Qualité des sources en fusion d'informations dans le cadre de la théorie des fonctions de croyance

Frédéric Pichon

UNIVERSITÉ D'ARTOIS

Mémoire scientifique

Pour l'obtention de

L'HABILITATION À DIRIGER DES RECHERCHES

UNIVERSITÉ D'ARTOIS

ÉCOLE DOCTORALE SCIENCES POUR L'INGÉNIEUR

LILLE-NORD DE FRANCE

DOMAINE DE RECHERCHE : Génie informatique et automatique

Présenté par

PICHON Frédéric

Qualité des sources en fusion d'informations dans le cadre de la théorie des fonctions de croyance.

Directeur de recherche : **LEFÈVRE Éric**

Soutenance le 7 novembre 2018
Devant la commission d'examen suivante

JURY

BERGE-CHERFAOUI Véronique	Professeur, Université de Technologie de Compiègne	Rapporteur
LE HÉGARAT-MASCLE Sylvie	Professeur, Université de Paris Sud	Rapporteur
PRADE Henri	Directeur de recherche, CNRS - IRIT	Rapporteur
ALLAOUI Hamid	Professeur, Université d'Artois	Examinateur
DENŒUX Thierry	Professeur, Université de Technologie de Compiègne	Président
DUBOIS Didier	Directeur de recherche, CNRS - IRIT	Examinateur
LEFÈVRE Éric	Professeur, Université d'Artois	Directeur

À Marie, Jeanne et Éloi.
À mes parents.

Table des matières

Première partie

Synthèse des travaux

Introduction

> *Quand deux témoins me disent une chose, il faut, pour que je me trompe en ajoutant foi à leur témoignage, que l'un & l'autre m'induisent en erreur ; si je suis sûr de l'un des deux, peu m'importe que l'autre soit croyable. Or la probabilité que l'un & l'autre me trompent, est une probabilité composée de deux probabilités, que le premier trompe, & que le second trompe. Celle du premier est 1/10 (puisque la probabilité que la chose est conforme à son rapport est 9/10) ; la probabilité que le second me trompe aussi, est encore 1/10 : donc la probabilité composée est la dixième d'une dixième ou 1/100 ; donc la probabilité du contraire, c'est-à-dire celle que l'un ou l'autre dit vrai, est 99/100.* Extrait de l'article "Probabilité", Encyclopédie de D'Alembert et Diderot [35], XVIII^{e} siècle.

Cadre général

Mes travaux de recherche depuis l'obtention de mon doctorat ont été réalisés en tant qu'ingénieur de recherche au Laboratoire de Mathématiques et Techniques de la Décision de Thales Research & Technology, puis en tant que maître de conférences au Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A) de l'Université d'Artois. Dans ces laboratoires, j'ai été membre des thèmes "Compréhension automatisée de la situation" et "Décision et fusion d'informations" respectivement. Une problématique commune à ces deux thèmes et dans laquelle la majorité de mes travaux de recherche s'intègrent est celle de la **fusion d'informations** ; problématique qui était déjà au cœur de mon travail de thèse. La fusion d'informations consiste en l'agrégation d'informations incertaines provenant de diverses sources (par exemple des capteurs ou des experts), afin d'extraire une connaissance véridique et aussi fine que possible à propos d'une entité d'intérêt. Ce problème, abordé dans différents formalismes et fondamentalement différent bien que partageant quelques similarités avec ceux de la révision d'informations et de l'agrégation de préférences [39], se retrouve dans de nombreux domaines tels que le traitement des images [66, 7], la robotique [83] et le renseignement [90]. Il est en fait aussi ancien que la théorie des probabilités : ses racines remontent au moins à la formalisation de la fiabilité des témoignages à la fin du XVIII^{e} siècle (voir par exemple l'article "Probabilité" dans l'Encyclopédie de D'Alembert et Diderot [35])

Les informations fournies par des sources ne peuvent être interprétées à moins de disposer de méta-connaissances sur ces sources, *i.e.*, de connaissances sur leur **qualité**. Typiquement, dans les modèles de fusion à ce jour, ces méta-connaissances reposent sur des hypothèses quant à la fiabilité des sources, où une source est dite fiable si elle est pertinente, c'est-à-dire si elle fournit des informations utiles sur l'entité d'intérêt ; par exemple, si la source est un capteur qui s'exprime au hasard car il est en panne, alors ce capteur n'est pas pertinent. En particulier, si une source

informe à propos de la valeur d'un paramètre $\boldsymbol{x}$ défini sur un domaine $\mathcal{X}$ sous la forme d'un témoignage $\boldsymbol{x} \in A \subset \mathcal{X}$ et que cette source est supposée pertinente avec une probabilité p, alors on considère que ce témoignage ne sert à rien avec la probabilité $1 - p$. Cela est connu, dans le cadre de la théorie des fonctions de croyance [14, 123, 131], en tant que l'affaiblissement d'une information [123, p. 251] [126], et l'état de connaissance résultant est représenté par une fonction à support simple [123] : le poids p est alloué au fait de savoir seulement que $\boldsymbol{x} \in A$ et le poids $1 - p$ est alloué à savoir que $\boldsymbol{x} \in \mathcal{X}$ (c'est la probabilité de ne rien savoir de la source). Si deux sources, avec des probabilités indépendantes p_1 et p_2 respectivement d'être pertinentes, donnent la même information $\boldsymbol{x} \in A$, alors il est justifié d'allouer la fiabilité $p_1 + p_2 - p_1 p_2$ à l'assertion $\boldsymbol{x} \in A$, puisque l'on doit déduire cela dès qu'au moins une des sources est pertinente ; cela est le résultat obtenu par la règle de combinaison de Dempster [14, 123] (résultat déjà connu dans l'Encyclopédie [35]).

Comme l'on peut s'en rendre compte avec la discussion précédente, la question de la fusion d'informations et de la prise en compte associée de la qualité des sources, est intimement liée à la modélisation de l'incertitude. Les informations fournies ainsi que la connaissance sur la qualité des sources sont en effet généralement incertaines et par conséquent le résultat de leur fusion l'est généralement également, et ce d'autant plus que quand bien même les informations seraient précises elles peuvent être incohérentes ce qui induit également de l'incertitude. Au-delà de la théorie des probabilités, il existe tout un panel de théories pour la modélisation de l'incertitude, avec des liens entre elles de mieux en mieux compris et permettant de pallier certaines difficultés, dues au recours à une mesure de probabilité unique, de la théorie des probabilités [44]. La théorie des fonctions de croyance a déjà été mentionnée et on peut ajouter deux autres formalismes également très importants que sont la théorie des possibilités [41] et la théorie des probabilités imprécises [141], dont les représentations de l'incertitude sont formellement moins et plus générales, respectivement, que celle de la théorie des fonctions de croyance [1]. Remarquons que ces théories peuvent être utilisées avantageusement de manière complémentaire [2] comme illustré, par exemple, dans [4].

La **théorie des fonctions de croyance** offre un compromis entre flexibilité et complexité qui semble intéressant en pratique, comme en témoigne son utilisation dans de nombreux problèmes tels que, récemment, la classification automatique d'un grand nombre de données [26], les classifications multi-labels [27] et multi-classes [84], l'association [28], et la classification basée sur des règles [71] ; le lecteur est renvoyé à [21] et [79, p.12] pour un bien plus large éventail d'applications de cette théorie. En outre, relevons qu'une raison sinon principale, du moins fréquente d'usage de cette théorie est sa pertinence pour la fusion d'informations. Cette problématique est en effet au cœur de cette théorie depuis ses origines et ses solutions

1. Brièvement, une mesure de possibilité (rigoureusement, de nécessité) est une fonction de croyance dont les ensembles focaux sont imbriqués, et une fonction de croyance est une mesure de probabilité inférieure dont la transformée de Möbius n'admet pas de valeurs négatives.

2. Elles peuvent également être mises en compétition [52].

dans ce cadre théorique n'ont cessé de s'étoffer depuis. En fait, comme le suggère la discussion précédente sur la fiabilité des témoignages ainsi qu'il le sera développé dans ce mémoire, la théorie des fonctions de croyance est particulièrement adaptée à la question fondamentale qu'est le traitement de la qualité des sources en fusion d'informations.

Thèmes de recherches

Mes travaux sur la **qualité des sources en fusion d'informations dans le cadre de la théorie des fonctions de croyance** se sont articulés selon trois axes principaux. Le premier concerne la formalisation et la prise en compte de connaissances sur la qualité des sources d'information. J'ai proposé **une approche générale pour la fusion de fonctions de croyance**, permettant de considérer d'autres facettes de la qualité des sources que leur pertinence. En particulier, le cas de sources pouvant manquer partiellement de sincérité a été investigué en profondeur. Cette approche permet de traiter également le cas où il s'agit d'interpréter l'information fournit par une seule source en fonction de sa qualité, problème connu sous l'appellation de *correction* [94] d'une fonction de croyance. Un des résultats de ce travail est que l'affaiblissement de Shafer et la règle de combinaison non normalisée de Dempster (aussi appelée règle conjonctive [130]), qui concernent seulement la pertinence des sources, se trouvent considérablement étendus. Cette dernière règle est notamment généralisée à tous les connecteurs logiques. L'approche proposée englobe également d'autres mécanismes de correction et de fusion importants, tels que l'affaiblissement contextuel [95] et les α-jonctions de Smets [128]. En outre, j'ai mené un travail spécifique sur des mécanismes de corrections dites contextuelles que l'on peut obtenir avec cette approche. Enfin, j'ai également montré que lorsque les comportements des sources sont indépendants, le résultat de la fusion de leur témoignage s'obtient simplement en combinant par la règle conjonctive le témoignage corrigé de chaque source.

La seconde voie que j'ai explorée concerne une question présente dès le livre fondateur de Shafer [123]. S'il l'on considère, comme Shafer [123], la théorie des fonctions de croyance essentiellement comme une approche de fusion de témoignages élémentaires partiellement fiables, chacun étant représenté par une fonction à support simple, alors il semble naturel que toute fonction de croyance puisse être décomposée en termes de telles fonctions. Une telle décomposition révélerait alors les composants élémentaires sous-jacents à un état de connaissance complexe représenté par une fonction de croyance. Dans ce sens, Shafer [123] a montré qu'une classe de fonctions de croyance (celles dites séparables) pouvaient se décomposer comme la combinaison par la règle de Dempster de fonctions à supports simples. De plus il a montré qu'il était possible, moyennant le recours à la notion de grossissement [123], d'exprimer toute fonction de croyance non séparable en terme de telles fonctions élémentaires. Cependant, comme défendu par Smets [127], la solution de Shafer pour les fonctions de croyance non séparables peut être jugée non satisfaisante au regard

du fait qu'elle rend l'utilisation de la règle de Dempster sinon injustifiée, du moins hasardeuse (l'argument repose sur le fait que la règle de Dempster ne commute pas avec le grossissement). Smets [127] a proposé une autre solution pour décomposer ces dernières fonctions de croyance. Précisément, il a montré que toute fonction de croyance peut se décomposer de manière unique comme la combinaison par la règle conjonctive de fonctions à supports simples *généralisées* (fonctions où les masses peuvent être négatives). Sa solution a eu un succès certain, en particulier elle a été à la base de propositions pour traiter les questions de la fusion [18, 74, 106], de la correction [95, 96, 113] et du regroupement [118] de fonctions de croyance. Cela montre que savoir décomposer une fonction de croyance a, au-delà de la satisfaction intellectuelle d'une théorie bien cohérente, un intérêt fondamental aux conséquences potentiellement importantes. Toutefois, malgré son succès et comme la solution de Shafer, la solution de Smets n'est pas exempte de tout reproche. L'interprétation donnée par Smets à ses fonctions à supports simples généralisées, bien qu'élégante, ainsi que leur existence posent en effet quelques questions et mériteraient de plus amples justifications. Aussi ai-je mis à jour une autre **décomposition unique de toute fonction de croyance** reposant seulement sur des concepts bien définis (et n'utilisant pas la notion de grossissement). Ma proposition est dans le même esprit que celles de Shafer et de Smets – une fonction de croyance est vue comme résultant de témoignages élémentaires partiellement fiables –, mais elle suit une approche complètement différente basée sur une utilisation conjointe de l'approche générale pour la fusion évoquée dans la paragraphe précédent et de résultats de Teugels concernant la représentation de la distribution de Bernoulli multivariée [137]. Brièvement, j'ai montré que toute fonction de croyance peut se décomposer en témoignages élémentaires partiellement fiables ayant une structure de dépendance ou, de manière analogue, en une combinaison conjonctive de fonctions à supports simples ayant une structure de dépendance. Enfin, je suis également revenu sur les résultats de Smets [127] et ai montré qu'au lieu d'interpréter, non sans interrogation, en termes de décomposition d'une fonction de croyance la fonction dite de poids qu'il a mise à jour, il est possible de donner à cette fonction une sémantique totalement différente et bien définie en termes de mesures d'information [51].

L'approche générale pour la correction et la fusion de fonctions de croyance que j'ai proposée est théorique en ce qu'elle n'inclut pas de moyens pratiques pour l'appliquer et en particulier de **moyens pour obtenir les méta-connaissances** sur les sources qu'elle nécessite. Mon troisième axe de travail a consisté à proposer de tels moyens. Plus précisément, je me suis intéressé à deux situations que l'on peut rencontrer. La première est celle où l'on dispose d'une certaine expérience préalable des sources à traiter. Cette expérience peut prendre la forme de données mettant en regard ce que les sources ont dit et les vraies réponses qui étaient attendues dans des situations passées, ou encore de connaissances expertes montrant comment qualifier des sources en fonction d'attributs les décrivant. Dans le cas où des données sont disponibles, deux approches ont été étudiées : une première inspirée de [97] permettant de déterminer efficacement la sincérité contextuelle d'une source par minimisation d'un critère d'erreur ; une autre prolongeant les travaux menés dans [49, 143] et per-

mettant de prédire la sincérité de sources à l'aide de résultats récents en inférence statistique et prédiction dans la cadre de la théorie des fonctions de croyance [20, 75]. Cette seconde approche a été développée en grande partie, et appliquée au problème de la détection de visages sur des vidéos, dans le cadre de la thèse de Pauline Minary. Le cas où on l'on dispose de connaissances expertes sur les sources se prête bien à des approches multi-critères et nous en avons proposé une basée sur l'intégrale de Choquet [56] permettant d'établir la pertinence d'une source. La seconde situation que l'on peut rencontrer est celle où l'on a peu, voire aucune, expérience préalable avec les sources. Ce cas est traité classiquement (voir, par exemple, [130, 42]) dans la théorie des fonctions de croyance en faisant d'abord une hypothèse forte sur la qualité des sources, telle que supposer que toutes les sources sont fiables, puis en relâchant progressivement cette hypothèse si elle induit trop de conflit [30, 91]. Il a été possible de proposer une formalisation générique de cette démarche itérative, où il s'agit de trouver une hypothèse sur la qualité des sources induisant un bon compromis entre spécificité et cohérence du résultat de la fusion, cette hypothèse étant prélevée dans un ensemble ordonné d'hypothèses au regard de leurs cohérence et spécificité. De plus, de tels ensembles ordonnés ont été mis à jour, rendant la démarche opérationnelle.

En dehors de ces trois axes de travaux directement liés à la qualité des sources en fusion d'informations, j'ai mené depuis mon doctorat quelques travaux sur d'autres problématiques, avec pour dénominateur commun une gestion fine des incertitudes, le plus souvent à l'aide de la théorie des fonctions de croyance. Parmi ceux-ci, le plus conséquent a été conduit ces dernières années. En effet, de manière quasi concomitante à mon arrivée au sein du LGI2A en septembre 2013, et en partie suite aux recommandations issues de la campagne d'évaluation 2013-2014 de l'AERES (visite de l'unité le 13 novembre 2013), il a été décidé par le laboratoire de mettre des moyens significatifs pour développer plus avant les collaborations entre ses thèmes, et en particulier de financer un projet de thèse bi-thèmes. Dans un souci d'intégration mais également d'ouverture, je me suis impliqué dans la proposition d'un tel projet, qui a été retenue et qui mettait en jeu les thèmes "Optimisation des systèmes logistiques" et "Décision et fusion d'informations". Précisément il s'agissait d'étudier le **problème de tournées de véhicules avec contrainte de capacité et demandes incertaines**, lorsque les incertitudes sont modélisées dans le cadre de la théorie des fonctions de croyance puisque, par exemple, les informations sur les demandes des clients peuvent être issues de sources partiellement fiables. Le défi scientifique de ce projet était double : permettre plus de flexibilité dans la representation des incertitudes qu'avec les approches existantes sans que la complexité induite ne devienne rédhibitoire, et créer un pont entre les deux approches classiques pour ce problème que sont l'optimisation stochastique et l'optimisation robuste. Les contributions apportées dans le cadre de la thèse de Nathalie Helal sur ce sujet ont été les suivantes. Deux modèles pour ce problème ont été proposés, étendant deux approches classiques en programmation stochastique que sont les programmations à base de contraintes en probabilité et à base de recours. De plus, diverses propriétés de ces modèles ont été mises en évidence. En particulier, ils se dégénèrent bien en

les modèles stochastiques lorsque les demandes sont stochastiques, mais également en l'approche ensembliste pessimiste pure lorsque les demandes sont connues sous forme d'intervalles. Ils présentent en sus des propriétés intéressantes de monotonie entre leurs paramètres et le coût de la solution optimale. Des fonctions de croyance particulières rendant gérables les calculs associés à ces modèles, ont par ailleurs été identifiées. Enfin, une méthode de résolution basée sur une métaheuristique a été développée pour chacun des deux modèles, et leur bon comportement a été validé expérimentalement sur des instances réalistes.

Organisation

Ce mémoire scientifique présentant une synthèse de mes principaux travaux de recherche depuis l'obtention de mon doctorat est organisé en cinq chapitres. Dans le chapitre 1, les notions sur les fonctions de croyance, qui seront nécessaires pour exposer dans les chapitres suivants mes travaux, sont rappelées brièvement. Mes contributions associées aux trois axes évoqués dans la section précédente sont abordées dans les trois chapitres suivants : l'approche générale pour la correction et fusion de fonctions de croyance est décrite dans le chapitre 2 ; la décomposition des fonctions de croyance est présentée dans le chapitre 3 ; les moyens pour obtenir les connaissances sur la qualité des sources sont discutés dans le chapitre 4. Dans le chapitre 5, l'étude du problème de tournées de véhicules avec contrainte de capacité et demandes incertaines représentées par des fonctions de croyance est synthétisée. Enfin, les différents thèmes dans lesquels j'envisage de poursuivre mes travaux au cours des prochaines années sont abordés dans le chapitre 6.

Chapitre 1

Rappels sur la théorie des fonctions de croyance

1.1 Introduction

La théorie des fonctions de croyance trouve ses origines dans les travaux de Dempster [14] situés dans le contexte de l'inférence statistique. Elle a été établie par Shafer [123] essentiellement comme une théorie mathématique des témoignages. Elle a ensuite été étoffée et utilisée par de nombreux chercheurs [21], le plus prolifique étant sans doute Smets qui a entre autres proposé une justification axiomatique cohérente de ses principaux concepts [131]. Elle est maintenant bien identifiée comme une des alternatives importantes à la théorie des probabilités pour la modélisation des incertitudes [44] et une communauté internationale, la *Belief Functions and Applications Society*[1], a même vu le jour en 2010 afin de promouvoir l'enseignement, la recherche et l'application de cette théorie.

Dans ce chapitre, les notions de cette théorie utilisées dans le cadre de mes travaux sont rappelées.

1.2 Représentation

Pour la représentation de l'incertitude d'un agent vis-à-vis de la vraie valeur d'un paramètre $\boldsymbol{x}$ à valeurs dans un ensemble fini $\mathcal{X} = \{x_1, \ldots, x_n\}$ appelé cadre de discernement, le concept fondamental est celui de fonction de masse, définie comme une fonction $m^{\mathcal{X}}$ de l'ensemble des parties de $\mathcal{X}$ dans $[0, 1]$, *i.e.*, $m^{\mathcal{X}} : 2^{\mathcal{X}} \to [0, 1]$, vérifiant :

$$\sum_{A \subseteq \mathcal{X}} m^{\mathcal{X}}(A) = 1. \tag{1.1}$$

La masse $m^{\mathcal{X}}(A)$ s'interprète comme la probabilité que $\boldsymbol{x} \in A$ représente fidèlement la connaissance disponible à propos de la vraie valeur de $\boldsymbol{x}$ [37, 3], ou pour faire plus court, c'est la probabilité de ne savoir que $\boldsymbol{x} \in A$ [12, 45]. L'ensemble des fonctions de masse sur $\mathcal{X}$ est noté $\mathcal{M}^{\mathcal{X}}$.

1. http://bfas.iut-lannion.fr

La définition de $m^{\mathcal{X}}$ permet que $m^{\mathcal{X}}(\emptyset) > 0$, auquel cas cela traduit une incohérence dans la connaissance représentée par m [33]. Cela peut arriver en particulier lorsque $m^{\mathcal{X}}$ est le résultat de la fusion d'informations [130, 39].

Tout ensemble $A \subseteq \mathcal{X}$ tel que $m^{\mathcal{X}}(A) > 0$ est appelé ensemble focal de $m^{\mathcal{X}}$. Une fonction de masse $m^{\mathcal{X}}$ est dite : dogmatique si $\mathcal{X}$ n'est pas un ensemble focal ; normalisée si $\emptyset$ n'est pas un ensemble focal ; catégorique si elle a un seul ensemble focal ; vide si $\mathcal{X}$ est son seul ensemble focal ; à support simple ou, pour faire plus court, simple si elle a au plus deux ensembles focaux, et si elle en a deux, $\mathcal{X}$ est un de ceux-là ; Bayésienne si ses ensembles focaux sont des singletons ; consonante si ses ensembles focaux sont imbriqués. De plus, la négation $\overline{m}^{\mathcal{X}}$ d'une fonction de masse $m^{\mathcal{X}}$ est définie par [40] : $\overline{m}^{\mathcal{X}}(A) = m^{\mathcal{X}}(\overline{A})$, $\forall A \subseteq \mathcal{X}$, où $\overline{A}$ est le complémentaire de A dans $\mathcal{X}$. Par ailleurs, une fonction de masse $m^{\mathcal{X}}$ non normalisée peut être transformée en une fonction de masse normalisée $m^{\mathcal{X}}_*$, par l'opération de normalisation suivante :

$$m^{\mathcal{X}}_*(A) = \begin{cases} \kappa \cdot m^{\mathcal{X}}(A) & \text{pour tout } A \subseteq \mathcal{X}, A \neq \emptyset, \\ 0 & \text{si } A = \emptyset, \end{cases} \tag{1.2}$$

avec $\kappa = (1 - m^{\mathcal{X}}(\emptyset))^{-1}$.

À partir d'une fonction de masse $m^{\mathcal{X}}$, on obtient une fonction de croyance $Bel^{\mathcal{X}}$, une fonction de plausibilité $Pl^{\mathcal{X}}$ et une fonction de communalité $q^{\mathcal{X}}$ définies, pour tout $A \subseteq \mathcal{X}$, par :

$$Bel^{\mathcal{X}}(A) = \sum_{\emptyset \neq B \subseteq A} m^{\mathcal{X}}(B), \tag{1.3}$$

$$Pl^{\mathcal{X}}(A) = \sum_{B \cap A \neq \emptyset} m^{\mathcal{X}}(B), \tag{1.4}$$

$$q^{\mathcal{X}}(A) = \sum_{B \supseteq A} m^{\mathcal{X}}(B). \tag{1.5}$$

Le degré de croyance $Bel^{\mathcal{X}}(A)$ évalue à quel point $\boldsymbol{x} \in A$ est impliqué par la connaissance disponible et $Pl^{\mathcal{X}}(A)$ évalue à quel degré $\boldsymbol{x} \in A$ est cohérent avec elle [46] ; la fonction de communalité joue quant à elle plus un rôle technique qui sera couvert plus tard. Ces fonctions sont des représentations équivalentes de la même information, la donnée de l'une d'entre elles permettant de retrouver les autres. On construit également la fonction de contour $pl^{\mathcal{X}}$ comme la restriction de $Pl^{\mathcal{X}}$ aux singletons : $pl^{\mathcal{X}}(x) = Pl^{\mathcal{X}}(\{x\})$, pour tout $x \in \mathcal{X}$.

Si une fonction de masse $m^{\mathcal{X}}$ est Bayésienne, alors $Bel^{\mathcal{X}}(A) = Pl^{\mathcal{X}}(A)$ pour tout $A \subseteq \mathcal{X}$ et $Bel^{\mathcal{X}}(A \cup B) = Bel^{\mathcal{X}}(A) + Bel^{X}(B)$ pour tout $A, B \subseteq \mathcal{X}$ tels que $A \cap B = \emptyset$, et cette fonction est une mesure de probabilité. Si $m^{\mathcal{X}}$ est consonante, alors $Pl^{\mathcal{X}}(A \cup B) = \max(Pl^{\mathcal{X}}(A), Pl^{X}(B))$, pour tout $A, B \subseteq \mathcal{X}$, et cette fonction est une mesure de possibilité.

Le calcul matriciel est utile pour simplifier les mathématiques de la théorie des fonctions de croyance [129]. Une fonction de masse $m^{\mathcal{X}}$ et ses fonctions associées,

telle que $q^{\mathcal{X}}$, peuvent être vues comme des vecteurs colonnes de taille $2^{|\mathcal{X}|} = 2^n$, dont les éléments sont ordonnés selon l'ordre dit binaire détaillé ci-après. Soit k un nombre entier tel que $1 \leq k \leq 2^n$. k peut être écrit sous forme d'une expansion binaire, *i.e.*,

$$k = 1 + \sum_{i=1}^{n} k_i 2^{i-1}, \tag{1.6}$$

où $k_i \in \{0, 1\}$. L'expansion (1.6) induit une correspondance biunivoque $k \leftrightarrow (k_1, \ldots, k_n)$, c'est-à-dire que l'équation (1.6) associe à chaque entier k, $1 \leq k \leq 2^n$, un vecteur binaire $(k_1, \ldots, k_n) \in \{0, 1\}^n$. Soit A_k, $1 \leq k \leq 2^n$, le sous-ensemble de $\mathcal{X}$, tel que $x_i \in A_k$ si $k_i = 1$ et $x_i \notin A_k$ si $k_i = 0$, avec k_i, $i = 1, \ldots, n$, les termes dans l'expansion binaire (1.6) de k. Dans l'ordre binaire, le k^{e} élément du vecteur $\mathbf{m}^{\mathcal{X}}$ correspond à l'ensemble A_k. Ainsi $\emptyset$ est le premier élément, $\{x_1\}$ le second, $\{x_2\}$ le troisième, $\{x_1, x_2\}$ le quatrième, *etc.* Par exemple, pour $n = 4$, $A_{14} = \{x_1, x_3, x_4\}$ puisque $k = 14$ est en correspondance biunivoque avec $(k_1 = 1, k_2 = 0, k_3 = 1, k_4 = 1)$. L'équation (1.5) devient sous forme matriciel :

$$\mathbf{q}^{\mathcal{X}} = \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right) \mathbf{m}^{\mathcal{X}}, \tag{1.7}$$

où $\otimes$ est le produit de Kronecker.

1.3 Comparaison

Le principe d'engagement minimum [126] dicte qu'étant donné un ensemble de fonctions de masse compatibles avec un ensemble de contraintes données, il faut choisir la fonction de masse la moins informative. Pour que ce principe soit opérationnel, il faut être capable de comparer le contenu informationnel de fonctions de masse. La notion de spécialisation [40], qui étend la comparaison naturelle du contenu informationnel d'ensembles ($\boldsymbol{x} \in A$ est plus spécifique que $\boldsymbol{x} \in B$ si $A \subset B$), permet cela. Elle définit la comparaison informationnelle de fonctions de masse de la manière suivante : une fonction de masse $m_1^{\mathcal{X}}$ est au moins autant informative (ou spécifique) qu'une autre fonction de masse $m_2^{\mathcal{X}}$, ce qui est noté $m_1^{\mathcal{X}} \sqsubseteq m_2^{\mathcal{X}}$, si et seulement si il existe une matrice stochastique (gauche) $S = [S(A, B)]$, $A, B \in 2^{\mathcal{X}}$, vérifiant

$$S(A, B) > 0 \Rightarrow A \subseteq B, \quad A, B \subseteq \mathcal{X}, \tag{1.8}$$

$$m_1^{\mathcal{X}}(A) = \sum_{B \subseteq \mathcal{X}} S(A, B) m_2^{\mathcal{X}}(B), \quad \forall A \subseteq \mathcal{X}. \tag{1.9}$$

Le terme $S(A, B)$ peut être vu comme la proportion de la masse $m_2^{\mathcal{X}}(B)$ qui est transférée de l'ensemble B vers l'ensemble au moins autant informatif A.

1.4 Combinaison

Lorsqu'un agent reçoit de plusieurs sources des informations à propos de $\boldsymbol{x}$ représentées sous forme de fonctions de croyance, il peut construire sa connaissance $m^{\mathcal{X}}$ à propos de $\boldsymbol{x}$ en fusionnant ces fonctions de croyance à l'aide d'une règle de combinaison.

Soit $m_1^{\mathcal{X}}$ et $m_2^{\mathcal{X}}$ deux fonctions de masse fournies à propos de $\boldsymbol{x}$ par deux sources $\mathfrak{s}_1$ et $\mathfrak{s}_2$ respectivement. La règle de combinaison conjonctive fusionne $m_1^{\mathcal{X}}$ et $m_2^{\mathcal{X}}$ pour donner $m^{\mathcal{X}}$ par l'équation suivante :

$$m^{\mathcal{X}}(A) \quad = \quad \sum_{B \cap C = A} m_1^{\mathcal{X}}(B) m_2^{\mathcal{X}}(C), \quad \forall A \subseteq \mathcal{X}. \tag{1.10}$$

Il est utile de noter cette règle par le symbole $\textcircled{\cap}$: le résultat de la combinaison de deux fonctions de masse $m_1^{\mathcal{X}}$ et $m_2^{\mathcal{X}}$ par la règle conjonctive s'écrit alors $m_{1\textcircled{\cap}2}^{\mathcal{X}} = m_1^{\mathcal{X}} \textcircled{\cap} m_2^{\mathcal{X}}$. Cette règle est : commutative ($m_1^{\mathcal{X}} \textcircled{\cap} m_2^{\mathcal{X}} = m_2^{\mathcal{X}} \textcircled{\cap} m_1^{\mathcal{X}}$, pour tout $m_1^{\mathcal{X}}, m_2^{\mathcal{X}} \in \mathcal{M}^{\mathcal{X}}$), associative ($(m_1^{\mathcal{X}} \textcircled{\cap} m_2^{\mathcal{X}}) \textcircled{\cap} m_3^{\mathcal{X}} = m_1^{\mathcal{X}} \textcircled{\cap} (m_2^{\mathcal{X}} \textcircled{\cap} m_3^{\mathcal{X}})$, pour tout $m_1^{\mathcal{X}}, m_2^{\mathcal{X}}, m_3^{\mathcal{X}} \in \mathcal{M}^{\mathcal{X}}$) et admet la fonction de masse vide pour unique élément neutre ($m_1^{\mathcal{X}} \textcircled{\cap} m_2^{\mathcal{X}} = m_1^{\mathcal{X}}$ si et seulement si $m_2^{\mathcal{X}}(\mathcal{X}) = 1$). De plus, elle vérifie

$$q_{1\textcircled{\cap}2}^{\mathcal{X}}(A) = q_1^{\mathcal{X}}(A) \cdot q_2^{\mathcal{X}}(A), \quad \forall A \subseteq \mathcal{X}, \tag{1.11}$$

avec $q_1^{\mathcal{X}}$, $q_2^{\mathcal{X}}$ et $q_{1\textcircled{\cap}2}^{\mathcal{X}}$ les fonctions de communalités associées respectivement à $m_1^{\mathcal{X}}, m_2^{\mathcal{X}}$ et $m_{1\textcircled{\cap}2}^{\mathcal{X}}$.

L'utilisation de cette règle suppose que les sources $\mathfrak{s}_1$ et $\mathfrak{s}_2$ soient considérées toutes les deux fiables et indépendantes [130]. Ces hypothèses, sur lesquelles le chapitre suivant reviendra, sont souvent faites en pratique, cette règle étant en effet prévalente dans la littérature.

Lorsqu'il est choisi d'utiliser cette règle pour combiner deux fonctions de masse $m_1^{\mathcal{X}}$ et $m_2^{\mathcal{X}}$, si l'on obtient $m_{1\textcircled{\cap}2}^{\mathcal{X}}(\emptyset) > 0$, alors deux attitudes sont généralement suivies. Il est possible, comme suggéré par Shafer [123], de normaliser $m_{1\textcircled{\cap}2}$ par l'équation (1.2)– on aura alors en fait combiné $m_1^{\mathcal{X}}$ et $m_2^{\mathcal{X}}$ par la règle de Dempster –, ce qui revient à considérer le conflit (nom donné à $m_{1\textcircled{\cap}2}^{\mathcal{X}}(\emptyset)$ [130]) comme une sorte de bruit dont il faut se défaire [32]. L'autre point de vue consiste à considérer la valeur positive de $m_{1\textcircled{\cap}2}^{\mathcal{X}}(\emptyset)$ comme une indication que les hypothèses associées à l'utilisation de la règle conjonctive ne sont pas valables, et cela d'autant plus que cette valeur est élevée. Des hypothèses moins fortes sont alors généralement faites. Par exemple, il peut être supposé seulement qu'au moins une des sources est fiable, auquel cas il convient d'utiliser la règle disjonctive [40, 126], notée $\textcircled{\cup}$, et dont la définition est la même que (1.10) à la différence que l'intersection $\cap$ est remplacée par l'union $\cup$.

Concernant les règles conjonctive et disjonctive, il est utile pour la suite de faire deux remarques. Premièrement, ces deux règles sont des cas particuliers des α-conjonctions et α-disjonctions respectivement. Ces deux familles de règles, issues d'un travail formel de Smets [128], représentent essentiellement l'ensemble des

règles de combinaison commutatives, associatives et linéaires admettant un élément neutre (la fonction de masse vide pour les α-conjonctions et sa négation pour les α-disjonctions). Smets a montré que chacune de ces familles dépend d'un paramètre $\alpha \in [0,1]$. En particulier, les α-conjonctions varient entre deux règles basées sur des opérateurs Booléens lorsque α décroit (un comportement similaire est observé pour les α-disjonctions) : la règle conjonctive (pour $\alpha = 1$) et la règle équivalence [104] (pour $\alpha = 0$) dont la définition est la même que (1.10) à la différence que l'intersection $\cap$ est remplacée par $\underline{\cap}$ correspondant à l'équivalence logique, *i.e.*, $B \underline{\cap} C = (B \cap C) \cup (\overline{B} \cap \overline{C})$ pour tout $B, C \subseteq \mathcal{X}$. Smets n'a toutefois pas fourni d'interprétation claire pour les cas $\alpha \in (0,1)$.

Deuxièmement, les inverses de ces deux règles, notées respectivement $\textcircled{\not\cap}$ et $\textcircled{\not\cup}$, peuvent être définies [127, 18]. L'inverse de la règle conjonctive restaure $m_1^{\mathcal{X}}$ de $m_{1\textcircled{\cap}2}^{\mathcal{X}}$ et $m_2^{\mathcal{X}}$, *i.e.*, $m_{1\textcircled{\cap}2}^{\mathcal{X}} \textcircled{\not\cap} m_2^{\mathcal{X}} = m_1^{\mathcal{X}}$. Elle est utile si le témoignage $m_2^{\mathcal{X}}$ est a posteriori remis en cause et devrait donc être enlevé de $m_{1\textcircled{\cap}2}^{\mathcal{X}}$. Soit $q_1^{\mathcal{X}}$ et $q_2^{\mathcal{X}}$ les fonctions de communalité associées respectivement à deux fonctions de masse $m_1^{\mathcal{X}}$ et $m_2^{\mathcal{X}}$. L'inverse de la règle conjonctive est définie par :

$$q_{1\textcircled{\not\cap}2}^{\mathcal{X}}(A) \quad = \quad \frac{q_1^{\mathcal{X}}(A)}{q_2^{\mathcal{X}}(A)}, \quad \forall A \subseteq \mathcal{X}. \tag{1.12}$$

Cette opération est bien définie tant que (i) $m_{1\textcircled{\not\cap}2}^{\mathcal{X}}$ est une fonction de masse, ce qui n'est pas nécessairement le cas puisque le quotient de deux fonctions de communalité n'est pas toujours une fonction de communalité, et (ii) $m_2^{\mathcal{X}}$ n'est pas dogmatique, ce qui garantit $q_2^{\mathcal{X}}(A) > 0$ pour tout $A \subseteq \mathcal{X}$. L'inverse de la règle disjonctive est définie de la même manière à partir de la fonction d'implicabilité $b^{\mathcal{X}}$ définie par

$$b^{\mathcal{X}}(A) = bel^{\mathcal{X}}(A) + m^{\mathcal{X}}(\emptyset), \quad \forall A \subseteq \mathcal{X}, \tag{1.13}$$

et dont le comportement est similaire à (1.11) pour la règle disjonctive.

1.5 Décomposition

Certaines fonctions de masse peuvent s'obtenir comme le résultat de la combinaison par la règle conjonctive de fonctions de masse simples ; elles sont appelées séparables dans ce mémoire [2]. Une fonction de masse simple ayant deux ensembles focaux $A \subset \mathcal{X}$ et $\mathcal{X}$, avec les masses respectives $1-w$ et w, $w \in [0,1]$, peut être simplement notée A^w. Toute fonction de masse séparable $m^{\mathcal{X}}$ peut alors s'exprimer sous la forme

$$m^{\mathcal{X}} \quad = \quad \textcircled{\cap}_{A \subset \mathcal{X}} A^{w(A)}, \tag{1.14}$$

2. Dans [123], le terme séparable désigne les fonctions de masse issues de la combinaison par la règle de Dempster de fonctions de masses simples, et celles issues de la combinaison par la règle conjonctive sont appelées u-séparables dans [18].

avec $w(A) \in [0,1]$ pour tout $A \subset \mathcal{X}$. Cette décomposition est unique si $m^{\mathcal{X}}$ est non dogmatique (auquel cas $w(A) > 0$ pour tout $A \subset \mathcal{X}$).

Smets [127] a étendu cette décomposition à toutes les fonctions de masse non dogmatiques. Pour y parvenir, il a introduit le concept de fonction de masse simple généralisée définie comme une fonction $\mu : 2^{\mathcal{X}} \to \mathbb{R}$ vérifiant :

$$\begin{aligned} \mu(A) &= 1 - w, \\ \mu(\mathcal{X}) &= w, \\ \mu(B) &= 0, \quad \forall B \in 2^{\mathcal{X}} \backslash \{A, \mathcal{X}\}, \end{aligned}$$

pour un $A \neq \mathcal{X}$ et un $w \in [0, +\infty)$. Toute fonction de masse simple généralisée μ peut être notée A^w avec $A \neq \mathcal{X}$ et $w \in [0, +\infty)$. Lorsque $w \leq 1$, A^w est une fonction de masse simple. Lorsque $w > 1$, A^w est appelée fonction de masse simple inverse ; Smets [127] a proposé d'interpréter une telle fonction comme des raisons de *ne pas* croire dans A (il parle aussi de "dette de croyance" dans A) puisque combiner $A^{1/w}$ avec A^w par $\textcircled{\cap}$ (en utilisant une extension triviale de la règle conjonctive pour combiner des fonctions de masse simples généralisées) produit la fonction de masse vide, c'est-à-dire que les raisons de croire en A représentée par $A^{1/w}$ sont contre-balancées par A^w.

En utilisant ce nouveau concept, Smets [127] a montré que pour toute fonction de croyance non dogmatique $m^{\mathcal{X}}$, nous avons :

$$m^{\mathcal{X}} = \textcircled{\cap}_{A \subset \mathcal{X}} A^{w(A)}, \tag{1.15}$$

avec $w(A) \in (0, +\infty)$ pour tout $A \subset \mathcal{X}$. La fonction dite de poids $w : 2^{\mathcal{X}} \backslash \{\mathcal{X}\} \to (0, +\infty)$ qui apparaît dans (1.15), est une représentation équivalente d'une fonction de masse non dogmatique. Elle peut être obtenue de la fonction de communalité comme suit :

$$w(A) = \prod_{B \supseteq A} q^{\mathcal{X}}(B)^{(-1)^{|B|-|A|+1}}, \quad \forall A \subset \mathcal{X}. \tag{1.16}$$

La règle conjonctive admet une expression simple en utilisant la fonction de poids :

$$w_{1\textcircled{\cap}2}(A) = w_1(A) \cdot w_2(A), \quad \forall A \subset \mathcal{X}. \tag{1.17}$$

De plus, toute fonction de masse non dogmatique peut s'écrire sous la forme [127] :

$$m^{\mathcal{X}} = m_c^{\mathcal{X}} \textcircled{\not\cap} m_d^{\mathcal{X}}, \tag{1.18}$$

avec $m_c^{\mathcal{X}}$ et $m_d^{\mathcal{X}}$ deux fonctions de masse séparables telles que

$$\begin{aligned} m_c^{\mathcal{X}} &= \textcircled{\cap}_{A \subset \mathcal{X}} A^{\min(1, w(A))}, \\ m_d^{\mathcal{X}} &= \textcircled{\cap}_{A \subset \mathcal{X}} A^{\min(1, \frac{1}{w(A)})}, . \end{aligned}$$

En d'autres termes, Smets [127] a montré avec (1.18) que toute fonction de masse non dogmatique peut être obtenue uniquement de fonctions de masse simples. Les fonctions de masse $m_c^{\mathcal{X}}$ et $m_d^{\mathcal{X}}$ dans (1.18) sont appelées les composants de confiance et défiance respectivement de $m^{\mathcal{X}}$ par Smets [127], qui a proposé d'interpréter $m_c^{\mathcal{X}}$ comme représentant "des bonnes raisons de croire" dans des propositions $A \subset \mathcal{X}$, et $m_d^{\mathcal{X}}$ comme représentant "des bonnes raisons de ne pas croire" dans d'autres propositions. Sa décomposition est illustrée par l'exemple 1.1.

Exemple 1.1 ("Example 2, continuation" de [127]) *Soit $\mathcal{X} = \{x_1, x_2, x_3\}$ le domaine d'un paramètre $\boldsymbol{x}$ et $m^{\mathcal{X}}$ la fonction de masse telle que*

$$m^{\mathcal{X}}(\{x_1, x_2\}) = m^{\mathcal{X}}(\{x_1, x_3\}) = m^{\mathcal{X}}(\{x_1, x_2, x_3\}) = 1/3. \tag{1.19}$$

Nous avons

$$\begin{aligned} w(\{x_1, x_2\}) &= 1/2, \\ w(\{x_1, x_3\}) &= 1/2, \\ w(\{x_1\}) &= 4/3, \end{aligned}$$

et $w(A) = 1$ pour tout $A \in 2^{\mathcal{X}} \backslash \{\mathcal{X}, \{x_1, x_2\}, \{x_1, x_3\}, \{x_1\}\}$. Ainsi, à partir de (1.15) et le fait que la fonction de masse vide, qui peut être notée A^1 pour n'importe quel $A \subset \mathcal{X}$, est l'élément neutre de $\textcircled{\cap}$, il vient

$$m^{\mathcal{X}} = \{x_1, x_2\}^{1/2} \textcircled{\cap} \{x_1, x_3\}^{1/2} \textcircled{\cap} \{x_1\}^{4/3},$$

et en utilisant (1.18)

$$m^{\mathcal{X}} = \{x_1, x_2\}^{1/2} \textcircled{\cap} \{x_1, x_3\}^{1/2} \textcircled{\cup} \{x_1\}^{3/4}.$$

Selon Smets [127], $m^{\mathcal{X}}$ est donc le résultat des éléments d'information indépendants suivants :

- *une première source fournit l'information qu'il faut croire en $\{x_1, x_2\}$ et une fiabilité de $1/2$ est donnée à cette source ;*
- *une seconde source fournit l'information qu'il faut croire en $\{x_1, x_3\}$ et une fiabilité de $1/2$ est donnée à cette source ;*
- *une troisième source fournit l'information qu'il ne faut pas croire en $\{x_1\}$ et une fiabilité de $3/4$ est donnée à cette source ;*

Une décomposition duale, basée sur la règle disjonctive, peut être obtenue [18]. Elle repose sur la fonction de poids disjonctifs $v : 2^{\mathcal{X}} \backslash \{\emptyset\} \to (0, +\infty)$, qui est une représentation équivalente d'une fonction de masse non normalisée $m^{\mathcal{X}}$, obtenue par

$$v(A) = \prod_{B \subseteq A} b^{\mathcal{X}}(B)^{(-1)^{|A|-|B|+1}}, \quad \forall A \neq \emptyset,$$

avec $b^{\mathcal{X}}$ la fonction d'implicabilité associée à $m^{\mathcal{X}}$. Nous avons

$$m^{\mathcal{X}} = \textcircled{\cup}_{A \neq \emptyset} A_{v(A)}, \tag{1.20}$$

avec $A_{v(A)} : 2^{\mathcal{X}} \rightarrow \mathbb{R}$ une fonction appelée fonction de masse simple généralisée négative allouant $v(A)$ à $\emptyset$, $1 - v(A)$ à $A \subseteq \mathcal{X}, A \neq \emptyset$, et 0 à tout $B \in 2^{\mathcal{X}} \backslash \{A, \emptyset\}$ (si $v(A) \leq 1$ alors cette fonction est une fonction de masse, appelée fonction de masse simple négative). Il n'y a pas d'interprétation claire fournie dans [18] pour les poids disjonctifs $v(A)$, $A \neq \emptyset$.

1.6 Correction

Lorsqu'un agent reçoit d'une source $\mathfrak{s}$ de l'information à propos de $\boldsymbol{x}$ sous forme d'une fonction de croyance $m^{\mathcal{X}}_{\mathfrak{S}}$, il peut construire, grâce à l'opération d'affaiblissement [123, 126], sa connaissance $m^{\mathcal{X}}$ à propos de $\boldsymbol{x}$ en corrigeant l'information fournie par la source en fonction de la fiabilité qu'il lui accorde. Soit $\beta \in [0, 1]$ le degré de croyance de l'agent dans le fait que la source est fiable. La connaissance $m^{\mathcal{X}}$ de l'agent issue de l'affaiblissement de $m^{\mathcal{X}}_{\mathfrak{S}}$ est définie par :

$$m^{\mathcal{X}}(A) = \begin{cases} \beta \cdot m^{\mathcal{X}}_{\mathfrak{S}}(A), & \text{si } A \subset \mathcal{X}, \\ \beta \cdot m^{\mathcal{X}}_{\mathfrak{S}}(A) + 1 - \beta, & \text{sinon.} \end{cases} \tag{1.21}$$

Remarquons que l'équation (1.21) peut aussi être obtenue si l'agent suppose que la source est fiable avec une masse β et non fiable avec une masse $1 - \beta$, plutôt que de supposer que la source est fiable avec un degré de croyance β [97].

L'agent peut disposer de connaissances plus fines sur la fiabilité de la source. Par exemple, il peut considérer que la fiabilité de la source dépend de la vraie valeur de $\boldsymbol{x}$. Mercier *et al.* [97] s'intéressent à ce problème. Plus précisément, ils considèrent une partition de $\mathcal{X}$ et que la source a une fiabilité différente en fonction de l'élément de cette partition auquel appartient la vraie valeur de $\boldsymbol{x}$. Par exemple, si $\boldsymbol{x}$ est la maladie d'un patient donné alors le diagnostic d'un médecin aura une certaine fiabilité si $\boldsymbol{x} \in B$, pour un certain $B \subset \mathcal{X}$, et une autre fiabilité si $\boldsymbol{x} \in \overline{B}$. Mercier *et al.* [97] proposent un mécanisme de correction de fonction de croyance, appelé affaiblissement contextuel basé sur un grossissement, afin de prendre en compte ce type de connaissance. Il est défini de la manière suivante. Soit $\beta_B \in [0, 1]$ le degré de croyance de l'agent que la source est fiable dans le contexte $B \subseteq \mathcal{X}$ et soit $\mathcal{B}$ l'ensemble de contextes pour lequel l'agent possède une telle connaissance contextuelle, où $\mathcal{B}$ forme une partition de $\mathcal{X}$. La connaissance $m^{\mathcal{X}}$ de l'agent issue de l'affaiblissement contextuel basé sur un grossissement de l'information $m^{\mathcal{X}}_{\mathfrak{S}}$ fournie par la source, est donnée par l'équation :

$$m^{\mathcal{X}} = m^{\mathcal{X}}_{\mathfrak{S}} \textcircled{\cup} (\textcircled{\cup}_{B \in \mathcal{B}} B_{\beta_B}),$$

ou, plus simplement (en utilisant un abus de notation usuel),

$$m^{\mathcal{X}} = m^{\mathcal{X}}_{\mathfrak{S}} \textcircled{\cup}_{B\in\mathcal{B}} B_{\beta_B}, \tag{1.22}$$

où B_{β_B} est la fonction de masse simple négative allouant la masse β_B à $\emptyset$ et la masse $1-\beta_B$ à B.

L'équation (1.22) étend l'affaiblissement défini par (1.21), ce dernier pouvant être exprimé par [97] :

$$m^{\mathcal{X}} = m^{\mathcal{X}}_{\mathfrak{S}} \textcircled{\cup} \mathcal{X}_\beta, \tag{1.23}$$

avec $\mathcal{X}_\beta$ la fonction de masse simple négative allouant la masse β à $\emptyset$ et la masse $1-\beta$ à $\mathcal{X}$. En d'autres termes, l'affaiblissement est un cas particulier de l'affaiblissement contextuel basé sur un grossissement, retrouvé pour $\mathcal{B} = \{\mathcal{X}\}$.

1.7 Marginalisation, extension et propagation

Comme remarqué par Shafer [123], la granularité du cadre de discernement $\mathcal{X}$ est toujours, dans une certaine mesure, une question de convention puisque tout élément $x \in \mathcal{X}$ représentant un état de la nature peut toujours être divisé en plusieurs possibilités. Aussi, il est important de savoir comment exprimer sur un cadre plus grossier ou, à l'inverse, plus fin une fonction de croyance définie sur un cadre donné. Cela est possible grâce aux opérations de marginalisation et extension vide rappelée ci-après.

Soit $\mathcal{X}$ et $\mathcal{Y}$ deux ensembles finis. Une application $\rho : 2^{\mathcal{X}} \to 2^{\mathcal{Y}}$ est appelée raffinement si elle vérifie les deux propriétés suivantes :

1. L'ensemble $\{\rho(\{x\}), x \in \mathcal{X}\}$ forme une partition de $\mathcal{Y}$.
2. Pour tout $B \subseteq \mathcal{X}$, nous avons :

$$\rho(B) = \bigcup_{x\in B} \rho(\{x\}). \tag{1.24}$$

Si un tel raffinement ρ existe, alors $\mathcal{Y}$ est appelé raffinement de $\mathcal{X}$ pour ρ, et $\mathcal{X}$ est appelé grossissement de $\mathcal{Y}$ pour ρ.

Il existe plusieurs façons de définir l'inverse de ρ, en particulier la réduction extérieure [123] ρ^* est définie par

$$\rho^*(A) = \{x \in \mathcal{X} | \rho(\{x\}) \cap A \neq \emptyset\}, \quad \forall A \subseteq \mathcal{Y}. \tag{1.25}$$

Les extensions des applications ρ et ρ^* aux fonctions de croyance sont appelées respectivement extension vide et marginalisation. L'extension vide d'une fonction

de masse $m^{\mathcal{X}}$ dans le cadre plus fin $\mathcal{Y}$ est notée $m^{\mathcal{X}\uparrow\mathcal{Y}}$ et donnée par l'équation

$$m^{\mathcal{X}\uparrow\mathcal{Y}}(A) = \begin{cases} m^{\mathcal{X}}(B) & \text{si } A = \rho(B) \text{ pour un } B \subseteq \mathcal{X}, \\ 0 & \text{sinon.} \end{cases} \tag{1.26}$$

Cette opération se justifie par le principe d'engagement minimum [126].

La marginalisation d'une fonction de masse $m^{\mathcal{Y}}$ dans le cadre plus grossier $\mathcal{X}$ est notée $m^{\mathcal{Y}\downarrow\mathcal{X}}$ et donnée par l'équation

$$m^{\mathcal{Y}\downarrow\mathcal{X}}(B) = \sum_{\rho^*(A)=B} m^{\mathcal{Y}}(A), \quad \forall B \subseteq \mathcal{X} \tag{1.27}$$

Le cadre de discernement $\mathcal{X}$ d'une fonction de masse $m^{\mathcal{X}}$ peut être le produit Cartésien d'autres ensembles finis, par exemple on peut avoir $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ avec $\mathcal{X}_1$ et $\mathcal{X}_2$ deux ensembles finis associés à des paramètres $\boldsymbol{x}^1$ et $\boldsymbol{x}^2$ respectivement. Si l'on s'intéresse seulement à l'incertitude contenue dans $m^{\mathcal{X}} = m^{\mathcal{X}_1\times\mathcal{X}_2}$ quant à la vraie valeur du paramètre $\boldsymbol{x}^2$, on peut marginaliser $m^{\mathcal{X}_1\times\mathcal{X}_2}$ sur le cadre $\mathcal{X}_2$ en transférant chaque masse $m^{\mathcal{X}_1\times\mathcal{X}_2}(A)$, $A \subseteq \mathcal{X}_1 \times \mathcal{X}_2$, vers la projection de A sur $\mathcal{X}_2$:

$$m^{\mathcal{X}_1\times\mathcal{X}_2\downarrow\mathcal{X}_2}(B) = \sum_{proj(A\downarrow\mathcal{X}_2)=B} m^{\mathcal{X}_1\times\mathcal{X}_2}(A), \quad \forall B \subseteq \mathcal{X}_2, \tag{1.28}$$

avec $proj(A \downarrow \mathcal{X}_2)$ le sous-ensemble de $\mathcal{X}_2$ résultant de la projection de A sur $\mathcal{X}_2$. À l'inverse, si l'on dispose d'une connaissance concernant seulement $\boldsymbol{x}^2$ représentée par une fonction de masse $m^{\mathcal{X}_2}$, on peut exprimer cette connaissance sur l'espace produit $\mathcal{X}_1 \times \mathcal{X}_2$ à l'aide de l'extension vide définie par

$$m^{\mathcal{X}_2\uparrow\mathcal{X}_1\times\mathcal{X}_2}(A) = \begin{cases} m^{\mathcal{X}_2}(B) & \text{si } A = \mathcal{X}_1 \times B, \text{ pour un } B \subseteq \mathcal{X}_2, \\ 0 & \text{sinon.} \end{cases} \tag{1.29}$$

Les équations (1.29) et (1.28) correspondent respectivement aux équations (1.26) et (1.27), en remarquant que $\mathcal{X}_2$ est le grossissement de $\mathcal{X}_1 \times \mathcal{X}_2$ pour le raffinement $\rho : 2^{\mathcal{X}_2} \to 2^{\mathcal{X}_1\times\mathcal{X}_2}$ défini par $\rho(\{x^2\}) = \mathcal{X}_1 \times \{x^2\}$ pour tout $x^2 \in \mathcal{X}_2$. Remarquons également que l'extension vide (1.29) ne crée pas de connaissance concernant $\boldsymbol{x}^1$ au sens que si l'on marginalise $m^{\mathcal{X}_2\uparrow\mathcal{X}_1\times\mathcal{X}_2}$ sur $\mathcal{X}_1$, on obtient la fonction de masse vide. Cette opération permet entre autres de définir la combinaison par la règle conjonctive de fonctions de masse définies sur des cadres différents :

$$m_1^{\mathcal{X}_1} \textcircled{\cap} m_2^{\mathcal{X}_2} := m_1^{\mathcal{X}_1\uparrow\mathcal{X}_1\times\mathcal{X}_2} \textcircled{\cap} m_2^{\mathcal{X}_2\uparrow\mathcal{X}_1\times\mathcal{X}_2}. \tag{1.30}$$

Concernant les fonctions de masse définies sur des espaces différents, un outil important est le *valuation based system* (VBS) introduit par Shenoy [124]. Ce formalisme permet de représenter et raisonner efficacement à partir de connaissances représentées par des fonctions de croyance définies sur des espaces produits, lorsque

celles-ci sont supposées fiables et indépendantes (auquel cas elles doivent être combinées par la règle conjonctive). Au niveau de la représentation, une description qualitative des connaissances est fournie graphiquement sous forme d'un *valuation network* qu'on traduira ici par réseau évidentiel, comme celui donné par la figure 1.1 : dans le réseau évidentiel, les cadres de discernement $\mathcal{X}_i$ associés à des paramètres $\boldsymbol{x}^i$ d'intérêt sont représentés dans des cercles, et les fonctions de croyance disponibles à propos de ces paramètres (portant généralement sur plus d'un paramètre à la fois) sont représentées dans des losanges. Au niveau du raisonnement, le VBS permet de résoudre efficacement le problème dit de l'inférence [81], *i.e.*, calculer la connaissance marginale à propos d'un sous-ensemble des paramètres étant données toutes les connaissances sur tous les paramètres. Pour résoudre ce problème, la solution directe consiste à d'abord combiner par la règle conjonctive toutes les fonctions de croyance disponibles puis à marginaliser le résultat de la combinaison sur le cadre approprié. Par exemple, pour le réseau évidentiel de la figure 1.1, si l'on s'intéresse au paramètre $\boldsymbol{x}^2$, alors la connaissance marginale sur ce paramètre est obtenue par l'opération :

$$\left(m^{\mathcal{X}_1} \textcircled{\cap} m^{\mathcal{X}_1 \times \mathcal{X}_2} \textcircled{\cap} m^{\mathcal{X}_2 \times \mathcal{X}_3}\right)^{\downarrow \mathcal{X}_2}. \tag{1.31}$$

La difficulté de procéder ainsi est la complexité exponentielle du calcul de la combinaison de toutes les fonctions de croyance, ce qui rend cette solution généralement infaisable. Le VBS exploite des propriétés de la combinaison conjonctive et de la marginalisation, qui rendent possible des calculs locaux évitant d'effectuer explicitement ce calcul (le lecteur est renvoyé vers [125] pour une description de ces calculs locaux).

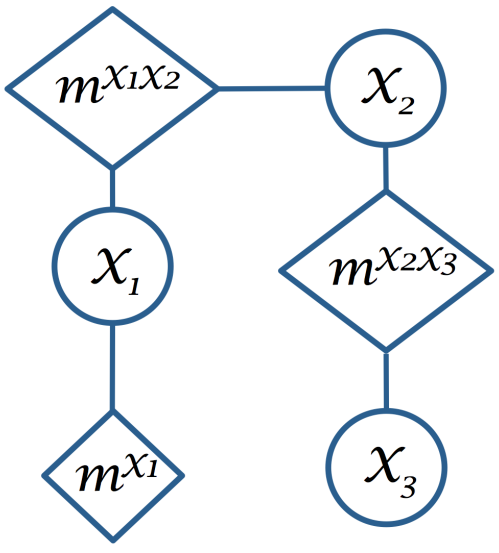


FIGURE 1.1 – Un exemple de réseau évidentiel.

Un dernier aspect utile à rappeler concernant les fonctions de masse définies sur

des espaces produits est la possibilité de propager via une application f l'incertitude à propos de paramètres. Soit $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N$ des paramètres définis sur les ensembles finis $\mathcal{X}_1, \ldots, \mathcal{X}_N$ respectivement et une connaissance jointe à propos des valeurs de ces paramètres représentée par la fonction de masse $m^{\mathcal{X}_1 \times \cdots \times \mathcal{X}_N}$. De plus, soit $\boldsymbol{y}$ un paramètre prenant ses valeurs dans l'ensemble fini $\mathcal{Y}$, et tel que $y = f(x^1, \ldots, x^N)$ pour une application $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \to \mathcal{Y}$. Comme montré dans [43], la connaissance incertaine $m^{\mathcal{X}_1 \times \cdots \times \mathcal{X}_N}$ à propos des paramètres $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N$, induit une fonction de masse $m^{\mathcal{Y}}$ à propos de la vraie valeur de $\boldsymbol{y}$ définie par

$$m^{\mathcal{Y}}(B) = \sum_{f(A)=B} m^{\mathcal{X}_1 \times \cdots \times \mathcal{X}_N}(A), \quad \forall B \subseteq \mathcal{Y}, \tag{1.32}$$

avec $f(A) = \{f(x^1, \ldots, x^N) | (x^1, \ldots, x^N) \in A\}$ pour tout $A \subseteq \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$.

1.8 Prédiction

La prédiction (statistique) s'attache à quantifier l'incertitude à propos de la vraie valeur d'une future réalisation $\boldsymbol{x}$ d'une variable aléatoire $X \sim P_X(\cdot; \boldsymbol{\theta})$, à partir d'une observation passée $\boldsymbol{y}$ d'une variable aléatoire $Y \sim P_Y(\cdot; \boldsymbol{\theta})$, où $\boldsymbol{\theta}$ est un paramètre dont on sait seulement qu'il appartient à un ensemble Θ. Ce problème a reçu différentes solutions dans le cadre de la théorie des fonctions de croyance (voir [23] pour une synthèse récente). La solution proposée par Kanjanatarakul *et al.* [77, 75], qui est compatible avec l'approche Bayésienne pour la prédiction [82], est brièvement rappelée ici.

Dans une première étape, l'approche de Kanjanatarakul *et al.* utilise la méthode pour l'inférence statistique basée sur la vraisemblance introduite par Shafer [123] et justifiée par Denœux [20], qui permet d'obtenir une fonction de croyance $Bel^{\Theta}_{\boldsymbol{y}}$ représentant la connaissance vis-à-vis de $\boldsymbol{\theta}$ étant donné l'observation $\boldsymbol{y}$. Cette fonction de croyance est consonante et sa fonction de contour $pl^{\Theta}_{\boldsymbol{y}}$, qui la caractérise, est la fonction de vraisemblance normalisée associée à $\boldsymbol{y}$:

$$pl^{\Theta}_{\boldsymbol{y}}(\theta) = \frac{L(\theta; \boldsymbol{y})}{L(\hat{\theta}; \boldsymbol{y})}, \quad \forall \theta \in \Theta, \tag{1.33}$$

avec $L(\theta; \boldsymbol{y}) = p(\boldsymbol{y}; \theta)$ et $\hat{\theta}$ un maximum de $L(\theta; \boldsymbol{y})$ (un estimateur du maximum de vraisemblance (EMV) de $\boldsymbol{\theta}$) et où il est supposé que $L(\hat{\theta}; \boldsymbol{y}) < +\infty$. De plus, les ensembles focaux de $Bel^{\Theta}_{\boldsymbol{y}}$ sont définis par :

$$\Gamma_{\boldsymbol{y}}(u) = \{\theta \in \Theta | pl^{\Theta}_{\boldsymbol{y}}(\theta) \geq u\}, \quad u \in [0, 1], \tag{1.34}$$

et l'ensemble aléatoire [101] $\Gamma_{\boldsymbol{y}}(U)$ avec $U \sim \mathcal{U}([0, 1])$ est équivalent à $Bel^{\Theta}_{\boldsymbol{y}}$ au sens que

$$Bel^{\Theta}_{\boldsymbol{y}}(A) = P_U(\{u \in [0, 1] | \Gamma_{\boldsymbol{y}}(u) \subseteq A\}). \tag{1.35}$$

Dans une seconde étape, X est exprimée (à la manière de Dempster [15]) sous

la forme

$$X = \phi(\boldsymbol{\theta}, V), \tag{1.36}$$

où V est une variable pivotale telle que $V \sim \mathcal{U}([0,1])$. Comme V ne dépend pas de $\boldsymbol{\theta}$, V et la variable aléatoire U sous-jacente à $Bel_{\boldsymbol{y}}^{\Theta}$ sont indépendantes. Elles suivent une distribution jointe uniforme sur $[0,1]^2$, et l'ensemble aléatoire $\phi(\Gamma_{\boldsymbol{y}}(U), V)$ induit une fonction de croyance $Bel_{\boldsymbol{y}}^{\mathcal{X}}$ quantifiant l'incertitude à propos de $\boldsymbol{x}$ définie par

$$Bel_{\boldsymbol{y}}^{\mathcal{X}}(A) = P_{U,V}(\{(u,v) \in [0,1]^2 | \phi(\Gamma_{\boldsymbol{y}}(u), v) \subseteq A\}). \tag{1.37}$$

En particulier, si X est une variable aléatoire binaire ($\mathcal{X} = \{0,1\}$) de distribution de Bernoulli $\mathcal{B}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in [0,1]$, alors elle peut être générée à l'aide de l'équation suivante :

$$X = \phi(\boldsymbol{\theta}, V) = \begin{cases} 1, & \text{si } V \leq \boldsymbol{\theta}, \\ 0, & \text{sinon.} \end{cases} \tag{1.38}$$

De plus, si $pl_{\boldsymbol{y}}^{\Theta}$ est uni-modale et continue (ce qui est le cas, par exemple, si Y suit une distribution binomiale), alors les ensembles $\Gamma_{\boldsymbol{y}}(u)$ sont des intervalles fermés $[\theta^-(u), \theta^+(u)]$. En composant $\Gamma_{\boldsymbol{y}}$ avec ϕ, on obtient, pour tout $(u,v) \in [0,1]^2$,

$$\phi(\Gamma_{\boldsymbol{y}}(u), v) = \phi([\theta^-(u), \theta^+(v)], V) = \begin{cases} \{1\}, & \text{si } v \leq \theta^-(u), \\ \{0\}, & \text{si } v > \theta^+(u), \\ \{0,1\}, & \text{sinon.} \end{cases} \tag{1.39}$$

De (1.37) et (1.39), on trouve :

$$Bel_{\boldsymbol{y}}^{\mathcal{X}}(\{1\}) = P_{U,V}(\{(u,v) \in [0,1]^2 | v \leq \theta^-(u)\}). \tag{1.40}$$

Kanjanatarakul *et al.* [77] montrent que $Bel_{\boldsymbol{y}}^{\mathcal{X}}(\{1\})$ vérifie alors

$$Bel_{\boldsymbol{y}}^{\mathcal{X}}(\{1\}) = \hat{\theta} - \int_0^{\hat{\theta}} pl_{\boldsymbol{y}}^{\Theta}(\theta) d\theta. \tag{1.41}$$

De manière similaire, on a

$$\begin{aligned} Pl_{\boldsymbol{y}}^{\mathcal{X}}(\{1\}) &= 1 - Bel_{\boldsymbol{y}}^{\mathcal{X}}(\{0\}) \\ &= 1 - P_{U,V}(\{(u,v) \in [0,1]^2 | v > \theta^+(u)\}) \\ &= P_{U,V}(\{(u,v) \in [0,1]^2 | v \leq \theta^+(u)\}) \end{aligned}$$

et on peut montrer que $Pl_{\boldsymbol{y}}^{\mathcal{X}}(\{1\})$ vérifie alors [77] :

$$Pl_{\boldsymbol{y}}^{\mathcal{X}}(\{1\}) = \hat{\theta} + \int_{\hat{\theta}}^{1} pl_{\boldsymbol{y}}^{\Theta}(\theta) d\theta. \tag{1.42}$$

$Bel_{\boldsymbol{y}}^{\mathcal{X}}(\{1\})$ est le dégre de croyance supportant strictement $\boldsymbol{x} = 1$ alors que la plausibilité $Pl_{\boldsymbol{y}}^{\mathcal{X}}(\{1\})$ est la part de croyance ne le contredisant pas. En outre, la différence $Pl_{\boldsymbol{y}}^{\mathcal{X}}(\{1\}) - Bel_{\boldsymbol{y}}^{\mathcal{X}}(\{1\})$, qui est égale à la masse $m_{\boldsymbol{y}}^{\mathcal{X}}(\{0,1\})$ affectée à l'ignorance, est simplement l'aire sous la fonction $pl_{\boldsymbol{y}}^{\Theta}$ et celle-ci tend vers 0 si, par exemple, Y suit une distribution binomiale de paramètres n et $\boldsymbol{\theta}$, et n tend vers l'infini [77].

1.9 Décision

Soit une fonction $f : \mathcal{X} \to \mathbb{R}$ et une distribution de probabilité P sur $\mathcal{X}$. L'espérance mathématique notée $E(f, P)$ de f relativement à P est donnée par :

$$E(f, P) = \sum_{x \in \mathcal{X}} P(\{x\}) f(x). \tag{1.43}$$

Ce concept probabiliste d'espérance admet plusieurs généralisations en théorie des fonctions de croyance [17]. Soit une fonction de masse $m^{\mathcal{X}}$ sur $\mathcal{X}$. Les espérances inférieure et supérieure de f relativement à $m^{\mathcal{X}}$ sont définies respectivement par :

$$E_*(f, m^{\mathcal{X}}) = \sum_{A \subseteq \mathcal{X}} m^{\mathcal{X}}(A) \min_{x \in A} f(x), \tag{1.44}$$

$$E^*(f, m^{\mathcal{X}}) = \sum_{A \subseteq \mathcal{X}} m^{\mathcal{X}}(A) \max_{x \in A} f(x). \tag{1.45}$$

Ces définitions sont obtenues en considérant les bornes inférieure et supérieure respectivement des espérances de f pour toutes les distributions de probabilités dites compatibles avec $m^{\mathcal{X}}$, une distribution P étant compatible si et seulement si $Bel^{\mathcal{X}}(A) \leq P(A) \leq Pl^{\mathcal{X}}(A)$ pour tout $A \subseteq \mathcal{X}$.

Une autre approche consiste à transformer $m^{\mathcal{X}}$ en une distribution de probabilité P puis à calculer l'espérance de f relativement à P. Une telle transformation peut être obtenue en normalisant la fonction de contour [140, 11]. Soit P_{pl} la distribution de probabilité résultant de cette transformation. Elle est définie par :

$$P_{pl}(\{x\}) = \frac{pl^{\mathcal{X}}(x)}{\sum_x pl^{\mathcal{X}}(x)}, \quad \forall x \in \mathcal{X}. \tag{1.46}$$

Un interêt de cette transformation est que lorsqu'il s'agit de la calculer pour une fonction de masse issue d'une combinaison par la règle conjonctive d'un ensemble de fonctions de masse, alors cela peut être fait efficacement puisque cette transformation commute avec la règle conjonctive. À partir de cette transformation, on obtient une autre espérance de f relativement à $m^{\mathcal{X}}$, notée $E_{pl}(f, m^{\mathcal{X}})$, qui n'est rien d'autre que l'espérance de f relativement à la distribution de probabilité P_{pl}, *i.e.*, $E_{pl}(f, m^{\mathcal{X}}) = E(f, P_{pl})$.

Si $m^{\mathcal{X}}$ est Bayésienne, alors $E_*(f, m^{\mathcal{X}})$, $E^*(f, m^{\mathcal{X}})$ et $E_{pl}(f, m^{\mathcal{X}})$ reviennent à l'espérance de f relativement à la distribution de probabilité P telle que $P(\{x\}) = m^{\mathcal{X}}(\{x\})$ pour tout $x \in \mathcal{X}$. C'est en ce sens que ce sont des généralisations du

concept probabiliste d'espérance.

Ces notions peuvent être utilisées pour établir des règles de décision en reconnaissance des formes généralisant l'approche classique Bayésienne [47]. Soit $\mathcal{X} = \{x_1, \ldots, x_n\}$ un ensemble de formes types, ou classes, auxquelles peuvent appartenir des objets d'intérêt. Une règle de décision associe à chaque objet une action parmi un ensemble fini $\mathcal{A} = \{a_1, \ldots, a_n\}$ d'actions, l'action a_i étant interprétée comme l'affectation de l'objet à la classe x_i. De plus, le choix de l'action a_i pour un objet appartenant à la classe x_j est supposé entraîner un coût $f_i(x_j)$. Dans l'approche classique, étant donné une connaissance sur la classe d'un objet représentée par une distribution de probabilité P sur $\mathcal{X}$, la règle de décision utilisée est celle consistant à affecter à l'objet l'action $a \in \mathcal{A}$ pour laquelle l'espérance de son coût relativement à P est la plus faible :

$$a = \arg\min_{a_i \in \mathcal{A}} E(f_i, P). \tag{1.47}$$

Notons qu'il est possible d'ajouter à l'ensemble $\mathcal{A}$ une action a_0, telle que $f_0(x_j) = \lambda \geq 0$ pour tout $x_j \in \mathcal{X}$. Une telle action a un coût espéré $E(f_0, P) = \lambda$ quelque soit P. Cette action, dite de rejet, permet d'éviter d'affecter un objet à une des classes lorsque les coûts espérés d'affectation aux classes sont trop élevées, c'est-à-dire qu'ils dépassent un seuil limite λ.

Dans le cadre de la théorie des fonctions de croyance, l'existence de plusieurs définitions de l'espérance d'une fonction permet de considérer différentes stratégies de décision. En effet, si l'on dispose d'une connaissance sur la classe de l'objet représentée non pas par une distribution de probabilité mais par une fonction de masse $m^{\mathcal{X}}$, on peut, par exemple, utiliser une règle de décision "pessimiste" consistant à affecter à l'objet l'action $a^* \in \mathcal{A}$ pour laquelle l'espérance supérieure de son coût relativement à $m^{\mathcal{X}}$ est la plus faible :

$$a^* = \arg\min_{a_i \in \mathcal{A}} E^*(f_i, m^{\mathcal{X}}). \tag{1.48}$$

1.10 Conclusion

Ce chapitre a rappelé les notions de la théorie des fonctions de croyance utiles à la présentation de mes travaux dans le reste de ce mémoire. Le chapitre suivant sera l'occasion de revenir en détail sur les mécanismes de correction et de fusion d'informations. Il sera montré entre autres que tous les mécanismes rappelés dans ce chapitre sont des cas particuliers d'une approche générale de correction et fusion d'informations basée sur des hypothèses explicites sur la qualité des sources.

Chapitre 2

Un modèle pour l'interprétation de témoignages

2.1 Introduction

Afin de pouvoir interpréter des informations fournies par des sources, il est nécessaire de disposer de connaissances sur la qualité de ces sources. Dans les mécanismes de correction et fusion existants, ces méta-connaissances portent classiquement sur des hypothèses concernant la fiabilité des sources, fiabilité entendue en termes de pertinence. Au-delà de la pertinence, j'ai proposé de considérer le cas où des connaissances sur d'autres facettes de la qualité des sources sont disponibles [103, 108, 109] (la référence [108] est fournie en Annexe A). Dans ce chapitre, la possibilité d'ajouter des hypothèses à propos de la sincérité des sources est d'abord considérée. Puis, un modèle général capable de gérer des hypothèses à propos de diverses formes de qualité des sources est présenté.

2.2 Pertinence et sincérité

Nous supposons dans cette section que la fiabilité d'une source inclut une autre dimension en plus de sa pertinence : sa sincérité. Une source sincère est une source qui délivre l'information qu'elle possède. Une source peut être non sincère de différentes manières. La forme la plus simple d'insincérité pour une source est de dire le contraire de ce qu'elle croit être la vérité. Elle peut aussi en déclarer moins ou même autre chose cohérent avec ses croyances. Par exemple, un biais systématique d'un capteur peut être vu comme une forme de manque de sincérité. Cependant, dans cette section, nous supposons seulement la forme la plus simple d'insincérité.

2.2.1 Cas d'une seule source

Considérons le cas où une seule source $\mathfrak{s}$ fournit une information à propos d'un paramètre $\boldsymbol{x}$ défini sur un domaine $\mathcal{X}$ et que cette information est de la forme $\boldsymbol{x} \in A$, où A est un sous-ensemble propre non vide [1] de $\mathcal{X}$. Si $\mathfrak{s}$ est supposée non pertinente, quelque soit sa sincérité, l'information qu'elle fournit est totalement inutile et peut être remplacée par l'information triviale $\boldsymbol{x} \in \mathcal{X}$. Au contraire, si $\mathfrak{s}$ est supposée pertinente et dire le contraire de ce qu'elle sait, alors une information valide à propos

1. Toute entité fournissant une information non-triviale et non-contradictoire est considérée comme source.

de $\boldsymbol{x}$ peut être retrouvée : il faut remplacer $\boldsymbol{x} \in A$ par $\boldsymbol{x} \in \overline{A}$. Naturellement, si $\mathfrak{s}$ est supposée fiable et sincère, alors on infère que $\boldsymbol{x} \in A$.

Formellement, soit $\mathcal{H} = \{(PE, SI), (PE, \neg SI), (\neg PE, SI), (\neg PE, \neg SI)\}$ l'espace des états possibles de la source au regard de sa pertinence et sa sincérité, où PE et SI signifient que $\mathfrak{s}$ est pertinente et sincère respectivement. Alors, en suivant l'approche de Demspter [14], le raisonnement précédent peut être représenté par une fonction multivoque $\Gamma_A : \mathcal{H} \to \mathcal{X}$ telle que

$$\begin{aligned} \Gamma_A(PE, SI) &= A; && (2.1)\\ \Gamma_A(PE, \neg SI) &= \overline{A}; && (2.2)\\ \Gamma_A(\neg PE, SI) &= \Gamma(\neg PE, \neg SI) = \mathcal{X}. && (2.3) \end{aligned}$$

$\Gamma_A(h)$ indique comment interpréter le témoignage $\boldsymbol{x} \in A$ dans chaque état $h \in \mathcal{H}$ de $\mathfrak{s}$.

En général, la connaissance à propos de la pertinence et sincérité de la source est incertaine. Précisément, chaque état $h \in \mathcal{H}$ peut être assigné une probabilité $prob(h)$ telle que $\sum_h prob(h) = 1$. Dans ce cas, l'information $\boldsymbol{x} \in A$ induit l'état de connaissance sur $\mathcal{X}$ représenté par la fonction de masse $m^{\mathcal{X}}$ définie par

$$\begin{aligned} m^{\mathcal{X}}(A) &= prob(PE, SI); && (2.4)\\ m^{\mathcal{X}}(\overline{A}) &= prob(PE, \neg SI); && (2.5)\\ m^{\mathcal{X}}(\mathcal{X}) &= prob(\neg PE) = prob(\neg PE, SI) + prob(\neg PE, \neg SI). && (2.6) \end{aligned}$$

Le témoignage fournit par la source peut lui-même être incertain et, en particulier, prendre la forme d'une fonction de masse $m^{\mathcal{X}}_{\mathfrak{S}}$ sur $\mathcal{X}$. Dans ce cas, puisque tout $m^{\mathcal{H}}_{\mathfrak{S}}(A)$ doit être transféré à $\Gamma_A(h)$ si $\mathfrak{s}$ est dans l'état h, l'état de connaissance sur $\mathcal{X}$ est donné par la fonction de masse suivante :

$$m^{\mathcal{X}}(B) = \sum_h prob(h) \sum_{A:\Gamma_A(h)=B} m^{\mathcal{X}}_{\mathfrak{S}}(A). \tag{2.7}$$

En particulier, supposer que $p = prob(PE)$ et $q = prob(SI)$, et que la pertinence de la source est indépendante de sa sincérité amène à transformer $m^{\mathcal{X}}_{\mathfrak{S}}$ en la fonction de masse $m^{\mathcal{X}}$ définie par :

$$m^{\mathcal{X}}(A) = pq\, m^{\mathcal{X}}_{\mathfrak{S}}(A) + p(1-q)\, \overline{m}^{\mathcal{X}}_{\mathfrak{S}}(A) + (1-p)\, m^{\mathcal{X}}_{\mathcal{X}}(A), \quad \forall A \subseteq \mathcal{X}, \tag{2.8}$$

où $\overline{m}^{\mathcal{X}}_{\mathfrak{S}}$ est la négation de $m^{\mathcal{X}}_{\mathfrak{S}}$ et $m^{\mathcal{X}}_{\mathcal{X}}$ est la fonction de masse vide.

L'opération d'affaiblissement (1.21) est un cas particulier de la transformation (2.8), retrouvé pour $q = 1$: cela correspond à une source partiellement pertinente qui est sincère. L'opération de "reniement" (*negating* en anglais [113]) est aussi un cas particulier, retrouvé pour $p = 1$, qui correspond à une source partiellement sincère qui est pertinente. En particulier, la négation d'une fonction de masse est obtenue pour $p = 1$ et $q = 0$: cela correspond à une source pertinente qui ment –

le terme "mentir" est utilisé comme un synonyme de "dire le contraire de ce qu'elle croit être la vérité", indépendamment de l'existence d'une intention de sa part de tromper.

D'autres formes de méta-connaissances incertaines peuvent être envisagées, telles que savoir seulement que l'état de la source appartient à un sous-ensemble H de $\mathcal{H}$. Cela arrive par exemple si la source est supposée être pertinente ou sincère mais pas les deux, *i.e.*, $H = \{(PE, \neg SI), (\neg PE, SI)\}$. Dans ce cas, on doit déduire que $\boldsymbol{x} \in \Gamma_A(H)$, où $\Gamma_A(H)$ est l'image de H par Γ_A :

$$\Gamma_A(H) = \bigcup_{h \in H} \{\Gamma_A(h)\}.$$

Ce type d'hypothèses non élémentaires n'est en fait pas très intéressant ($\Gamma_A(H) = \mathcal{X}$ si $|H| > 1$), mais il est important dans le cas où plusieurs sources fournissent des informations.

2.2.2 Cas de plusieurs sources

Afin d'interpréter des informations fournies par deux sources, il faut faire des hypothèses sur leur état conjoint en termes de pertinence et de sincérité. Soit $\mathcal{H}_i$ l'ensemble des états possibles de la source $\mathfrak{s}_i$, $i = 1, 2$. L'ensemble des hypothèses élémentaires sur les sources est alors $\mathcal{H}_{1:2} = \mathcal{H}_1 \times \mathcal{H}_2$ (nous avons $|\mathcal{H}_{1:2}| = 16$).

Considérons un cas simple où chaque source $\mathfrak{s}_i$ fournit l'information $\boldsymbol{x} \in A_i$, $i = 1, 2$, et où elles sont supposées être dans l'état $\mathbf{h} = (h^1, h^2) \in \mathcal{H}_{1:2}$, avec $h^i \in \mathcal{H}_i$ l'état de $\mathfrak{s}_i$, $i = 1, 2$. Le résultat de la combinaison des informations $\mathbf{A} = (A_1, A_2) \subseteq \mathcal{X} \times \mathcal{X}$ dépend de l'hypothèse $\mathbf{h}$ faite quant à leur comportement, et peut être représenté par une fonction multivoque $\Gamma_{\mathbf{A}} : \mathcal{H}_{1:2} \to \mathcal{X}$.

Puisque l'on doit déduire $\boldsymbol{x} \in \Gamma_{A_i}(h^i)$ quand $\mathfrak{s}_i$ dit $x \in A_i$ et est dans l'état $h_i \in \mathcal{H}_i$, il est clair que

$$\Gamma_{\mathbf{A}}(\mathbf{h}) = \Gamma_{A_1}(h^1) \cap \Gamma_{A_2}(h^2),$$

et pour des hypothèses non-élémentaires $H \subseteq \mathcal{H}_{1:2}$, on a $\Gamma_{\mathbf{A}}(H) = \bigcup_{\mathbf{h} \in H} \{\Gamma_{\mathbf{A}}(\mathbf{h})\}$.

Ainsi, par exemple, si l'on suppose que les deux sources sont pertinentes, et que $\mathfrak{s}_1$ est sincère si et seulement si $\mathfrak{s}_2$ l'est aussi, alors on doit conclure que $\boldsymbol{x} \in (A_1 \cap A_2) \cup (\overline{A_1} \cap \overline{A_2})$, ce qui correspond au connecteur d'équivalence logique. Plus généralement, on montre que tous les connecteurs de la logique propositionnelle peuvent être retrouvés avec des hypothèses sur la qualité des sources en termes de pertinence et de sincérité.

Supposons maintenant que les témoignages des sources $\mathfrak{s}_1$ et $\mathfrak{s}_2$ sont incertains et prennent la forme de fonctions de masse $m_1^{\mathcal{X}}$ et $m_2^{\mathcal{X}}$ respectivement. Supposons de plus que les sources sont indépendantes au sens suivant : si on interprète $m_i^{\mathcal{X}}(A_i)$ comme la probabilité que $\mathfrak{s}_i$ fournisse l'information $\boldsymbol{x} \in A_i$, alors la probabilité que $\mathfrak{s}_1$ fournisse l'information $\boldsymbol{x} \in A_1$ et $\mathfrak{s}_2$ fournisse conjointement l'information $\boldsymbol{x} \in A_2$ est $m_1^{\mathcal{X}}(A_1) \cdot m_2^{\mathcal{X}}(A_2)$. Enfin, faisons l'hypothèse supplémentaire que la

méta-connaissance sur les sources est incertaine et représentée par une fonction de masse $m^{\mathcal{H}_{1:2}}$ sur $\mathcal{H}_{1:2}$. Dans ce cas général, on montre que la connaissance induite sur $\mathcal{X}$ est représentée par la fonction de masse suivante :

$$m^{\mathcal{X}}(B) \quad = \quad \sum_{H} m^{\mathcal{H}_{1:2}}(H) \sum_{\mathbf{A}:\Gamma_{\mathbf{A}}(H)=B} m_1^{\mathcal{X}}(A_1) m_2^{\mathcal{X}}(A_2). \tag{2.9}$$

Cette approche peut être formellement étendue au cas de sources dépendantes, en utilisant le cadre de Destercke et Dubois [32]. Elle peut aussi être directement étendue au cas de $N > 2$ sources qui sont partiellement pertinentes et sincères comme fait dans [107], bien que dans ce cas la complexité algorithmique induite implique qu'en pratique la méta-connaissance sur les sources utilisée doive rester relativement simple.

La règle de combinaison (2.9) inclut toutes les règles basées sur des opérateurs Booléens, et en particulier les règles conjonctive et disjonctive. La première est retrouvée en supposant que les deux sources sont pertinentes et sincères et la seconde en supposant, *e.g.*, que les deux sources sont pertinentes et au moins une d'entre elles est sincère. Une autre règle intéressante basée sur les opérateurs Booléens que l'on peut obtenir est celle que nous avons étudiée dans [107], nommée depuis *evidential q-relaxation* [145], et correspondant à l'hypothèse que les N sources à combiner sont sincères mais que seulement q d'entre elles sont pertinentes (hypothèse commune en analyse par intervalles [70]). On peut également montrer que la règle (2.9) comprend d'autres méthodes de fusion usuelles : celle utilisée dans diverses approches (voir, *e.g.*, [80, 119, 144, 16]) qui consiste à affaiblir les sources puis à les combiner par la règle conjonctive, ainsi qu'une règle voisine qu'est la moyenne pondérée de fonctions de croyance.

2.3 Hypothèses générales de comportement

Jusqu'ici les méta-connaissances concernaient la pertinence et la forme la plus simple de manque de sincérité. Toutefois, dans certaines applications, le manque de sincérité peut prendre des formes plus fines. De plus, les connaissances à propos de la qualité des sources peuvent même concerner d'autres aspects que leur pertinence et sincérité. Une approche permettant de prendre en compte des hypothèses générales à propos du comportement des sources est donc nécessaire. Une telle approche est présentée dans cette section.

2.3.1 Modèle

Supposons qu'une source $\mathfrak{s}$ informe sur la valeur d'un paramètre $\boldsymbol{y}$ défini sur un domaine $\mathcal{Y}$ et que l'information prenne la forme $\boldsymbol{y} \in A$, pour un $A \subseteq \mathcal{Y}$. On suppose de plus que $\mathfrak{s}$ puisse être dans un état parmi v états élémentaires, au lieu de quatre comme c'est le cas dans la section 2.2, *i.e.*, on généralise l'espace des états de $\mathcal{H} = \{(PE, SI), (PE, \neg SI), (\neg PE, SI), (\neg PE, \neg SI)\}$ à $\mathcal{H} = \{h_1, \ldots, h_v\}$. Par

ailleurs, on est intéressé par la valeur prise par un paramètre connexe $\boldsymbol{x}$ défini sur un domaine $\mathcal{X}$ et on a à notre disposition une méta-connaissance qui lie l'information $\boldsymbol{y} \in A$ fournie par $\mathfrak{s}$ à une information de la forme $\boldsymbol{x} \in B$, pour un $B \subseteq X$, pour chaque état possible $h \in \mathcal{H}$ de $\mathfrak{s}$. En d'autres termes, pour chaque $A \subseteq \mathcal{Y}$, il y a une fonction multivoque $\Gamma_A : \mathcal{H} \to \mathcal{X}$ prescrivant, pour chaque hypothèse élémentaire $h \in \mathcal{H}$, comment interpréter sur $\mathcal{X}$ l'information $\boldsymbol{y} \in A$ fournie par $\mathfrak{s}$. On ajoute également la condition naturelle qu'il existe $h \in \mathcal{H}$ tel que $\Gamma_{A_1}(h) \neq \Gamma_{A_2}(h)$ pour n'importe quels sous-ensembles distincts A_1 et A_2 de $\mathcal{Y}$. Des hypothèses incomplètes $H \subseteq \mathcal{H}$ peuvent aussi être considérées, auquel cas l'information $\boldsymbol{y} \in A$ est interprétée comme $\boldsymbol{x} \in \Gamma_A(H) = \cup_{h \in H} \Gamma_A(h)$.

Le cadre de la section 2.2 est un cas particulier, retrouvé en choisissant $v = 4$, $\boldsymbol{y} = \boldsymbol{x}$ et, *e.g.*, $h_1 = (PE, SI)$, $h_2 = (PE, \neg SI)$, $h_3 = (\neg PE, SI)$, $h_4 = (\neg PE, \neg SI)$. L'exemple 2.1 donne une autre illustration de cette approche générale.

Exemple 2.1 (Inspiré de Janez et Appriou [68]) *On cherche à déterminer le type $\boldsymbol{x}$ d'une voie routière, avec $\boldsymbol{x}$ défini sur $\mathcal{X} = \{chemin, route, autoroute\}$. Une source $\mathfrak{s}$ fournit de l'information sur le type, mais elle a une perception limitée des types possibles et en particulier ne connait pas le type "route", de sorte à ce qu'elle fournit de l'information sur l'espace $\mathcal{Y} = \{chemin, autoroute\}$. De plus, on sait que cette source discrimine entre les voies routières en utilisant soit leur largeur soit leur revêtement. Si la source utilise la largeur, alors lorsqu'elle dit "chemin", on peut seulement inférer sans risque que le type est "chemin ou route" puisque les chemins et les routes ont des largeurs similaires, et lorsqu'elle dit "autoroute", on peut conclure "autoroute". D'un autre côté, si elle utilise le revêtement, alors lorsqu'elle déclare "chemin", on peut déduire "chemin", et lorsqu'elle dit "autoroute", on peut seulement déduire "route ou autoroute" puisque ces deux types de voies ont des revêtements similaires.*

Ce problème peut être formalisé en utilisant les fonctions multivoques Γ_{chemin}, $\Gamma_{autoroute}$, et $\Gamma_{\mathcal{Y}}$ de $\mathcal{H} = \{largeur, revêtement\}$ vers $\mathcal{X}$ définies par

$$\begin{aligned}
\Gamma_{chemin}(largeur) &= \{chemin, route\}, \\
\Gamma_{chemin}(revêtement) &= \{chemin\}, \\
\Gamma_{autoroute}(largeur) &= \{autoroute\}, \\
\Gamma_{autoroute}(revêtement) &= \{route, autoroute\}, \\
\Gamma_{\mathcal{Y}}(largeur) &= \mathcal{X}, \\
\Gamma_{\mathcal{Y}}(revêtement) &= \mathcal{X}.
\end{aligned}$$

Lorsque le témoignage de la source et la méta-connaissance sur elle sont incertains et représentés par les fonctions de masse $m^{\mathcal{Y}}_{\mathfrak{s}}$ et $m^{\mathcal{H}}$ respectivement, la connaissance sur $\mathcal{X}$ étant donné $m^{\mathcal{Y}}_{\mathfrak{s}}$ et $m^{\mathcal{H}}$ est obtenu par l'équation

$$m^{\mathcal{X}}(B) = \sum_{H} m^{\mathcal{H}}(H) \sum_{A:\Gamma_A(H)=B} m^{\mathcal{Y}}_{\mathfrak{s}}(A), \quad \forall B \subseteq \mathcal{X}. \tag{2.10}$$

La transformation (2.10) est appelée correction basée sur le comportement (*Behaviour-Based Correction* (BBC) en anglais) puisqu'elle corrige l'information fournie par la source étant donné notre connaissance sur son comportement. Elle généralise (2.7) ainsi que la méthode de déconditionnement par association des hypothèses les plus compatibles [69], qui a son tour généralise une operation de la théorie des fonctions de croyance appelée *ballooning extension* [126].

Passer au cas de deux sources $\mathfrak{s}_1$ et $\mathfrak{s}_2$ fournissant des informations à propos de $\boldsymbol{y}$ et pouvant être chacune dans un parmi v états, avec $\mathcal{H}_1$ et $\mathcal{H}_2$ les espaces des états de ces sources, ne pose pas de difficulté particulière. On obtient ainsi que si elles fournissent des témoignages incertains $m_1^{\mathcal{Y}}$ et $m_2^{\mathcal{Y}}$ et que l'on a une méta-connaissance incertaine $m^{\mathcal{H}_{1:2}}$ sur elles, avec $\mathcal{H}_{1:2} = \mathcal{H}_1 \times \mathcal{H}_2$, alors la connaissance induite sur $\mathcal{X}$ est donnée par l'équation :

$$m^{\mathcal{X}}(B) \quad = \quad \sum_{H} m^{\mathcal{H}_{1:2}}(H) \sum_{\mathbf{A}:\Gamma_{\mathbf{A}}(H)=B} m_1^{\mathcal{Y}}(A_1) m_2^{\mathcal{Y}}(A_2), \tag{2.11}$$

avec $\Gamma_{\mathbf{A}}(\mathbf{h}) = \Gamma_{A_1}(h^1) \cap \Gamma_{A_2}(h^2)$ où $\Gamma_{A_i}(h^i)$ indique comment interpréter sur $\mathcal{X}$ l'information $\boldsymbol{y} \in A_i$ fournie par $\mathfrak{s}_i$ lorsqu'elle est dans l'état $h^i \in \mathcal{H}_i$. La combinaison (2.11) généralise (2.9) et est appelée fusion basée sur le comportement (*Behaviour-Based Fusion* (BBF) en anglais).

2.3.2 Cas particuliers

Deux cas particuliers du modèle présenté dans la section 2.3.1 sont intéressants à discuter. Le premier cas concerne une propriété particulière que peut respecter la méta-connaissance $m^{\mathcal{H}_{1:2}}$. Le second cas concerne la possibilité de représenter des hypothèses h fines concernant la sincérité des sources.

2.3.2.1 Sources méta-indépendantes

La règle BBF (2.11) suppose l'indépendance des sources mais n'impose pas l'indépendance de leurs comportements. Si on suppose cette "méta-indépendence", on aura $m^{\mathcal{H}_{1:2}}(H) = m^{\mathcal{H}_1}(H_1) m^{\mathcal{H}_2}(H_2)$ si $H = H_1 \times H_2$ et $m^{\mathcal{H}_{1:2}}(H) = 0$ sinon [2]. Une question naturelle qui se pose dans ce cas est de savoir s'il est équivalent de combiner les témoignages $m_1^{\mathcal{Y}}$ et $m_2^{\mathcal{Y}}$ par la règle BBF, ou de les corriger individuellement avec la procédure BBC (2.10) en utilisant les méta-connaissances $m^{\mathcal{H}_1} = m^{\mathcal{H}_{1:2}\downarrow\mathcal{H}_1}$ et $m^{\mathcal{H}_2} = m^{\mathcal{H}_{1:2}\downarrow\mathcal{H}_2}$ respectivement, puis de combiner par la règle conjonctive les témoignages corrigés (Figure 2.1).

Le théorème 2.1 donne la réponse à cette question :

Théorème 2.1 *Si les sources sont méta-indépendantes, il est équivalent de combiner $m_1^{\mathcal{Y}}$ et $m_2^{\mathcal{Y}}$ par la règle BBF, ou de combiner par la règle conjonctive chacune de ces informations corrigées avec la procédure BBC.*

2. Cela correspond à l'indépendance évidentielle [123] entre les espaces $\mathcal{H}_1$ et $\mathcal{H}_2$ par rapport à $m^{\mathcal{H}_{1:2}}$.

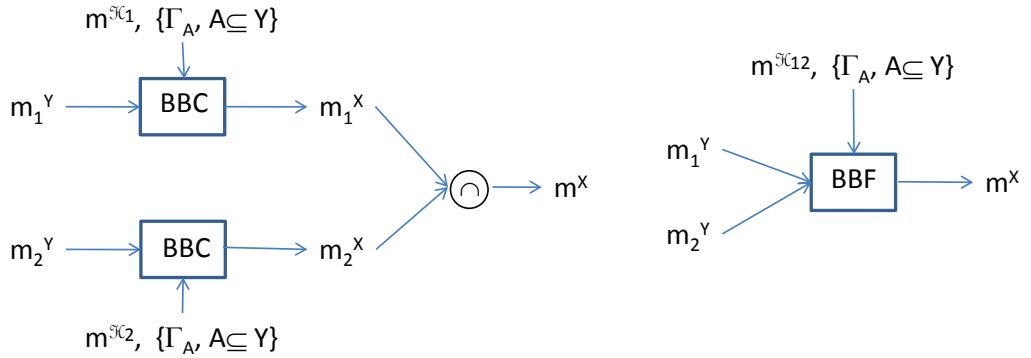


FIGURE 2.1 – Deux façons de combiner des témoignages $m_1^{\mathcal{Y}}$ et $m_2^{\mathcal{Y}}$: en utilisant la procédure BBC (à gauche) et en utilisant la règle BBF (à droite).

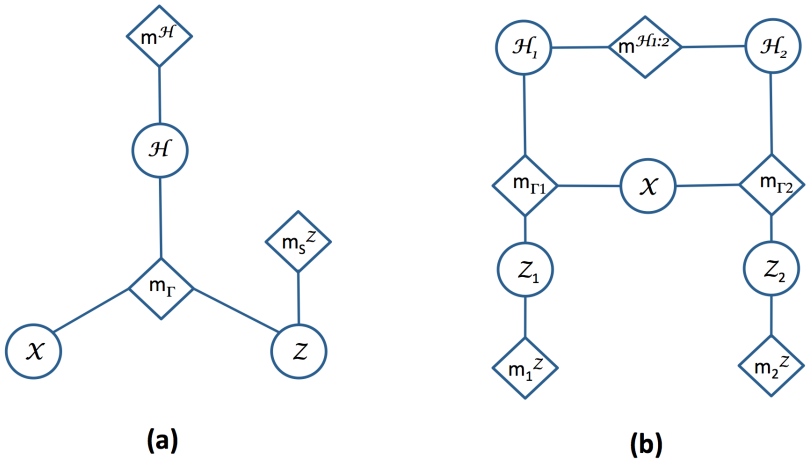


FIGURE 2.2 – Réseaux évidentiels correspondant à la procédure BBC (a) et la règle BBF (b).

Au-delà de l'intérêt propre que ce théorème peut avoir, sa démonstration est instructive. En effet, elle a nécessité d'exprimer la procédure BBC et la règle BBF sous forme de réseaux évidentiels (Figures 2.2a et 2.2b respectivement) représentant toutes les informations disponibles. En substance, les fonctions de masse sur $\mathcal{Z} = 2^{\mathcal{Y}}$ correspondent aux informations incertaines fournies par les sources, les fonctions de masse avec Γ en indice encodent les fonctions multivoques Γ_A d'interprétation des témoignages en fonction des états des sources, et les fonctions de masse sur $\mathcal{H}$ représentent les méta-connaissances incertaines. On montre que la correction et la fusion basées sur les comportements sont retrouvées en marginalisant sur $\mathcal{X}$ les connaissances jointes que ces réseaux contiennent [109].

2.3.2.2 Sincérité contextuelle

Le modèle général présenté à la section 2.3.1 permet de modéliser des formes fines de manque de sincérité, comme je l'ai montré dans [104, 113] (la référence [113] est fournie en Annexe B) et le présente dans cette section.

En regardant de plus près l'état de non sincérité $\neg SI$ considéré dans la section 2.2, on peut remarquer qu'il correspond à supposer qu'une source $\mathfrak{s}$ dit le contraire de ce qu'elle sait, quelque soit ce qu'elle dise concernant chacune des valeurs possibles $x_i \in \mathcal{X}$ qu'admet le paramètre $\boldsymbol{x}$, puisque on doit inverser ce que $\mathfrak{s}$ dit pour chacune de ces valeurs. Par exemple, soit $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ et supposons que $\mathfrak{s}$ déclare $\boldsymbol{x} \in A = \{x_1, x_3\}$, *i.e.*, elle dit que x_1 et x_3 sont des valeurs possibles pour $\boldsymbol{x}$ et que x_2 et x_4 n'en sont pas. Dès lors, si $\mathfrak{s}$ est supposée être dans l'état $\neg SI$, on doit déduire que $\boldsymbol{x} \in \overline{A} = \{x_2, x_4\}$, *i.e.*, x_1 et x_3 ne sont pas des valeurs possibles pour $\boldsymbol{x}$ et x_2 et x_4 en sont.

Cette discussion suggère d'introduire la notion de sincérité d'une source *pour une valeur* $x_i \in \mathcal{X}$: une source sincère (resp. non sincère) pour une valeur $x_i \in \mathcal{X}$ est une source qui dit ce qu'elle sait (resp. le contraire de ce qu'elle sait) pour cette valeur. L'état $\neg SI$ correspond alors à l'hypothèse qu'une source n'est pas sincère pour *toutes* les valeurs $x_i \in \mathcal{X}$. C'est donc un modèle plutôt extrême du manque de sincérité d'une source et il semble intéressant de considérer des formes plus subtiles d'insincérité.

En particulier une source $\mathfrak{s}$ peut être non sincère seulement pour *quelques* valeurs $x_i \in \mathcal{X}$ (et sincère pour les autres valeurs $x_i \in \mathcal{X}$), *i.e.*, un manque contextuel de sincérité. Soit ℓ_B l'état signifiant que la source est sincère pour toutes les valeurs dans $B \subseteq \mathcal{X}$, et qu'elle ne l'est pas pour toutes les valeurs dans $\overline{B}$ (donc $\neg SI = \ell_\emptyset$, et $SI = \ell_\mathcal{X}$). On montre que si la source dit $\boldsymbol{x} \in A$ et est supposée être dans l'état ℓ_B, alors on doit déduire que $\boldsymbol{x} \in \Gamma_A(\ell_B) = (A \cap B) \cup (\overline{A} \cap \overline{B})$ (des exemples où un tel état peut être intéressant sont fournis dans [113]).

En étant encore plus précis à propos des hypothèses sous-jacentes à l'état $\neg SI$, un modèle raffiné du manque contextuel de sincérité peut être obtenu, en utilisant les notions de clauses positive et négative [64, Chapter 8]. Par exemple, lorsque $\mathfrak{s}$ déclare que x_1 est une valeur possible pour $\boldsymbol{x}$, c'est une clause positive communiquée par par la source, et lorsque $\mathfrak{s}$ déclare que x_2 n'est pas une valeur possible pour $\boldsymbol{x}$, c'est une clause négative. Partant de là, la sincérité d'une source pour chaque $x_i \in \mathcal{X}$ peut être caractérisée au regard de la *polarité* des clauses qu'elle dit. Plus précisément, une source est dite *positivement* sincère (resp. non sincère) pour une valeur $x_i \in \mathcal{X}$, lorsqu'elle déclare que x_i est une valeur possible $\boldsymbol{x}$ et sait qu'elle l'est (resp. qu'elle ne l'est pas). Par conséquent, quand une source est supposée positivement sincère (resp. non sincère) pour $x_i \in \mathcal{X}$ et déclare que x_i est une valeur possible pour $\boldsymbol{x}$, alors on doit déduire qu'elle l'est (resp. qu'elle ne l'est pas). De manière similaire, une source est *négativement* sincère (resp. non sincère) pour $x_i \in \mathcal{X}$, quand elle déclare que x_i n'est pas une valeur possible pour $\boldsymbol{x}$ et sait qu'elle ne l'est pas (resp. qu'elle l'est). Aussi, quand une source est supposée négativement sincère (resp. non sincère) pour $x_i \in \mathcal{X}$ et déclare que x_i n'est pas une valeur possible pour $\boldsymbol{x}$, on doit

déduire qu'elle ne l'est pas (resp. qu'elle l'est). Équipé de ces nouvelles notions, on voit que l'état $\neg SI$ correspond à une source qui est positivement *et* négativement non sincère pour *toutes* les valeurs $x_i \in \mathcal{X}$. En utilisant cet état, on fait donc deux hypothèses fortes : le contexte (ensemble de valeurs) concerné par le manque de sincérité est *tout* le cadre de discernement, et les *deux* polarités sont concernées par le manque de sincérité.

Cela suggère à nouveau de considérer des hypothèses plus faibles concernant le manque de sincérité. Deux états sont particulièrement intéressants. Le premier, noté p_B, correspond à l'hypothèse qu'une source est (positivement et négativement) sincère pour tout $x_i \in B$, et positivement non sincère et négativement sincère pour tout $x_i \in \overline{B}$. Le second, noté n_B, correspond à l'hypothèse qu'une source est positivement sincère et négativement non sincère pour tout $x_i \in B$, et (positivement et négativement) sincère pour tout $x_i \in \overline{B}$. On montre qu'étant donné un témoignage $\boldsymbol{x} \in A$, on déduit pour ces deux états $\boldsymbol{x} \in \Gamma_A(p_B) = A \cap B$ et $\boldsymbol{x} \in \Gamma_A(n_B) = A \cup B$ respectivement. Remarquons que, plus généralement, on peut obtenir tous les connecteurs logiques entre le témoignage A et le contexte B, à partir d'hypothèses portant sur la polarité de la (in)sincérité dans B et dans $\overline{B}$.

En utilisant dans le cadre de la procédure BBC, les mensonges contextuels représentés par les états ℓ_B (menteur dans $\overline{B}$), p_B (menteur positif dans $\overline{B}$) et n_B (menteur négatif dans B), $B \subseteq \mathcal{X}$, on obtient des généralisations de l'affaiblissement contextuel basé sur un grossissement et du reniement [3] rappelés en section 1.6, ainsi qu'un mécanisme de correction dual à l'affaiblissement contextuel (introduit d'un point de vue purement formel dans [95] et appelé renforcement contextuel). Je m'attarde brièvement ici seulement sur les résultats concernant l'affaiblissement contextuel basé sur un grossissement, qui est le plus connu d'entre ces mécanismes.

Soit la méta-connaissance suivante : la source $\mathfrak{s}$ est sincère avec une masse β_B et ment négativement dans B (*i.e.*, est dans l'état n_B) avec une masse $1 - \beta_B$, pour un $B \subseteq \mathcal{X}$. Soit un ensemble de telles méta-connaissances pour des $B \in \mathcal{B}$ avec $\mathcal{B}$ un sous-ensemble de $2^{\mathcal{X}}$. On montre que si l'on suppose qu'au moins une d'entre ces méta-connaissances est jugée fiable, alors il faut interpréter un témoignage incertain $m^{\mathcal{X}}_{\mathfrak{s}}$ fourni par $\mathfrak{s}$ par la fonction de masse définie par

$$m^{\mathcal{X}} = m^{\mathcal{X}}_{\mathfrak{s}} \textcircled{\cup}_{B \in \mathcal{B}} B_{\beta_B}. \tag{2.12}$$

La correction représentée par (2.12) généralise celle de l'affaiblissement contextuel basé sur un grossissement (1.22), au sens que $\mathcal{B}$ n'est pas nécessairement une partition de $\mathcal{X}$. L'interprétation qui en est donnée est toutefois bien différente.

Remarquons que l'équation (2.12) avait déjà été considérée par Mercier *et al.* [95], qui l'avaient nommée simplement affaiblissement contextuel et en avaient fourni une interprétation similaire à celle de l'affaiblissement contextuel basé sur un grossissement rappelée en section 1.6, la différence étant que l'ensemble $\mathcal{B}$ pouvait être arbitraire et non plus seulement une partition. Cette interprétation était en fait erronée comme nous l'avons montré dans [96].

3. Précisément, on obtient une version contextuelle du reniement.

Notons également qu'à l'aide des états n_B et de la procédure BBC, on peut même proposer une interprétation à une version plus générale de (2.12) où des β_B peuvent être tels $\beta_B > 1$ et non plus tels que $\beta_B \in [0, 1]$ (on généralise alors l'opération de *de-discounting* [25], qui est l'inverse de l'affaiblissement).

Les mensonges contextuels peuvent également être intéressants dans le cadre de la règle BBF. En particulier, j'ai montré que les α-conjonctions de Smets (section 1.4) sont un cas spécial de la règle BBF : elles correspondent à supposer que les deux sources sont sincères ou commettent le même mensonge ℓ_B, avec un poids particulier dépendant de α et de B [104]. On obtient une interprétation similaire pour les α-disjonctions.

2.3.3 Modèles connexes

Smets [132] a proposé un mécanisme de correction pour une source pertinente et partiellement sincère, qui est clairement étendu par notre approche. La procédure BBC est également plus générale que le modèle de sources partiellement pertinentes proposé par Haenni et Hartmann [61], comme expliqué dans [108].

Une extension de l'affaiblissement a été proposée dans [94], dans laquelle des méta-connaissances incertaines sur la source $\mathfrak{s}$ sont représentées par une fonction de masse Bayésienne $m^{\mathcal{H}}$ sur l'espace $\mathcal{H} = \{h_1, \ldots, h_v\}$ d'états possibles de la source. L'interprétation des états $h \in \mathcal{H}$ est donnée dans cette extension par des transformations $m_h^{\mathcal{X}}$ de $m_{\mathfrak{s}}^{\mathcal{X}}$: si la source fournit le témoignage incertain $m_{\mathfrak{s}}^{\mathcal{X}}$ et est dans l'état h, alors la connaissance induite à propos de $\boldsymbol{x}$ est représentée par $m_h^{\mathcal{X}}$. On montre que cette extension et la procédure BBC coïncident si $m_h^{\mathcal{X}}$ est définie à partir de $m_{\mathfrak{s}}^{\mathcal{X}}$ en utilisant les fonctions multivoques Γ_A comme suit

$$m_h^{\mathcal{X}}(B) = \sum_{A:\Gamma_A(h)=B} m_{\mathfrak{s}}^{\mathcal{Y}}(A), \quad \forall B \subseteq \mathcal{X},$$

qui est la connaissance déduite sur $\mathcal{X}$ par la procédure BBC lorsque la source est supposée être dans l'état h. Cette extension est toutefois limitée au cas de méta-connaissances Bayésiennes et le passage au cas de méta-connaissances générales (non nécessairement Bayésiennes), comme permis par la procédure BBC, pose des difficultés [4].

Enfin, remarquons que des alternatives à la règle conjonctive, où l'intersection est remplacée par d'autres opérations ensemblistes, ont déjà été considérées. Toutefois, le modèle présenté dans ce chapitre est le premier à fournir une interprétation explicite à ces règles. De plus, l'idée d'utiliser des réseaux évidentiels pour retrouver la procédure BBC et la règle BBF est inspirée d'approches similaires [126, 60] utilisées pour retrouver l'affaiblissement et la règle disjonctive.

4. La généralisation naïve aux méta-connaissances non nécessairement Bayésiennes, de l'approche suivie dans [94] pour dériver cette extension, n'est valide que si les fonctions de masse m_h, $h \in \mathcal{H}$, sont indépendantes [126]. Or, cela n'est en général pas le cas. On peut donner comme exemple le cas où les états h concernent la sincérité des sources [132] – on obtient effectivement un résultat aberrant pour ces états [103].

2.4 Conclusion

Ce chapitre a présenté un résumé de mes travaux concernant une approche générale pour la correction et la fusion de fonctions de croyance. Cette approche intègre explicitement les méta-connaissances sur les sources, c'est-à-dire les connaissances à propos de leur qualité, cette qualité pouvant concerner différents aspects et notamment la pertinence et la sincérité. Cette approche inclut et étend en particulier l'affaiblissement de Shafer et sa version contextualisée, la règle non normalisée de Dempster et les α-junctions de Smets.

Elle est également à la base d'une nouvelle décomposition des fonctions de croyance, qui est le sujet du prochain chapitre.

Chapitre 3

Décomposition d'un témoignage

3.1 Introduction

La possibilité de pouvoir décomposer de manière unique un témoignage représenté par une fonction de croyance, en une fusion de témoignages élémentaires est une question fondamentale. Smets [127] a trouvé une nouvelle représentation d'une fonction de croyance – la fonction de poids w (1.16) – pour laquelle il a proposé une interprétation en termes de décomposition. Sa solution pour le problème de la décomposition est formellement élégante et également attrayante d'un point de vue intuitif. Elle a aussi eu un certain succès. Elle pose toutefois certaines questions, pour lesquelles les réponses avancées ne semblent pas encore totalement satisfaisantes (section 3.2). J'ai proposé dans [105] (disponible en Annexe C) une solution alternative pour décomposer toute fonction de croyance (résumée en section 3.3), ainsi qu'une lecture totalement différente de la fonction de poids w (section 3.4). Ma proposition, surtout concernant l'interprétation de la fonction w, est peut-être moins attrayante que celle de Smets, mais elle pose aussi moins de questions d'ordre théorique car elle repose sur des notions plus classiques.

3.2 Revue critique de la solution de Smets

L'examen de la solution de Smets ainsi que ma proposition pour la décomposition des fonctions de croyance et l'interprétation de la fonction w, reposent sur un cas particulier très simple du modèle présenté au chapitre 2, où les méta-connaissances sur les sources concernent leur fiabilité entendue au sens classique, c'est-à-dire n'incluant pas la dimension supplémentaire qu'est la sincérité, et où le témoignage de chaque source est un sous-ensemble du domaine du paramètre d'intérêt. Pour ce cas particulier, j'utilise une notation qui s'avère pratique et qui est présentée à la section 3.2.1.

3.2.1 Sources partiellement fiables : notation

Soit une source $\mathfrak{s}_1$ informant à propos de la valeur d'un paramètre $\boldsymbol{x}$ défini sur un domaine $\mathcal{X} = \{x_1, \ldots, x_n\}$, sous la forme d'une information $\boldsymbol{x} \in A_{\mathfrak{s}_1}$ avec $A_{\mathfrak{s}_1} \subseteq \mathcal{X}$. De plus, on suppose que cette source peut être seulement dans un des deux états : fiable ou non fiable. Soit R_1 la variable désignant la fiabilité de $\mathfrak{s}_1$ et définie sur le domaine $\mathcal{R}_1 = \{0, 1\}$ où 0 signifie que $\mathfrak{s}_1$ est fiable et 1 signifie que $\mathfrak{s}_1$ n'est pas fiable. La fiabilité de la source peut être encodée par une fonction multivoque Γ_1 de

$\mathcal{R}_1$ vers $\mathcal{X}$ telle que

$$\Gamma_1(0) = A_{\mathfrak{s}_1}, \tag{3.1}$$
$$\Gamma_1(1) = \mathcal{X}. \tag{3.2}$$

$\Gamma_1(k_1)$ donne l'interprétation du témoignage $\boldsymbol{x} \in A_{\mathfrak{s}_1}$ dans chaque configuration $k_1 \in \mathcal{R}_1$ de $\mathfrak{s}_1$.

Si $\mathfrak{s}_1$ est supposée fiable avec une probabilité $1 - \pi_1$ et non fiable avec une probabilité π_1, $\pi_1 \in [0, 1]$, alors la connaissance induite à propos de $\boldsymbol{x}$ par l'information que $\mathfrak{s}_1$ a transmise est représentée par la fonction de masse simple [1] :

$$\begin{aligned} m(A_{\mathfrak{s}_1}) &= 1 - \pi_1, \\ m(\mathcal{X}) &= \pi_1, \end{aligned} \tag{3.3}$$

également notée plus simplement $A^{\pi_1}_{\mathfrak{s}_1}$.

Considérons maintenant le cas où N sources $\mathfrak{s}_i$, $i = 1, \ldots, N$, informent à propos de $\boldsymbol{x}$ sous la forme des témoignages respectifs $\boldsymbol{x} \in A_{\mathfrak{s}_i}$, avec $A_{\mathfrak{s}_i} \subseteq \mathcal{X}$, $i = 1, \ldots, N$, et que chacune de ces sources peut être soit fiable soit non fiable. Soit R_i la variable désignant la fiabilité de $\mathfrak{s}_i$ et définie sur le domaine $\mathcal{R}_i = \{0, 1\}$, $i = 1, \ldots, N$. Soit Γ_i la fonction multivoque de $\mathcal{R}_i$ vers $\mathcal{X}$ encodant la fiabilité de $\mathfrak{s}_i$ et définie de manière similaire à (3.1) – (3.2) en remplaçant 1 par i. L'ensemble des états joints élémentaires sur la fiabilité de ces sources est $\mathcal{R}_{1:N} := \times_{i=1}^{N} \mathcal{R}_i$. À tout état $(k_1, \ldots, k_N) \in \mathcal{R}_{1:N}$, on associe le nombre k, $1 \leq k \leq 2^N$, tel que

$$k = 1 + \sum_{i=1}^{N} k_i 2^{i-1}, \tag{3.4}$$

i.e., on a $k \leftrightarrow (k_1, \ldots, k_N) \in \mathcal{R}_{1:N}$. De plus, pour tout $k \leftrightarrow (k_1, \ldots, k_N) \in \mathcal{R}_{1:N}$ on définit la fonction multivoque Γ de $\mathcal{R}_{1:N}$ vers $\mathcal{X}$ par

$$\Gamma(k) = \bigcap_{i=1}^{N} \Gamma_i(k_i). \tag{3.5}$$

$\Gamma(k)$ représente l'information déduite sur $\mathcal{X}$ à partir des témoignages $(A_{\mathfrak{s}_1}, \ldots, A_{\mathfrak{s}_N})$ fournis par $\mathfrak{s}_1, \ldots, \mathfrak{s}_N$, quand elles sont dans les états $(k_1, \ldots, k_N)$.

Si à chaque état joint $k \leftrightarrow (k_1, \ldots, k_N) \in \mathcal{R}_{1:N}$ est alloué une probabilité $p_k = p_{k_1,\ldots,k_N}$ tel que $\sum_{k=1}^{2^N} p_k = 1$, alors la connaissance sur $\mathcal{X}$ est donnée par l'équation :

$$m(B) = \sum_{k:\Gamma(k)=B} p_k, \quad \forall B \subseteq \mathcal{X}. \tag{3.6}$$

De plus, si pour chaque $k \in \mathcal{R}_{1:N}$ la probabilité p_k que les sources soient dans

1. Dans ce chapitre, toutes les fonctions de masse considérées sont définies sur $\mathcal{X}$, on peut donc omettre l'exposant $^{\mathcal{X}}$ et écrire simplement m au lieu $m^{\mathcal{X}}$.

l'état joint $(k_1, \ldots, k_N)$ est égale au produit des probabilités marginales des états individuels k_i, $i = 1, \ldots, N$, *i.e.*, les probabilités p_k satisfont

$$p_k \quad = \quad \prod_{i=1}^{N} (1 - \pi_i)^{1-k_i} \pi_i^{k_i}, \quad \forall k \in \mathcal{R}_{1:N}, \tag{3.7}$$

avec k_i, $i = 1, \ldots, N$, les termes de l'expansion binaire (3.4) de k, et π_i la probabilité marginale que $\mathfrak{s}_i$ est non fiable, *i.e.*,

$$\pi_i \quad = \quad \sum_{k:k_i=1} p_k, \tag{3.8}$$

alors les sources sont méta-indépendantes et donc, en utilisant le Théorème 2.1, on obtient que la fonction de masse (3.6) satisfait dans ce cas :

$$m \quad = \quad \textcircled{\cap}_{i=1}^{N} A_{\mathfrak{s}_i}^{\pi_i}. \tag{3.9}$$

3.2.2 Discussion sur la solution de Smets

Il est clair que la décomposition (1.14) est une instance de (3.9) et peut donc être interprétée à l'aide du modèle de sources partiellement fiables de la section précédente. Précisément, toute fonction de masse séparable m sur le domaine $\mathcal{X} = \{x_1, \ldots, x_n\}$ avec fonction de poids w associée peut être vue comme issue des éléments d'information suivants :

- il y a $2^n - 1$ sources, avec $\mathfrak{s}_i$ la source fournissant l'information $\boldsymbol{x} \in A_i$, où A_i est le i^{e} sous-ensemble de $\mathcal{X}$ selon l'ordre binaire (section 1.2) ;
- chaque source $\mathfrak{s}_i$ est non fiable avec une probabilité marginale $w(A_i)$;
- les sources sont méta-indépendantes.

Cette interprétation de la décomposition (1.14) est totalement en ligne avec celle de Smets pour cette décomposition, comme l'on peut s'en rendre compte en comparant les éléments d'information ci-dessus avec ceux en jeu dans l'exemple 1.1. Le fait qu'elle suive l'approche de Dempster [14] pour induire une fonction de masse, la rend un peu plus lourde mais lui donne en contrepartie l'avantage de reposer sur des concepts bien connus et de rendre explicites les hypothèses faites sur les sources.

De par la similarité entre (1.14) et (1.15), on peut se demander si une extension directe de l'interprétation ci-dessus au cas de fonctions de masse non dogmatiques est possible. La réponse est négative puisque cela requerrait de permettre des probabilités marginales $w(A_i)$ supérieures à 1. On peut estimer que cela constitue un inconvénient de la décomposition de Smets que d'être incompatible avec le modèle de sources partiellement fiables, alors que ce modèle repose sur des concepts bien connus et semble être totalement en accord avec sa vision de sa décomposition.

Indépendamment de ce commentaire, la décomposition de Smets soulève des questions. Sa décomposition implique des fonctions de masse simples inverses A^w, $w > 1$ qui, rappelons-le, ne sont pas des fonctions de masse puisqu'elles ne satisfont

pas $m(A) \in [0,1]$ pour tout $A \subseteq \mathcal{X}$. Smets [127] a proposé d'interpréter ces fonctions comme des éléments d'information de la forme *une source fournit l'information* qu'il ne faut pas croire en A *et une fiabilité de* $1/w$ *est donnée à cette source*. Bien que Smets [127] fournisse une intuition sur la pertinence de cette sémantique, elle n'a pas de définition opérationnelle, comme cela est nécessaire pour toute modélisation de l'incertitude (voir, par exemple, [46]). De plus, si l'on accepte l'existence de cette notion de "bonne raisons de *ne pas croire*" et sa représentation mathématique associée, et que l'on veut modéliser tout état de croyance impliquant des "bonnes raisons de croire" et "des bonnes raisons de ne pas croire" dans des propositions, alors il est nécessaire de considérer ce que Smets appelle des structures de croyances latentes [127], qui sont des couples de fonctions de croyance (représentant respectivement la confiance et la défiance) qui ne peuvent pas être réduites en général à une seule fonction de croyance. Par exemple, modifier seulement un peu, de 3/4 à 0.74, la fiabilité de la source dans le troisième élément d'information de l'exemple 1.1 ne donne plus une fonction de croyance.

Une théorie reposant sur les structures de croyance latentes et impliquant des fonctions de masse simples inverses va clairement au-delà d'une théorie basée seulement sur les fonctions de croyance. Une théorie opérationnelle et basée sur cet objet mathématique plus riche reste à proposer. Le travail très récent de Dubois *et al.* [38], qui réinterprète la notion de dette de croyance de Smets en termes de biais cognitif représentant un préjugé, mènera peut-être à une telle théorie.

En tout cas, que la notion de dette de croyance représentée par des fonctions de masse simples inverses trouve ou non une existence propre, elle n'est pas strictement nécessaire pour résoudre les problèmes de la decomposition d'une fonction de croyance et de l'interprétation de la fonction de poids w puisque, comme cela est montré dans les sections 3.3 et 3.4 respectivement, on peut résoudre ces problèmes à l'aide d'autres notions plus classiques.

3.3 Une nouvelle décomposition

Dans cette section, une nouvelle décomposition des fonctions de croyance est présentée, d'abord en termes de témoignages fournis par des sources partiellement fiables (section 3.3.2) puis en termes d'une combinaison conjonctive de fonctions de masses simples (section 3.3.3), similairement à la décomposition de Smets. Cette décomposition est également comparée brièvement à celle de Smets (section 3.3.4). Elle se base sur un cas particulier du modèle décrit à la section 3.2.1 et qui permet d'induire toute fonction de masse sur $\mathcal{X}$ (section 3.3.1).

3.3.1 Fonction de masse induite par $|\mathcal{X}|$ témoignages

Soit m une fonction de masse sur $\mathcal{X}$. Considérons un cas particulier du modèle décrit à la section 3.2.1, où l'on a $N = |\mathcal{X}| = n$ sources et $A_{\mathfrak{s}_i} = \mathcal{X} \backslash \{x_i\} = \overline{\{x_i\}}$, $i = 1, \ldots, n$, et où la méta-connaissance incertaine sur les sources est telle que $p_k = m(A_k)$, $1 \leq k \leq 2^n$, avec A_k le k^{e} sous-ensemble de $\mathcal{X}$ selon l'ordre binaire. On

montre alors que la connaissance sur $\mathcal{X}$ (donnée par l'équation (3.6)) induite par la méta-connaissance incertaine sur les sources et leur témoignage, n'est rien d'autre que la fonction de masse m. En d'autres termes, toute fonction de masse sur $\mathcal{X}$ peut être induite par un certain cas particulier du modèle de sources partiellement fiables de la section 3.2.1. Cela est illustré par l'exemple 3.1.

Exemple 3.1 (Suite de l'exemple 1.1) *Soit m la fonction de masse de l'exemple 1.1 (définie par (1.19)). Elle peut être écrite de manière équivalente*

$$m(A_4) = m(A_6) = m(A_8) = 1/3,$$

puisque $A_4 = \{x_1, x_2\}$, $A_6 = \{x_1, x_3\}$ et $A_8 = \{x_1, x_2, x_3\}$.

Soit trois sources $\mathfrak{s}_i$ déclarant respectivement $\boldsymbol{x} \in \overline{\{x_i\}}$, $i = 1, 2, 3$. Soit $p_k = p_{k_1,k_2,k_3}$ la probabilité que les sources soient dans l'état $k \leftrightarrow (k_1, k_2, k_3) \in \mathcal{R}_{1:3}$ ($\mathcal{R}_{1:3} := \times_{i=1}^{3} \mathcal{R}_i$). Supposons que les probabilités $p_k = p_{k_1,k_2,k_3}$, $1 \leq k \leq 8$, valent :

$$\begin{aligned} p_4 &= p_{110} = m(A_4), \\ p_6 &= p_{101} = m(A_6), \\ p_8 &= p_{111} = m(A_8), \end{aligned}$$

et $p_k = 0$ pour $k = 1, 2, 3, 5, 7$.

Nous avons (selon (3.5))

$$\begin{aligned} \Gamma(4) &= \Gamma[(1,1,0)] \\ &= \Gamma_1(1) \bigcap \Gamma_2(1) \bigcap \Gamma_3(0) \\ &= \mathcal{X} \cap \mathcal{X} \cap \overline{\{x_3\}} \\ &= \{x_1, x_2\} \end{aligned}$$

et, de manière similaire, $\Gamma(6) = \{x_1, x_3\}$ et $\Gamma(8) = \{x_1, x_2, x_3\}$.

Lorsque les témoignages des sources sont interprétés en tenant compte de la méta-connaissance sur elles, la probabilité p_4 doit être allouée à $\boldsymbol{x} \in \Gamma(4) = \{x_1, x_2\} = A_4$, la probabilité p_6 à $\boldsymbol{x} \in A_6$ et la probabilité p_8 à $\boldsymbol{x} \in A_8$, i.e., *la fonction de masse m définie par (1.19) est retrouvée.*

La fonction de masse m peut donc être vue comme résultant de trois sources $\mathfrak{s}_i$ déclarant respectivement $\boldsymbol{x} \in \overline{\{x_i\}}$, $i = 1, 2, 3$, et pour lesquelles la connaissance sur leur fiabilité est la suivante : avec une probabilité p_4, $\mathfrak{s}_1$ et $\mathfrak{s}_2$ ne sont pas fiables et $\mathfrak{s}_3$ est fiable ; avec une probabilité p_6, $\mathfrak{s}_1$ et $\mathfrak{s}_3$ ne sont pas fiables et $\mathfrak{s}_2$ est fiable ; avec une probabilité p_8, les trois sources ne sont pas fiables.

Comme nous allons le voir dans la prochaine section, cette approche permettant d'induire toute fonction de masse est particulièrement utile lorsqu'elle est utilisée conjointement avec des résultats de Teugels [137] concernant la représentation de la distribution de Bernoulli multivariée.

3.3.2 Fiabilité individuelle et dépendances entre les fiabilités

Soit $\{R_i : i = 1, \ldots, n\}$ une séquence de variables aléatoires de Bernoulli prenant leurs valeurs dans $\mathcal{R}_i = \{0, 1\}$, $i = 1, \ldots, n$, *i.e.*, pour $i = 1, \ldots, n$,

$$P(R_i = 1) = \pi_i, \;\; P(R_i = 0) = \xi_i,$$

où $0 \leq \pi_i = 1 - \xi_i \leq 1$. Rappelons que $\mathbb{E}[R_i] = \pi_i$.

Considérons la distribution de Bernoulli multivariée (DBM)

$$p_{k_1,\ldots,k_n} := P(R_1 = k_1, \ldots, R_n = k_n) \tag{3.10}$$

où $k_i \in \{0, 1\}$, $i = 1, \ldots, n$. Cette DBM peut être représentée par le vecteur $\mathbf{p}$ contenant 2^n composants, avec $p_k = p_{k_1,\ldots,k_n}$ (en utilisant la correspondance $k \leftrightarrow (k_1, \ldots, k_n)$) son k^{e} composant, $1 \leq k \leq 2^n$.

Dans [137], Teugels considère deux vecteurs associées à la DBM (3.10) : le vecteur $\boldsymbol{\mu}$ (dit vecteur des *moments* [137]) et le vecteur $\boldsymbol{\sigma}$ (dit vecteur des *moments centrés* [137]) définis respectivement par

$$\begin{aligned} \boldsymbol{\mu} &= (\mu_1, \ldots, \mu_{2^n})^T, \\ \boldsymbol{\sigma} &= (\sigma_1, \ldots, \sigma_{2^n})^T, \end{aligned}$$

où, pour $1 \leq k \leq 2^n$,

$$\begin{aligned} \mu_k &= \mathbb{E}\left[\prod_{i=1}^{n} R_i^{k_i}\right], \\ \sigma_k &= \mathbb{E}\left[\prod_{i=1}^{n} (R_i - \pi_i)^{k_i}\right], \end{aligned}$$

avec k_i, $i = 1, \ldots, n$, les termes dans l'expansion binaire de k.

Remarquons que μ_k revient à la probabilité (marginale) que chaque variable dans $\{R_i : i, k_i = 1\}$ soit égale à 1. Cela ne doit pas être confondu avec p_k, qui est la probabilité que chaque variable dans $\{R_i : i, k_i = 1\}$ soit égale à 1 *et* chaque variable dans $\{R_i : i, k_i = 0\}$ soit égale à 0. De plus, notons que le vecteur $\boldsymbol{\mu}$ contient les moments d'ordre 1 des v.a. dans $\{R_i : i = 1, \ldots, n\}$, et il contient également les moments produits (ou, moments joints) [133, 117] de tous les sous-ensembles d'au moins deux v.a. dans $\{R_i : i = 1, \ldots, n\}$. Le vecteur $\boldsymbol{\sigma}$ contient quant à lui les moments centrés d'ordre 1 des v.a. dans $\{R_i : i = 1, \ldots, n\}$, qui sont égaux à 0, ainsi que les moments centrés produits (ou joints) [133, 117] de tous les sous-ensembles d'au moins deux v.a. dans $\{R_i : i = 1, \ldots, n\}$, et en particulier les covariances entre toute paire de v.a. dans $\{R_i : i = 1, \ldots, n\}$.

Teugels [137, Théorème 1] montre que des relations intéressantes existent entre

les vecteurs $\boldsymbol{\mu}$ et $\boldsymbol{\sigma}$ et le vecteur $\mathbf{p}$ de la DBM :

$$\mathbf{p} = \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \right) \boldsymbol{\mu} \,, \tag{3.11}$$

$$\boldsymbol{\mu} = \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right) \mathbf{p} \,, \tag{3.12}$$

$$\mathbf{p} = \begin{bmatrix} \xi_n & -1 \\ \pi_n & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} \xi_1 & -1 \\ \pi_1 & 1 \end{bmatrix} \boldsymbol{\sigma} \,, \tag{3.13}$$

$$\boldsymbol{\sigma} = \begin{bmatrix} 1 & 1 \\ -\pi_n & \xi_n \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 1 \\ -\pi_1 & \xi_1 \end{bmatrix} \mathbf{p} \,. \tag{3.14}$$

Comme cela est détaillé dans [137, Section 2.3 (i)], les deux représentations (3.11) et (3.13) contiennent autant $(2^n - 1)$ de paramètres que $\mathbf{p}$, puisque $\mu_1 = \sigma_1 = 1$. Plus précisément, dans (3.13), $2^n - n - 1$ paramètres sont donnés par $\boldsymbol{\sigma}$, puisque $\sigma_k = 0$ lorsque $k_1 + \cdots + k_n = 1$, et les n paramètres restants sont les π_i, $i = 1, \ldots, n$. Les composant non nuls de $\boldsymbol{\sigma}$ représentent les $2^n - n - 1$ dépendances possibles entre les v.a. appartenant à tout sous-ensemble (d'au moins deux) des v.a. dans $\{R_i : i = 1, \ldots, n\}$; si ces v.a. sont toutes indépendantes, nous avons $\boldsymbol{\sigma} = \mathbf{e}_1$. D'ailleurs, Teugels [137] appelle également $\boldsymbol{\sigma}$ le vecteur des *dépendances* – on trouvera une application dans [138] des moments centrés stockés dans $\boldsymbol{\sigma}$, où ils sont utilisés pour formuler des tests d'hypothèses à propos de l'interaction entre des traitements par divers types de médicaments.

On peut remarquer que dans l'approche présentée à la section 3.3.1 permettant d'induire toute fonction de masse m à partir de n sources fournissant les témoignages $\boldsymbol{x} \in \overline{\{x_i\}}$, $i = 1, \ldots, n$, la méta-connaissance sur ces sources est représentée par la DBM (3.10) de vecteur $\mathbf{p}$ tel que $p_k = m(A_k)$, $1 \leq k \leq 2^n$. Cette remarque nous permet d'appliquer les résultats de Teugels [137] rappelés ci-dessus et en particulier de décomposer cette connaissance jointe sur la fiabilité des sources en composants plus élémentaires concernant à la fois la fiabilité individuelle de chaque source et les dépendances entre leur fiabilité. Plus précisément, on montre qu'il est possible de voir toute fonction de masse m comme issue des composants basiques suivants :

— des témoignages $\boldsymbol{x} \in \overline{\{x_i\}}$ fournis par des sources $\mathfrak{s}_i$, $i = 1, \ldots, n$;

— pour la source $\mathfrak{s}_i$, $i = 1, \ldots, n$, une connaissance sur sa fiabilité individuelle prenant la forme de l'espérance mathématique $\mathbb{E}[R_i] = \pi_i$ (qui est formellement égale à la probabilité marginale que $\mathfrak{s}_i$ ne soit pas fiable) ;

— pour le sous-ensemble de sources $\{\mathfrak{s}_i : i, k_i = 1\}$ associé (par $k \leftrightarrow (k_1, \ldots, k_n)$) à k, $1 \leq k \leq 2^n$,[2] une connaissance sur la dépendance entre leur fiabilité prenant la forme du moment centré $\mathbb{E}\left[\prod_{i=1}^{n}(R_i - \pi_i)^{k_i}\right] = \sigma_k$;

où π_i et σ_k sont obtenus par les équations

$$\pi_i = pl(x_i), \tag{3.15}$$

et

$$\sigma_k = \left(\begin{bmatrix} 1 & 0 \\ -q(\{x_n\}) & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ -q(\{x_1\}) & 1 \end{bmatrix} \mathbf{q}\right)(A_k), \tag{3.16}$$

avec pl et q les fonctions de contour et de communalité associées à m.

Ces éléments de base permettent de retrouver toute fonction de masse m car de telles connaissances sur les fiabilités individuelles des sources et sur les dépendances entre leur fiabilité, induisent un vecteur $\mathbf{p}$ de connaissance jointe sur leur fiabilité vérifiant $p_k = m(A_k)$, auquel cas m est induite des témoignages $\boldsymbol{x} \in \overline{\{x_i\}}$, $i = 1, \ldots, n$, fournis par les sources selon le résultat décrit à la section 3.3.1.

Cette nouvelle décomposition d'une fonction de masse est illustrée par l'exemple 3.2.

Exemple 3.2 (Suite de l'exemple 3.1) *Soit la fonction de masse m de l'exemple 3.1. On trouve en utilisant (3.15)*

$$\begin{aligned} \pi_1 &= 1, \\ \pi_2 &= 2/3, \\ \pi_3 &= 2/3, \end{aligned}$$

De plus on obtient par (3.16)[3]

$$\begin{aligned} \sigma_4 &= 0, \\ \sigma_6 &= 0, \\ \sigma_7 &= -1/9, \\ \sigma_8 &= 0. \end{aligned}$$

En d'autres termes, la fonction de masse m peut être vue comme issue des éléments d'informations suivants :

— *Il y a trois sources $\mathfrak{s}_i$, chacune déclarant respectivement $\boldsymbol{x} \in \overline{\{x_i\}}$, $i = 1, 2, 3$;*

— *$\mathfrak{s}_1$, $\mathfrak{s}_2$ et $\mathfrak{s}_3$ sont supposées non fiables avec les probabilités (marginales) 1, $2/3$ et $2/3$ respectivement ;*

2. En réalité, il n'y a pas besoin de considérer les cas $k = 1$ et k tel que $k_1 + \cdots + k_n = 1$, qui vérifient nécessairement $\sigma_1 = 1$ et $\sigma_k = 0$ et qui correspondent respectivement à un sous-ensemble de sources vide et aux sous-ensembles contenant une seule source. Ces cas sont toutefois inclus ici pour éviter des subtilités techniques ultérieures qui nuiraient à la présentation.

3. Il n'y a pas besoin de considérer σ_1, σ_2, σ_3 et σ_5, cf note de bas de page précédente.

— *La covariance entre les fiabilités de $\mathfrak{s}_2$ et $\mathfrak{s}_3$ vaut $-1/9$, et les covariances entre les fiabilités de $\mathfrak{s}_1$ et $\mathfrak{s}_2$, et de $\mathfrak{s}_1$ et $\mathfrak{s}_3$, ainsi que le moment centré entre les fiabilités de $\mathfrak{s}_1$, $\mathfrak{s}_2$ et $\mathfrak{s}_3$, sont nuls.*

Cela est clairement une décomposition différente de celle de Smets fournie pour la même fonction de masse dans l'exemple 1.1.

Notons que les démonstrations des équations (3.15) et (3.16) reposent, entre autres, sur l'égalité $\boldsymbol{\mu} = \mathbf{q}$, où $\boldsymbol{\mu}$ est le vecteur des moments issu de $\mathbf{p}$ par (3.12) avec $\mathbf{p}$ le vecteur représentant les méta-connaissances sur les sources tel que $p_k = m(A_k)$, $1 \leq k \leq 2^n$, et où $\mathbf{q}$ est le vecteur des communalités issue de la fonction de masse m par (1.7). Cette égalité, qui découle directement de $\mathbf{p} = \mathbf{m}$ et des équations (3.12) et (1.7), nous indique que la quantité $q(A_k)$ est égale à la valeur du moment μ_k entre les v.a. $\{R_i : i, k_i = 1\}$, représentant les fiabilités des sources $\{\mathfrak{s}_i : i, k_i = 1\}$. Plus simplement, la quantité $q(A_k)$ est égale à la probabilité marginale μ_k que chaque source dans $\{\mathfrak{s}_i : i, k_i = 1\}$ ne soit pas fiable.

3.3.3 Présentation en termes de fonctions de masse simples

L'approche générale proposée dans [32] pour la combinaison conjonctive de fonctions de masse, permet d'autres structures de dépendance entre elles que l'indépendance. Cette approche est brièvement rappelée ici. Soit n fonctions de masse $m_1, \ldots, m_n$, sur $\mathcal{X}$. Leur combinaison conjonctive résulte en la fonction de masse $m_\cap$ sur $\mathcal{X}$ obtenue par la procédure suivante [32] :

1. Une fonction de masse dite jointe $jm : \times_{i=1}^{n} 2^{\mathcal{X}} \to [0, 1]$ est construite, préservant $m_1, \ldots, m_n$, en tant que marginales, ce qui signifie que $\forall A_i \in \mathcal{F}_i$, avec $\mathcal{F}_i$ l'ensemble des ensembles focaux de m_i,

$$m_i(A_i) \quad = \sum_{A_1 \in \mathcal{F}_1, \ldots, A_{i-1} \in \mathcal{F}_{i-1}, A_{i+1} \in \mathcal{F}_{i+1}, \ldots, A_n \in \mathcal{F}_n} jm(A_1, \ldots, A_{i-1}, A_i, A_{i+1}, \ldots, A_n). \tag{3.17}$$

2. Chaque masse jointe $jm(A_1, \ldots, A_n)$ est allouée au sous-ensemble $\bigcap_{i=1}^{n} A_i$ dans la fonction de masse finale $m_\cap$, *i.e.*, pour tout $A \subseteq \mathcal{X}$

$$m_\cap(A) \quad = \sum_{\bigcap_{i=1}^{n} A_i = A} jm(A_1, \ldots, A_n).$$

La fonction jm inclut une représentation de la dépendance mutuelle entre les informations m_i [32]. En particulier, la combinaison de $m_1, \ldots, m_n$ par la règle conjonctive $\textcircled{\cap}$ est retrouvée pour $jm(A_1, \ldots, A_n) = m_1(A_1)m_2(A_2)\ldots m_n(A_n)$.

Considérons un cas particulier de cette approche où chaque fonction de masse m_i a seulement deux ensembles focaux (auquel cas elle est dite *binaire*), *i.e.*, $m_i(A_{i_1}) = \pi_i$ et $m_i(A_{i_0}) = 1 - \pi_i = \xi_i$ pour un $\pi_i \in [0, 1]$ et des $A_{i_0}, A_{i_1} \subseteq \mathcal{X}$. Dans ce cas, seuls les sous-ensembles $\mathbf{A}_k := (A_{1_{k_1}}, A_{2_{k_2}}, \ldots, A_{n_{k_n}}) \subseteq \times_{i=1}^{n} \mathcal{X}$, $1 \leq k \leq 2^n$ avec $k \leftrightarrow (k_1, \ldots, k_n) \in \{0, 1\}^n$, peuvent recevoir une masse jointe non nulle dans jm.

Soit **jm** le vecteur contenant 2^n composants avec $jm_k = jm(\mathbf{A}_k)$ son k^{e} composant. Soit $\boldsymbol{\sigma}$ le vecteur obtenu par

$$\begin{aligned}\boldsymbol{\sigma} &= \begin{bmatrix} 1 & 1 \\ -m_n(A_{n_1}) & m_n(A_{n_0}) \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 1 \\ -m_1(A_{1_1}) & m_1(A_{1_0}) \end{bmatrix} \mathbf{jm} \\ &= \begin{bmatrix} 1 & 1 \\ -\pi_n & \xi_n \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 1 \\ -\pi_1 & \xi_1 \end{bmatrix} \mathbf{jm}\,. \end{aligned} \tag{3.18}$$

Des résultats de [137] rappelés dans la section précédente, on obtient que la structure de dépendance encodée dans jm entre les fonctions de masse binaires m_i est caractérisée entièrement par le vecteur $\boldsymbol{\sigma}$ puisque toute fonction de masse jointe avec des marginales données est obtenue pour un unique vecteur $\boldsymbol{\sigma}$. En fait, on justifie même de regarder σ_k comme représentant la dépendance entre les fonctions de masse $\{m_i : i, k_i = 1\}$.

Puisque la combinaison conjonctive de n fonctions de masse binaires $m_1, \ldots, m_n$, étant donnée la structure de dépendance jm, est complètement déterminée par le vecteur $\boldsymbol{\sigma}$ associé à jm, alors elle peut être exprimée comme une règle de combinaison (conjonctive) paramétrée, de paramètre $\boldsymbol{\sigma}$ représentant la structure de dépendance. Formellement, soit $\mathcal{B}$ l'ensemble de toutes les fonctions de masse binaires sur $\mathcal{X}$ et $\mathcal{M}$ l'ensemble de toutes les fonctions de masse sur $\mathcal{X}$. On définit la règle de combinaison $\textcircled{\cap}_{\boldsymbol{\sigma}} : \mathcal{B}^n \to \mathcal{M}$ par

$$\textcircled{\cap}_{\boldsymbol{\sigma}}(m_1, \ldots, m_n) := m_{\cap}, \tag{3.19}$$

avec $m_\cap$ le résultat de la combinaison conjonctive des m_i étant donnée la structure de dépendance jm déterminée par le vecteur $\boldsymbol{\sigma}$. L'opération $\textcircled{\cap}_{\boldsymbol{\sigma}}(m_1, \ldots, m_n)$ est bien définie tant qu'il existe une fonction de masse jointe jm compatible avec le vecteur $\boldsymbol{\sigma}$ et les fonctions de masses binaires m_i. De plus, nous pouvons remarquer que la règle conjonctive est un cas particulier de cette règle retrouvée pour $\boldsymbol{\sigma} = \mathbf{e}_1$:

$$\textcircled{\cap}_{\mathbf{e}_1}(m_1, \ldots, m_n) = m_1 \textcircled{\cap} \ldots \textcircled{\cap} m_n.$$

Lorsque $A_{i_1} = \mathcal{X}$ pour $i = 1, \ldots, n$, *i.e.*, les fonctions de masse m_i sont simples ($m_i = A_{i_0}^{\pi_i}$), leur combinaison par la règle $\textcircled{\cap}_{\boldsymbol{\sigma}}$ est notée $\textcircled{\cap}_{\boldsymbol{\sigma}}\left(A_{1_0}^{\pi_1}, \ldots, A_{n_0}^{\pi_n}\right)$.

On montre que la fonction de masse (3.6) représentant l'interprétation de témoignages élémentaires $(A_{\mathfrak{s}_1}, \ldots, A_{\mathfrak{s}_N})$ émis par des sources de fiabilités partielles décrites par la DBM $p_k, 1 < k < 2^N$, peut être exprimée sous forme d'une combinaison $\textcircled{\cap}_{\boldsymbol{\sigma}}$ de N fonctions de masse simples $A_{\mathfrak{s}_i}^{\pi_i}$, avec π_i la fiabilité marginale de $\mathfrak{s}_i$ (obtenue par (3.8)) et $\boldsymbol{\sigma}$ le vecteur des moments centrés de cette DBM représentant les dépendances entre les fiabilités des sources [105, Theorem 2]. Ce résultat couplé avec ceux de la section précédente permettent d'obtenir le théorème suivant :

Théorème 3.1 *Toute fonction de masse m définie sur le domaine $\mathcal{X} = \{x_1, \ldots, x_n\}$*

satisfait

$$m = \textcircled{\cap}_{\boldsymbol{\sigma}} \left(\overline{\{x_1\}}^{\pi_1}, \ldots, \overline{\{x_n\}}^{\pi_n} \right), \tag{3.20}$$

avec π_i, $i = 1, \ldots, n$ *et* $\boldsymbol{\sigma}$ *définis par* (3.15) *et* (3.16) *respectivement.*

Le théorème 3.1 montre que toute fonction de masse peut se décomposer de manière unique comme la combinaison conjonctive de $|\mathcal{X}|$ fonctions de masse simples ayant une certaine structure de dépendance. Cela est illustré par l'exemple 3.3.

Exemple 3.3 (Suite de l'exemple 3.2) *La fonction de masse* m *de l'exemple 3.1 résulte de la combinaison conjonctive des fonctions de masse simples* $\overline{\{x_1\}}^1$, $\overline{\{x_2\}}^{2/3}$ *et* $\overline{\{x_3\}}^{2/3}$ *ayant la structure de dépendance* $\boldsymbol{\sigma} = (1, 0, 0, 0, 0, 0, -\frac{1}{9}, 0)'$, i.e.,

$$m = \textcircled{\cap}_{(1,0,0,0,0,0,-\frac{1}{9},0)'} \left(\overline{\{x_1\}}^1, \overline{\{x_2\}}^{2/3}, \overline{\{x_3\}}^{2/3} \right).$$

La décomposition présentée dans la section précédente dit en substance que toute fonction de masse s'obtient de témoignages simples fournies par $|\mathcal{X}|$ sources, de connaissances sur la fiabilité individuelle (marginale) de chaque source et de connaissances sur la dépendance entre leur fiabilité. Le théorème 3.1 offre une perspective intéressante sur cette décomposition : toute fonction de masse s'obtient en corrigeant (affaiblissant) d'abord chacun de ces témoignages simples selon la fiabilité marginale de sa source émettrice, puis en combinant selon la structure de dépendance $\boldsymbol{\sigma}$ ces témoignages corrigés, la structure $\boldsymbol{\sigma}$ n'étant rien d'autre que la méta-dépendance supposée des sources.

3.3.4 Comparaison avec la décomposition de Smets

Cette nouvelle décomposition des fonctions de croyance ressemble à celle de Smets en ce qu'elles voient toutes les deux une fonction de croyance comme résultant de sources partiellement fiables fournissant des témoignages simples. D'ailleurs, elles coïncident dans le cas particulier suivant.

Soit m une fonction de masse telle que sa décomposition selon le théorème 3.1 vérifie $\boldsymbol{\sigma} = \mathbf{e}_1$. On a alors

$$\begin{aligned} m &= \textcircled{\cap}_{\mathbf{e}_1} \left(\overline{\{x_1\}}^{\pi_1}, \ldots, \overline{\{x_n\}}^{\pi_n} \right), \\ &= \textcircled{\cap}_{i=1}^{n} \overline{\{x_i\}}^{\pi_i}. \end{aligned} \tag{3.21}$$

Les fonctions de masse satisfaisant (3.21) seront appelées $\mathbf{e}_1$-*séparables*.

La décomposition de Smets pour m donne :

$$m = \textcircled{\cap}_{i=1}^{n} \overline{\{x_i\}}^{w(\overline{\{x_i\}})},$$

avec $w(\overline{\{x_i\}}) = \pi_i$, $i = 1, \ldots, n$. Donc, selon les deux décompositions, toute fonction de masse $\mathbf{e}_1$-séparable est obtenue de la combinaison conjonctive de $|\mathcal{X}|$ éléments d'information indépendants représentés par les fonctions de masse simples $\overline{\{x_i\}}^{\pi_i}$.

Notons cependant que l'on peut montrer que ces deux décompositions ne coïncident pas en général pour les fonctions de masse séparables.

On peut aussi relever que ces décompositions sont semblables en ce qu'elles commutent toutes les deux avec la règle conjonctive (cf [105, Section 4.3 (ii)]).

De manière générale, la décomposition de Smets et la décomposition présentée dans ce chapitre diffèrent toutefois significativement sur divers aspects. Premièrement, la décomposition de Smets met en jeu 2^n-1 sources méta-indépendantes alors que la nouvelle décomposition implique n sources généralement méta-dépendantes. En termes d'éléments d'information élémentaires, cela signifie que la nouvelle solution décompose une fonction de masse en une combinaison conjonctive de n fonctions de masse simples ayant une structure de dépendance, alors que la solution de Smets la décompose en une combinaison conjonctive de 2^n-1 fonctions de masse simples (généralisées) indépendantes. Deuxièmement, la nouvelle décomposition peut être obtenue pour toute fonction de masse, alors que celle de Smets est restreinte aux fonctions de masse non dogmatiques. Troisièmement, la nouvelle décomposition repose sur des concepts plus établis – en particulier les moments et moments centrés de variables de Bernoulli – que celle de Smets.

3.4 Une perspective alternative sur la fonction de poids

Au lieu d'utiliser la fonction de poids w, la décomposition de Smets (1.15) peut être présentée de manière équivalente en utilisant la fonction $s : 2^{\mathcal{X}}\backslash\{\mathcal{X}\} \to (-\infty, +\infty)$ telle que $s(A) = -\ln w(A)$ pour tout $A \subset \mathcal{X}$ (voir, *e.g.*, [59]). Les fonctions de masse simples correspondent alors au cas $s(A) \geq 0$ et les fonctions de masse simples inverses au cas $s(A) < 0$. La fonction s est en fait la fonction que Shafer a utilisé originellement dans son livre [123] pour présenter la décomposition des fonctions de masse séparables (1.14). Par conséquent et en suivant [59], $s(A)$, $A \subset \mathcal{X}$, peuvent être appelés les *poids de Shafer*.

Soit m une fonction de masse sur $\mathcal{X} = \{x_1, x_2\}$. Comme cela a été expliqué à la section 3.3.1, m peut être induite en considérant deux sources $\mathfrak{s}_1$ et $\mathfrak{s}_2$ fournissant respectivement les témoignages $\boldsymbol{x} \in \overline{\{x_1\}}$ et $\boldsymbol{x} \in \overline{\{x_2\}}$ et en supposant que les sources sont dans l'état $k \leftrightarrow (k_1, k_2)$, $(k_1, k_2) \in \mathcal{R}_1 \times \mathcal{R}_2$, avec une probabilité $p_k = P(R_1 = k_1, R_2 = k_2) := m(A_k)$, $1 \leq k \leq 4$.

Rappelons (cf [51, p. 28]) que *l'information mutuelle* $I(R_1 = 1; R_2 = 1)$ entre les événements $R_1 = 1$ et $R_2 = 1$ est donnée par l'équation

$$I(R_1 = 1; R_2 = 1) \quad = \quad \ln \frac{P(R_1 = 1, R_2 = 1)}{P(R_1 = 1)P(R_2 = 1)}.$$

De plus, *l'information propre conditionnelle* (*conditional self information* en anglais) $I(R_1 = 1|R_2 = 1)$ de l'événement $R_1 = 1$ étant donné l'événement $R_2 = 1$ est [51, p. 36]

$$I(R_1 = 1|R_2 = 1) \quad = \quad -\ln \frac{P(R_1 = 1, R_2 = 1)}{P(R_2 = 1)},$$

et l'information conditionnelle propre $I(R_2 = 1|R_1 = 1)$ de l'événement $R_2 = 1$ étant donné l'événement $R_1 = 1$ est

$$I(R_2 = 1|R_1 = 1) = -\ln \frac{P(R_1 = 1, R_2 = 1)}{P(R_1 = 1)}.$$

On montre que

$$\begin{aligned} s(\emptyset) &= I(R_1 = 1; R_2 = 1), \\ s(\{x_1\}) &= I(R_2 = 1|R_1 = 1), \\ s(\{x_2\}) &= I(R_1 = 1|R_2 = 1), \end{aligned}$$

i.e, $s(\emptyset)$ est égal à l'information mutuelle que les deux sources sous-jacentes à m ne soient pas fiables, $s(\{x_1\})$ est égal à l'information propre conditionnelle que la seconde source ne soit pas fiable sachant que la première source ne l'est pas non plus, et $s(\{x_2\})$ est égal à l'information propre conditionnelle que la première source ne soit pas fiable sachant que la seconde source ne l'est pas non plus.

Pour toute fonction de masse m sur $\mathcal{X} = \{x_1, x_2\}$, on a $s(\{x_i\}) \geq 0$, $i = 1, 2$, et $s(\emptyset) \in (-\infty, +\infty)$. En d'autres termes lorsque $\mathcal{X}$ est binaire on peut avoir $s(A) < 0$ – un cas de dette de croyance selon l'interprétation de Smets – seulement pour $A = \emptyset$. La nouvelle perspective ci-dessus concernant s donne une sémantique complètement différente à $s(\emptyset) < 0$ de celle de dette de croyance, et précisément une sémantique bien établie qui est celle d'information mutuelle.

Lorsque $\mathcal{X} = \{x_1, x_2, x_3\}$, toute fonction de masse m sur $\mathcal{X}$ peut être induite par trois sources fournissant les témoignages $\boldsymbol{x} \in \overline{\{x_i\}}$, $i = 1, \ldots, n$, et tel que la probabilité qu'elles soient dans l'état $k \in \mathcal{R}_{1:3}$ est $p_k = P(R_1 = k_1, R_2 = k_2, R_3 = k_3) := m(A_k)$, $1 \leq k \leq 8$. On obtient

$$\begin{aligned} s(\emptyset) &= I(R_1 = 1; R_2 = 1; R_3 = 1), \\ s(\{x_1\}) &= I(R_2 = 1; R_3 = 1|R_1 = 1), \\ s(\{x_1, x_2\}) &= I(R_3 = 1|R_1 = 1, R_2 = 1), \end{aligned}$$

où $I(R_1 = 1; R_2 = 1; R_3 = 1)$ est l'information mutuelle entre les événements $R_1 = 1$, $R_2 = 1$ et $R_3 = 1$; $I(R_2 = 1; R_3 = 1|R_1 = 1)$ est l'information mutuelle conditionnelle entre les événements $R_2 = 1$ et $R_3 = 1$ étant donné l'événement $R_1 = 1$; et $I(R_3 = 1|R_1 = 1, R_2 = 1)$ est l'information propre conditionnelle de l'événement $R_3 = 1$ étant donné les événements $R_1 = 1$ et $R_2 = 1$. On obtient également

$$\begin{aligned} s(\{x_2\}) &= I(R_1 = 1; R_3 = 1|R_2 = 1), \\ s(\{x_3\}) &= I(R_1 = 1; R_2 = 1|R_3 = 1), \\ s(\{x_1, x_3\}) &= I(R_2 = 1|R_1 = 1, R_3 = 1), \\ s(\{x_2, x_3\}) &= I(R_1 = 1|R_2 = 1, R_3 = 1). \end{aligned}$$

En conséquence, par exemple, $s(\{x_2\})$ est égal à l'information mutuelle conditionnelle que les première et troisième sources sous-jacentes à m ne soient pas fiables sachant que la seconde ne l'est pas non plus.

Plus généralement, on montre ([105, Theorem 4]) que ce type d'interprétation pour la fonction s en termes de mesures d'information peut être obtenue pour toute taille du cadre de discernement $\mathcal{X}$. Cette interprétation est peut-être moins attrayante que celle de Smets, et en particulier elle n'est pas liée au concept de décomposition, mais elle repose sur des notions plus établies.

Pour finir, soulignons que la question de la décomposition *disjonctive* de toute fonction de croyance a été également abordée dans mes travaux. J'ai montré ([105, Theorem 6]) que l'on peut obtenir une contrepartie disjonctive du théorème 3.1. J'ai également fourni ([105, Theorem 5]) une interprétation en termes de mesures d'information, similaire à celle donnée ci-dessus pour s, pour les poids de Shafer *disjonctifs* définis par $r(A) = -\ln v(A)$, pour tout $A \subseteq \mathcal{X}, A \neq \emptyset$, où v est la fonction de poids disjonctifs (cf section 1.5), et pour lesquels il n'existait pas d'interprétation.

3.5 Conclusion

Ce chapitre a présenté une synthèse de mes résultats concernant la décomposition des fonctions de croyance. J'ai montré que toute fonction de masse définie sur un cadre $\mathcal{X}$ résulte de la fusion de $|\mathcal{X}|$ témoignages élémentaires partiellement fiables ayant une structure de dépendance, ou encore elle peut être exprimée de manière unique comme le résultat de la combinaison conjonctive de $|\mathcal{X}|$ fonctions de masse simples ayant une structure de dépendance et caractérisées par la fonction de contour. J'ai aussi montré que l'on peut donner à la fonction de poids une interprétation en termes de mesures d'information. Ces résultats constituent une alternative à la proposition de Smets pour la décomposition des fonctions de croyance et l'interprétation de la fonction w, dont la faiblesse principale est qu'elle repose sur une notion qui mériteraient de plus amples justifications.

Le modèle présenté au chapitre 2 et sur lequel se basent mes résultats sur la décomposition, permet d'interpréter des témoignages en fonction de méta-connaissances sur les sources fournissant ces témoignages. Dans le chapitre suivant, la question importante de la détermination de ces méta-connaissances est abordée.

Chapitre 4

Détermination des méta-connaissances

4.1 Introduction

Lorsque des témoignages à propos d'un paramètre $\boldsymbol{x}$ sont reçus, il faut, pour pouvoir les interpréter, tenir compte de la qualité des sources émettrices. Le modèle présenté au chapitre 2 donne un moyen de représenter des connaissances sur la qualité des sources et d'exploiter ces connaissances pour interpréter leur témoignage. Il n'indique toutefois pas *quelles* méta-connaissances sur les sources utiliser pour interpréter les témoignages reçus. On peut distinguer deux situations par rapport à ce problème, qui appellent des solutions différentes. La première de ces situations est celle où l'on dispose d'une expérience préalable importante des sources émettrices, auquel cas cette expérience va pouvoir être exploitée pour se projeter sur leur qualité. La seconde situation est celle où l'expérience antérieure est faible et où donc les seules informations significatives disponibles sont les témoignages reçus. Dans ce second cas, le choix des méta-connaissances à utiliser doit reposer sur d'autres considérations telles que des principes généraux.

Dans mes travaux, j'ai abordé ces deux situations et mes contributions vis-à-vis d'elles sont synthétisées aux sections 4.2 et 4.3 respectivement.

4.2 À partir d'une expérience préalable

L'expérience préalable peut prendre diverses formes. Il est possible, par exemple, de disposer de données mettant en regard ce que les sources ont dit (ou ce que *la* source a dit, dans le cas d'un problème de correction d'information) et les vraies réponses qui étaient attendues dans des situations passées. Dans ce cas, une solution consiste à sélectionner la méta-connaissance parmi un ensemble de méta-connaissances candidates telle que, sur ces données, les témoignages des sources fusionnés (ou corrigés, dans le cas d'une seule source) selon cette méta-connaissance sont les plus proches des réponses attendues. Une autre solution consiste à exploiter ces données pour produire une estimation (au sens statistique) de la qualité des sources qui peut ensuite être utilisée pour prédire leur qualité vis-à-vis des témoignages en jeu.

J'ai étudié ces deux solutions dans mes travaux. Précisément, j'ai considéré la première solution pour le cas où les méta-connaissances candidates sont celles associées aux mécanismes de corrections contextuelles évoqués à la section 2.3.2.2, et

Tableau 4.1 – Sorties de deux capteurs, notés $\mathfrak{s}_1$ et $\mathfrak{s}_2$, à propos de la classe de 4 objets pouvant être des avions (x_1), des hélicoptères (x_2) ou des rockets (x_3). Les données viennent de [50, Table 1].

		x_1	x_2	x_3	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$	$\mathcal{X}$	Vraie valeur
Capteur 1	$m_{\mathfrak{s}_1}\{o_1\}$	0	0	0.5	0	0	0.3	0.2	x_1
	$m_{\mathfrak{s}_1}\{o_2\}$	0	0.5	0.2	0	0	0	0.3	x_2
	$m_{\mathfrak{s}_1}\{o_3\}$	0	0.4	0	0	0.6	0	0	x_1
	$m_{\mathfrak{s}_1}\{o_4\}$	0	0	0	0	0.6	0.4	0	x_3
Capteur 2	$m_{\mathfrak{s}_2}\{o_1\}$	0	0	0	0.7	0	0	0.3	x_1
	$m_{\mathfrak{s}_2}\{o_2\}$	0.3	0	0	0.4	0	0	0.3	x_2
	$m_{\mathfrak{s}_2}\{o_3\}$	0.2	0	0	0	0	0.6	0.2	x_1
	$m_{\mathfrak{s}_2}\{o_4\}$	0	0	0	0	0	1	0	x_3

j'ai utilisé l'autre solution pour établir la connaissance sur la sincérité de sources. Ces études sont synthétisées aux sections 4.2.1 et 4.2.2 respectivement.

Notons avant cela que l'expérience préalable peut prendre d'autres formes et en particulier celle de dires d'experts. Dans [112], nous nous sommes intéressés au cas où ces dires correspondaient à des comparaisons de sources du point de vue de leur pertinence, ces sources étant caractérisées par des attributs. Nous avons proposé une approche basée sur l'intégrale de Choquet exploitant ces connaissances et permettant d'apprécier le degré de pertinence d'une source étant donné l'observation des valeurs de ses attributs. De plus, nous avons appliqué cette approche au problème de l'évaluation de la fiabilité de comptes Twitter, ces comptes étant vus comme des sources caractérisées par différents attributs tels que les nombres de *retweets* et de fautes d'orthographe.

4.2.1 Sincérité par minimisation d'un critère d'erreur

Supposons un ensemble d'apprentissage contenant les témoignages d'une source (exprimés sous la forme de fonctions de masse) à propos de la vraie valeur d'un paramètre $\boldsymbol{x}$ d'intérêt défini sur un domaine $\mathcal{X} = \{x_1, \ldots, x_n\}$ pour L objets o_ℓ, $\ell \in \{1, \ldots, L\}$. Un exemple illustratif de tels ensembles d'apprentissage est donné dans le tableau 4.1 pour deux capteurs en charge de reconnaître des objets volants qui peuvent être des avions (x_1), hélicoptères (x_2) ou des rockets (x_3).

En s'inspirant d'un travail précédent [50] concernant l'apprentissage de la fiabilité d'une source, Mercier *et al.* [97] ont proposé une méthode pour déterminer, à

partir d'un tel ensemble d'apprentissage, le vecteur $\boldsymbol{\beta} = (\beta_B \in [0,1], B \in \mathcal{B})$ des degrés de fiabilité associés à l'affaiblissement contextuel basé sur un grossissement (cf section 1.6), étant donné une partition $\mathcal{B}$ de $\mathcal{X}$ fixée. Le vecteur $\boldsymbol{\beta}$ est choisi comme celui minimisant le critère d'erreur suivant entre les témoignages corrigés d'une source $\mathfrak{s}$ et la réalité

$$E_{pl}(\boldsymbol{\beta}) = \sum_{\ell=1}^{L} \sum_{i=1}^{n} (pl\{o_\ell\}(x_i) - \delta_{\ell,i})^2 \ , \tag{4.1}$$

où $\forall \ell \in \{1, \ldots, L\}$, $pl\{o_\ell\}$ est la fonction de contour associée à la fonction de masse $m\{o_\ell\}$ obtenue d'un affaiblissement contextuel basé sur un grossissement $\mathcal{B}$ du témoignage $m_{\mathfrak{s}}\{o_\ell\}$ selon le vecteur $\boldsymbol{\beta} = (\beta_B \in [0,1], B \in \mathcal{B})$ de degrés de fiabilité, et où $\delta_{\ell,i}$ est la variable binaire indiquant la vraie valeur du paramètre $\boldsymbol{x}$ pour l'objet o_ℓ de la manière suivante : $\forall i \in \{1, \ldots, n\}$, $\delta_{\ell,i} = 1$ si la valeur de $\boldsymbol{x}$ pour l'objet o_ℓ est x_i, et $\delta_{\ell,i} = 0$ sinon.

L'idée de cette méthode est de trouver les degrés de fiabilité associés à l'affaiblissement contextuel basé sur un grossissement qui vont en moyenne, après correction, amener les témoignages de la source le plus proche de la réalité. Notons que le problème de trouver la partition optimale $\mathcal{B}$ de $\mathcal{X}$ pour une source donnée, c'est-à-dire celle minimisant (4.1), restait ouvert dans [97].

Cette méthode peut être utilisée pour apprendre les méta-connaissances associées à d'autres mécanismes de correction et peut même être étendue, si l'on dispose de témoignages de plusieurs sources vis-à-vis des mêmes objets comme c'est le cas dans le tableau 4.1, à l'apprentissage des méta-connaissances associées à des mécanismes de fusion comme cela est illustré dans [97].

Dans [113, Section 8] (disponible en Annexe B), nous avons étudié son application aux mécanismes de corrections contextuelles évoqués à la section 2.3.2.2, c'est-à-dire à l'affaiblissement contextuel (AC) (qui généralise l'affaiblissement contextuel basé sur un grossissement au sens où $\mathcal{B}$ ne doit pas nécessairement être une partition de $\mathcal{X}$), au renforcement contextuel (RC) et au reniement contextuel (NC).

On montre que pour chacun de ces mécanismes – qui correspondent à des connaissances particulières sur la sincérité contextuelle d'une source – il existe un ensemble $\mathcal{B}$ optimal : pour AC il s'agit de l'ensemble formé des singletons de $\mathcal{X}$, et pour RC et NC il s'agit de l'ensemble formé des complémentaires des singletons. De plus, on montre que pour chacun de ces mécanismes, la minimisation de (4.1) est un problème des moindres carrés, qui peut donc être résolu de manière efficace. Par exemple, pour AC, on a

$$E_{pl}(\boldsymbol{\beta}) = \|\boldsymbol{Q}\boldsymbol{\beta} - \boldsymbol{d}\|^2 \text{ avec } \boldsymbol{Q} = \begin{bmatrix} diag(\boldsymbol{pl}_1 - 1) \\ \vdots \\ diag(\boldsymbol{pl}_L - 1) \end{bmatrix} \text{ et } \boldsymbol{d} = \begin{bmatrix} \boldsymbol{\delta}_1 - 1 \\ \vdots \\ \boldsymbol{\delta}_L - 1 \end{bmatrix} , \tag{4.2}$$

pour $\boldsymbol{\beta} = (\beta_{\{x_i\}} \in [0,1], i \in \{1, \ldots, n\})$ et avec $diag(\boldsymbol{v})$ la matrice diagonale compo-

sée des éléments du vecteur $\boldsymbol{v}$ sur sa diagonale, $\boldsymbol{pl}_\ell = (pl_{\mathfrak{S}}\{o_\ell\}(x_1), \ldots, pl_{\mathfrak{S}}\{o_\ell\}(x_n))^T$, et $\boldsymbol{\delta}_\ell = (\delta_{\ell,1}, \ldots, \delta_{\ell,n})^T$. Remarquons que pour obtenir ces résultats, il a fallu trouver les expressions des fonctions de contour resultant de l'application de ces trois mécanismes de correction, ce qui n'était pas trivial pour NC (cf [113, Appendix C]) et qui nous a permis en outre de mettre en évidence leurs capacités respectives de correction d'une source au regard du critère (4.1) (cf [113, Section 8.5]).

L'exemple 4.1 donne un argument expérimental que suivre une telle approche afin de déterminer la méta-connaissance à utiliser pour interpréter les témoignages fournie par une source, produit des méta-connaissances valides au sens qu'elles permettent d'améliorer les performances de cette source. Dans cet exemple, le rôle de la source est joué par un classifieur et le paramètre d'intérêt $\boldsymbol{x}$ est la classe d'objets.

Exemple 4.1 *Considérons un problème de classification supervisé à* 5 *classes* ($\mathcal{X} = \{x_1, x_2, x_3, x_4, x_5\}$) *qui suivent* 5 *distributions normales à* 2 *variables de moyennes respectives* $\mu_{x_1} = (0,0)$, $\mu_{x_2} = (2,0)$, $\mu_{x_3} = (0,2)$, $\mu_{x_4} = (2,2)$, $\mu_{x_5} = (1,1)$ *avec pour matrice de covariance commune :*

$$\Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}.$$

On génère 1000 *instances pour chacune des* 5 *classes, illustrées par la figure 4.1.*

On considère pour ce problème le classifieur évidentiel des 3 *plus proches voisins (ev-*3*ppv) dont les paramètres sont obtenus selon les heuristiques données dans [16]. Un tel classifieur produit pour toute instance à classer, une fonction de masse sur* $\mathcal{X}$ *représentant son incertitude quant à la vraie classe de l'instance.*

Les 5000 *instances générées sont divisées en trois parties :*

1. *Le premier tiers constitue l'ensemble d'apprentissage du classifieur ev-*3*ppv* i.e.*, l'ensemble permettant d'apprendre les paramètres du classifieur et constituant l'ensemble des voisins permettant de classer les nouvelles instances ;*
2. *Le second tiers est utilisé pour apprendre, à partir des sorties du classifieur ev-*3*ppv pour les instances de ce tiers, les meilleures (au sens de* (4.1)*) méta-connaissances associées aux mécanismes AC, RC, et NC ;*
3. *Le dernier tiers forme l'ensemble de test.*

*Les performances du classifieur ev-*3*ppv seul et du classifieur ev-*3*ppv avec des corrections par AC, RC et NC sont données sur la figure 4.2 en utilisant des courbes ROC (les décisions ont été prises en utilisant la transformation* (1.46)*).*

Les classes x_2 *et* x_3*, étant globalement "disjointes" des autres (cf figure 4.1), sont bien reconnues par le classifieur ev-*3*pp (cf figures 4.2(b) et 4.2(c)). De plus, les corrections AC, RC et NC n'améliorent ni ne détériorent les performances vis-à-vis de ces classes.*

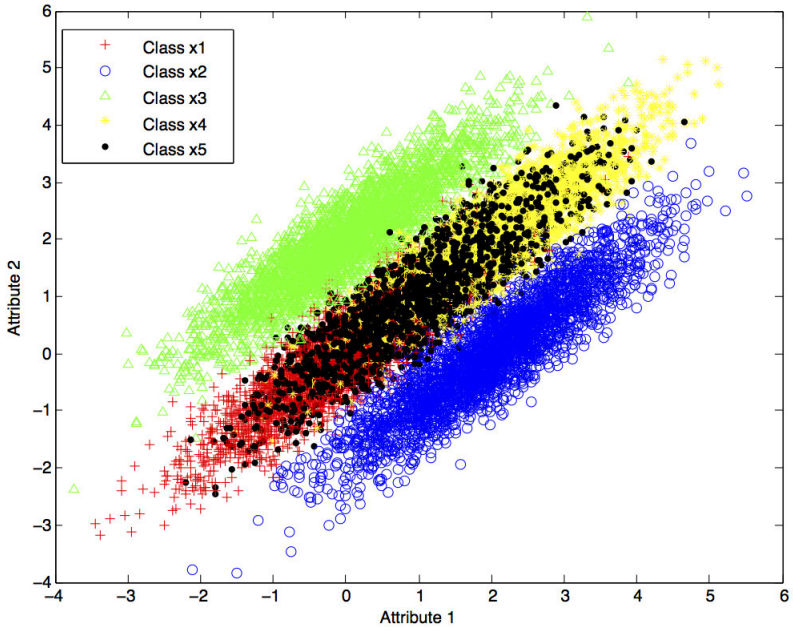


FIGURE 4.1 – Illustration des données générées pour un problème de classification à 5 classes avec 2 attributs.

Les classes x_1, x_4 et x_5 se recouvrent (cf figure 4.1) et ne sont pas si bien reconnues par le classifieur ev-3pp (cf figures 4.2(a), 4.2(d) et 4.2(e)). Par contre, les corrections AC, RC et NC parviennent à améliorer le classifieur dans ces cas plus difficiles.

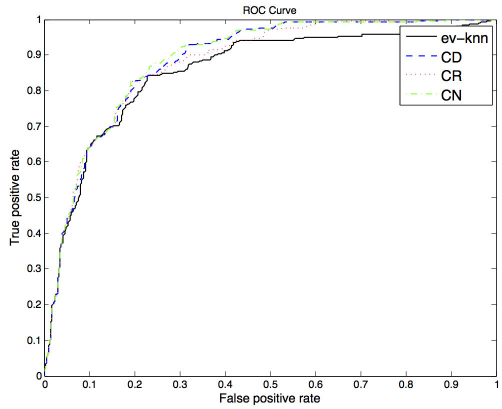
4.2.2 Sincérité par prédiction

Dans cette section, des façons d'établir la sincérité de sources par des approches basées sur la prédiction (cf section 1.8) sont discutées, pour deux situations différentes en termes de données d'apprentissage.

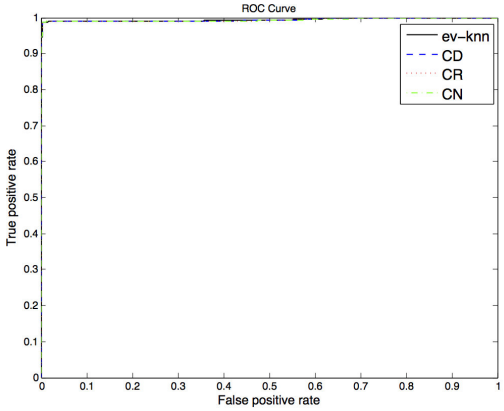
4.2.2.1 Exploitation du taux d'erreur

Considérons un cas particulier de l'ensemble d'apprentissage décrit à la section 4.2.1, où les témoignages d'une source $\mathfrak{s}$ ne contiennent pas d'incertitude, *i.e.*, pour chaque o_ℓ, $\ell \in \{1, \ldots, L\}$, il existe $x_i \in \mathcal{X}$ tel que $m_{\mathfrak{s}}\{o_\ell\}(\{x_i\}) = 1$. Dans ce cas, on peut compter le nombre e de témoignages incorrects, *i.e.*, "d'erreurs" commises par la source sur les L objets, où une erreur est commise pour un objet o_ℓ si $m_{\mathfrak{s}}\{o_\ell\}(\{x_i\}) = 1$ et $\delta_{\ell,i} = 0$. Le nombre de témoignages corrects est alors $L - e$.

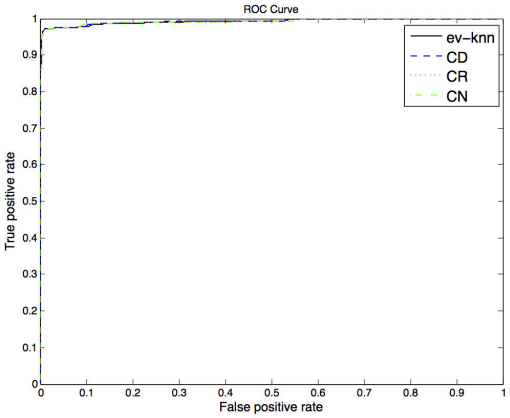
Il a été proposé dans [49] de considérer la valeur $\varepsilon := e/L$ comme la probabilité que $\mathfrak{s}$ ne soit pas pertinente (et $1-\varepsilon$ comme la probabilité qu'elle le soit) et d'utiliser cette méta-connaissance sur $\mathfrak{s}$ pour interpréter son prochain témoignage. On peut remarquer qu'une telle définition est celle que l'on obtient si :



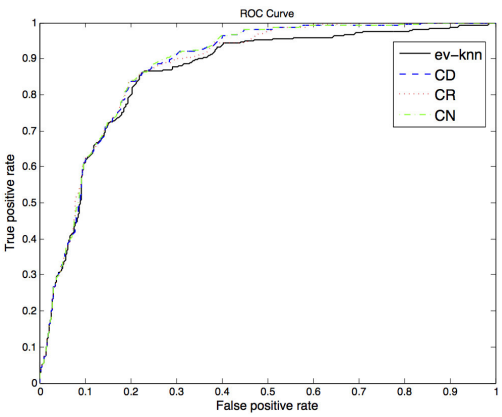
(a) Avec x_1 comme classe positive.



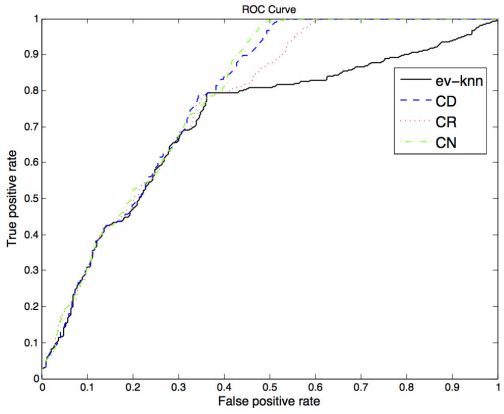
(b) Avec x_2 comme classe positive.



(c) Avec x_3 comme classe positive.



(d) Avec x_4 comme classe positive.



(e) Avec x_5 comme classe positive.

FIGURE 4.2 – Performances sous forme de courbes ROC pour le classifieur ev-3ppv seul et le classifieur ev-3ppv corrigé par AC, RC et NC (appelés, dans les figures, *ev-knn*, *CD*, *CR*, et *CN* respectivement).

1. on suppose que les erreurs sont des réalisations de l'état $\neg PE$ et les témoignages corrects des réalisations de l'état PE ;
2. on estime à partir de ces données par l'EMV la probabilité de pertinence de la source [1] ;
3. on effectue une prédiction sur sa pertinence selon cette estimation [82, Chapter 17].

Utiliser de cette manière les données d'apprentissage semble toutefois discutable puisqu'une source peut très bien avoir fourni une information correcte tout en ayant été dans l'état $\neg PE$ (*e.g.*, une montre cassée donne deux fois par jour la bonne heure). Il semble plus approprié, si l'on peut raisonnablement supposer que $\mathfrak{s}$ est pertinente, de considérer les erreurs comme des réalisations de la non-sincérité, *i.e.*, de l'état $\neg SI$ de la source, et les témoignages corrects comme des réalisations de sa sincérité SI. Dans ce cas, on peut alors utiliser des méthodes de prédiction (*e.g.*, celle basée sur l'EMV ou celle de Kanjanatarakul *et al.* rappelée à la section 1.8) pour établir la connaissance sur la sincérité de la source, à partir des valeurs e et L.

C'est ce que nous avons proposé dans [88] où nous avons en particulier étudier la prédiction de la sincérité par l'EMV [2], auquel cas la valeur ε devient la probabilité que $\mathfrak{s}$ ne soit pas sincère (et $1-\varepsilon$ est la probabilité qu'elle soit sincère). Précisément, nous avons montré, dans le cadre d'un problème de fusion de classifieurs et sur plusieurs jeux de données classiques du répertoire UCI [3], que d'aussi bonnes voire parfois de meilleures performances sont obtenues en corrigeant (reniant) la sortie de chaque classifieur selon sa sincérité prédite par l'EMV puis en combinant par la règle conjonctive les sorties corrigées de tous les classifieurs, plutôt qu'en fusionnant les décisions des classifieurs par le vote à la majorité.

4.2.2.2 Étalonnage

Dans les données d'apprentissage considérées jusqu'ici, les témoignages des sources prennent la forme de fonctions de masse sur $\mathcal{X}$. Toutefois, il se peut que les informations issues des sources prennent une autre forme. En particulier, lorsque $\mathcal{X}$ est binaire, *i.e.*, $\mathcal{X} = \{x_1, x_2\}$, une source peut retourner pour tout objet un nombre réel, que l'on appellera *score*, représentant sa confiance dans le fait que la vraie valeur du paramètre $\boldsymbol{x}$ pour l'objet est x_1 (et, par symétrie, sa confiance dans le fait que $\boldsymbol{x} = x_2$). Cela est une situation que l'on rencontre typiquement en classification binaire – où pour chaque objet, $\boldsymbol{x}$ correspond à sa classe – lorsque par exemple la source (classifieur) utilisée est un SVM. L'ensemble d'apprentissage pour une telle source est alors $\mathcal{L} = \{(s_1, \boldsymbol{\delta}_1), \ldots, (s_L, \boldsymbol{\delta}_L)\}$, où s_ℓ est le score retourné par la source pour l'objet o_ℓ et $\boldsymbol{\delta}_\ell$ encode la vraie valeur de $\boldsymbol{x}$ pour cet objet. Si l'on dispose de plusieurs (N) sources de cette nature, l'ensemble d'apprentissage est

1. Les données nous donnant le nombre $L - e$ que l'on interprète comme une réalisation d'une expérience binomiale de paramètres θ et L, avec θ la probabilité de pertinence de la source.
2. L'approche de prédiction de Kanjanatarakul *et al.* n'était pas encore publiée au moment de nos travaux.
3. http://archive.ics.uci.edu/ml/

$\mathcal{L}' = \{(s_{11}, s_{21}, \ldots, s_{N1}, \boldsymbol{\delta}_1), \ldots, (s_{1L}, s_{2L}, \ldots, s_{NL}, \boldsymbol{\delta}_L)\}$ où $s_{j\ell}$ correspond au score retourné par la j^{e} source pour le ℓ^{e} objet.

Pour des sources s'exprimant de la sorte, on peut faire appel à des méthodes dites d'étalonnage (*calibration* en anglais) [102, 146] afin d'interpréter le témoignage (score) s fourni par une source pour un nouvel objet o (et, dans le cas de plusieurs sources, pour interpréter les témoignages $\boldsymbol{s} = (s_1, s_2, \ldots, s_N)$ fournis par les sources pour o). Ces méthodes interprètent l'information s émise par la source, sous la forme d'une distribution de probabilité sur $\mathcal{X}$ ou, de manière équivalente, par une fonction de masse Bayésienne notée $m^{\mathcal{X}}_{\mathcal{L},s}$, représentant l'incertitude vis-à-vis de la vraie valeur de $\boldsymbol{x}$ pour o étant donné s et $\mathcal{L}$.

Une limitation de ces méthodes est qu'elles ne prennent pas en compte l'incertitude, due à la quantité de données d'apprentissage, dans leurs estimations des probabilités, et en particulier : moins il y a de données d'apprentissage, plus les probabilités estimées sont incertaines [143]. Afin de traiter cette difficulté, le problème de l'étalonnage d'une (seule) source $\mathfrak{s}$ a récemment été considéré dans le cadre de la théorie des fonctions de croyance, donnant lieu aux méthodes dites évidentielles d'étalonnage [143] qui interprètent le score s fourni par $\mathfrak{s}$ sous forme d'une fonction de masse (non Bayésienne) $m^{\mathcal{X}}_{\mathcal{L},s}$. Ces dernières méthodes sont capables de représenter explicitement l'incertitude due à la quantité de données d'apprentissage, ce qui est important dans des domaines d'application critiques tels que la médecine [120] et qui conduit également à de meilleures performances que les méthodes d'étalonnage probabilistes dans des problèmes de classification, comme montré dans [143].

On peut remarquer que ces méthodes probabilistes et évidentielles d'étalonnage peuvent être présentées en termes d'un témoignage sur $\mathcal{X}$ émis par $\mathfrak{s}$ qui est interprété en fonction d'une connaissance sur la sincérité de $\mathfrak{s}$ déterminée par une approche de prédiction. En particulier, la fonction de masse $m^{\mathcal{X}}_{\mathcal{L},s}$ produite par les méthodes les plus élaborées d'étalonnage que sont les méthodes évidentielles reposant sur l'approche de prédiction de Kanjanatarakul *et al.* (cf section 1.8) proposées par Xu *et al.* [143], peut être retrouvée en considérant un témoignage $m^{\mathcal{X}}_{\mathfrak{s}}(\{x_1\}) = 1$ fourni par $\mathfrak{s}$ qui est corrigé selon une connaissance $m^{\mathcal{H}}_{\mathcal{L},s}$ sur la sincérité ($\mathcal{H} = \{SI, \neg SI\}$) de $\mathfrak{s}$ ($\mathfrak{s}$ étant supposée pertinente), où $m^{\mathcal{H}}_{\mathcal{L},s}$ est la fonction de masse résultant de la prédiction (1.41)-(1.42) effectuée à partir d'une fonction de contour $pl^{\Theta}_{\mathcal{L},s}$ représentant une inférence basée sur $\mathcal{L}$ concernant la probabilité θ de sincérité de la source lorsqu'elle dit s ($pl^{\Theta}_{\mathcal{L},s}$ étant défini différemment en fonction de la méthode évidentielle d'étalonnage utilisée).

Par exemple, considérons un découpage du domaine des scores possibles émis par $\mathfrak{s}$ en J intervalles et que, lorsque s appartient au j^{e} de ces intervalles, la fonction $pl^{\Theta}_{\mathcal{L},s}$ est définie par

$$pl^{\Theta}_{\mathcal{L},s}(\theta) = \frac{\theta^{L_j - e_j}(1-\theta)^{e_j}}{\hat{\theta}^{L_j - e_j}(1-\hat{\theta})^{e_j}}, \tag{4.3}$$

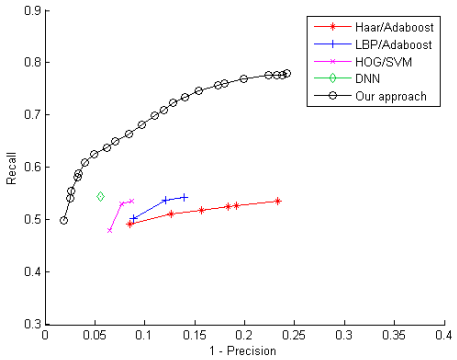
avec $\hat{\theta} = \frac{L_j - e_j}{L_j}$, où L_j est le nombre de couples $(s_\ell, \boldsymbol{\delta}_\ell) \in \mathcal{L}$ tels que s_ℓ appartient au j^{e} intervalle, et e_j est le nombre de couples $(s_\ell, \boldsymbol{\delta}_\ell) \in \mathcal{L}$ tels que s_ℓ appartient au j^{e} intervalle et $\delta_{\ell,2} = 1$. Une telle définition pour $pl^{\Theta}_{\mathcal{L},s}$ s'obtient en appliquant

aux données de $\mathcal{L}$ appartenant au j^{e} intervalle, un raisonnement similaire à celui suivi dans la section 4.2.2.1 pour prédire la sincérité d'une source. Précisément, en considérant que le témoignage de $\mathfrak{s}$ est $m^{\mathcal{X}}_{\mathfrak{S}}(\{x_1\}) = 1$ pour tout objet de l'ensemble d'apprentissage, on compte le nombre de fois où son témoignage est correct (*i.e.*, $\mathfrak{s}$ est sincère) dans le j^{e} intervalle, ce qui correspond au nombre $L_j - e_j$, et le nombre de fois où elle commet une erreur (*i.e.*, $\mathfrak{s}$ n'est pas sincère) dans cet intervalle, ce qui correspond au nombre e_j. La fonction (4.3) est alors celle résultant de l'utilisation de la méthode pour l'inférence statistique basée sur la vraisemblance de Shafer (cf section 1.8) pour le cas où on a observé $L_j - e_j$ succès pour une variable aléatoire suivant une loi binomiale de paramètres θ et L_j (θ étant ici la probabilité que la source soit sincère).

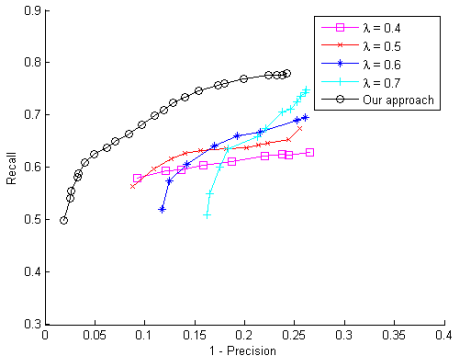
Pour une telle définition (4.3), on vérifie que la fonction de masse sur $\mathcal{X}$ obtenue en corrigeant $m^{\mathcal{X}}_{\mathfrak{S}}$ par la méta-connaissance $m^{\mathcal{H}}_{\mathcal{L},s}$ résultant d'une prediction sur la sincérité de la source étant donné $pl^{\Theta}_{\mathcal{L},s}$, est la fonction de masse $m^{\mathcal{X}}_{\mathcal{L},s}$ obtenue par l'extension évidentielle, basée sur la méthode d'inférence de Shafer, de la technique d'étalonnage probabiliste dite *binning* [143]. Les autres techniques d'étalonnage évidentiel proposées par Xu *et al.* [143], reviennent à d'autres définitions de $pl^{\Theta}_{\mathcal{L},s}$ et donc à d'autres façons d'exploiter les données d'apprentissage pour inférer sur la sincérité de la source.

Dans le cadre de la thèse de Pauline Minary, nous avons appliqué et étendu les méthodes d'étalonnage évidentiel existantes. Plus précisément, dans un premier travail [98] (disponible à l'annexe D), nous avons utilisé les méthodes existantes pour proposer une approche de détection des pixels dans une image appartenant à des visages. Brièvement, cette approche consiste à étalonner individuellement des détecteurs de visages (deux variantes du détecteur de Viola et Jones [139] qui repose sur Adaboost, un détecteur basé sur un SVM et un autre [73] basé sur un réseau de neurones profond) puis à combiner par la règle de Dempster leur sortie étalonnée pour produire une fonction de croyance par pixel représentant l'incertitude vis-à-vis du fait que ce pixel appartienne ou non à un visage, et enfin à prendre une décision pour chaque pixel à l'aide de la règle de décision pessimiste (cf section 1.9). Nous avons montré qu'au prix d'une complexité légèrement plus importante et non problématique en pratique, notre approche présentait un certain nombre d'avantages conceptuels par rapport à une méthode de détection proposée dans [142] [4] et basée également sur l'étalonnage évidentiel des détecteurs suivi de leur fusion. De plus, nous avons montré sur un jeu de données de la littérature (le jeu "FDDB" [67]) et un autre de la SNCF, que notre approche présentait de meilleures performances que les détecteurs individuels sur lesquels elle se basait et que la méthode de détection issue de [142]. La figure 4.3 illustre ces résultats. Notons que notre approche de détection est la base d'un outil semi-automatique pour l'anonymisation des personnes présentes sur des vidéos de quais de gares, utilisé à la SNCF et développé par Pauline Minary durant sa thèse.

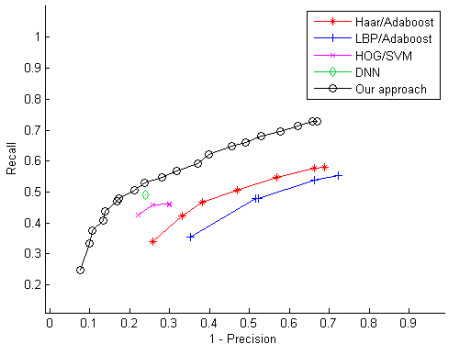
4. Plus précisément, nous avons comparé notre approche à une version améliorée de [142], l'amélioration venant du fait que nous avons remplacé la technique d'étalonnage utilisée dans [142] par un étalonnage plus évolué introduit dans [143].



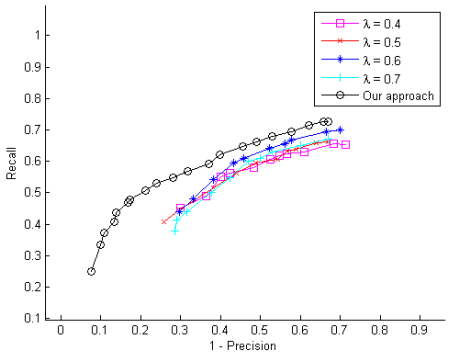
(a) vs détecteurs (FDDB).



(b) vs méthode issue de [142] (FDDB).



(c) vs détecteurs (SNCF).



(d) vs méthode issue de [142] (SNCF).

FIGURE 4.3 – Performances de notre approche de détection de pixels appartenant à des visages, versus les détecteurs (4.3(a) (données FDDB) et 4.3(c) (données SNCF) – le réseau de neurones profond est appelé DNN) et versus la méthode issue de [142] (4.3(b) (données FDDB) et 4.3(d) (données SNCF) – λ est un paramètre de la méthode proposée dans [142]). Les courbes sont obtenues en faisant varier le coût de décider à tort qu'un pixel n'appartient pas à un visage.

Dans un second temps [99] (disponible à l'annexe E), nous avons proposé des extensions des méthodes évidentielles de Xu *et al.* [143], pour le cas où il s'agit d'étalonner un ensemble de sources, plutôt qu'une seule, c'est-à-dire que l'on reçoit pour un objet un vecteur de scores $\boldsymbol{s} = (s_1, s_2, \ldots, s_N)$, avec s_j le score émis par la j^{e} source, et il s'agit d'interpréter ce vecteur par une fonction de masse $m^{\mathcal{X}}_{\mathcal{L}',\boldsymbol{s}}$ représentant l'incertitude vis-à-vis de la vraie valeur de $\boldsymbol{x}$ pour cet objet étant donné $\boldsymbol{s}$ et l'ensemble d'apprentissage $\mathcal{L}'$.

Remarquons que tout comme il est possible d'exprimer les méthodes d'étalonnage d'une seule source en termes de méthodes de prédiction de sa sincérité, les méthodes d'étalonnage de plusieurs sources peuvent également être présentées sous cette angle. Brièvement, la fonction de masse $m^{\mathcal{X}}_{\mathcal{L}',\boldsymbol{s}}$ peut être retrouvée en considérant des témoignages $m^{\mathcal{X}}_{\mathfrak{S}_j}(\{x_1\}) = 1$, $j = 1, \ldots, N$, fournis par les N sources et que les sources peuvent être seulement dans deux états : soit toutes sincères (noté SI^N) soit toutes non sincères (noté $\neg SI^N$). De plus, la connaissance sur ces deux états, permettant d'interpréter les témoignages $m^{\mathcal{X}}_{\mathfrak{S}_j}$, est induite par (1.41)-(1.42) d'une fonction de contour $pl^{\Theta}_{\mathcal{L}',\boldsymbol{s}}$ représentant une inférence, faite à partir de $\mathcal{L}'$, concernant la probabilité $\theta = p(SI^N)$ $(1 - \theta = p(\neg SI^N))$ de sincérité jointe des sources lorsqu'elles disent $\boldsymbol{s}$.

Nous avons montré sur des jeux de données de classification issus du répertoire UCI, que ces méthodes d'étalonnage évidentiel d'un ensemble de sources permettent d'obtenir de meilleures performances que l'approche qui consiste à étalonner individuellement chaque source puis à combiner par la règle de Dempser leur témoignage étalonné. Ces méthodes apportent également une amélioration par rapport aux méthodes d'étalonnage probabiliste d'un ensemble de sources, lorsque l'action de rejet (cf section 1.9) est permise et que le nombre de données d'apprentissage est faible.

4.3 Selon des principes

Si l'expérience préalable que l'on a des sources est faible, une stratégie permettant de déterminer la méta-connaissance à utiliser pour fusionner leur témoignage consiste, comme pour les méthodes décrites à la section 4.2, à considérer un ensemble d'hypothèses candidates sur leur qualité, à partir duquel choisir, ainsi que des critères de sélection permettant d'effectuer ce choix, la différence principale avec les méthodes de la section 4.2 étant que ces critères de sélection ne peuvent plus reposer sur les données mais seulement sur des principes généraux.

Dans [107] (disponible à l'annexe F), nous avons proposé une ligne directrice pour définir un tel ensemble, liée à deux critères formalisant deux caractéristiques principales que l'on peut rechercher à propos de notre connaissance quant à un paramètre $\boldsymbol{x}$: premièrement, la rendre aussi spécifique que possible et, deuxièmement, aussi cohérente que possible.

En effet, le résultat d'une fusion est d'autant mieux qu'il satisfait ces deux critères. D'un côté, un résultat spécifique mais peu cohérent n'est pas désirable, puisque l'on ne peut lui accorder une grande confiance : dans la théorie des fonctions de

croyance, il est usuel de questionner l'usage de la règle conjonctive, qui correspond à l'hypothèse que les sources sont toutes pertinentes et sincères (cf section 2.2.2), lorsque l'incohérence (le conflit) [30] en résultant est trop élevé. La gestion de l'incohérence est elle-même un problème difficile, comme le montre la littérature sur l'analyse du conflit (*e.g.*, [130, 91]). D'un autre côté, un résultat cohérent mais peu spécifique n'est pas désirable non plus, puisqu'il est indécis : cela explique pourquoi certaines règles de combinaison qui correspondent à des hypothèses plus faibles sur les sources, telle que la règle disjonctive, sont peu utilisées. Notons que les objectifs de cohérence et de spécificité sont quelque peu antagonistes, car plus les sources sont informatives, plus il y a de chances qu'elles soient en conflit à propos de la vraie valeur de $\boldsymbol{x}$. Remarquons également que le besoin de chercher un compromis entre cohérence et informativité de la connaissance se retrouve dans d'autres domaines : par exemple, pour gérer l'incohérence dans une base de connaissance [57, 58], Grant et Hunter ont proposé une procédure par étape afin d'améliorer la cohérence tout en minimisant la perte d'information.

Étant donné ces observations, nous avons proposé de considérer un ensemble ordonné de méta-connaissances sur les sources induisant des résultats de fusion de moins en moins spécifiques, mais garantissant en contrepartie que leur cohérence augmente, et de sélectionner dans cet ensemble la méta-connaissance induisant la connaissance la plus spécifique et satisfaisant un seuil minimal de cohérence. L'idée en elle-même n'est pas totalement nouvelle (cf, *e.g.*, [42, 89, 87]). L'originalité de notre travail réside essentiellement dans la formalisation de cette proposition via un ensemble de définitions et de résultats, rendant l'approche à la fois générique, justifiée et pratique. Une présentation sommaire en est donnée dans la suite.

4.3.1 Cohérence et spécificité induites

Considérons le cadre général de la section 2.3.1 pour le cas particulier où $\boldsymbol{y} = \boldsymbol{x}$. Soit $m_i^{\mathcal{X}}$ les témoignages issus de sources indépendantes $\mathfrak{s}_i$, $i = 1, \ldots, N$ et $\mathbf{cm}^{\mathcal{X}} := (m_1^{\mathcal{X}}, \ldots, m_N^{\mathcal{X}})$ la collection de ces témoignages. Soit $m^{\mathcal{H}_{1:N}}$ une méta-connaissance sur ces sources. Soit $m[\mathbf{cm}^{\mathcal{X}}; m^{\mathcal{H}_{1:N}}]^{\mathcal{X}}$ la fonction de masse résultant de la combinaison de ces témoignages selon la méta-connaissance $m^{\mathcal{H}_{1:N}}$ par la règle BBF (2.11).

Rappelons qu'une mesure $\phi(m^{\mathcal{X}}) \in [0, 1]$ de cohérence d'une fonction de masse $m^{\mathcal{X}}$ définie et justifiée dans [30] est

$$\phi(m^{\mathcal{X}}) = \max_{x \in \mathcal{X}} pl^{\mathcal{X}}(x), \tag{4.4}$$

où $pl^{\mathcal{X}}$ est la fonction de contour associée à $m^{\mathcal{X}}$. Cette mesure vérifie $\phi(m^{\mathcal{X}}) = 0$ si et seulement si $m^{\mathcal{X}}(\emptyset) = 1$, et $\phi(m^{\mathcal{X}}) = 1$ si et seulement si l'intersection des éléments focaux de $m^{\mathcal{X}}$ est non vide.

On définit une mesure $\phi(\mathbf{cm}^{\mathcal{X}}; m^{\mathcal{H}_{1:N}})$ de $m^{\mathcal{H}_{1:N}}$*-cohérence* de la collection de

témoignages $\mathbf{cm}^{\mathcal{X}}$ par :

$$\phi(\mathbf{cm}^{\mathcal{X}}; m^{\mathcal{H}_{1:N}}) := \phi(m[\mathbf{cm}^{\mathcal{X}}; m^{\mathcal{H}_{1:N}}]^{\mathcal{X}}). \tag{4.5}$$

Cette mesure évalue à quel point la méta-connaissance $m^{\mathcal{H}_{1:N}}$ est une hypothèse valide concernant le comportement des sources lorsqu'elles fournissent les témoignages $m_i^{\mathcal{X}}$, $i, \ldots, N$. On montre que son calcul n'est pas nécessairement coûteux, en particulier si les sources sont méta-indépendantes selon $m^{\mathcal{H}_{1:N}}$. La quantité $1 - \phi(\mathbf{cm}^{\mathcal{X}}; m^{\mathcal{H}_{1:N}})$ constitue en outre une extension de la mesure de conflit proposée par Destercke et Burger [30], à toutes les règles de combinaison qui peuvent être obtenues par (2.11).

Soit $m_1^{\mathcal{H}_{1:N}}$ et $m_2^{\mathcal{H}_{1:N}}$ deux méta-connaissances sur les sources. Si pour toute collection $\mathbf{cm}^{\mathcal{X}} = (m_1^{\mathcal{X}}, \ldots, m_N^{\mathcal{X}})$ de fonctions de masse, nous avons

$$m[\mathbf{cm}^{\mathcal{X}}; m_1^{\mathcal{H}_{1:N}}]^{\mathcal{X}} \sqsubseteq m[\mathbf{cm}^{\mathcal{X}}; m_2^{\mathcal{H}_{1:N}}]^{\mathcal{X}} \tag{4.6}$$

alors $m_1^{\mathcal{H}_{1:N}}$ est dite au moins autant *méta-spécifique* que $m_2^{\mathcal{H}_{1:N}}$, ce qui est noté $m_1^{\mathcal{H}_{1:N}} \sqsubseteq_{\mathcal{H}} m_2^{\mathcal{H}_{1:N}}$. Cette notion nous permet de comparer des méta-connaissances en termes de spécificité des résultats de fusion qu'elles induisent.

On montre que si une méta-connaissance est au moins autant méta-spécifique qu'une autre, alors elle induit une connaissance sur $\mathcal{X}$ au plus autant cohérente que l'autre pour toute collection de témoignages :

$$m_1^{\mathcal{H}_{1:N}} \sqsubseteq_{\mathcal{H}} m_2^{\mathcal{H}_{1:N}} \Rightarrow \phi(\mathbf{cm}^{\mathcal{X}}; m_1^{\mathcal{H}_{1:N}}) \leq \phi(\mathbf{cm}^{\mathcal{X}}; m_2^{\mathcal{H}_{1:N}}), \quad \forall \mathbf{cm}^{\mathcal{X}} \in \times_{i=1}^{N} \mathcal{M}^{\mathcal{X}}. \tag{4.7}$$

Cette relation indique clairement qu'atteindre la cohérence et la spécificité sont des objectifs quelque peu contraires.

Remarque 4.1 *La mesure (4.4) de cohérence d'une fonction de masse correspond à une des deux définitions – la version forte – de la cohérence totale d'une fonction de masse considérées dans [30], cette dernière notion dans sa forme faible étant associée à une autre mesure de cohérence $\phi_2(m^{\mathcal{X}}) := 1 - m^{\mathcal{X}}(\emptyset)$. Nous avons choisi la mesure $\phi(m^{\mathcal{X}})$ plutôt que $\phi_2(m^{\mathcal{X}})$ dans notre approche essentiellement car cette dernière implique des calculs plus lourds.*

La mesure de la cohérence d'une fonction de masse est une question importante car le conflit entre des fonctions de masse se définit à partir d'elle : il est égal à l'incohérence de leur combinaison conjonctive [30]. Destercke et Burger [30] montrent d'ailleurs qu'avec les deux mesures ϕ et ϕ_2 on obtient deux mesures de conflit (dont la mesure classique $m_{1\textcircled{\cap}2}^{\mathcal{X}}(\emptyset)$) respectant un ensemble de propriétés désirables pour de telles mesures.

Ces dernières années, nous nous sommes intéressés à cette contribution fondamentale de Destercke et Burger sur la notion de conflit et en avons proposé une vision géométrique dans un premier travail [110], qui a donné lieu ensuite à une étude plus poussée [111] (disponible à l'annexe G) sur la mesure de la cohérence d'une fonction de masse. Nous avons en particulier mis à jour le fait que l'on peut

*définir une famille ordonnée de mesures de cohérence, correspondant à des définitions plus ou moins fortes de la notion de cohérence totale et dont les deux définitions de [30] sont des cas extrêmes. De plus, cette famille induit une famille de mesures de conflit respectant les propriétés désirables pour de telles mesures énoncées dans [30]. Elle se prête également à une perspective géométrique : on montre essentiellement que l'incohérence d'une fonction de masse est égale à une certaine distance entre elle et la fonction de masse représentant l'incohérence totale (*i.e., $m^{\mathcal{X}}(\emptyset) = 1$*), résultat qui permet de donner un nouvel éclairage sur la relation entre les mesures de conflit et de distance.*

4.3.2 Sélection de l'hypothèse sur le comportement des sources

Nous avons proposé de considérer une méta-connaissance $m_1^{\mathcal{H}_{1:N}}$ initiale pour fusionner les sources telle que $m_1^{\mathcal{H}^{1:N}}(\mathbf{h}) = 1$, avec $\mathbf{h} = (h^1, \ldots, h^N) \in \mathcal{H}^{1:N}$ et $\Gamma_A(h^i) = A$ pour tout $A \subseteq \mathcal{X}$ et $i = 1, ..., N$, *i.e.*, une hypothèse sur le comportement des sources qui n'induit aucune transformation de leur témoignage. Cette hypothèse correspond donc à accepter les témoignages tels qu'ils sont, sans les altérer en aucune façon. C'est l'hypothèse la plus classique en fusion d'informations et en théorie des fonctions de croyance en particulier, puisqu'elle correspond à la règle conjonctive. C'est donc une méta-connaissance par défaut naturelle.

Grâce à (4.5), nous avons une évaluation de la validité de cette hypothèse par rapport aux témoignages en présence. Comme cela est classiquement préconisé en théorie des fonctions de croyance, cette hypothèse peut être utilisée pour combiner les témoignages si la cohérence induite (4.5) par cette hypothèse est suffisamment élevée, c'est-à-dire au dessus d'un certain seuil τ, et rejetée comme invalide sinon. Dans ce dernier cas, il faut chercher d'autres hypothèses amenant à une cohérence supérieure.

Le résultat (4.7) est instrumental pour cela, puisque choisir une méta-connaissance $m_2^{\mathcal{H}_{1:N}}$ telle que $m_1^{\mathcal{H}_{1:N}} \sqsubseteq_{\mathcal{H}} m_2^{\mathcal{H}_{1:N}}$ garantit un accroissement de la cohérence. Nous avons donc proposé la stratégie en deux étapes suivante pour sélectionner la méta-connaissance à utiliser :

1. Définir une collection de méta-connaissances $\mathbf{cm}^{\mathcal{H}} := (m_1^{\mathcal{H}^{1:N}}, ..., m_M^{\mathcal{H}^{1:N}})$ telle que pour tout $1 \leq j < M$, $m_j^{\mathcal{H}^{1:N}} \sqsubset_{\mathcal{H}} m_{j+1}^{\mathcal{H}^{1:N}}$, et avec $m_1^{\mathcal{H}^{1:N}}$ comme définie ci-dessus ;
2. Tester chaque $m_j^{\mathcal{H}^{1:N}}$ itérativement avec $j = 1, \ldots, M$, jusqu'à ce que

$$\phi(\mathbf{cm}^{\mathcal{X}}; m_j^{\mathcal{H}^{1:N}}) \geq \tau.$$

Cela garantit qu'à chaque itération de j à $j+1$, la spécificité va décroître ($m_j^{\mathcal{H}^{1:N}} \sqsubset_{\mathcal{H}} m_{j+1}^{\mathcal{H}^{1:N}}$) et la cohérence croître ($\phi(\mathbf{cm}^{\mathcal{X}}; m_j^{\mathcal{H}^{1:N}}) \leq \phi(\mathbf{cm}^{\mathcal{X}}; m_{j+1}^{\mathcal{H}^{1:N}})$), le processus s'arrêtant dès que le résultat est jugé suffisamment cohérent et donc digne de confiance. En d'autres termes, cette stratégie fait décroître graduellement la spécificité jusqu'à ce qu'un niveau satisfaisant de cohérence soit atteint.

Cette stratégie générique peut être instanciée pour plusieurs cas pratiques de collections de méta-connaissances $\mathbf{cm}^{\mathcal{H}}$. Une collection peut par exemple être construite à partir de l'hypothèse que q sur les N sources sont pertinentes et qu'elles sont toutes sincères (déjà mentionnée en section 2.2.2) : $m_j^{\mathcal{H}^{1:N}}$, $1 \leq j \leq N$, correspond à l'hypothèse que $N - j + 1$ sur les N sources sont pertinentes et qu'elles sont toutes sincères. Le cas $q = N$ (règle conjonctive) est alors représenté par $m_1^{\mathcal{H}^{1:N}}$ et $q = 1$ (règle disjonctive) par $m_N^{\mathcal{H}^{1:N}}$. Cette collection traite toutes les sources de la même manière, ce qui semble intéressant en l'absence d'expérience préalable avec elles.

Un autre cas intéressant est obtenu en considérant l'hypothèse $m^{\mathcal{H}^{1:N}}$ que les sources $\mathfrak{s}_i$ sont pertinentes avec des probabilités p_i, $i = 1, \ldots, N$, et qu'elles sont méta-indépendantes. Cette hypothèse revient à affaiblir le témoignage de la source $\mathfrak{s}_i$ selon p_i, puis à combiner par la règle conjonctive les témoignages affaiblis des sources. Pour cette hypothèse, on obtient une collection de méta-connaissances satisfaisant la stratégie ci-dessus en définissant $m_j^{\mathcal{H}^{1:N}}$ et $m_{j+1}^{\mathcal{H}^{1:N}}$ tels que $p_i^j \leq p_i^{j+1}$ avec l'inégalité stricte pour au moins un i. Remarquons que les approches [119, 80, 144] basées sur un affaiblissement séquentiel sont alors toutes incluses dans cette approche.

Bien qu'elles semblent d'un intérêt pratique plus limité, on montre que les α-conjonctions peuvent également être utilisées pour définir une collection $\mathbf{cm}^{\mathcal{H}}$, en faisant décroître le paramètre α de 1 à 0.

Enfin, notons que nous avons illustré dans [107, Section VII] l'intérêt de cette approche générique de sélection de méta-connaissances sur les sources, sur un problème de sûreté nucléaire où il s'agit de fusionner des sources (qui sont des organisations tels que le CEA ou l'IRSN) produisant des estimations incertaines de la valeur de paramètres d'un réacteur nucléaire dans des conditions transitoires. Dans ce problème, les données sont manquantes (et coûteuses) et les phénomènes impliqués sont complexes, ce qui fait qu'il n'y a pas de moyen fiable pour connaître la qualité des sources. La collection de méta-connaissances que nous avons utilisée pour ce problème qui comportait 10 sources, est celle mentionnée plus haut correspondant à l'hypothèse que q sur les N sources sont pertinentes. Nous avons pu montré que notre méthode donnait des résultats aux propriétés souhaitables, c'est-à-dire des résultats : (1) cohérents ; (2) informatifs ; (3) lisibles par un utilisateur final (ce qui est essentiel dans ce type d'applications).

4.4 Conclusion

Afin de pouvoir appliquer le modèle présenté au chapitre 2, il est nécessaire de déterminer des méta-connaissances concrètes sur les sources. Ce chapitre a donné un résumé de mes contributions vis-à-vis de ce problème. Je me suis intéressé aux deux situations principales que l'on peut rencontrer : d'un côté le cas où l'on dispose d'une expérience préalable importante des sources, et d'un autre côté le cas où l'expérience préalable des sources est faible. Deux types d'approches ont été présentées pour le premier cas : une basée sur la minimisation d'un critère d'erreur et une autre reposant sur la prédiction. Il a été montré, via des problèmes de classification, que

ces deux types d'approches produisent des méta-connaissances utiles. Pour le second cas, une solution correspondant à la recherche explicite d'une méta-connaissance offrant un compromis cohérence-spécificité satisfaisant, a été décrite. Cette solution est à la fois générique, bien fondée et pratique. Sa pertinence dans une application réelle a également été évoquée.

Le prochain chapitre est consacré à la présentation d'un travail aux liens plus ténus avec les précédents chapitres, puisqu'il concerne l'utilisation de la théorie des fonctions de croyance pour représenter l'incertitude dans un problème d'optimisation classique.

Chapitre 5

Problème de tournées de véhicules avec demandes évidentielles et contrainte de capacité

5.1 Introduction

Comme mentionné dans l'introduction de ce mémoire, la théorie des fonctions de croyance a été utilisée dans de nombreux problèmes. Toutefois, une inspection plus poussée de ces applications révèle que dans un grand nombre de cas, si cette théorie est utilisée, c'est essentiellement pour fusionner des informations. Dans un certain sens cela n'est pas surprenant, puisque là réside l'un des points forts de cette théorie depuis ses origines.

Dans les problèmes où la fusion d'informations occupe une place moins importante, la présence de cette théorie est donc moindre. En particulier, elle demeure assez confidentielle pour ce qui concerne l'optimisation sous incertitude et ses applications (à ma connaissance, les quelques travaux notables sont [93, 100, 134]). Sa capacité à modéliser finement l'incertitude semble pourtant intéressante pour ces problèmes importants, qui sont traités classiquement dans le cadre probabiliste ou le cadre ensembliste et donc avec des représentations moins riches de l'incertitude.

Ce constat, couplé à ma prise de fonction au sein du LGI2A qui avait à ce moment-là la volonté de renforcer les collaborations entre ses thèmes via le financement d'une thèse bi-thèmes, m'a amené à proposer et à conduire avec des collègues du laboratoire une étude dans ce domaine. Précisément, nous nous sommes intéressés au problème de tournées de véhicules avec contrainte de capacité où les demandes des clients sont incertaines et représentées par des fonctions de croyance.

Les résultats de cette étude, réalisée dans le cadre de la thèse de Nathalie Helal, ont été présentés dans [63] (disponible à l'annexe H) et sont synthétisés dans ce chapitre. Notre point de départ fut l'approche suivie par de nombreux auteurs (voir, *e.g.*, [54, 55] et les références à l'intérieur) pour représenter l'incertitude sur les demandes, qui est d'utiliser des variables aléatoires (les demandes sont alors dites stochastiques). Le problème résultant se traite via la programmation stochastique [6], qui en offre deux modèles principaux rappelés à la section 5.2. Notre contribution a été pour l'essentiel de proposer une extension de ces modèles au cas de demandes "évidentielles", c'est-à-dire de demandes incertaines représentées par des fonctions de croyance. Cette contribution est résumée à la section 5.3. Enfin, nous avons mis en œuvre une métaheuristique pour résoudre les modèles obtenus,

que nous avons testée sur quelques instances réalistes ; ces tests sont brièvement discutés à la section 5.4.

5.2 Problème de tournées de véhicules avec demandes stochastiques et contrainte de capacité

Les tournées de véhicules représentent une importante classe de problèmes en recherche opérationnelle et en optimisation combinatoire. Dans le problème de tournées de véhicules avec contrainte de capacité (CVRP, en abrégé, pour *Capacitated Vehicle Routing Problem*), une flotte de m véhicules de même capacité Q et localisée dans un dépôt, doit collecter [1] les demandes de n clients, avec d_i la demande (déterministe) du client i telle que $0 < d_i \leq Q$, $i = 1, \ldots, n$. L'objectif dans le CVRP est de trouver un ensemble de m routes (une par véhicule) tel que la somme des coûts des routes soit minimale, la somme des demandes des clients sur chaque route ne dépasse pas Q, chaque route se termine au dépôt, chaque client soit visité par une seule route et sa demande soit entièrement satisfaite. Ce problème est une extension du problème du voyageur de commerce et fait partie de la classe des problèmes NP-complets. Il peut être écrit sous forme d'un programme linéaire en nombres entiers.

Formellement, soit $c_{i,j}$ le coût du trajet entre le client i et j (le dépôt est représenté par le client artificiel $i = 0$ et on a $c_{i,i} = +\infty$, $i = 0, \ldots, n$). Soit R_k la route du véhicule k et $w_{i,j}^k$ (avec $w_{ii}^k = 0$) une variable binaire qui vaut 1 si le véhicule k se déplace de i à j et sert j (sauf si j est le dépôt). L'objectif du CVRP s'exprime alors par [8, 85] :

$$\min \sum_{k=1}^{m} C(R_k), \tag{5.1}$$

où

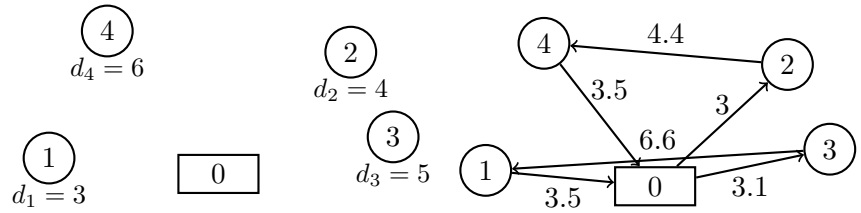
$$C(R_k) = \sum_{i=0}^{n} \sum_{j=0}^{n} c_{i,j} w_{i,j}^k, \tag{5.2}$$

cette minimisation étant sujette à un ensemble de contraintes sur les variables $w_{i,j}^k$ correspondant aux contraintes informelles données dans le paragraphe précédent et en particulier la contrainte que la somme des demandes des clients sur chaque route ne doit pas dépasser Q s'exprime par

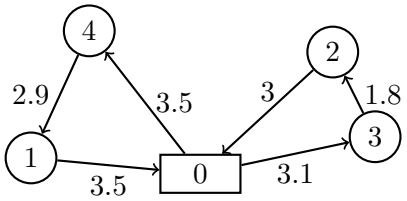
$$\sum_{i=1}^{n} d_i \sum_{j=0}^{n} w_{i,j}^k \leq Q, \quad k = 1, \ldots, m. \tag{5.3}$$

Exemple 5.1 *Supposons $m = 2$ véhicules de capacité limite $Q = 10$, qui doivent collecter les demandes de $n = 4$ clients ayant les demandes $d_1 = 3, d_2 = 4, d_3 = 5, d_4 = 6$. Ces clients sont illustrés sur la figure 5.1(a) (le dépôt est noté "0"). Les*

1. Une formulation équivalente a pour but de livrer des demandes au lieu de les collecter.



(a) Les quatre clients avec leur demande. (b) Une solution candidate.



(c) Une solution optimale.

FIGURE 5.1 – Un CVRP simple.

Tableau 5.1 – Matrice des coûts de trajet.

	0	1	2	3	4
0	$+\infty$	3.5	3	3.1	3.5
1	3.5	$+\infty$	6	6.6	2.9
2	3	6	$+\infty$	1.8	4.4
3	3.1	6.6	1.8	$+\infty$	5.7
4	3.5	2.9	4.4	5.7	$+\infty$

coûts de trajet sont donnés par le tableau 5.1. Une solution candidate, i.e.*, un ensemble de routes satisfaisant les contraintes du CVRP et en particulier* (5.3)*, est montrée sur la figure 5.1(b) ; le coût total de trajet (la valeur de la fonction objectif* (5.1)*) de cette solution est* 24.1*. Une solution optimale,* i.e.*, une solution candidate de coût minimal parmi les solutions candidates, est donnée sur la figure 5.1(c) ; son coût total de trajet est* 17.8*.*

Le CVRP avec demandes stochastiques (CVRPSD, en abrégé, pour *CVRP with Stochastic Demands*) est une modification de ce problème où les demandes d_i, $i = 1, \ldots, n$, des clients sont des variables aléatoires telles que $P(d_i \leq Q) = 1$ (ces variables sont généralement supposées indépendantes). Le CVRPSD est typiquement abordé via le cadre de la programmation stochastique, qui modélise les programmes stochastiques en deux étapes : une solution "a priori" est établie dans une première étape, puis dans une seconde étape les réalisations des variables aléatoires – les

demandes réelles des clients dans le cas du CVRPSD – sont révélées et des mesures correctives, ou *actions de recours*, sont prises si nécessaire sur la solution de la première étape – dans le cas du CVRPSD, des retours au dépôt pour décharger, par exemple, peuvent être effectués si les contraintes de capacités ne sont pas satisfaites. Plus précisément, le CVRPSD est modélisé soit sous la forme d'un programme à base de contraintes en probabilité [9] (section 5.2.1) soit sous la forme d'un programme à base de recours [6] (section 5.2.2).

5.2.1 Programme à base de contraintes en probabilité

La programmation à base de contraintes en probabilité (CCP, en abrégé, pour *Chance Constrained Programming*) consiste à trouver une solution de première étape telle que la probabilité que la demande totale sur toute route dépasse la capacité, soit inférieure à un seuil. Formellement, une formulation CCP du CVRPSD correspond au même problème d'optimisation que pour le CVRP à la différence que les contraintes (5.3) sont remplacées par les contraintes dites en probabilité suivantes :

$$P\left(\sum_{i=0}^{n} d_i \sum_{j=0}^{n} w_{i,j}^{k} \leq Q\right) \geq 1-\beta, \qquad k=1,\ldots,m, \tag{5.4}$$

avec $1-\beta$ la borne minimale sur la probabilité que la contrainte de capacité soit respectée.

Ce modèle ne considère pas le coût des actions correctives qui peuvent être nécessaires lorsque cette solution de première étape est mise en œuvre. En effet, lors de sa mise en œuvre, il est peu probable mais toutefois possible que la capacité des véhicules soit dépassée, *i.e.*, des échecs sur les routes se produisent, et donc il se peut qu'il faille effectuer des actions correctives quand les demandes réelles sont révélées dans la deuxième étape.

5.2.2 Programme à base de recours

La programmation à base de recours (SPR, en abrégé, pour *Stochastic Programming with Recourse*) gère explicitement la possibilité d'échec de la solution de première étape, en intégrant dans la fonction objectif le coût des actions de recours. Plus précisément, dans le modèle SPR du CVRPSD, le coût *espéré* des actions de recours intervenant dans la deuxième étape est considéré, et le problème est de trouver un ensemble de routes qui a un coût total espéré minimal, ce coût étant défini comme la somme entre le coût de la solution de première étape si aucun échec ne se produit et le coût espéré des actions de recours de la deuxième étape. Formellement, soit $C_{\mathrm{E}}(R_k)$ le coût espéré de la route R_k défini par

$$C_{\mathrm{E}}(R_k) = C(R_k) + C_{\mathrm{P}}(R_k), \tag{5.5}$$

avec $C(R_k)$ le coût (5.2) de la route R_k si aucune échec ne s'y produit, et $C_{\mathrm{P}}(R_k)$ le coût espéré des recours (pénalités) sur R_k – ce dernier coût peut être défini de

nombreuses façons suivant la politique de recours utilisée [10, 53, 86, 36].

Une formulation SPR du CVRPSD correspond alors au même problème d'optimisation que pour le CVRP à la différence que l'objectif est modifié en

$$\min \sum_{k=1}^{m} C_{\mathrm{E}}(R_k), \tag{5.6}$$

et les contraintes de capacité (5.3) sont supprimées.

5.3 Modèles dans le cas de demandes évidentielles

Considérons à présent le problème où la connaissance quant aux demandes des clients est représentée dans le cadre de la théorie des fonctions de croyance, plutôt que la théorie des probabilités. On obtient alors le problème de tournées de véhicules avec demandes "évidentielles" et contrainte de capacité (CVRPED, en abrégé, pour *CVRP with evidential demands*). Soit Θ_i le domaine de la variable d_i – nous limitons notre étude au cas où les demandes des clients sont des entiers positifs (donc $\Theta_i = \{1, 2, \ldots, Q\}$, $i = 1, \ldots, n$). La connaissance sur les demandes des clients est alors représentée par une fonction de masse $m^{\Theta_{1:n}}$ sur $\Theta_{1:n} := \times_{i=1}^{n} \Theta_i$. Notons qu'en pratique, il se peut que seulement des connaissances marginales sous la forme de fonctions de masse m^{Θ_i}, $i = 1, \ldots, n$, soient disponibles à propos des demandes individuelles des clients. Dans ce cas, $m^{\Theta_{1:n}}$ peut être obtenue par $m^{\Theta_{1:n}} = \textcircled{\cap}_{i=1}^{n} m^{\Theta_i}$ (cf équation (1.30)) en supposant indépendantes les demandes, de manière similaire à ce qui est généralement fait pour le CVRPSD.

En suivant ce qui a été proposé pour les programmes linéaires sous incertitude représentée dans le cadre de la théorie des fonctions de croyance [93], les approches CCP et SPR de la programmation stochastique peuvent être étendues au CVRPED qui est un programme linéaire en nombres entiers impliquant des fonctions de croyance : le modèle CCP du CVRPSD est généralisé en un programme à base de contraintes en croyance (BCP, en abrégé, pour *Belief Constrained Programming* [93]) pour le CVPRED (section 5.3.1), et le modèle SPR du CVRPSD est généralisé en un programme à base de recours pour le CVRPED (section 5.3.2).

5.3.1 Programme à base de contraintes en croyance

5.3.1.1 Formalisation

La formulation BCP du CVRPED que nous avons proposée correspond au même problème d'optimisation que le modèle CCP du CVRPSD à la différence que les

contraintes (5.4) sont remplacées par les contraintes *en croyance* suivantes :

$$Bel\left(\sum_{i=1}^{n} d_i \sum_{j=0}^{n} w_{i,j}^k \leq Q\right) \geq 1-\underline{\beta}, \qquad k=1,\ldots,m, \tag{5.7}$$

$$Pl\left(\sum_{i=1}^{n} d_i \sum_{j=0}^{n} w_{i,j}^k \leq Q\right) \geq 1-\overline{\beta}, \qquad k=1,\ldots,m, \tag{5.8}$$

avec $\underline{\beta} \geq \overline{\beta}$ et $1-\underline{\beta}$ (resp. $1-\overline{\beta}$) la borne minimale sur la croyance (resp. plausibilité) que la contrainte de capacité soit respectée.

Pour évaluer les contraintes (5.7)-(5.8), la demande totale sur chaque route doit être déterminée en additionnant toutes les demandes des clients sur cette route. Cela se fait pour une route R contenant N clients en utilisant l'équation (1.32), avec f l'addition d'entiers et où $m^{\mathcal{X}_1\times\cdots\times\mathcal{X}_N}$ représente dans ce cas la connaissance sur les demandes des clients appartenant à R, avec $\mathcal{X}_i$ le domaine de la demande du i^{e} client sur R, cette connaissance s'obtenant par marginalisation de $m^{\Theta_{1:n}}$ sur $\mathcal{X}_1 \times \cdots \times \mathcal{X}_N$. Si $m^{\mathcal{X}_1\times\cdots\times\mathcal{X}_N}$ a au plus c ensembles focaux, alors la complexité de l'évaluation des contraintes (5.7)-(5.8) est de $\mathcal{O}(N\cdot Q^N\cdot c)$. Remarquons toutefois que si ces ensembles focaux sont tous des produits Cartésiens d'intervalles, alors cette complexité est seulement de $\mathcal{O}(N\cdot c)$. On met également en évidence [63, Remark 3] que d'autres ensembles focaux aux formes particulières, inspirées de l'optimisation robuste [5], induisent des complexité acceptables.

Les sections 5.3.1.2 et 5.3.1.3 qui suivent donnent quelques propriétés du modèle BCP du CVRPED.

5.3.1.2 Cas particuliers

En fonction des valeurs de $\underline{\beta}$ et de $\overline{\beta}$ ainsi que de la nature de $m^{\Theta_{1:n}}$, le modèle BCP du CVRPED peut se dégénérer en divers problèmes d'optimisation connus.

En particulier, si $m^{\Theta_{1:n}}$ est Bayésienne et donc nous avons en fait affaire à un CVRPSD, alors le modèle BCP du CVRPED est équivalent au modèle CCP de ce CVRPSD pour β dans (5.4) égal à $\overline{\beta}$.

Si $m^{\Theta_{1:n}}$ est catégorique et que son seul ensemble focal est le produit Cartésien de n intervalles, *i.e.*, la demande d_i de chaque client est connue sous la forme d'un interval $[\![\underline{d_i};\overline{d_i}]\!]$, alors les contraintes (5.7) and (5.8) se réduisent lorsque $\underline{\beta} < 1$ à

$$\sum_{i=0}^{n} \overline{d_i} \sum_{j=0}^{n} w_{i,j}^k \leq Q, \qquad k=1,\ldots,m. \tag{5.9}$$

Le modèle BCP revient donc dans ce cas à chercher la solution qui minimise le coût total de servir les clients (5.1) en supposant leur demande maximale, ce qui correspond à une procédure de type minimax rencontrée en optimisation robuste [135].

Si $m^{\Theta_{1:n}}$ est consonante, alors le modèle BCP peut être rapproché des travaux qui représentent l'incertitude sur les demandes par des ensembles flous comme

dans [136].

Si $\underline{\beta} = \overline{\beta}$ et les demandes sont indépendantes, alors le modèle BCP est équivalent au modèle CCP du CVRPSD où les demandes stochastiques d_i, $i = 1, \ldots, n$, sont indépendantes et définies par $P(d_i = \max(A)) = m^{\Theta_i}(A)$, pour tout ensemble focal A de m^{Θ_i}, avec $m^{\Theta_i} := m^{\Theta_{1:n}\downarrow\Theta_i}$. Un résultat similaire est obtenu pour le cas $\underline{\beta} = 1 > \overline{\beta}$.

5.3.1.3 Monotonie

On montre que le coût d'une solution optimale d'un CVRPED modélisé par BCP est non croissant en Q, $\underline{\beta}$ et $\overline{\beta}$. Cela signifie que si le décideur est disposé à acheter des véhicules de plus grande capacité ou à ce que la capacité sur toute route soit dépassée plus souvent, alors il obtiendra des solutions (optimales) au plus autant coûteuses, comme cela devrait être le cas.

Soit $\boldsymbol{x}$ un paramètre défini sur $\mathcal{X}$. Comme cela a été rappelé à la section 1.3, des fonctions de croyance sur $\mathcal{X}$ peuvent être comparées en termes de contenu informationnel grâce à la relation $\sqsubseteq$, qui généralise la comparaison de la spécificité d'ensembles (A est au moins autant spécifique que B, $\forall A, B \subseteq \mathcal{X}$, si $A \subseteq B$).

Lorsque $\mathcal{X}$ est totalement ordonné, on peut comparer deux sous-ensembles A et B de $\mathcal{X}$ en termes d'ordre (plutôt que de spécificité) et, en particulier, en s'inspirant de la comparaison d'intervalles selon l'ordre latticiel [31], on peut dire que A est au moins aussi petit que B, noté $A \leq_{lo} B$, si $\underline{a} \leq \underline{b}$ et $\overline{a} \leq \overline{b}$, avec $\underline{a}$ et $\underline{b}$ (resp. $\overline{a}$ et $\overline{b}$) les indices des plus petites (resp. grandes) valeurs dans A et B. En suivant l'extension $\sqsubseteq$ aux fonctions de masse de l'inclusion ($\subseteq$) entre ensembles, nous avons proposé d'étendre $\leq_{lo}$ aux fonctions de masse de la manière suivante : une fonction de masse $m_1^{\mathcal{X}}$ est dite au moins aussi petite qu'une autre fonction de masse $m_2^{\mathcal{X}}$, ce qui est noté $m_1^{\mathcal{X}} \preceq m_2^{\mathcal{X}}$, si et seulement si il existe une matrice stochastique (gauche) $R = [R(A, B)]$, $A, B \in 2^{\mathcal{X}}$, vérifiant

$$R(A, B) > 0 \Rightarrow A \leq_{lo} B, \quad A, B \subseteq \mathcal{X}, \tag{5.10}$$

$$m_1^{\mathcal{X}}(A) = \sum_{B \subseteq \mathcal{X}} R(A, B) m_2^{\mathcal{X}}(B), \quad \forall A \subseteq \mathcal{X}. \tag{5.11}$$

Nous sommes ainsi munis de deux moyens pour comparer deux fonctions de masse sur $\mathcal{X}$: $m_1^{\mathcal{X}} \sqsubseteq m_2^{\mathcal{X}}$ signifie que $m_2^{\mathcal{X}}$ représente une connaissance moins précise que $m_1^{\mathcal{X}}$ à propos de $\boldsymbol{x}$, alors que $m_1^{\mathcal{X}} \preceq m_2^{\mathcal{X}}$ signifie que $m_2^{\mathcal{X}}$ représente une connaissance disant que $\boldsymbol{x}$ prend des plus grandes valeurs que ce qu'en dit $m_1^{\mathcal{X}}$.

Soit $m^{\Theta_{1:n}}$ et $m_+^{\Theta_{1:n}}$ deux fonctions de masse sur les demandes des n clients. De plus, soit $m^{\Theta_i} := m^{\Theta_{1:n}\downarrow\Theta_i}$ et $m_+^{\Theta_i} := m_+^{\Theta_{1:n}\downarrow\Theta_i}$, $i = 1, \ldots, n$. Si les demandes des clients sont jugées indépendantes et si $m^{\Theta_i} \preceq m_+^{\Theta_i}, i = 1, \ldots, n$, alors on montre que le coût de la solution optimale d'un CVRPED modélisé par BCP où la connaissance sur la demande est $m^{\Theta_{1:n}}$, est inférieur ou égal au coût de sa solution optimale si la connaissance sur les demandes est $m_+^{\Theta_{1:n}}$. Cela signifie en substance que plus la connaissance est pessimiste vis-à-vis des demandes des clients, c'est-à-dire plus on

croit que leur demande est élevée, plus le coût de la solution optimale est grand. Remarquons que pour le modèle BCP, le coût de la solution optimale du CVRPED n'est pas nécessairement plus grand si la connaissance sur les demandes est moins spécifique, contrairement au modèle à base de recours qui est le sujet de la prochaine section.

5.3.2 Programme à base de recours

5.3.2.1 Formalisation

Dans le modèle SPR du CVRPSD, l'incertitude sur les demandes des clients est prise en compte via les quantités $C_{\mathrm{P}}(R_k)$, $k = 1, \ldots, m$, qui représentent les coûts espérés des recours (pénalités) sur les routes étant donné l'incertitude sur les demandes. Pour étendre ce modèle en un programme à base de recours pour le CVRPED, il suffit donc de proposer une définition de $C_{\mathrm{P}}(R_k)$ adaptée à l'incertitude sur les demandes considérée dans le CVRPED. Nous avons proposé une telle définition, pour la politique de recours la plus utilisée qui consiste à permettre au véhicule de faire un aller-retour au dépôt pour décharger s'il atteint un client et qu'il ne peut collecter qu'une partie de la demande de ce client car sa capacité est dépassée [36]. Le reste de cette section présente cette définition.

Incertitude sur les recours Considérons une route R contenant N clients et, sans perte de généralité, que le i^{e} client sur R soit le client i. La connaissance sur les demandes de ces clients est alors représentée par le fonction de masse $m^{\Theta_{1:N}} = m^{\Theta_{1:n}\downarrow\Theta_{1:N}}$ où $\Theta_{1:N} := \times_{i=1}^{N}\Theta_i$.

Étant donné la politique de recours considérée et le fait que les demandes des clients sont au maximum de Q, il vient que pour toutes demandes déterministes $(\theta_1, \ldots, \theta_N) \in \Theta_{1:N}$ des clients sur R la capacité ne peut être dépassée au premier client sur R et donc un recours ne peut survenir à ce client, par contre des recours peuvent être nécessaires pour plusieurs des autres clients sur R (au pire, si $\theta_i = Q$ pour $i = 1, \ldots, N$, il y a aura un recours à chaque client excepté le premier).

Formellement, soit r_j la variable binaire valant 1 si un recours se produit au j^{e} client sur R et 0 sinon. Les situations possibles de recours sur R correspondent alors aux vecteurs $(r_2, \ldots, r_N) \in \{0,1\}^{N-1}$. De plus, les recours survenant sur R étant donné les demandes déterministes $(\theta^1, \ldots, \theta^N)$ et la capacité Q peuvent être représentés par une application $f_Q : \Theta_{1:N} \to \Omega$ avec $\Omega := \{0,1\}^{N-1}$. Par exemple, si $N = 4$ et $\theta_i = Q$, $i = 1, \ldots, 4$, on a $f_Q(\theta_1, \theta_2, \theta_3, \theta_4) = (r_2, r_3, r_4)$ avec $r_j = 1$ pour $j \in \{2, 3, 4\}$. Déterminer l'expression générale de f_Q ne pose pas de difficulté majeure (le lecteur est renvoyé à [63, Section 3.2.2] pour cette expression).

En propageant via l'application f_Q l'incertitude $m^{\Theta_{1:N}}$ sur les demandes des clients, on obtient la fonction de masse m^{Ω} représentant l'incertitude sur les recours survenant sur R, définie par

$$m^{\Omega}(B) = \sum_{f_Q(A)=B} m^{\Theta_{1:N}}(A), \quad \forall B \subseteq \Omega. \tag{5.12}$$

Si $m^{\Theta_{1:N}}$ a au plus c ensembles focaux, la complexité de calculer (5.12) est $\mathcal{O}(Q^N \cdot c)$. Toutefois, si les ensembles de focaux de $m^{\Theta_{1:N}}$ sont tous des produits Cartésiens d'intervalles, on peut faire descendre cette complexité à $\mathcal{O}(2^N \cdot c)$, grâce à un algorithme basé sur un arbre binaire que nous avons proposé et qui permet de calculer $f_Q(A)$ en $\mathcal{O}(2^N)$ (cf [63, Section 3.2.3] pour cet algorithme).

Coût espéré des recours Soit $g : \Omega \to \mathbb{R}^+$ la fonction représentant le coût de toute situation de recours $\omega \in \Omega$ sur R, avec ω le vecteur binaire $(r_2, r_3, \ldots, r_N)$. Puisque le coût de la pénalité de dépassement de la capacité au client i est $2c_{0,i}$ (coût d'un aller-retour au dépôt depuis le client i), le coût associé à la situation de recours ω est

$$g(\omega) = \sum_{i=2}^{N} r_i 2c_{0,i}. \tag{5.13}$$

Soit $C^*_{\mathrm{P}}(R) := E^*(g, m^\Omega)$ le coût espéré supérieur des recours sur R (calculé en utilisant (1.45)). En adoptant de manière similaire à [93] une attitude pessimiste, nous avons proposé de prendre comme définition du coût espéré $C_{\mathrm{P}}(R_k)$ des recours sur R_k pour le programme à base de recours du CVRPED, le coût espéré supérieur de ces recours, *i.e.*,

$$C_{\mathrm{P}}(R_k) := C^*_{\mathrm{P}}(R_k), \quad k = 1, \ldots, m. \tag{5.14}$$

5.3.2.2 Propriétés

Comme pour le programme à base de contraintes en croyances, quelques propriétés intéressantes du programme à base de recours pour le CVPRED peuvent être exhibées.

Cas particuliers Si $m^{\Theta_{1:n}}$ est Bayésienne, *i.e.*, nous avons en fait affaire à un CVRPSD, alors $C^*_{\mathrm{P}}(R)$ se réduit à l'espérance mathématique de la fonction de coût g relativement à la distribution de probabilité m^Ω et donc le programme à base de recours du CVRPED revient à celui de ce CVRPSD.

Si $m^{\Theta_{1:n}}$ est catégorique, alors $C^*_{\mathrm{P}}(R)$ est le pire coût *possible* de R. Dans ce cas, le programme à base de recours du CVRPED revient à optimiser contre les pires coûts possibles des solutions candidates, et il partage donc des similarités avec la protection contre le pire cas populaire en optimisation robuste [135].

Monotonie Soit $m^{\Theta_{1:n}}$ et $m_*^{\Theta_{1:n}}$ deux fonctions de masse sur les demandes des n clients. De plus, soit $m^{\Theta_i} := m^{\Theta_{1:n}\downarrow\Theta_i}$ et $m_*^{\Theta_i} := m_*^{\Theta_{1:n}\downarrow\Theta_i}$, $i = 1, \ldots, n$. Si les demandes des clients sont jugées indépendantes et si $m^{\Theta_i} \sqsubseteq m_*^{\Theta_i}, i = 1, \ldots, n$, alors on montre que le coût de la solution optimale d'un CVPRED modélisé à base de recours et où la connaissance sur la demande est $m^{\Theta_{1:n}}$, est inférieur ou égal au coût de sa solution optimale si la connaissance sur les demandes est $m_*^{\Theta_{1:n}}$. Cela signifie que moins la connaissance à propos des demandes des clients est spécifique, plus le coût de la solution optimale est élevé.

5.4 Tests numériques

5.4.1 Présentation

Le CVRP étant un problème NP-complet, les méthodes de résolution exactes peuvent nécessiter un temps prohibitif pour résoudre des instances de grandes tailles. Les métaheuristiques, tel que l'algorithme de recuit simulé [78], permettent de déterminer des solutions approchées pour de telles instances dans un temps raisonnable.

Pour résoudre des instances du CVPRED modélisé par les deux approches présentées à la section 5.3, nous avons donc choisi d'utiliser une métaheuristique. Plus précisément, nous avons développé des versions modifiées de l'algorithme de recuit simulé proposé dans [62] pour le CVRP. La modification principale introduite pour le modèle BCP est le calcul des contraintes (5.7)-(5.8) pour vérifier la faisabilité de chaque route. Le modèle à base de recours supprime quant à lui la contrainte de capacité, mais change la façon dont la fonction objectif est calculée.

Nous avons généré deux ensembles d'instances AE et AE^+ à partir de l'ensemble A d'instances CVRP d'Augerat *et al.* [121]. Chaque instance dans ces deux ensembles correspond à une instance de l'ensemble A et en particulier elle a les mêmes coordonnées des clients et la même capacité Q.

Pour chaque instance de l'ensemble AE, la connaissance $m^{\Theta_{1:n}}$ sur la demande des clients est obtenue en supposant que les demandes sont indépendantes. De plus, chaque d_i est associée à la fonction de masse m^{Θ_i} définie par

$$m^{\Theta_i}(\{d_i^{det}\}) = 0.8, \tag{5.15}$$
$$m^{\Theta_i}([\underline{z_i}, \overline{z_i}]) = 0.2, \tag{5.16}$$

avec d_i^{det} la demande déterministe originale du client i dans l'instance correspondante de l'ensemble A, et $\underline{z_i}$ et $\overline{z_i}$ tirés aléatoirement dans $(d_i^{det}, Q]$ et $[\underline{z_i}, Q]$, respectivement.

Pour chaque instance de l'ensemble AE^+, les demandes sont également supposées indépendantes, et leur fonction de masse associée est notée $m_+^{\Theta_i}$ et définie à partir de m^{Θ_i} par

$$m_+^{\Theta_i}([d_i^{det}, d_i^{det} + a_i^+]) = 0.8, \tag{5.17}$$
$$m_+^{\Theta_i}([\underline{z_i}, \overline{z_i}]) = 0.2, \tag{5.18}$$

avec a_i^+ tiré aléatoirement dans $[0, \underline{z_i} - d_i^{det} - 1]$.

Notons que $m^{\Theta_i} \sqsubseteq m_+^{\Theta_i}$ et $m^{\Theta_i} \preceq m_+^{\Theta_i}$, $i = 1, \ldots, n$. De plus, la fonction de masse représentant la connaissance incertaine à propos des demandes des clients pour toute route de toute instance de ces deux ensembles d'instances, est telle que ses ensembles focaux sont tous des produits Cartésiens d'intervalles.

5.4.2 Résultats

Le Tableau 5.2 présente un extrait des résultats obtenus en exécutant 30 fois le recuit simulé pour le modèle BCP. Nous avons testé l'approche pour deux valeurs différentes du couple $(\underline{\beta},\ \overline{\beta})$: $(0.4, 0.25)$ et une autre valeur plus contraignante $(0.2, 0.15)$. Les valeurs affichées sont celles des meilleures solutions sur ces 30 exécutions. On peut constater que le couple $(\underline{\beta},\ \overline{\beta})$ le plus contraignant induit des meilleurs coûts supérieurs à l'autre couple, pour les deux ensembles d'instances. De même, l'ensemble d'instances le plus pessimiste vis-à-vis des demandes des clients, *i.e.*, l'ensemble AE^+, induit des meilleurs coûts supérieurs à l'autre ensemble, pour les deux couples $(\underline{\beta},\ \overline{\beta})$.

Le Tableau 5.3 présente un extrait des résultats obtenus pour le modèle à base de recours en lançant 30 fois le recuit simulé. La colonne "Pénalité" indique la part des recours dans le coût de la meilleure solution trouvée. On peut remarquer que l'ensemble d'instances le moins spécifique à propos des demandes des clients, *i.e.*, l'ensemble AE^+, induit des meilleurs coûts supérieurs à l'autre ensemble.

Au global, ces résultats expérimentaux indiquent que nos méthodes de résolutions pour les deux modèles du CVRPED se comportent de manière appropriée, bien qu'elles ne soient pas des méthodes exactes, vis-à-vis des propriétés de monotonie exhibées théoriquement.

5.5 Conclusion

Ce chapitre a présenté les résultats d'une étude à laquelle j'ai contribué concernant le problème de tournées de véhicules avec contrainte de capacité et demandes incertaines représentées dans le cadre de la théorie des fonctions de croyance. Deux modèles pour ce problème ont été présentés. Ils étendent les deux approches de la programmation stochastique que sont la programmation à base de contraintes en probabilité et à base de recours, et ils peuvent également être rapprochés de l'approche ensembliste pessimiste pure. Ils présentent de plus quelques propriétés de monotonie intéressantes. La question de leur complexité a aussi été abordée et un cas particulier important, rendant leurs calculs possibles, a été identifié. Enfin, des méthodes de résolution pour ces modèles ont été testées et validées à partir d'instances classiques.

Tableau 5.2 – Résultats du recuit simulé pour le modèle BCP

Instance	Ensemble AE		Ensemble AE^+	
	Meilleur coût $\underline{\beta} = 0.4, \overline{\beta} = 0.25$	Meilleur coût $\underline{\beta} = 0.2, \overline{\beta} = 0.15$	Meilleur coût $\underline{\beta} = 0.4, \overline{\beta} = 0.25$	Meilleur coût $\underline{\beta} = 0.2, \overline{\beta} = 0.15$
1	1418,3	1850,9	1830,8	2225,4
3	1073,1	1480,2	1196	1502
5	1318,9	1718,6	1755,3	2145,7
7	1597,9	2113,9	1957	2542,6
9	1485	1944,8	1915,5	2203,5
11	1693,4	2158,2	2234,2	2796,6
13	1890,1	2573,2	2287,7	3099,4
15	1872,4	2397,5	2359,8	2956,3
17	2052,6	2636,8	2591	3165,7
19	2263,9	2969,1	2744,8	3415,5
21	2532,3	3207,2	3217,7	4291
23	2179	2881,1	2755	3638,6
25	2214,7	3070,9	2748,1	3509,1
27	3507,2	4524,9	4995,8	5986,3

Tableau 5.3 – Résultats du recuit simulé pour le modèle à base de recours

	Ensemble AE		Ensemble AE^+	
Instance	Meilleur coût	Pénalité	Meilleur coût	Pénalité
1	1750,3	16,8%	2252,6	18,3%
3	1296,1	18%	1490,3	16,6%
5	1670,1	24,2%	2205,3	16,9%
7	1895,6	24,4%	2561,5	17%
9	1851	21,1%	2319,9	19,9%
11	2127,8	20,7%	2858,5	11,8%
13	2530,2	22,7%	3084,7	15,7%
15	2499,5	21,2%	3135,5	15,8%
17	2709,6	19,8%	3366,9	16,6%
19	3083	23,4%	3696,4	16%
21	3317,8	23,6%	4437,5	17,5%
23	2966,7	19,6%	3578,1	19,4%
25	2889,7	22,4%	3665,9	15,4%
27	5016,2	24,2%	6790,5	16,5%

Chapitre 6

Perspectives

La théorie des fonctions de croyance a été introduite originellement par Shafer comme une approche de fusion de témoignages élémentaires partiellement fiables. Shafer et Smets ont fourni des arguments dans ce sens, en montrant en particulier que toute fonction de croyance peut se décomposer en termes de fonctions à supports simples indépendantes, moyennant toutefois quelques subtilités qui peuvent interroger. Je pense avoir apporté un nouvel argument en mettant en évidence qu'en plus de permettre la représentation des témoignages élémentaires partiellement fiables et indépendants, la théorie des fonctions de croyance représente aussi les témoignages élémentaires partiellement fiables et dépendants. Plus précisément, quel que soit l'ensemble de témoignages élémentaires partiellement fiables (et en particulier quelles que soient leurs dépendances), il existe une unique fonction de croyance le représentant, et à toute fonction de croyance, on peut associer de manière unique un ensemble de témoignages élémentaires partiellement fiables l'induisant. Cette décomposition d'une fonction de croyance peut en outre être présentée en termes d'une combinaison conjonctive de fonctions à supports simples ayant une structure de dépendance. Au-delà de la fiabilité, nous avons montré que la théorie des fonctions de croyance permet plus généralement la fusion de témoignages étant donné des connaissances sur d'autres facettes de leur qualité, telle que leur sincérité. Ces conclusions reposent sur une approche générale et naturelle pour la fusion d'informations, où les connaissances sur la qualité des sources sont explicites. D'un point de vue théorique, je pense que l'on dispose maintenant d'un ensemble d'arguments, certes perfectibles, mais relativement solides indiquant que cette théorie est particulièrement adaptée à la fusion de témoignages.

Pour la mise en pratique de cette approche générale de fusion, et spécifiquement la détermination des méta-connaissances sur les sources, nous avons proposé un ensemble de méthodes complétant et étendant des résultats de la littérature, et dont l'efficacité a été étudiée et montrée principalement dans des problèmes de classification supervisée. Cependant, on peut remarquer que nos méthodes basées sur l'exploitation d'une expérience préalable des sources, sont restreintes au contexte classique où l'ensemble d'apprentissage est "dur" et cela à double titre : il ne contient pas d'incertitude quant aux réponses attendues et la possibilité qu'il puisse évoluer n'est pas considérée. Or, dans bien des cas réels, il est possible que les données d'apprentissage soient entachées d'incertitude au niveau des réponses attendues, ou que de nouvelles données d'apprentissage puissent être acquises. En ce qui concerne notre méthode destinée au cas où l'expérience préalable des sources est faible, nous avons utilisé une opérationnalisation des principes généraux de spécificité et de co-

hérence à la fois pratique, mais là aussi d'une certaine manière assez classique. Ces deux notions ne sont en effet pas définies de manière unique dans la théorie des fonctions de croyance et, si cela ne semble pas poser trop de problèmes en général pour la spécificité, la question de la mesure de la cohérence, qui revient à celle de la mesure du conflit, a suscité beaucoup d'intérêt et est aujourd'hui encore un sujet de recherche.

Outre son intérêt pour traiter des problèmes faisant intervenir la fusion d'informations, le formalisme des fonctions de croyance offre plus généralement un cadre de représentation de l'incertitude permettant d'aborder dans une perspective nouvelle des problèmes où l'incertitude joue un rôle important. Par exemple, dans le cas du problème de tournées de véhicules avec contrainte de capacité, elle permet à la fois de représenter l'incertitude sur les demandes des clients de manière plus précise que les approches classiques que sont l'optimisation stochastique et l'optimisation robuste, et d'unir ces deux approches qui sont en général présentées comme distinctes.

Aussi, j'envisage au cours des prochaines années de continuer à étudier ce que cette théorie peut apporter dans de tels problèmes, tout en gardant des activités autour de la fusion d'informations qui seront destinées dans un premier temps principalement aux limites de mes propositions évoquées plus haut. Une liste non exhaustive des thèmes que je compte plus particulièrement aborder est donnée ci-dessous.

Thème 1 : Théorie des fonctions de croyance et apprentissage partiellement supervisé. Le problème de l'estimation paramétrique à partir d'observations incertaines peut être traité de différentes manières [19, 13]. En particulier, si l'incertitude sur les observations prend la forme de fonctions de croyance, on peut recourir à la vraisemblance *évidentielle* [19]. Cette dernière notion est particulièrement utile pour étendre des classifieurs au cas de données partiellement étiquetées, comme montré par exemple dans [114] pour les cas de l'analyse discriminante linéaire et de la régression logistique, et dans [76] pour le cas d'une nouvelle version du classifieur évidentiel des K plus proches voisins basée sur l'affaiblissement contextuel. L'application de ces développements récents au problème de la détermination des méta-connaissances en présence d'incertitude au niveau des réponses attendues dans les données d'apprentissage, me semble pouvoir être envisagée.

Thème 2 : Théorie des fonctions de croyance et apprentissage actif. En apprentissage actif [122], un classifieur est d'abord appris à partir d'un ensemble de données labellisées, puis la possibilité lui est offerte de demander itérativement à un oracle de lui fournir le label de données non étiquetées afin qu'il puisse s'améliorer. L'enjeu est alors d'interroger l'oracle de manière parcimonieuse. À cette fin, le classifieur interroge typiquement l'oracle vis-à-vis des données non étiquetées pour lesquelles il est le plus incertain, le raisonnement étant que cela lui permettra de diminuer le plus rapidement possible son incertitude en général et donc d'améliorer ses performances. Cette approche, appliquée classiquement à des classifieurs proba-

bilistes, semble particulièrement intéressante à considérer dans le cadre de la théorie des fonctions de croyance. En effet, en reconnaissant qu'il existe deux origines à l'incertitude vis-à-vis d'un phénomène – le caractère intrinsèquement aléatoire du phénomène et son observation limitée (voir, *e.g.*, [120]) –, et donc que seulement une part de l'incertitude est réductible – celle liée à la seconde origine, appelée incertitude épistémique –, et en remarquant que les classifieurs évidentiels permettent, contrairement aux classifieurs probabilistes, de distinguer clairement ces origines, alors on s'offre la possibilité d'interroger l'oracle de manière encore plus judicieuse en se focalisant sur la part d'incertitude qui peut être vraiment diminuée, ce qui concrètement revient à l'interroger vis-à-vis des données non étiquetées pour lesquelles l'incertitude épistémique est la plus élevée. Cette idée a déjà été exploitée dans [116] et [92], même si sous des formes différentes (la classification, plutôt que l'apprentissage, et la désambiguïsation de données d'apprentissage partiellement étiquetées, plutôt que l'ajout de nouvelles données, étant faites de manière active dans [116] et [92] respectivement). Nous avons pour notre part réalisé récemment un premier travail [115] suivant cette idée, dans le contexte de l'étalonnage évidentiel dont le fondement est justement de distinguer les origines de l'incertitude. Les résultats préliminaires obtenus nous encouragent à poursuivre dans cette voie.

Thème 3 : Théorie des fonctions de croyance et réseaux de neurones. Denœux considère dans [24, 22] la classe des classifieurs correspondant à la régression logistique ainsi qu'à son application à des transformations non linéaires des attributs des objets à classer. Cette classe est large et contient en particulier les réseaux de neurones à propagation directe (*feedforward*). Il montre que ces classifieurs peuvent être vus comme des classifieurs évidentiels. Plus précisément, pour une régression logistique (généralisée) donnée, chaque attribut (ou transformation d'attribut) avec son paramètre de régression associé induisent une fonction de masse, et la combinaison par la règle de Dempster de ces fonctions de masse suivie de la transformation en probabilité par (1.46) permet de retrouver les probabilités en sortie de cette régression. L'intérêt de cette nouvelle perspective sur ces classifieurs réside dans les fonctions de masse qu'elle met en évidence. En particulier, celle résultant de la combinaison de Dempster est plus informative que les probabilités en sortie. Elle permet de distinguer les situations d'ignorance et de conflit quant aux classes des objets, ce qui a des conséquences pour la prise de décision. Ce travail pionnier ouvre un programme de recherche complet dont l'objectif est de mieux utiliser les classifieurs existants et d'en developper de nouveaux, et dans lequel je compte m'investir. De plus, j'ai également l'intention de m'intéresser à l'approche en elle-même, ce qui pourrait donner lieu à une premiere utilisation des résultats autour de la décomposition d'une fonction de croyance présentée à la section 3.3. En effet, en se restreignant au cas binaire pour simplifier la discussion, on peut montrer que l'on peut utiliser n'importe quelle règle $\textcircled{\cap}_{\boldsymbol{\sigma}}$ (3.19) pour combiner les deux fonctions de masse simples $\{\theta_1\}^{\exp(-w^+)}$ et $\{\theta_2\}^{\exp(-w^-)}$ sur la classe d'un objet qui apparaissent au cœur de l'approche de Denœux (cf [22, Section 3.1.1]), sans que cela n'affecte

le fait de pouvoir retrouver les probabilités en sortie de la régression [1]. Par contre, la structure de dépendance $\boldsymbol{\sigma}$ utilisée (et précisément σ_4 qui est le seul terme non trivial de ce vecteur) détermine le degré de conflit évoqué dans [22, Section 3.1.1] et qui représente une part importante de l'intérêt de l'approche. Actuellement, le degré de conflit considéré revient à choisir $\sigma_4 = 0$. Il me semble intéressant d'étudier l'impact de la valeur σ_4 dans les analyses d'un classifieur permises par l'approche de Denœux, et d'étendre ces considérations au cas général.

Thème 4 : Conflit et distance entre fonctions de croyance. En partie depuis le fameux contre-exemple de Zadeh, la question de la mesure du conflit entre fonctions de croyance, et de sa résolution, a toujours suscité un certain intérêt dans la communauté des fonctions de croyance (voir Smets [130]). Comme évoqué à la remarque 4.1, Destercke et Burger [30] ont réalisé une analyse profonde de cette notion et ont proposé une série de propriétés qu'une mesure de conflit doit respecter, ainsi que des mesures les vérifiant basées sur des mesures de la cohérence d'une fonction de croyance. Ils ont également précisé de manière convaincante la position des mesures de distance [72] entre fonctions de croyance par rapport à celle des mesures de conflit, la conclusion étant qu'une distance entre deux fonctions de croyance ne devrait pas être utilisée en tant qu'une mesure de leur conflit. Ces questions m'intéressent également, en particulier vis-à-vis de leur rôle dans la sélection des hypothèses sur la qualité des sources d'information. Nous poursuivons actuellement l'étude [111] mentionnée à la remarque 4.1 et précisément nous venons de mettre à jour une autre famille ordonnée de mesures de cohérence, qui induit une famille de mesures de conflit [2, 1] et qui ne contient rien d'autre que les racines n-ièmes, $n \geq 1$, des mesures de la famille proposée dans [111]. Un des intérêts de la nouvelle famille de mesures de conflit, outre le fait qu'elle vérifie également l'axiomatique de [30], est qu'elle inclut les mesures de conflit probabiliste et logique considérées par Destercke et Burger, comme des cas particuliers coïncidant respectivement avec la borne inférieure et la limite asymptotique supérieure de cette famille. Nous avons également pu faire apparaître un lien avec les ensembles minimaux inconsistants [65]. Une des perspectives qui nous intéresse est que cette famille pourrait peut-être se prêter de manière plus naturelle à une interprétation géométrique que celle dans [111]. Le choix de la "meilleure" mesure en considérant les coûts computationnels et la sémantique, entre autres critères, est un autre problème que nous souhaitons étudier.

Thème 5 : Théorie des fonctions de croyance et optimisation sous incertitude. Il me semble intéressant de poursuivre l'étude de l'utilisation de la théorie des fonctions de croyance comme formalisme de représentation de l'incertitude dans des problèmes d'optimisation. Plus précisément, les problèmes de tournées de véhicules, qui sont tout à fait en ligne avec les domaines d'application privilégiés du LGI2A, offrent un cadre de réflexion sur ces questions qui me semble suffisamment

1. Cela est dû aux faits suivants : la fonction de contour de la combinaison conjonctive de ces fonctions de masse simples est déterminée par elles, quelle que soit leur structure de dépendance $\boldsymbol{\sigma}$, et la normalisation de cette fonction de contour est égale aux probabilités en sortie de la régression.

riche pour l'instant. En effet, ils soulèvent des questions pour lesquelles les réponses apportées seront peut-être pertinentes pour d'autres problèmes. Un premier aspect à creuser et qui pourrait donner potentiellement une valeur supplémentaire au travail que nous avons effectué, est la question de la comparaison, au-delà de considérations purement formelles, d'un traitement probabiliste des incertitudes versus un traitement évidentiel. En particulier, on peut considérer la situation où l'on dispose de *quelques* données statistiques sur les demandes de clients et comparer les prescriptions de ces deux approches dans ce cas selon des critères qui restent à définir. Un second aspect potentiellement fécond à considérer est le cas où l'incertitude ne se situe pas (ou pas seulement) dans les variables apparaissant dans les contraintes mais dans celles présentes dans la fonction objectif. C'est le cas particulièrement de la variante du problème de tournées de véhicules où les temps de trajet sont incertains [54]. Traiter cette variante nécessitera peut-être des modèles différents de ceux que l'on a proposés. Enfin, un troisième sujet d'intérêt est la question de la résolution des modèles que l'on obtient, puisque nous utilisons pour l'instant un algorithme basé sur le recuit simulé certes fonctionnel mais toutefois assez simple.

Thème 6 : Théorie des fonctions de croyance et propagation des incertitudes. La propagation des incertitudes (cf section 1.7) a des applications en analyse des risques [4], aide à la décision multicritère [48], mais aussi en optimisation sous incertitude, comme l'illustrent les programmes à base de contraintes en croyance et à base de recours pour le CVRPED. Dans un travail récent [34], nous avons généralisé les relations $\sqsubseteq$ et $\preceq$, qui étendent aux fonctions de croyance les relations entre ensembles $\subseteq$ et $\leq_{lo}$, à toute relation $\mathbf{R}$ entre ensembles. Nous avons étudié quelles propriétés communes de $\mathbf{R}$ sont conservées lorsqu'elle est étendue aux fonctions de croyance. De plus, nous avons montré que si une fonction $f : \mathcal{X} \to \mathcal{Y}$ vérifie $A\mathbf{R}^{\mathcal{X}}B \Rightarrow f(A)\mathbf{R}^{\mathcal{Y}}f(B)$ pour tout $A, B \subseteq \mathcal{X}$ et pour $\mathbf{R}^{\mathcal{X}}$ et $\mathbf{R}^{\mathcal{Y}}$ des relations binaires sur $2^{\mathcal{X}}$ et $2^{\mathcal{Y}}$, alors elle conserve cette propriété si A et B sont remplacés par des fonctions de masse sur $\mathcal{X}$. Une élégante application de ce résultat pour le cas de $\sqsubseteq$ a été fournie dans [43] : il permet un calcul approché (encadrant) de la somme de variables aléatoires de manière plus efficace qu'une simulation de Monte-Carlo. Nous l'avons pour notre part exploité dans [63] pour les cas particuliers de $\sqsubseteq$ et $\preceq$. L'application de ce résultat général à des problèmes de propagation des incertitudes où une telle propriété de monotonie jouerait un rôle, constitue une perspective intéressante.

Thème 7 : Théorie des fonctions de croyance et autres formalismes de representation de l'incertitude. Les liens entre les multiples formalismes pour la représentation de l'incertitude, et en particulier entre les plus importants que sont les théories des probabilités, des possibilités, des fonctions de croyance et des probabilités imprécises, sont de mieux en mieux compris [44]. Renforcer cette compréhension à diverses vertus telles que fournir de nouveaux éclairages sur des notions dans une théorie à partir de concepts dans une autre, permettre de les utiliser de

manière conjointe en tirant partie des avantages de chacune, ou encore résoudre des problèmes dans une théorie à partir d'avancées faites dans une autre. À ce sujet, le pont que j'utilise au chapitre 3 entre la distribution de Bernoulli multivariée et la fonction de masse m'a permis de proposer une relecture de certaines notions et en particulier de la fonction de poids, et de fournir une nouvelle solution au problème de la décomposition des fonctions de croyance. Il serait intéressant d'étudier si d'autres profits peuvent être tirés de ce pont. Dans le même ordre d'idées, l'interprétation donnée à la fonction de poids apporte un regard nouveau sur les fonctions de masses simples généralisées qui interviennent dans la décomposition de Smets et qui allouent pour certaines une masse négative à un ensemble. Cela peut peut-être être utile vis-à-vis du problème ouvert [29] qu'est donner une interprétation aux masses négatives rencontrées lors du calcul de la transformée de Möbius d'une mesure de probabilité inférieure.

La validation dans le cadre d'applications industrielles de ces recherches essentiellement théoriques, se fera dans les domaines d'application du LGI2A que sont la logistique durable et la mobilité intelligente, mais également dans les domaines du transport ferroviaire et de la sûreté maritime. Des projets de collaboration avec la SNCF et le CMRE (Centre for Maritime Research and Experimentation) de l'OTAN sont en effet en cours d'élaboration. En outre, des échanges avec la communauté scientifique nationale seront développés plus avant, en particulier par le biais des GdRs IA (Aspects Formels et Algorithmiques de l'Intelligence Artificielle) et ISIS (Information, Signal, Image et ViSion) de l'INS2I du CNRS.

Deuxième partie

Sélection d'articles

Annexe A

Relevance and truthfulness in information correction and fusion

---

Article paru dans *International Journal of Approximate Reasoning*, 53(2) : 159-175, 2012.

International Journal of Approximate Reasoning 53 (2012) 159–175

Contents lists available at SciVerse ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Relevance and truthfulness in information correction and fusion ☆

Frédéric Pichon [a,*], Didier Dubois [b], Thierry Denœux [c]

[a] Thales Research and Technology, Campus Polytechnique, 1 Avenue Augustin Fresnel, 91767 Palaiseau Cedex, France
[b] Université Paul Sabatier, IRIT – Bureau 308, 118 Route de Narbonne, 31062 Toulouse Cedex 09, France
[c] Université de Technologie de Compiègne, UMR CNRS 6599 Heudiasyc, Centre de Recherches de Royallieu, BP20529, 60205 Compiègne Cedex, France

ARTICLE INFO

Article history:
Available online 24 February 2011

Keywords:
Dempster–Shafer theory
Belief functions
Evidence theory
Boolean logic
Information fusion
Discounting

ABSTRACT

A general approach to information correction and fusion for belief functions is proposed, where not only may the information items be irrelevant, but sources may lie as well. We introduce a new correction scheme, which takes into account uncertain metaknowledge on the source's relevance and truthfulness and that generalizes Shafer's discounting operation. We then show how to reinterpret all connectives of Boolean logic in terms of source behavior assumptions with respect to relevance and truthfulness. We are led to generalize the unnormalized Dempster's rule to all Boolean connectives, while taking into account the uncertainties pertaining to assumptions concerning the behavior of sources. Eventually, we further extend this approach to an even more general setting, where source behavior assumptions do not have to be restricted to relevance and truthfulness. We also establish the commutativity property between correction and fusion processes, when the behaviors of the sources are independent.

Quand deux témoins me disent une chose, il faut, pour que je me trompe en ajoutant foi à leur témoignage, que l'un & l'autre m'induisent en erreur; si je suis sûr de l'un des deux, peu m'importe que l'autre soit croyable. Or la probabilité que l'un & l'autre me trompent, est une probabilité composée de deux probabilités, que le premier trompe, & que le second trompe. Celle du premier est 1/10 (puisque la probabilité que la chose est conforme à son rapport est 9/10); la probabilité que le second me trompe aussi, est encore 1/10: donc la probabilité composée est la dixième d'une dixième ou 1/100; donc la probabilité du contraire, c'est-à-dire celle que l'un ou l'autre dit vrai, est 99/100. Entry "Probabilité", Encyclopedia of D'Alembert and Diderot, XVIIIth century.

1. Introduction

The problem of constructing an agent's knowledge on the value taken by a parameter x defined on a domain X, where the agent's sole information on the parameter comes from one or many sources, has gained increased interest in the last 20 years with the development of various kinds of information systems. This problem is actually as old as probability theory: its roots can be traced at least back to the formalization of the reliability of testimonies (see, for instance, the entry "Probabilité" in D'Alembert and Diderot's famous XVIIIth century Encyclopedia[1]).

It is not possible for an agent to evaluate the pieces of information provided by several sources, unless some meta-knowledge on the sources is available to this agent. Typically, meta-knowledge on the sources amounts to assumptions about

☆ This paper is an extended and revised version of [6,20].

* Corresponding author.
E-mail addresses: Frederic.Pichon@thalesgroup.com (F. Pichon), Didier.Dubois@irit.fr (D.Dubois), Thierry.Denoeux@utc.fr (T. Denœux).

[1] Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers. D. Diderot and J. le Rond D'Alembert, editors. University of Chicago: ARTFL Encyclopédie Project (Winter 2008 Edition), Robert Morrissey (ed). http://artflx.uchicago.edu/cgi-bin/philologic/getobject.pl?c.99:87.encyclopedie0110.362669.

0888-613X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2011.02.006

their relevance. If a source providing a testimony of the form $x \in A$ is relevant with probability p, then one assumes that the corresponding information is not useful with probability $1 - p$. In the context of the theory of belief functions [4,22,31], this is known as the *discounting* of a piece of information [22,27] and the resulting state of knowledge is represented by a simple support function [22]: the weight p is allocated to the fact of being able to state $x \in A$ with certainty, and the weight $1 - p$ is allocated to the tautology (it becomes the probability of knowing nothing from the source). If the agent receives the piece of information $x \in A$ from two independent sources, with respective reliabilities p_1 and p_2, then Dempster's rule of combination [4,22], justifies attaching reliability $p_1 + p_2 - p_1 p_2$ to the statement $x \in A$ (this was already explained in full details in the D'Alembert and Diderot Encyclopedia, see the above-mentioned entry).

In this paper, it is proposed to also take into account some meta-knowledge on the truthfulness of the sources. We study how the information provided by a single source is modified, or *corrected* [16], when the agent has some uncertain meta-knowledge on relevance and truthfulness of the source. The case where multiple sources provide information is also thoroughly investigated. This study is performed in the framework of the theory of belief functions. It leads to a general approach to the correction (single source case) and fusion (multiple sources case) of belief functions. This exploration is then pushed forward and further extended to an even more general setting, where assumptions about information sources do not have to be restricted to relevance and truthfulness.

The rest of this paper is organized as follows. The notion of truthfulness is added to the notion of relevance in Section 2, where a thorough study of what this addition brings to the problems of information correction and fusion is conducted. In Section 3, this investigation is pursued by allowing for general source behavior assumptions that go beyond the notions of relevance and truthfulness. A link between information correction and fusion processes, when the behaviors of the sources are independent, is exhibited in Section 4. Some relationships with previous works are outlined in Section 5. Section 6 concludes the paper.

2. Relevance and truthfulness

It is assumed here that the reliability of a source of information involves two dimensions: its *relevance* and its *truthfulness*. A source is said to be relevant if it provides useful information regarding a given question of interest. If the source is a human agent, irrelevance means that the provided information does not pertain to the question it answers, for instance because the agent is actually ignorant. If the source is a sensor, the sensor response is typically irrelevant when it is out of order. For instance, it is useless to try and find the time it is from a clock that is not working since there is no way to know whether the supplied information is correct or not (the hour read on a broken watch can even be correct). In contrast, a source is said to be truthful if it actually supplies the information it possesses. There are various forms of lack of truthfulness. A source may declare the contrary of what it knows, or just say less, or something different, even if consistent with its knowledge. Lack of truthfulness for a sensor may take the form of a systematic bias. Note that if the the agent receiving information does not know in which way the source lies, the difference between irrelevance and lack of truthfulness of a source becomes itself less significant from the standpoint of this agent.

2.1. The case of a single source

Suppose a single source provides information on the value of some deterministic parameter x ranging on a set X of possible values (for instance, somebody's birth-date). Such a piece of information may be of the form "All the source knows is that $x \in A$" where A is a proper non-empty subset of X, supposedly containing the actual value of x. We assume that $\emptyset \subset A \subset X$ because we consider as a source any entity that supplies a non-trivial and non-self-contradictory input. If the source declares not to know the value of x, this would be modeled by $A = X$. However, such information is immaterial for the purpose of information fusion. For simplicity, in the following, we shall assume a crude description of the lack of truthfulness, namely that the source declares the opposite of what it knows to be true. The difference between a source known to lie in this way, and a source known to be irrelevant is that it is possible to retrieve the actual information from the former, while the latter is totally useless.

Knowledge about whether a source is reliable or not, truthful or not differs from the knowledge supplied by the source. It is higher order knowledge and is called *meta-knowledge*. If the source that declares $x \in A$ is known to be irrelevant, the agent receiving this information can always replace it by the trivial information $x \in X$, whether the source is truthful or not. If the source is relevant and is known to lie in the way assumed above, the agent should replace it by $x \in \overline{A}$, where $\overline{A}$ is the complement of A.

2.1.1. Crisp testimony and uncertain meta-knowledge

However, the difficulty is that, in general, the meta-information is uncertain. Consider the frame of discernment $\mathcal{H}$ describing the possible states of the source. Define $\mathcal{H} = \mathcal{R} \times \mathcal{T}$, with $\mathcal{R} = \{R, \neg R\}$ and $\mathcal{T} = \{T, \neg T\}$, as the domain of the pair of Boolean variables (h_R, h_T), where R means relevant and T means truthful. Meta-knowledge about a source may take the form of subjective probabilities $prob(h_R, h_T)$ about the state of the source. Following Dempster's approach [4], a multiple-valued function Γ_A from $\mathcal{H}$ to X can be defined such that:

$$\Gamma_A(R, T) = A;$$
$$\Gamma_A(R, \neg T) = \overline{A};$$
$$\Gamma_A(\neg R, T) = \Gamma(\neg R, \neg T) = X.$$

$\Gamma_A(h)$ interprets the testimony $x \in A$ in each configuration h of the source. Hence, this piece of information will be systematically interpreted by a belief function in the sense of Shafer [22], with mass function m^X on X defined by

$$m^X(A) = prob(R, T),$$
$$m^X(\overline{A}) = prob(R, \neg T),$$
$$m^X(X) = prob(\neg R) = prob(\neg R, T) + prob(\neg R, \neg T).$$

A mass function m^X on X is formally a probability distribution on the power set of X (hence $\sum_{A\subseteq X} m^X(A) = 1$). In this uncertainty theory, the mass $m^X(A)$ is assigned to the possibility of stating $x \in A$ as a faithful representation of the available knowledge; it does not evaluate the likelihood of event A like does a subjective probability $prob(A)$. Philosophically, and in analogy to modal logic, the probability $prob(A)$ could be called a *de re* probability, while $m^X(A)$ can be understood as a *de dicto* probability (in opposition to the usual probabilistic tradition).

Let $q = prob(T|R)$ and $p = prob(R)$. Assuming $\emptyset \subset A \subset X$, it is easily found that

$$m^X(A) = p \cdot q; \tag{1}$$
$$m^X(\overline{A}) = p \cdot (1 - q); \tag{2}$$
$$m^X(X) = 1 - p, \tag{3}$$

corresponding, respectively, to the cases where the source is relevant and truthful, relevant and untruthful, and irrelevant. In practice, it can be assumed that the relevance of a source is independent of its truthfulness, although Eqs. (1)–(3) show that this is not necessary in our approach. In this case, the probability distribution on $\mathcal{H}$ is defined from the probability $p = prob(R)$ that it is relevant and $q = prob(T)$ the probability of its being truthful.

2.1.2. Uncertain testimony and meta-knowledge

More generally, one may assume that the information supplied by a source already takes the form of any kind of mass function m_S^X on X (especially, $m_S^X(X) > 0$ and/or $m_S^X(\emptyset) > 0$ could be allowed). Assuming that the source is in a given state h, then each mass $m_S^X(A)$ should be transferred to $\Gamma_A(h)$, yielding the following mass function:

$$m^X(B|h) = \sum_{A:\Gamma_A(h)=B} m_S^X(A), \quad \forall B \subseteq X. \tag{4}$$

When meta-knowledge on the source is uncertain and each state h has a probability $prob(h)$, then (4) implies that:

$$m^X(B) = \sum_h m^X(B|h) prob(h) = \sum_h prob(h) \sum_{A:\Gamma_A(h)=B} m_S^X(A). \tag{5}$$

Let us already remark that (5) may also be recovered using standard operations of belief function theory (i.e., vacuous extension, Dempter's rule of combination and marginalization) on the considered pieces of evidence (namely the uncertain testimony and metaknowledge), as will be shown in Section 4.2 (Lemma 1).

Assuming the uncertain meta-knowledge of the preceding section, i.e., $prob(R, T) = pq$, $prob(R, \neg T) = p(1 - q)$, $prob(\neg R, T) = (1 - p)q$ and $prob(\neg R, \neg T) = (1 - p)(1 - q)$, leads then to transforming the mass function m_S^X into a new mass function denoted by m^X and defined by:

$$m^X = pq\ m_S^X + p(1 - q)\ \overline{m}_S^X + (1 - p)\ m_X^X, \tag{6}$$

where $\overline{m}_S^X$ is the (random) set complement of m [7], defined by $\overline{m}_S^X(A) = m_S^X(\overline{A})$, $\forall A \subseteq X$, and m_X^X the vacuous mass function defined by $m_X^X(X) = 1$. We thus get

$$m^X(A) = pq\ m_S^X(A) + p(1 - q)\ m_S^X(\overline{A})$$

for all $A \neq X$ and

$$m^X(X) = pq\ m_S^X(X) + p(1 - q) m_S^X(\emptyset) + 1 - p.$$

This is clearly a generalization of the notion of discounting of a belief function proposed by Shafer [22] to integrate the reliability of information sources. In the model underlying the discounting operation, the lack of reliability of a source is assumed to originate in some flaw making it irrelevant. Our approach adds the possibility of the source lacking truthfulness,

i.e., lying.[2] Let us also remark that the complement of a mass function is recovered as a special case of this approach: it corresponds to a relevant source that is lying.

One may as well consider more complex assumptions corresponding to subsets of $\mathcal{H}$, representing epistemic states of the receiving agent about the source state. For instance, the agent may know that:

- The source is either relevant or truthful but not both, that is $R \oplus T = (R \wedge \neg T) \vee (\neg R \wedge T)$, corresponding to the disjunction of two mutually exclusive states.
- The source is relevant or truthful, which is the disjunction $R \vee T$ of three possible mutually exclusive states $\neg R \wedge T$, $R \wedge \neg T$ and $R \wedge T$.

Let $H \subseteq \mathcal{H}$ be some assumption of this type about the source. Using a common abuse of notation, the image of H under Γ_A will be denoted as $\Gamma_A(H)$. It is defined as

$$\Gamma_A(H) = \bigcup_{h \in H} \{\Gamma_A(h)\}.$$

The above results can be applied to such non-elementary assumptions. However, this is not so useful in the case of a single source, since $\Gamma_A(H) = X$ as long as H is not elementary. The major appeal of non-elementary assumptions in the case of several sources will become patent in the sequel.

2.2. The case of multiple sources

If there are two sources of information, two approaches can be envisaged:

1. Modifying information items supplied by each source, then merging the resulting belief functions (using Dempster's rule [22] or its unnormalized version [29]).
2. Embedding meta-knowledge about the source state inside the merging process.

It is clear that the latter option looks more general and more convincing. In this case, a joint assumption on the relevance and truthfulness of sources is in order. Denoting by $\mathcal{H}_1$ and $\mathcal{H}_2$, the set of possible state configurations of each source, the set of elementary joint state assumptions on sources will be $\mathcal{H}_{12} = \mathcal{H}_1 \times \mathcal{H}_2$. Hence, there are 16 possible states of the pair of sources $(h_R^1, h_T^1, h_R^2, h_T^2)$. Uncertain meta-knowledge about the state of sources must be expressed on $\mathcal{H}_{12}$. In the following we describe the result of making elementary assumptions on sources; then we consider the case of uncertain meta-knowledge and uncertain sources, and we are led to equip the assumption space itself with a belief structure.

2.2.1. Crisp testimonies and precise meta-knowledge

Suppose that source 1 asserts $x \in A$ and source 2 asserts $x \in B$ where $A, B \neq X$. How to combine these pieces of information depends on the chosen assumption on the state of sources. Namely, there is a multiple-valued mapping $\Gamma_{A,B} : \mathcal{H}_{12} \to 2^X$ prescribing, for each elementary assumption, the result of the process of merging the two information items.

1. Suppose both sources are truthful.
   (a) If they are both relevant, then one must conclude that $x \in A \cap B$.
   (b) If source 2 (resp. 1) is irrelevant, then one must conclude that $x \in A$ (resp. B).
   (c) Else $x \in X$.
2. Suppose source 1 truthful and source 2 lies.
   (a) If they are both relevant, then one must conclude that $x \in A \cap \overline{B}$.
   (b) If source 2 (resp. 1) is irrelevant, then one must conclude that $x \in A$ (resp. $\overline{B}$).
   (c) Else $x \in X$.
3. Suppose source 2 is truthful and source 1 lies.
   (a) If they are both relevant, then one must conclude that $x \in \overline{A} \cap B$.
   (b) If source 2 (resp. 1) is irrelevant, then one must conclude that $x \in \overline{A}$ (resp. B).
   (c) Else $x \in X$.
4. Suppose both sources lie.
   (a) If they are both relevant, then one must conclude that $x \in \overline{A} \cap \overline{B}$.
   (b) If source 2 (resp. 1) is irrelevant, then one must conclude that $x \in \overline{A}$ (resp. $\overline{B}$).
   (c) Else $x \in X$.

Obviously, the four binary connectives $A \cap B$, $\overline{A} \cap B$, $A \cap \overline{B}$, and $\overline{A} \cap \overline{B}$ are obtained, depending on the truthfulness of supposedly relevant sources. Note that elementary assumptions may be incompatible with some available pieces of

[2] We use the term "lying" here as a synonym of "not telling the truth", irrespective of the existence of any intention of an agent to deceive.

information. For instance, in case of conflicting information ($A \cap B = \emptyset$), case 1a is obviously impossible: either one of the sources is irrelevant, or one of them lies. Likewise, case $A \cap \overline{B} = \emptyset$ excludes assumption 2a, and case $A \cup B = X$ excludes the assumption that both sources lie.

2.2.2. Crisp testimonies and incomplete meta-knowledge

Other Boolean binary connectives can be retrieved by considering non-trivial non-elementary assumptions $H \subset \mathcal{H}_{12}$ on the state of sources, namely disjunctions of elementary assumptions. In theory, the number of such composite assumptions is huge (2^{16}). In practice, only a few assumptions are interesting to study. Indeed, the resulting information is trivial ($x \in X$ because $\Gamma_{A,B}(H) = X$) as soon as H contains an elementary assumption of the form ($\neg R$, h_T^1, $\neg R$, h_T^2), for instance. Some forms of non trivial incomplete meta-knowledge are worth considering.

A first kind of non-trivial meta-knowledge consists in guessing the number of truthful and/or relevant sources, by lack of knowledge on the reliability of individual sources. The interesting cases are as follows (alternative weaker assumptions of this form generate no information):

- Both sources are relevant, and at least one of them is truthful. This is the disjunction of assumptions 1a, 2a, and 3a. Then, $x \in A \cup B$ follows.
- Both sources are relevant, exactly one of which is truthful. This is the disjunction of assumptions 2a and 3a. Then, $x \in A \triangle B$ (*exclusive or*).
- Both sources are relevant, at most one of which truthful. This is the disjunction of assumptions 2a, 3a, and 4a. Then, $x \in \overline{A} \cup \overline{B}$.
- Both sources are truthful, and at least one of them is relevant. This is the disjunction of assumptions 1b and 1a. Then, again $x \in A \cup B$. The same results would be obtained by assuming truthful sources truthful, *exactly* one of them being relevant.

Another kind of meta-knowledge pertains to logical dependence between source states. For instance, one may know that both sources are relevant, but source 1 is truthful if and only if source 2 is so too. This is the disjunction of assumptions 1a and 4a, yielding $x \in (A \cap B) \cup (\overline{A} \cap \overline{B})$, which corresponds to the Boolean equivalence connective. One could likewise retrieve the connective $A \cup \overline{B}$ postulating that:

- Both sources are relevant, but it is impossible that at the same time source 1 lies and source 2 is truthful.
- Or yet that either source 1 is truthful while the other is irrelevant, or source 2 lies and the first one is irrelevant.

This assumption is captured by the implication *if B then A*, which boils down to the following piece of meta-knowledge: "If source 2 is truthful, then source 1 is truthful too" (in the case of relevant sources).

It is possible to retrieve *almost all* binary Boolean connectives of propositional logic (except $A \perp B = \emptyset$, already ruled out in the case of a single source, if one requires that the result of the merging process should be logically consistent). This is not surprising at all, in some sense. However the point here is that *each* logical connective can be derived from an assumption about the global quality of information sources, in terms of truthfulness and relevance. This kind of interpretation has been known for a long time for union and intersection only [8].

Actually, when modeling a complex assumption on quality of sources by means of the appropriate connective yielding the correct ensuing information drawn from these sources, part of the actual meta-information is lost. For instance, $x \in A \cup B$ is obtained in several distinct situations. However, the information that would result from a finer representation of the complex assumptions would be different in each case. Namely:

- If sources are both truthful, and *exactly* one is relevant, then either one should know that $x \in A$ or one should know that $x \in B$ (had we known which source is relevant).
- If both sources are relevant, and at least one is truthful, then either one should know that $x \in A \cap \overline{B}$, or one should know that $x \in \overline{A} \cap B$ or yet that $x \in A \cap B$ (had we known which source is truthful).

In both cases, *one can derive* that we know $x \in A \cup B$, which is weaker that the most precise pieces of information one could derive in each case. In order to express these subtle distinctions, modal logic could be instrumental since it is more expressive than propositional logic. Denoting $\Box$ the modality "to know", it is widely known that $\Box A \vee \Box B$ is not equivalent to (and weaker than) the formula $\Box(A \cap \overline{B}) \vee \Box(\overline{A} \cap B) \vee \Box(A \cap B)$ in a standard modal logic. This line of study would require an investigation in epistemic logic [13], and is left for further research. Nevertheless, the reader is referred to Banerjee and Dubois [1] for a more refined representation in a modal logic framework (and in terms of subsets of the power set of X) of what an agent knows about the epistemic state of another agent acting as a source of information.

2.2.3. Uncertain testimonies and sure meta-knowledge

Suppose now that one incomplete assumption $H \subset \mathcal{H}_{12}$ on the quality of the sources is known to be true. Let $\otimes_H$ be the set-theoretic connective associated to the assumption H in agreement with the above described assignment. Suppose that source 1 (resp. 2) supplies a mass function m_1^X (resp. m_2^X). Moreover we postulate that sources are *independent* in the

following sense: interpreting $m_i^X(A)$ as the probability that source i supplies information item $x \in A$, then the probability that source 1 supplies information item $x \in A$ and source 2 supplies at the same time information item $x \in B$ is the product $m_1^X(A) \cdot m_2^X(B)$.

In this framework, the probability that should be assigned to the possibility of interpreting the joint information supplied by the sources by the statement $x \in C \subseteq X$ is equal to

$$m^X(C) = \sum_{A,B:C=A\otimes_H B} m_1^X(A) \cdot m_2^X(B). \tag{7}$$

This result is a straightforward consequence of the claim that if source 1 asserts $x \in A$ and source 2 asserts $x \in B$, then under assumption H, the conclusion should be that $x \in A \otimes_H B$ is what we actually know. There are 15 variants of this combination rule including the unnormalized version of Dempster's rule (also called conjunctive rule) [29] and the disjunctive rule [7]. Observe that when $A \otimes_H B = \emptyset$ for two focal sets A and B, each coming from a distinct source, this conflict no longer pertains to a disagreement inside X between the two sources, but to a conflict between the information items supplied by the two sources and the meta-assumption H, in space $\mathcal{H}_{12} \times X$. Several approaches make sense to cope with this conflict:

- Either renormalize the resulting belief function like with Dempster's rule, which amounts to assuming the correctness of assumption H, and conditioning on the assumption that sources should not contradict each other.
- Or reject assumption H and prefer one that is compatible with the information supplied by the sources.

Remark. When belief functions built from mass functions m_i^X are consonant, hence fully represented by their contour functions considered as possibility distributions $\pi_i^X : X \to [0, 1]$, one could choose to perform the fusion operation inside the possibilistic framework [9], replacing combination rule (7) by a fuzzy logic connective that extends $\otimes_H$ from the Boolean to the multiple-valued setting.

2.2.4. Crisp testimonies and uncertain meta-knowledge

Now we assume some uncertainty about the meta-knowledge regarding source quality. It is natural to try and represent this meta-uncertainty by means of a mass function $m^{\mathcal{H}}$ on the space $\mathcal{H}$ of incomplete assumptions, rather than a probability distribution on $\mathcal{H}_{12}$. At this point, we limit ourselves to the case where information supplied by sources are simple testimonies of the form $x \in A$ and $x \in B$ respectively. The result of the merging is a mass function m^X on X defined by:

$$m^X(C) = \sum_{H:A\otimes_H B=C} m^{\mathcal{H}_{12}}(H). \tag{8}$$

This mass function actually induces a probability distribution over the 15 Boolean binary connectives attached to assumptions H:

$$p^{\mathcal{H}_{12}}(\otimes) = \sum_{H:\otimes_H=\otimes} m^{\mathcal{H}_{12}}(H). \tag{9}$$

If one of the sources is non-informative ($B = X$), only three connectives remain possible (they reduce to $C = A, \overline{A}$ or X) and the interpretation of information supplied by a single source is recovered (Section 2.1.1).

This approach is general in the sense that, even if the information supplied by each source is independent from the information supplied by the other one, pieces of meta-knowledge regarding the states of each source may not be independent. Such a "meta-independence" between sources may be modeled by assuming that

$$m^{\mathcal{H}_{12}}(H) = \begin{cases} m^{\mathcal{H}_1}(H_1)m^{\mathcal{H}_2}(H_2), & \text{if } H = H_1 \times H_2, \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$

which corresponds to evidential independence [22] between frames $\mathcal{H}_1$ and $\mathcal{H}_2$ with respect to $m^{\mathcal{H}_{12}}$. We note that this notion should not be confused with other notions of independence in evidence theory, as outlined, e.g., in [3,5].

For instance, assume truthful sources with independent probabilities of relevance p_1 and p_2: for $i = 1, 2$,

$$m^{\mathcal{H}_i}(\{(R_i, T_i)\}) = p_i, \tag{11}$$

$$m^{\mathcal{H}_i}(\{(\neg R_i, T_i)\}) = 1 - p_1. \tag{12}$$

We then have

$$m^{\mathcal{H}_{12}}(\{(R_1, T_1, R_2, T_2)\}) = m^{\mathcal{H}_1}(\{(R_1, T_1)\})m^{\mathcal{H}_2}(\{(R_2, T_2)\}) = p_1p_2 \tag{13}$$

and, similarly,

$$m^{\mathcal{H}_{12}}(\{(R_1, T_1, \neg R_2, T_2)\}) = p_1(1-p_2), \tag{14}$$

$$m^{\mathcal{H}_{12}}(\{(\neg R_1, T_1, R_2, T_2)\}) = (1-p_1)p_2, \tag{15}$$

$$m^{\mathcal{H}_{12}}(\{(\neg R_1, T_1, \neg R_2, T_2)\}) = (1-p_1)(1-p_2), \tag{16}$$

and $m^{\mathcal{H}_{12}}(H) = 0$ for all other $H \subseteq \mathcal{H}_{12}$.

Furthermore, it is easy to verify, under these specific hypotheses, that it is equivalent to combine discounted testimonies from each source (with discounting factors p_1 and p_2) by means of the unnormalized Dempster's rule, or to use the combination rule (8) proposed above using the mass function $m^{\mathcal{H}_{12}}$ defined by (13)-(16). Indeed, both methods yield the same mass function m^X:

$$\begin{aligned}
&m^X(A \cap B) = p_1 p_2 && (H = (R_1, T_1, R_2, T_2));\\
&m^X(A) = p_1(1-p_2) && (H = (R_1, T_1, \neg R_2, T_2));\\
&m^X(B) = (1-p_1)p_2 && (H = (\neg R_1, T_1, R_2, T_2));\\
&m^X(X) = (1-p_1)(1-p_2) && (H = (\neg R_1, T_1, \neg R_2, T_2)).
\end{aligned}$$

A more general form of this property will be studied in Section 4.

2.2.5. General case

Consider now the general case where information forwarded by independent sources are belief functions defined by independent mass functions m_1^X and m_2^X. The two merging operations (7) and (8) can be extended jointly by first selecting a merging operation $\otimes$ with probability $p^{\mathcal{H}_{12}}(\otimes)$, and then applying combination $\otimes$ between focal sets of m_1^X and m_2^X:

$$m^X(C) = \sum_H m^{\mathcal{H}_{12}}(H) \sum_{A,B:C=A\otimes_H B} m_1^X(A) m_2^X(B) \tag{17}$$

$$= \sum_{\otimes} p^{\mathcal{H}_{12}}(\otimes) \sum_{A,B:C=A\otimes B} m_1^X(A) m_2^X(B). \tag{18}$$

Here again, we may remark that a more formal derivation of the above result in a more general setting will be presented in Section 4.2 (Lemma 2).

The extension of this approach to the case of $n > 2$ sources that are more or less certainly truthful and/or relevant does not raise any theoretical issue. However the computational complexity will increase exponentially (since there will be 4^n elementary assumptions on the global state of sources, hence a 2^{4^n} complexity for the belief function expressing meta-knowledge on the sources, in the general case).

3. Beyond relevance and truthfulness: a general model of meta-knowledge

In the preceding section, we have seen that considering meta-knowledge on the relevance and truthfulness of information sources leads to some interesting results. In particular, a new correction scheme has been introduced, which generalizes the notions of discounting and complement of a belief function. It also becomes possible to reinterpret all connectives of Boolean logic in terms of assumptions with respect to the relevance and truthfulness of information sources. Furthermore, a general combination rule has been derived, which generalizes the unnormalized version of Dempster's rule to all Boolean connectives and that integrates the uncertainties pertaining to assumptions concerning the possible behavior or state of the sources in the fusion process itself.

In some applications, it may happen that one has finer or even different meta-knowledge on the sources than knowing their relevance and truthfulness. It seems interesting to be able to use such meta-knowledge. In this section, an approach to account for general source behavior assumptions is proposed, through a generalization of the preceding section. We study first the single source case before continuing with the multiple sources case.

3.1. The case of a single source

The notions of relevance and truthfulness were formalized in Section 2.1 using multivalued mappings Γ_A from $\mathcal{H} = \mathcal{R} \times \mathcal{T}$ to X, for each $A \subseteq X$. In this section, we propose a generalization of this setting to account for general source behavior assumptions.

3.1.1. Crisp testimony and certain meta-knowledge

Let us suppose that a source S provides a piece of information on the value taken by a variable y defined on a domain Y. We suppose that this piece of information takes the form $y \in A$, for some $A \subseteq Y$. Let us further assume that the source may be in N elementary states instead of four (as is the case in Section 2.1.1), i.e., we generalize the frame from $\mathcal{H} = \{(R,T), (R,\neg T), (\neg R,T), (\neg R,\neg T)\}$ to $\mathcal{H} = \{h_1, \ldots, h_N\}$ (N does not need to be greater than or equal to four, as

illustrated in Example 1 below). In addition, we consider that we are not so much interested in the value taken by y, as by the related value taken by a variable x defined on a domain X (x and y may or may not be the same parameter). Let us also assume that we have at our disposal some meta-knowledge that relate the piece of information $y \in A$ provided by the source on Y to an information of the form $x \in B$, for some $B \subseteq X$, for each possible state $h \in \mathcal{H}$ of the source.

The reasoning described in the previous paragraph can be formalized as follows. For each $A \subseteq Y$, we define a multivalued mapping Γ_A from $\mathcal{H}$ to X. $\Gamma_A(h)$ indicates how to interpret on X the piece of information $y \in A$ provided by the source in each configuration h of the source. As done in Section 2.1.2, we may also consider non elementary hypotheses $H \subseteq \mathcal{H}$, whose image by Γ_A is $\Gamma_A(H) = \cup_{h \in H} \Gamma_A(h)$.

It is easy to see that the setting introduced in Section 2.1.1 is a particular case of this general scheme, with $N = 4$ and $y = x$ and where the multivalued mappings Γ_A are defined by, for all $A \subseteq X$:

$$\begin{aligned} &\Gamma_A(h_1) = A, \\ &\Gamma_A(h_2) = \overline{A}, \\ &\Gamma_A(h_3) = \Gamma_A(h_4) = X. \end{aligned} \tag{19}$$

The states h_1, h_2, h_3 and h_4 then respectively correspond to the hypotheses (R, T), $(R, \neg T)$, $(\neg R, T)$ and $(\neg R, \neg T)$. More generally this framework also covers known canonical examples for belief function design such as the randomly coded message example, provided by Shafer and Tversky [24].

Furthermore, let us illustrate this general setting using two examples, where meta-knowledge on sources is not limited to notions of relevance and truthfulness.

Example 1 (*Case $y = x$, inspired from Shafer* [23])**.** Let us assume that we are interested by the amount of money Glenn paid for his coffee dues. Besides, we consider that there are only four possible amounts: 0, \$1, \$5 or \$10. The only information we have on this amount comes from a person, named Bill, that we do not know very well and that may be informed, approximately informed or unreliable. If Bill is informed, whatever amount he provides should be accepted. If Bill is approximately informed, the amount he provides should be expanded using the lowest and highest closest amounts (e.g., \$1 is expanded to {0, \$1, \$5}). If Bill is unreliable, the amount he provides cannot be used and we are left in our state of ignorance.

Using the general scheme proposed above, we may formalize this problem as follows. We have $\mathcal{H} = \{\text{informed}, \text{approximately informed}, \text{unreliable}\} = \{h_1, h_2, h_3\}$ and $X = \{0, \$1, \$5, \$10\} = \{x_1, x_2, x_3, x_4\}$. Let $A_{k,r}$ denote the subset $\{x_k, \ldots, x_r\}$, for $1 \leq k \leq r \leq 4$ and let I denote the set of intervals of X: $I = \{A_{k,r}, 1 \leq k \leq r \leq 4\}$. By convention, we consider that the piece of information provided by Bill is one of the intervals in I. We may then define the various states of the source as follows:

$$\begin{aligned} &\Gamma_{A_{k,r}}(\text{informed}) = A_{k,r}, \\ &\Gamma_{A_{k,r}}(\text{ap-informed}) = \begin{cases} \{x_{k-1}\} \cup A_{k,r} \cup \{x_{r+1}\} & \text{if } k > 1 \text{ and } r < 4, \\ A_{k,r} \cup \{x_{r+1}\} & \text{if } k = 1 \text{ and } r < 4, \\ \{x_{k-1}\} \cup A_{k,r} & \text{if } k > 1 \text{ and } r = 4, \\ A_{k,r} & \text{if } k = 1 \text{ and } r = 4, \end{cases} \\ &\Gamma_{A_{k,r}}(\text{unreliable}) = X. \end{aligned}$$

Example 2 (*Case $Y \neq X$, inspired from Janez and Appriou* [14])**.** Let us assume that we are interested in finding the type of a given road, which can only be a track, a lane or a highway. We have a source at our disposal that provides information on this type. However, the source has but a limited perception of the possible types of road and in particular is not aware of the existence of the type "lane". In addition, we know that this source discriminate between roads either using their width or their texture (width and texture are called attributes in [14]). If the source uses the road width, then when it says "track", it is clear that we may only safely infer that the type is "track or lane" since tracks and lanes have similar width, and when it says "highway", we may infer "highway". On the other hand, if the source uses the road texture, then when it says "track", we may infer "track", and when it says "highway", we may only infer "highway or lane" since highways and lanes have similar textures.

Using the approach proposed above, we may formalize this problem as follows. We have $Y = \{\text{track}, \text{highway}\}$, $X = \{\text{track}, \text{lane}, \text{highway}\}$, $\mathcal{H} = \{\text{width}, \text{texture}\}$ and

$$\begin{aligned} &\Gamma_{\text{track}}(\text{width}) = \{\text{track}, \text{lane}\}, \\ &\Gamma_{\text{highway}}(\text{width}) = \{\text{highway}\}, \\ &\Gamma_Y(\text{width}) = X, \\ &\Gamma_{\text{track}}(\text{texture}) = \{\text{track}\}, \end{aligned}$$

$$\Gamma_{\text{highway}}(\text{texture}) = \{\text{lane, highway}\},$$
$$\Gamma_Y(\text{texture}) = X.$$

3.1.2. Behavior-based correction scheme

The approach described in the previous section may be generalized to the case where the source provides uncertain information in the form of a mass function m_S^Y and meta-knowledge on the source are uncertain. Assuming some hypothesis $H \subseteq \mathcal{H}$ on the behavior of the source, then each mass $m_S^Y(A)$ should be transferred to $\Gamma_A(H)$, yielding the following mass function:

$$m^X(B|H) = \sum_{A:\Gamma_A(H)=B} m_S^Y(A), \tag{20}$$

for all $B \subseteq X$.

In the more general situation where we have uncertain meta-knowledge described by a mass function $m^{\mathcal{H}}$ on $\mathcal{H}$, then we get

$$m^X(B) = \sum_H m^X(B|H) m^{\mathcal{H}}(H) = \sum_H m^{\mathcal{H}}(H) \sum_{A:\Gamma_A(H)=B} m_S^Y(A), \tag{21}$$

for all $B \subseteq X$, which clearly generalizes (5). The correction mechanism defined by (21) will be hereafter referred to as *Behavior-Based Correction* (BBC). A more formal derivation of (21) will be provided in Section 4.2 (Lemma 1).

In addition, let us remark that the BBC procedure generalizes a familiar operation of Dempster–Shafer theory, called *conditional embedding* or *ballooning extension* [26,27]. Let us explicitly reinterpret this operation in terms of source behavior assumptions. The ballooning extension is the process that transforms a mass function m^Y defined on a domain Y into a mass function on an extended space X, where $X \supseteq Y$. Let $m^{Y\Uparrow X}$ denote the ballooning extension of m^Y to X. It is defined as $m^{Y\Uparrow X}(B) = m^Y(A)$ if $B = A \cup (X\backslash Y)$ and $m^{Y\Uparrow X}(B) = 0$ otherwise. Suppose that a source S provides a piece of information on the value taken by a parameter x defined on a domain X. We assume further that the information provided by S takes the form of a mass function m_S^Y on the domain $Y \subseteq X$. We consider that there may be two reasons why the source provides a piece of information on the value taken by x on the domain Y instead of X: either the source has a limited perception of the actual domain of x or it knows that the values in $X\backslash Y$ are impossible. Let h_1 denote the state where the source has a limited perception of the actual domain of x and let h_2 denote the state where the source knows the values in $X\backslash Y$ to be impossible. We associate to these two states the multivalued mappings Γ_A, $A \subseteq Y$, from $\mathcal{H} = \{h_1, h_2\}$ to X defined by, for all $A \subseteq Y$:

$$\Gamma_A(h_1) = A \cup (Y\backslash X), \tag{22}$$
$$\Gamma_A(h_2) = A. \tag{23}$$

$\Gamma_A(h_1)$ translates the idea that when the source states $x \in A$, $A \subseteq Y$, we may only safely conclude that $x \in A \cup (X\backslash Y)$, due to the limited perception of the source. Let $m^{\mathcal{H}}$ represent our meta-knowledge on the behavior of the source. If $m^{\mathcal{H}}$ is such that $m^{\mathcal{H}}(\{h_1\}) = 1$ and if we use the BBC procedure to transform m_S^Y into a mass function on X, then the ballooning extension is recovered. The ballooning extension can thus be seen as a correction scheme corresponding to a particular assumption on the behavior of the source with respect to its limited perception of the actual domain of a variable.

The ballooning extension is the most well-known representative of so-called deconditioning methods [15]. To complete the picture on the relationship between the BBC and these methods, we may remark that another deconditioning method, known as the method by association of highest compatible hypotheses [15] and that generalizes the ballooning extension, can also be seen as a particular case of the BBC scheme. Similarly to the ballooning extension, this method transforms a mass function m^Y defined on a domain Y into a mass function on an extended space X, where X contains all the elements of Y and some new elements. However, this transformation is guided by a compatibility relation $\omega : 2^Y \to 2^{X\backslash Y}$, where $\omega(A)$, $A \subseteq Y$, represents the set of hypotheses in $X\backslash Y$ with which the hypotheses of Y contained in A are strongly compatible. For instance, in Example 2 above, the transformation based on road width of the type "track" to "track or lane" and of the type "highway" to "highway" relies on such compatibility relation ω, where $Y = \{\text{track, highway}\}$, $X = \{\text{track, lane, highway}\}$, and $\omega(\text{track}) = \text{lane}$ and $\omega(\text{highway}) = \emptyset$.

Let $m^{Y\Uparrow_\omega X}$ denote the extension of m^Y to X using the method by association of highest compatible hypotheses. It is defined as $m^{Y\Uparrow_\omega X}(B) = m^Y(A)$ if $B = A \cup \omega(A)$ and $m^{Y\Uparrow_\omega X}(B) = 0$ otherwise. The ballooning extension is recovered when $\omega(A) = X\backslash Y$, for all $A \subseteq Y$. As the ballooning extension, it is easy to see that the method by association of highest compatible hypothesesis is a particular case of the BBC (simply replace $\Gamma_A(h_1) = A \cup (Y\backslash X)$ in (22) by $\Gamma_A(h_1) = A \cup \omega(A)$). The state h_1 then corresponds to a particular attribute, such as road width, used by the source to discriminate the hypotheses in Y, as an attribute defines a particular compatibility relation. This deconditioning method can thus also be seen as a correction scheme corresponding to a hypothesis on the behavior of the source.

3.2. The case of multiple sources

Let us now consider that we have two sources S_1 and S_2, each of which may be in one of N elementary states (those N states are the same for both sources). It is convenient to denote by h_j^i the state j of source S_i, for $i = 1, 2$ and $j = 1, \ldots, N$. Accordingly, let $\mathcal{H}_i = \left\{h_1^i, \ldots, h_N^i\right\}$ denote the possible states of source S_i, $i = 1, 2$. The set of elementary hypotheses on the source behaviors will be $\mathcal{H}_{12} = \mathcal{H}_1 \times \mathcal{H}_2$.

3.2.1. Crisp testimonies and certain meta-knowledge

Let us assume that source S_1 states $y \in A$ and S_2 states $y \in B$, $A, B \subseteq Y$. What can be concluded about X after merging these pieces of information will depend on the hypothesis made on the behavior of the sources. We can define a multivalued mapping $\Gamma_{A,B}$ from $\mathcal{H}_{12}$ to X, which assigns to each elementary hypothesis $h = (h^1, h^2)$, $h \in \mathcal{H}_{12}$, the result of the fusion of the two pieces of information $y \in A$ and $y \in B$. As we must conclude $\Gamma_A(h^1)$ when S_1 is in state $h^1 \in \mathcal{H}_1$, and we must conclude $\Gamma_B(h^2)$ when S_2 is in state $h^2 \in \mathcal{H}_2$, where Γ_A and Γ_B are the mappings defined in Section 3.1.1, it is clear that we must conclude $\Gamma_A(h^1) \cap \Gamma_B(h^2)$ when the sources are in state $(h^1, h^2) \in \mathcal{H}_{12}$. Hence, the mapping $\Gamma_{A,B}$ is defined by

$$\Gamma_{A,B}(h) = \Gamma_A(h^1) \cap \Gamma_B(h^2),$$

for all $h \in \mathcal{H}_{12}$.

3.2.2. Behavior-based fusion scheme

Following the same path as that of Section 2.2, we can generalize the above approach by allowing both the information provided by the source and our meta-knowledge about the source to be uncertain.

Let us assume that S_1 and S_2 provide information on Y in the form of two mass functions m_1^Y and m_2^Y, respectively, and that they are independent. If we know that hypothesis $H \subseteq \mathcal{H}_{12}$ holds, then the mass $m_1^Y(A)m_2^Y(B)$ should be transferred to the set

$$C = \Gamma_{A,B}(H) = \bigcup_{(h^1,h^2)\in H} (\Gamma_A(h^1) \cap \Gamma_B(h^2)).$$

The result of the fusion of m_1^Y and m_2^Y given $H \subseteq \mathcal{H}_{12}$ is then the mass function $m^X(\cdot|H)$ defined by:

$$m^X(C|H) = \sum_{A,B:C=\Gamma_{A,B}(H)} m_1^Y(A)\ m_2^Y(B), \tag{24}$$

for all $C \subseteq X$. When meta-knowledge on $\mathcal{H}_{12}$ is represented by a mass function $m^{\mathcal{H}_{12}}$, we then get:

$$m^X(C) = \sum_H m^X(C|H)m^{\mathcal{H}_{12}}(H) = \sum_H m^{\mathcal{H}_{12}}(H) \sum_{A,B:C=\Gamma_{A,B}(H)} m_1^Y(A)m_2^Y(B), \tag{25}$$

for all $C \subseteq X$. Eq. (25) will be referred to as *Behavior-Based Fusion* (BBF). It is clearly a generalization of the general combination rule proposed in Section 2.2.5. A more formal derivation of this rule will be presented in Section 4.2 (Lemma 2).

As a final remark in this section, we may note that the fusion process can be cast in more general settings than those considered here. In particular, one may face problems where sources S_i, $i = 1, \ldots, n$, provide information on different frames Y_i and admit different numbers N_i of elementary states. It is interesting to note that in such a setting, an equation similar to (25) can easily be obtained and a result similar to the one that will be shown in the next section still holds. Although this setting is more general, we have refrained from introducing it in this paper in order to improve readability.

4. Commutativity between correction and fusion schemes with meta-independent sources

As we did at the end of Section 2.2.4, let us now assume sources are meta-independent, i.e., that the mass function $m^{\mathcal{H}_{12}}$ expressing our uncertain meta-knowledge satisfies (10). Under this assumption, we may wonder whether it is equivalent to combine two mass functions m_1^Y and m_2^Y using the BBF rule, or to apply the BBC procedure to m_1^Y and m_2^Y, and combine the transformed mass functions by the unnormalized Dempster's rule (Fig. 1). In order to answer this question, we need first to recall the definitions of some operations related to the use of belief functions defined on product spaces.

4.1. Operations on product spaces

Let $m^{X\times Y}$ denote a mass function defined on the Cartesian product $X \times Y$ of the domains of two parameters x and y. The marginal mass function $m^{X\times Y\downarrow X}$ is defined as

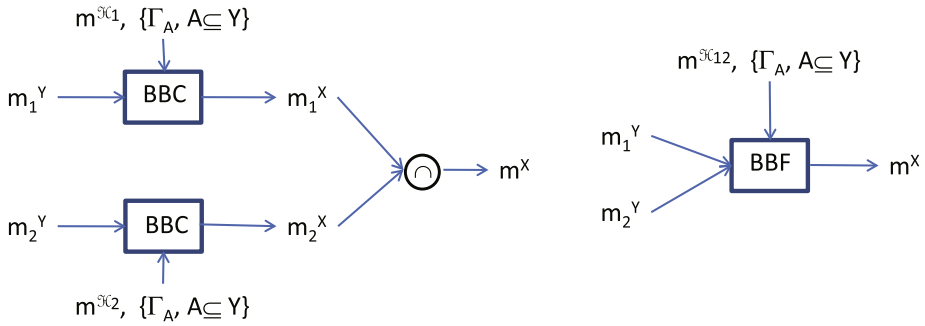


Fig. 1. Two ways of combining two mass functions m_1^Y and m_2^Y using meta-knowledge about the sources: using the BBC procedure (left) and using the BBF rule (right). The equivalence between these two methods under the meta-independence assumption is proved in this section.

$$m^{X\times Y\downarrow X}(A) = \sum_{\{B\subseteq X\times Y,(B\downarrow X)=A\}} m^{X\times Y}(B), \quad \forall A\subseteq X,$$

where $(B\downarrow X)$ denotes the projection of B onto X.

Conversely, let m^X be a mass function defined on X. Its vacuous extension [22] on $X\times Y$ is defined as:

$$m^{X\uparrow X\times Y}(B) = \begin{cases} m^X(A) & \text{if } B = A\times Y \text{ for some } A\subseteq X,\\ 0, & \text{otherwise.}\end{cases}$$

Given two mass functions m_1^X and m_2^Y, their combination by the unnormalized Dempster's rule on $X\times Y$ can be obtained by combining their vacuous extensions on $X\times Y$. Formally:

$$m_1^X \textcircled{\cap} m_2^Y = m_1^{X\uparrow X\times Y} \textcircled{\cap} m_2^{Y\uparrow X\times Y}.$$

Let m^U and m^V be two mass functions on product spaces U and V. The following property, referred to as "Distributivity of marginalization over combination" [25], holds:

$$(m^U \textcircled{\cap} m^V)^{\downarrow U} = m^U \textcircled{\cap} m^{V\downarrow U\cap V}, \tag{26}$$

where $U\cap V$ denotes, by convention, the Cartesian product of frames common to U and V.

4.2. Meta-independence result

Let us consider again the setting of Section 3.1, in which three distinct pieces of evidence are defined:

1. A mass function m_S^Y on Y provided by source S.
2. A mass function $m^{\mathcal{H}}$ on $\mathcal{H} = \{h_1,\ldots,h_N\}$ representing our uncertain meta-knowledge on the source.
3. For each $A\subseteq Y$, a multivalued mapping Γ_A from $\mathcal{H}$ to X indicating how to interpret on X the piece of information $y\in A\subseteq Y$ provided by the source in each configuration $h\in\mathcal{H}$.

The last piece of evidence defines a relation between spaces $\mathcal{H}$, Y and X, which may be represented by the following categorical mass function on $\mathcal{H}\times Y\times X$:

$$m_\Gamma^{\mathcal{H}\times Y\times X}\left[\bigcup_{h\in\mathcal{H},A\subseteq Y}(\{h\}\times A\times\Gamma_A(h))\right] = 1. \tag{27}$$

The three mass functions m_S^Y, $m^{\mathcal{H}}$ and $m_\Gamma^{\mathcal{H}\times Y\times X}$ can be seen as defining an evidential network, as shown in Fig. 2a. As will be shown below, the BBC procedure is equivalent to combining these three mass functions using Dempter's rule, and marginalizing the result on X.

By combining m_S^Y with $m_\Gamma^{\mathcal{H}\times Y\times X}$ and marginalizing on $\mathcal{H}\times X$, we get a mass function $m_{S\Gamma}^{\mathcal{H}\times X}$ on $\mathcal{H}\times X$, defined by:

$$m_{S\Gamma}^{\mathcal{H}\times X}\left[\bigcup_{h\in\mathcal{H}}(\{h\}\times\Gamma_A(h))\right] = m_S^Y(A), \quad \forall A\subseteq Y. \tag{28}$$

For instance, let $\mathcal{H}$ be the space $\mathcal{H} = \mathcal{R}\times\mathcal{T}$, let $Y = X$ and let the multivalued mappings Γ_A be defined by (19) for all $A\subseteq Y$. The mass function $m_{S\Gamma}^{\mathcal{H}\times X}$ is then given by

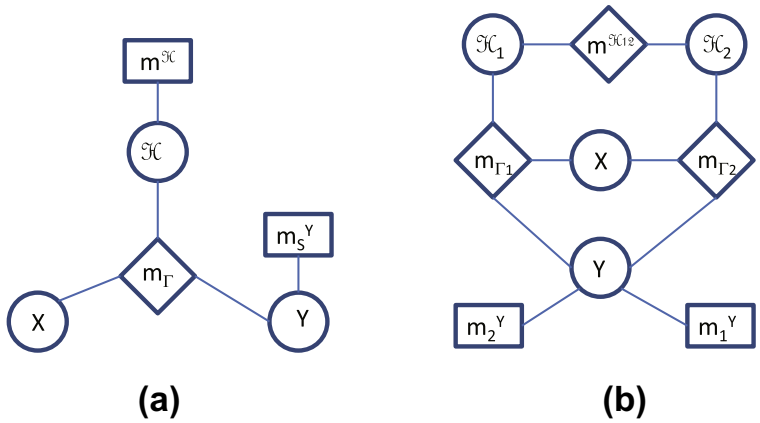


Fig. 2. Evidential networks corresponding to the BBC procedure (a) and the BBF rule (b).

$$m_{S\Gamma}^{\mathcal{H}\times X}((\{h_1\}\times A)\cup(\{h_2\}\times\overline{A})\cup(\{h_3\}\times X)\cup(\{h_4\}\times X))=m_S^Y(A),$$

for all $A\subseteq Y$.

The following lemma states that the mass function given by the BBC (21) can be obtained by combining $m_{S\Gamma}^{\mathcal{H}\times X}$ with $m^{\mathcal{H}}$, and marginalizing on X.

Lemma 1. *We have, for all* $B\subseteq X$

$$\left(m_{S\Gamma}^{\mathcal{H}\times X}\textcircled{\cap}m^{\mathcal{H}}\right)^{\downarrow X}(B)=m^X(B),$$

where m^X *is the mass function defined by* (21).

Proof. Let $m^{\mathcal{H}\times X}=m_{S\Gamma}^{\mathcal{H}\times X}\textcircled{\cap}m^{\mathcal{H}}$. It can be computed as follows:

$$m^{\mathcal{H}\times X}(C)=\begin{cases}m^{\mathcal{H}}(H)\cdot m_S^Y(A) & \text{if } C=(\bigcup_{h\in\mathcal{H}}\{h\}\times\Gamma_A(h))\cap(H\times X),\\ 0 & \text{otherwise.}\end{cases}$$

Now, for all $H\subseteq\mathcal{H}$ and all $A\subseteq Y$,

$$\left[\left(\bigcup_{h\in\mathcal{H}}\{h\}\times\Gamma_A(h)\right)\cap(H\times X)\right]\downarrow X=\bigcup_{h\in H}\Gamma_A(h)=\Gamma_A(H).$$

Therefore, $m^{\mathcal{H}\times X\downarrow X}(B)$ for any $B\subseteq X$ can be obtained by summing over all $H\subseteq\mathcal{H}$ and all $A\subseteq Y$ such that $\Gamma_A(H)=B$:

$$m^{\mathcal{H}\times X\downarrow X}(B)=\sum_{H,A:\Gamma_A(H)=B}m^{\mathcal{H}}(H)\cdot m_S^Y(A),$$

which is equivalent to (21). □

This lemma shows that BBC, and thus deconditioning methods, the discounting operation, the complement of a belief function and the correction scheme introduced in Section 2, can be obtained by defining an evidential network on $\mathcal{H}\times Y\times X$ and by propagating uncertainty in this network using the unnormalized Dempster's rule.

Let us now consider the setting of Section 3.2. We consider two sources S_1 and S_2, which provide items of evidence m_1^Y and m_2^Y on Y, respectively. Let $m^{\mathcal{H}_{12}}$ be a mass function on $\mathcal{H}_{12}=\mathcal{H}_1\times\mathcal{H}_2$ representing our uncertain meta-knowledge on the sources. As before, the mappings Γ_A for all $A\subseteq Y$ induce mass functions $m_{\Gamma i}^{\mathcal{H}_i\times Y\times X}$, $i=1,2$ of the form (27). These mass functions define the evidential network shown in Fig. 2b. By combining $m_{\Gamma i}^{\mathcal{H}_i\times Y\times X}$ with m_i^Y and marginalizing on $\mathcal{H}_i\times X$, we get mass functions $m_{i\Gamma}^{\mathcal{H}_i\times X}$, $i=1,2$ with the following expressions:

$$m_{i\Gamma}^{\mathcal{H}_i\times X}\left[\bigcup_{h\in\mathcal{H}_i}(\{h\}\times\Gamma_A(h))\right]=m_i^Y(A),\quad\forall A\subseteq Y.$$

As expressed by the following lemma, the mass function m^X computed by the BBF rule (25) can be obtained by combining mass functions $m_{i\Gamma}^{\mathcal{H}_i\times X}$, $i=1,2$ with $m^{\mathcal{H}_{12}}$, and marginalizing the result on X:

Lemma 2. *We have, for all* $B \subseteq X$

$$\left(m_{1\Gamma}^{\mathcal{H}_1\times X} \textcircled{\cap} m_{2\Gamma}^{\mathcal{H}_2\times X} \textcircled{\cap} m^{\mathcal{H}_{12}}\right)^{\downarrow X}(B) = m^X(B), \tag{29}$$

where m^X *is the mass function defined by* (25).

Proof. Let $m_{12\Gamma}^{\mathcal{H}_{12}\times X}$ be the mass function obtained by combining the first two mass functions in (29). We have, for all $A, B \subseteq Y$:

$$m_{12\Gamma}^{\mathcal{H}_{12}\times X}(C) = \begin{cases} m_1^Y(A)\cdot m_2^Y(B) & \text{if } C = \bigcup\limits_{(h^1,h^2)\in\mathcal{H}_{12}} \{(h^1,h^2)\}\times(\Gamma_A(h^1)\cap\Gamma_B(h^2)) \\ 0 & \text{otherwise.} \end{cases}$$

By combining the above mass function with $m^{\mathcal{H}_{12}}$, we get a new mass function $m^{\mathcal{H}_{12}\times X}$ defined by

$$m^{\mathcal{H}_{12}\times X}(C) = m^{\mathcal{H}_{12}}(H)\cdot m_1^Y(A)\cdot m_2^Y(B)$$

if

$$C = \left(\bigcup_{(h^1,h^2)\in\mathcal{H}_{12}} \left\{(h^1,h^2)\right\}\times(\Gamma_A(h^1)\cap\Gamma_B(h^2))\right)\cap(H\times X)$$

and $m^{\mathcal{H}_{12}\times X}(C) = 0$ otherwise.

Now, for all $H \subseteq \mathcal{H}_{12}$ and for all $A, B \subseteq Y$,

$$\begin{aligned}&\left[\left(\textstyle\bigcup_{(h^1,h^2)\in\mathcal{H}_{12}} \left\{(h^1,h^2)\right\}\times(\Gamma_A(h^1)\cap\Gamma_B(h^2))\right)\cap(H\times X)\right]\downarrow X \\ &= \textstyle\bigcup_{(h^1,h^2)\in H}(\Gamma_A(h^1)\cap\Gamma_B(h^2)) = \Gamma_{A,B}(H).\end{aligned}$$

Therefore, $m^{\mathcal{H}_{12}\times X\downarrow X}(C)$ for $C \subseteq X$ can be obtained by summing over all $H \subseteq \mathcal{H}_1\times\mathcal{H}_2$ and all $A, B \subseteq Y$ such that $\Gamma_{A,B}(H) = C$:

$$m^{\mathcal{H}_{12}\times X\downarrow X}(C) = \sum_{H,A,B:\Gamma_{A,B}(H)=C} m^{\mathcal{H}_{12}}(H)\cdot m_1^Y(A)\cdot m_2^Y(B)$$

which is equivalent to (25). □

This lemma shows that the BBF rule, and thus the generalization of the unnormalized version of Dempster's rule to all Boolean connectives, can be obtained by defining an evidential network on $\mathcal{H}_{12}\times Y\times X$ and by propagating uncertainty in this network using the unnormalized Dempster's rule. In addition, this implies that the fusion schemes studied in this paper can be recovered using the unnormalized Dempster's rule and marginalization.

Theorem 1. *With meta-independent sources, it is equivalent to combine the uncertain information* m_1^Y *and* m_2^Y *by the BBF rule or to combine by the unnormalized Dempster's rule each of these pieces of information corrected using the BBC procedure.*

Proof. Let $m^{\mathcal{H}_1}$ and $m^{\mathcal{H}_2}$ represent our uncertain meta-knowledge on the behaviors of two sources S_1 and S_2, respectively. Meta-independence of S_1 and S_2 is equivalent to $m^{\mathcal{H}_{12}} = m^{\mathcal{H}_1}\textcircled{\cap} m^{\mathcal{H}_2}$, where $m^{\mathcal{H}_{12}}$ represent our uncertain meta-knowledge on the sources. Under this assumption, we thus have, with the same notations as above:

$$m_{1\Gamma}^{\mathcal{H}_1\times X}\textcircled{\cap} m_{2\Gamma}^{\mathcal{H}_2\times X}\textcircled{\cap} m^{\mathcal{H}_{12}} = m_{1\Gamma}^{\mathcal{H}_1\times X}\textcircled{\cap} m_{2\Gamma}^{\mathcal{H}_2\times X}\textcircled{\cap} m^{\mathcal{H}_1}\textcircled{\cap} m^{\mathcal{H}_2}.$$

Marginalizing on $\mathcal{H}_1\times X$, we get, using (26):

$$\left(m_{1\Gamma}^{\mathcal{H}_1\times X}\textcircled{\cap} m_{2\Gamma}^{\mathcal{H}_2\times X}\textcircled{\cap} m^{\mathcal{H}_1}\textcircled{\cap} m^{\mathcal{H}_2}\right)^{\downarrow\mathcal{H}_1\times X} = m_{1\Gamma}^{\mathcal{H}_1\times X}\textcircled{\cap} m^{\mathcal{H}_1}\textcircled{\cap}\left(m_{2\Gamma}^{\mathcal{H}_2\times X}\textcircled{\cap} m^{\mathcal{H}_2}\right)^{\downarrow X},$$

which, after further marginalization on X, becomes:

$$\left(m_{1\Gamma}^{\mathcal{H}_1\times X}\textcircled{\cap} m_{2\Gamma}^{\mathcal{H}_2\times X}\textcircled{\cap} m^{\mathcal{H}_1}\textcircled{\cap} m^{\mathcal{H}_2}\right)^{\downarrow X} = \left(m_{1\Gamma}^{\mathcal{H}_1\times X}\textcircled{\cap} m^{\mathcal{H}_1}\right)^{\downarrow X}\textcircled{\cap}\left(m_{2\Gamma}^{\mathcal{H}_2\times X}\textcircled{\cap} m^{\mathcal{H}_2}\right)^{\downarrow X}.$$

The theorem then follows from Lemmas 1 and 2. □

Let us note that a similar theorem holds for the case of N sources instead of 2. This is a direct consequence of Lemma 2 being straightforwardly generalizable to the case of N sources.

This theorem leads to an interesting remark: the method, often used in applications, that consists in discounting sources and then combining them by the unnormalized Dempster's rule, can be seen as a particular case of the BBF rule. Indeed, without lack of generality, consider the case of two sources S_1 and S_2. The discount and combine method corresponds to

truthful sources with independent probabilities p_1 and p_2 of relevance, i.e., to a meta-knowledge $m^{\mathcal{H}_{12}}$ on the sources such that $m^{\mathcal{H}_{12}} = m^{\mathcal{H}_1} \textcircled{\cap} m^{\mathcal{H}_2}$, with $m^{\mathcal{H}_1}$, $m^{\mathcal{H}_2}$ and $m^{\mathcal{H}_{12}}$ defined by (11)–(16).

Another popular method for taking into account meta-knowledge on the reliability of the sources is to compute a weighted average of the mass functions to be combined. Indeed, as remarked by Shafer [22, p. 253], this methods yield results similar to those obtained by Dempster's rule applied to equally discounted mass functions, when the discount rate tends to 1. Interestingly, the weighted average rule is also a special case of the BBF rule. The mass function m resulting from the weighted average of two mass functions m_1 and m_2 provided by two sources S_1 and S_2 is defined by $m = w \cdot m_1 + (1 - w) \cdot m_2$, $w \in [0, 1]$, where w is the relative reliability of S_1. It is clear that the weighted average results from meta-knowledge on the sources described by the following mass function

$$m^{\mathcal{H}_{12}}(\{(R_1, T_1, \neg R_2, T_2)\}) = w,$$
$$m^{\mathcal{H}_{12}}(\{(\neg R_1, T_1, R_2, T_2)\}) = 1 - w,$$

which is clearly different from that defined by (13)–(16) and associated to the discount and combine method.

5. Relation to previous work

The idea of exploiting meta-knowledge about the sources of information for correcting or combining belief functions has been explored by several researchers. This section discusses the relation between the notions introduced in Sections 2 and 3 and previous work on similar topics.

5.1. Related work on information correction

As already mentioned, the approach developed in Section 2.1 extends the discounting operation, introduced by Shafer [22] and formalized by Smets [27]. This basic model corresponds to the case where the source is known to be truthful, but has only a probability of being relevant. In [30], Smets proposed a counterpart to this model, in which the source is relevant but may not be truthful. Smets described a scenario in which a "deceiver agent" may replace a belief function by its complement, and he proposed solutions to detect and remedy such a situation. The model introduced in Section 2.1 clearly subsumes these two basic models.

An extension of the discounting operation, called *contextual discounting*, was introduced by Mercier et al. in [18]. In this approach, a binary frame $\mathcal{R}$ for the relevance of the source is introduced as in classical discounting. Additionally, a coarsening Θ of X is defined, and conditional mass functions $m^{\mathcal{R}}(\cdot|\theta)$ on $\mathcal{R}$ given θ, for each $\theta \in \Theta$, are postulated. A discounted mass function on X is obtained by combining the mass function m_S^X provided by the source with the conditional mass functions $m^{\mathcal{R}}(\cdot|\theta), \theta \in \Theta$. In [17], Mercier et al. further generalize this model by allowing the user to specify conditional mass functions $m^{\mathcal{R}}(\cdot|A)$ for any $A \subseteq X$. A crucial assumption in the contextual discounting model and its variants is that of independence between the items of evidence introduced in the model. This correction scheme is, in a sense, simpler than the one introduced in Section 2.1, in that it has no "truthfulness" component. On the other hand, it is based on more complex meta-knowledge about the source, as beliefs on $\mathcal{R}$ are assessed conditionally on different contexts, corresponding to different hypotheses about the variable x of interest. A more complex model incorporating both $\mathcal{R}$ and $\mathcal{T}$ components, and conditional mass functions on $\mathcal{R} \times \mathcal{T}$ given hypotheses about X could obviously be defined, if required by applications.

In [16], Mercier et al. also proposed another extension of the discounting operation, in which uncertain meta-knowledge on the source S is quantified by a mass function $m^{\mathcal{H}}$ on the space $\mathcal{H} = \{h_1, \ldots, h_N\}$ of possible states of the source. The interpretation of those states $h \in \mathcal{H}$ is given by transformations m_h^X of m_S^X: if the source is in state h and if it provides the mass function m_S^X, then we must adopt m_h^X as the representation of our state of belief. This is formalized using conditional mass function by postulating that $m^X(\cdot|h, m_S^X) = m_h^X$, where $m^X(\cdot|h, m_S^X)$ represents our uncertainty on X in a context where h holds and the source provides information m_S^X. This correction scheme is comparable to BBC introduced in Section 3.1, in that it expresses meta-knowledge about the source in a frame of N arbitrary states. The two models coincide in the special case where the mass function $m^{\mathcal{H}}$ on $\mathcal{H}$ is Bayesian, and m_h^X is defined from m_S^X using multivalued mappings Γ_A as:

$$m_h^X(B) = \sum_{A:\Gamma_A(h)=B} m_S^Y(A), \quad \forall B \subseteq X,$$

in which case both models yield

$$m^X = \sum_{h \in \mathcal{H}} m^{\mathcal{H}}(\{h\}) \cdot m_h^X.$$

However, the two models are distinct in the general case, and the choice of one model or another should be guided by the nature of available knowledge in each specific application.

Finally, we should also mention in this section the work of Haenni and Hartmann [12], who proposed a model of partially relevant information sources. In this model, each source S_i is assumed to provide information on a binary variable *HYP* in

Table 1
Mappings Γ_A corresponding to the PD model of Haenni and Hartmann [12].

h	$\Gamma_{\{0\}}(h)$	$\Gamma_{\{1\}}(h)$
100	{0}	{1}
110	{0}	{1}
101	{0}	{1}
111	{0}	{1}
000	{0, 1}	∅
010	{0}	{1}
001	{1}	{0}
011	∅	{0, 1}

the form of a binary report REP_i. Each source generates its report according to s independent variables, possibly including the hypothesis HYP in question. Based on various hypotheses about the relation between REP_i and the underlying variables, a taxonomy of models is generated. Although, at first glance, this formalism seems to be different from ours, our approach happens upon closer examination to be more general. Consider, for instance, the PD model, which is one of the most complex models described in [12]. In this model, the report is generated by the source as follows: if the source is reliable ($REL = 1$), then $REP = HYP$. IF $REL = 0$, then REP is equal to random variable P if $HYP = 1$, and it is equal to a random variable Q if $HYP = 0$. The three random variables REL, P and Q are assumed to be independent, and the model has three parameters $\rho = \Pr(REL = 1)$, $p = \Pr(P = 1)$ and $q = \Pr(Q = 1)$. With our notations, this model can be translated as follows. Let $Y = \{0, 1\}$ the frame of REP, and $\mathcal{H} = \{0, 1\}^3$ the frame of the triple (REL, P, Q). The joint probability distribution of this triple defines a Bayesian mass function $m^{\mathcal{H}}$ on $\mathcal{H}$; for instance, $m^{\mathcal{H}}(\{(1, 0, 1)\}) = \rho(1 - p)q$. Finally, the mappings Γ_A for $A = \{0\}$ and $A = \{1\}$ are given in Table 1. All other models described by Haenni and Hartmann could be translated in a similar way.

5.2. Related work on information fusion

The idea of defining alternatives to Dempster's rule by replacing intersection with other set-theoretic operations can be traced to Smets' 1978 thesis [26], in which he introduced the disjunctive rule of combination together with the Generalised Bayes Theorem (see also [27] for a more accessible reference). This approach was generalized to arbitrary set operations by Dubois and Prade [7] and Yager [32]. In [11], Haenni noticed that the disjunctive rule could be deduced by defining an evidential network with two binary frames $\mathcal{R}_1$ and $\mathcal{R}_2$ for the reliability of the two sources, and combining mass functions m_1^X and m_2^X with a categorical mass function on $\mathcal{R}_1 \times \mathcal{R}_2$ expressing that at least one of the two sources is reliable. The model defined in Section 2.2 is clearly an extension of this simple framework.

In [28], Smets introduced two families of combination rules depending on a parameter α, which he called α-conjunctions and α-disjunctions. These two families are basically the only sets of linear operators with a commutative monoid structure. The α-conjunctions include the unnormalized Dempster's rule (for $\alpha = 1$) and admit the vacuous mass function as neutral element. The α-disjunctions range between the disjunctive rule and a rule corresponding to the exclusive OR, and admit the contradiction ($m(\emptyset) = 1$) as neutral element. In [28], Smets derived these rules from axiomatic requirements, but admitted that they lacked a clear interpretation for $\alpha \in (0, 1)$. In [19,21], Pichon provided such an interpretation in terms of truthfulness of the sources. For instance, he showed that the α-conjunction results from the following assumptions:

1. The two sources are relevant, and either both truthful, or both non truthful (i.e., the operator $\otimes$ is logical equivalence).
2. The degree of belief in the hypothesis that at least one of sources is truthful, conditionally to each value $x \in X$, is equal to α.

Under these assumptions, the α-conjunction can be obtained by defining a belief network in $\mathcal{T}_1 \times \mathcal{T}_2 \times X$, and combining all pieces of evidence, assumed to be independent. The α-disjunction can be obtained in a similar way, starting from different assumptions about the truthfulness of the sources.

As shown in Section 4, the approach developed in Section 2.2, and extended in Section 3.2 can also be derived from uncertainty propagation in an evidential network, in which some variables may be related to the truthfulness of the sources. Although each of the two families of α-junctions relies on a single parameter, the interpretation of this parameter is not easy to disclose, and the independence assumptions involved in the model do not seem very natural. In contrast, the model developed in this paper allows us to represent richer forms of meta-knowledge and it lends itself to easier interpretation.

6. Conclusion

We have proposed a general approach to the correction and fusion of belief functions, which integrates an agent's meta-knowledge on the truthfulness and relevance of the sources of information. This formalism considerably extends Shafer's discounting operation, which deals only with the relevance of sources, as well as the unnormalized Dempster's rule. The

obtained results can be applied, in particular, to all domains where information sources are intelligent agents able to lie, independently of their competence to provide information.

We have further extended this approach by allowing for general source behavior assumptions that go beyond the notions of relevance and truthfulness. This extension is potentially useful for various applications and, in particular, those involving information sources defined on different frames.

We have then shown that the correction and fusion schemes introduced in this paper can be obtained by defining particular evidential networks and by propagating uncertainty in these networks using the unnormalized Dempster's rule. Using a well-known property of belief functions defined on product spaces, we have proved that commutativity between correction and fusion processes holds, when the behaviors of the sources are independent.

Finally, the proposed formal representation of meta-knowledge on the behavior of information sources turns out to be somewhat similar to, but arguably more general and flexible than other approaches introduced in the Dempster–Shafer framework.

One line for further research is to extend the framework to the case of sources reporting to the agent what other sources reported to them. In other words, instead of considering several parallel testimonies, one may consider a series of agents, each reporting to the next one what the previous agent reported. There are then several uncertain information distortion steps in a row by sources having uncertain behavior. Interestingly, this alternative line of research is already present in the entry "Probabilité" in D'Alembert and Diderot Encyclopedia. Recently, Cholvy [2] also investigated this issue. Eventually one may consider the case of series-parallel networks of more or less reliable sources with uncertain information flows.

Another interesting perspective is the possibility to learn the behavior of sources by comparing the pieces of information provided by those sources with the ground truth, as done in a simple framework for discounting [10] and contextual discounting [18].

Acknowledgments

Frédéric Pichon was supported in part by a grant from the French National Research Agency through the CSOSG research program (project CAHORS).

References

[1] M. Banerjee, D. Dubois, A simple modal logic for reasoning about revealed beliefs, in: C. Sossai, G. Chemelle, (Eds.), European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2009), LNAI, Verona, Italy, vol. 5590, Springer, 2009, pp. 805–816.
[2] L. Cholvy, Evaluation of information reported: a model in the theory of evidence, in: E. Hüllermeier, R. Kruse, F. Hoffmann, (Eds.), IPMU 2010, Dortmund, Part I, CCIS, vol. 80, Springer-Verlag, Berlin, 2010, pp. 258–267.
[3] I. Couso, S. Moral, Independence concepts in evidence theory, International Journal of Approximate Reasoning 51 (7) (2010) 748–758.
[4] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Annals of Mathematical Statistics 38 (1967) 325–339.
[5] S. Destercke, Independence concepts in evidence theory: some results about epistemic irrelevance and imprecise belief functions, in: Workshop on the Theory of Belief Functions, Brest, France, 2010.
[6] D. Dubois, T. Denœux, Pertinence et sincérité en fusion d'informations, in: Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2009), Annecy, France, 2009. Cépaduès-Editions, pp. 23–30.
[7] D. Dubois, H. Prade, A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets, International Journal of General Systems 12 (3) (1986) 193–226.
[8] D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, Computational Intelligence 4 (4) (1988) 244–264.
[9] D. Dubois, H. Prade, Possibility theory and data fusion in poorly informed environments, Control Engineering Practice 2 (5) (1994) 811–823.
[10] Z. Elouedi, K. Mellouli, Ph. Smets, Assessing sensor reliability for multisensor data fusion with the transferable belief model, IEEE Transactions on System Man and Cybernetics B 34 (1) (2004) 782–787.
[11] R. Haenni, Uncover Dempster's rule where it is hidden, in: Proceedings of the 9th International Conference on Information Fusion (FUSION 2006), Florence, Italy, 2006.
[12] R. Haenni, S. Hartmann, Modeling partially reliable information sources: a general approach based on Dempster–Shafer theory, Information Fusion 7 (4) (2006) 361–379.
[13] J. Hintikka, Knowledge and Belief, Cornell University Press, 1962.
[14] F. Janez, A. Appriou, Théorie de l'évidence et cadres de discernement non exhaustifs, Traitement du Signal 13 (3) (1996) 237–250.
[15] F. Janez, A. Appriou, Theory of evidence and non-exhaustive frames of discernment: Plausibilities correction methods, International Journal of Approximate Reasoning 18 (1998) 1–19.
[16] D. Mercier, T. Denœux, M.-H. Masson, Belief function correction mechanisms, in: B. Bouchon-Meunier, R.R. Yager, J.-L. Verdegay, M. Ojeda-Aciego, L. Magdalena (Eds.), Foundations of Reasoning under Uncertainty, Studies in Fuzziness and Soft Computing, vol. 249, Springer-Verlag, Berlin, 2010, pp. 203–222.
[17] D. Mercier, E. Lefèvre, F. Delmotte, Belief functions contextual discounting and canonical decompositions, International Journal of Approximate Reasoning 53 (2) (2012) 146–158.
[18] D. Mercier, B. Quost, T. Denœux, Refined modeling of sensor reliability in the belief function framework using contextual discounting, Information Fusion 9 (2) (2008) 246–258.
[19] F. Pichon, Belief functions: canonical decompositions and combination rules, Ph.D. Thesis, Université de Technologie de Compiègne, Compiègne, France, 2009.
[20] F. Pichon, Equivalence of correction and fusion schemes under meta-independence of sources, in: Workshop on the Theory of Belief Functions (BELIEF 2010), Brest, France, April 2010.
[21] F. Pichon, T. Denœux, Interpretation and computation of α-junctions for combining belief functions, in: Sixth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA'09), Durham, United Kingdom, July 2009.
[22] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
[23] G. Shafer, A betting interpretation for probabilities and Dempster–Shafer degrees of belief (invited talk), in: Workshop on the Theory of Belief Functions (BELIEF 2010), Brest, France, April 2010.

[24] G. Shafer, A. Tversky, Languages and designs for probability, Cognitive Science 9 (1985) 309–339.
[25] P.P. Shenoy, G. Shafer, Axioms for probability and belief-function propagation, in: R.D. Shachter, T. Levitt, J.F. Lemmer, L.N. Kanal (Eds.), Uncertainty in Artificial Intelligence, vol. 4, North-Holland, 1990, pp. 169–198.
[26] Ph. Smets, Un modèle mathématico-statistique simulant le processus du diagnostic médical, Ph.D. Thesis, Université Libre de Bruxelles, Brussels, Belgium, 1978 (in French).
[27] Ph. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, International Journal of Approximate Reasoning 9 (1) (1993) 1–35.
[28] Ph. Smets, The α-junctions: combination operators applicable to belief functions, in: D.M. Gabbay, R. Kruse, A. Nonnengart, H.J. Ohlbach (Eds.), First International Joint Conference on Qualitative and Quantitative Practical Reasoning (ECSQARU-FAPR'97), Bad Honnef (Germany), 09/06/1997–12/06/1997, Lecture Notes in Computer Science, vol. 1244, Springer, 1997, pp. 131–153.
[29] Ph. Smets, The transferable belief model for quantified belief representation, in: D.M. Gabbay, Ph. Smets (Eds.), Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, Kluwer, Dordrecht, 1998, pp. 267–301.
[30] Ph. Smets, Managing deceitful reports with the transferable belief model, in: 8th International Conference on Information Fusion, Philadelphia, USA, July 2005. doi:10.1109/ICIF.2005.1591953.
[31] Ph. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (1994) 191–243.
[32] R.R. Yager, Arithmetic and other operations on Dempster–Shafer structures, International Journal of Man–Machines Studies 25 (1986) 357–366.

Annexe B

Proposition and learning of some belief function contextual correction mechanisms

---

Article paru dans *International Journal of Approximate Reasoning*, 72 : 4-42, 2016.

International Journal of Approximate Reasoning 72 (2016) 4–42

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Proposition and learning of some belief function contextual correction mechanisms ☆

Frédéric Pichon *, David Mercier, Éric Lefèvre, François Delmotte

Univ. Artois, EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), Béthune F-62400, France

ARTICLE INFO

Article history:
Received 31 December 2014
Received in revised form 12 December 2015
Accepted 12 December 2015
Available online 22 December 2015

Keywords:
Dempster–Shafer theory
Belief functions
Information correction
Discounting

ABSTRACT

Knowledge about the quality of a source can take several forms: it may for instance relate to its truthfulness or to its relevance, and may even be uncertain. Of particular interest in this paper is that such knowledge may also be contextual; for instance the reliability of a sensor may be known to depend on the actual object observed. Various tools, called correction mechanisms, have been developed within the theory of belief functions, to take into account knowledge about the quality of a source. Yet, only a single tool is available to account for contextual knowledge about the quality of a source, and precisely about the relevance of a source. There is thus some lack of flexibility since contextual knowledge about the quality of a source does not have to be restricted to its relevance. The first aim of this paper is thus to try and enlarge the set of tools available in belief function theory to deal with contextual knowledge about source quality. This aim is achieved by (1) providing an interpretation to each one of two contextual correction mechanisms introduced initially from purely formal considerations, and (2) deriving extensions – essentially by uncovering contextual forms – of two interesting and non-contextual correction mechanisms. The second aim of this paper is related to the origin of contextual knowledge about the quality of a source: due to the lack of dedicated approaches, it is indeed not clear how to obtain such specific knowledge in practice. A sound, easy to interpret and computationally simple method is therefore provided to learn from data contextual knowledge associated with the contextual correction mechanisms studied in this paper.

1. Introduction

In today's society, a lot of information is accessible. Yet, for a piece of information to be useful, it must be interpreted with respect to the source that provides it, and in particular in the light of the quality of the source. Clearly, this is no easy task. First, the quality of a source may come in many guises: a source can for instance be biased, or even be totally irrelevant. Second, this quality may be only known with some uncertainty by the agent who has to interpret the piece of information [28].

The theory of belief functions [32,39,36] is a flexible framework to model and deal with uncertainty. Various tools have been developed within this framework to take into account uncertain knowledge about the quality of a source and to modify, or *correct* [19,28], a piece of information provided by the source according to this knowledge. The most common,

☆ This paper is an extended and revised version of [30] and [22].

* Corresponding author.
E-mail addresses: frederic.pichon@univ-artois.fr (F. Pichon), david.mercier@univ-artois.fr (D. Mercier), eric.lefevre@univ-artois.fr (É. Lefèvre), francois.delmotte@univ-artois.fr (F. Delmotte).

http://dx.doi.org/10.1016/j.ijar.2015.12.012
0888-613X/

and historically the first, of such tools is the discounting operation [32,33], which corresponds to the situation where the agent has some knowledge regarding the relevance of the source [28]. The discounting operation is central in numerous and diverse applications of belief function theory, such as classification [5] and information fusion [31,16,40] (see [27, Remarks 5 and 6] for more details on the role of discounting in these applications).

Since its inception, the discounting operation has been extended in different ways. Notably, its inverse, called de-discounting, is introduced and used in [7] to show that two well-known and apparently quite different classifiers based on belief functions, produce actually similar outputs in an important special case. This mechanism allows one to retract a discounting which is judged no longer valid or justified; it has the effect of strengthening, rather than weakening as is the case with discounting, a piece of information. It is applied successfully in a mailing address recognition system [19], where it is used in conjunction with discounting to correct outputs of postal address readers.

Another interesting extension is the correction mechanism proposed recently by Pichon et al. [28], in order to take into account knowledge about the truthfulness of a source, besides its relevance. Its interest resides in the fact that it offers a means to deal with sources that may lie, or that are biased in the case where the source is a sensor. As shown in [28], truthfulness assumptions are also quite interesting in that they can be used to reinterpret all connectives of Boolean logic, which in turn leads to generalize the unnormalized Dempster's rule [4,32] to all Boolean connectives – this rule being the pivotal and most often used combination rule in belief function theory.

Of particular interest in this paper is the fact that the quality of a source may also be contextual; for instance,[1] a thermometer is relevant to measure a temperature which falls within its range, but is typically useless if the temperature is outside of it; if we let $\mathcal{X} = \{-100\,^\circ\text{C}, \ldots, 1000\,^\circ\text{C}\}$ be the possible temperatures, then the context here is the range, which could be, e.g., $\{-38\,^\circ\text{C}, \ldots, 356\,^\circ\text{C}\}$ (range of mercury thermometers). Furthermore, such contextual quality may also be known with some uncertainty; for instance one may believe to some degree that a source is relevant for a given context.

To deal with such contextual knowledge, yet another extension of discounting is introduced by Mercier et al. [23], who consider the case where one has some knowledge about the relevance of the source, conditionally on different subsets (contexts) A of $\mathcal{X}$ such that the set $\mathcal{A}$ of these contexts forms a partition of $\mathcal{X}$, leading to an operation called *contextual discounting based on a coarsening*. Formally, contextual discounting based on a coarsening relies on the disjunctive rule of combination [9,33] and is related to the canonical decomposition of a belief function [34] as highlighted in [20]. This contextual correction mechanism was extended recently by Mercier et al. [21]: the set of contexts $\mathcal{A}$ for which one has some knowledge about the relevance of the source can be arbitrary (it no longer needs to form a partition of $\mathcal{X}$).

Contextual discounting based on a coarsening [23] and its extension uncovered in [21] are, to the best of our knowledge, the only contextual correction mechanisms that have been thoroughly studied in the literature. There is thus clearly a lack of tools to deal with contextual quality, since it does not have to be restricted to contextual relevance. As a matter of fact, Mercier et al. [20] introduce two other contextual correction mechanisms, which are quite interesting from a formal point of view: the first one, referred simply as *contextual discounting* in [20, Theorem 1][2], can be viewed as a generalization of contextual discounting based on a coarsening in that it has the same formal definition as this latter mechanism except that the set $\mathcal{A}$ that appears in its definition can be arbitrary; the second one is a dual reinforcement process to contextual discounting, which has a similar definition as contextual discounting, except that it relies on the unnormalized Dempster's rule, and it can also be linked to the canonical decomposition of a belief function. However, Mercier et al. [20] do not provide an interpretation for this latter correction mechanism, nor do they provide an interpretation for contextual discounting as shown in [21], hence the practical usefulness of these two contextual correction mechanisms remains unknown.

As a first step toward enlarging the set of tools dedicated to handling contextual quality, one may thus try and provide an interpretation to each one of Mercier et al. [20] contextual discounting and reinforcement processes. Mimicking what has been done for discounting with the introduction of contextual discounting based on a coarsening, one may also try and derive contextual forms of correction mechanisms that have already proved interesting in their non-contextual versions; in particular one may attempt to "contextualize" the two extensions of discounting recalled above that are the de-discounting operation and Pichon et al. [28] truthfulness-based correction mechanism. The first aim of this paper is to explore these different routes and to find out whether they can yield useful complements to contextual discounting based on a coarsening and its recent extension [21], with respect to the problem of handling contextual knowledge about the quality of a source. As will be seen, this exploration rests on a detailed analysis of Pichon et al. [28] truthfulness model.

In addition to the above issue of being able to take into account contextual knowledge about the quality of a source, an associated issue is the origin of such knowledge; it is indeed not totally clear how to obtain such specific knowledge in practice. Two different approaches [11,23] have been proposed to find out the contextual quality of a source, and more precisely to discover it from available labeled data. Elouedi et al. [11] approach is based on the use of confusion matrices. Its simplicity makes it quite appealing. However, it is restricted to the case of contextual discounting based on a coarsening, where the coarsening is fixed to the partition of singletons. Besides, it basically amounts to assuming that a source makes a correct prediction only when it is relevant, which is debatable (a non-relevant source may provide correct information, see, e.g., [28]). Mercier et al. [23] approach on the other hand, relies on the minimization of an error criterion. It is quite

[1] Other examples of contextual quality will be given in later sections of this paper.

[2] The operation referred to as contextual discounting in [20, Theorem 1] was thought – erroneously as shown in [21] – to be the extension to an arbitrary set of contexts, of contextual discounting based on a coarsening, hence its name. To ensure consistency with previous published works, the same name is used for this operation in this paper, although the results in [21] suggest this name may be somewhat of a misnomer.

interesting since, in addition to learning the contextual quality of a source, it may be potentially useful to improve the performance of a source in, e.g., a classification application. However, it is restricted to the case of contextual discounting based on a coarsening, where a partition (set of contexts) of $\mathcal{X}$ has been fixed beforehand. The second aim of this paper is therefore to alleviate this latter restriction and more generally to extend Mercier et al. [23] learning approach to the other contextual correction mechanisms studied in this paper, such as Mercier et al. [20] contextual discounting and reinforcement processes.

This paper is organized as follows. Necessary notions on belief function theory and on existing correction mechanisms are recalled in Section 2. A new framework for handling detailed assumptions about the truthfulness of a source is then obtained from a careful analysis of Pichon et al. [28] truthfulness model, which is carried out in two steps (Section 3 then Section 4). Using this framework, an interpretation for each one of Mercier et al. [20] contextual discounting and reinforcement processes is derived (Section 5). Contextual de-discounting is introduced and then used in conjunction with the canonical decomposition, to define an extension of contextual discounting (Section 6). A contextual version of Pichon et al. [28] truthfulness-based correction mechanism is uncovered in Section 7. Learning contextual correction mechanisms from labeled data is addressed in Section 8. Finally, Section 9 concludes the paper.

2. Belief function theory: necessary notions

In this section, we first recall basic concepts of belief function theory. Then, we present existing correction mechanisms that are of interest for this paper.

2.1. Basic concepts

We review in this section the following basic concepts: the representation, combination and canonical decomposition of beliefs.

2.1.1. Representation of beliefs

In this paper, we adopt Smets' Transferable Belief Model (TBM) [39,36], where the beliefs held by an agent regarding the actual value taken by a parameter $\mathbf{x}$ defined on a finite domain, called *frame of discernment*, $\mathcal{X} = \{x_1, \ldots, x_K\}$, are modeled using a belief function [32] and represented using an associated *mass function*. A mass function (MF) on $\mathcal{X}$ is defined as a mapping $m : 2^{\mathcal{X}} \rightarrow [0, 1]$ verifying $\sum_{A \subseteq \mathcal{X}} m(A) = 1$. The mass $m(A)$ represents the subjective probability that the agent knows that the value of $\mathbf{x}$ lies somewhere in set A, and nothing more specific [3,10].

Subsets A of $\mathcal{X}$ such that $m(A) > 0$ are called *focal sets* of m. A MF is said to be: *vacuous* if $\mathcal{X}$ is its only focal set, in which case it is denoted by $m_{\mathcal{X}}$; *inconsistent* if $\emptyset$ is its only focal set, in which case it is denoted by $m_{\emptyset}$; *dogmatic* if $\mathcal{X}$ is not a focal set; *normal* if $\emptyset$ is not a focal set. A non-normal MF m can be transformed into a normal MF m^* by the normalization operation defined as follows, for all $A \subseteq \mathcal{X}$:

$$m^*(A) = \begin{cases} k \cdot m(A) & \text{if } A \neq \emptyset, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

with $k = (1 - m(\emptyset))^{-1}$.

Equivalent representations of a MF m exist. In particular the *belief*, *plausibility*, *commonality* and *implicability* functions are defined, respectively, as:

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B),$$

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B),$$

$$q(A) = \sum_{B \supseteq A} m(B),$$

and

$$b(A) = \sum_{B \subseteq A} m(B),$$

for all $A \subseteq \mathcal{X}$. MF m can be recovered from any of these functions. In particular, we have:

$$m(A) = \sum_{B \supseteq A} (-1)^{|B|-|A|} q(B), \tag{2}$$

for all $A \subseteq \mathcal{X}$, with $|A|$ denoting the cardinality of A. The degree of belief $bel(A)$ evaluates to what extent event A is logically implied by the available evidence and $pl(A)$ evaluates to what extent event A is consistent with the available evidence [10]; the commonality and implicability functions play more of a technical role as will be seen in the next section.

2.1.2. Combination of beliefs

Beliefs can be aggregated using so-called combination rules. In particular, the conjunctive combination rule, or *conjunctive rule* for short, which is the unnormalized version of Dempster's rule [4], is defined as follows. Let m_1 and m_2 be two MFs, and let $m_{1\textcircled{\cap}2}$ be the MF resulting from their combination by the conjunctive rule denoted by $\textcircled{\cap}$. We have:

$$m_{1\textcircled{\cap}2}(A) = \sum_{B\cap C=A} m_1(B)\, m_2(C), \quad \forall A \subseteq \mathcal{X}. \tag{3}$$

The conjunctive rule admits a simple expression in terms of commonality functions:

$$q_{1\textcircled{\cap}2}(A) = q_1(A) \cdot q_2(A), \quad \forall A \subseteq \mathcal{X}, \tag{4}$$

where q_1, q_2 and $q_{1\textcircled{\cap}2}$ denote the commonality functions associated to m_1, m_2 and $m_{1\textcircled{\cap}2}$, respectively. The conjunctive rule is commutative, associative and admits the vacuous MF $m_{\mathcal{X}}$ as neutral element.

Assume now that $m_{1\textcircled{\cap}2}$ has been obtained by combining MFs m_1 and m_2, and then it appears that m_2 is actually not supported by evidence and should thus be removed from $m_{1\textcircled{\cap}2}$. This operation is possible using the inverse of the conjunctive rule [34,6], which may be called the conjunctive decombination rule and denoted by $\textcircled{\cap\!\!\!/}$. We have:

$$m_{1\textcircled{\cap}2}\textcircled{\cap\!\!\!/} m_2 = m_1.$$

Let q_1 and q_2 be the commonality functions associated respectively to any two MFs m_1 and m_2, the conjunctive decombination rule is defined as:

$$q_{1\textcircled{\cap\!\!\!/}2}(A) = \frac{q_1(A)}{q_2(A)}, \quad \forall A \subseteq \mathcal{X}. \tag{5}$$

This operation is well-defined as long as m_2 is non-dogmatic (in which case we have $q_2(A) > 0$ for all A) and $m_{1\textcircled{\cap\!\!\!/}2}$ is a MF (this is not necessarily the case since the quotient of two commonality functions is not always a commonality function).

Other combination rules of interest for this paper are the disjunctive rule $\textcircled{\cup}$ [9,33] and the equivalence rule $\textcircled{\underline{\cap}}$ [35,25]. Their definitions are similar to that of the conjunctive rule: one merely needs to replace $\cap$ in (3) by, respectively, $\cup$ and $\underline{\cap}$, where $\underline{\cap}$ denotes logical equality, i.e., $B \underline{\cap} C = (B \cap C) \cup (\overline{B} \cap \overline{C})$ for all $B, C \subseteq \mathcal{X}$, where $\overline{A}$ denotes the complement of some $A \subseteq \mathcal{X}$. The interpretations of these three rules are discussed in detail in [28].

The disjunctive rule has a simple expression in terms of implicability functions, which is the counterpart of (4):

$$b_{1\textcircled{\cup}2}(A) = b_1(A) \cdot b_2(A), \quad \forall A \subseteq \mathcal{X}.$$

The disjunctive rule is commutative, associative and admits the inconsistent MF $m_\emptyset$ as neutral element. Besides, as for the conjunctive rule, an inverse operation may be defined for $\textcircled{\cup}$ [6]:

$$b_{1\textcircled{\cup\!\!\!/}2}(A) = \frac{b_1(A)}{b_2(A)}, \quad \forall A \subseteq \mathcal{X}.$$

This operation, referred to as disjunctive decombination, is well-defined as long as m_2 is non-normal, since in this case we have $b_2(A) > 0$ for all A.

We may note that a similar expression as (4), i.e., a simple pointwise product expression, exists for the rule $\textcircled{\underline{\cap}}$. We defer its introduction for clarity of presentation.

2.1.3. Canonical decompositions

Following Shafer [32, Chapter 4] (see also [34,6]), a MF m may be called *conjunctively separable*, or, for short, $\textcircled{\cap}$*-separable*, if it can be obtained as the result of the combination by the conjunctive rule of so-called *simple* MFs, which are MFs having at most two focal sets, including the frame of discernment $\mathcal{X}$ [6].

A simple MF having focal sets $\mathcal{X}$ and $A \subset \mathcal{X}$, with respective masses w and $1-w$, $w \in [0,1]$, may be simply denoted by A^w; for instance, a MF defined on $\mathcal{X} = \{x_1, x_2, x_3\}$ and having focal sets $\mathcal{X}$ and $\{x_1, x_3\}$, with respective masses 0.7 and 0.3, may be denoted by $\{x_1, x_3\}^{0.7}$. Using this notation, every non-dogmatic $\textcircled{\cap}$-separable MF m may be uniquely expressed as [34,6]:

$$m = \textcircled{\cap}_{A\subset\mathcal{X}} A^{w(A)}, \tag{6}$$

with $w(A) \in (0,1]$ for all $A \subset \mathcal{X}$. Example 1 illustrates Equation (6).

Example 1. Let $\mathcal{X} = \{x_1, x_2, x_3\}$ and m be a non-dogmatic MF defined on $\mathcal{X}$ by:

$$m(\{x_1\}) = 0.6,$$
$$m(\{x_1, x_3\}) = 0.12,$$
$$m(\mathcal{X}) = 0.28.$$

We have

$$m = \emptyset^1 \textcircled{\cap} \{x_1\}^{0.4} \textcircled{\cap} \{x_2\}^1 \textcircled{\cap} \{x_3\}^1 \textcircled{\cap} \{x_1, x_2\}^1 \textcircled{\cap} \{x_1, x_3\}^{0.7} \textcircled{\cap} \{x_2, x_3\}^1,$$

or, equivalently, $m = \textcircled{\cap}_{A \subset \mathcal{X}} A^{w(A)}$, with

$$w(A) = \begin{cases} 0.4 & \text{if } A = \{x_1\}, \\ 0.7 & \text{if } A = \{x_1, x_3\}, \\ 1 & \forall A \in 2^{\mathcal{X}} \setminus \{\{x_1\}, \{x_1, x_3\}, \mathcal{X}\}. \end{cases}$$

Hence, m is a $\textcircled{\cap}$-separable MF.

For any non-dogmatic $\textcircled{\cap}$-separable MF m uniquely expressed as (6), let us define the sets $\mathcal{C} = \{A | A \subset \mathcal{X}, w(A) < 1\}$ and $\mathcal{W} = \{A^{w(A)} | A \in \mathcal{C}\}$; for instance, for the MF m in Example 1, we have $\mathcal{C} = \{\{x_1\}, \{x_1, x_3\}\}$ and $\mathcal{W} = \{\{x_1\}^{0.4}, \{x_1, x_3\}^{0.7}\}$. We will refer to set $\mathcal{C}$ associated to a non-dogmatic $\textcircled{\cap}$-separable MF m, as its *conjunctive core*. Clearly, since A^1 is equivalent to the vacuous mass function $m_{\mathcal{X}}$ for any $A \subset \mathcal{X}$ and as $m_{\mathcal{X}}$ is a neutral element for $\textcircled{\cap}$, Equation (6) reduces to

$$m = \textcircled{\cap}_{A \in \mathcal{C}} A^{w(A)}, \tag{7}$$

for any non-dogmatic $\textcircled{\cap}$-separable MF m such that $m \neq m_{\mathcal{X}}$,[3] i.e., m can be uniquely expressed as the conjunctive combination of the simple MFs in $\mathcal{W}$. For instance, for MF m in Example 1, we have $m = \{x_1\}^{0.4} \textcircled{\cap} \{x_1, x_3\}^{0.7}$.

Smets [34] further shows that in fact any non-dogmatic MF can be obtained from simple MFs. More precisely, let m be a non-dogmatic MF, then it may be uniquely expressed as the conjunctive decombination of two non-dogmatic $\textcircled{\cap}$-separable MFs, that is, as:

$$m = m^c \textcircled{\not\cap} m^d, \tag{8}$$

where m^c and m^d are non-dogmatic $\textcircled{\cap}$-separable MFs, such that their conjunctive cores denoted respectively by $\mathcal{C}^c$ and $\mathcal{C}^d$ satisfy $\mathcal{C}^c \cap \mathcal{C}^d = \emptyset$, as illustrated by Example 2 below. This decomposition into simple MFs of a non-dogmatic MF m is referred to as the *conjunctive canonical decomposition* of m. The m^c and m^d components in (8) are called the *confidence and diffidence components*, respectively, of m by Smets [34], who proposed to view m^c as representing positive evidence ("good reasons to believe") in some propositions $A \subseteq \mathcal{X}$, and m^d as representing negative evidence ("good reasons not to believe") in some other propositions.

Example 2. (Based on Example 2 of [34].) Let $\mathcal{X} = \{x_1, x_2, x_3\}$ and m be a MF defined on $\mathcal{X}$ by:

$$m(\{x_1, x_2\}) = 1/3,$$
$$m(\{x_1, x_3\}) = 1/3,$$
$$m(\mathcal{X}) = 1/3.$$

We have $m = m^c \textcircled{\not\cap} m^d$, with $m^c = \textcircled{\cap}_{A \subset \mathcal{X}} A^{w^c(A)}$ where

$$w^c(A) = \begin{cases} 0.5 & \text{if } A = \{x_1, x_2\}, \\ 0.5 & \text{if } A = \{x_1, x_3\}, \\ 1 & \forall A \in 2^{\mathcal{X}} \setminus \{\{x_1, x_2\}, \{x_1, x_3\}, \mathcal{X}\}, \end{cases}$$

and with $m^d = \textcircled{\cap}_{A \subset \mathcal{X}} A^{w^d(A)}$ where

$$w^d(A) = \begin{cases} 0.75 & \text{if } A = \{x_1\}, \\ 1 & \forall A \in 2^{\mathcal{X}} \setminus \{\{x_1\}, \mathcal{X}\}. \end{cases}$$

Therefore, $\mathcal{C}^c = \{\{x_1, x_2\}, \{x_1, x_3\}\}$, $\mathcal{C}^d = \{\{x_1\}\}$ and $\mathcal{C}^c \cap \mathcal{C}^d = \emptyset$.

As shown in [6], it is possible to obtain a relation based on the disjunctive rule $\textcircled{\cup}$ that is the counterpart to (8). Let us call $\textcircled{\cup}$-*separable* a MF that can be obtained as the result of the combination by the disjunctive rule of so-called *negative simple* MFs, which are MFs having at most two focal sets, including the empty set $\emptyset$ [6]. A negative simple MF having focal sets $\emptyset$ and $A \neq \emptyset$, with respective masses v and $1 - v$, $v \in [0, 1]$, may be simply denoted by A_v; for instance, a MF defined on $\mathcal{X} = \{x_1, x_2, x_3\}$ and having focal sets $\emptyset$ and $\{x_1, x_3\}$, with respective masses 0.7 and 0.3, may be denoted by $\{x_1, x_3\}_{0.7}$. Every non-normal $\textcircled{\cup}$-separable MF m may then be uniquely expressed as [6]:

[3] If $m = m_{\mathcal{X}}$, i.e., m is vacuous, then we have $w(A) = 1$ for all $A \subset \mathcal{X}$ and thus $\mathcal{C} = \emptyset$.

$$m = \textcircled{\cup}_{A \neq \emptyset} A_{v(A)}, \tag{9}$$

with $v(A) \in (0, 1]$ for all $A \neq \emptyset$. For any non-normal $\textcircled{\cup}$-separable MF m uniquely expressed as (9), we refer to the set $\mathcal{C}^{disj} = \{A | A \neq \emptyset, v(A) < 1\}$ as the *disjunctive core* of m.

Then, any non-normal MF m may be uniquely expressed as the disjunctive decombination of two non-normal $\textcircled{\cup}$-separable MFs, that is, as:

$$m = m^{c,disj} \textcircled{\emptyset} m^{d,disj}, \tag{10}$$

where $m^{c,disj}$ and $m^{d,disj}$ are non-normal $\textcircled{\cup}$-separable MFs, such that their disjunctive cores denoted respectively by $\mathcal{C}^{c,disj}$ and $\mathcal{C}^{d,disj}$ satisfy $\mathcal{C}^{c,disj} \cap \mathcal{C}^{d,disj} = \emptyset$. This decomposition into negative simple MFs of a non-normal MF is referred to as its *disjunctive canonical decomposition*.

2.2. Correction mechanisms

Knowledge about the reliability of a source is classically taken into account in the TBM through the *discounting* operation [32,33]. Suppose a source S providing a piece of information represented by a MF m_S. Let β, with $\beta \in [0, 1]$, be the agent's degree of belief that the source is reliable. The agent's belief m on $\mathcal{X}$ is then defined by:

$$m(A) = \beta\, m_S(A) + (1 - \beta) m_{\mathcal{X}}(A), \quad \forall A \subseteq \mathcal{X}. \tag{11}$$

Remarkably, Equation (11) is also obtained if the agent assumes that the source is reliable with mass β and not reliable with mass $1 - \beta$, rather than if he assumes that the source is reliable with degree of belief β [23,28]. This alternative interpretation of discounting may be sometimes more instructive when discounting needs to be compared with other correction mechanisms.

Denœux and Smets [7] introduce a correction mechanism, called *de-discounting*, which is basically the inverse of the discounting operation. Assume an agent that receives a mass function m_S from a source, m_S being the result of a discounting with degree β, $1 - m_S(\mathcal{X}) \leq \beta \leq 1$, of some mass function m. Assume further that the agent believes that this discounting is not valid. He can then recover m using de-discounting:

$$m(A) = \frac{m_S(A) - (1 - \beta) m_{\mathcal{X}}(A)}{\beta}, \quad \forall A \subseteq \mathcal{X}. \tag{12}$$

Mercier et al. [23] consider the case where the agent has some knowledge about the reliability of a source, conditionally on different subsets (contexts) A of $\mathcal{X}$ such that the set of these contexts forms a partition of $\mathcal{X}$. Precisely, let β_A, with $\beta_A \in [0, 1]$, be the agent's degree of belief that the source is reliable in context $A \subseteq \mathcal{X}$ and let $\mathcal{A}$ be the set of contexts for which the agent possesses such contextual knowledge, where $\mathcal{A}$ forms a partition of $\mathcal{X}$. The agent's belief m on $\mathcal{X}$ is then defined by [23]

$$m = m_S \textcircled{\cup} (\textcircled{\cup}_{A \in \mathcal{A}} A_{\beta_A}),$$

or, for short (with an obvious abuse of notation already used in [20]),

$$m = m_S \textcircled{\cup}_{A \in \mathcal{A}} A_{\beta_A}, \tag{13}$$

where A_{β_A} denotes the negative simple MF having focal sets $\emptyset$ and A with respective masses β_A and $1 - \beta_A$. Equation (13) is known as *contextual discounting based on a coarsening*. It extends discounting defined by (11), which can be expressed as [23]:

$$m = m_S \textcircled{\cup} \mathcal{X}_\beta, \tag{14}$$

where $\mathcal{X}_\beta$ denotes the negative simple MF having focal sets $\emptyset$ and $\mathcal{X}$ with respective masses β and $1 - \beta$, i.e., discounting is a particular case of contextual discounting based on a coarsening, which is recovered for $\mathcal{A} = \{\mathcal{X}\}$. Moreover, Mercier et al. [20] provide an equivalent representation for (13) using the fact that the term $\textcircled{\cup}_{A \in \mathcal{A}} A_{\beta_A}$ in (13) constitutes a MF whose canonical decomposition into negative simple MFs is direct (from its definition one can see that it is a $\textcircled{\cup}$-separable MF). This other representation is based on the so-called disjunctive weight function [6] (we refer the interested reader to [20] for details on this other representation).

Recently, Mercier et al. [20,21] consider the more general case where the agent has some knowledge about the reliability of a source, conditionally on different subsets (contexts) A of $\mathcal{X}$, but where the set of these contexts can be arbitrary, that is do not need to form a partition of $\mathcal{X}$. Let β_A, with $\beta_A \in [0, 1]$, be the agent's degree of belief that the source is reliable in context $A \subseteq \mathcal{X}$ and let $\mathcal{A}$ be the set of contexts for which the agent possesses such contextual knowledge. The agent's belief m on $\mathcal{X}$ is then defined by [21]:

$$m = m_S \textcircled{\cup} (\textcircled{\cap}_{A \in \mathcal{A}} \overline{A}^{1 - \beta_A}). \tag{15}$$

In the correction schemes recalled above, the reliability of a source is assimilated to its relevance as explained in [28]. In [28], Pichon et al. assume that the reliability of a source involves in addition another dimension: its truthfulness. Pichon et al. [28] note that there exists various forms of lack of truthfulness for a source. However, Pichon et al. [28] study only the crudest description of the lack of truthfulness, where a non-truthful source is a source that declares the contrary of what it knows. According to this definition, from a piece of information of the form $\mathbf{x} \in B$ for some $B \subseteq \mathcal{X}$ provided by a relevant source S, one must conclude that $\mathbf{x} \in B$ or $\mathbf{x} \in \overline{B}$, depending on whether the source S is assumed to be truthful or not. More generally, suppose that S provides a piece of information represented by a MF m_S and that the agent thinks that the source is truthful with mass β and non-truthful with mass $1-\beta$. Then, his belief m on $\mathcal{X}$ is defined by [28]:

$$m(A) = \beta \cdot m_S(A) + (1-\beta) \cdot \overline{m}_S(A), \quad \forall A \subseteq \mathcal{X}, \tag{16}$$

where $\overline{m}_S$ denotes the *negation* of MF m_S defined as $\overline{m}_S(A) = m_S(\overline{A})$, $\forall A \subseteq \mathcal{X}$ [9]. The operation defined by (16) may be called *negating* of a belief function, since m becomes closer to the negation $\overline{m}_S$ of m_S as β approaches 0. This is in contrast with the discounting operation, for which m becomes closer to the vacuous mass function $m_{\mathcal{X}}$ as β approaches 0.

Finally, Mercier et al. [20] introduce formally two contextual correction mechanisms, called *contextual discounting* (CD) and *contextual reinforcement* (CR) hereafter. They are defined as follows. Let m_S be a MF provided by a source S. Then, the CD of m_S is the MF m defined by:

$$m = m_S \textcircled{\cup}_{A \in \mathcal{A}} A_{\beta_A}, \tag{17}$$

and the CR of m_S is the MF m defined by:

$$m = m_S \textcircled{\cap}_{A \in \mathcal{A}} A^{\beta_A}, \tag{18}$$

with $\beta_A \in [0,1]$, $A \in \mathcal{A}$, for some subset $\mathcal{A}$ of $2^{\mathcal{X}}$. CD (17) is clearly a straightforward formal generalization of contextual discounting based on a coarsening (13), the mere difference being that $\mathcal{A}$ in (17) do not need to form a partition of $\mathcal{X}$ contrary to that in (13). Mercier et al. [20] show that CR amounts to the negation of the CD of the negation of m_S. However, they do not go further in providing a clear explanation as to what knowledge about the behavior of the source this correction of m_S correspond, nor do they provide an interpretation for CD as shown by [21]. One of the main results of this paper is to provide an interpretation for both CD and CR; it relies on an extension of Pichon et al. [28] truthfulness model, which is introduced in two steps (Section 3 then Section 4).

3. Contextual truthfulness

As recalled in Section 2.2, Pichon et al. [28] study only a rudimentary form of non-truthfulness. In this section, a detailed analysis of Pichon et al. [28] truthfulness model is first conducted. This analysis then leads naturally to a refined model of source truthfulness that allows the integration of more subtle knowledge about the lack of truthfulness of an information source.

3.1. Analysis of Pichon et al. truthfulness model

Let us review and analyze in some details Pichon et al. [28] truthfulness model. This analysis will be informed by the following example (Example 3) borrowed from [14], which will be subsequently adapted to try and provide new insights on Pichon et al. truthfulness model.

Example 3. (See Example 1[4] of [14].) Suppose a murder has been committed. There are three suspects: *Peter*, *John*, and *Mary*. In the belief function framework, the set $\mathcal{X} = \{Peter, John, Mary\}$ can be seen as the frame of discernment associated to the parameter $\mathbf{x}$ representing the murderer.

Suppose a witness who is aware that the three suspects are *Peter*, *John*, and *Mary*, and that tells that the murderer was a man. This piece of information is equivalent to an agent who receives it, to the testimony $\mathbf{x} \in B$, with $B = \{Peter, John\}$. Hence, based on this evidence, the following MF representing the agent's belief on the murderer can be constructed:

$$m(\{Peter, John\}) = 1. \tag{19}$$

This kind of example is quite common in the literature on belief functions and is often used to illustrate notions of the framework (see in particular the original problem "The murder of Mr. Jones" in Smets and Kennes [39]). Example 3 is admittedly quite simple, yet if we take a closer look at it, a fact that seems trivial but that will nonetheless be instrumental for our analysis, can be noticed. Indeed, by remarking that the witness is actually providing information on the parameter

[4] This is actually only an excerpt of [14, Example 1], which has furthermore been slightly modified here to fit the formalism and terminology used in the present paper, and to serve our purpose.

of interest $\mathbf{x}$ via an auxiliary variable $\mathbf{y}$ with domain $\mathcal{Y} = \{male, \neg male\}$, we can notice that an implicit assumption is made in this example: the witness must have the *same* view as the agent on who is a man among the suspects, i.e., *Peter* and *John*, for the conclusion reached by the agent to be proper, that is, formally, they must both think that the variables $\mathbf{y}$ and $\mathbf{x}$ are related by the mapping $\rho : \mathcal{Y} \to 2^{\mathcal{X}}$ defined by:

$$\rho(\{man\}) = \{Peter, John\}, \quad \rho(\{\neg man\}) = \{Mary\},$$

and known as a *refining* [32]. Obviously, in this example, the witness most certainly has the same view as the agent on who is a man, thus this assumption can be safely kept implicit and the conclusion reached by the agent is sound.

Now, as recalled in Section 2.2, Pichon et al. [28] consider a truthfulness model where an agent should deduce that $\mathbf{x} \in \overline{B}$ from a piece of information $\mathbf{x} \in B$ provided by a non-truthful source, assuming that a non-truthful source is a source that declares the contrary of what it knows. For instance, if the witness in Example 3 had declared that the murderer was *Mary* or *John*, and if the agent had believed that the witness was non-truthful, then the agent should have deduced that the murderer was *Peter*. Let us further note that Pichon et al. [28] also call a non-truthful source, a *lying* source, and they explicitly write that they "use the term *lying* as a synonym of *not telling the truth*, irrespective of the existence of any intention of a source to deceive" [28]. The same applies in this paper, where we use this term to ensure continuity with previous related works. Yet, we may also use the term *biased*, which is perhaps a more appropriate and less connoted term. Hence, a source that is considered to be lying may equivalently be said to be biased.

We can see at least two practical situations where one would need to use such a truthfulness model and negate the information provided by a source. The first one is perhaps the most obvious and easy to understand: when a source lies intentionally, i.e., has the intention to deceive, and chooses the most simple and common strategy to deceive, i.e., the crudest kind of lie, which is to tell the contrary of what it knows. The second situation is one where the source is non-truthful unintentionally, which may be the case when the source has a *different* (precisely opposite) view from that of the agent on the relation between an auxiliary variable and the parameter of interest, as illustrated by Example 4.

Example 4. Suppose a murder has been committed. There are four male suspects: *Eloy*, *Conrad*, *Linus* and *Aeneas*. Let $\mathcal{X} = \{Eloy, Conrad, Linus, Aeneas\}$ be the domain associated to the parameter $\mathbf{x}$ representing the murderer.

Suppose an agent who sees the suspects and that only *Eloy* and *Conrad* have a beard. Assume further that a witness, *Jane*, is aware that the suspects are *Eloy*, *Conrad*, *Linus* and *Aeneas*, and that tells that the murderer had a beard. This piece of information provided by *Jane* is thus equivalent to the agent to the testimony $\mathbf{x} \in B$, with $B = \{Eloy, Conrad\}$.

Suppose the agent learns some time after receiving this testimony that *Jane* has actually not met the suspects recently and in particular she is not aware that since she has last seen them, each of the suspects has changed his beard situation (those that did not have a beard have grown one and those that had one have cut it). In other words, her knowledge ρ_{Jane} about the relation between the auxiliary variable $\mathbf{y}$ defined on $\mathcal{Y} = \{beard, \neg beard\}$ and the parameter of interest $\mathbf{x}$ is outdated and is actually the opposite of the agent's knowledge ρ_{Ag} on this relation:

$$\rho_{Jane}(\{beard\}) = \rho_{Ag}(\{\neg beard\}) = \{Linus, Aeneas\},$$
$$\rho_{Jane}(\{\neg beard\}) = \rho_{Ag}(\{beard\}) = \{Eloy, Conrad\}.$$

This means that through the piece of information "the murderer had a beard", that was equivalent to the agent to the testimony $\mathbf{x} \in B = \{Eloy, Conrad\}$, *Jane* mislead (unintentionally) the agent about what she knows of the guilt of each suspect, and in particular she told the opposite of what she knows, since she actually knows that $\mathbf{x} \in \{Linus, Aeneas\}$.

Hence, from his knowledge on her bias with respect to the beard situation of all the suspects, and in particular that she tells the opposite of what she knows, i.e., is non-truthful, the agent should deduce from *Jane*'s piece of information $\mathbf{x} \in B = \{Eloy, Conrad\}$ that in fact $\mathbf{x} \in \overline{B} = \{Linus, Aeneas\}$, which is indeed what *Jane* actually knows about the murderer.

Further insight on Pichon et al. [28] truthfulness model may be gained by examining precisely what happens when one deduces $\mathbf{x} \in \overline{B}$ from assuming that a source providing testimony $\mathbf{x} \in B$ is lying, that is, tells the contrary of what it knows. In details, it means that the source is assumed to be lying, i.e., telling the contrary of what it knows, whatever it is telling concerning each of the possible values $x \in \mathcal{X}$ that admits parameter $\mathbf{x}$, since one must invert what the source tells for each of these values, as illustrated by Example 5.

Example 5. Without lack of generality, assume for instance $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ and that the source tells $\mathbf{x} \in B = \{x_3, x_4\}$, i.e., it tells that x_3 and x_4 are possible values for $\mathbf{x}$ and it tells that x_1 and x_2 are not possible values for $\mathbf{x}$, then one must deduce that $\mathbf{x} \in \overline{B} = \{x_1, x_2\}$, i.e., x_3 and x_4 are not possible values for $\mathbf{x}$ and x_1 and x_2 are possible values for $\mathbf{x}$.

This leads us to introduce the following definition.

Definition 1. A source is said to be truthful (resp. non-truthful) *for a value* $x \in \mathcal{X}$, when it tells what it knows (resp. the contrary of what it knows) for this value.

Table 1
Non-truthful in $\overline{A}$.

$x \in B$	$x \in A$	ℓ_A
0	0	1
0	1	0
1	0	0
1	1	1

According to this new terminology, a non-truthful source in Pichon et al. [28] truthfulness model is then a source that is non-truthful for *ALL* values $x \in \mathcal{X}$ (and a truthful source is a source that is truthful for all values $x \in \mathcal{X}$).

This analysis allows us to highlight that the crude form of non-truthfulness studied in [28] is actually a quite strong model of the lack of truthfulness of an information source, and as such might only be suitable for a limited number of practical situations such as the ones discussed above. It seems thus interesting to study more subtle variants of this model, and in particular to relax its assumption about the truthfulness of the source for each $x \in \mathcal{X}$: a source could be non-truthful only for *SOME* values $x \in \mathcal{X}$ (and truthful for all other values $x \in \mathcal{X}$).

This alternative should add some flexibility in terms of knowledge that can be taken into account about a source lack of truthfulness, and could thus be interesting from an applicative point of view. Such a study is carried out in the next section.

3.2. Contextual liar

Let us consider the case of a source assumed to be non-truthful for *some* values $x \in \mathcal{X}$, and to be truthful for all other values $x \in \mathcal{X}$. Let $A \subseteq \mathcal{X}$ be the set of values for which the source is assumed to be truthful, and $\overline{A}$ the set of values for which it is assumed to be non-truthful. For short, we may say that the source is truthful in A and non-truthful in $\overline{A}$, or even more simply, when no confusion is possible, that the source is non-truthful in $\overline{A}$ (or biased in $\overline{A}$) – it will then be implicit that the source is truthful in A.

Definition 2 *(Non-truthful in $\overline{A}$).* A source is said to be *non-truthful in* $\overline{A}$ if it is truthful for all $x \in A$, and non-truthful for all $x \in \overline{A}$. This state of the source is denoted by ℓ_A.

Intuitively, this state corresponds simply to a source that lies only for a subset of values, that is, it tells the opposite of what it knows for each value in this set, and tells what it knows for the values outside of this set. If we call this latter set a *context*, then the source may be seen and referred to as a contextual liar.

A sensible question is then: what must one conclude about $\mathbf{x}$ when the source tells $\mathbf{x} \in B$ and is assumed to be in state ℓ_A? The answer is provided by Proposition 1.

Proposition 1. *If a source tells $\mathbf{x} \in B$ and is assumed to be non-truthful in $\overline{A}$, one must deduce that $\mathbf{x} \in B \underline{\cap} A$.*

Proof. To prove this proposition, one merely needs to look in turn at each $x \in \mathcal{X}$ and to find which one of the following four cases applies:

1. If the source tells x is possibly the actual value of $\mathbf{x}$, i.e., the information $\mathbf{x} \in B$ provided by the source is such that $x \in B$,
   (a) And if the source is assumed to be truthful for x, i.e., $x \in A$, then one must conclude that x is possibly the actual value of $\mathbf{x}$;
   (b) And if the source is assumed to be non-truthful for x, i.e., $x \in \overline{A}$, then one must conclude that x is not a possibility for the actual value of $\mathbf{x}$;
2. If the source tells x is not a possibility for the actual value of $\mathbf{x}$, i.e., $x \notin B$,
   (a) And if the source is assumed to be truthful for x, i.e., $x \in A$, then one must conclude that x is not a possibility for the actual value of $\mathbf{x}$;
   (b) And if the source is assumed to be non-truthful for x, i.e., $x \in \overline{A}$, then one must conclude that x is possibly the actual value of $\mathbf{x}$.

Table 1 synthesizes these four cases: it lists exhaustively, i.e., for all possible cases with respect to the membership of a given value $x \in \mathcal{X}$ to the sets B and A, whether one should deduce that this value x is possibly the actual value of $\mathbf{x}$ or not – the former is indicated by a 1 and the latter by a 0 in column ℓ_A. According to Table 1, when the source is assumed to be in state ℓ_A, then one should deduce that $x \in \mathcal{X}$ is a possible value for $\mathbf{x}$ iff x belongs to both B and A or does not belong to both B and A (which corresponds to logical equality), and therefore, since this holds for all $x \in \mathcal{X}$, one should deduce that $\mathbf{x} \in (B \cap A) \cup (\overline{B} \cap \overline{A}) = B \underline{\cap} A$. □

Example 6. As an illustration of Proposition 1, assume for instance $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ and that the source tells $\mathbf{x} \in B = \{x_3, x_4\}$. Furthermore, assume the source is in state $\ell_{\{x_1,x_3\}}$, i.e., is non-truthful for x_2 and for x_4. Then, one should deduce that $\mathbf{x} \in \{x_3, x_4\} \underline{\cap} \{x_1, x_3\} = \{x_2, x_3\}$.

Remark 1. The non-truthful state considered by Pichon et al. [28], which corresponds to a source that is non-truthful for all values $x \in \mathcal{X}$, is equivalent to the state $\ell_\emptyset$, and is thus a particular case of the states ℓ_A, $A \subseteq \mathcal{X}$.

Remark 2. The states ℓ_A, $A \subseteq \mathcal{X}$, correspond to the states used by Pichon [25] to provide an interpretation for the α-conjunctions [35,37], which is to our knowledge the first one to consider such states (but they are not introduced and discussed with as many details in [25] as they are in this paper).

In practice, similarly to what we have done in Section 3.1 for the non-truthful state considered by Pichon et al. [28], we can distinguish two situations where states ℓ_A may be useful. First, the source may be intentionally non-truthful in $\overline{A}$, simply to deceive an agent, yet in a more subtle way than what is allowed by Pichon et al. [28] non-truthful state. Indeed, the source may think that to better deceive the agent, it is going to lie, i.e., tell the opposite of what it knows, but only for a subset of values rather than for all possible values that can take the parameter of interest. Second, in the unintentional case, such a state ℓ_A can be explained by a *difference* between the source and the agent who receives the piece of information provided by the source, on the relation they respectively believe holds between an auxiliary variable and the parameter of interest. More precisely, such a state can be encountered when the source is "wrong" on this relation – wrong with respect to the agent's knowledge – but only for a subset of values in $\mathcal{X}$, as illustrated by Example 7.

Example 7 *(Example 4 continued).* Consider again the setting of Example 4, but this time assume the agent learns some time after receiving Jane's testimony that she has actually not met recently SOME of the suspects, and in particular she is unaware that since she has last seen these suspects, each of them has changed his beard situation. In other words, her knowledge ρ_{Jane} about the relation between $\mathbf{y}$ and $\mathbf{x}$ is partially outdated and is actually partially the opposite of the agent's knowledge ρ_{Ag} on this relation. Indeed, let *Conrad* and *Linus* be these suspects that she has not met recently, we have then on the one hand

$$\rho_{Ag}(\{beard\}) = \{Eloy, Conrad\}, \quad \rho_{Ag}(\{\neg beard\}) = \{Linus, Aeneas\},$$

and on the other hand

$$\rho_{Jane}(\{beard\}) = \{Eloy, Linus\}, \quad \rho_{Jane}(\{\neg beard\}) = \{Conrad, Aeneas\}.$$

This means that through the piece of information "the murderer had a beard", that was equivalent to the agent to the testimony $\mathbf{x} \in B = \{Eloy, Conrad\}$, *Jane* mislead unintentionally (because of her partially outdated knowledge about the beard situation of the suspects) the agent about what she knows of the guilt of the two suspects *Conrad* and *Linus*, since she actually knows that $\mathbf{x} \in \{Eloy, Linus\}$. Indeed, she told the opposite of what she knows for these two suspects, since, e.g., she told that *Conrad* was possibly the murderer ($Conrad \in B$) whereas she knows that he is not. On the contrary, she told the truth about *Eloy* and *Aeneas*, since, e.g., she told that *Eloy* was possibly the murderer ($Eloy \in B$) and she knows that he is possibly the murderer.

Hence, since she told the truth for the suspects in $A = \{Eloy, Aeneas\}$ and lied (i.e., was non-truthful) for the suspects in $\overline{A}$, or in other words is in state $\ell_{\{Eloy,Aeneas\}}$, the agent should deduce from *Jane*'s piece of information $\mathbf{x} \in B$ that in fact

$$\mathbf{x} \in B \underline{\cap} A = \{Eloy, Conrad\} \underline{\cap} \{Eloy, Aeneas\} = \{Eloy, Linus\},$$

which is indeed what the source actually knows about the murderer.

This section has introduced a refined model of source truthfulness, which allows one to account for a contextual lack of truthfulness of a source. This model has been obtained by relaxing a strong assumption underlying Pichon et al. [28] truthfulness model, which was brought to light by a detailed analysis of this latter model. The next section will show that it is possible to push this analysis further and reveal another strong assumption underlying Pichon et al. [28] truthfulness model.

4. Polarized truthfulness

In this section, the analysis of Pichon et al. [28] truthfulness model started in Section 3 is pursued. This analysis then yields a further refined model of source truthfulness.

4.1. Analysis of Pichon et al. truthfulness model (continued)

Let us consider again Example 5: according to Pichon et al. [28] truthfulness model, when a source tells $\mathbf{x} \in B = \{x_3, x_4\}$ and is assumed to be non-truthful, one must deduce that $\mathbf{x} \in \overline{B} = \{x_1, x_2\}$, that is:

- the source tells that x_3 is a possible value for **x**, and one must deduce that x_3 is actually not a possible value for **x**;
- it tells that x_4 is a possible value for **x**, and one must deduce that it is not;
- it tells that x_1 is not a possible value for **x**, and one must deduce that it is;
- it tells that x_2 is not a possible value for **x**, and one must deduce that it is.

To characterize what is at stake in the above reasoning, we introduced Definition 1, that is, the notion of truthfulness for a value $x \in \mathcal{X}$, and revealed that a non-truthful source in Pichon et al. [28] sense, is a source that is non-truthful for each value $x \in \mathcal{X}$ since it amounts to assuming that it tells the contrary of what it knows for each of those values.

Actually, one can be even more specific about the assumptions underlying Pichon et al. [28] truthfulness model, by distinguishing between positive clauses and negative clauses, also known as clauses having positive polarity and negative polarity (in a grammatical sense; see, e.g., [13, Chapter 8]), told by the source. For instance, when the source tells that x_3 is a possible value for **x**, this is a positive clause told by the source, and when the source tells that x_1 is not a possible value for **x**, it is a negative clause.

We may then characterize more finely the truthfulness of the source for each $x \in \mathcal{X}$, that is, with respect to the polarity of the clauses it tells.

Definition 3. A source is said to be *positively* truthful (resp. *positively* non-truthful) for a value $x \in \mathcal{X}$, when it tells that x is a possible value for **x** and knows that it is (resp. it is not) a possible value for **x**.

Definition 4. A source is said to be *negatively* truthful (resp. *negatively* non-truthful) for a value $x \in \mathcal{X}$, when it tells that x is not a possible value for **x** and knows that it is not (resp. it is) a possible value.

According to this terminology, a source which is assumed to be non-truthful for $x \in \mathcal{X}$ (Definition 1), is assumed to be positively *AND* negatively non-truthful for x, since whatever it may tell about x (be it a positive clause or a negative clause), it is assumed to tell the contrary of what it knows. Most importantly, a non-truthful source in Pichon et al. [28] truthfulness model is then a source that is assumed to be positively *AND* negatively non-truthful, for *ALL* values $x \in \mathcal{X}$ (and a truthful source is a source that is positively and negatively truthful, for all values $x \in \mathcal{X}$).

This finer analysis reinforces the statement made in Section 3.1: the crude form of non-truthfulness studied in [28] is actually a rather strong model of the lack of truthfulness of an information source. It make two assumptions, one on the context (set of values) concerned by the lack of truthfulness and one on the polarity of the lack of truthfulness, both of which are strong: the values concerned by the lack of truthfulness are *all* the values of the frame, and *both* polarities (positive and negative) are concerned by the lack of truthfulness.

Here again, it seems interesting to search for and study more subtle variants of this crude model, to obtain more versatility. This amounts to relaxing further the model, that is, relaxing the two above assumptions about the truthfulness of the source for each $x \in \mathcal{X}$: a source could be positively OR negatively non-truthful for SOME values $x \in \mathcal{X}$ (and truthful for all other values $x \in \mathcal{X}$).

Concretely, such relaxation comes down to three cases: a source could be

1. positively and negatively non-truthful for some values $x \in \mathcal{X}$ (and truthful for all other values $x \in \mathcal{X}$);
2. positively non-truthful and negatively truthful for some values $x \in \mathcal{X}$ (and truthful for all other values $x \in \mathcal{X}$);
3. positively truthful and negatively non-truthful for some values $x \in \mathcal{X}$ (and truthful for all other values $x \in \mathcal{X}$).

The first case was the object of the study carried out in Section 3.2, since, as already mentioned above, a source which is non-truthful for $x \in \mathcal{X}$ (Definition 1) is more precisely said to be, using Definitions 3 and 4, positively and negatively non-truthful for x. Cases 2 and 3 are treated in Sections 4.2 and 4.3, respectively.

4.2. Positive contextual liar

Let us turn our attention to case 2 above, i.e., the case where a source is assumed to be positively non-truthful and negatively truthful for some values $x \in \mathcal{X}$, and to be truthful for all other values $x \in \mathcal{X}$. Let $A \subseteq \mathcal{X}$ be the set of values for which the source is assumed to be truthful, and $\overline{A}$ the set of values for which it is assumed to be positively non-truthful and negatively truthful. For short, we may say that the source is truthful in A, and positively non-truthful and negatively truthful in $\overline{A}$, or even more simply, when no confusion is possible, that the source is positively non-truthful in $\overline{A}$ (or positively biased in $\overline{A}$), hence mentioning explicitly only the situation where the source commits a lie.

Definition 5 *(Positively non-truthful in $\overline{A}$).* A source is said to be *positively non-truthful in* $\overline{A}$ if it is truthful for all $x \in A$, and positively non-truthful and negatively truthful for all $x \in \overline{A}$. This state of the source is denoted by p_A.

This state corresponds to a source that lies (i.e., tells the contrary of what it knows) only for a subset of values and only when it tells for any of these values that it is a possibility for the actual value of **x**.

Table 2
Positively non-truthful in $\overline{A}$.

$x \in B$	$x \in A$	p_A
0	0	0
0	1	0
1	0	0
1	1	1

Let us note that this is yet again a more elaborate, thus more interesting, strategy for a source to deceive an agent than to simply tell the opposite of what it knows as in Pichon et al. [28], and thus state p_A may be useful when faced with intentionally deceitful sources (the potential usefulness of state p_A in the unintentional case will be commented later in this section). A source lying in this way may be referred to as a positive contextual liar in the sequel.

Proposition 2. *If a source tells* $\mathbf{x} \in B$ *and is assumed to be positively non-truthful in* $\overline{A}$*, one must deduce that* $\mathbf{x} \in B \cap A$.

Proof. The proof is similar to the proof of Proposition 1 and based on the fact that when the source is in state p_A, the four possible cases with respect to the membership of a given value $x \in \mathcal{X}$ to the sets B and A must be treated according to Table 2. □

Example 8. As an illustration of Proposition 2, assume for instance $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ and that the source tells $\mathbf{x} \in B = \{x_3, x_4\}$. Furthermore, assume the source is in state $p_{\{x_1,x_3\}}$, i.e., is positively non-truthful for x_2 and for x_4. Then, one should deduce that $\mathbf{x} \in \{x_3, x_4\} \cap \{x_1, x_3\} = \{x_3\}$.

In the unintentional case, similarly as state ℓ_A, state p_A can be explained by a difference between the source and the agent, on the relation they respectively believe holds between an auxiliary variable and the parameter of interest, as illustrated by Example 9.

Example 9 *(Example 4 continued).* Consider again the setting of Example 4, but this time assume the agent learns some time after receiving Jane's testimony that she has actually not met recently some of the suspects, and more specifically she is unaware that among these suspects that she has not met recently, those that have changed their beard situation are only those that have a beard.

Let *Conrad* and *Linus* be these suspects that she has not met recently. Among these two suspects, only *Conrad* has a beard, and thus he is the only one who has changed his beard situation since *Jane* last saw these two suspects, which means that both *Conrad* and *Linus* did not have a beard when *Jane* last saw them. We have then

$$\rho_{Ag}(\{beard\}) = \{Eloy, Conrad\}, \quad \rho_{Ag}(\{\neg beard\}) = \{Linus, Aeneas\},$$

and

$$\rho_{Jane}(\{beard\}) = \{Eloy\}, \quad \rho_{Jane}(\{\neg beard\}) = \{Conrad, Linus, Aeneas\}.$$

This means that through the piece of information "the murderer had a beard", that was equivalent to the agent to the testimony $\mathbf{x} \in B = \{Eloy, Conrad\}$, *Jane* mislead unintentionally the agent about what she actually knows of the guilt of the suspects. Precisely, due to her partially outdated knowledge on the beard situation of the suspects, *Jane* told what she knows for the suspects in $A = \{Eloy, Aeneas\}$, and told the opposite of what she knows for each of the suspects in $\overline{A}$ only when she told that he is possibly the murderer.

Hence, since she was truthful for the suspects in $A = \{Eloy, Aeneas\}$ and was positively non-truthful and negatively truthful for the suspects in $\overline{A}$, or in other words is in state $p_{\{Eloy,Aeneas\}}$, the agent should deduce from *Jane*'s piece of information $\mathbf{x} \in B$ that in fact

$$\mathbf{x} \in B \cap A = \{Eloy, Conrad\} \cap \{Eloy, Aeneas\} = \{Eloy\},$$

which is indeed what the source actually knows about the murderer.

4.3. Negative contextual liar

To provide a full picture, case 3 mentioned at the end of Section 4.1, is studied in this section, yet more briefly since it is quite similar to case 2.

Let us recall that case 3 corresponds to assuming that a source is positively truthful and negatively non-truthful for some values $x \in \mathcal{X}$, and is truthful for all other values $x \in \mathcal{X}$. Let $A \subseteq \mathcal{X}$ be the set of values for which the source is assumed to

Table 3
Negatively non-truthful in A.

$x \in B$	$x \in A$	n_A
0	0	0
0	1	1
1	0	1
1	1	1

be positively truthful and negatively non-truthful, and $\overline{A}$ the set of values for which it is assumed to be truthful.[5] For short, the source may be said to be negatively non-truthful in A (or negatively biased in A).

Definition 6 *(Negatively non-truthful in A).* A source is said to be *negatively non-truthful in A* if it is positively truthful and negatively non-truthful for all $x \in A$, and truthful for all $x \in \overline{A}$. This state is denoted by n_A.

This state corresponds to a source that lies only for a subset of values and only when it tells for any of these values that it is not a possibility for the actual value of $\mathbf{x}$. A source lying in this way may therefore be called a negative contextual liar.

Proposition 3. *If a source tells* $\mathbf{x} \in B$ *and is assumed to be negatively non-truthful in A, one must deduce that* $\mathbf{x} \in B \cup A$.

Proof. The proof is similar to the proof of Proposition 1 and based on the fact that when the source is in state n_A, the four possible cases with respect to the membership of a given value $x \in \mathcal{X}$ to the sets B and A must be treated according to Table 3. □

Example 10. As an illustration of Proposition 3, assume for instance $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ and that the source tells $\mathbf{x} \in B = \{x_3, x_4\}$. Furthermore, assume the source is in state $n_{\{x_1,x_3\}}$, i.e., is negatively non-truthful for x_1 and for x_3. Then, one should deduce that $\mathbf{x} \in \{x_3, x_4\} \cup \{x_1, x_3\} = \{x_1, x_3, x_4\}$.

For the same reason as state p_A, state n_A may be useful when faced with intentionally deceitful sources. A source may also happen to be a negative contextual liar unintentionally, as illustrated by Example 11.

Example 11 *(Example 4 continued).* Consider again the setting of Example 4, but this time assume the agent learns some time after receiving Jane's testimony that she has actually not met recently some of the suspects, and more specifically she is unaware that among these suspects that she has not met recently, those that have changed their beard situation are only those that do not have a beard.

Let *Conrad* and *Linus* be these suspects that she has not met recently. Among these two suspects, only *Linus* does not have a beard, and thus he is the only one who has changed his beard situation since *Jane* last saw these two suspects, which means that both *Conrad* and *Linus* had a beard when *Jane* last saw them. We have then

$$\rho_{Ag}(\{beard\}) = \{Eloy, Conrad\}, \quad \rho_{Ag}(\{\neg beard\}) = \{Linus, Aeneas\},$$

and

$$\rho_{Jane}(\{beard\}) = \{Eloy, Conrad, Linus\}, \quad \rho_{Jane}(\{\neg beard\}) = \{Aeneas\}.$$

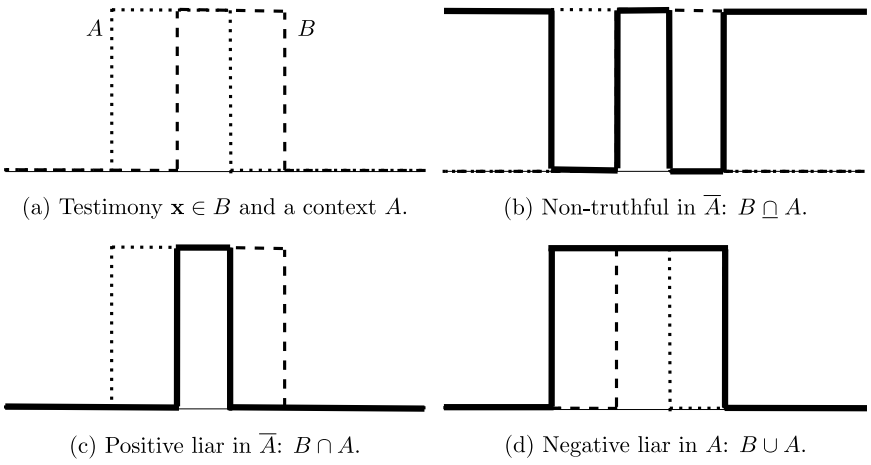
Jane told the opposite of what she knows for each of the suspects in $A = \{Conrad, Linus\}$ only when she told that he is not the murderer, and she told what she knows for the suspects in $\overline{A}$. In other words, she was positively truthful and negatively non-truthful for the suspects in A, and truthful for the suspects in $\overline{A}$. Since she was thus in state $n_{\{Conrad,Linus\}}$, the agent should deduce from *Jane*'s piece of information $\mathbf{x} \in B$ that in fact

$$\mathbf{x} \in B \cup A = \{Eloy, Conrad\} \cup \{Conrad, Linus\} = \{Eloy, Conrad, Linus\},$$

which is indeed what the source actually knows about the murderer.

The three kinds – contextual liar, positive contextual liar, and negative contextual liar – of liar studied so far are summarized and illustrated in Fig. 1. They constitute natural relaxations of the strong assumptions underlying the state of non-truthfulness of Pichon et al. [28]. They seem at least as interesting as this state when considering intentionally deceitful sources and may also be used to account for various ways a source may lack truthfulness unintentionally. Yet, the setting

[5] Contrary to the states ℓ_A and p_A, where A is the set of values for which the source is assumed to be truthful. This slight difference in denoting which set is the set where the source is truthful, is useful to present more elegantly results in the next sections, but it has no fundamental consequence.



(a) Testimony $\mathbf{x} \in B$ and a context A. (b) Non-truthful in $\overline{A}$: $B \underline{\cap} A$.

(c) Positive liar in $\overline{A}$: $B \cap A$. (d) Negative liar in A: $B \cup A$.

Fig. 1. The three kinds of contextual biases of interest in this paper. (a) Indicator functions of a testimony $\mathbf{x} \in B \subseteq \mathcal{X}$ and of a context $A \subseteq \mathcal{X}$. (b) In bold, result of the transformation of the testimony $\mathbf{x} \in B$ given contextual lie ℓ_A; in such state, the source is believed truthful in A, hence whatever it says about $\mathbf{x}$ within A should be kept as is, and non-truthful in $\overline{A}$, hence whatever it says about $\mathbf{x}$ outside of A should be reversed. (c) Contextual lie p_A (whenever the source says that a value outside of A is possible, it lies, and whatever it says within A should be kept as is). (d) Contextual lie n_A (whenever the source says that a value within A is not possible, it lies, and whatever it says outside of A should be kept as is).

considered is quite basic: the testimony provided by the source is crisp ($\mathbf{x} \in B$) and the state of the source is assumed to be known precisely. More generally, both the testimony provided by the source and the knowledge of the agent about the source truthfulness (referred to as *meta-knowledge* in [28]) may be uncertain. This is the next section topic, which in addition uses this more general setting to provide an interpretation for contextual discounting as well as an interpretation for contextual reinforcement.

5. Interpretation of CD and of CR

In this section, uncertainty is first added to the setting considered in Sections 3 and 4, resulting in a general framework able to handle various situations with respect to knowledge about the contextual biases of a source. Then, an interpretation for CR is proposed using this framework. In addition, it is shown that it is possible to provide a similar perspective on CD.

5.1. Uncertain testimony and meta-knowledge

Let $\mathcal{H}$ denote the possible states of a source S with respect to its polarized contextual truthfulness, i.e., $\mathcal{H} = \mathcal{H}_\ell \cup \mathcal{H}_p \cup \mathcal{H}_n$, where $\mathcal{H}_\ell = \{\ell_A | A \subseteq \mathcal{X}\}$, $\mathcal{H}_p = \{p_A | A \subseteq \mathcal{X}\}$ and $\mathcal{H}_n = \{n_A | A \subseteq \mathcal{X}\}$.

Following [28], we can define a multivalued mapping Γ_B from $\mathcal{H}$ to $\mathcal{X}$ that encodes the three kinds of contextual lies studied in Sections 3 and 4:

$$\Gamma_B(\ell_A) = B \underline{\cap} A, \tag{20}$$

$$\Gamma_B(p_A) = B \cap A, \tag{21}$$

$$\Gamma_B(n_A) = B \cup A, \tag{22}$$

for all $A \subseteq \mathcal{X}$. $\Gamma_B(h)$ indicates how to interpret the piece of information $\mathbf{x} \in B$ provided by the source, when the source is assumed to be in some state $h \in \mathcal{H}$. In addition, if the knowledge about the source state is imprecise and given by $H \subseteq \mathcal{H}$, then one should deduce that $\mathbf{x} \in \Gamma_B(H)$, where $\Gamma_B(H)$ denotes the image of H by Γ_B, defined by $\Gamma_B(H) := \bigcup_{h \in H} \Gamma_B(h)$.

Remark 3. We have

$$\Gamma_B(p_{\mathcal{X}}) = \Gamma_B(n_\emptyset) = \Gamma_B(\ell_{\mathcal{X}}) = B, \quad \forall B \subseteq \mathcal{X}.$$

This is so because these three states correspond actually to the same assumption of a truthful source. As such they may be simply denoted by t in the sequel (in accordance with the notation used for this state in the original paper [28]).

Remark 4. States ℓ_A, p_A, n_A, $A \subseteq \mathcal{X}$, and their associated transformations (20) (logical equality), (21) (conjunction), (22) (disjunction), of a testimony $\mathbf{x} \in B$, are particular cases of a more general formal model of truthfulness assumptions yielding all possible binary Boolean connectives, as shown in Appendix A.

As already mentioned, both the testimony of the source and the meta-knowledge of the agent may be uncertain. Let m_S be the uncertain testimony and $m^{\mathcal{H}}$ the uncertain meta-knowledge. In such case, the *Behavior-Based Correction* (BBC)

procedure[6] introduced by Pichon et al. [28], can be used to derive the knowledge of the agent on $\mathcal{X}$. It is represented by the MF m defined for all $C \subseteq \mathcal{X}$ as [28]:

$$m(C) = \sum_{H \subseteq \mathcal{H}} m^{\mathcal{H}}(H) \sum_{B:\Gamma_B(H)=C} m_S(B). \tag{23}$$

For convenience, we may denote by $f_{m^{\mathcal{H}}}(m_S)$ the Behavior-Based Correction of MF m_S according to meta-knowledge $m^{\mathcal{H}}$, i.e., we have $m = f_{m^{\mathcal{H}}}(m_S)$ with m the MF defined by (23). The BBC procedure is illustrated by Example 12.

Example 12. Let $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$. Assume a source S provides the following uncertain testimony:

$$m_S(\{x_1, x_3\}) = 0.7,$$
$$m_S(\{x_1, x_2, x_3\}) = 0.3.$$

Suppose further the following uncertain knowledge about the quality of the source:

$$m^{\mathcal{H}}\left(\left\{p_{\{x_3,x_4\}}\right\}\right) = 0.4,$$
$$m^{\mathcal{H}}\left(\left\{p_{\{x_3,x_4\}}, n_{\{x_1,x_4\}}\right\}\right) = 0.6,$$

that is, the source is assumed to be positively biased in $\{x_1, x_2\}$ with mass 0.4, and positively biased in $\{x_1, x_2\}$ or negatively biased in $\{x_1, x_4\}$ with mass 0.6.

Since

$$\begin{aligned}
\Gamma_{\{x_1,x_3\}}\left(p_{\{x_3,x_4\}}\right) &= \{x_1, x_3\} \cap \{x_3, x_4\} = \{x_3\}, \\
\Gamma_{\{x_1,x_3\}}\left(\left\{p_{\{x_3,x_4\}}, n_{\{x_1,x_4\}}\right\}\right) &= \Gamma_{\{x_1,x_3\}}\left(p_{\{x_3,x_4\}}\right) \bigcup \Gamma_{\{x_1,x_3\}}\left(n_{\{x_1,x_4\}}\right) \\
&= (\{x_3\}) \bigcup (\{x_1, x_3\} \cup \{x_1, x_4\}) = \{x_1, x_3, x_4\}, \\
\Gamma_{\{x_1,x_2,x_3\}}\left(p_{\{x_3,x_4\}}\right) &= \{x_1, x_2, x_3\} \cap \{x_3, x_4\} = \{x_3\}, \\
\Gamma_{\{x_1,x_2,x_3\}}\left(\left\{p_{\{x_3,x_4\}}, n_{\{x_1,x_4\}}\right\}\right) &= (\{x_3\}) \bigcup (\{x_1, x_2, x_3\} \cup \{x_1, x_4\}) = \mathcal{X},
\end{aligned}$$

the agent knowledge m on $\mathcal{X}$ is defined, according to the BBC procedure (23), by:

$$\begin{aligned}
m(\{x_3\}) &= m^{\mathcal{H}}\left(\left\{p_{\{x_3,x_4\}}\right\}\right) \cdot (m_S(\{x_1, x_3\}) + m_S(\{x_1, x_2, x_3\})) \\
&= 0.4 \cdot (0.7 + 0.3) = 0.4, \\
m(\{x_1, x_3, x_4\}) &= m^{\mathcal{H}}\left(\left\{p_{\{x_3,x_4\}}, n_{\{x_1,x_4\}}\right\}\right) \cdot m_S(\{x_1, x_3\}) \\
&= 0.6 \cdot 0.7 = 0.42, \\
m(\{x_1, x_2, x_3, x_4\}) &= m^{\mathcal{H}}\left(\left\{p_{\{x_3,x_4\}}, n_{\{x_1,x_4\}}\right\}\right) \cdot m_S(\{x_1, x_2, x_3\}) \\
&= 0.6 \cdot 0.3 = 0.18.
\end{aligned}$$

The next section will show how BBC may be used to provide an interpretation for CR.

5.2. Contextual reinforcement

Let us consider a particular kind of contextual lie among those introduced in Sections 3 and 4: the states p_A, $A \subseteq \mathcal{X}$, corresponding to the assumptions that the source is a positive liar in $\overline{A}$.

The next proposition, which is based on theses states and that uses the notation introduced in Remark 3, will be instrumental to provide our interpretation of CR.

Proposition 4. *Let m_S be the MF provided by a source S and let $m^{\mathcal{H}}_{A,\cap}$ be our meta-knowledge on the source defined by*

$$m^{\mathcal{H}}_{A,\cap}(\{t\}) = \beta_A, \quad m^{\mathcal{H}}_{A,\cap}(\{p_A\}) = 1 - \beta_A, \tag{24}$$

i.e., with mass β_A the source is truthful, and with mass $1 - \beta_A$ it is a positive liar in $\overline{A}$.

[6] The BBC procedure is a general mechanism allowing one to derive an agent's knowledge on $\mathcal{X}$ from an uncertain testimony m_S, when the agent has some uncertain meta-knowledge $m^{\mathcal{H}}$ about the source, and where $\mathcal{H}$ may represent various state spaces, not necessarily related to the notion of truthfulness. See [28] for details.

We have

$$f_{m^{\mathcal{H}}_{A,\cap}}(m_S) = m_S \textcircled{\cap} A^{\beta_A}.$$

Proof. From the definition of $\textcircled{\cap}$, we have that, for all $B \subseteq \mathcal{X}$, the quantity $m_S(B) \cdot (A^{\beta_A})(\mathcal{X}) = m_S(B) \cdot \beta_A$ is allocated to set $B \cap \mathcal{X} = B$ and the quantity $m_S(B) \cdot A^{\beta_A}(A) = m_S(B) \cdot (1 - \beta_A)$ is allocated to set $B \cap A$.

Similarly, from the definition of the BBC procedure, the quantity $m_S(B) \cdot \beta_A$ is allocated to set $\Gamma_B(t) = B$ and the quantity $m_S(B) \cdot (1 - \beta_A)$ is allocated to set $\Gamma_B(p_A) = B \cap A$, for all $B \subseteq \mathcal{X}$. □

We may then show one of our main results.

Proposition 5. *Let m_S be a MF. We have, $\forall \mathcal{A}$ and with $\beta_A \in [0, 1]$, $\forall A \in \mathcal{A}$:*

$$m_S \textcircled{\cap}_{A \in \mathcal{A}} A^{\beta_A} = (\circ_{A \in \mathcal{A}} f_{m^{\mathcal{H}}_{A,\cap}})(m_S), \tag{25}$$

where $\circ$ denotes function composition and where mass functions $m^{\mathcal{H}}_{A,\cap}$, $A \in \mathcal{A}$, are defined by (24).

Proof. Without lack of generality, let us index the elements in $\mathcal{A}$: $A_1, \ldots, A_N$, where $N = |\mathcal{A}|$. Thus, Equation (25) can be rewritten as

$$m_S \textcircled{\cap}_{i=1}^{N} A_i^{\beta_{A_i}} = (\circ_{i=1}^{N} f_{m^{\mathcal{H}}_{A_i,\cap}})(m_S), \tag{26}$$

From Proposition 4, we have: $m_S \textcircled{\cap} A_1^{\beta_{A_1}} = f_{m^{\mathcal{H}}_{A_1,\cap}}(m_S)$. Hence (26) holds for $N = 1$.

Assume now that (26) holds for $N = Q$. To prove this proposition, it suffices then to show that (26) holds for $N = Q + 1$. We have:

$$\begin{aligned} m_S \textcircled{\cap}_{i=1}^{Q+1} A_i^{\beta_{A_i}} &= m_S \textcircled{\cap}_{i=1}^{Q} A_i^{\beta_{A_i}} \textcircled{\cap} A_{Q+1}^{\beta_{A_{Q+1}}} \\ &= (\circ_{i=1}^{Q} f_{m^{\mathcal{H}}_{A_i,\cap}})(m_S) \textcircled{\cap} A_{Q+1}^{\beta_{A_{Q+1}}}. \end{aligned}$$

From Proposition 4, we obtain:

$$\begin{aligned} (\circ_{i=1}^{Q} f_{m^{\mathcal{H}}_{A_i,\cap}})(m_S) \textcircled{\cap} A_{Q+1}^{\beta_{A_{Q+1}}} &= f_{m^{\mathcal{H}}_{A_{Q+1},\cap}}((\circ_{i=1}^{Q} f_{m^{\mathcal{H}}_{A_i,\cap}})(m_S)) \\ &= (\circ_{i=1}^{Q+1} f_{m^{\mathcal{H}}_{A_i,\cap}})(m_S). \qquad \square \end{aligned}$$

Proposition 5 is important in that it constitutes the first known interpretation for CR. It shows that CR, which appears on the left side of (25), corresponds to independent behavior-based corrections – one for each context $A \in \mathcal{A}$ – where for each context A, the source is assumed to be truthful with mass β_A, and to be a positive liar in $\overline{A}$ with mass $1 - \beta_A$.

Proposition 5 is illustrated by Examples 13 and 14, which show that CR may be encountered when dealing with intentionally and unintentionally lying sources, respectively.

Example 13. Let m_S be an uncertain testimony provided by a source S on $\mathcal{X} = \{x_1, x_2, x_3\}$. An agent believes that S lies intentionally, and more specifically that it positively lies for x_3 with mass 0.4, and, independently, positively lies for x_1 with mass 0.2.

In other words, S is subject to independent contextual lies of the form "positively non-truthful in $\overline{A}$" for contexts $\mathcal{A} = \{A_1, A_2\}$ with $A_1 = \{x_1, x_2\}$ and $A_2 = \{x_2, x_3\}$, and with masses $1 - \beta_{A_1} = 0.4$ and $1 - \beta_{A_2} = 0.2$.

From Proposition 5, the agent's belief m on $\mathcal{X}$ is then defined by

$$m = m_S \textcircled{\cap} \{x_1, x_2\}^{0.6} \textcircled{\cap} \{x_2, x_3\}^{0.8}.$$

Example 14 *(Example 9 continued).* Consider again the setting of Example 4, but this time assume the agent learns some time after receiving Jane's testimony that she may actually not have met recently some of the suspects, and more specifically she is unaware that among these suspects that she may not have met recently, those that have changed their beard situation are only those that have a beard. Besides, he is unsure of who are these suspects: it could be *Conrad* and *Linus* with mass 0.4, or independently, with mass 0.2, *Linus* and *Eloy*.

From Proposition 5, the agent's belief m on who is the murderer is then obtained by

$$m = \{Eloy, Conrad\}^0 \textcircled{\cap} \{Eloy, Aeneas\}^{0.6} \textcircled{\cap} \{Conrad, Aeneas\}^{0.8},$$

where $\{Eloy, Conrad\}^0$ is the MF representing testimony $\mathbf{x} \in B = \{Eloy, Conrad\}$.

5.3. Contextual discounting

We show in this section that it is possible to obtain a similar perspective on CD as the interpretation proposed for CR in Section 5.2.

Let us consider another kind of contextual lie in this section: the states n_A, $A \subseteq \mathcal{X}$, corresponding to the assumptions that the source is a negative liar in A.

Proposition 6. *Let m_S be the MF provided by a source S and let $m^{\mathcal{H}}_{A,\cup}$ be our meta-knowledge on the source defined by*

$$m^{\mathcal{H}}_{A,\cup}(\{t\}) = \beta_A, \quad m^{\mathcal{H}}_{A,\cup}(\{n_A\}) = 1 - \beta_A, \tag{27}$$

i.e., with mass β_A the source is truthful, and with mass $1 - \beta_A$ it is a negative liar in A.

We have

$$f_{m^{\mathcal{H}}_{A,\cup}}(m_S) = m_S \textcircled{\cup} A_{\beta_A}.$$

Proof. The proof is similar to that of Proposition 4. □

Proposition 7. *Let m_S be a MF. We have, $\forall \mathcal{A}$ and with $\beta_A \in [0, 1]$, $\forall A \in \mathcal{A}$:*

$$m_S \textcircled{\cup}_{A \in \mathcal{A}} A_{\beta_A} = (\circ_{A \in \mathcal{A}} f_{m^{\mathcal{H}}_{A,\cup}})(m_S), \tag{28}$$

where mass functions $m^{\mathcal{H}}_{A,\cup}$, $A \in \mathcal{A}$, are defined by (27).

Proof. The proof is similar to that of Proposition 5, using Proposition 6 instead of Proposition 4. □

Proposition 7 shows that, similarly to CR, CD (left side of (28)) amounts to independent BBCs – one for each context – corresponding to simple contextual biases. The only difference between the two correction mechanisms is what is assumed with mass $1 - \beta_A$: with the former that the source is a positive liar in $\overline{A}$, whereas with the latter that the source is a negative liar in A. This latter finding concerning the difference between CR and CD suggests that CR seems as interesting as CD to handle contextual knowledge about the quality of a source, since the truthfulness assumptions associated to CR seem as useful in practice as those associated to CD.

Remark 5. This interpretation of CD implies that classical discounting (11) amounts simply to assuming that the source is truthful with mass β, and negatively non-truthful in $\mathcal{X}$ with mass $1 - \beta$, or for short, truthful with mass β and *negatively* non-truthful with mass $1 - \beta$. Hence, discounting (11) may be seen as a relaxation of negating (16), since this latter correction amounts to assuming that the source is truthful with mass β, and (positively *and* negatively) non-truthful with mass $1 - \beta$.

Remark 6. This interpretation of CD provides a new perspective on contextual discounting based on a coarsening, which is a particular case of CD. Moreover, it brings a new element to the discussion entertained in [20, Section 5.2], where Mercier et al. distinguish two kinds of contextual knowledge about the quality of a source: one may have some knowledge on the quality of the source given that the true value of the parameter of interest $\mathbf{x}$ is in some set $A \subseteq \mathcal{X}$, which is the kind of contextual knowledge used in the original derivation of contextual discounting based on a coarsening; and one may have some knowledge on the quality of the source with respect to what the source declares about $\mathbf{x}$, which is the kind of contextual knowledge at play when one considers the interpretation of CD uncovered in this section.

Remark 7. Each of CD and CR can also be viewed as a single BBC corresponding to some particular knowledge on the truthfulness of the information source, which is basically the one obtained by combining together, that is for all $A \in \mathcal{A}$, the simple pieces of meta-knowledge $m^{\mathcal{H}}_{A,\cup}$ (27) and $m^{\mathcal{H}}_{A,\cap}$ (24), respectively, as shown in Appendix B.

This section has provided an interpretation for CR as well as an interpretation for CD using our proposed refined model of source truthfulness. From a formal point of view, these new results were also obtained from the very definitions of these two mechanisms: both of these contextual corrections amount to the combination with some separable MF (conjunctive combination with a $\textcircled{\cap}$-separable MF in the case of CR, and disjunctive combination with a $\textcircled{\cup}$-separable MF in the case of CD), and the simple MFs composing those separable MFs are directly related to the assumptions on the truthfulness of the source made by CR and CD as shown by Propositions 4 and 6. In the next section, we study the possibility of extending CD and CR by going beyond the combination with separable MFs.

6. Canonical decompositions and contextual corrections

In this section, a more general form of CD is derived by exploiting the canonical decomposition and by contextualizing the de-discounting operation. As will be seen, this generalization may be useful in that it allows one to take into account even more situations with respect to knowledge about the quality of a source. First, the de-discounting operation is contextualized. Then, contextual de-discounting is used to extend CD. Similar results are also obtained for CR.

6.1. Contextual de-corrections

De-discounting (12) is the inverse of discounting (11), which can also be expressed as (14). As recalled in Section 2.2, de-discounting is useful to remove a discounting that is considered no longer valid. To remove a discounting, one merely needs to use the disjunctive decombination rule. Indeed, we have

$$m \textcircled{\cup} \mathcal{X}_\beta \textcircled{\emptyset} \mathcal{X}_\beta = m.$$

In other words, de-discounting defined by (12) admits a simple expression as:

$$m = m_S \textcircled{\emptyset} \mathcal{X}_\beta. \tag{29}$$

Now, much as discounting is a particular case of CD, it is natural to view de-discounting as a particular case of the following operation that may be called *contextual de-discounting* (CdD).

Definition 7 *(Contextual de-discounting).* Let m_S be a MF. Its correction using contextual de-discounting, given a set $\mathcal{A}$ of contexts with associated parameters $\beta_A \in (0, 1]$, for all $A \in \mathcal{A}$, is defined as the following MF m:

$$m = m_S \textcircled{\emptyset}_{A \in \mathcal{A}} A_{\beta_A}. \tag{30}$$

The interpretation of CdD is similar to that of de-discounting and of CD: it amounts to the removals of $|\mathcal{A}|$ independent BBCs, where for each context A the source was assumed to be truthful with mass β_A and a negative liar in A with mass $1 - \beta_A$. Example 15 illustrates CdD.

Example 15. Let m_S be an uncertain testimony provided by a source S on $\mathcal{X} = \{Peter, John, Mary\}$. An agent Ag believes that S lies intentionally, and more specifically that it negatively lies for *Peter* and *John* with mass 0.4.

In other words, Ag assumes that S is subject to a contextual lie of the form "negatively non-truthful in A" for context $\{Peter, John\}$ with mass $1 - \beta_{\{Peter,John\}} = 0.4$. From Proposition 7, the agent's belief m_{Ag} on $\mathcal{X}$ is then defined by:

$$m_{Ag} = m_S \textcircled{\cup} \{Peter, John\}_{0.6}.$$

Suppose that an agent Ag_2 receives from Ag the piece of information m_{Ag}. Suppose further that Ag_2 does not know what S told to Ag, i.e., he does not know m_S, but he knows Ag's meta-knowledge on the source, and he thinks that it is wrong, since he believes that the source tells a negative lie for *Peter* and *John* with mass 0.2.

This amounts to applying to m_{Ag} a contextual de-discounting given context $\{Peter, John\}$ with parameter $\beta_{\{Peter,John\}} = 0.6$ (in order to remove the correction performed by Ag on the testimony of the source), and then a contextual discounting given context $\{Peter, John\}$ with parameter $\beta_{\{Peter,John\}} = 0.8$, i.e.,

$$m_{Ag_2} = (m_{Ag} \textcircled{\emptyset} \{Peter, John\}_{0.6}) \textcircled{\cup} \{Peter, John\}_{0.8},$$

or, equivalently, applying to m_{Ag} a contextual de-discounting given context $\{Peter, John\}$ with parameter $\beta_{\{Peter,John\}} = 0.75$, i.e.,

$$m_{Ag_2} = m_{Ag} \textcircled{\emptyset} \{Peter, John\}_{0.75},$$

since for any two functions[7] $A_{v_1} : 2^{\mathcal{X}} \to \mathbb{R}$ and $A_{v_2} : 2^{\mathcal{X}} \to \mathbb{R}$ defined for $i = 1, 2$ by:

$$\begin{aligned} A_{v_i} : A &\mapsto 1 - v_i, \\ \emptyset &\mapsto v_i, \\ B &\mapsto 0 \ \ \forall B \in 2^{\mathcal{X}} \setminus \{A, \emptyset\}, \end{aligned} \tag{31}$$

for some $A \supset \emptyset$ and some $v_i \in (0, +\infty)$, we have [6]:

[7] Such function is called a negative generalized MF in [6]. It is used here only as a formal and useful tool to simplify the presentation.

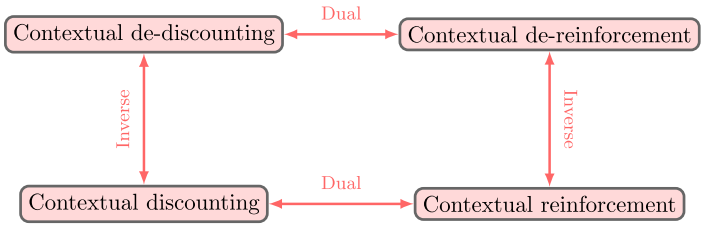


Fig. 2. Relationships between the four contextual correction mechanisms.

$$A_{v_1} \textcircled{\cup} A_{v_2} = A_{v_1 \cdot v_2}, \tag{32}$$

$$A_{v_1} \textcircled{\cap} A_{v_2} = A_{v_1 / v_2}. \tag{33}$$

Let us now consider CR. CR for $\mathcal{A} = \{\emptyset\}$ amounts to $m = m_S \textcircled{\cap} \emptyset^{\beta}$, i.e., a process that redistributes a portion $1 - \beta$ of the masses given to the non-empty sets, to the empty set. In other words, CR for $\mathcal{A} = \{\emptyset\}$ is the dual of discounting and may thus simply be called *reinforcement* hereafter. Its inverse is defined by $m = m_S \textcircled{\cap} \emptyset^{\beta}$, and may be called *de-reinforcement*.

Remark 8. Normalization (1) may be expressed as [6]: $m^* = m \textcircled{\cap} \emptyset^{1-m(\emptyset)}$ and corresponds thus to the de-reinforcement of MF m with degree $\beta = 1 - m(\emptyset)$. Its dual is called *maximal* de-discounting [7], and corresponds to setting $\beta = 1 - m_S(\mathcal{X})$ in (29).

Similarly as for de-discounting and CdD, de-reinforcement is a particular case of the following operation that may be called contextual de-reinforcement (CdR):

Definition 8 *(Contextual de-reinforcement).* Let m_S be a MF. Its correction using contextual de-reinforcement, given a set $\mathcal{A}$ of contexts with associated parameters $\beta_A \in (0, 1]$, for all $A \in \mathcal{A}$, is defined as the following MF m:

$$m = m_S \textcircled{\cap}_{A \in \mathcal{A}} A^{\beta_A}. \tag{34}$$

The interpretation of CdR is similar to that of CdD, and CdR may be illustrated using a similar example as Example 15.

Remark 9. As for other computations involving decombination rules $\textcircled{\cap}$ and $\textcircled{\cup}$, we note that CdD (30) and CdR (34) may not always yield a belief function, hence they should be used with care. For instance, as recalled in Section 2.2, for de-discounting (29) to yield a belief function, it is necessary that $1 - m_S(\mathcal{X}) \leq \beta \leq 1$.

Fig. 2 synthesizes the relations between CD, CR, CdD and CdR.

6.2. Contextual corrections based on the canonical decompositions

Let us now consider a correction of a MF m_S resulting in a MF m, and involving both a CD (17) and a CdD (30), with associated sets of contexts $\mathcal{A}^c$ and $\mathcal{A}^d$, respectively, and associated degrees of belief $\beta_A^c \in (0, 1)$, $A \in \mathcal{A}^c$, and $\beta_A^d \in (0, 1)$, $A \in \mathcal{A}^d$, respectively, i.e., the operation

$$m = m_S \textcircled{\cup}_{A \in \mathcal{A}^c} A_{\beta_A^c} \textcircled{\cup}_{A \in \mathcal{A}^d} A_{\beta_A^d}, \tag{35}$$

such that $\mathcal{A}^c \cap \mathcal{A}^d = \emptyset$ and such that these CD and CdD together form a disjunctive canonical decomposition, i.e., the function $m_\cup$ defined by

$$m_\cup = \textcircled{\cup}_{A \in \mathcal{A}^c} A_{\beta_A^c} \textcircled{\cup}_{A \in \mathcal{A}^d} A_{\beta_A^d},$$

is a (non-normal) mass function.

Equation (35) defines clearly an extension of CD, which is recovered if $\mathcal{A}^d = \emptyset$. It will be referred to as *contextual discounting based on the canonical decomposition* in the remainder of this paper and abbreviated by CD+.

Definition 9 *(Contextual discounting based on the canonical decomposition).* Let m_S be a MF. Its correction using contextual discounting based on the canonical decomposition, given sets of contexts $\mathcal{A}^c$ and $\mathcal{A}^d$ and associated degrees $\beta_A^c \in (0, 1)$,

$A \in \mathcal{A}^c$, and $\beta_A^d \in (0,1)$, $A \in \mathcal{A}^d$, such that $\mathcal{A}^c \cap \mathcal{A}^d = \emptyset$[8] and such that

$$m_{\cup} = \textcircled{\cup}_{A \in \mathcal{A}^c} A_{\beta_A^c} \textcircled{\emptyset}_{A \in \mathcal{A}^d} A_{\beta_A^d},$$

is a non-normal mass function, is defined as the following MF m:

$$m = m_S \textcircled{\cup} m_{\cup}.$$

Let us note that, by definition, a given CD based on the canonical decomposition can be applied to any MF m_S since it amounts simply to the disjunctive combination of m_S with another (non-normal) MF, contrary to a given contextual de-discounting, which validity is dependent on the MF m_S to be corrected (cf. Remark 9).

CD+ is a correction that is relevant if an uncertain testimony must be both discounted for some contexts and de-discounted for some other contexts, which may happen when one believes that this testimony is the result of an initial piece of information that has not been properly contextually discounted as illustrated by Example 16.

Example 16 *(Example 15 continued).* Let us consider again the setting of Example 15, but this time assumes that Ag_2 thinks Ag is wrong about the behavior of S, and in particular that the source tells actually a negative lie for *Peter* with mass 0.6, and independently a negative lie for *John* with mass 0.7.

This amounts to applying to m_{Ag} a contextual de-discounting given context $\{Peter, John\}$ with parameter $\beta_{\{John\}} = 0.6$ (in order to remove the correction performed by Ag on the testimony of the source), and then a contextual discounting given contexts $\mathcal{A} = \{\{Peter\}, \{John\}\}$ with associated parameters $\beta_{\{Peter\}} = 0.4$ and $\beta_{\{John\}} = 0.3$, i.e.,

$$m_{Ag_2} = (m_{Ag} \textcircled{\emptyset} \{Peter, John\}_{0.6}) \textcircled{\cup} \{Peter\}_{0.4} \textcircled{\cup} \{John\}_{0.3},$$

or, equivalently, combining disjunctively m_{Ag} with a non-normal MF $m_{\cup}$:

$$m_{Ag_2} = m_{Ag} \textcircled{\cup} m_{\cup},$$

with $m_{\cup}$ defined by

$$m_{\cup} = (\{Peter\}_{0.4} \textcircled{\cup} \{John\}_{0.3}) \textcircled{\emptyset} (\{Peter, John\}_{0.6}).$$

Remark 10. From a formal point of view, CD+ (35) can be equivalently presented as, using (32) and (33):

$$m = m_S \textcircled{\cup}_{A \in \mathcal{A}} A_{\beta_A}, \tag{36}$$

with $\mathcal{A} = \mathcal{A}^c \cup \mathcal{A}^d$ and with $\beta_A = \beta_A^c$ if $A \in \mathcal{A}^c$ and $\beta_A = 1/\beta_A^d$ if $A \in \mathcal{A}^d$, for all $A \in \mathcal{A}$ (hence $\beta_A \in (0, +\infty)$). In other words, CD+ has the same definition as CD (17), except that we may have $\beta_A > 1$, for some $A \in \mathcal{A}$. This technical remark will be useful later for a technical result (Remark 12 in Section 8).

Of course, a similar reasoning can be followed to introduce the dual notion of *contextual reinforcement based on the canonical decomposition* denoted by CR+: formally, it is simply the combination by $\textcircled{\cap}$ of MF m_S with a non-dogmatic MF $m_{\cap}$.

Definition 10 *(Contextual reinforcement based on the canonical decomposition).* Let m_S be a MF. Its correction using contextual reinforcement based on the canonical decomposition, given sets of contexts $\mathcal{A}^c$ and $\mathcal{A}^d$ and associated degrees $\beta_A^c \in (0,1)$, $A \in \mathcal{A}^c$, and $\beta_A^d \in (0,1)$, $A \in \mathcal{A}^d$, such that $\mathcal{A}^c \cap \mathcal{A}^d = \emptyset$ and such that

$$m_{\cap} = \textcircled{\cap}_{A \in \mathcal{A}^c} A^{\beta_A^c} \textcircled{\emptyset}_{A \in \mathcal{A}^d} A^{\beta_A^d},$$

is a non-dogmatic mass function, is defined as the following MF m:

$$m = m_S \textcircled{\cap} m_{\cap}.$$

[8] This condition is not strictly needed. It is imposed in this definition, so that each couple $(\mathcal{A}^c, \mathcal{A}^d)$ with associated degrees $\beta_A^c, A \in \mathcal{A}^c$, and $\beta_A^d, A \in \mathcal{A}^d$, uniquely defines a CD+ correction of a MF m_S. Indeed, for each couple $(\mathcal{A}^c, \mathcal{A}^d)$ such that $\mathcal{A}^c \cap \mathcal{A}^d = \emptyset$, there exists an infinity of couples $(\mathcal{A}^{c'}, \mathcal{A}^{d'})$ with associated degrees $\beta_A^{c'} \in (0,1), A \in \mathcal{A}^{c'}$, and $\beta_A^{d'} \in (0,1), A \in \mathcal{A}^{d'}$, such that $\mathcal{A}^{c'} \cap \mathcal{A}^{d'} \neq \emptyset$, satisfying $\textcircled{\cup}_{A \in \mathcal{A}^c} A_{\beta_A^c} \textcircled{\emptyset}_{A \in \mathcal{A}^d} A_{\beta_A^d} = \textcircled{\cup}_{A \in \mathcal{A}^{c'}} A_{\beta_A^{c'}} \textcircled{\emptyset}_{A \in \mathcal{A}^{d'}} A_{\beta_A^{d'}}$, which is a consequence of (33). Besides, for any two couples $(\mathcal{A}^c, \mathcal{A}^d)$ and $(\mathcal{A}^{c'}, \mathcal{A}^{d'})$ such that $\mathcal{A}^c \cap \mathcal{A}^d = \emptyset$ and $\mathcal{A}^{c'} \cap \mathcal{A}^{d'} = \emptyset$, we have $\textcircled{\cup}_{A \in \mathcal{A}^c} A_{\beta_A^c} \textcircled{\emptyset}_{A \in \mathcal{A}^d} A_{\beta_A^d} = \textcircled{\cup}_{A \in \mathcal{A}^{c'}} A_{\beta_A^{c'}} \textcircled{\emptyset}_{A \in \mathcal{A}^{d'}} A_{\beta_A^{d'}}$ iff $\mathcal{A}^c = \mathcal{A}^{c'}$, $\mathcal{A}^d = \mathcal{A}^{d'}$, $\beta_A^c = \beta_A^{c'}$ for each $A \in \mathcal{A}^c$ and $\beta_A^d = \beta_A^{d'}$ for each $A \in \mathcal{A}^d$, which follows from the uniqueness of the disjunctive canonical decomposition.

CR+ extends CR since CR amounts to the combination with a $\textcircled{\cap}$-separable MF, that is, CR is a CR+ such that $\mathcal{A}^d = \emptyset$. CR+ is a correction that is sensible if MF m_S provided by the source must be both reinforced for some contexts $\mathcal{A}^c$ and de-reinforced for some other contexts $\mathcal{A}^d$, and may be illustrated similarly as CD+ (cf. Example 16 for the illustration of CD+).

Remark 11. A counterpart to Remark 10 exists for CR+. Indeed, by extending the notation A^w to the case where $w \in (0, +\infty)$, similarly as notation A_v is extended in (31) to $v \in (0, +\infty)$, and using the $\textcircled{\cap}$-counterparts of (32) and (33) shown in [34], CR+ can be equivalently presented as:

$$m = m_S \textcircled{\cap}_{A \in \mathcal{A}} A^{\beta_A}, \tag{37}$$

with $\mathcal{A} = \mathcal{A}^c \cup \mathcal{A}^d$ and with $\beta_A = \beta_A^c$ if $A \in \mathcal{A}^c$ and $\beta_A = 1/\beta_A^d$ if $A \in \mathcal{A}^d$, for all $A \in \mathcal{A}$ (hence $\beta_A \in (0, +\infty)$). CR+ has then the same definition as CR, except that we may have $\beta_A > 1$, for some $A \in \mathcal{A}$. This will be useful later for a technical result (Remark 13 in Section 8).

This section has introduced extensions of CD and CR, by exploiting the fact that the assumptions made by a given CD or CR, can be readily seen through the definition of this CD or CR: the assumptions correspond to the simple MFs in said definition. Hence, thanks to the inverse rules $\textcircled{\cup}$ and $\textcircled{\cap}$, it becomes possible to remove such assumptions if one believes that they are no longer tenable, which amounts to a so-called contextual de-correction. In the next section, yet another way of extending CD and CR is studied: it is based on exploiting the state ℓ_A introduced in Section 3.

7. Contextual negating

In this section, we introduce a new contextual correction scheme, which is formally similar to the two existing ones and that is related to the negating operation.

7.1. Non-truthful in $\overline{A}$

As shown in Section 5, CD and CR result from corrections induced by simple pieces of meta-knowledge $m^{\mathcal{H}}_{A,\cup}$ (27) and $m^{\mathcal{H}}_{A,\cap}$ (24) relying on contextual biases n_A and p_A, respectively. In practice, these two states transform a testimony $\mathbf{x} \in B$ into $B \cup A$ and $B \cap A$, respectively.

In Section 3, a third natural contextual lie, state ℓ_A, which corresponds to assuming that the source is non-truthful in $\overline{A}$, was brought to light and studied. This state yields $\mathbf{x} \in B \underline{\cap} A$ from a testimony $\mathbf{x} \in B$. Interestingly, the properties satisfied by $\underline{\cap}$ (associativity, commutativity, neutral element) allow us to obtain similar propositions as those obtained for CR and CD (the proofs of those propositions are similar to the ones of CR and CD, and are thus omitted).

Proposition 8. *Let m_S be the MF provided by a source S and let $m^{\mathcal{H}}_{A,\underline{\cap}}$ be our meta-knowledge on the source defined by*

$$m^{\mathcal{H}}_{A,\underline{\cap}}(\{t\}) = \beta_A, \quad m^{\mathcal{H}}_{A,\underline{\cap}}(\{\ell_A\}) = 1 - \beta_A, \tag{38}$$

i.e., with mass β_A the source is truthful, and with mass $1 - \beta_A$ it is non-truthful in $\overline{A}$.

We have

$$f_{m^{\mathcal{H}}_{A,\underline{\cap}}}(m_S) = m_S \underline{\textcircled{\cap}} A^{\beta_A}.$$

Proposition 9. *Let m_S be a MF. We have, $\forall \mathcal{A}$ and with $\beta_A \in [0, 1]$, $\forall A \in \mathcal{A}$:*

$$m_S \underline{\textcircled{\cap}}_{A \in \mathcal{A}} A^{\beta_A} = (\circ_{A \in \mathcal{A}} f_{m^{\mathcal{H}}_{A,\underline{\cap}}})(m_S), \tag{39}$$

where mass functions $m^{\mathcal{H}}_{A,\underline{\cap}}$, $A \in \mathcal{A}$, are defined by (38).

Equation (39) is the $\underline{\cap}$ counterpart to Equations (28) and (25), which are based on $\cup$ and $\cap$, respectively. It constitutes a contextual correction, which, similarly to CD and CR, amounts to independent BBCs – one for each context – corresponding to simple contextual biases. The only difference with CD and CR is what is assumed with mass $1 - \beta_A$: that the source is non-truthful in $\overline{A}$.

An interesting fact can be brought to light about this contextual correction.

Proposition 10. *For any MF m_S, we have*

$$m_S \underline{\textcircled{\cap}} \emptyset^{\beta} = \beta \cdot m_S + (1 - \beta) \cdot \overline{m}_S. \tag{40}$$

Proof. This proposition follows from Proposition 8 and Remark 1. □

This proposition shows that negating (16) is a particular case of the contextual correction (39): it is recovered for $\mathcal{A} = \{\emptyset\}$, as should be since contextual correction (39) reduced to $\mathcal{A} = \{\emptyset\}$ corresponds to assuming that the source is truthful with mass β and non-truthful with mass $1 - \beta$, which is the meta-knowledge associated to the negating operation. Contextual correction (39) constitutes thus a similar extension with respect to negating, than CD is to discounting and CR is to reinforcement. It may therefore be seen as a contextual version of negating and be called *contextual negating* (CN).

Definition 11 *(Contextual negating).* Let m_S be a MF. Its correction using contextual negating, given a set $\mathcal{A}$ of contexts with associated parameters $\beta_A \in [0, 1]$, for all $A \in \mathcal{A}$, is defined as the following MF m:

$$m = m_S \textcircled{\cap}_{A \in \mathcal{A}} A^{\beta_A}.$$

We note that the computational complexity of CN is similar to that of CD and CR: it merely corresponds to the complexity of applying $|\mathcal{A}|$ combinations by the $\textcircled{\cap}$ rule.

Furthermore, similar examples as Examples 13 and 14 can easily be constructed to illustrate situations where CN may be needed.

7.2. Discussion

Let us briefly emphasize the similarities and differences between CD and the new contextual correction mechanism uncovered above that is CN.

In their non-contextual version, they reduce to discounting and negating, respectively. Discounting is the correction mechanism derived from basic assumptions about the relevance of the source, whereas negating originates from basic assumptions about the truthfulness of the source. Alternatively, as pointed out by Remark 5, discounting can also be viewed as a relaxed form of negating, in that it may also be recovered from assumptions about the truthfulness of the source that are less strong than those yielding negating. These two mechanisms differ partially in how they treat a testimony $\mathbf{x} \in B$: they both keep mass β on B, and discounting transfers mass $1 - \beta$ to $\mathcal{X}$ whereas negating transfers it to $\overline{B}$; this comes from the fact that we deduce either $\mathbf{x} \in \mathcal{X}$ or $\mathbf{x} \in \overline{B}$ depending on whether we think the source is non-relevant (or negatively non-truthful, see Remark 5) or non-truthful. As extensions of discounting and negating, CD and CN are thus clearly fundamentally different operations.

Furthermore, in essence, given a context A and testimony $\mathbf{x} \in B$, both CD and CN keep mass β_A on B, and CD transfers mass $1 - \beta_A$ to $B \cup A$, whereas CN transfers it to $B \underline{\cap} A$, so that in practice the difference between the two mechanisms is what happens with mass $1 - \beta_A$: with CD, it is assumed that at least one of the pieces of information $\mathbf{x} \in B$ and $\mathbf{x} \in A$ is true, whereas with CN, we have that either both or none of these pieces of information are true.

Let us finally remark that, similarly as inverses of $\textcircled{\cap}$ and $\textcircled{\cup}$ can be defined from pointwise divisions of commonality and implicability functions, respectively, it is in principle possible to define the inverse of the rule $\textcircled{\underline{\cap}}$ from pointwise division of the so-called 0-commonality function [26,24], which is the counterpart of the commonality and implicability functions for the rule $\textcircled{\underline{\cap}}$. Hence, in principle, it is possible to define a notion of contextual de-negating as well as a more general form of CN similar to the more general forms of CD and CR obtained in Section 6. However, for the definition of the inverse of the rule $\textcircled{\underline{\cap}}$ to be usable in practice, one needs to know under which simple conditions the 0-commonality function does not equal to 0, since divisions by zeros must be avoided, similarly as it is known that the commonality and implicability functions are different from 0 as long as the MF is non-dogmatic or non-normal, respectively. Unfortunately, we have not been able so far to find such simples conditions for the 0-commonality function. This is left for further research.

This section has introduced a new contextual correction, which is formally similar to CD and CR and that is related to the negating operation. All three of these mechanisms stem from uncertain knowledge about the behavior of the source, which reduce to a set of context $\mathcal{A}$ and associated parameters β_A, $A \in \mathcal{A}$. The next section presents a method to derive such knowledge from labeled data.

8. Learning contextual biases of a source from labeled data

For contextual correction mechanisms to be useful in applications, one needs practical means to choose the set of contexts $\mathcal{A}$ and to determine the associated vector $\boldsymbol{\beta} = (\beta_A, A \in \mathcal{A})$. As already mentioned in Section 1, the set $\mathcal{A}$ and vector $\boldsymbol{\beta}$ could be learned from available labeled data, and in particular using methods based on the minimization of an error criterion. This latter type of methods is considered in this section.

8.1. Description of the learning process

Let us assume that a training set describing the outputs of a source (expressed in the form of a MF) regarding the classes in $\mathcal{X} = \{x_1, \ldots, x_K\}$ of n objects o_i, $i \in \{1, \ldots, n\}$, is available. Small illustrative examples of such training sets are given in Table 4 for two sensors in charge of recognizing flying objects which can be airplanes (a), helicopters (h) or rockets (r).

Table 4
Outputs of two sensors regarding the classes of 4 objects which can be airplanes (a), helicopters (h) or rockets (r). Data come from [12, Table 1].

		a	h	r	$\{a,h\}$	$\{a,r\}$	$\{h,r\}$	$\mathcal{X}$	Ground truth
Sensor 1	$m_{S_1}\{o_1\}$	0	0	0.5	0	0	0.3	0.2	a
	$m_{S_1}\{o_2\}$	0	0.5	0.2	0	0	0	0.3	h
	$m_{S_1}\{o_3\}$	0	0.4	0	0	0.6	0	0	a
	$m_{S_1}\{o_4\}$	0	0	0	0	0.6	0.4	0	r
Sensor 2	$m_{S_2}\{o_1\}$	0	0	0	0.7	0	0	0.3	a
	$m_{S_2}\{o_2\}$	0.3	0	0	0.4	0	0	0.3	h
	$m_{S_2}\{o_3\}$	0.2	0	0	0	0	0.6	0.2	a
	$m_{S_2}\{o_4\}$	0	0	0	0	0	1	0	r

Inspired from previous work in pattern recognition [41], Elouedi et al. [12] propose a method to automatically compute, from such a training set, the degree of reliability $\beta \in [0, 1]$ of the classical discounting operation (11). This scalar β is chosen as the one which minimizes the following measure of discrepancy between the corrected source outputs and the reality:

$$E_{bet}(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{K} (BetP\{o_i\}(x_k) - \delta_{i,k})^2, \tag{41}$$

where $\forall i \in \{1, \ldots, n\}$, $BetP\{o_i\}$ is the pignistic probability [38] associated with the MF $m\{o_i\}$ obtained from a discounting with a degree of reliability β of the output $m_S\{o_i\}$ of the source S regarding the class of object o_i, and $\delta_{i,k}$ is a binary variable that indicates the class of object o_i as follows: $\forall k \in \{1, \ldots, K\}$, $\delta_{i,k} = 1$ if object o_i belongs to the class x_k, and $\delta_{i,k} = 0$ otherwise.

The main idea is to find the reliability degree $\beta \in [0, 1]$ that will bring on average, after correction, the outputs of a source closer to the reality.

In [23], it has been shown that this measure of discrepancy (41) can serve as well to learn the vector $\boldsymbol{\beta} = (\beta_A \in [0, 1], A \in \mathcal{A})$, $\mathcal{A}$ forming a partition of $\mathcal{X}$, of reliability degrees of a contextual discounting based on a coarsening, once a partition (a set of contexts) $\mathcal{A}$ has been fixed.

Moreover, it has also been proposed in [23] to learn this latter vector $\boldsymbol{\beta}$ using another measure of discrepancy based on the plausibility function and defined by:

$$E_{pl}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k=1}^{K} (pl\{o_i\}(\{x_k\}) - \delta_{i,k})^2, \tag{42}$$

where $\forall i \in \{1, \ldots, n\}$, $pl\{o_i\}$ is the plausibility function obtained from a contextual discounting based on a coarsening $\mathcal{A}$ of $\mathcal{X}$, with a vector $\boldsymbol{\beta} = (\beta_A \in [0, 1], A \in \mathcal{A})$ of reliability degrees, of $m_S\{o_i\}$. This measure allows one to express the problem of reliability degrees computation in a more efficient way than (41) does, specifically a constrained least-squares problem [23, Section 5.1]. However, the problem of finding the optimal partition $\mathcal{A}$ of $\mathcal{X}$ for a given source, that is the one minimizing (42), was left open.

In the remainder of this section, we propose to study and refine the above process to automatically learn CD, CR and CN. Precisely, in Section 8.2, we consider the learning of CD with an arbitrary set of contexts $\mathcal{A}$ (17) (i.e., the more general CD that does not require the set of contexts $\mathcal{A}$ to form a partition of $\mathcal{X}$), and of CD+ (35) (i.e., CD based on the canonical decomposition), using the measure of discrepancy E_{pl} (42). In Sections 8.3 and 8.4, similar investigations are pursued for the learning of CR and CN respectively. Learning of CD, CR and CN is then illustrated and commented in Section 8.5. Finally, we show in Section 8.6 that this learning approach may be useful to improve the performance of a source in a classification application.

We remark beforehand that our choice to use measure E_{pl} (42) in our investigations is explained by four main reasons. First, this measure was the preferred one in the approach proposed in [23], which we are clearly extending with this current work. Second, using the plausibility on singletons is in accordance with the Shafer [32] and Smets [39] *singular* [8] interpretation of belief functions adopted in this paper, where one searches to know the actual value of $\mathbf{x}$. Third, as will be seen later, it is possible to obtain simple analytical expressions showing how the parameters β_A of CD, CR and CN, affect the plausibility of singletons, which can be quite helpful when analyzing the respective capacities of these mechanisms to improve a source performance. At last, as it will also be seen later, it can be shown that there exists actually an optimal set of contexts for each of CD, CR and CN, that ensures the minimization of the measure, and that finding this minimum amounts to a computationally simple optimization problem (a constrained least-squares problem with $|\mathcal{X}|$ unknowns). To sum up, we chose E_{pl} to ensure: continuity with previous works, conformance with the belief function theory interpretation used in this paper, ease of analysis and ease of optimization. Yet, we note that other measures of discrepancy could be used, e.g., the measure E_{bet} (41) or a measure based on a distance [15], but then it is neither guaranteed that their minimization can be performed efficiently nor guaranteed that it will be easy to analyze how CD, CR and CN affect them. Let us finally mention that if the measure of discrepancy is used to optimize some decision system, then this measure should be related to the chosen decision rule; for instance, the measure E_{bet} (41) should be used in conjunction with decisions based on pignistic

probability, and E_{pl} (42) should be used, as done in Section 8.6, for decisions based on the plausibility transformation [2], which transforms a belief function into a probability distribution by normalizing the plausibilities on singletons.

8.2. Learning contextual discounting

In this section, the learning of CD according to the plausibility based measure of discrepancy (42) is studied, first for the case of CD defined by (17), which means a contextual discounting of the following form: $m = m_S \textcircled{\cup} m_C$, with m_C a $\textcircled{\cup}$-separable MF.

The main issue here is to decide which set $\mathcal{A}$ to consider and to find the associated vector $\boldsymbol{\beta}$ that minimize measure (42). To try and solve this issue, we can first remark that this latter measure requires that the plausibilities on singletons after having applied such a CD defined by (17), be known. These plausibilities are given in the following proposition.

Proposition 11. *Let $m = m_S \textcircled{\cup}_{A\in\mathcal{A}} A_{\beta_A}$, $\beta_A \in [0,1]$, for all $A \in \mathcal{A}$, be the CD of a MF m_S. The plausibility function associated with m is defined for all $x \in \mathcal{X}$ by:*

$$pl(\{x\}) = 1 - (1 - pl_S(\{x\})) \prod_{A\in\mathcal{A}, x\in A} \beta_A \,. \tag{43}$$

Proof. As $m = m_S \textcircled{\cup} m_C$, with $m_C = \textcircled{\cup}_{A\in\mathcal{A}} A_{\beta_A}$, CD is given in terms of implicability functions by:

$$\begin{aligned} b &= b_S \cdot b_C \\ &= b_S \prod_{A\in\mathcal{A}} b_{\beta_A}, \end{aligned} \tag{44}$$

with, for all $B \subseteq \mathcal{X}$, $b_{\beta_A}(B) = 1$ if $A \subseteq B$, $b_{\beta_A}(B) = \beta_A$ otherwise. Thus, for all $B \subseteq \mathcal{X}$:

$$b(B) = b_S(B) \prod_{A\in\mathcal{A}, A\not\subseteq B} \beta_A. \tag{45}$$

Consequently, for all $x \in \mathcal{X}$:

$$\begin{aligned} pl(\{x\}) &= 1 - b(\overline{\{x\}}) = 1 - b_S(\overline{\{x\}}) \prod_{A\in\mathcal{A}, A\not\subseteq \overline{\{x\}}} \beta_A \quad \text{(from (45))} \\ &= 1 - b_S(\overline{\{x\}}) \prod_{A\in\mathcal{A}, x\in A} \beta_A = 1 - (1 - pl_S(\{x\})) \prod_{A\in\mathcal{A}, x\in A} \beta_A. \qquad \square \end{aligned} \tag{46}$$

Proposition 11 is illustrated by Example 17.

Example 17. Let $\mathcal{X} = \{x_1, x_2, x_3\}$ and consider the CD of a MF m_S with various sets of contexts $\mathcal{A}$:

- If $\mathcal{A} = \{\{x_1\}, \{x_2\}, \{x_3\}\}$, then we have:

$$\begin{aligned} pl(\{x_1\}) &= 1 - (1 - pl_S(\{x_1\}))\beta_{\{x_1\}}, \\ pl(\{x_2\}) &= 1 - (1 - pl_S(\{x_2\}))\beta_{\{x_2\}}, \\ pl(\{x_3\}) &= 1 - (1 - pl_S(\{x_3\}))\beta_{\{x_3\}}. \end{aligned}$$

- If $\mathcal{A} = 2^{\mathcal{X}}$, then we have:

$$\begin{aligned} pl(\{x_1\}) &= 1 - (1 - pl_S(\{x_1\}))\beta_{\{x_1\}}\beta_{\{x_1,x_2\}}\beta_{\{x_1,x_3\}}\beta_{\{x_1,x_2,x_3\}}, \\ pl(\{x_2\}) &= 1 - (1 - pl_S(\{x_2\}))\beta_{\{x_2\}}\beta_{\{x_1,x_2\}}\beta_{\{x_2,x_3\}}\beta_{\{x_1,x_2,x_3\}}, \\ pl(\{x_3\}) &= 1 - (1 - pl_S(\{x_3\}))\beta_{\{x_3\}}\beta_{\{x_1,x_3\}}\beta_{\{x_2,x_3\}}\beta_{\{x_1,x_2,x_3\}}. \end{aligned}$$

- If $\mathcal{A} = \{\{x_2\}\}$, then we have:

$$\begin{aligned} pl(\{x_1\}) &= pl_S(\{x_1\}), \\ pl(\{x_2\}) &= 1 - (1 - pl_S(\{x_2\}))\beta_{\{x_2\}}, \\ pl(\{x_3\}) &= pl_S(\{x_3\}). \end{aligned}$$

The next proposition indicates that the minimization of E_{pl} when CD has been applied, is obtained using the vector $\boldsymbol{\beta}$ composed of the K parameters $\beta_{\{x_k\}}$, which means the parameters associated with the singletons of $\mathcal{X}$. Moreover the minimization of E_{pl} using this vector constitutes a constrained least-squares problem which can then be solved efficiently using standard algorithms.

Proposition 12. *The minimization of E_{pl} with CD is obtained using the vector $\boldsymbol{\beta} = (\beta_{\{x_k\}} \in [0,1], k \in \{1,\ldots,K\})$ and constitutes a constrained least-squares problem as (42) can then be rewritten as:*

$$E_{pl}(\boldsymbol{\beta}) = \|\mathbf{Q}\,\boldsymbol{\beta} - \mathbf{d}\|^2 \quad \textit{with} \quad \mathbf{Q} = \begin{bmatrix} diag(\mathbf{pl_1} - 1) \\ \vdots \\ diag(\mathbf{pl_n} - 1) \end{bmatrix} \quad \textit{and} \quad \mathbf{d} = \begin{bmatrix} \boldsymbol{\delta_1} - 1 \\ \vdots \\ \boldsymbol{\delta_n} - 1 \end{bmatrix}, \tag{47}$$

with $diag(\boldsymbol{v})$ a square diagonal matrix with the elements of vector $\boldsymbol{v}$ on the main diagonal, and with $\mathbf{pl_i} = (pl_S\{o_i\}(\{x_1\}), \ldots, pl_S\{o_i\}(\{x_K\}))^T$, and $\boldsymbol{\delta_i} = (\delta_{i,1}, \ldots, \delta_{i,K})^T$ the column vector of 0–1 class indicator variables for object o_i.

Proof. From Proposition 11, after having applied CD on m_S, the discrepancy measure E_{pl} (42) can be written as: $E_{pl}(\boldsymbol{\beta}) = \sum_{k=1}^{K} E_{pl}(\boldsymbol{\beta}, x_k)$, with for all $k \in \{1,\ldots,K\}$:

$$E_{pl}(\boldsymbol{\beta}, x_k) := \sum_{i=1}^{n} \left(\left(1 - (1 - pl_S\{o_i\}(\{x_k\})) \prod_{A \in \mathcal{A}, x_k \in A} \beta_A \right) - \delta_{i,k} \right)^2 . \tag{48}$$

As $E_{pl}(\boldsymbol{\beta}, x_k) \geq 0$ for all $k \in \{1,\ldots,K\}$, the minimum value of $E_{pl}(\boldsymbol{\beta})$ is obtained when each $E_{pl}(\boldsymbol{\beta}, x_k)$ reaches its minimum.

Besides, from (44), (45) and (46), the product $\prod_{A \in \mathcal{A}, x_k \in A} \beta_A$ of coefficients β_A in $E_{pl}(\boldsymbol{\beta}, x_k)$ (48) is equal to $b_C(x_k)$ for all $k \in \{1,\ldots,K\}$, b_C being the implicability function associated with $m_C = \bigcirc_{A \in \mathcal{A}} A_{\beta_A}$, and belongs then to $[0,1]$ and can be denoted by a variable $\beta_k \in [0,1]$. Hence, for each $k \in \{1,\ldots,K\}$, the minimum of $E_{pl}(\boldsymbol{\beta}, x_k)$ is reached for a particular value of β_k.

Now, we can remark that each coefficient $\beta_{\{x_k\}}$, $k \in \{1,\ldots,K\}$, only appears in the expression of $E_{pl}(\boldsymbol{\beta}, x_k)$ (48), $k \in \{1,\ldots,K\}$. Hence, choosing $\beta_k = \beta_{\{x_k\}}$ for all k (which means choosing $\mathcal{A}$ composed of the set of singletons of $\mathcal{X}$) constitutes then a solution, i.e., a set of contexts for which the minimum value of $E_{pl}(\boldsymbol{\beta})$ is reached.

Each value of E_{pl} is then reachable using the vector $\boldsymbol{\beta}$ of coefficients $\beta_k := \beta_{\{x_k\}}$, $k \in \{1,\ldots,K\}$, and as already mentioned in [23, Section 5.1], the computation of the coefficient $\boldsymbol{\beta}$ with CD based on the singletons is a constrained least-squares problem. Indeed, for all $k \in \{1,\ldots,K\}$, and for all $i \in \{1,\ldots,n\}$:

$$\begin{aligned} pl\{o_i\}(\{x_k\}) - \delta_{i,k} &= 1 - (1 - pl_S\{o_i\}(\{x_k\}))\beta_k - \delta_{i,k} \\ &= (pl_S\{o_i\}(\{x_k\}) - 1)\beta_k - (\delta_{i,k} - 1) \,. \end{aligned}$$

Then (42) can be rewritten as (47). □

This answers a prospect given in [23] concerning the study of the set of contexts which yields the best possible value for the measure of discrepancy E_{pl}. The answer given here is that there will be no smaller value reachable for E_{pl} than the one obtained with the set of the singletons of $\mathcal{X}$ with associated coefficients $\boldsymbol{\beta} = (\beta_{\{x_k\}}, k \in \{1,\ldots,K\})$.

Remark 12. If the more general CD based on the canonical decomposition (cf. Section 6.2) is applied, namely a CD defined by: $m = m_S \bigcirc m_C$ with m_C a non-normal MF, the results of this learning (optimization in the sense of (42)) will still be the same, that is to say a constrained least-squares problem of the form (47) with $|\mathcal{X}|$ unknowns $\beta_{\{x_k\}} \in [0,1]$. Indeed, from (44), (45) and (46), the product $\prod_{A \in \mathcal{A}, x_k \in A} \beta_A$ is equal to $b_C(x_k)$, b_C being the implicability function associated with m_C, and belongs then to $[0,1]$ and thus Proposition 12 also holds for this more general CD. This allows us to remark that the degrees of freedom added by the possibility of having $\beta_A > 1$ are actually useless[9] to improve the measure (42).

8.3. Learning contextual reinforcement

With the same idea but in a reinforcement context, we study in this section the learning of CR according to the plausibility based measure of discrepancy (42).

Plausibilities on the singletons after having applied CR are given in the next proposition.

[9] However, this more general form of CD may be useful considering other discrepancy measures (and it may be, of course, potentially useful outside of a learning context, to model richer knowledge about the behavior of the source than what is allowed by the less general forms of CD).

Proposition 13. *Let $m = m_S \textcircled{\cap}_{A\in\mathcal{A}} A^{\beta_A}$, $\beta_A \in [0,1]$, for all $A \in \mathcal{A}$, be the CR of a MF m_S. The plausibility function associated with m is defined for all $x \in \mathcal{X}$ by:*

$$pl(\{x\}) = pl_S(\{x\}) \prod_{A\in\mathcal{A}, x\notin A} \beta_A \,.$$

Proof. As $m = m_S \textcircled{\cap} m_C$, with $m_C = \textcircled{\cap}_{A\in\mathcal{A}} A^{\beta_A}$, the CR is determined in terms of commonality functions by:

$$\begin{aligned} q &= q_S \cdot q_C \\ &= q_S \prod_{A\in\mathcal{A}} q^{\beta_A} \end{aligned} \tag{49}$$

with, for all $B \subseteq \mathcal{X}$, $q^{\beta_A}(B) = 1$ if $B \subseteq A$, $q^{\beta_A}(B) = \beta_A$ otherwise. Then, for all $B \subseteq \mathcal{X}$:

$$q(B) = q_S(B) \prod_{A\in\mathcal{A}, B\not\subseteq A} \beta_A, \tag{50}$$

which means that after having applied CR, plausibilities on singletons are defined, for all $x \in \mathcal{X}$, by:

$$\begin{aligned} pl(\{x\}) = q(\{x\}) &= q_S(\{x\}) \prod_{A\in\mathcal{A}, x\notin A} \beta_A \\ &= pl_S(\{x\}) \prod_{A\in\mathcal{A}, x\notin A} \beta_A. \qquad \square \end{aligned} \tag{51}$$

Example 18. Let $\mathcal{X} = \{x_1, x_2, x_3\}$ and consider the CR of a MF m_S with various sets of contexts $\mathcal{A}$:

- If $\mathcal{A} = \{\{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_3\}\}$, then we have:

$$\begin{aligned} pl(\{x_1\}) &= pl_S(\{x_1\})\beta_{\{x_2,x_3\}}, \\ pl(\{x_2\}) &= pl_S(\{x_2\})\beta_{\{x_1,x_3\}}, \\ pl(\{x_3\}) &= pl_S(\{x_3\})\beta_{\{x_1,x_2\}}. \end{aligned}$$

- If $\mathcal{A} = 2^{\mathcal{X}}$, then we have:

$$\begin{aligned} pl(\{x_1\}) &= pl_S(\{x_1\})\beta_\emptyset\beta_{\{x_2\}}\beta_{\{x_3\}}\beta_{\{x_2,x_3\}}, \\ pl(\{x_2\}) &= pl_S(\{x_2\})\beta_\emptyset\beta_{\{x_1\}}\beta_{\{x_3\}}\beta_{\{x_1,x_3\}}, \\ pl(\{x_3\}) &= pl_S(\{x_3\})\beta_\emptyset\beta_{\{x_1\}}\beta_{\{x_2\}}\beta_{\{x_1,x_2\}}. \end{aligned}$$

Proposition 14. *The minimization of E_{pl} with CR is obtained using the vector $\boldsymbol{\beta} = (\beta_{\overline{\{x_k\}}} \in [0,1], k \in \{1,\dots,K\})$ and constitutes a constrained least-squares problem as (42) can then be written as:*

$$E_{pl}(\boldsymbol{\beta}) = \|\boldsymbol{P}\boldsymbol{\beta} - \boldsymbol{\delta}\|^2, \quad with \quad \boldsymbol{P} = \begin{bmatrix} diag(\boldsymbol{pl_1}) \\ \vdots \\ diag(\boldsymbol{pl_n}) \end{bmatrix} \quad and \quad \boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\delta_1} \\ \vdots \\ \boldsymbol{\delta_n} \end{bmatrix}, \tag{52}$$

with the same notations as in Proposition 12.

Proof. From Proposition 13, for each $k \in \{1,\dots,K\}$, coefficient $\beta_{\overline{\{x_k\}}}$ only appears in $pl(x_k)$ when a CR has been applied. Then, with the same reasoning as for the CD case, the minimum value of E_{pl} with CR can be reached using the set of contexts $\{\overline{x_k} = \mathcal{X} \setminus \{x_k\}, k \in \{1,\dots,K\}\}$.

The minimization of E_{pl} with CR based on the vector $\boldsymbol{\beta} = (\beta_k := \beta_{\overline{\{x_k\}}}, k \in \{1,\dots,K\})$ is also a constrained least-squares problem as (42) can be written as (52) (as $\forall k \in \{1,\dots,K\}$ and $\forall i \in \{1,\dots,n\}$, $pl\{o_i\}(\{x_k\}) - \delta_{i,k} = pl_S\{o_i\}(\{x_k\})\beta_k - \delta_{i,k}$). $\square$

Remark 13. As was the case for CD (Remark 12), if CR based on the canonical decomposition is used, namely: $m = m_S \textcircled{\cap} m_C$ with m_C a non-dogmatic MF, the learning in the sense of (42) will still yield to the learning of $|\mathcal{X}|$ parameters $\beta_{\overline{\{x_k\}}} \in [0,1]$, as from (49), (50) and (51), for all $x \in \mathcal{X}$, $\prod_{A\in\mathcal{A}, x\notin A} \beta_A = q_C(\{x\}) \in [0,1]$, and thus Proposition 14 also holds for this more general form of CR.

8.4. Learning contextual negating

The learning of CN is explored in this section.
The plausibilities on the singletons after having applied CN are given in the next proposition.

Proposition 15. *Let $m = m_S \textcircled{\cap}_{A\in\mathcal{A}} A^{\beta_A}$ with $\beta_A \in [0,1]$, for all $A \in \mathcal{A}$, be the CN of a MF m_S. The plausibility function associated with m is defined for all $x \in \mathcal{X}$ by:*

$$pl(\{x\}) = 0.5 + 0.5 \cdot (2 \cdot pl_S(\{x\}) - 1) \prod_{A\in\mathcal{A}, x\notin A} (2 \cdot \beta_A - 1). \tag{53}$$

Proof. See Appendix C. □

Proposition 16. *The minimization of E_{pl} with CN is obtained using the vector $\boldsymbol{\beta} = (\beta_{\overline{\{x_k\}}} \in [0,1], k \in \{1,\dots,K\})$ and constitutes a constrained least-squares problem as (42) can then be written as:*

$$E_{pl}(\boldsymbol{\beta}) = \|\boldsymbol{D\beta} - \boldsymbol{z}\|^2, \quad with \quad \boldsymbol{D} = \begin{bmatrix} diag(2\,\boldsymbol{pl_1} - 1) \\ \vdots \\ diag(2\,\boldsymbol{pl_n} - 1) \end{bmatrix} \quad and \quad \boldsymbol{z} = \begin{bmatrix} \boldsymbol{pl_1} + \boldsymbol{\delta_1} - 1 \\ \vdots \\ \boldsymbol{pl_n} + \boldsymbol{\delta_n} - 1 \end{bmatrix}, \tag{54}$$

with the same notations as in Proposition 12.

Proof. The proof is similar to those of Propositions 12 and 14, and uses the fact that from Proposition 15 we obtain $pl\{o_i\}(\{x_k\}) = 0.5 + (pl_S\{o_i\}(\{x_k\}) - 0.5)(2\beta_{\overline{\{x_k\}}} - 1)$, $\forall k \in \{1,\dots,K\}$ and $\forall i \in \{1,\dots,n\}$, which allows us to rewrite (42) as (54). □

8.5. Comments and illustration

In this section, some comments are first made on the respective correction capacities of CD, CR and CN, with respect to the discrepancy measure (42) and taking into consideration the findings of Propositions 12, 14 and 16. Then, an illustrative example of the proposed learning of CD, CR and CN, is given. This example is also useful to make insightful additional remarks on CD, CR and CN, with respect to source performance improvement.

8.5.1. Correction capacities

The plausibility ranges on singletons after having applied CD, CR and CN are given by Remarks 14, 15 and 16, respectively.

Remark 14. With CD, as $pl(\{x\}) = 1 - (1 - pl_S(\{x\}))\beta_{\{x\}}$ for each $x \in \mathcal{X}$, with $\beta_{\{x\}}$ varying in $[0,1]$, $pl(\{x\})$ can take any values in the interval $[pl_S(\{x\}), 1]$. It means that with CD, the plausibility $pl_S(\{x\})$ on each singleton can be shifted as close to 1 as required.

Remark 15. With CR, as $pl(\{x\}) = pl_S(\{x\})\beta_{\overline{\{x\}}}$ for each $x \in \mathcal{X}$, with $\beta_{\overline{\{x\}}}$ varying in $[0,1]$, $pl(\{x\})$ can take any values in $[0, pl_S(\{x\})]$. In other words, with CR, the plausibility $pl_S(\{x\})$ on each singleton can be carried as close to 0 as necessary.

Remark 16. With CN, as $pl(\{x\}) = 0.5 + (pl_S(\{x\}) - 0.5)(2\beta_{\overline{\{x\}}} - 1)$ for each $x \in \mathcal{X}$, with $\beta_{\overline{\{x\}}}$ varying in $[0,1]$, $pl(\{x\})$ can take any values in the interval $[\min(pl_S(\{x\}), 1 - pl_S(\{x\})), \max(pl_S(\{x\}), 1 - pl_S(\{x\}))]$. It is of particular interest if $pl_S(\{x\})$ is close to 0 or on the contrary close to 1 meaning that in these cases $pl_S(\{x\})$ can be moved to any values in $[0,1]$. However if $pl_S(\{x\})$ is close to 0.5, CN is not able to change its value which is confined around 0.5.

The following example illustrates these different capacities of adjustment on simple scenarios.

Example 19. Let $\mathcal{X} = \{a,b,c\}$ and suppose, without lack of generality, that the ground truth is a.

Suppose a source n°1 outputs a mass $m_S(\{b,c\}) = 1$ which means that $pl_S(\{a\}) = 0$ and $pl_S(\{b\}) = pl_S(\{c\}) = 1$. From Remarks 14, 15 and 16, to bring source n°1 output closer to the ground truth: CD can increase $pl_S(\{a\})$ to 1; CR can decrease $pl_S(\{b\})$ to 0 and $pl_S(\{c\})$ to 0; CN can increase $pl_S(\{a\})$ to 1 and decrease $pl_S(\{b\})$ to 0 and $pl_S(\{c\})$ to 0.

This case is presented again in Table 5, where two more situations are considered: a source n°2 giving $m_S(\{c\}) = 1$, that is $pl_S(\{a\}) = pl_S(\{b\}) = 0$ and $pl_S(\{c\}) = 1$, and a source n°3 giving $m_S(\{a,b\}) = 1$, that means $pl_S(\{a\}) = pl_S(\{b\}) = 1$ and $pl_S(\{c\}) = 0$.

As it can be observed in Table 5, CD can improve only one value of plausibility: the plausibility on the ground truth by increasing it as close as possible to 1, whereas CR can improve the other plausibility values (the ones not associated

Table 5
Attainable plausibilities with CD, CR and CN for three sources outputs.

	Ground truth	Source n°1	CD	CR	CN	Source n°2	CD	CR	CN	Source n°3	CD	CR	CN
$pl(\{a\})$	1	0	1	0	1	0	1	0	1	1	1	1	1
$pl(\{b\})$	0	1	1	0	0	0	0	0	0	1	1	0	0
$pl(\{c\})$	0	1	1	0	0	1	1	0	0	0	0	0	0
		CD: $E_{pl}=2$ CR: $E_{pl}=1$ CN: $E_{pl}=0$				CD: $E_{pl}=1$ CR: $E_{pl}=1$ CN: $E_{pl}=0$				CD: $E_{pl}=1$ CR: $E_{pl}=0$ CN: $E_{pl}=0$			

Table 6
Results for the minimization of E_{pl} with the data in Table 4 for each contextual correction mechanism for both sensors 1 and 2.

Contextual correction	Sensor 1	Sensor 2
CD	$\boldsymbol{\beta}=(0.76, 1.00, 1.00)$ $E_{pl}(\boldsymbol{\beta})=3.39$	$\boldsymbol{\beta}=(0.74, 1.00, 1.00)$ $E_{pl}(\boldsymbol{\beta})=4.81$
CR	$\boldsymbol{\beta}=(0.94, 0.66, 0.38)$ $E_{pl}(\boldsymbol{\beta})=2.33$	$\boldsymbol{\beta}=(0.65, 0.22, 0.55)$ $E_{pl}(\boldsymbol{\beta})=2.39$
CN	$\boldsymbol{\beta}=(0.33, 1.00, 0.45)$ $E_{pl}(\boldsymbol{\beta})=2.59$	$\boldsymbol{\beta}=(0.63, 0.06, 0.86)$ $E_{pl}(\boldsymbol{\beta})=2.25$

with the ground truth) by decreasing them as near as possible to 0. In contrast, CN can improve all plausibility values, by increasing the plausibility on the ground truth up to 1 and by decreasing the other possibilities down to 0. CR has then more degrees of flexibility than CD to improve the plausibility output of the source, and CN has in turn one more degree of flexibility than CR, on these three particular cases. As a result, we can see in Table 5 that for each source, CN has a lower (or equal) E_{pl} than CR, which in turn has a lower (or equal) E_{pl} than CD.

Let us note however that there exist situations where CD may be of more help than CR and CN, in particular those where all the plausibilities on singletons which are not the ground truth are below 0.5, i.e., $pl_S(\{b\}) < 0.5$ and $pl_S(\{c\}) < 0.5$, and where the plausibility on the ground truth is above 0.5, i.e., $pl_S(\{a\}) > 0.5$. In such a situation, using Remark 14, the lowest possible E_{pl} for CD, denoted by E_{pl}^{CD}, is attained with vector $\boldsymbol{\beta}=(0,1,1)$, in which case we have $E_{pl}^{CD}=(pl_S(\{b\}))^2+(pl_S(\{c\}))^2$. Similarly, the lowest possible E_{pl} for CR is attained with, using Remark 15, vector $\boldsymbol{\beta}=(1,0,0)$, in which case $E_{pl}^{CR}=(1-pl_S(\{a\}))^2$. Concerning CN, using Remark 16, the lowest possible E_{pl} is $E_{pl}^{CN}=(pl_S(\{b\}))^2+(pl_S(\{c\}))^2+(1-pl_S(\{a\}))^2$ which is attained for $\boldsymbol{\beta}=(1,1,1)$. We thus see that in such a situation, we have necessarily $E_{pl}^{CD}<E_{pl}^{CN}$ and $E_{pl}^{CR}<E_{pl}^{CN}$. Besides, we also have $E_{pl}^{CD}<E_{pl}^{CR}$ if $(pl_S(\{b\}))^2+(pl_S(\{c\}))^2<(1-pl_S(\{a\}))^2$; an example of a source output satisfying this latter inequality is: $m_S(\{a\})=0.6$ and $m_S(\{b\})=m_S(\{c\})=0.2$.

8.5.2. An illustrative example

Let us consider the data given in Table 4.

Results of the minimization of E_{pl} for CD, CR and CN are summarized in Table 6 for both sensors 1 and 2. Let us recall that $\boldsymbol{\beta}=(\beta_{\{a\}},\beta_{\{h\}},\beta_{\{r\}})$ for CD, and $\boldsymbol{\beta}=(\beta_{\overline{\{a\}}},\beta_{\overline{\{b\}}},\beta_{\overline{\{c\}}})$ for CR and CN, with different meanings and associated transformations for each vector.

In order to analyze the results presented in Table 6, one may look at the mass transfers associated with these learned vectors, that is the transformations of the sensor outputs induced by CD, CR and CN. For instance, the CD learned vector $\boldsymbol{\beta}=(0.76, 1.00, 1.00)$ for sensor 1 indicates that for each $B\subseteq\mathcal{X}$, a portion $(1-0.76=0.24)$ of mass $m_S(B)$ should be transferred to $B\cup\{a\}$. One may also use the interpretation given to the parameters β_A; for instance having $\beta_{\{a\}}=0.76$ indicates that sensor 1 negatively lies for airplanes with mass 0.24. Yet, we find it more instructive as well as more appropriate considering the discrepancy measure used, to analyze those results in light of the changes they suggest on the plausibilities given to the singletons by the sensors. We focus on such an analysis hereafter.

For CD it can be observed that $\beta_{\{h\}}=\beta_{\{r\}}=1$ for both sensors, which means from the previous Remark 14 that $pl_S(\{r\})$ and $pl_S(\{h\})$ do not have to be increased to improve E_{pl}, which is not the case for $pl_S(\{a\})$ as it has to be increased since $\beta_{\{h\}}<1$ ($pl_S(\{a\})$ will be increased slightly more for sensor 2 than for sensor 1 since $0.76>0.74$). This means that the sensors may be considered cautious enough concerning the plausibilities they allocate to objects of type h and r, but may be a bit too bold concerning objects of type a. In particular, if for any object to be classified they give a low plausibility to h or to r then it should be left as is, but if they give a low plausibility to a then it should be increased.

Based on Remark 15, the CR learned vector for sensor 1 indicate that $pl_S(\{a\})$ should (almost) not be decreased, whereas $pl_S(\{h\})$ and even more $pl_S(\{r\})$ have to be decreased for this sensor. This means that sensor 1 is bold enough for a, but too cautious for h and especially too cautious for r. So if sensor 1 gives a high plausibility to h or to r, then it should be decreased (even more so for r), but a high plausibility on a should remain roughly as is. Conclusions are different for

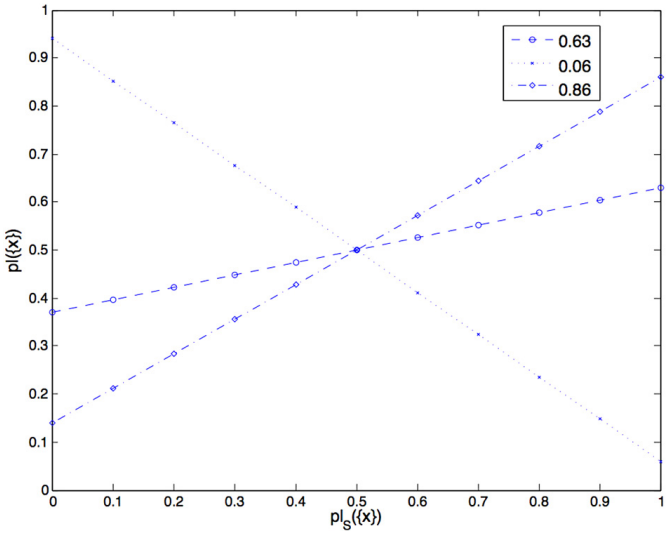


Fig. 3. Plausibility on a singleton $pl(\{x\})$ after CN correction of $pl_S(\{x\})$, for three different values of $\beta_{\overline{\{x\}}}$.

sensor 2: all plausibilities should be substantially decreased. Precisely, only 65% of $pl_S(\{a\})$ should be left on a, 22% of $pl_S(\{h\})$ should be left on h and 55% of $pl_S(\{r\})$ should be left on r.

For CN, we know from Remark 16 that we have $pl(\{x\}) = 0.5 + (pl_S(\{x\}) - 0.5)(2\beta_{\overline{\{x\}}} - 1)$ for each $x \in \mathcal{X}$. Fig. 3 shows this latter equation for the three values of $\beta_{\overline{\{x\}}}$ learned for sensor 2. As it can be seen on this figure, if we have $pl_S(\{h\}) < 0.5$ for sensor 2, then this plausibility on h should be increased, for instance if $pl_S(\{h\}) = 0.2$ then this value should be increased to $pl(\{h\}) = 0.5 + (pl_S(\{h\}) - 0.5)(2 \cdot \beta_{\overline{\{h\}}} - 1) = 0.5 + (0.2 - 0.5)(2 \cdot 0.06 - 1) = 0.764$. Conversely, if $pl_S(\{h\}) > 0.5$, then this plausibility on h should be decreased. For a and r, the situation is similar: if the plausibilities on these singletons are lower than 0.5, then they should be increased, and if they are greater than 0.5, they should be decreased, albeit with different rates. Of note for sensor 1, is the fact that $pl_S(\{h\})$ should not be altered at all since $\beta_{\overline{\{h\}}} = 1$ for this sensor, which is quite different from the situation observed for sensor 2.

This small illustrative example shows that CD, CR and CN yield different tunings of a source. Perhaps most importantly, on a practical side, Table 6 shows that CR and CN also permit to obtain lower values for E_{pl} than those reached with CD, for both sensors 1 and 2, which confirms the advantages in some cases of CR and CN over CD exposed in Example 19 concerning the minimization of E_{pl}. This demonstrates the interest of investigating alternative correction mechanisms distinct from a discounting process. A second observation is that a lower value for E_{pl} using CN has been obtained for sensor 2 compared to the value obtained with CR, and conversely a lower value for E_{pl} using CR has been obtained for sensor 1 compared to the value obtained with CN, which shows the potential utility of both mechanisms in terms of performance improvement.

Finally, let us note that even if CR and CD are related (CR amounts to the negation of the CD of the negation of the information provided by the source [20]), CR and CD parameters minimizing E_{pl} (42) cannot be deduced analytically from each other, as illustrated by Example 20, which shows that knowing the vector $\boldsymbol{\beta}$ minimizing E_{pl} for CD does not imply knowing the vector $\boldsymbol{\beta}$ minimizing E_{pl} for CR.

Example 20. Let us modify in Table 4, MF $m_{S_1}\{o_1\}$ by

$$m_{S_1}\{o_1\}(\{r\}) = 0.5282,$$
$$m_{S_1}\{o_1\}(\{h, r\}) = 0.3000,$$
$$m_{S_1}\{o_1\}(\mathcal{X}) = 0.1718,$$

i.e., information coming from sensor 1 is slightly deteriorated, the truth being a. Then, learnings of CD parameters for sensors 1 and 2 yield the same vector $\boldsymbol{\beta} = (0.74, 1.00, 1.00)$, while learnings of CR parameters yield $\boldsymbol{\beta} = (0.92, 0.68, 0.38)$ for sensor 1 and $\boldsymbol{\beta} = (0.65, 0.22, 0.55)$ for sensor 2.

8.6. Application in classification

Despite the formal elegance of CD, CR and CN, their usefulness in belief function based applications might be challenged. Therefore, an experiment is conducted in this section to demonstrate their ability to improve the performances of an evidential classifier.

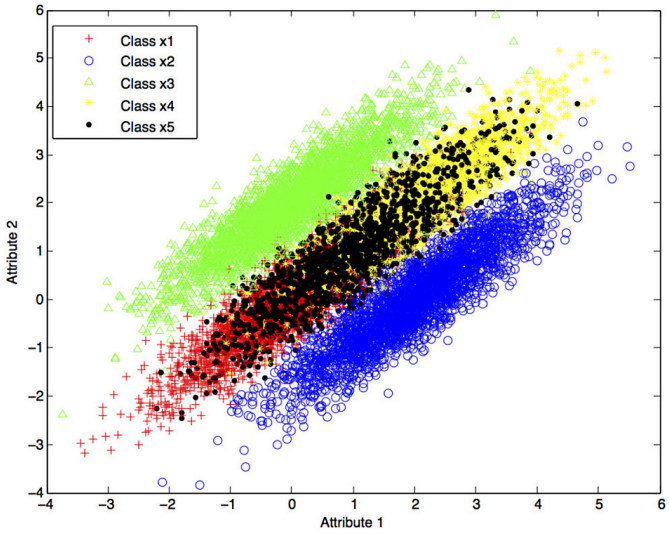


Fig. 4. Illustration of the data generated for a 5-class classification problem with 2 features.

The chosen evidential classifier is the evidential k-nearest neighbor classifier (*ev-knn*) introduced by Denœux in [5]. It is used on a 5-class classification problem with data generated from 5 bivariate normal distributions with respective means $\mu_{x_1} = (0, 0)$, $\mu_{x_2} = (2, 0)$, $\mu_{x_3} = (0, 2)$, $\mu_{x_4} = (2, 2)$, $\mu_{x_5} = (1, 1)$ and common variance matrix

$$\Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}.$$

For each class $x \in \mathcal{X} = \{x_1, x_2, x_3, x_4, x_5\}$, 1000 instances have been generated. The total amount of data is then composed of 5000 instances and is illustrated in Fig. 4.

The principle of ev-knn [5] is to consider the k-nearest neighbors n_i (according to a distance measure d, e.g., the Euclidean distance), $i \in \{1, \ldots, k\}$, of a new instance a to be classified, and to build a MF regarding the class of a, which results from the combination by Dempster's rule of k MFs m_i, $i \in \{1, \ldots, k\}$, each one of these MFs being associated with a neighbor n_i of a, and reflecting a piece of evidence regarding the class of the instance a to be classified. With $x^i \in \mathcal{X}$ the class of n_i, each m_i, $i \in \{1, \ldots, k\}$, is defined in the following manner:

$$m_i(A) = \begin{cases} \tau \cdot e^{-\gamma_{x^i} (d(a,n_i))^2} & \text{if } A = \{x^i\}, \\ 1 - \tau \cdot e^{-\gamma_{x^i} (d(a,n_i))^2} & \text{if } A = \mathcal{X}, \\ 0 & \text{otherwise,} \end{cases} \tag{55}$$

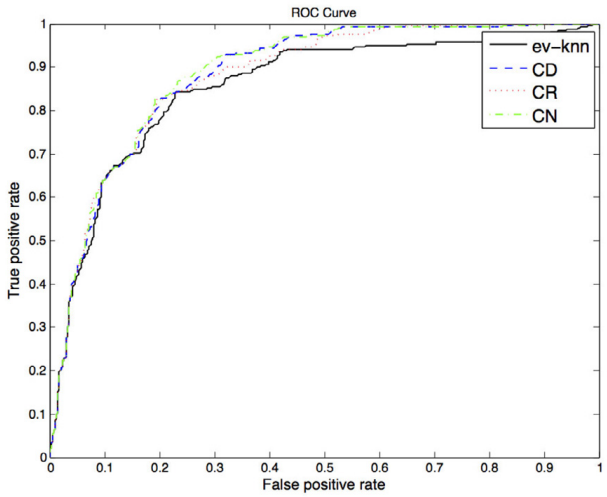
with τ a constant in $[0, 1]$, γ_x a positive constant depending on class $x \in \mathcal{X}$, and d a distance. For each neighbor n_i of a, the knowledge of the class x^i of n_i is a piece of evidence increasing the belief that the class of a is x^i depending on the distance $d(a, n_i)$ between a and n_i.

Following the simple heuristic used in [5], τ has been fixed to 0.95 and γ_x has been chosen equal to the inverse of the mean distance between each sample of class x, which belongs to the training set of the classifier ev-knn. The number k of neighbors has been chosen equal to 3.

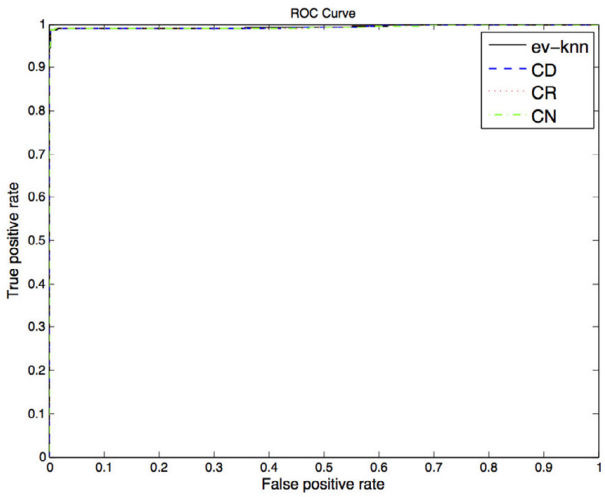
Let us note that these latter choices for the parameters of the ev-knn classifier may not be the most optimal, i.e., the ones leading to the best possible performances with respect to the experiment described in this section; in particular, an approach has been proposed in [41] to optimize these parameters. Similarly, ev-knn may not be the (evidential) classifier, with the best possible performances on this experiment. The idea in this section is not to show that correction mechanisms yield the best possible classifier on a given classification problem. Rather, the aim is to illustrate the fact that given a real-life situation where one has access to an information source, such as a sensor, which cannot really be controlled (i.e., is like a black box) and which is not necessarily the best possible sensor out there, it may be possible to improve this source using correction mechanisms. Hence, the role played by ev-knn in our experiment is simply to simulate such a source.

The 5000 instances displayed in Fig. 4 have been divided into three parts:

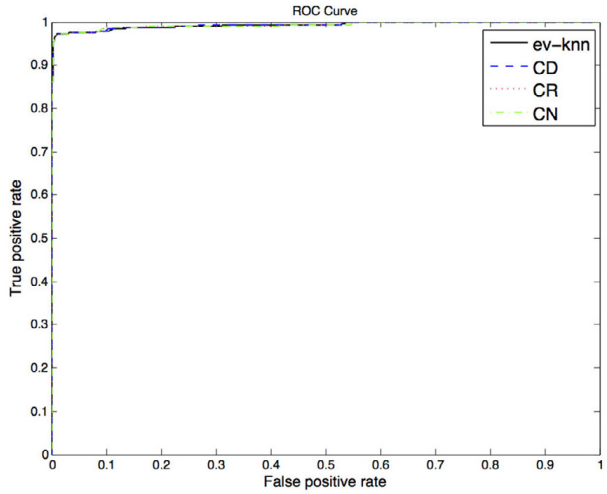
1. The first third of the data constitutes the training set of ev-knn, i.e., the set used to learn the parameters γ_x for each class $x \in \mathcal{X}$, as well as the set of neighbors used to classify samples;
2. The second third of the data constitutes the training set for the correction mechanisms, i.e., it is dedicated to learn the best CD, CR and CN of ev-knn outputs;
3. The last third of the data constitutes the test set, i.e., it is used to test these learned corrections of ev-knn, as well as ev-knn.



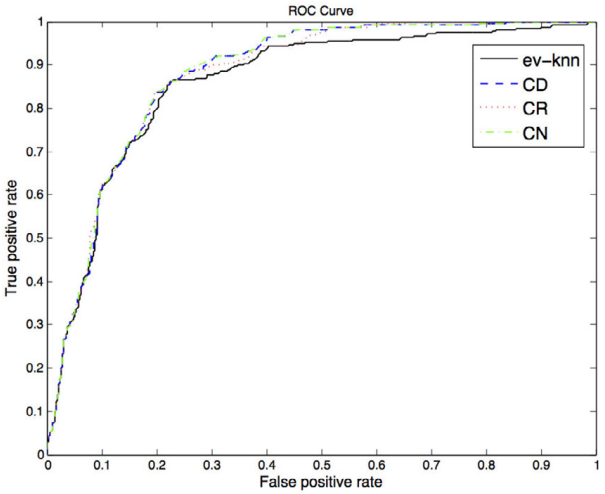
(a) Positive class: class x_1.



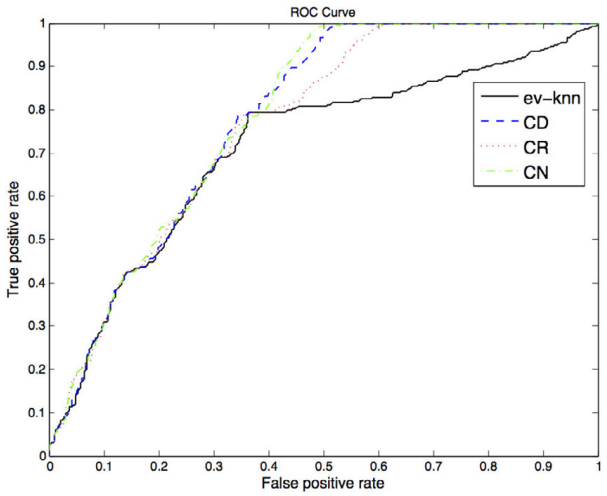
(b) Positive class: class x_2.



(c) Positive class: class x_3.



(d) Positive class: class x_4.



(e) Positive class: class x_5.

Fig. 5. ROC curves for each of the 5 classes.

The performances of ev-knn without correction (ev-knn), ev-knn with CD (CD), ev-knn with CR (CR), and ev-knn with CN (CN), are reported in Fig. 5 using ROC curves (a higher curve corresponds to higher performance); the ROC curves were obtained using the plausibility transformation.

As may be expected, samples of classes x_2 and x_3, which are clearly disjoint from the samples of the other classes (cf. Fig. 4), are very well classified by ev-knn (cf. Figs. 5b and 5c). Besides, CD, CR and CN neither improve nor deteriorate ev-knn outputs for these classes.

Samples of classes x_1, x_4 and x_5, overlap (cf. Fig. 4) and are not so well classified by ev-knn (cf. Figs. 5a, 5d and 5e). Most interestingly, CD, CR and CN succeed in improving ev-knn outputs for these more difficult classes, which is an experimental evidence of the interest of these correction mechanisms to improve the performance of a source.

9. Conclusion

The aim of this study was twofold: (1) to enlarge the set of tools available to deal with contextual knowledge about the quality of a source, and (2) to provide a practical means to obtain such kind of knowledge.

Several conclusions may be drawn from our works and with respect to this aim.

First, it is indeed possible to find useful complements to contextual discounting based on a coarsening and its recent extension to an arbitrary set of contexts: as illustrated through numerous examples, CD, CR, CN, CdD, CdR, CD+ and CR+ may be used to account for various situations with respect to meta-knowledge about a source, and some of them have even been shown to be also interesting to improve a source performance.

The pitfall is that these contextual correction mechanisms correspond to quite specific meta-knowledge about a source, and accordingly to rather precise and different interpretations of a given piece of information provided by the source, and thus one must be careful in choosing one or the other mechanism and in setting its associated parameters $\mathcal{A}$ and $\boldsymbol{\beta}$.

When not much is known about the quality of a source, this variety and associated difficulty may seem daunting. Fortunately, if labeled data are available, we have shown that it is possible to learn which mechanism is the best (in terms of performance), and its associated parameters. In this case, the subtlety resides in using an error criterion that is meaningful, that leads itself to easy analysis and that also leads to an optimization problem, which can be solved efficiently.

This study may be continued in various directions, which are left for further research.

In [28], it was envisioned to derive a contextual version of negating by considering a situation where the agent holds beliefs concerning the truthfulness of the source conditionally on different subsets of $\mathcal{X}$, much as contextual discounting based on a coarsening was originally derived in [23] and is extended to an arbitrary set of subsets in [21]. It would be interesting to study this alternative path to derive a contextual version of negating and to compare the resulting correction mechanism to CN.

Compared to CD and CR, some formal results remain to be obtained for CN. In particular, since the inverse of the rule $\cap$, and precisely the 0-commonality function, is at the base of what would be contextual de-negating and thus of contextual negating based on the canonical decomposition, it would be interesting to study further this function in order to know under which conditions it does not equal to 0, similarly as we know that commonality $q(A) \neq 0$ for all $A \subseteq \mathcal{X}$ if the mass function is non-dogmatic.

The conjunctive and equivalence rules, which are at the base of CR and CN respectively, are actually two extreme members of a family of rules known as the α-conjunctions [35,25] depending on a parameter $\alpha \in [0, 1]$. Hence, formally, CR and CN may easily be presented as two members of a family of contextual correction mechanisms based on the α-conjunctions. While it might be possible to find an interpretation for all the members in this family, their main interest might reside in applications, where they could be used as flexible tools to improve a source. This would require extending to this family of correction mechanisms, the efficient means proposed in this paper to learn CR and CN.

Concerning the learning of contextual correction parameters, other discrepancy measures (in particular distances [15,18]) than the one based on the plausibility function may be studied. Other approaches for obtaining these parameters may also be investigated, such as expert elicitation procedures (see, e.g., [29] for an expert elicitation procedure of the knowledge associated with discounting) or methods using confusion matrices, which may be appropriate for discovering truthfulness [17]. Finally, these approaches should also be looked at to obtain the parameters associated with the extension of contextual discounting based on a coarsening uncovered recently in [21].

Acknowledgements

The authors thank the anonymous referees for their valuable comments that helped to improve this paper. Thanks also to Stéphane Meilliez for providing a shorter proof of Lemma 6.

Appendix A. A general model of source truthfulness

In this appendix, a general model of source truthfulness is presented. It includes as particular cases the three states ℓ_A, p_A, and n_A introduced in Sections 3 and 4.

A.1. Elementary-level truthfulness

Assume that a source S provides a piece of information on the value taken by $\mathbf{x}$ of the form $\mathbf{x} \in B$, for some $B \subseteq \mathcal{X}$.

Table 7
Interpretations of the source testimony according to couples (t_1, t_2).

$x \in B$	$x \in A$	$\neg t_x^p, \neg t_x^p$	$t_x, \neg t_x^p$	$\neg t_x^p, t_x$	t_x, t_x	$\neg t_x, \neg t_x^p$	$\neg t_x^n, \neg t_x^p$	$\neg t_x, t_x$	$\neg t_x^n, t_x$	$\neg t_x^p, \neg t_x$	$t_x, \neg t_x$	$\neg t_x^p, \neg t_x^n$	$t_x, \neg t_x^n$	$\neg t_x, \neg t_x$	$\neg t_x^n, \neg t_x$	$\neg t_x, \neg t_x^n$	$\neg t_x^n, \neg t_x^n$
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

Let us now consider a particular value $x \in \mathcal{X}$. In Sections 3.1 and 4.1, the notions of (non) truthful, *positively* (non) truthful and *negatively* (non) truthful for a value $x \in \mathcal{X}$ have been defined (Definitions 1, 3 and 4, respectively). Let us denote by $\mathbf{t}_x$ a variable with associated frame $\mathcal{T}_x = \{t_x, \neg t_x, \neg t_x^p, \neg t_x^n\}$ allowing us to model the global truthfulness of the source with respect to the value x, where:

- t_x corresponds to the case where the source tells the truth whatever it says about the value x, i.e., to a (positively and negatively) truthful source for x;
- $\neg t_x$ corresponds to the case where the source lies whatever it says about the value x, i.e., to a (positively and negatively) non-truthful source for x;
- $\neg t_x^p$ corresponds to the case of a source that lies only when it says that x is a possibility for $\mathbf{x}$, i.e., to a positively non-truthful and negatively truthful source for x;
- $\neg t_x^n$ corresponds to the case of a source that lies only when it tells that x is not a possibility for $\mathbf{x}$, i.e., to a positively truthful and negatively non-truthful source for x.

Thus, there are four possible cases:

1. Suppose the source tells x is possibly the actual value of $\mathbf{x}$, i.e., the information $\mathbf{x} \in B$ provided by the source is such that $x \in B$.
   (a) If the source is assumed to be in state t_x or $\neg t_x^n$, then one must conclude that x is possibly the actual value of $\mathbf{x}$;
   (b) If the source is assumed to be in state $\neg t_x^p$ or $\neg t_x$, then one must conclude that x is not a possibility for the actual value of $\mathbf{x}$;
2. Suppose the source tells x is not a possibility for the actual value of $\mathbf{x}$, i.e., $x \notin B$.
   (a) If the source is assumed to be in state t_x or $\neg t_x^p$, then one must conclude that x is not a possibility for the actual value of $\mathbf{x}$;
   (b) If the source is assumed to be in state $\neg t_x^n$ or $\neg t_x$, then one must conclude that x is possibly the actual value of $\mathbf{x}$;

A.2. Two truthfulness assumptions

Let $\mathcal{T}$ denote the possible states of S with respect to its truthfulness for all $x \in X$. By definition, $\mathcal{T} = \times_{x \in \mathcal{X}} \mathcal{T}_x$.

Let $h_A^{t_1,t_2} \in \mathcal{T}$, $A \subseteq \mathcal{X}$, denote the state where the source is in state $t_1 \in \mathcal{T}_x$ for all $x \in A$, and in state $t_2 \in \mathcal{T}_x$ for all $x \notin A$. For instance, let $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, $A = \{x_3, x_4\}$, $t_1 = t_x$ and $t_2 = \neg t_x^p$, then

$$h_A^{t_1,t_2} = h_{\{x_3,x_4\}}^{t_x,\neg t_x^p} = \left(\neg t_{x_1}^p, \neg t_{x_2}^p, t_{x_3}, t_{x_4}\right),$$

i.e., the source is truthful for x_3 and x_4, and is positively non-truthful and negatively truthful for x_1 and x_2.

Consider now the following question: what must one conclude about $\mathbf{x}$ when the source tells $\mathbf{x} \in B$ and is assumed to be in some state $h_A^{t_1,t_2}$? To answer this question, one merely needs to look in turn at each $x \in \mathcal{X}$ and:

1. to find for each of those $x \in \mathcal{X}$, whether $x \notin B$ or $x \in B$;
2. to find for each of those $x \in \mathcal{X}$, whether $x \notin A$ or $x \in A$, and, accordingly, which of t_1 or t_2 applies for each of those $x \in \mathcal{X}$;
3. and then, using information obtained at steps 1 and 2, to find out for each of those $x \in \mathcal{X}$, which one of the four cases 1.a), 1.b), 2.a) or 2.b) described at the very end of Section A.1 applies.

Table 7 lists exhaustively, i.e., for all possible cases with respect to the membership of a given value x to the sets B and A, and for all possible couples (t_1, t_2), whether one should deduce that a given value $x \in \mathcal{X}$ is possibly the actual value of $\mathbf{x}$ or not – the former is indicated by a 1 and the latter by a 0 in columns (t_1, t_2). According to Table 7, when the source is assumed to be in, e.g., state $h_A^{t_x,\neg t_x^p}$, then one should deduce that $x \in \mathcal{X}$ is a possible value for $\mathbf{x}$ iff $x \in B$ and $x \in A$, and therefore, since this holds for all $x \in \mathcal{X}$, one should deduce that $\mathbf{x} \in B \cap A$. We may then remark that state $h_A^{t_x,\neg t_x^p}$ is actually nothing but state p_A discussed in Section 4.2. Similarly, state $h_A^{\neg t_x^n,t_x}$ corresponds to state n_A of Section 4.3, and state $h_A^{t_x,\neg t_x}$ correspond to state ℓ_A of Section 3.2.

In addition, let us remark that states ℓ_A, p_A, n_A, induce that a testimony $\mathbf{x} \in B$ should be combined with subset A using operators $\underline{\cap}$, $\cap$, $\cup$, respectively. This comes from the fact that these states are associated respectively to logical equality, conjunction, and disjunction, as can be seen from Table 7. More generally, this latter table shows that the couples $(t_1, t_2) \in \mathcal{T}_x^2$ actually yield all possible binary Boolean connectives; a formally interesting result pertaining to information correction which has some similarity with what was done recently in information fusion in [28] and in [27], where the conjunctive rule and the notion of conflict, respectively, are extended to other binary Boolean connectives than the conjunction.

Remark 17. In this paper, we are only interested by the states based on the couples $(t_1, t_2) \in \{(t_x, \neg t_x^p), (\neg t_x^n, t_x), (t_x, \neg t_x)\}$ since they are the only ones needed for our developments. Yet, let us note that the other couples may be useful. For instance, the state $h_{x_i}^{\neg t_x^n, \neg t_x^p}$, which is such that $\Gamma_B(h_{x_i}^{\neg t_x^n, \neg t_x^p}) = x_i$, for all $B \subseteq \mathcal{X}$, allows one to recover the correction mechanism used in [1] to favor a given element $x_i \in \mathcal{X}$.

Appendix B. Single correction perspective

Interestingly, the BBCs corresponding to simple contextual biases put forward by Proposition 7, are equivalent to a single BBC corresponding to some particular knowledge on the truthfulness of the information source as shown by Proposition 17.

Proposition 17.

$$(\circ_{A\in\mathcal{A}}\, f_{m_{A,\cup}^{\mathcal{H}}})(m_S) = f_{m_{\mathcal{A},\cup}^{\mathcal{H}}}(m_S), \tag{56}$$

with $m_{\mathcal{A},\cup}^{\mathcal{H}} = \textcircled{\cup}_{A\in\mathcal{A}} m_{A,\cup}^{\mathcal{H}}$.

Proof. Let us first consider the left side of Equation (56). From Proposition 7, we have

$$(\circ_{A\in\mathcal{A}}\, f_{m_{A,\cup}^{\mathcal{H}}})(m_S) = m_S \textcircled{\cup}_{A\in\mathcal{A}} A_{\beta_A}. \tag{57}$$

From the definition of $\textcircled{\cup}$, we have that Equation (57) allocates, for all $\mathcal{C} \subseteq \mathcal{A}$ and all $B \subseteq \mathcal{X}$, the quantity

$$m_S(B) \cdot \prod_{A\in\mathcal{C}} (1-\beta_A) \cdot \prod_{D\in\mathcal{A}\setminus\mathcal{C}} \beta_D$$

to $B \cup (\cup_{A\in\mathcal{C}} A) \subseteq \mathcal{X}$, and a null mass to all $D \subseteq \mathcal{X}$, $D \neq B \cup (\cup_{A\in\mathcal{C}} A), \forall \mathcal{C} \subseteq \mathcal{A}, \forall B \subseteq \mathcal{X}$.

Let us now consider the right side of Equation (56). Let $\mathcal{H}_{\mathcal{A},n} = \{n_A, A \in \mathcal{A}\}$. The mass function $m_{\mathcal{A},\cup}^{\mathcal{H}}$ clearly has $2^{|\mathcal{H}_{\mathcal{A},n}|}$ focal sets, which are $\mathcal{H}_{\mathcal{A},n}$ and the sets $\{t \cup \{n_A, A \in \mathcal{C}\}\}$, for all $\mathcal{C} \subset \mathcal{A}$. Besides, from the definition of $\textcircled{\cup}$, the mass function $m_{\mathcal{A},\cup}^{\mathcal{H}}$ is such that

$$m_{\mathcal{A},\cup}^{\mathcal{H}}(\mathcal{H}_{\mathcal{A},n}) = \prod_{A\in\mathcal{A}} (1-\beta_A),$$
$$m_{\mathcal{A},\cup}^{\mathcal{H}}(\{t \cup \{n_A, A \in \mathcal{C}\}\}) = \prod_{A\in\mathcal{C}} (1-\beta_A) \cdot \prod_{D\in\mathcal{A}\setminus\mathcal{C}} \beta_D, \quad \forall \mathcal{C} \subset \mathcal{A}.$$

Furthermore, from the definition of the BBC procedure, we have that $f_{m_{\mathcal{A},\cup}^{\mathcal{H}}}(m_S)$ allocates,

- for all $\mathcal{C} \subset \mathcal{A}$ and all $B \subseteq \mathcal{X}$, the quantity

  $$m_S(B) \cdot \prod_{A\in\mathcal{C}} (1-\beta_A) \cdot \prod_{D\in\mathcal{A}\setminus\mathcal{C}} \beta_D$$

  to

  $$\begin{aligned}\Gamma_B(\{t \cup \{n_A, A \in \mathcal{C}\}\}) &= \Gamma_B(t) \cup (\cup_{A\in\mathcal{C}} \Gamma_B(n_A)) \\ &= B \cup (\cup_{A\in\mathcal{C}} (B \cup A)) \\ &= B \cup (\cup_{A\in\mathcal{C}} A)\end{aligned}$$

- and for $\mathcal{C} = \mathcal{A}$ and all $B \subseteq \mathcal{X}$, the quantity

  $$m_S(B) \cdot \prod_{A\in\mathcal{A}} (1-\beta_A) = m_S(B) \cdot \prod_{A\in\mathcal{C}} (1-\beta_A) \cdot \prod_{D\in\mathcal{A}\setminus\mathcal{C}} \beta_D$$

  to

$$\begin{aligned}\Gamma_B(\mathcal{H}_{\mathcal{A},n}) &= \Gamma_B(\{n_A, A \in \mathcal{C}\})\\ &= \cup_{A\in\mathcal{C}}\Gamma_B(n_A)\\ &= \cup_{A\in\mathcal{C}}(B \cup A)\\ &= B \cup (\cup_{A\in\mathcal{C}} A),\end{aligned}$$

which completes the proof. □

In other words, CD can also be viewed as a single correction under meta-knowledge $m^{\mathcal{H}}_{\mathcal{A},\cup}$ obtained by combining disjunctively the pieces of meta-knowledge $m^{\mathcal{H}}_{A,\cup}$, $A \in \mathcal{A}$.

A similar proposition (Proposition 18) to Proposition 17 exists for CR (although it is not its strict counterpart). It relies on the following technical lemma, which puts forward a particular subset of $\mathcal{H}$, denoted by P_A, that induces the same transformation to a testimony $\mathbf{x} \in B$ as contextual lie p_A.

Lemma 1.

$$\Gamma_B(p_A) = \Gamma_B(P_A), \quad \forall A \subseteq \mathcal{X}, \quad \forall B \subseteq \mathcal{X},$$

with P_A defined by $P_A = \{p_x | x \in A\} \cup p_\emptyset$.

Proof. We have

$$\begin{aligned}\Gamma_B(P_A) &= \Gamma_B(p_\emptyset)\bigcup_{x\in A}\Gamma_B(p_x) = (B\cap\emptyset)\bigcup_{x\in A}(B\cap x)\\ &= (B\cap\emptyset)\bigcup_{x\in A}(B\cap x) = \bigcup_{x\in A}(B\cap x)\\ &= B\cap(\bigcup_{x\in A} x) = B\cap A = \Gamma_B(p_A). \qquad \square\end{aligned}$$

Proposition 18.

$$m_S \textcircled{\cap}_{A\in\mathcal{A}} A^{\beta_A} = f_{m^{\mathcal{H}}_{\mathcal{A},\cap}}(m_S) \tag{58}$$

with $m^{\mathcal{H}}_{\mathcal{A},\cap} = \textcircled{\cap}_{A\in\mathcal{A}} m^{\mathcal{H}}_{A,P,\cap}$, where

$$m^{\mathcal{H}}_{A,P,\cap}(P_{\mathcal{X}}) = \beta_A, \quad m^{\mathcal{H}}_{A,P,\cap}(P_A) = 1-\beta_A,$$

and $P_A = \{p_x | x \in A\} \cup p_\emptyset$, for all $A \subseteq \mathcal{X}$.

Proof. First, let us study the left side of (58). From the definition of $\textcircled{\cap}$, we have that, for all $B \subseteq \mathcal{X}$, the quantity $m_S(B) \cdot (\textcircled{\cap}_{A\in\mathcal{A}} A^{\beta_A})(C)$, with C a focal set of $\textcircled{\cap}_{A\in\mathcal{A}} A^{\beta_A}$, is allocated to set $B \cap C$.

Moreover, from the definitions of A^{β_A} and $m^{\mathcal{H}}_{A,P,\cap}$, it is straightforward to see that

$$(\textcircled{\cap}_{A\in\mathcal{A}} A^{\beta_A})(C) = (\textcircled{\cap}_{A\in\mathcal{A}} m^{\mathcal{H}}_{A,P,\cap})(P_C)$$

holds for any focal set C of $\textcircled{\cap}_{A\in\mathcal{A}} A^{\beta_A}$.

Studying now the right side of (58). From the definition of the BBC procedure and of $m^{\mathcal{H}}_{\mathcal{A},\cap}$, the quantity

$$m_S(B) \cdot (\textcircled{\cap}_{A\in\mathcal{A}} m^{\mathcal{H}}_{A,P,\cap})(P_C) = m_S(B) \cdot (\textcircled{\cap}_{A\in\mathcal{A}} A^{\beta_A})(C)$$

is allocated to set $\Gamma_B(P_C) = \Gamma_B(p_C)$ (using Lemma 1) which is in turn equal to $B \cap C$. □

In other words, CR can be viewed as a single correction under meta-knowledge $m^{\mathcal{H}}_{\mathcal{A},\cap}$ obtained by combining conjunctively the pieces of meta-knowledge $m^{\mathcal{H}}_{A,P,\cap}$, each of them inducing the same correction as the pieces of meta-knowledge $m^{\mathcal{H}}_{A,\cap}$ (24) underlying CR.[10]

[10] This statement holds because $\Gamma_B(P_X) = \Gamma_B(p_X)$ (using Lemma 1), and state p_X corresponds to state t (Remark 3).

Appendix C. Proof of Proposition 15

A proof for Proposition 15 is provided in Appendix C.2. It uses the matrix notation for belief functions and in particular some technical results using this notation and concerning rule ⓝ, which are first recalled in Appendix C.1.

C.1. Matrix notation

Matrix calculus can be applied to belief functions in order to simplify their mathematics [37]. A MF m (and its associated functions, e.g., q) can be seen as a column vector of size $2^{|\mathcal{X}|}$, whose elements are ordered according to the so-called binary order: the ith element of the vector $\mathbf{m}$ corresponds to the set with elements indicated by 1 in the binary representation of $i-1$. For instance, let $\mathcal{X}=\{x_1,x_2,x_3,x_4\}$. The first element ($i=1$) of the vector $\mathbf{m}$ corresponds to $\emptyset$ since the binary representation of $1-1$ is 0000. The twelfth element ($i=12$) corresponds to $\{x_1,x_2,x_4\}$ since the binary representation of $12-1$ is 1011.

Let us denote by $\mathbf{Kron}(\mathbf{A},\mathbf{B})$ the $mp\times nq$ matrix resulting from the Kronecker product of an $m\times n$ matrix $\mathbf{A}$ with a $p\times q$ matrix $\mathbf{B}$. The matrix $\mathbf{Kron}(\mathbf{A},\mathbf{B})$ is defined by:

$$\mathbf{Kron}(\mathbf{A},\mathbf{B})=\begin{bmatrix} A(1,1)\mathbf{B} & \cdots & A(1,n)\mathbf{B} \\ \vdots & \ddots & \vdots \\ A(m,1)\mathbf{B} & \cdots & A(m,n)\mathbf{B} \end{bmatrix}.$$

The transformation of a MF m into its associated commonality function q admits a simple expression using matrix notation. We have:

$$\mathbf{q}=\mathbf{Q}\cdot\mathbf{m},$$

with $\mathbf{Q}$ a matrix that can be obtained in a simple way using Kronecker multiplication, from the building block $\begin{bmatrix}1&1\\0&1\end{bmatrix}$:

$$\mathbf{Q}^{i+1}=\mathbf{Kron}\left(\begin{bmatrix}1&1\\0&1\end{bmatrix},\ \mathbf{Q}^{i}\right),\quad \mathbf{Q}^{1}=1,$$

where $\mathbf{Q}^{i+1}$ denotes the matrix $\mathbf{Q}$ when $|\mathcal{X}|=i$. On the other hand, m can be recovered from q as follows:

$$\mathbf{m}=\mathbf{Q}^{-1}\cdot\mathbf{q}.$$

As mentioned in Section 2.1.2, combination by the rule ⓝ can be expressed similarly as combination by ∩ and ∪, that is by a simple pointwise product expression, as shown by Smets [35,37]. The counterpart of the commonality and implicability functions on which the pointwise product expression of ⓝ is based, is called 0-commonality in [26,24]. Let $\underline{q}$ denote the 0-commonality function associated to a MF m. We have, for any two MFs m_1 and m_2 [35,37]:

$$\underline{q}_{1\textcircled{\cap}2}(A)=\underline{q}_1(A)\cdot\underline{q}_2(A),\quad \forall A\subseteq\mathcal{X}.$$

Smets [35,37] showed that function $\underline{q}$ can be obtained as follows:

$$\underline{\mathbf{q}}=\underline{\mathbf{Q}}\cdot\mathbf{m},$$

with $\underline{\mathbf{Q}}$ a matrix, which as shown by Pichon and Denœux [26,24] can easily be obtained by Kronecker multiplication, similarly as $\mathbf{Q}$, but using building block $\begin{bmatrix}1&1\\-1&1\end{bmatrix}$.

In the sequel, we will also denote by $\underline{\mathbf{B}}$ the matrix obtained by Kronecker multiplication, similarly as $\mathbf{Q}$, but using building block $\begin{bmatrix}1&-1\\1&1\end{bmatrix}$. This matrix plays only a technical role in the proof that follows.

C.2. Plausibility on singletons after CN

The proof of Proposition 15 requires the following technical Lemmas 2, 3, 4, 5 and 6.

Lemma 2.

$$\underline{Q}(B,A)=(-1)^{|\bar{A}\cap B|},\quad \forall A,B\subseteq\mathcal{X}. \tag{59}$$

Proof. From [37, page 26 (**Q** corresponding to **G** with $\alpha = 0$)], column A of matrix $\mathbf{Q}$ is $\mathbf{V}_A \cdot \mathbf{1}$ ($\mathbf{1}$ denotes the column vector which components are 1), with $\mathbf{V}_A$ a matrix defined by $\mathbf{V}_A = \prod_{x \notin A} \mathbf{V}_{\bar{x}}$, where $\mathbf{V}_{\bar{x}} = [V_{\bar{x}}(A, B)]$, $\forall x \in \mathcal{X}$, $\forall A, B \subseteq \mathcal{X}$, with:

$$V_{\bar{x}}(A,B) = \begin{cases} 1 & \text{if } x \notin A, \quad A = B, \\ -1 & \text{if } x \in A, \quad A = B, \\ 0 & \text{if } A \neq B. \end{cases} \tag{60}$$

Matrices $\mathbf{V}_{\bar{x}}$ are diagonal, hence we have for all $A, B \subseteq \mathcal{X}$:

$$\begin{aligned} \underline{Q}(B,A) &= (\mathbf{V}_A \cdot \mathbf{1})(B) = V_A(B,B) = \prod_{x \notin A} V_{\bar{x}}(B,B) \\ &= \left(\prod_{x \notin A, x \in B} V_{\bar{x}}(B,B) \right) \cdot \left(\prod_{x \notin A, x \notin B} V_{\bar{x}}(B,B) \right) \\ &= \prod_{x \notin A, x \in B} V_{\bar{x}}(B,B) = \prod_{x \in \bar{A} \cap B} V_{\bar{x}}(B,B) = (-1)^{|\bar{A} \cap B|}. \end{aligned}$$ □

Lemma 3. *For all $A \subset \mathcal{X}$, the 0-commonality function $\underline{q}_A$ associated to the simple MF A^{β_A}, $\beta_A \in [0,1]$, is defined for all $B \subseteq \mathcal{X}$ by:*

$$\underline{q}_A(B) = \begin{cases} 1 & \text{if } |\bar{A} \cap B| \text{ is even,} \\ 2 \cdot \beta_A - 1 & \text{otherwise.} \end{cases} \tag{61}$$

Proof.

$$\begin{aligned} \underline{q}_A(B) &= \sum_{C \subseteq \mathcal{X}} \underline{Q}(B,C) \cdot (A^{\beta_A})(C) = \underline{Q}(B,A) \cdot (1 - \beta_A) + \underline{Q}(B,\mathcal{X}) \cdot \beta_A \\ &= (-1)^{|\bar{A} \cap B|} \cdot (1 - \beta_A) + \beta_A \quad \text{(using Lemma 2).} \end{aligned}$$ □

Lemma 4.

$$\underline{\mathbf{Q}}^{-1} = 0.5^{|\mathcal{X}|} \cdot \underline{\mathbf{B}}. \tag{62}$$

Proof. From [24, Corollary 6.1], $\underline{\mathbf{Q}}^{-1}$ may be obtained by Kronecker multiplication using the building block $0.5 \cdot \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. The lemma follows from property $\mathbf{Kron}(k \cdot \mathbf{A}, \mathbf{B}) = \mathbf{Kron}(\mathbf{A}, k \cdot \mathbf{B}) = k \cdot \mathbf{Kron}(\mathbf{A}, \mathbf{B})$, k scalar, of Kronecker multiplication and the fact that $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ is the building block of $\underline{\mathbf{B}}$. □

Lemma 5.

$$\underline{B}(B,A) = (-1)^{|A \cap \bar{B}|}, \quad \forall A, B \subseteq \mathcal{X}. \tag{63}$$

Proof. Using [24, Proposition 6.7], we have $\underline{\mathbf{Q}} = \mathbf{J} \cdot \underline{\mathbf{B}} \cdot \mathbf{J}$ with $\mathbf{J}$ the square matrix, the elements of which are zeros except those on the secondary diagonal which are ones [37]. Placed before a matrix $\mathbf{M}$, matrix $\mathbf{J}$ inverses the rows of $\mathbf{M}$, which implies that $(\mathbf{J} \cdot \mathbf{M})(\bar{B}, A) = \mathbf{M}(B, A)$, $\forall A, B \subseteq \mathcal{X}$. Placed after a matrix $\mathbf{M}$, $\mathbf{J}$ inverses the columns of $\mathbf{M}$ which yields to $(\mathbf{M} \cdot \mathbf{J})(B, \bar{A}) = \mathbf{M}(B, A)$, $\forall A, B \subseteq \mathcal{X}$. Thus: $\underline{Q}(\bar{B}, \bar{A}) = \underline{B}(B, A)$, $\forall A, B \subseteq \mathcal{X}$, which, using Lemma 2, gives (63). □

Lemma 6. *For all $x \in \mathcal{X}$, for all $C \subseteq \mathcal{X}$ such that $C \neq \emptyset$ and $C \neq \{x\}$:*

$$\sum_{B \subseteq \overline{\{x\}}} (-1)^{|C \cap B|} = 0. \tag{64}$$

Proof. Let $x \in \mathcal{X}$, $\mathcal{X}^\star = \mathcal{X} \setminus \{x\} = \overline{\{x\}}$ and $C^\star = C \setminus \{x\}$, $\forall C \subseteq \mathcal{X}$. We have $B \cap C = B \cap C^\star \subseteq \mathcal{X}^\star$, $\forall C \subseteq \mathcal{X}$, $\forall B \subseteq \mathcal{X}^\star$. Thus, $\forall C \subseteq \mathcal{X}$ s.t. $C \neq \emptyset$ and $C \neq \{x\}$, $\sum_{B \subseteq \overline{\{x\}}} (-1)^{|C \cap B|}$ is equal to $\sum_{B \subseteq \mathcal{X}^\star} (-1)^{|C^\star \cap B|}$ with $C^\star \subseteq \mathcal{X}^\star$ and $C^\star \neq \emptyset$. Let $m = |C^\star|$, $n = |\overline{C^\star}|$, $\mathcal{P}_{even} = \{B \subseteq \mathcal{X}^\star, |C^\star \cap B| \text{ is even}\}$, $\mathcal{P}_{odd} = \{B \subseteq \mathcal{X}^\star, |C^\star \cap B| \text{ is odd}\}$. Let us recall that there are $\binom{m}{k} 2^n$ subsets of $\mathcal{X}^\star$ with k elements in $C^\star$. We then have: $\sum_{B \subseteq \mathcal{X}^\star} (-1)^{|C^\star \cap B|} = |\mathcal{P}_{even}| - |\mathcal{P}_{odd}| = \sum_{k \text{ even}} \binom{m}{k} 2^n - \sum_{k \text{ odd}} \binom{m}{k} 2^n = 2^n \sum_{k=0}^{m} \binom{m}{k} (-1)^k = 0$ (Binomial theorem). □

Proposition 15 can then be proved as follows.

Proof. Let pl and $\underline{q}$ be, respectively, the plausibility and 0-commonality functions associated to MF m defined by $m = m_S \textcircled{\cap}_{A\in\mathcal{A}} A^{\beta_A}$ with $\beta_A \in [0,1]$, for all $A \in \mathcal{A}$.

For all $x \in \mathcal{X}$, we then have:

$$\begin{aligned}
pl(\{x\}) &= \sum_{A\cap\{x\}\neq\emptyset} m(A) = \sum_{x\in A} (\underline{\mathbf{Q}}^{-1}\cdot\underline{\mathbf{q}})(A) = \sum_{x\in A} (0.5^{|\mathcal{X}|}\cdot\underline{\mathbf{B}}\cdot\underline{\mathbf{q}})(A) \quad \text{(using Lemma 4)}\\
&= 0.5^{|\mathcal{X}|}\cdot\sum_{x\in A}(\underline{\mathbf{B}}\cdot\underline{\mathbf{q}})(A) = 0.5^{|\mathcal{X}|}\cdot\sum_{x\in A}\sum_{C}\underline{B}(A,C)\cdot\underline{q}(C)\\
&= 0.5^{|\mathcal{X}|}\cdot\sum_{C}\underline{q}(C)\sum_{x\in A}\underline{B}(A,C) = 0.5^{|\mathcal{X}|}\cdot\sum_{C}\underline{q}(C)\sum_{x\in A}(-1)^{|C\cap\overline{A}|} \quad \text{(using Lemma 5)}\\
&= 0.5^{|\mathcal{X}|}\cdot\sum_{C}\underline{q}(C)\sum_{B\subseteq\overline{\{x\}}}(-1)^{|C\cap B|}\\
&= 0.5^{|\mathcal{X}|}\cdot\left(\underline{q}(\emptyset)\sum_{B\subseteq\overline{\{x\}}}(-1)^{|\emptyset\cap B|} + \underline{q}(\{x\})\sum_{B\subseteq\overline{\{x\}}}(-1)^{|\{x\}\cap B|} + \sum_{C\neq\emptyset,\{x\}}\underline{q}(C)\sum_{B\subseteq\overline{\{x\}}}(-1)^{|C\cap B|}\right).
\end{aligned}$$

Since there are $2^{|\mathcal{X}|-1}$ subsets of $\overline{\{x\}}$, this last equation becomes:

$$\begin{aligned}
&0.5^{|\mathcal{X}|}\cdot\left(2^{|\mathcal{X}|-1}\cdot\underline{q}(\emptyset) + 2^{|\mathcal{X}|-1}\cdot\underline{q}(\{x\}) + \sum_{C\neq\emptyset,\{x\}}\underline{q}(C)\sum_{B\subseteq\overline{\{x\}}}(-1)^{|C\cap B|}\right)\\
&\quad = 0.5\cdot\underline{q}(\emptyset) + 0.5\cdot\underline{q}(\{x\}) + 0.5^{|\mathcal{X}|}\cdot\sum_{C\neq\emptyset,\{x\}}\underline{q}(C)\sum_{B\subseteq\overline{\{x\}}}(-1)^{|C\cap B|}. \qquad (65)
\end{aligned}$$

Using [24, Proposition 6.5], which tells us that $\underline{q}(\emptyset) = 1$, and Lemma 6, Equation (65) reduces to:

$$pl(\{x\}) = 0.5 + 0.5\cdot\underline{q}(\{x\}).$$

Besides, using Lemma 3 and the definition of $\underline{q}$:

$$\underline{q}(B) = \underline{q}_S(B)\cdot\prod_{A\in\mathcal{A},\,|\overline{A}\cap B| \text{ is odd}}(2\cdot\beta_A - 1), \quad \text{for all } B\subseteq\mathcal{X}.$$

Thus, for all $x \in \mathcal{X}$:

$$\underline{q}(\{x\}) = \underline{q}_S(\{x\})\cdot\prod_{A\in\mathcal{A},\,|\overline{A}\cap\{x\}|\text{is odd}}(2\cdot\beta_A - 1) = \underline{q}_S(\{x\})\cdot\prod_{A\in\mathcal{A},A\subseteq\overline{\{x\}}}(2\cdot\beta_A - 1).$$

At last:

$$\begin{aligned}
\underline{q}_S(\{x\}) &= (\underline{\mathbf{Q}}\cdot\mathbf{m}_S)(\{x\}) = \sum_{A\subseteq\mathcal{X}}\underline{Q}(x,A)\cdot m_S(A)\\
&= \sum_{A\cap\{x\}\neq\emptyset}\underline{Q}(x,A)\cdot m_S(A) + \sum_{A\subseteq\overline{\{x\}}}\underline{Q}(x,A)\cdot m_S(A)\\
&= \sum_{A\cap\{x\}\neq\emptyset}(-1)^{|\overline{A}\cap\{x\}|}\cdot m_S(A) + \sum_{A\subseteq\overline{\{x\}}}(-1)^{|\overline{A}\cap\{x\}|}\cdot m_S(A) \quad \text{(using Lemma 2)}\\
&= \sum_{A\cap\{x\}\neq\emptyset} m_S(A) - \sum_{A\subseteq\overline{\{x\}}} m_S(A)\\
&= pl_S(\{x\}) - b_S(\overline{\{x\}})\\
&= 2\cdot pl_S(\{x\}) - 1,
\end{aligned}$$

which completes the proof. □

References

[1] M. Bou Farah, D. Mercier, E. Lefèvre, F. Delmotte, A high-level application using belief functions for exchanging and managing uncertain events on the road in vehicular ad hoc networks, Ann. Télécommun. 69 (3–4) (2014) 185–199.
[2] B.R. Cobb, P.P. Shenoy, On the plausibility transformation method for translating belief function models to probability models, Int. J. Approx. Reason. 41 (3) (April 2006) 314–330.
[3] I. Couso, D. Dubois, Statistical reasoning with set-valued information: ontic vs. epistemic views, Int. J. Approx. Reason. 55 (7) (2014) 1502–1518.
[4] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.
[5] T. Denœux, A *k*-nearest neighbor classification rule based on Dempster–Shafer theory, IEEE Trans. Syst. Man Cybern. 25 (5) (1995) 804–813.
[6] T. Denœux, Conjunctive and disjunctive combination of belief functions induced by non-distinct bodies of evidence, Artif. Intell. 172 (2008) 234–264.
[7] T. Denœux, P. Smets, Classification using belief functions: relationship between case-based and model-based approaches, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 36 (6) (2006) 1395–1406.
[8] S. Destercke, T. Burger, Toward an axiomatic definition of conflict between belief functions, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 43 (2) (2013) 585–596.
[9] D. Dubois, H. Prade, A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets, Int. J. Gen. Syst. 12 (3) (1986) 193–226.
[10] D. Dubois, H. Prade, P. Smets, A definition of subjective possibility, Int. J. Approx. Reason. 48 (2) (2008) 352–364.
[11] Z. Elouedi, E. Lefèvre, D. Mercier, Discountings of a belief function using a confusion matrix, in: 22nd IEEE International Conference on Tools with Artificial Intelligence, ICTAI, vol. 1, Oct. 2010, pp. 287–294.
[12] Z. Elouedi, K. Mellouli, P. Smets, Assessing sensor reliability for multisensor data fusion within the transferable belief model, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 34 (1) (2004) 782–787.
[13] R.D. Huddleston, G.K. Pullum, A Student's Introduction to English Grammar, Cambridge University Press, 2005.
[14] L. Jiao, Q. Pan, T. Denœux, Y. Liang, X. Feng, Belief rule-based classification system: extension of FRBCS in belief functions framework, Inf. Sci. 309 (2015) 26–49.
[15] A.-L. Jousselme, P. Maupin, Distances in evidence theory: comprehensive survey and generalizations, Int. J. Approx. Reason. 53 (2) (2012) 118–145.
[16] J. Klein, O. Colot, Automatic discounting rate computation using a dissent criterion, in: Proceedings of the 1st Workshop on the Theory of Belief Functions, BELIEF, 2010, pp. 1–6.
[17] E. Lefèvre, F. Pichon, D. Mercier, Z. Elouedi, B. Quost, Estimation de sincérité et pertinence à partir de matrices de confusion pour la correction de fonctions de croyance, in: Rencontres Francophones sur la Logique Floue et ses Applications, LFA, vol. 1, Cépaduès, Nov. 2014, pp. 287–294.
[18] M. Loudahi, J. Klein, J.-M. Vannobel, O. Colot, New distances between bodies of evidence based on dempsterian specialization matrices and their consistency with the conjunctive combination rule, Int. J. Approx. Reason. 55 (5) (2014) 1093–1112.
[19] D. Mercier, G. Cron, T. Denœux, M.-H. Masson, Decision fusion for postal address recognition using belief functions, Expert Syst. Appl. 36 (3) (2009) 5643–5653.
[20] D. Mercier, E. Lefèvre, F. Delmotte, Belief functions contextual discounting and canonical decompositions, Int. J. Approx. Reason. 53 (2) (2012) 146–158.
[21] D. Mercier, F. Pichon, E. Lefèvre, Corrigendum to "Belief functions contextual discounting and canonical decompositions" [International Journal of Approximate Reasoning 53 (2012) 146–158], Int. J. Approx. Reason. 70 (2016) 137–139.
[22] D. Mercier, F. Pichon, E. Lefèvre, F. Delmotte, Learning contextual discounting and contextual reinforcement from labelled data, in: S. Destercke, T. Denœux (Eds.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, in: Lecture Notes in Computer Science, vol. 9161, Springer International Publishing, 2015, pp. 472–481.
[23] D. Mercier, B. Quost, T. Denœux, Refined modeling of sensor reliability in the belief function framework using contextual discounting, Inf. Fusion 9 (2) (2008) 246–258.
[24] F. Pichon, Belief functions: canonical decompositions and combination rules, PhD thesis, Université de Technologie de Compiègne, Compiègne, France, 2009.
[25] F. Pichon, On the α-conjunctions for combining belief functions, in: T. Denœux, M.-H. Masson (Eds.), Belief Functions: Theory and Applications, in: Advances in Intelligent and Soft Computing, vol. 164, Springer, Berlin/Heidelberg, 2012, pp. 285–292.
[26] F. Pichon, T. Denœux, Interpretation and computation of α-junctions for combining belief functions, in: Sixth Int. Symp. on Imprecise Probability: Theories and Applications, ISIPTA'09, Durham, United Kingdom, July 2009.
[27] F. Pichon, S. Destercke, T. Burger, A consistency–specificity trade-off to select source behavior in information fusion, IEEE Trans. Cybern. 45 (4) (2015) 598–609.
[28] F. Pichon, D. Dubois, T. Denœux, Relevance and truthfulness in information correction and fusion, Int. J. Approx. Reason. 53 (2) (2012) 159–175.
[29] F. Pichon, C. Labreuche, B. Duqueroie, T. Delavallade, Multidimensional approach to reliability evaluation of information sources, in: P. Capet, T. Delavallade (Eds.), Information Evaluation, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2014, pp. 129–156.
[30] F. Pichon, D. Mercier, F. Delmotte, E. Lefèvre, Truthfulness in contextual information correction, in: Fabio Cuzzolin (Ed.), Belief Functions: Theory and Applications, in: Lecture Notes in Computer Science, vol. 8764, Springer International Publishing, 2014, pp. 11–20.
[31] J. Schubert, Conflict management in Dempster–Shafer theory using the degree of falsity, Int. J. Approx. Reason. 52 (3) (2011) 449–460.
[32] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
[33] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, Int. J. Approx. Reason. 9 (1) (1993) 1–35.
[34] P. Smets, The canonical decomposition of a weighted belief, in: Proceedings of the 14th Int. Joint Conf. on Artificial Intelligence, IJCAI'95, San Mateo, California, USA, 1995, Morgan Kaufmann, 1995, pp. 1896–1901.
[35] P. Smets, The α-junctions: combination operators applicable to belief functions, in: D.M. Gabbay, R. Kruse, A. Nonnengart, H.J. Ohlbach (Eds.), First International Joint Conference on Qualitative and Quantitative Practical Reasoning, ECSQARU-FAPR'97, Bad Honnef, Germany, 09/06/1997–12/06/1997, in: Lecture Notes in Computer Science, vol. 1244, Springer, 1997, pp. 131–153.
[36] P. Smets, The transferable belief model for quantified belief representation, in: D.M. Gabbay, Ph. Smets (Eds.), Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, Kluwer, Dordrecht, 1998, pp. 267–301.
[37] P. Smets, The application of the matrix calculus to belief functions, Int. J. Approx. Reason. 31 (1–2) (2002) 1–30.
[38] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, Int. J. Approx. Reason. 38 (2) (2005) 133–147.
[39] P. Smets, R. Kennes, The transferable belief model, Artif. Intell. 66 (1994) 191–243.
[40] Y. Yang, D. Han, C. Han, Discounted combination of unreliable evidence using degree of disagreement, Int. J. Approx. Reason. 54 (8) (2013) 1197–1216.
[41] L.M. Zouhal, T. Denœux, An evidence-theoretic *k*-NN rule with parameter optimization, IEEE Trans. Syst. Man Cybern. 28 (2) (1998) 263–271.

Annexe C

Canonical decomposition of belief functions based on Teugels' representation of the multivariate Bernoulli distribution

---

Article paru dans *Information Sciences*, 428 : 76-104, 2018.

Information Sciences 428 (2018) 76–104

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Canonical decomposition of belief functions based on Teugels' representation of the multivariate Bernoulli distribution

Frédéric Pichon

Univ. Artois, EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), Béthune, F-62400, France

ARTICLE INFO

Article history:
Received 6 December 2016
Revised 12 August 2017
Accepted 13 October 2017
Available online 16 October 2017

Keywords:
Dempster-Shafer theory
Belief function
Canonical decomposition
Multivariate Bernoulli distribution
Moment
Mutual information
Random set.

ABSTRACT

A canonical decomposition of belief functions is a unique decomposition of belief functions into elementary pieces of evidence. Smets found an equivalent representation of belief functions, which he interpreted as a canonical decomposition. However, his proposal is not entirely satisfactory as it involves elementary pieces of evidence, corresponding to a generalisation of belief function axioms, whose semantics lacks formal justifications. In this paper, a new canonical decomposition relying only on well-defined concepts is proposed. In particular, it is based on a means to induce belief functions from the multivariate Bernoulli distribution and on Teugels' representation of this distribution, which consists of the means and the central moments of the underlying Bernoulli random variables. According to our decomposition, a belief function results from as many crisp pieces of information as there are elements in its domain, and from simple probabilistic knowledge concerning their marginal reliability and the dependencies between their reliability. In addition, we show that instead of interpreting with some difficulty Smets' representation of belief functions as a canonical decomposition, it is possible to give it a different and well-defined semantics in terms of measures of information associated with the reliability of the pieces of information in our decomposition.

1. Introduction

A belief function [1,2] is a rich mathematical object for representing uncertain information on the actual value taken by a variable. Belief functions have been successfully applied to numerous problems such as, recently, clustering of large dissimilarity data [3], multi-label classification [4], object association [5,6], and rule-based classification [7]; we refer the reader to [8] and the references therein for a much larger account of belief function applications.

Belief functions were originally introduced by Shafer [1], as a formal object for the representation of evidence. An important result shown in [1] is that the so-called separable belief functions can be canonically decomposed, that is, decomposed uniquely into elementary pieces of evidence. Smets [9] extended this result and proposed a solution to canonically decompose any belief function. The concept of canonical decomposition is important at a fundamental level in that it lays bare the underlying elementary components of a complex belief state. It also raises the interesting question as to how can one accumulate evidence progressively to construct probability judgements on the value taken by a variable of interest. Moreover, it can be useful to tackle several problems, as exemplified in recent years where proposals based on Smets' decomposition have appeared to address the issues of belief function combination [10–12], correction [13–15] and clustering [16].

E-mail address: frederic.pichon@univ-artois.fr

https://doi.org/10.1016/j.ins.2017.10.030
0020-0255/

In this paper, this concept is deeply revisited. First, a critical review of Smets' canonical decomposition is conducted (Section 2): we challenge his solution and argue that it is not entirely satisfactory. Then, a new canonical decomposition of belief functions is proposed (Section 3). This new decomposition is in the same spirit as Smets' – a belief function is viewed as resulting from partially reliable sources providing crisp pieces of information –, but it follows a completely different approach based on some results concerning the representation of the multivariate Bernoulli distribution [17]. Next, some comments on our solution are provided and, in particular, our solution is compared to Smets' and is considered in the context of random sets [18,19] (Section 4). Finally, a new perspective on Smets' decomposition is brought to light using measures of information [20] associated with the multivariate Bernoulli distribution (Section 5). Section 6 concludes the paper.

2. Review of Smets' canonical decomposition

In this section, necessary background on belief functions and on their canonical decomposition as proposed by Smets is provided. Then, a general information fusion approach based on explicit assumptions about the reliability of information sources is recalled and used for critical examination of Smets' canonical decomposition.

2.1. Basics of belief function theory

Belief function theory [1,2] is a framework for uncertainty modelling and reasoning. In this theory, uncertainty regarding the actual value taken by a variable $\boldsymbol{y}$ defined on a finite domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$, is represented by a so-called *mass* function (MF) defined as a mapping $m : 2^{\mathcal{Y}} \to [0, 1]$ satisfying $\sum_{A \subseteq \mathcal{Y}} m(A) = 1$. The quantity $m(A)$ may be interpreted as the probability of knowing only that $\boldsymbol{y} \in A$, $A \subseteq \mathcal{Y}$ [21]. Subsets A of $\mathcal{Y}$ such that $m(A) > 0$ are called *focal sets* of m. A mass function is called: *dogmatic* if $\mathcal{Y}$ is not a focal set; *normal* if $\emptyset$ is not a focal set; *vacuous* if $\mathcal{Y}$ is its only focal set; *simple* if it has at most two focal sets, and if it has two, $\mathcal{Y}$ is one of those; *Bayesian* if its focal sets are singletons. A non normal MF m can be normalised, *i.e.*, transformed into a normal MF m^*, by the following operation:

$$m^*(A) = \begin{cases} \kappa \cdot m(A) & \text{for all } A \subseteq \mathcal{Y}, A \neq \emptyset, \\ 0 & \text{if } A = \emptyset, \end{cases} \tag{1}$$

with $\kappa = (1 - m(\emptyset))^{-1}$.

The *belief* function *bel* is an equivalent representation of a mass function m. It is defined as

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B), \quad \forall A \subseteq \mathcal{Y}.$$

The degree of belief $bel(A)$ evaluates to what extent event A is logically implied by the available evidence [22]. Other equivalent representations of m that are of interest for this paper are the *plausibility* pl and *commonality* q functions:

$$\begin{aligned} pl(A) &= \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \subseteq \mathcal{Y}, \\ q(A) &= \sum_{B \supseteq A} m(B), \quad \forall A \subseteq \mathcal{Y}. \end{aligned} \tag{2}$$

The degree of plausibility $pl(A)$ evaluates to what extent event A is consistent with the available evidence [22]. The commonality function has a technical role, which will be described later. Functions m, *bel*, pl and q are in one-to-one correspondence, in particular mass function m can be recovered from any of these functions. For instance, we have:

$$m(A) = \sum_{B \supseteq A} (-1)^{|B|-|A|} q(B), \quad \forall A \subseteq \mathcal{Y}, \tag{3}$$

with $|A|$ denoting the cardinality of $A \subseteq \mathcal{Y}$. Let us note that the plausibility function restricted to the singletons of $\mathcal{Y}$ is called the *contour* function. Besides, the plausibility of any singleton of $\mathcal{Y}$ is equal to its commonality, *i.e.*, $pl(\{y\}) = q(\{y\})$ for all $y \in \mathcal{Y}$.

Matrix calculus is useful to simplify the mathematics of belief function theory [23]. A mass function m and its associated functions, such as q, can be seen as column vectors of size $2^{|\mathcal{Y}|} = 2^n$, whose elements are ordered according to the so-called binary order detailed hereafter. Let k be an integer such that $1 \leq k \leq 2^n$. k can be written in a binary expansion, *i.e.*,

$$k = 1 + \sum_{i=1}^{n} k_i 2^{i-1}, \tag{4}$$

where $k_i \in \{0, 1\}$. Expansion (4) induces a one-to-one correspondence $k \leftrightarrow (k_1, \ldots, k_n)$, that is Eq. (4) associates to each integer k, $1 \leq k \leq 2^n$, a binary vector $(k_1, \ldots, k_n) \in \{0, 1\}^n$. Let A_k, $1 \leq k \leq 2^n$, be the subset of $\mathcal{Y}$, such that $y_i \in A_k$ if $k_i = 1$ and $y_i \notin A_k$ if $k_i = 0$, with k_i, $i = 1, \ldots, n$, the terms in the binary expansion (4) of k. In the binary order, the kth element of the vector $\mathbf{m}$ corresponds to the set A_k. Thus $\emptyset$ is the first element, $\{y_1\}$ the second element, $\{y_2\}$ the third element,

$\{y_1, y_2\}$ the fourth element, etc. For instance, for $n = 4$, $A_{14} = \{y_1, y_3, y_4\}$ since $k = 14$ is in one-to-one correspondence with $(k_1 = 1, k_2 = 0, k_3 = 1, k_4 = 1)$. Eqs. (2) and (3) become in matrix form [23]:

$$\mathbf{q} = \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\right)\mathbf{m}, \qquad \mathbf{m} = \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}\right)\mathbf{q}, \tag{5}$$

where $\otimes$ denotes Kronecker product.

Several so-called belief function combination rules have been proposed to combine multiple pieces of information about a variable [24]. The most classical ones are Dempster's rule of combination [1,25] and its unnormalised version called conjunctive rule [24]. We denote the former by $\oplus$ and the latter by $\textcircled{\cap}$. They are defined as follows. Let m_1 and m_2 be two mass functions representing uncertainty about a variable $\boldsymbol{y}$. Let $m_{1\oplus 2} = m_1 \oplus m_2$ and $m_{1\textcircled{\cap} 2} = m_1 \textcircled{\cap} m_2$ denote the mass functions resulting from the combination of m_1 and m_2 by $\oplus$ and by $\textcircled{\cap}$ respectively. We have

$$m_{1\textcircled{\cap} 2}(A) = \sum_{B\cap C = A} m_1(B) m_2(C), \quad \forall A \subseteq \mathcal{Y}, \tag{6}$$

and, assuming that $m_{1\textcircled{\cap} 2}(\emptyset) \neq 1$, $m_{1\oplus 2} = m^*_{1\textcircled{\cap} 2}$, *i.e.*, combination by Dempster's rule amounts to combination by the conjunctive rule (6) followed by normalisation (1). Both rules are commutative, associative and admit the vacuous mass function as only neutral element. These rules are appropriate when m_1 and m_2 represent independent bodies of evidence, and the conjunctive rule is sensible under the so-called open world assumption, whereas Dempter's rule corresponds to the closed world assumption [24]. The conjunctive rule has a simple expression using the commonality function:

$$q_{1\textcircled{\cap} 2}(A) = q_1(A) \cdot q_2(A), \quad \forall A \subseteq \mathcal{Y}, \tag{7}$$

with q_1, q_2 and $q_{1\textcircled{\cap} 2}$ the commonality functions associated to m_1, m_2 and $m_{1\textcircled{\cap} 2}$, respectively.

The inverse of the conjunctive rule, denoted by $\textcircled{\emptyset}$, can be defined [9,10]. It is the rule which restores m_1 from $m_{1\textcircled{\cap} 2}$ and m_2, *i.e.*, $m_{1\textcircled{\cap} 2} \textcircled{\emptyset} m_2 = m_1$. It is sensible if we learn that m_2 is actually not supported by evidence and should thus be removed from $m_{1\textcircled{\cap} 2}$. Let q_1 and q_2 be the commonality functions associated respectively to any two mass functions m_1 and m_2, the inverse of the conjunctive rule is defined as:

$$q_{1\textcircled{\emptyset} 2}(A) = \frac{q_1(A)}{q_2(A)}, \quad \forall A \subseteq \mathcal{Y}. \tag{8}$$

This operation is well-defined as long as (i) $m_{1\textcircled{\emptyset} 2}$ is a mass function, which is not necessarily the case since the quotient of two commonality functions is not always a commonality function, and (ii) m_2 is non dogmatic, which ensures $q_2(A) > 0$ for all $A \subseteq \mathcal{Y}$.

2.2. Smets' canonical decomposition of belief functions

Some mass functions can be obtained as the result of the combination by Dempster's rule of simple mass functions; they are called *separable* by Shafer [1]. A simple mass function having two focal sets $A \subset \mathcal{Y}$ and $\mathcal{Y}$, with respective masses $1 - w$ and w, $w \in [0, 1]$, may be simply denoted A^w. For every separable mass function m, one has

$$m = \bigoplus_{\emptyset \neq A \subset \mathcal{Y}} A^{w(A)}, \tag{9}$$

with $w(A) \in [0, 1]$ for all $A \subset \mathcal{Y}, A \neq \emptyset$. This *canonical decomposition* of m is unique if m is non dogmatic (in which case $w(A) > 0$ for all $A \subset \mathcal{Y}, A \neq \emptyset$).

Smets [9] extended the concept of separability to the more general case of non normal mass functions. This extended concept is called u-separability in [10] (u stands for unnormalised) and relies on the unnormalised version of Dempster's rule, *i.e.*, the conjunctive rule. A mass function m is thus u-separable if it can be expressed as:

$$m = \bigcirc\!\!\!\!\!\!\cap_{A \subset \mathcal{Y}} A^{w(A)}, \tag{10}$$

with $w(A) \in [0, 1]$ for all $A \subset \mathcal{Y}$. This decomposition is also unique if m is non dogmatic.

Smets [9] extended this latter decomposition to all non dogmatic mass functions. The decomposition of non dogmatic mass functions relies on the concept of *generalised* simple mass function defined as a function $\mu : 2^{\mathcal{Y}} \to \mathbb{R}$ verifying:

$$\begin{aligned} &\mu(A) = 1 - w, \\ &\mu(\mathcal{Y}) = w, \\ &\mu(B) = 0, \quad \forall B \in 2^{\mathcal{Y}} \backslash \{A, \mathcal{Y}\}, \end{aligned}$$

for some $A \neq \mathcal{Y}$ and some $w \in [0, +\infty)$. Any generalised simple mass function μ can be noted A^w for some $A \neq \mathcal{Y}$ and $w \in [0, +\infty)$. When $w \leq 1$, A^w is a simple mass function. When $w > 1$, A^w is called *inverse* simple mass function; Smets [9] proposed to interpret such function as reasons *not* to believe in A (he also speaks of "debt of belief" in A) since combining $A^{1/w}$ with A^w by $\textcircled{\cap}$ (using a trivial extension of the conjunctive rule to combine generalised simple mass functions) yields the vacuous mass function, that is, the reasons to believe in A represented by $A^{1/w}$ are counter-balanced by A^w.

Using the concept of generalised simple mass function, Smets showed that for any non dogmatic mass function m, we have:

$$m = \bigcircledcap_{A \subset \mathcal{Y}} A^{w(A)}, \tag{11}$$

with $w(A) \in (0, +\infty)$ for all $A \subset \mathcal{Y}$. The *weight* function $w : 2^{\mathcal{Y}} \backslash \{\mathcal{Y}\} \to (0, +\infty)$ that appears in (11), is an equivalent representation of a non dogmatic mass function. It can be obtained from the commonality function as follows:

$$w(A) = \prod_{B \supseteq A} q(B)^{(-1)^{|B|-|A|+1}}, \quad \forall A \subset \mathcal{Y}. \tag{12}$$

Note that the conjunctive rule has a simple expression using the weight function:

$$w_{1\textcircled{\cap}2}(A) = w_1(A) \cdot w_2(A), \quad \forall A \subset \mathcal{Y}. \tag{13}$$

Using function w, Smets [9] further showed that any non dogmatic mass function m can be written as

$$m = m^c \textcircled{\cap} m^d, \tag{14}$$

with m^c and m^d two u-separable mass functions such that

$$m^c = \bigcircledcap_{A \subset \mathcal{Y}} A^{1 \wedge w(A)},$$

and

$$m^d = \bigcircledcap_{A \subset \mathcal{Y}} A^{1 \wedge \frac{1}{w(A)}},$$

where $\wedge$ denotes the minimum operator. In other words, Smets [9] showed with (14) that any non dogmatic mass function can be uniquely obtained from simple mass functions. The m^c and m^d mass functions in (14) are called the *confidence and diffidence components*, respectively, of m by Smets [9], who proposed to view m^c as representing "good reasons to believe" in some propositions $A \subset \mathcal{Y}$, and m^d as representing "good reasons not to believe" in some other propositions.

Smets' canonical decomposition of non dogmatic mass functions is illustrated by Example 1.

Example 1. ("Example 2, continuation" of [9]) Let $\mathcal{Y} = \{y_1, y_2, y_3\}$ be the domain of a variable $\boldsymbol{y}$ and let m be a mass function on $\mathcal{Y}$ such that

$$m(\{y_1, y_2\}) = m(\{y_1, y_3\}) = m(\{y_1, y_2, y_3\}) = 1/3. \tag{15}$$

We have

$$\begin{aligned} w(\{y_1, y_2\}) &= 1/2, \\ w(\{y_1, y_3\}) &= 1/2, \\ w(\{y_1\}) &= 4/3, \end{aligned}$$

and $w(A) = 1$ for all $A \in 2^{\mathcal{Y}} \backslash \{\mathcal{Y}, \{y_1, y_2\}, \{y_1, y_3\}, \{y_1\}\}$. Hence, using (11) and the fact that the vacuous mass function is the neutral element of $\textcircled{\cap}$ and can be noted A^1 for any $A \subset \mathcal{Y}$, we have

$$m = \{y_1, y_2\}^{1/2} \textcircled{\cap} \{y_1, y_3\}^{1/2} \textcircled{\cap} \{y_1\}^{4/3},$$

and using (14) we have

$$m = \{y_1, y_2\}^{1/2} \textcircled{\cap} \{y_1, y_3\}^{1/2} \textcircled{\cap} \{y_1\}^{3/4}.$$

According to Smets [9], m is thus the result of the following independent pieces of evidence:

- a first source provides the evidence *believe* $\{y_1, y_2\}$ and this source is given reliability 1/2;
- a second source provides the evidence *believe* $\{y_1, y_3\}$ and this source is given reliability 1/2;
- a third source provides the evidence *do not believe* $\{y_1\}$ and this source is given reliability 3/4.

Smets' decomposition is quite elegant and has sparked several contributions such as fusion [10–12], correction [13–15] and clustering schemes [16]. Nonetheless, despite its appeal and the success that it has enjoyed, we will try and argue in Section 2.4 that Smets' decomposition is not entirely satisfactory and we will provide an alternative decomposition in

Section 3. Both our critique of Smets' decomposition and our alternative to it rely in part on a general approach to information fusion, which is recalled in the next section.

2.3. Behaviour-based fusion

In [26], Pichon et al. introduced a general approach to information fusion. In this approach, an agent builds his belief about the actual value of a variable $\boldsymbol{y}$ defined on a domain $\mathcal{Y}$, given pieces of information about $\boldsymbol{y}$ provided by some sources and given his knowledge about the behaviour of these sources (called *meta-knowledge* in [26] as it is higher order knowledge that differs from the knowledge supplied by the sources). This approach is recalled hereafter in the particular case where the sources provide crisp pieces of information about $\boldsymbol{y}$ and meta-knowledge on the sources concern their reliability, as this particular case is instrumental to the present paper.

Let us consider the simple situation where there is a single source $\mathfrak{s}_1$ providing to an agent a piece of information about $\boldsymbol{y}$ and this piece of information is crisp, *i.e.*, it is of the form $\boldsymbol{y} \in A$ for some $A \subseteq \mathcal{Y}$. Besides, the agent assumes that the source can be in only one of two states: reliable or not reliable. If the source is reliable, then the agent should deduce that $\boldsymbol{y} \in A$ from the piece of information provided by $\mathfrak{s}_1$. If the source is not reliable, then the piece of information provided by $\mathfrak{s}_1$ is useless and the agent knows only that $\boldsymbol{y} \in \mathcal{Y}$, *i.e.*, he knows nothing.

Let $\mathcal{X}_1 = \{0, 1\}$ be the space denoting the reliability of $\mathfrak{s}_1$, where 0 means that $\mathfrak{s}_1$ is reliable and 1 means that $\mathfrak{s}_1$ is not reliable. The above reasoning can be encoded by multi-valued mappings Γ_A, $A \subseteq \mathcal{Y}$, from $\mathcal{X}_1$ to $\mathcal{Y}$ such that

$$\begin{aligned} &\Gamma_A(0) = A, \\ &\Gamma_A(1) = \mathcal{Y}. \end{aligned}$$

$\Gamma_A(k_1)$ interprets the testimony $\boldsymbol{y} \in A$ in each configuration $k_1 \in \mathcal{X}_1$ of the source $\mathfrak{s}_1$.

More generally, meta-knowledge on the source may be uncertain and specifically $\mathfrak{s}_1$ may be assumed to be reliable with probability $1 - \pi_1$ and not reliable with probability π_1, $\pi_1 \in [0, 1]$. In such case, if $\mathfrak{s}_1$ provides the piece of information $\boldsymbol{y} \in A$, then probability $1 - \pi_1$ will be transferred to $\Gamma_A(0)$ and probability π_1 to $\Gamma_A(1)$, *i.e.*, the knowledge of the agent about $\boldsymbol{y}$ is represented by a mass function defined as [26][1]:

$$\begin{aligned} &m(A) = 1 - \pi_1, \\ &m(\mathcal{Y}) = \pi_1. \end{aligned} \tag{16}$$

We may remark that mass function defined by (16) is a simple mass function, which may be denoted by A^{π_1}. Besides, the transformation of testimony $\boldsymbol{y} \in A$ according to (16) is nothing but the discounting operation proposed by Shafer [1] to integrate the reliability of information sources.

Let us now consider the case where there are N sources providing crisp pieces of information about $\boldsymbol{y}$ and each source can either be reliable or not reliable. Let $\mathcal{X}_i = \{0, 1\}$ be the space denoting the reliability of source $\mathfrak{s}_i$, $i = 1, \ldots, N$, where 0 means that $\mathfrak{s}_i$ is reliable and 1 means that $\mathfrak{s}_i$ is not reliable. The set of elementary joint states on the sources is therefore the Cartesian product $\mathcal{X}^N := \times_{i=1}^N \mathcal{X}_i$. To any state $(k_1, \ldots, k_N) \in \mathcal{X}^N$, we associate the number k, $1 \leq k \leq 2^N$, such that

$$k = 1 + \sum_{i=1}^{N} k_i 2^{i-1}, \tag{17}$$

i.e., we have $k \leftrightarrow (k_1, \ldots, k_N) \in \mathcal{X}^N$. Furthermore for any $k \leftrightarrow (k_1, \ldots, k_N) \in \mathcal{X}^N$, we define a mapping $\Gamma_{\mathbf{A}}$ for any $\mathbf{A} = (A_{\mathfrak{s}_1}, \ldots, A_{\mathfrak{s}_N}) \subseteq \times_{i=1}^N \mathcal{Y}$ as

$$\Gamma_{\mathbf{A}}(k) = \bigcap_{i=1}^{N} \Gamma_{A_{\mathfrak{s}_i}}(k_i).$$

$\Gamma_{\mathbf{A}}(k)$ represents the information deduced on $\mathcal{Y}$ from crisp testimonies $(A_{\mathfrak{s}_1}, \ldots, A_{\mathfrak{s}_N})$ provided by sources $\mathfrak{s}_1, \ldots, \mathfrak{s}_N$, when they are in states $(k_1, \ldots, k_N)$ [26]. We have $\Gamma_{\mathbf{A}}(k) \subseteq \mathcal{Y}$ since $\Gamma_{\mathbf{A}}(k)$ is defined as the intersection of subsets of $\mathcal{Y}$.

In the more general case where meta-knowledge on the sources is uncertain and each joint state $k \in \mathcal{X}^N$ has a probability p_k, the knowledge of the agent about $\mathcal{Y}$ is represented by a mass function defined as [26]:

$$m(B) = \sum_{k:\Gamma_{\mathbf{A}}(k)=B} p_k, \quad \forall B \subseteq \mathcal{Y}. \tag{18}$$

An important special case of uncertain meta-knowledge is when for each $k \in \mathcal{X}^N$, the probability p_k that the sources are in joint state $(k_1, \ldots, k_N)$ is equal to the product of the marginal probabilities of the individual states k_i, $i = 1, \ldots, N$, *i.e.*, the

[1] Mass function m may be induced formally as follows. $(\mathcal{X}_1, 2^{\mathcal{X}_1}, P)$ is a probability space, with P a probability measure on $\mathcal{X}_1$ such that $P(\{0\}) = 1 - \pi_1$ and $P(\{1\}) = \pi_1$, and $(\mathcal{Y}, 2^{\mathcal{Y}})$ is a measurable space. Since Γ_A is *strongly measurable* [27] with respect to $2^{\mathcal{Y}}$ and $2^{\mathcal{X}_1}$, the four-tuple $(\mathcal{X}_1, 2^{\mathcal{X}_1}, P, \Gamma_A)$ induces a belief function on $\mathcal{Y}$ (obtained by composition of P and the *lower inverse* [27] of Γ_A) with associated mass function m defined by (16) [27,28].

probabilities p_k satisfy the following property

$$p_k = \prod_{i=1}^{N} (1-\pi_i)^{1-k_i} \pi_i^{k_i}, \quad \forall k \in \mathcal{X}^N, \tag{19}$$

with k_i, $i = 1, \ldots, N$, the terms in the binary expansion (17) of k, and π_i the marginal probability that $\mathfrak{s}_i$ is not reliable, *i.e.*,

$$\pi_i = \sum_{k:k_i=1} p_k. \tag{20}$$

This property is a case of so-called *meta-independence* between the sources [26], which corresponds to the situation where pieces of meta-knowledge regarding the states of each source are independent. From [26, Theorem 1], if the probabilities p_k satisfy (19), then the mass function m given by (18) can be equivalently written as:

$$m = \bigcirc_{i=1}^{N} A_{\mathfrak{s}_i}^{\pi_i}, \tag{21}$$

with $A_{\mathfrak{s}_i}^{\pi_i}$ the simple mass function such that $m(A_{\mathfrak{s}_i}) = 1 - \pi_i$ and $m(\mathcal{Y}) = \pi_i$.

2.4. Discussion on Smets' decomposition

It is clear that the canonical decomposition (10) is an instance of (21) and thus can be given an interpretation using Pichon et al.'s scheme recalled in the preceding section. Precisely, any u-separable mass function m defined on a domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$ and with associated weight function w, can be seen as originating from the following pieces of evidence:

- there are $2^n - 1$ sources, with source $\mathfrak{s}_i$, $i = 1, \ldots, 2^n - 1$, providing the crisp piece of information $\boldsymbol{y} \in A_i$, with A_i the ith subset of $\mathcal{Y}$ according to the binary order;
- each source $\mathfrak{s}_i$ is non reliable with marginal probability $w(A_i)$;
- the sources are meta-independent.

This interpretation of the canonical decomposition (10), obtained using Pichon et al.'s scheme, is totally in line with Smets' interpretation of this decomposition, as can be recognised when comparing the above pieces of evidence to the ones at play in Example 1. The above interpretation is a bit more cumbersome due to its heavier formalisation but it has the advantage to rely on well-known concepts since it follows Dempster's original approach [25], where belief functions are induced from a space equipped with a probability measure and a multi-valued mapping from this space to another one. Overall, one may argue that both interpretation are essentially one and the same, and that Pichon et al.'s scheme merely provides a formal ground to Smets' interpretation using well-known concepts.

Due to the similarity between (10) and (11), one may wonder whether the straightforward extension of the above interpretation to non-dogmatic mass functions decomposed using (11) is possible. The answer is negative since it would require allowing marginal probabilities $w(A_i)$ greater than 1. We feel that it is a drawback of Smets' decomposition that it is incompatible with the partially reliable source model relying only on well-known concepts and leading to (21), while this formal model seems to be totally in line with his view on his decomposition.

This leads us to discuss further Smets' decomposition of non-dogmatic mass functions, irrespective of our previous comment. His decomposition involves so-called inverse simple mass functions A^w, $w > 1$, which we recall are *not* mass functions since they do not satisfy $m(A) \in [0, 1]$ for all $A \subseteq \mathcal{Y}$. Smets [9] proposed to interpret these functions as pieces of evidence of the form *a source provides the evidence* do not believe *A and this source is given reliability* $1/w$. While Smets [9] provides some intuition on why such semantics might suit such functions, it lacks an operational definition,[2] as is necessary for any uncertainty representation (see, *e.g.*, [22]). Moreover, if one is to accept the existence of this notion of "good reasons *not to believe*" and its associated mathematical representation, and one wants to model every possible state of belief involving "good reasons to believe" and "good reasons not to believe" in some propositions, then one actually needs to consider what Smets calls latent belief structures [9], which are couples of belief functions (representing respectively confidence and diffidence) that cannot be reduced in general to a single belief function. For instance, replacing the third piece of evidence in Example 1 *a third source provides the evidence* do not believe $\{y_1\}$ *and this source is given reliability* 3/4 by *a third source provides the evidence* do not believe $\{y_2\}$ *and this source is given reliability* 3/4 does not yield a mass function (one can easily check that $\{y_1, y_2\}^{1/2} \bigcirc \{y_1, y_3\}^{1/2} \bigcirc \{y_2\}^{3/4}$ is not a mass function); even modifying only slightly the third piece of evidence to *a third source provides the evidence* do not believe $\{y_1\}$ *and this source is given reliability 0.74* does not yield a mass function. A theory based on latent belief structures and involving inverse simple mass functions clearly goes beyond a theory based only on belief functions. A fully operationalised theory based on this richer mathematical object remains to be proposed.

This section has reviewed Smets' canonical decomposition and has also recalled Pichon et al.'s fusion scheme, which allows one to account explicitly for uncertain knowledge about source reliability. In the next section, we use this latter scheme to unveil a new canonical decomposition of belief functions. As will be seen, this new canonical decomposition relies only on well-defined concepts, contrary to Smets' decomposition.

[2] Their relevance has also been questioned in [29].

3. A new canonical decomposition of belief functions

In this section, a specific connection between the multivariate Bernoulli distribution and belief functions is established and used in conjunction with Pichon et al.'s fusion scheme to lay the foundations for a new canonical decomposition of belief functions. Then, Teugels' representations of the multivariate Bernoulli distribution [17] are recalled and used together with the aforementioned connection to propose a new canonical decomposition of belief functions.

3.1. Multivariate Bernoulli distribution induced belief function

In this section, one way is provided to induce from the multivariate Bernoulli distribution any belief function on domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$.

Let $\{X_i : i = 1, \ldots, n\}$ be a sequence of Bernoulli random variables with ranges $\mathcal{X}_i = \{0, 1\}$, $i = 1, \ldots, n$, *i.e.*, for $i = 1, \ldots, n$,

$$P(X_i = 1) = \pi_i, \ P(X_i = 0) = \xi_i,$$

where $0 \leq \pi_i = 1 - \xi_i \leq 1$. Recall that $\mathbb{E}[X_i] = \pi_i$.

Consider the multivariate Bernoulli distribution (MBD)

$$p_{k_1,\ldots,k_n} := P(X_1 = k_1, \ldots, X_n = k_n) \tag{22}$$

where $k_i \in \{0, 1\}$, $i = 1, \ldots, n$. Let k be an integer such that $1 \leq k \leq 2^n$. Using the binary expansion of k based on the k_i, $i = 1, \ldots, n$, that is the correspondence $k \leftrightarrow (k_1, \ldots, k_n)$, we can write

$$p_k = p_{k_1,\ldots,k_n}, \quad 1 \leq k \leq 2^n,$$

and represent a MBD p by the vector $\mathbf{p}$ containing 2^n components, with p_k its kth component, $1 \leq k \leq 2^n$. This latter convenient notation for the MBD is borrowed from [17].

Definition 1 (MBD-equivalent)**.** Let m be any mass function on $\mathcal{Y}$ and let p be the MBD (22) such that $p_k = m(A_k)$, $1 \leq k \leq 2^n$, with A_k the kth subset of $\mathcal{Y}$ according to the binary order. p is called the *MBD-equivalent* of m.

Accordingly, the vector $\mathbf{p}$ associated to the MBD-equivalent of a mass function m, is the same vector as $\mathbf{m}$.

In addition, for $i = 1, \ldots, n$, let Γ_i be a multi-valued mapping from $\mathcal{X}_i$ to $\mathcal{Y}$ such that

$$\begin{aligned} &\Gamma_i(0) = \mathcal{Y} \backslash \{y_i\} = \overline{\{y_i\}}, \\ &\Gamma_i(1) = \mathcal{Y}. \end{aligned} \tag{23}$$

Let $\mathcal{X}^n := \times_{i=1}^n \mathcal{X}_i$ and T be a multi-valued mapping from $\mathcal{X}^n$ to $\mathcal{Y}$ s.t.

$$\mathrm{T}(k) = \bigcap_{i=1}^{n} \Gamma_i(k_i), \tag{24}$$

for all $k \leftrightarrow (k_1, \ldots, k_n) \in \mathcal{X}^n$. Remark 1 provides a setting in which mapping T naturally occurs.

Remark 1. Let $\mathcal{Y} = \{y_1, \ldots, y_n\}$ be the domain of a variable $\boldsymbol{y}$. Assume there are n sources of information $\mathfrak{s}_i$, $i = 1, \ldots, n$, with $\mathfrak{s}_i$ providing the following piece of information about $\boldsymbol{y}$: $\boldsymbol{y} \in \overline{\{y_i\}}$. Furthermore, an agent who receives these pieces of information, considers that each source can either be reliable or not reliable. Let $\mathcal{X}_i = \{0, 1\}$ be the space denoting the reliability of $\mathfrak{s}_i$, where 0 means that $\mathfrak{s}_i$ is reliable and 1 means that $\mathfrak{s}_i$ is not reliable. Following Pichon et al.'s fusion scheme recalled in Section 2.3, a multi-valued mapping Γ_i from $\mathcal{X}_i$ to $\mathcal{Y}$ can be defined as (23); $\Gamma_i(k_i)$ interprets the testimony $\boldsymbol{y} \in \overline{\{y_i\}}$ in each configuration $k_i \in \mathcal{X}_i$ of the source $\mathfrak{s}_i$. More generally, a multi-valued mapping T from $\mathcal{X}^n$ to $\mathcal{Y}$ can be defined as (24); $\mathrm{T}(k)$ interprets the testimonies $\boldsymbol{y} \in \overline{\{y_i\}}$, $i = 1, \ldots, n$, in each joint configuration $k \in \mathcal{X}^n$ of the sources $\mathfrak{s}_i$, $i = 1, \ldots, n$.

Proposition 1. *Let m be any mass function on $\mathcal{Y}$ and let p be its MBD-equivalent. Then transferring p via the multi-valued mapping* T *defined by (24) induces m.*

Proof. Following [25], transferring distribution p via mapping T yields a mass function m' on $\mathcal{Y}$ defined, for $1 \leq k \leq 2^n$, by

$$m'(A_k) = \sum_{k' \in \mathcal{X}^n : \mathrm{T}(k') = A_k} p_{k'}.$$

For any $1 \leq k \leq 2^n$, we have

$$\begin{aligned}
\mathrm{T}(k) &= \bigcap_{i=1}^{n} \Gamma_i(k_i) \\
&= \left(\bigcap_{i:k_i=0} \Gamma_i(0) \right) \bigcap \left(\bigcap_{i:k_i=1} \Gamma_i(1) \right) \\
&= \bigcap_{i:k_i=0} \Gamma_i(0) \\
&= \bigcap_{i:k_i=0} \mathcal{Y} \backslash \{y_i\} \\
&= \mathcal{Y} \backslash \{y_i : i, k_i = 0\} \\
&= \{y_i : i, k_i = 1\} = A_k.
\end{aligned}$$

Hence $\sum_{k' \in \mathcal{X}^n : \mathrm{T}(k') = A_k} p_{k'} = p_k$ and thus m' is m. □

Example 2 illustrates Proposition 1.

Example 2 (Example 1 continued)**.** Consider the same mass function m as in Example 1. This mass function can be equivalently written as

$$m(A_4) = m(A_6) = m(A_8) = 1/3,$$

since $A_4 = \{y_1, y_2\}$, $A_6 = \{y_1, y_3\}$ and $A_8 = \{y_1, y_2, y_3\}$.

In addition, let p be the MBD-equivalent of m, *i.e.*, the MBD defined by

$$\begin{aligned}
p_4 &= p_{110} = m(\{y_1, y_2\}), \\
p_6 &= p_{101} = m(\{y_1, y_3\}), \\
p_8 &= p_{111} = m(\{y_1, y_2, y_3\}),
\end{aligned}$$

and $p_k = 0$ for $k = 1, 2, 3, 5, 7$.

We have

$$\begin{aligned}
\mathrm{T}(4) &= \mathrm{T}[(1,1,0)] \\
&= \Gamma_1(1) \bigcap \Gamma_2(1) \bigcap \Gamma_3(0) \\
&= \mathcal{Y} \cap \mathcal{Y} \cap \overline{\{y_3\}} \\
&= \{y_1, y_2\}
\end{aligned}$$

and, similarly, $\mathrm{T}(6) = \{y_1, y_3\}$ and $\mathrm{T}(8) = \{y_1, y_2, y_3\}$. Hence, probability p_4 is transferred to $\mathrm{T}(4) = \{y_1, y_2\} = A_4$, probability p_6 to A_6 and probability p_8 to A_8, *i.e.*, the mass function m defined by (15) is recovered.

Considering the setting of Remark 1, this means that mass function m can be viewed as resulting from having three sources $\mathfrak{s}_i$, each telling $\boldsymbol{y} \in \overline{\{y_i\}}$, $i = 1, 2, 3$, and such that the agent has the following knowledge on the reliability of the sources: with probability p_4, $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are not reliable and $\mathfrak{s}_3$ is reliable; with probability p_6, $\mathfrak{s}_1$ and $\mathfrak{s}_3$ are not reliable and $\mathfrak{s}_2$ is reliable; with probability p_8, the three sources are not reliable.

Let us finally state a result, which links the contour function of a mass function and the marginal expectations (called means in [17]) of its MBD-equivalent.

Lemma 1. *Let m be any mass function on $\mathcal{Y}$ and let p be its MBD-equivalent. The contour function of m is equal to the expectations of the r.v. $\{X_i : i = 1, \ldots, n\}$:*

$$pl(\{y_i\}) = \pi_i, \quad 1 \leq i \leq n.$$

Proof. For any $1 \leq i \leq n$, we have

$$\pi_i = \sum_{k:k_i=1} p_k$$

and

$$pl(\{y_i\}) = \sum_{k:k_i=1} m(A_k).$$

The lemma holds since $p_k = m(A_k)$ for all $1 \leq k \leq 2^n$, by definition of p. □

This section has provided a way to meaningfully induce any belief function from the MBD. This relation between the MBD and belief functions will be exploited in conjunction with Teugels' representations of the MBD [17] recalled in the next section, to introduce a new canonical decomposition of belief functions in Section 3.3.

3.2. Teugels' representations of the MBD

In [17], Teugels considers two vectors associated to the MBD (22): the vector $\boldsymbol{\mu}$ and the vector $\boldsymbol{\sigma}$ defined respectively as

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{2^n})',$$
$$\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_{2^n})',$$

where, for $1 \leq k \leq 2^n$,

$$\mu_k = \mathbb{E}\left[\prod_{i=1}^{n} X_i^{k_i}\right],$$
$$\sigma_k = \mathbb{E}\left[\prod_{i=1}^{n} (X_i - \pi_i)^{k_i}\right],$$

with k_i, $i = 1, \ldots, n$, the terms in the binary expansion of k. Let us remark that μ_k comes down to the (marginal) probability that each of the variables in $\{X_i : i, k_i = 1\}$ equals 1. This is not to be confused with p_k, which is the probability that each of the variables in $\{X_i : i, k_i = 1\}$ equals 1 *and* each of the variables in $\{X_i : i, k_i = 0\}$ equals 0. In addition, we note that vector $\boldsymbol{\mu}$ contains the moments of order 1 of the r.v. in $\{X_i : i = 1, \ldots, n\}$, and it contains also so-called product moments (or, joint moments) [30,31] of all subsets of at least two r.v. in $\{X_i : i = 1, \ldots, n\}$. Following [17], $\boldsymbol{\mu}$ is simply referred to in this paper as the vector of *moments* of the MBD (22). Vector $\boldsymbol{\sigma}$ contains the central moments of order 1 of the r.v. in $\{X_i : i = 1, \ldots, n\}$, which equal 0, as well as product (or, joint) central moments [30,31] of all subsets of at least two r.v. in $\{X_i : i = 1, \ldots, n\}$, and in particular the covariances of all pairs of r.v. in $\{X_i : i = 1, \ldots, n\}$ (this will be illustrated later by Example 3). Following again [17], $\boldsymbol{\sigma}$ is simply referred to as the vector of *central moments* of the MBD (22).

Teugels [17] shows that interesting relations hold between vectors $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and the MBD vector $\mathbf{p}$. We reproduce his results for convenience:

Theorem 1 (Theorem 1 and p. 261 of [17]). *We have*

(i)

$$\mathbf{p} = \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}\right) \boldsymbol{\mu} \tag{25}$$

and

$$\boldsymbol{\mu} = \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\right) \mathbf{p}. \tag{26}$$

(ii)

$$\mathbf{p} = \begin{bmatrix} \xi_n & -1 \\ \pi_n & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} \xi_1 & -1 \\ \pi_1 & 1 \end{bmatrix} \boldsymbol{\sigma} \tag{27}$$

and

$$\boldsymbol{\sigma} = \begin{bmatrix} 1 & 1 \\ -\pi_n & \xi_n \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 1 \\ -\pi_1 & \xi_1 \end{bmatrix} \mathbf{p}. \tag{28}$$

(iii)

$$\boldsymbol{\mu} = \begin{bmatrix} 1 & 0 \\ \pi_n & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ \pi_1 & 1 \end{bmatrix} \boldsymbol{\sigma} \tag{29}$$

and

$$\boldsymbol{\sigma} = \begin{bmatrix} 1 & 0 \\ -\pi_n & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ -\pi_1 & 1 \end{bmatrix} \boldsymbol{\mu}. \tag{30}$$

Of particular interest with respect to the new canonical decomposition of a belief function to be introduced in the next section, are relations (27) and (28); the former is illustrated by Example 3.

Example 3 (Example 1.2 of [17]). Let $n = 3$. In this case, we have

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \\ \sigma_7 \\ \sigma_8 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \mathbb{E}[(X_1 - \pi_1)(X_2 - \pi_2)] \\ 0 \\ \mathbb{E}[(X_1 - \pi_1)(X_3 - \pi_3)] \\ \mathbb{E}[(X_2 - \pi_2)(X_3 - \pi_3)] \\ \mathbb{E}[(X_1 - \pi_1)(X_2 - \pi_2)(X_3 - \pi_3)] \end{bmatrix}.$$

σ_4 is nothing but the covariance between r.v. X_1 and X_2; accordingly we may denote σ_4 by $\sigma_{1,2}$. Similarly, we may denote σ_6 by $\sigma_{1,3}$, σ_7 by $\sigma_{2,3}$, and σ_8 by $\sigma_{1,2,3}$. Thus, we can write:

$$\boldsymbol{\sigma} = (1, 0, 0, \sigma_{1,2}, 0, \sigma_{1,3}, \sigma_{2,3}, \sigma_{1,2,3})'.$$

From (27), one obtains

$$\mathbf{p} = \begin{bmatrix} p_{000} \\ p_{100} \\ p_{010} \\ p_{110} \\ p_{001} \\ p_{101} \\ p_{011} \\ p_{111} \end{bmatrix} = \left(\begin{bmatrix} \xi_3 & -1 \\ \pi_3 & 1 \end{bmatrix} \otimes \begin{bmatrix} \xi_2 & -1 \\ \pi_2 & 1 \end{bmatrix} \otimes \begin{bmatrix} \xi_1 & -1 \\ \pi_1 & 1 \end{bmatrix} \right) \boldsymbol{\sigma}$$

$$= \begin{bmatrix} \xi_1\xi_2\xi_3 + \xi_3\sigma_{1,2} + \xi_2\sigma_{1,3} + \xi_1\sigma_{2,3} - \sigma_{1,2,3} \\ \pi_1\xi_2\xi_3 - \xi_3\sigma_{1,2} - \xi_2\sigma_{1,3} + \pi_1\sigma_{2,3} + \sigma_{1,2,3} \\ \xi_1\pi_2\xi_3 - \xi_3\sigma_{1,2} + \pi_2\sigma_{1,3} - \xi_1\sigma_{2,3} + \sigma_{1,2,3} \\ \pi_1\pi_2\xi_3 + \xi_3\sigma_{1,2} - \pi_2\sigma_{1,3} - \pi_1\sigma_{2,3} - \sigma_{1,2,3} \\ \xi_1\xi_2\pi_3 + \pi_3\sigma_{1,2} - \xi_2\sigma_{1,3} - \xi_1\sigma_{2,3} + \sigma_{1,2,3} \\ \pi_1\xi_2\pi_3 - \pi_3\sigma_{1,2} + \xi_2\sigma_{1,3} - \pi_1\sigma_{2,3} - \sigma_{1,2,3} \\ \xi_1\pi_2\pi_3 - \pi_3\sigma_{1,2} - \pi_2\sigma_{1,3} + \xi_1\sigma_{2,3} - \sigma_{1,2,3} \\ \pi_1\pi_2\pi_3 + \pi_3\sigma_{1,2} + \pi_2\sigma_{1,3} + \pi_1\sigma_{2,3} + \sigma_{1,2,3} \end{bmatrix}.$$

As detailed in [17, Section 2.3 (i)], both representations (25) and (27) contain as many parameters as $\mathbf{p}$, that is $2^n - 1$, since $\mu_1 = \sigma_1 = 1$. Specifically, in (27), $2^n - n - 1$ parameters are given by $\boldsymbol{\sigma}$, as $\sigma_k = 0$ when $k_1 + \cdots + k_n = 1$, and the n remaining parameters are π_i, $i = 1, \ldots, n$. The non-null components of $\boldsymbol{\sigma}$ represent the $2^n - n - 1$ possible dependencies between any subset (of at least two) of the r.v. $\{X_i : i = 1, \ldots, n\}$; under independence of all these r.v., we have $\boldsymbol{\sigma} = \mathbf{e}_1$. As a matter of fact, Teugels [17] refers to $\boldsymbol{\sigma}$ as the *dependency vector*. The central moments stored in $\boldsymbol{\sigma}$ may be particularly useful as illustrated by Teugels and Van Horebeek [32], who use them to formulate hypothesis tests about the interactions between treatments by several kinds of drugs.

Representation (27) can be insightfully depicted graphically using a Venn diagram,[3] as illustrated in Fig. 1 for the case $n = 3$:

- each Bernoulli random variable X_i underlying the MBD is represented by a circle (and more generally a closed curve for the cases $n > 3$);
- the dependency (central moment) σ_k between the variables $\{X_i : i, k_i = 1\}$ is in one-to-one correspondence with an overlap region, *i.e.*, the non-null components of $\boldsymbol{\sigma}$ are mapped to overlap regions;
- the mean π_i of each variable X_i is in one-to-one correspondence with a non overlap region, *i.e.*, the means π_i, $i = 1, \ldots, n$, are mapped to non overlap regions.

3.3. A new canonical decomposition of belief functions based on Teugels' representation of the MBD

It will be convenient to introduce the following definition associated with representation (27).

Definition 2 (Teugels' representation of the MDB). The Teugels' representation of the MBD (22) is the vector $\boldsymbol{\tau}$ of size 2^n such that, for $1 \leq k \leq 2^n$,

$$\tau_k = \begin{cases} \pi_{\arg\max_{1 \leq i \leq n} k_i} & \text{if } k_1 + \cdots + k_n = 1, \\ \sigma_k & \text{otherwise.} \end{cases} \tag{31}$$

[3] Although one must be aware that some individual (specifically overlap) regions can have positive or negative value.

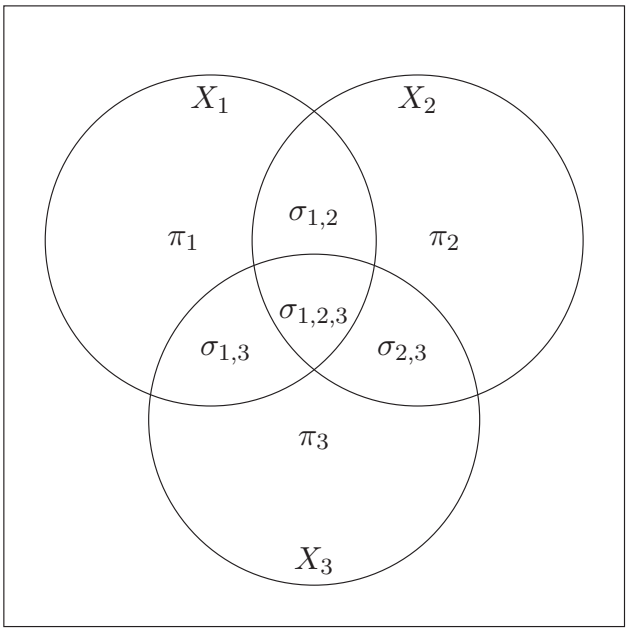


Fig. 1. Venn diagram for representation (27) of the multivariate Bernoulli distribution, using the notation for $\boldsymbol{\sigma}$ introduced in Example 3.

Definition 2 basically introduces a vector which stores the $2^n - 1$ parameters of representation (27), *i.e*, the (marginal) mean values π_i, $i = 1, \ldots, n$, and the non-null components of $\boldsymbol{\sigma}$. Note that $\tau_1 = \sigma_1 = 1$. For instance, for $n = 2$, Eq. (27) is

$$\begin{aligned}\mathbf{p} &= \begin{bmatrix} \xi_2 & -1 \\ \pi_2 & 1 \end{bmatrix} \otimes \begin{bmatrix} \xi_1 & -1 \\ \pi_1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \end{bmatrix} \\ &= \begin{bmatrix} \xi_2 & -1 \\ \pi_2 & 1 \end{bmatrix} \otimes \begin{bmatrix} \xi_1 & -1 \\ \pi_1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \sigma_4 \end{bmatrix}\end{aligned}$$

and it can be rewritten using $\boldsymbol{\tau}$ as

$$\mathbf{p} = \begin{bmatrix} 1-\tau_3 & -1 \\ \tau_3 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1-\tau_2 & -1 \\ \tau_2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \tau_4 \end{bmatrix}.$$

Proposition 1and Definition 2 allow us to propose the following decomposition of a mass function into elementary items, which have well-defined semantics.

Definition 3 (t-canonical decomposition)**.** Let m be a mass function defined on a domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$ and let τ be the Teugels' representation of its MBD-equivalent. Its t-canonical decomposition (or t-decomposition for short) is the mapping t on $2^{\mathcal{Y}} \backslash \{\emptyset\}$, called the Teugels function and defined by:

$$t(A_k) = \tau_k, \quad 1 < k \leq 2^n, \tag{32}$$

with A_k the kth subset of $\mathcal{Y}$ according to the binary order.

The t-canonical decomposition of a mass function m is thus basically Teugels' representation of its MBD-equivalent. It is unique since the Teugels function t is in one-to-one correspondence with function m (function τ is in such correspondence with the MBD-equivalent p of m and the vector $\mathbf{p}$ is equal to the vector $\mathbf{m}$).

The Teugels function t can be computed from Eqs. (28),(31) and (32), and Lemma 1, which comes down to the following definition for t, for all $A \in 2^{\mathcal{Y}} \backslash \{\emptyset\}$:

$$t(A) = \begin{cases} pl(A), & \text{if } |A| = 1, \\ \left(\begin{bmatrix} 1 & 1 \\ -pl(\{y_n\}) & 1 - pl(\{y_n\}) \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 1 \\ -pl(\{y_1\}) & 1 - pl(\{y_1\}) \end{bmatrix} \mathbf{m} \right)(A), & \text{otherwise.} \end{cases} \tag{33}$$

Adopting the setting of Remark 1, the t-decomposition of a mass function can be interpreted in the same vein as Smets' decomposition, *i.e.*, a mass function can be viewed as resulting from partially reliable sources providing crisp pieces of information. More precisely, any mass function m can be recovered from the following basic components:

- crisp testimonies $\boldsymbol{y} \in \overline{\{y_i\}}$ provided by sources $\mathfrak{s}_i$, $i = 1, \ldots, n$;
- knowledge on the individual (marginal) reliability of each source $\mathfrak{s}_i$, represented by the mean $t(\{y_i\})$ (which happens to be given by the contour function, *i.e.*, $t(\{y_i\}) = pl(\{y_i\})$);
- knowledge on the dependency between the source reliabilities for each subset of sources $\{\mathfrak{s}_i : i, k_i = 1\}$, represented by the central moment $t(A_k)$, $|A_k| > 1$, with k_i, $i = 1, \ldots, n$, the terms in the binary expansion of k.

This is illustrated by Examples 4 and 5.

Example 4. Let m be a mass function about a variable $\boldsymbol{y}$ defined on $\mathcal{Y} = \{y_1, y_2\}$, with associated Teugels function t. From (33) we obtain

$$\begin{aligned} t(\{y_1\}) &= pl(\{y_1\}) \\ &= m(\{y_1\}) + m(\{y_1, y_2\}), \end{aligned} \tag{34}$$

$$\begin{aligned} t(\{y_2\}) &= pl(\{y_2\}) \\ &= m(\{y_2\}) + m(\{y_1, y_2\}), \end{aligned} \tag{35}$$

$$\begin{aligned} t(\{y_1, y_2\}) &= \begin{bmatrix} pl(\{y_2\})pl(\{y_1\}) \\ -pl(\{y_2\})(1 - pl(\{y_1\})) \\ -pl(\{y_1\})(1 - pl(\{y_2\})) \\ (1 - pl(\{y_1\}))(1 - pl(\{y_2\})) \end{bmatrix}' \mathbf{m} \\ &= m(\{y_1, y_2\})m(\emptyset) - m(\{y_1\})m(\{y_2\}). \end{aligned} \tag{36}$$

Let us now consider the following pieces of evidence:

- There are two sources $\mathfrak{s}_1$ and $\mathfrak{s}_2$, with $\mathfrak{s}_i$ providing the crisp piece of information $\boldsymbol{y} \in \overline{\{y_i\}}$, *i.e.*, $\mathfrak{s}_1$ tells $\boldsymbol{y} \in \{y_2\}$ and $\mathfrak{s}_2$ tells $\boldsymbol{y} \in \{y_1\}$;
- $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are believed to be non reliable with probabilities $t(\{y_1\})$ and $t(\{y_2\})$ respectively;
- The dependence between their reliabilities is given by covariance $t(\{y_1, y_2\})$.

These latter two pieces of evidence yield, using (27), the following knowledge about the reliability of the sources

$$\begin{aligned} P(X_1 = 0, X_2 = 0) &= (1 - t(\{y_1\}))(1 - t(\{y_2\})) + t(\{y_1, y_2\}), \\ P(X_1 = 1, X_2 = 0) &= t(\{y_1\})(1 - t(\{y_2\})) - t(\{y_1, y_2\}), \\ P(X_1 = 0, X_2 = 1) &= (1 - t(\{y_1\}))t(\{y_2\}) - t(\{y_1, y_2\}), \\ P(X_1 = 1, X_2 = 1) &= t(\{y_1\})t(\{y_2\}) + t(\{y_1, y_2\}). \end{aligned}$$

Besides, the mapping T representing the interpretations of the source testimonies in each of their joint configurations in terms of reliability is:

$$\begin{aligned} \mathrm{T}[(0,0)] &= \Gamma_1(0) \cap \Gamma_2(0) \\ &= \{y_2\} \cap \{y_1\} \\ &= \emptyset, \\ \mathrm{T}[(1,0)] &= \Gamma_1(1) \cap \Gamma_2(0) \\ &= \mathcal{Y} \cap \{y_1\} \\ &= \{y_1\}, \\ \mathrm{T}[(0,1)] &= \{y_2\}, \\ \mathrm{T}[(1,1)] &= \mathcal{Y}. \end{aligned}$$

Using Eqs. (34)–(36), it can easily be checked that

$$\begin{aligned} (1 - t(\{y_1\}))(1 - t(\{y_2\})) + t(\{y_1, y_2\}) &= m(\emptyset), \\ t(\{y_1\})(1 - t(\{y_2\})) - t(\{y_1, y_2\}) &= m(\{y_1\}), \\ (1 - t(\{y_1\}))t(\{y_2\}) - t(\{y_1, y_2\}) &= m(\{y_2\}), \\ t(\{y_1\})t(\{y_2\}) + t(\{y_1, y_2\}) &= m(\{y_1, y_2\}). \end{aligned}$$

Mass function m is thus clearly recovered when transferring P to $\mathcal{Y}$ via T. Remark also that the greater the covariance $t(\{y_1, y_2\})$ is, the greater are the probabilities $P(X_1 = 1, X_2 = 1)$ and $P(X_1 = 0, X_2 = 0)$ that the sources are jointly non reliable and that they are jointly reliable, respectively (and the smaller are the probabilities $P(X_1 = 0, X_2 = 1)$ and $P(X_1 = 1, X_2 = 0)$ that the first source is reliable whereas the second is not and that the first source is not reliable whereas the second is, respectively). Hence, the greater this covariance is, the greater are the masses on $\mathcal{Y}$ and on $\emptyset$ (and the smaller are the masses on $\{y_1\}$ and on $\{y_2\}$).

Example 5 (Example 2 continued)**.** Let us compute the Teugels function t associated to the mass function m of Example 2. We find using (33)

$$\begin{aligned} t(\{y_1\}) &= pl(\{y_1\}) = 1, \\ t(\{y_2\}) &= pl(\{y_2\}) = 2/3, \\ t(\{y_3\}) &= pl(\{y_3\}) = 2/3, \end{aligned}$$

and since (using $\pi_i = pl(\{y_i\})$ and $\xi_i = 1 - \pi_i$, $1 \le i \le 3$, to shorten expressions)

$$\begin{aligned} &\begin{bmatrix} 1 & 1 \\ -pl(\{y_3\}) & 1 - pl(\{y_3\}) \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ -pl(\{y_2\}) & 1 - pl(\{y_2\}) \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ -pl(\{y_1\}) & 1 - pl(\{y_1\}) \end{bmatrix} \mathbf{m} \\ &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -\pi_1 & \xi_1 & -\pi_1 & \xi_1 & -\pi_1 & \xi_1 & -\pi_1 & \xi_1 \\ -\pi_2 & -\pi_2 & \xi_2 & \xi_2 & -\pi_2 & -\pi_2 & \xi_2 & \xi_2 \\ \pi_2\pi_1 & -\pi_2\xi_1 & -\pi_1\xi_2 & \xi_2\xi_1 & \pi_2\pi_1 & -\pi_2\xi_1 & -\pi_1\xi_2 & \xi_2\xi_1 \\ -\pi_3 & -\pi_3 & -\pi_3 & -\pi_3 & \xi_3 & \xi_3 & \xi_3 & \xi_3 \\ \pi_3\pi_1 & -\pi_3\xi_1 & \pi_3\pi_1 & -\pi_3\xi_1 & -\pi_1\xi_3 & \xi_3\xi_1 & -\pi_1\xi_3 & \xi_3\xi_1 \\ \pi_3\pi_2 & \pi_3\pi_2 & -\pi_3\xi_2 & -\pi_3\xi_2 & -\pi_2\xi_3 & -\pi_2\xi_3 & \xi_3\xi_2 & \xi_3\xi_2 \\ -\pi_1\pi_3\pi_2 & \pi_3\pi_2\xi_1 & \pi_1\pi_3\xi_2 & -\pi_3\xi_2\xi_1 & \pi_1\pi_2\xi_3 & -\pi_2\xi_3\xi_1 & -\pi_1\xi_3\xi_2 & \xi_3\xi_2\xi_1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ m(\{y_1, y_2\}) \\ 0 \\ m(\{y_1, y_3\}) \\ 0 \\ m(\{y_1, y_2, y_3\}) \end{bmatrix}, \end{aligned}$$

we find

$$\begin{aligned} t(\{y_1, y_2\}) &= m(\{y_1, y_2\})\xi_2\xi_1 - m(\{y_1, y_3\})\pi_2\xi_1 + m(\{y_1, y_2, y_3\}\xi_2\xi_1 = 0 \quad (\text{since } \xi_1 = 0), \\ t(\{y_1, y_3\}) &= -m(\{y_1, y_2\})\pi_3\xi_1 - m(\{y_1, y_3\})\xi_3\xi_1 + m(\{y_1, y_2, y_3\}\xi_3\xi_1 = 0 \quad (\text{since } \xi_1 = 0), \\ t(\{y_2, y_3\}) &= -m(\{y_1, y_2\})\pi_3\xi_2 - m(\{y_1, y_3\})\pi_2\xi_3 + m(\{y_1, y_2, y_3\}\xi_3\xi_2 \\ &= -(1/3)(2/3)(1/3) - (1/3)(2/3)(1/3) + (1/3)(1/3)(1/3) = -1/9, \\ t(\{y_1, y_2, y_3\}) &= -m(\{y_1, y_2\})\pi_3\xi_2\xi_1 - m(\{y_1, y_3\})\pi_2\xi_3\xi_1 + m(\{y_1, y_2, y_3\}\xi_3\xi_2\xi_1 = 0 \quad (\text{since } \xi_1 = 0). \end{aligned}$$

In other words, considering as in Example 2 the setting of Remark 1, mass function m can be seen as originating from the following pieces of evidence:

- There are three sources $\mathfrak{s}_i$, each telling $\boldsymbol{y} \in \overline{\{y_i\}}$, $i = 1, 2, 3$;
- $\mathfrak{s}_1$, $\mathfrak{s}_2$ and $\mathfrak{s}_3$ are believed to be non reliable with probabilities 1, 2/3 and 2/3 respectively;
- The covariance between the reliabilities of $\mathfrak{s}_2$ and $\mathfrak{s}_3$ is $-1/9$, and all other subsets of variables representing the reliabilities of the sources have null central moments.

Note that this is clearly a different decomposition to that of Smets provided for the same mass function in Example 1.

Since it relies on representation (27) of the MBP, the t-decomposition can also be represented by a Venn diagram. For instance, the pieces of evidence underlying mass function m in Example 5 are shown in Fig. 2.

Up until now, the Teugels function t has been computed using (33), where t is obtained from the contour function and the mass function. However, as will be explained in the remainder of this section, t can also conveniently be obtained using only the commonality function.

Theorem 1 leads to the following straightforward remark.

Remark 2. Let m be a mass function on $\mathcal{Y}$ and p its MBD-equivalent. By comparing Eqs. (5) and (26), it is clear that the vector $\boldsymbol{\mu}$ of moments of p is equal to the commonality vector $\mathbf{q}$ associated to m, *i.e.*, $\mu_k = q(A_k)$ for all $1 \le k \le 2^n$.

Adopting the setting of Remark 1 and using Remark 2, a new interpretation is obtained for the commonality function: $q(A_k)$ is the moment between the random variables $\{X_i : i, k_i = 1\}$, representing the reliabilities of sources $\{\mathfrak{s}_i : i, k_i = 1\}$. More simply, $q(A_k)$ is the marginal probability that each of the sources in $\{\mathfrak{s}_i : i, k_i = 1\}$ is not reliable.

In particular, Remark 2 yields an alternative and direct proof for Lemma 1. Indeed, for k such that $k_1 + \cdots + k_n = 1$, we have

$$\begin{aligned} q(A_k) &= \mathbb{E}\left[X_{\arg\max_{1 \le i \le n} k_i}\right] \\ &= \pi_{\arg\max_{1 \le i \le n} k_i}, \end{aligned}$$

and thus $q(\{y_i\}) = \pi_i$, for all $1 \le i \le n$.

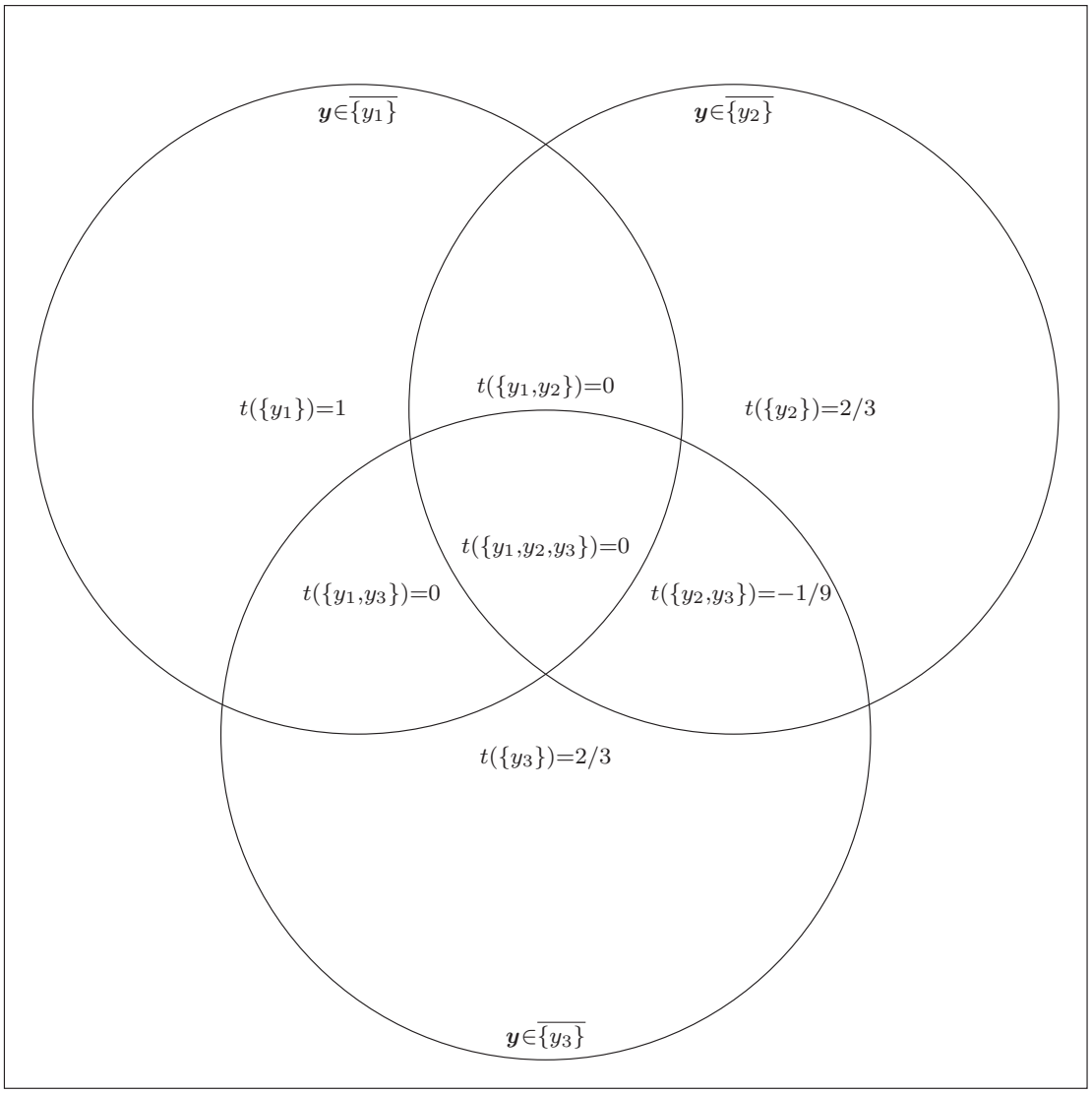


Fig. 2. t-decomposition of mass function m in Example 5, with $t(\{y_i\})$ the marginal probability that source $\mathfrak{s}_i$ is not reliable, $t(A_k)$, $|A_k|>1$, the central moment between the reliabilities of the subset of sources $\{\mathfrak{s}_i : i, k_i = 1\}$, and $\boldsymbol{y} \in \overline{\{y_i\}}$ the testimony provided by source $\mathfrak{s}_i$.

In addition, from Lemma 1 and Eqs. (30)–(32) , the following equivalent definition based only on the commonality function is obtained for t:

$$t(A) = \begin{cases} q(A), & \text{if } |A| = 1, \\ \left(\begin{bmatrix} 1 & 0 \\ -q(\{y_n\}) & 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 & 0 \\ -q(\{y_1\}) & 1 \end{bmatrix} \mathbf{q}\right)(A), & \text{otherwise.} \end{cases} \tag{37}$$

For instance, for $n = 2$, the Teugels function t associated to a mass function with commonality function q is

$$\begin{aligned} t(\{y_1\}) &= q(\{y_1\}), \\ t(\{y_2\}) &= q(\{y_2\}), \\ t(\{y_1, y_2\}) &= \begin{bmatrix} q(\{y_1\})q(\{y_2\}) & -q(\{y_2\}) & -q(\{y_1\}) & 1 \end{bmatrix} \mathbf{q} \\ &= q(\{y_1, y_2\} - q(\{y_1\})q(\{y_2\}). \end{aligned}$$

This section has proposed a new solution to decompose a belief function (Definition 3). This solution relies on the following building blocks: (i) the notion of MBD-equivalent of a belief function (Definition 1) and a particular multi-valued mapping allowing one to recover a belief function from its MBD-equivalent (Proposition 1), (ii) a setting that provides a meaning to this connection between the MBD and belief functions (Remark 1), and (iii) a particular representation of the MBD (Definition 2). Next section provides some comments on this solution.

4. Some comments on the *t*-Canonical decomposition

In this section, some comments are provided on the t-canonical decomposition. First, we study a particular case of the general approach proposed in [33] for the conjunctive combination of bodies of evidence with known dependence structure. Then, this study is used to express the t-canonical decomposition in terms of simple mass functions, similarly as Smets' decomposition. Next, the t-canonical decomposition is compared further with Smets' decomposition. Finally, the t-canonical decomposition is considered in the context of random sets.

4.1. Conjunctive combination with known dependence structure

The general approach proposed in [33] for the conjunctive combination of bodies of evidence represented by mass functions, allows for other dependence structures among them besides independence. This approach is the following. Let $m_1, \dots, m_N$ be N mass functions on $\mathcal{Y}$. The approach [33] defines a mass function $m_\cap$ on $\mathcal{Y}$ resulting from a conjunctive combination of $m_1, \dots, m_N$ as the result of the following procedure:

1. A so-called joint mass function $jm : \times_{i=1}^{N} 2^{\mathcal{Y}} \to [0,1]$ is built, preserving $m_1, \dots, m_N$ as marginals, which means that $\forall A_i \in \mathcal{F}_i$, with $\mathcal{F}_i$ the set of focal sets of m_i,
$$m_i(A_i) = \sum_{A_1 \in \mathcal{F}_1, \dots, A_{i-1} \in \mathcal{F}_{i-1}, A_{i+1} \in \mathcal{F}_{i+1}, \dots, A_N \in \mathcal{F}_N} jm(A_1, \dots, A_{i-1}, A_i, A_{i+1}, \dots, A_N). \tag{38}$$
2. Each joint mass $jm(A_1, \dots, A_N)$ is allocated to the subset $\bigcap_{i=1}^{N} A_i$ in the final mass function $m_\cap$, *i.e.*, for all $A \subseteq \mathcal{Y}$
$$m_\cap(A) = \sum_{\bigcap_{i=1}^{N} A_i = A} jm(A_1, \dots, A_N).$$

As noted in [33], Eq. (38) indicates that the information m_i representing each body of evidence can be retrieved from the richer information jm that includes a representation of their mutual dependence. In particular, the combination of $m_1, \dots, m_N$ by the conjunctive rule $\textcircled{\cap}$ is a particular case of this approach, retrieved for $jm(A_1, \dots, A_N) = m_1(A_1) m_2(A_2) \dots m_N(A_N)$ in Step 1; this latter equality corresponds to the assumption that the bodies of evidence represented by mass functions $m_1, \dots, m_N$ are independent.

Let us now consider a particular case of this general approach where each mass function m_i has only two focal sets, denoted A_{i_0} and A_{i_1}, for some $A_{i_0}, A_{i_1} \subseteq \mathcal{Y}$, such that $m(A_{i_1}) = \pi_i$ and $m(A_{i_0}) = 1 - \pi_i$ for some $\pi_i \in [0,1]$. Such kind of mass function will be called *binary* hereafter and may be simply denoted $(A_{i_0}, A_{i_1})^{\pi_i}$.

When each mass function m_i is binary and such that $m_i = (A_{i_0}, A_{i_1})^{\pi_i}$, it is clear that only subsets $\mathbf{A}_k := (A_{1_{k_1}}, A_{2_{k_2}}, \dots, A_{N_{k_N}}) \subseteq \times_{i=1}^{N} \mathcal{Y}$, $1 \le k \le 2^N$ with $k \leftrightarrow (k_1, \dots, k_N) \in \{0,1\}^N$, can receive a non null joint mass in jm. Moreover, by associating a Bernoulli r.v. X_i to each m_i, such that $P(X_i = 1) = m_i(A_{i_1}) = \pi_i$ and $P(X_i = 0) = m_i(A_{i_0}) = 1 - \pi_i$, we can establish a one-to-one correspondence between the MBD with underlying r.v. X_i, $i = 1, \dots, N$, and the joint mass function jm, by setting

$$P(X_1 = k_1, \dots, X_N = k_N) = jm(A_{1_{k_1}}, A_{2_{k_2}}, \dots, A_{N_{k_N}}),$$

or for short, using $k \leftrightarrow (k_1, \dots, k_N)$, $p_k = jm(\mathbf{A}_k)$ with $p_k := P(X_1 = k_1, \dots, X_N = k_N)$. From Theorem 1, we have that the joint mass function jm can be fully specified by parameters π_i, $i = 1, \dots, N$, and vector $\boldsymbol{\sigma}$ associated with its corresponding MBD. Most interestingly, the dependence structure encoded in jm between the binary mass functions m_i is actually entirely captured by vector $\boldsymbol{\sigma}$, since any MBD (and thus any corresponding joint mass function) with given marginals is obtained for a unique vector $\boldsymbol{\sigma}$. More specifically, since σ_k, $k_1 + \dots + k_N > 1$, represents the dependency among variables $\{X_i : i, k_i = 1\}$, and since variable X_i is associated to mass function m_i, $i = 1, \dots, N$, we may regard σ_k as capturing the dependency among mass functions $\{m_i : i, k_i = 1\}$.

Since the conjunctive combination of N binary mass functions with dependence structure represented by a joint mass function jm, is completely determined by vector $\boldsymbol{\sigma}$ associated to jm, then this combination can be expressed as a parameterised combination rule for binary mass functions, with parameter $\boldsymbol{\sigma}$ representing the dependence structure. Formally, let $\mathcal{B}$ denote the set of all binary mass functions on $\mathcal{Y}$ and $\mathcal{M}$ the set of all mass functions on $\mathcal{Y}$. In addition, let $m_i = (A_{i_0}, A_{i_1})^{\pi_i}$, $i = 1, \dots, N$, be N binary mass functions. Then, we refer to the operator $\textcircled{\cap}_{\boldsymbol{\sigma}} : \mathcal{B}^N \to \mathcal{M}$ such that

$$\textcircled{\cap}_{\boldsymbol{\sigma}}(m_1, \dots, m_N) = \textcircled{\cap}_{\boldsymbol{\sigma}}\left((A_{1_0}, A_{1_1})^{\pi_1}, \dots, (A_{N_0}, A_{N_1})^{\pi_N}\right) := m_\cap,$$

with $m_\cap$ the result of the conjunctive combination of mass functions m_i with dependence represented by the joint mass function jm determined by vector $\boldsymbol{\sigma}$ and having marginals m_i, $i = 1, \dots, N$, as the *conjunctive combination with dependence* $\boldsymbol{\sigma}$ (or, for short, $\boldsymbol{\sigma}$*-conjunctive combination*) of these mass functions. This is illustrated by Example 6.

Example 6. Let $m_1 = (A_{1_0}, A_{1_1})^{\pi_1}$ and $m_2 = (A_{2_0}, A_{2_1})^{\pi_2}$ be two binary mass functions on $\mathcal{Y} = \{y_1, y_2, y_3\}$ such that $A_{1_0} = \{y_2\}, A_{1_1} = \mathcal{Y}, \pi_1 = 0.4$, and $A_{2_0} = \{y_2, y_3\}, A_{2_1} = \{y_1, y_2\}, \pi_2 = 0.5$. In other words, $m_1 = (\{y_2\}, \mathcal{Y})^{0.4}$ and $m_2 = (\{y_2, y_3\}, \{y_1, y_2\})^{0.5}$. Let their dependence structure be represented by the joint mass function jm such that

$$jm(\mathbf{A}_1) = jm(\{y_2\}, \{y_2, y_3\}) = 0.4,$$

$$jm(\mathbf{A}_2) = jm(\mathcal{Y}, \{y_2, y_3\}) = 0.1,$$
$$jm(\mathbf{A}_3) = jm(\{y_2\}, \{y_1, y_2\}) = 0.2,$$
$$jm(\mathbf{A}_4) = jm(\mathcal{Y}, \{y_1, y_2\}) = 0.3.$$

One can easily check that jm satisfies (38) for $i = 1, 2$.

Performing Step 2 of the conjunctive combination of m_1 and m_2, we obtain:

$$m_\cap(\{y_2\}) = jm(\{y_2\}, \{y_2, y_3\}) + jm(\{y_2\}, \{y_1, y_2\}) = 0.6,$$
$$m_\cap(\{y_2, y_3\}) = jm(\mathcal{Y}, \{y_2, y_3\}) = 0.1,$$
$$m_\cap(\{y_1, y_2\}) = jm(\mathcal{Y}, \{y_1, y_2\}) = 0.3.$$

Now, the MBD in one-to-one correspondence with jm is

$$P(X_1 = 0, X_2 = 0) = p_1 = jm(\{y_2\}, \{y_2, y_3\}),$$
$$P(X_1 = 1, X_2 = 0) = p_2 = jm(\mathcal{Y}, \{y_2, y_3\}),$$
$$P(X_1 = 0, X_2 = 1) = p_3 = jm(\{y_2\}, \{y_1, y_2\}),$$
$$P(X_1 = 1, X_2 = 1) = p_4 = jm(\mathcal{Y}, \{y_1, y_2\}).$$

The vector $\boldsymbol{\sigma}$ associated to this MBD is, using (28),

$$\boldsymbol{\sigma} = \begin{bmatrix} 1 & 1 \\ -\pi_2 & 1-\pi_2 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ -\pi_1 & 1-\pi_1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix}$$
$$= (1, 0, 0, 0.1)'$$

From the definition of $\textcircled{\cap}_{\boldsymbol{\sigma}}$, we have then

$$m_\cap = \textcircled{\cap}_{\boldsymbol{\sigma}}(m_1, m_2) = \textcircled{\cap}_{(1,0,0,0.1)'}\left((\{y_2\}, \mathcal{Y})^{0.4}, (\{y_2, y_3\}, \{y_1, y_2\})^{0.5}\right).$$

The operation $\textcircled{\cap}_{\boldsymbol{\sigma}}(m_1, \ldots, m_N)$ is well-defined as long as there exists a joint mass function jm compatible with vector $\boldsymbol{\sigma}$ and binary mass functions m_i.

In addition, it is clear that the case where the bodies of evidence, represented by binary mass functions $m_1, \ldots, m_N$, are independent, comes down to $\boldsymbol{\sigma} = \mathbf{e}_1$, which means that the following equality holds:

$$\textcircled{\cap}_{\mathbf{e}_1}(m_1, \ldots, m_N) = m_1 \textcircled{\cap} \ldots \textcircled{\cap}\, m_N,$$

that is, the conjunctive rule is a particular case of the $\boldsymbol{\sigma}$-conjunctive rule, recovered for $\boldsymbol{\sigma} = \mathbf{e}_1$.

Finally, note that $\boldsymbol{\sigma}$-conjunctive combination $\textcircled{\cap}_{\boldsymbol{\sigma}}\left((A_{1_0}, A_{1_1})^{\pi_1}, \ldots, (A_{N_0}, A_{N_1})^{\pi_N}\right)$ of N binary mass functions $(A_{i_0}, A_{i_1})^{\pi_i}$ may be denoted for short $\textcircled{\cap}_{\boldsymbol{\sigma}}\left(A_{1_0}^{\pi_1}, \ldots, A_{N_0}^{\pi_N}\right)$ when $A_{i_1} = \mathcal{Y}$, $i = 1, \ldots, N$, since a binary mass function $(A_{i_0}, A_{i_1})^{\pi_i}$ such that $A_{i_1} = \mathcal{Y}$ is nothing but the simple mass function $A_{i_0}^{\pi_i}$.

4.2. t-*Canonical decomposition in terms of simple mass functions*

In Section 2.3, it was recalled that when some sources $\mathfrak{s}_1, \ldots, \mathfrak{s}_N$ provide crisp testimonies $\mathbf{A} = (A_{\mathfrak{s}_1}, \ldots, A_{\mathfrak{s}_N})$, and when knowledge about the source reliability is uncertain such that each joint state $k \in \mathcal{X}^N$ has a probability p_k , then the induced knowledge of the agent about $\mathcal{Y}$ is represented by a mass function m defined by

$$m(B) = \sum_{k:\Gamma_{\mathbf{A}}(k)=B} p_k, \quad \forall B \subseteq \mathcal{Y}. \tag{39}$$

This induced knowledge may be seen as the 'trace' on $\mathcal{Y}$ of the available pieces of information, that is, of the source testimonies and of the knowledge about their reliability. Besides, let us note that the vector $\boldsymbol{\sigma}$, obtained using (28) from the MBD such that $P(X_1 = k_1, \ldots, X_N = k_N) = p_k$ with X_i representing the reliability of $\mathfrak{s}_i$, represents the meta-dependences among the sources, that is, the dependences (in terms of central moments) between the reliabilities of all subsets of (at least two) sources.

Considering solely the testimony of source $\mathfrak{s}_i$ and what is known of its reliability, *i.e.*, it is not reliable with marginal probability π_i given by (20), then the induced knowledge on $\mathcal{Y}$ given this testimony is represented according to (16) by the simple mass function $A_{\mathfrak{s}_i}^{\pi_i}$. This latter mass function is the trace on $\mathcal{Y}$ of the pieces of information pertaining only to $\mathfrak{s}_i$, that is, of the testimony of $\mathfrak{s}_i$ and of what is known of its reliability. Accordingly, it will be referred to as the individual (or, marginal) trace associated with $\mathfrak{s}_i$.

Theorem 2. *The trace m defined by (39) verifies*

$$m = \textcircled{\cap}_{\boldsymbol{\sigma}}(A_{\mathfrak{s}_1}^{\pi_1}, \ldots, A_{\mathfrak{s}_N}^{\pi_N}), \tag{40}$$

with $A^{\pi_i}_{\mathfrak{s}_i}$ *the marginal trace associated with* $\mathfrak{s}_i$ *and* $\boldsymbol{\sigma}$ *the vector obtained using (28) from the MBD such that* $P(X_1 = k_1, \ldots, X_N = k_N) = p_k$ *with* X_i *representing the reliability of* $\mathfrak{s}_i$.

Proof. Since $\Gamma_{\mathbf{A}}(k) = \bigcap_{i=1}^{N} \Gamma_{A_{\mathfrak{s}_i}}(k_i)$, the mass function m defined by (39) is recovered by considering a joint mass function jm defined by

$$jm(\Gamma_{A_{\mathfrak{s}_1}}(k_1), \ldots, \Gamma_{A_{\mathfrak{s}_N}}(k_N)) = p_k,$$

and by allocating the joint mass $jm(\Gamma_{A_{\mathfrak{s}_1}}(k_1), \ldots, \Gamma_{A_{\mathfrak{s}_N}}(k_N))$ to $\bigcap_{i=1}^{N} \Gamma_{A_{\mathfrak{s}_i}}(k_i)$. Besides, it is direct to see that the marginals of jm are the marginal traces $A^{\pi_i}_{\mathfrak{s}_i}$, $i = 1, \ldots, N$ associated with sources $\mathfrak{s}_1, \ldots, \mathfrak{s}_N$.

Hence, the mass function m defined by (39) may be obtained as the conjunctive combination of elementary pieces of evidence represented by the simple mass functions $A^{\pi_i}_{\mathfrak{s}_i}$ with dependence structure jm, or, equivalently, we have

$$m = \textcircled{\scriptsize}_{\boldsymbol{\sigma}}(A^{\pi_1}_{\mathfrak{s}_1}, \ldots, A^{\pi_N}_{\mathfrak{s}_N}),$$

since the MBD in one-to-one correspondence with jm is formally equivalent to the one representing the meta-knowledge about the sources and thus the dependence structure in jm is represented by $\boldsymbol{\sigma}$. □

The fact that the mass function given by (39) can be equivalently obtained with (40) corresponds to the intuition. It means that a situation where some sources provide crisp testimonies $A_{\mathfrak{s}_i}$, are assumed to be marginally non reliable with probabilities π_i and to have a meta-dependence structure $\boldsymbol{\sigma}$, can be equivalently regarded with respect to the knowledge it induces on $\mathcal{Y}$, as a situation where one has accumulated elementary pieces of evidence, represented by the marginal traces $A^{\pi_i}_{\mathfrak{s}_i}$, and one assumes that these pieces of evidence have a dependence structure $\boldsymbol{\sigma}$ among them. In particular, we remark that the dependence structure at the meta level 'flows down' to the trace level, and specifically σ_k, $k_1 + \cdots + k_N > 1$, which represents the dependency among the marginal traces $\{A^{\pi_i}_{\mathfrak{s}_i} : i, k_i = 1\}$ as explained in Section 4.1, is thus nothing but the inheritance of the dependency σ_k between the reliability of the sources $\{\mathfrak{s}_i : i, k_i = 1\}$ that are at the origin of these marginal traces.

An important particular case of this is when the dependence structure at the meta level is that of independence, which corresponds to $\boldsymbol{\sigma} = \mathbf{e}_1$ (in which case Eq. (40) reduces to (21)). Indeed, in such a case where the sources $\mathfrak{s}_1, \ldots, \mathfrak{s}_N$ are meta-independent, then the induced knowledge on $\mathcal{Y}$ is given by (21), as explained in Section 2.3. Yet, according to the conjunctive combination approach recalled in Section 4.1, m defined by (21) is also obtained if one receives N elementary pieces of evidence represented by simple mass functions $A^{\pi_i}_{\mathfrak{s}_i}$, $i = 1, \ldots, N$, *i.e.*, one receives the marginal traces associated with the sources, and assumes that these pieces of evidence are independent. That is, one obtains m by assuming the same dependence structure at the trace level than that at the meta level.

Theorem 2 is further illustrated by Example 7.

Example 7. Let $\mathfrak{s}_1$ and $\mathfrak{s}_2$ be two sources providing the testimonies $A_{\mathfrak{s}_1} = A$ and $A_{\mathfrak{s}_2} = B$, for some $A, B \subseteq \mathcal{Y}$, $A \neq B$. Assume that $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are both reliable with probability p_1, $\mathfrak{s}_1$ is not reliable and $\mathfrak{s}_2$ is reliable with probability p_2, $\mathfrak{s}_1$ is reliable and $\mathfrak{s}_2$ is not reliable with probability p_3, and they are both not reliable with probability p_4, *i.e.*, we have

$$\begin{aligned} P(X_1 = 0, X_2 = 0) &= p_1, \\ P(X_1 = 1, X_2 = 0) &= p_2, \\ P(X_1 = 0, X_2 = 1) &= p_3, \\ P(X_1 = 1, X_2 = 1) &= p_4. \end{aligned}$$

Then, from (39), the induced knowledge on $\mathcal{Y}$ is represented by the mass function m defined as

$$m(A \cap B) = p_1, m(B) = p_2, m(A) = p_3, m(\mathcal{Y}) = p_4.$$

Now, the marginal probabilities that sources $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are non reliable are $\pi_1 = p_2 + p_4$ and $\pi_2 = p_3 + p_4$, respectively. In addition, their meta-dependence structure is $\boldsymbol{\sigma} = (1, 0, 0, \sigma_4)'$, with $\sigma_4 = p_4 \cdot p_1 - p_2 \cdot p_3$. Hence, from Theorem 2, m can be equivalently obtained as the conjunctive combination of the simple mass functions A^{π_1} and B^{π_2} with dependence structure $\boldsymbol{\sigma}$, that is, as

$$m = \textcircled{\scriptsize}_{(1,0,0,\sigma_4)'}(A^{\pi_1}, B^{\pi_2}).$$

Theorem 2 is particularly useful to obtain an expression of the t-decomposition in terms of a conjunctive combination of some simple mass functions:

Theorem 3. *Any mass function* m *defined on a domain* $\mathcal{Y} = \{y_1, \ldots, y_n\}$ *with Teugels function* t *satisfies*

$$m = \textcircled{\scriptsize}_{\boldsymbol{\sigma}}\left(\overline{\{y_1\}}^{t(\{y_1\})}, \ldots, \overline{\{y_n\}}^{t(\{y_n\})}\right), \tag{41}$$

with $\boldsymbol{\sigma}$ *the vector such that* $\sigma_1 = 1$ *and, for* $1 < k \leq 2^n$, $\sigma_k = t(A_k)$ *if* $|A_k| > 1$, *and* $\sigma_k = 0$ *otherwise.*

Proof. As shown in Section 3.3, the t-canonical decomposition allows one to view any mass function on a domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$ as resulting from crisp testimonies $\overline{\{y_i\}}$ provided by sources $\mathfrak{s}_i$, $i = 1, \ldots, n$, assumed to be non reliable with marginal probabilities $t(\{y_i\})$ and having some meta-dependence represented by central moments $t(A_k)$, $|A_k| > 1$. The theorem follows from this latter fact and Theorem 2. □

Theorem 3 shows that according to the t-decomposition, any mass function results from the conjunctive combination of $|\mathcal{Y}|$ elementary pieces of evidence represented by simple mass functions having some dependence structure. This is illustrated by Examples 8 and 9.

Example 8 (Example 4 continued)**.** Let m be a mass function defined on $\mathcal{Y} = \{y_1, y_2\}$, with associated Teugels function t. From Theorem 3, m results from the conjunctive combination of simple mass functions $\overline{\{y_1\}}^{t(\{y_1\})}$ and $\overline{\{y_2\}}^{t(\{y_2\})}$ with dependence structure $\boldsymbol{\sigma} = [1, 0, 0, t(\{y_1, y_2\})]'$, *i.e.*, we have

$$m = \textcircled{\cap}_{[1,0,0,t(\{y_1,y_2\})]'}\left(\overline{\{y_1\}}^{t(\{y_1\})}, \overline{\{y_2\}}^{t(\{y_2\})}\right).$$

Example 9 (Example 5 continued)**.** From Theorem 3, the mass function m of Example 2, with function t provided in Example 5, results from the conjunctive combination of simple mass functions $\overline{\{y_1\}}^{1}$, $\overline{\{y_2\}}^{2/3}$ and $\overline{\{y_3\}}^{2/3}$ with dependence structure $\boldsymbol{\sigma} = (1, 0, 0, 0, 0, 0, -\frac{1}{9}, 0)'$, *i.e.*, we have

$$m = \textcircled{\cap}_{(1,0,0,0,0,0,-\frac{1}{9},0)'}\left(\overline{\{y_1\}}^{1}, \overline{\{y_2\}}^{2/3}, \overline{\{y_3\}}^{2/3}\right).$$

Remark 3 (Bayesian mass function on a binary frame)**.** Let $\boldsymbol{y}$ be a variable defined on $\mathcal{Y} = \{y_1, y_2\}$. Let $\mathfrak{s}_1$ and $\mathfrak{s}_2$ be two sources providing the testimonies $\boldsymbol{y} \in \overline{\{y_1\}} = \{y_2\}$ and $\boldsymbol{y} \in \overline{\{y_2\}} = \{y_1\}$, respectively. Assume that $\mathfrak{s}_1$ is reliable and $\mathfrak{s}_2$ is not reliable with probability α, and that $\mathfrak{s}_1$ is not reliable and $\mathfrak{s}_2$ is reliable with probability $1 - \alpha$, for some $\alpha \in [0, 1]$, *i.e.*, we have

$$\begin{aligned}
&P(X_1 = 0, X_2 = 0) = 0,\\
&P(X_1 = 1, X_2 = 0) = 1 - \alpha,\\
&P(X_1 = 0, X_2 = 1) = \alpha,\\
&P(X_1 = 1, X_2 = 1) = 0.
\end{aligned}$$

Then, from (39), the induced knowledge on $\mathcal{Y}$ is represented by the Bayesian mass function m such that

$$m(\{y_1\}) = 1 - \alpha, \quad m(\{y_2\}) = \alpha. \tag{42}$$

The marginal probabilities that sources $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are non reliable are $\pi_1 = 1 - \alpha$ and $\pi_2 = \alpha$, respectively. In addition, their meta-dependence structure is $\boldsymbol{\sigma} = (1, 0, 0, \alpha^2 - \alpha)'$. Hence, from Theorem 2, m can be equivalently obtained as

$$m = \textcircled{\cap}_{(1,0,0,\alpha^2-\alpha)'}(\overline{\{y_1\}}^{1-\alpha}, \overline{\{y_2\}}^{\alpha}),$$

which is nothing but the t-decomposition of the Bayesian mass function (42).

Now, suppose we are given only the marginal probabilities $\pi_1 = 1 - \alpha$ and $\pi_2 = \alpha$. Then, assume that there is a *perfect dependence* [34] between $\mathfrak{s}_1$ being reliable and $\mathfrak{s}_2$ being not reliable, which means that the probability $P(X_1 = 0, X_2 = 1)$ that $\mathfrak{s}_1$ is reliable and $\mathfrak{s}_2$ is not reliable is such that

$$P(X_1 = 0, X_2 = 1) = P(X_1 = 0) \wedge P(X_2 = 1) = (1 - \pi_1) \wedge \pi_2,$$

where $\wedge$ denotes the minimum operator. This latter assumption is equivalent[4] to assuming $\boldsymbol{\sigma} = (1, 0, 0, \alpha^2 - \alpha)'$, since by definition $\sigma_1 = 1$, $\sigma_2 = 0$, $\sigma_3 = 0$ and from (27)

$$\begin{aligned}
P(X_1 = 0, X_2 = 1) &= (1 - \pi_1)\pi_2 - \sigma_4\\
\Leftrightarrow \sigma_4 &= (1 - \pi_1)\pi_2 - (1 - \pi_1) \wedge \pi_2\\
&= \alpha^2 - \alpha.
\end{aligned}$$

Hence, any Bayesian mass function on a binary domain is obtained, according to the t-decomposition, from two sources such that the marginal probability that the first source is reliable is equal to the marginal probability that the second source is not reliable, and such that there is a perfect dependence between the first source being reliable and the second source being not reliable.

Remark 4. The cautious rule $\textcircled{\wedge}$, which is based on Smets' decomposition, was proposed in [10] for the conjunctive combination of bodies of evidence, which cannot be assumed to be independent. The definition of this rule for the combination of two simple mass functions is the following. For all $A, B \subset \mathcal{Y}$ such that $A \neq B$, and $\pi_1, \pi_2 \in (0, 1]$, we have

$$A^{\pi_1} \textcircled{\wedge} A^{\pi_2} = A^{\pi_1 \wedge \pi_2}, \tag{43}$$

[4] Vector $\boldsymbol{\sigma}$ is equivalently obtained by three other assumptions: (i) perfect dependence between $\mathfrak{s}_1$ being not reliable and $\mathfrak{s}_2$ being reliable, (ii) *opposite dependence* [34] between $\mathfrak{s}_1$ being reliable and $\mathfrak{s}_2$ being reliable, (iii) opposite dependence between $\mathfrak{s}_1$ being not reliable and $\mathfrak{s}_2$ being not reliable.

$$A^{\pi_1} \textcircled{\wedge} B^{\pi_2} = A^{\pi_1} \textcircled{\cap} B^{\pi_2}. \tag{44}$$

The behaviour of the cautious rule can be analysed in the light of our framework of meta-dependent and partially reliable sources. Specifically, viewing A^{π_i} as the trace of a source $\mathfrak{s}_i$ providing testimony $\boldsymbol{y} \in A$ and assumed to be not reliable with marginal probability π_i, $i = 1, 2$, then Eq. (43) is recovered by assuming that there is a *perfect dependence* between $\mathfrak{s}_1$ not being reliable and $\mathfrak{s}_2$ not being reliable, which is equivalent to assuming that their meta-dependence is $\boldsymbol{\sigma} = (1, 0, 0, \pi_1 \wedge \pi_2 - \pi_1\pi_2)'$. From Theorem 2, we obtain then:

$$A^{\pi_1} \textcircled{\wedge} A^{\pi_2} = \textcircled{\cap}_{(1,0,0,\pi_1\wedge\pi_2-\pi_1\pi_2)'}(A^{\pi_1}, A^{\pi_2}).$$

Furthermore, viewing B^{π_2} as the trace of a source $\mathfrak{s}_3$ providing testimony $\boldsymbol{y} \in B$ and assumed to be not reliable with marginal probability π_2, then Eq. (44) is recovered by assuming *independence* between $\mathfrak{s}_1$ not being reliable and $\mathfrak{s}_3$ not being reliable, which is equivalent to assuming that their meta-dependence is $\boldsymbol{\sigma} = \mathbf{e}_1$. In other words, we have:

$$A^{\pi_1} \textcircled{\wedge} B^{\pi_2} = \textcircled{\cap}_{\mathbf{e}_1}(A^{\pi_1}, B^{\pi_2}).$$

Hence, the combination by the cautious rule of two elementary pieces of evidence is recovered as different particular cases of the rule $\textcircled{\cap}_\sigma$, that is, by assuming different dependence structures among these pieces of evidence – the dependence structure to be used in a given situation being determined by whether the pieces of evidence support the same subset or not.

4.3. Comparison with Smets' decomposition

(i). The *t*-canonical decomposition of belief functions bears some resemblance with that of Smets, in that both of these decompositions view a belief function defined on a domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$ as resulting from partially reliable sources providing crisp pieces of information about a variable of interest $\boldsymbol{y}$. As a matter of fact, these decompositions coincide in the following special case.

Let m be a mass function with Teugels function t such that for all $A \subseteq \mathcal{Y}$, $|A| > 1$, $t(A) = 0$. Then, we have

$$\begin{aligned} m &= \textcircled{\cap}_{\mathbf{e}_1}\left(\overline{\{y_1\}}^{t(\{y_1\})}, \ldots, \overline{\{y_n\}}^{t(\{y_n\})}\right), \\ &= \bigcirc\!\!\!\!\!\!\cap_{i=1}^{n} \overline{\{y_i\}}^{t(\{y_i\})}. \end{aligned} \tag{45}$$

Mass functions satisfying (45) will be called $\mathbf{e}_1$-separable hereafter. Moreover, for mass function m, Smets' decomposition yields:

$$m = \bigcirc\!\!\!\!\!\!\cap_{i=1}^{n} \overline{\{y_i\}}^{w(\overline{\{y_i\}})},$$

with $w(\overline{\{y_i\}}) = t(\{y_i\})$, $i = 1, \ldots, n$. Hence, according to both decompositions, any $\mathbf{e}_1$-separable mass function m is obtained from the conjunctive combination of $|\mathcal{Y}|$ independent elementary pieces of evidence represented by simple mass functions $\overline{\{y_i\}}^{t(\{y_i\})}$. In sum, the decompositions coincide for $\mathbf{e}_1$-separable mass functions.

In contrast, it may be interesting to note that the decompositions do not coincide in general for u-separable mass functions as shown by Example 10.

Example 10. Let m be a u-separable mass function on $\mathcal{Y} = \{y_1, y_2\}$ such that $m = \emptyset^{w(\emptyset)} \textcircled{\cap} \{y_1\}^{w(\{y_1\})}$, with $w(\emptyset), w(\{y_1\}) \in (0, 1)$. We may remark that

$$m = \textcircled{\cap}_{\mathbf{e}_1}\left(\emptyset^{w(\emptyset)}, \{y_1\}^{w(\{y_1\})}\right),$$

but that does not mean that m is $\mathbf{e}_1$-separable. Indeed, for m to be $\mathbf{e}_1$-separable, we must have $m = \textcircled{\cap}_{\mathbf{e}_1}\left(\overline{\{y_1\}}^{t(\{y_1\})}, \overline{\{y_2\}}^{t(\{y_2\})}\right)$.

Function t associated with m is

$$t(\{y_1\}) = w(\emptyset), \quad t(\{y_2\}) = w(\emptyset)w(\{y_1\}), \quad t(\{y_1, y_2\}) = w(\emptyset)w(\{y_1\})(1 - w(\emptyset)).$$

Hence, $m = \textcircled{\cap}_{[1,0,0,w(\emptyset)w(\{y_1\})(1-w(\emptyset))]'}\left(\overline{\{y_1\}}^{w(\emptyset)}, \overline{\{y_2\}}^{w(\emptyset)w(\{y_1\})}\right)$, that is, m results according to the *t*-decomposition from the conjunctive combination of simple mass functions $\overline{\{y_1\}}^{w(\emptyset)}$ and $\overline{\{y_2\}}^{w(\emptyset)w(\{y_1\})}$ with dependence structure $\boldsymbol{\sigma} = [1, 0, 0, w(\emptyset)w(\{y_1\})(1 - w(\emptyset))]'$, which is different from Smets' decomposition for which m results from the conjunctive combination of the independent simple mass functions $\emptyset^{w(\emptyset)}$ and $\{y_1\}^{w(\{y_1\})}$.

The fact that the decompositions coincide for $\mathbf{e}_1$-separable mass functions and do not coincide for u-separable mass functions actually follows from Lemma 2.

Lemma 2. *Let m be a mass function on $\mathcal{Y}$. We have*

- *m is* $\mathbf{e}_1$*-separable* $\Rightarrow$ *m is u-separable.*
- *m is* $\mathbf{e}_1$*-separable* $\not\Leftarrow$ *m is u-separable.*

Proof. $\Rightarrow$ follows from the definitions of $\mathbf{e}_1$-separability and u-separability.

$\not\Leftarrow$: Example 10 shows that a u-separable mass function m admits the t-decomposition $m = \textcircled{\cap}_{[1,0,0,w(\emptyset)w(\{y_1\})(1-w(\emptyset))]'}\left(\overline{\{y_1\}}^{w(\emptyset)}, \overline{\{y_2\}}^{w(\emptyset)w(\{y_1\})}\right) \neq \textcircled{\cap}_{\mathbf{e}_1}\left(\overline{\{y_1\}}^{w(\emptyset)}, \overline{\{y_2\}}^{w(\emptyset)w(\{y_1\})}\right)$. □

In general, Smets' decomposition and the t-decomposition significantly differ on various aspects. First, Smets' decomposition involves $2^n - 1$ sources while the t-decomposition involves n sources, and the sources are in general not meta-independent in the t-decomposition whereas Smets' decomposition involves meta-independence. In terms of elementary pieces of evidence, this means that the t-decomposition breaks down a mass function as the result of the conjunctive combination of n simple mass functions that have a dependency structure, whereas with Smets' decomposition it is the result of the conjunctive combination of $2^n - 1$ (generalised) simple mass functions that are independent. Second, the t-decomposition can be obtained for any mass function, whereas Smets' decomposition is restricted to non dogmatic mass function.[5] Third and most importantly, the t-decomposition involves only well-defined concepts with clear semantics, in particular means and central moments of Bernoulli variables, which is not the case of Smets' decomposition.

(ii). Smets' decomposition has a nice behaviour (13) with respect to the unnormalised Dempster's rule. It seems thus interesting to comment on the behaviour of our solution with respect to this rule. It is actually easy to uncover the assumptions associated with the use of this rule in our solution since Dempster's rule was originally introduced in a similar setting as ours [25]. Let m_1 and m_2 be two mass functions defined on $\mathcal{Y} = \{y_1, \ldots, y_n\}$. In our approach, mass function m_j, $j = 1, 2$, can be viewed as resulting from having n sources S_i^j, $i = 1, \ldots, n$, each telling $\boldsymbol{y} \in \overline{\{y_i\}}$, and such that uncertainty with respect to the reliability of these sources is represented by a probability distribution, with p_{k^j} the probability that sources S_i^j, $i = 1, \ldots, n$, are in the joint state k^j. Assuming that the sources S_i^1, $i = 1, \ldots, n$, underlying m_1 are meta-independent from the sources S_i^2, $i = 1, \ldots, n$, underlying m_2, we obtain that the joint probability $p_{(k^1,k^2)}$ that sources S_i^1 and S_i^2, $i = 1, \ldots, n$, are in the joint state (k^1, k^2) is $p_{(k^1,k^2)} = p_{k^1} \cdot p_{k^2}$. Besides, if the sources S_i^1 and S_i^2, $i = 1, \ldots, n$, are in the joint state (k^1, k^2), then we should deduce that $\boldsymbol{y} \in \mathrm{T}(k^1) \cap \mathrm{T}(k^2)$. Hence assuming meta-independence between the sources underlying m_1 and m_2 yield the mass function m_{12} such that

$$m_{12}(A) = \sum_{k^1,k^2:\mathrm{T}(k^1)\cap\mathrm{T}(k^2)=A} p_{(k^1,k^2)}. \tag{46}$$

It can easily be checked that $m_{12} = m_{1\textcircled{\cap}2}$ and thus the unnormalised Dempster's rule has a nice interpretation in our approach: it corresponds simply to assuming meta-independence between the sources underlying m_1 and the sources underlying m_2.

In addition, let us remark that the mass function $m_{1\textcircled{\cap}2}$ can be viewed using our solution as originating from n sources, which we denote by S_i^{12}, $i = 1, \ldots, n$, and an associated probability distribution p^{12} representing uncertainty with respect to the reliability of these latter sources. We may remark that due to (7) the vector $\boldsymbol{\mu}^{12}$ of moments associated with p^{12} is such that $\mu^{12} = \mu^1 \cdot \mu^2$, with $\boldsymbol{\mu}^i$ the vector of moments associated with the MBD-equivalent of m_i, $i = 1, 2$. In other words, any moment between the reliabilities of the sources underlying $m_{1\textcircled{\cap}2}$ is equal to the pointwise product of the same moments between the reliabilities of the sources underlying m_1 and m_2. This means that in our approach the unnormalised Dempster's rule has a similar behaviour to the one it has in Smets' solution: it comes down to a simple pointwise product of some functions (μ^1 and μ^2 in our case, w_1 and w_2 in Smets' case).

4.4. t-*Canonical decomposition of random sets*

A random set is a random element taking values as subsets of some space [18,19]. A random set is thus defined in the finite case by a probability distribution m on the power set $2^{\mathcal{X}}$ of some finite set $\mathcal{X} = \{x_1, \ldots, x_n\}$ such that $\sum_{A\subseteq\mathcal{X}} m(A) = 1$. Distribution m is formally equivalent to a mass function on $\mathcal{X}$ [27], but it has different semantics [35]. For instance, borrowing from [21], let $\mathcal{X} = \{\text{English, French, Spanish}\}$ denote the set of languages that a person can speak, then $m(A)$ is the probability that someone speaks *exactly* all the languages in A (and not other ones), $A \subseteq \mathcal{X}$. Furthermore, the commonality degree $q(A) = \sum_{B\supseteq A} m(B)$ is the probability that someone speaks *at least* all the languages in A; for instance, the probability $q(\{\text{English}\}$ that someone speaks at least English is obtained by summing the probabilities $m(\{\text{English}\}$ of speaking only English, $m(\{\text{English, French}\}$ of speaking only English and French, $m(\{\text{English, Spanish}\}$ of speaking only English and Spanish, and $m(\{\text{English, French, Spanish}\}$ of speaking the three languages.

Definition 4 (MBD-RS-equivalent)**.** Let m be the probability distribution of some finite random set, defined on the power set of some set $\mathcal{X} = \{x_1, \ldots, x_n\}$ and let p be the MBD (22) such that $p_k = m(A_k)$, $1 \leq k \leq 2^n$, with A_k the kth subset of $\mathcal{X}$ according to the binary order. p is called the *MBD-Random Set-equivalent* (MBD-RS-equivalent for short) of m.

[5] Smets [9] proposed a technical means to decompose a dogmatic mass function m, which consists essentially in assigning an ϵ to $\mathcal{Y}$. However, this comes down to approximating m by a non dogmatic mass function.

Although they seem fairly similar, the bond between the probability distribution of a random set and its MBD-RS-equivalent is much stronger than that of a mass function and its MBD-equivalent. Indeed, the notion of MBD-equivalent becomes interesting mostly in conjunction with Proposition 1. In contrast, the semantics of a random set and its MBD-RS-equivalent are in immediate correspondence. For instance, let m be the preceding probability distribution representing the languages that someone speaks among the $n = 3$ languages English, French and Spanish. Furthermore, let p be the MBD-RS-equivalent of m, where Bernoulli random variables X_1, X_2 and X_3 underlying p represent respectively whether someone speaks English, French and Spanish, with $X_i = 1$ meaning that someone speaks the ith language and $X_i = 0$ meaning that someone does not speak the ith language. Then p_k has the same interpretation as $m(A_k)$, $1 \leq k \leq 2^n$: it is the probability that someone speaks exactly all the languages in A_k (and not other ones). We have for instance $A_4 = \{\text{English, French}\}$ and the probability $m(A_4)$ that someone speaks only English and French is equal to $p_4 = p_{110} = P(X_1 = 1, X_2 = 1, X_3 = 0)$.

It is clear that the vector $\boldsymbol{\mu}$ computed through (26) from the MBD-RS-equivalent p of the probability distribution m of a random set, is the vector $\mathbf{q}$ of commonality degrees associated to m. Most importantly, thanks to the Teugels' representation τ of p, we can propose a canonical decomposition of a random set, which has a well-defined semantics. This canonical decomposition is given by function t computed using (37). For instance, continuing the language example, let m be such that

$$m(A_4) = m(A_6) = m(A_8) = 1/3, \tag{47}$$

with $A_4 = \{\text{English, French}\}$, $A_6 = \{\text{English, Spanish}\}$ and $A_8 = \{\text{English, French, Spanish}\}$. We have

$$\begin{aligned} t(\{\text{English}\} &= 1, \\ t(\{\text{French}\} &= 2/3, \\ t(\{\text{English,French}\} &= 0, \\ t(\{\text{Spanish}\} &= 2/3, \\ t(\{\text{English,Spanish}\} &= 0, \\ t(\{\text{French,Spanish}\} &= -1/9, \\ t(\{\text{English,French,Spanish}\} &= 0. \end{aligned}$$

Hence, the random set defined by m can be seen as originating from the following pieces of information

- Someone speaks English, French and Spanish with probabilities 1, 2/3 and 2/3 respectively;
- The covariance between someone speaking French and speaking Spanish is $-1/9$, and all other subsets of variables representing the languages spoken by someone have null central moments.

Let us conclude this section by remarking that the correspondence established between finite random sets and the MBD can be used together with the multi-valued mapping T (24) to relate random sets and belief functions in a different way than in [27], and specifically as a means to obtain a belief function from a random set. Indeed, let p be the MBD-RS-equivalent of the probability distribution m of some random set, defined on $2^{\mathcal{X}}$ with $\mathcal{X} = \{x_1, \ldots, x_n\}$. Then, transferring MBD p via multi-valued mapping T (24) yield a mass function m' on $\mathcal{Y} = \{y_1, \ldots, y_n\}$ defined, for $1 \leq k \leq 2^n$, by

$$m'(A_k) = \sum_{k' \in \mathcal{X}^n : \mathrm{T}(k') = A_k} p_{k'},$$

with A_k the kth subset of $\mathcal{Y}$. For instance, consider again probability distribution m defined by (47) and let p be its MBD-RS-equivalent, where Bernoulli random variables X_i, $i = 1, 2, 3$, underlying p represent respectively whether someone speaks English, French and Spanish. Let $\boldsymbol{y}$ be a variable defined on

$$\begin{aligned} \mathcal{Y} &= \{\text{British citizen, French citizen, Spanish citizen}\} \\ &= \{y_1, y_2, y_3\} \end{aligned}$$

and denoting the citizenship of someone (we assume someone has only one citizenship). Furthermore, assume the following pieces of information:

- Someone who does not speak the language of a given country, is surely not a citizen of that country;
- A citizen of a given country may speak the language of another country (*e.g.*, a Spanish man may speak French).

For each language i, $i = 1, 2, 3$, these pieces of information can be represented by a multi-valued mapping Γ_i defined by (23). For instance, we have $\Gamma_2(0) = \{y_1, y_3\}$ and $\Gamma_2(1) = \mathcal{Y}$ since someone who does not speak French ($X_2 = 0$) is surely not a French citizen ($\boldsymbol{y} \in \{y_1, y_3\}$), and since someone who speaks French ($X_2 = 1$) may be a citizen of any country ($\boldsymbol{y} \in \mathcal{Y}$). More generally, what can be deduced about the citizenship of someone given the languages that she speaks can be represented by multi-valued mapping T (24). For instance, someone who speaks English and French but not Spanish is a British citizen or a French citizen, as

$$\begin{aligned} \mathrm{T}[(1,1,0)] &= \Gamma_1(1) \bigcap \Gamma_2(1) \bigcap \Gamma_3(0) \\ &= \mathcal{Y} \cap \mathcal{Y} \cap \overline{\{y_3\}} \\ &= \{y_1, y_2\}. \end{aligned}$$

Hence, if knowledge about the languages spoken by someone is represented by probability distribution m defined by (47), we obtain that the citizenship of someone is represented by mass function m' defined on $\mathcal{Y}$ by $m'(\{y_1, y_2\}) = m'(\{y_1, y_3\}) = m'(\mathcal{Y}) = 1/3$.

5. A MBD-based perspective on the weight function

We have argued that the semantics provided by Smets for the function w, as weights associated with statements of the form *believe* and *do not believe* in some propositions, is not totally satisfactory. However, this does not mean that this function does not represent some facet of the information contained in a mass function, similarly as, *e.g.*, *pl* or *bel* – it just means that the interpretation given by Smets for this function as representing a canonical decomposition, that is, a decomposition into elementary pieces of evidence, is not totally acceptable, at least for the time being, and deserves further justifications.

As a matter of fact, a completely new perspective on the weight function w associated to a mass function m is brought to light in this section, using measures of information associated with the MBD-equivalent of m. As will be seen, the proposed interpretation is not as appealing as Smets', and especially is unrelated to the concept of canonical decomposition, but it relies only on well-defined concepts. Then, a similar result is provided for the disjunctive counterpart of w known as the *disjunctive* weight function [10] and a disjunctive counterpart of the expression (41) is also unveiled for the t-decomposition.

5.1. Weights as measures of information

Instead of using function w, Smets' decomposition (11) can be equivalently presented using a function $s : 2^{\mathcal{Y}} \setminus \{\mathcal{Y}\} \to (-\infty, +\infty)$ such that $s(A) = -\ln w(A)$ for all $A \subset \mathcal{Y}$ (see, *e.g.*, [36]). We note that simple mass functions correspond to the case where $s(A) \geq 0$ and that inverse simple mass functions correspond to the case where $s(A) < 0$. Function s is actually the function Shafer originally used in his monograph [1] to present the canonical decomposition of separable mass functions (9). Accordingly and following [36], $s(A)$, $A \subset \mathcal{Y}$, may be referred to as *Shafer's weights*.

Let p be the MBD-equivalent of some mass function m and let μ be the moments of p. By replacing q in (12) by μ, we obtain the following expression for function s, for all $A_k \subset \mathcal{Y}$:

$$\begin{aligned} s(A_k) &= -\ln \left(\prod_{A_{k'} \supseteq A_k} q(A_{k'})^{(-1)^{|A_{k'}|-|A_k|+1}} \right), \\ &= -\ln \left(\prod_{A_{k'} \supseteq A_k} \mu_{k'}^{(-1)^{|A_{k'}|-|A_k|+1}} \right). \end{aligned} \tag{48}$$

In particular, let m be a mass function on $\mathcal{Y} = \{y_1, y_2\}$. We have:

$$\begin{aligned} s(\emptyset) &= -\ln \frac{\mu_2 \mu_3}{\mu_1 \mu_4} \\ &= -\ln \frac{\mu_2 \mu_3}{\mu_4} \\ &= \ln \frac{\mu_4}{\mu_2 \mu_3}, \\ s(\{y_1\}) &= -\ln \frac{\mu_4}{\mu_2}, \\ s(\{y_2\}) &= -\ln \frac{\mu_4}{\mu_3}. \end{aligned}$$

Interestingly, these latter combinations of the marginal probabilities $\mu_{k'}$, and thus Shafer's weights, correspond to known quantities associated with the MBD. This is detailed hereafter.

Let X_1 and X_2 be two Bernoulli random variables and let p be the MBD such that $p_k = p_{k_1,k_2}$, $1 \leq k \leq 4$, with $p_{k_1,k_2} := P(X_1 = k_1, X_2 = k_2)$ where $k_i \in \{0, 1\}$, $i = 1, 2$, are the terms in the binary expansion of k. Furthermore, let $\boldsymbol{\mu}$ be the vector of moments of p. Recall (cf [20, p. 28]) that the *mutual information* $I(X_1 = k_1; X_2 = k_2)$ between two events $X_1 = k_1$ and $X_2 = k_2$ is[6]

$$\begin{aligned} I(X_1 = k_1; X_2 = k_2) &= \ln \frac{P(X_1 = k_1, X_2 = k_2)}{P(X_1 = k_1) P(X_2 = k_2)} \\ &= \ln \frac{p_k}{(\sum_{k' : k'_1 = k_1} p_{k'})(\sum_{k' : k'_2 = k_2} p_{k'})}. \end{aligned}$$

[6] Usually, log rather than ln is used, but this is just a change of unit of information: the unit is bits when using log and nats when using ln [20].

Besides, the *conditional self information* $I(X_1 = k_1|X_2 = k_2)$ of event $X_1 = k_1$ given event $X_2 = k_2$ is [20, p. 36]

$$\begin{aligned} I(X_1 = k_1|X_2 = k_2) &= -\ln \frac{P(X_1 = k_1, X_2 = k_2)}{P(X_2 = k_2)} \\ &= -\ln \frac{p_k}{\sum_{k':k'_2=k_2} p_{k'}}, \end{aligned}$$

and the conditional self information $I(X_2 = k_2|X_1 = k_1)$ of event $X_2 = k_2$ given event $X_1 = k_1$ is

$$\begin{aligned} I(X_2 = k_2|X_1 = k_1) &= -\ln \frac{P(X_1 = k_1, X_2 = k_2)}{P(X_1 = k_1)} \\ &= -\ln \frac{p_k}{\sum_{k':k'_1=k_1} p_{k'}}. \end{aligned}$$

In particular, we have for $k = 4 \leftrightarrow (k_1 = 1, k_2 = 1)$

$$\begin{aligned} I(X_1 = 1; X_2 = 1) &= \ln \frac{p_4}{(\sum_{k':k'_1=1} p_{k'})(\sum_{k':k'_2=1} p_{k'})} \\ &= \ln \frac{\mu_4}{\mu_2\mu_3}, \\ I(X_2 = 1|X_1 = 1) &= -\ln \frac{p_4}{\sum_{k':k'_1=1} p_{k'}} \\ &= -\ln \frac{\mu_4}{\mu_2}, \\ I(X_1 = 1|X_2 = 1) &= -\ln \frac{p_4}{\sum_{k':k'_2=1} p_{k'}} \\ &= -\ln \frac{\mu_4}{\mu_3}. \end{aligned}$$

Hence, when m is a mass function defined on the binary domain $\mathcal{Y} = \{y_1, y_2\}$, with MBD-equivalent p, we have

$$\begin{aligned} s(\emptyset) &= I(X_1 = 1; X_2 = 1), \\ s(\{y_1\}) &= I(X_2 = 1|X_1 = 1), \\ s(\{y_2\}) &= I(X_1 = 1|X_2 = 1). \end{aligned}$$

Adopting the setting of Remark 1, $s(\emptyset)$ is then equal to the mutual information between both sources underlying m not being reliable, $s(\{y_1\})$ is equal to the conditional self information of the second source not being reliable given that the first source is not reliable, and $s(\{y_2\})$ is equal to the conditional self information of the first source not being reliable given that the second source is not reliable.

Remark that for any mass function m on $\mathcal{Y} = \{y_1, y_2\}$, we have $s(\{y_i\}) = -\ln \frac{q(\{y_1,y_2\})}{q(\{y_i\})} \geq 0$, $i = 1, 2$, and $s(\emptyset) = \ln \frac{q(\{y_1,y_2\})}{q(\{y_1\})q(\{y_2\})} \in (-\infty, +\infty)$. In other words, when $\mathcal{Y}$ is binary, only for $A = \emptyset$ one can have $s(A) < 0$, which is a case of debt of belief for A according to Smets' interpretation of s. Our new perspective on s provides a completely different meaning to $s(\emptyset) < 0$ than that of debt of belief, and especially a well-established meaning which is that of mutual information. More generally, our alternative interpretation for the whole function s is admittedly less appealing than the one proposed by Smets, but it is rigorous.

When m is a mass function on $\mathcal{Y} = \{y_1, y_2, y_3\}$, its associated weight function s can also be interpreted using measures of information, as explained hereafter. The MBD-equivalent p of m relies in this case on three Bernoulli random variables $\{X_1, X_2, X_3\}$. Let us recall that the mutual information $I(X_1 = 1; X_2 = 1; X_3 = 1)$ between the three events $X_1 = 1$, $X_2 = 1$ and $X_3 = 1$ is [20, p. 57–58]

$$\begin{aligned} I(X_1 = 1; X_2 = 1; X_3 = 1) &= \ln \frac{P(X_1 = 1, X_2 = 1)P(X_1 = 1, X_3 = 1)P(X_2 = 1, X_3 = 1)}{P(X_1 = 1)P(X_2 = 1)P(X_3 = 1)P(X_1 = 1, X_2 = 1, X_3 = 1)} \\ &= \ln \frac{\mu_4\mu_6\mu_7}{\mu_2\mu_3\mu_5\mu_8}. \end{aligned}$$

with $\boldsymbol{\mu}$ the vector of moments of p. The *conditional mutual information* $I(X_2 = 1; X_3 = 1|X_1 = 1)$ between events $X_2 = 1$ and $X_3 = 1$ given event $X_1 = 1$ is [20, p. 28]

$$\begin{aligned} I(X_2 = 1; X_3 = 1|X_1 = 1) &= \ln \frac{P(X_1 = 1)P(X_1 = 1, X_2 = 1, X_3 = 1)}{P(X_1 = 1, X_2 = 1)P(X_1 = 1, X_3 = 1)} \\ &= \ln \frac{\mu_2\mu_8}{\mu_4\mu_6}. \end{aligned}$$

The conditional self information $I(X_3 = 1 |\ X_1 = 1, X_2 = 1)$ of event $X_3 = 1$ given events $X_1 = 1$ and $X_2 = 1$ is

$$\begin{aligned} I(X_3 = 1 |\ X_1 = 1, X_2 = 1) &= -\ln \frac{P(X_1 = 1, X_2 = 1, X_3 = 1)}{P(X_1 = 1, X_2 = 1)} \\ &= -\ln \frac{\mu_8}{\mu_4}. \end{aligned}$$

For m a mass function on $\mathcal{Y} = \{y_1, y_2, y_3\}$, we have thus

$$\begin{aligned} s(\emptyset) &= -\ln \frac{\mu_2\mu_3\mu_5\mu_8}{\mu_1\mu_4\mu_6\mu_7} \\ &= I(X_1 = 1; X_2 = 1; X_3 = 1), \text{ (since} \mu_1 = 1) \\ s(\{y_1\}) &= -\ln \frac{\mu_4\mu_6}{\mu_2\mu_8} \\ &= I(X_2 = 1; X_3 = 1 | X_1 = 1), \\ s(\{y_1, y_2\}) &= -\ln \frac{\mu_8}{\mu_4} \\ &= I(X_3 = 1 |\ X_1 = 1, X_2 = 1), \end{aligned}$$

and similarly we obtain

$$\begin{aligned} s(\{y_2\}) &= I(X_1 = 1; X_3 = 1 | X_2 = 1), \\ s(\{y_3\}) &= I(X_1 = 1; X_2 = 1 | X_3 = 1), \\ s(\{y_1, y_3\}) &= I(X_2 = 1 |\ X_1 = 1, X_3 = 1), \\ s(\{y_2, y_3\}) &= I(X_1 = 1 |\ X_2 = 1, X_3 = 1). \end{aligned}$$

Hence, for instance, $s(\{y_2\})$ is equal to the conditional mutual information between the first and third sources underlying m not being reliable given that the second source is not reliable.

The above interpretation for function s in terms of measures of information can be extended to belief functions defined on domains $\mathcal{Y}$ of any cardinality. This extension is based on the notion of conditional mutual information between an arbitrary number of events [20] recalled hereafter.

Let $\{V_i : i = 1, \ldots, \ell\}$ be a sequence of discrete random variables with ranges $\mathcal{V}_i$, $i = 1, \ldots, \ell$. Consider the multivariate distribution $P(V_1 = v_1, \ldots, V_\ell = v_\ell)$ with $v_i \in \mathcal{V}_i, i = 1, \ldots, \ell$. Then, the *conditional mutual information* between events $V_i = v_i, i = 1, \ldots, \ell - 1$, given event $V_\ell = v_\ell$, for some $v_i \in \mathcal{V}_i, i = 1, \ldots, \ell$, is [20, p. 58]

$$I(V_1 = v_1; \ldots; V_{\ell-1} = v_{\ell-1} | V_\ell = v_\ell) = \ln \frac{\prod_{E \subseteq \Xi, (V_\ell = v_\ell) \in E, |E| \notin 2\mathbb{N}} P(E)}{\prod_{E \subseteq \Xi, (V_\ell = v_\ell) \in E, |E| \in 2\mathbb{N}} P(E)}, \tag{49}$$

with $\Xi := \{V_i = v_i : i = 1, \ldots, \ell\}$. It will be convenient to denote the conditional mutual information between all the events belonging to some set $\mathcal{S}$, given another event E, as $I(Mut(\mathcal{S}))|E)$. For instance, the quantity (49) may be equivalently denoted as $I(Mut(\mathcal{S})|V_\ell = v_\ell)$ with $\mathcal{S} := \{V_i = v_i : i = 1, \ldots, \ell - 1\}$.

Theorem 4. *Let m be a mass function on $\mathcal{Y} = \{y_1, \ldots, y_n\}$ with MBD-equivalent p and with $\{X_i : i = 1, \ldots, n\}$ the Bernoulli random variables underlying p. We have*

$$s(A_k) = I(Mut(\mathcal{S}_k)|C_k = 1), \quad 1 \le k < 2^n,$$

with, using $k \leftrightarrow (k_1, \ldots, k_n)$, $C_k := \prod_{i=1}^n X_i^{k_i}$ and $\mathcal{S}_k := \{X_i = 1 : i, k_i = 0\}$.

Proof. Eq. (48) can be rewritten as

$$s(A_k) = \sum_{A_{k'} \supseteq A_k} (-1)^{|A_{k'}| - |A_k|} \ln \mu_{k'}. \tag{50}$$

For all $A_{k'}, A_k \subseteq \mathcal{Y}$, we have $A_{k'} \supseteq A_k \Leftrightarrow \forall 1 \le i \le n$ s.t. $k_i = 1$, we have $k'_i = 1$, using $k \leftrightarrow (k_1, \ldots, k_n)$ and $k' \leftrightarrow (k'_1, \ldots, k'_n)$.

Let $1 \le k \le 2^n$ and $1 \le k' \le 2^n$. If $\forall 1 \le i \le n$ s.t. $k_i = 1$, we have $k'_i = 1$, we will write $k' \sqsupseteq k$. Hence, for any $A_{k'}, A_k \subseteq \mathcal{Y}$, $A_{k'} \supseteq A_k \Leftrightarrow k' \sqsupseteq k$.

Let $|k| := |\{i : i, k_i = 1\}|$. Hence, for all $A_k \subseteq \mathcal{Y}$, we have $|A_k| = |k|$.

Using these notations, Eq. (50) can be rewritten as

$$s(A_k) = \sum_{k' \sqsupseteq k} (-1)^{|k'| - |k|} \ln \mu_{k'}. \tag{51}$$

For $1 \le k < 2^n$, let $\Xi_k := \mathcal{S}_k \cup \{C_k = 1\}$. Since for $1 \le k < 2^n$, $C_k = 1$ is equivalent to the event $\{X_i = 1 : i, k_i = 1\}$ with $k \leftrightarrow (k_1, \ldots, k_n)$, any $E \subseteq \Xi_k$ s.t. $(C_k = 1) \in E$ is equivalent to the event $E \backslash \{(C_k = 1)\} \cup \{X_i = 1 : i, k_i = 1\}$ and has thus probability $P(E) = \mu_{k'}$ with $\boldsymbol{\mu}$ the vector of moments of the MBD p and $k' \leftrightarrow (k'_1, \ldots, k'_n)$ s.t., for all $1 \le i \le n$, $k'_i = 1$ if $k_i = 1$ or $(X_i =$

$1) \in E \backslash \{(C_k = 1)\}$, and $k'_i = 0$ otherwise. Hence, for the conditional mutual information between events in $\mathcal{S}_k$ given event $C_k = 1$, we have

$$\begin{aligned} I(Mut(\mathcal{S}_k)|C_k = 1) &= \ln \frac{\prod_{E \subseteq \Xi_k, (C_k=1) \in E, |E| \notin 2\mathbb{N}} P(E)}{\prod_{E \subseteq \Xi_k, (C_k=1) \in E, |E| \in 2\mathbb{N}} P(E)} \\ &= \ln \frac{\prod_{k' \sqsupseteq k, |k'|-|k| \in 2\mathbb{N}} \mu_{k'}}{\prod_{k' \sqsupseteq k, |k'|-|k| \notin 2\mathbb{N}} \mu_{k'}} \\ &= \sum_{k' \sqsupseteq k, |k'|-|k| \in 2\mathbb{N}} \ln \mu_{k'} - \sum_{k' \sqsupseteq k, |k'|-|k| \notin 2\mathbb{N}} \ln \mu_{k'} \\ &= \sum_{k' \sqsupseteq k} (-1)^{|k'|-|k|} \ln \mu_{k'}. \end{aligned}$$

□

Since for $1 \leq k < 2^n$, $C_k = 1$ is equivalent to the event $\{X_i = 1 : i, k_i = 1\}$, Theorem 4 shows, using the setting of Remark 1, that $s(A_k)$ is equal to the conditional mutual information between sources $\mathfrak{s}_i$, i such that $k_i = 0$, not being reliable, given that the sources $\mathfrak{s}_i$, i such that $k_i = 1$, are not reliable.

5.2. Disjunctive counterparts

The *implicability* function b is another equivalent representation of a mass function m, defined as

$$b(A) = \sum_{B \subseteq A} m(B), \quad \forall A \subseteq \mathcal{Y}.$$

It allows an expression similar to Eq. (7) for the combination by the disjunctive rule $\textcircled{\cup}$ [37,38], which is a combination rule that has the same definition as $\textcircled{\cap}$ except that $\cap$ is replaced by $\cup$ in (6): we have $b_{1 \textcircled{\cup} 2}(A) = b_1(A) \cdot b_2(A)$, for all $A \subseteq \mathcal{Y}$. In matrix form, we have [23]

$$\mathbf{m} = \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \right) \mathbf{b}. \tag{52}$$

Let us consider the MBD p (22) and its associated vector $\boldsymbol{\lambda}$ defined as

$$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{2^n}),$$

where, for $1 \leq k \leq 2^n$,

$$\lambda_k = \mathbb{E}\left[\prod_{i=1}^{n} (1 - X_i)^{1-k_i} \right],$$

with k_i, $i = 1, \ldots, n$, the terms in the binary expansion of k. λ_k comes down to the (marginal) probability that each of the variables in $\{X_i : i, k_i = 0\}$ equals 0.

Using a similar[7] proof to that of (25) provided in [17], it is straightforward to show that the following relation holds

$$\mathbf{p} = \left(\bigotimes_{i=1}^{n} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \right) \boldsymbol{\lambda}. \tag{53}$$

Let m be a mass function on $\mathcal{Y}$ with associated implicability function b, and let p be its MBD-equivalent, with associated vector $\boldsymbol{\lambda}$. By comparing Eqs. (52) and (53), it is clear that $\lambda_k = b(A_k)$ for all $1 \leq k \leq 2^n$. Hence, adopting the setting of Remark 1, $b(A_k)$ is equal to the (marginal) probability that each of the sources in $\{\mathfrak{s}_i : i, k_i = 0\}$ is reliable.

As shown in [10], there exists a decomposition of non normal belief functions based on the disjunctive rule. This decomposition relies on the *disjunctive* weight function $v : 2^{\mathcal{Y}} \backslash \{\emptyset\} \to (0, +\infty)$, which is an equivalent representation of a non normal mass function m, defined as

$$v(A) = \prod_{B \subseteq A} b(B)^{(-1)^{|A|-|B|+1}}, \quad \forall A \neq \emptyset,$$

[7] The main difference with the proof of (25) is that one needs to use

$$\begin{bmatrix} 1 - X_i \\ X_i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 - X_i \\ 1 \end{bmatrix}, \quad i = 1, \ldots, n.$$

with b the implicability function associated to m. We have, for any non normal mass function m,

$$m = \bigcirc\!\!\!\!\!\!\cup_{A\neq\emptyset} A_{v(A)}, \tag{54}$$

with v the disjunctive weight function associated to m and $A_{v(A)} : 2^{\mathcal{Y}} \to \mathbb{R}$ a mapping called a *negative* generalised simple mass function allocating $v(A)$ to $\emptyset$, $1 - v(A)$ to $A \subseteq \mathcal{Y}, A \neq \emptyset$, and 0 to all $B \in 2^{\mathcal{Y}}\backslash\{A, \emptyset\}$ (if $v(A) \leq 1$, then this mapping is a proper mass function called a negative simple mass function). Decomposition (54) is unique and is the disjunctive counterpart of (11). The interpretation of disjunctive weights $v(A)$, $A \neq \emptyset$, is not discussed in [10]. However, using a similar reasoning to the one followed in Section 5.1 for the weight function w, disjunctive weights can be related to measures of information, as detailed below.

Much as it is possible to present decomposition (11) using function s rather than w, the decomposition (54) can be presented using a function $r : 2^{\mathcal{Y}}\backslash\{\emptyset\} \to (-\infty, +\infty)$ such that $r(A) = -\ln v(A)$. Let m be a mass function defined on $\mathcal{Y} = \{y_1, y_2\}$, with MBD-equivalent p. We obtain

$$\begin{aligned}
r(\{y_1\}) &= -\ln \frac{b(\emptyset)}{b(\{y_1\})} \\
&= -\ln \frac{\lambda_1}{\lambda_2} \\
&= I(X_1 = 0|X_2 = 0), \\
r(\{y_2\}) &= -\ln \frac{b(\emptyset)}{b(\{y_2\})} \\
&= -\ln \frac{\lambda_1}{\lambda_3} \\
&= I(X_2 = 0|X_1 = 0), \\
r(\mathcal{Y}) &= -\ln \frac{b(\{y_1\})b(\{y_2\})}{b(\emptyset)b(\mathcal{Y})} \\
&= \ln \frac{\lambda_1}{\lambda_2\lambda_3} \\
&= I(X_1 = 0; X_2 = 0),
\end{aligned}$$

since we have, for $k = 1 \leftrightarrow (k_1 = 0, k_2 = 0)$,

$$\begin{aligned}
I(X_1 = 0; X_2 = 0) &= \ln \frac{P(X_1 = 0, X_2 = 0)}{P(X_1 = 0)P(X_2 = 0)} \\
&= \ln \frac{p_1}{(\sum_{k':k'_1=0} p_{k'})(\sum_{k':k'_2=0} p_{k'})} \\
&= \ln \frac{\lambda_1}{\lambda_3\lambda_2}, \\
I(X_2 = 0|X_1 = 0) &= -\ln \frac{P(X_1 = 0, X_2 = 0)}{P(X_1 = 0)} \\
&= -\ln \frac{\lambda_1}{\lambda_3}, \\
I(X_1 = 0|X_2 = 0) &= -\ln \frac{P(X_1 = 0, X_2 = 0)}{P(X_2 = 0)} \\
&= -\ln \frac{\lambda_1}{\lambda_2}.
\end{aligned}$$

Adopting the setting of Remark 1, we have for instance that $r(\mathcal{Y})$ is thus equal to the mutual information between both sources underlying m being reliable.

More generally, the above interpretation for function r in terms of measures of information can be extended to belief functions defined on domains $\mathcal{Y}$ of any cardinality, thanks to Theorem 5, which is the counterpart to Theorem 4 for function r.

Theorem 5. *Let m be a mass function on $\mathcal{Y} = \{y_1, \ldots, y_n\}$ with MBD-equivalent p and with $\{X_i : i = 1, \ldots, n\}$ the Bernoulli random variables underlying p. We have*

$$r(A_k) = I(Mut(\mathcal{R}_k)|D_k = 1), \quad 1 < k \leq 2^n,$$

with, using $k \leftrightarrow (k_1, \ldots, k_N)$, $D_k := \prod_{i=1}^{n}(1 - X_i)^{1-k_i}$ and $\mathcal{R}_k := \{X_i = 0 : i, k_i = 1\}$.

Proof. The proof is similar to that of Theorem 4. □

Since for $1 < k \leq 2^n$, $D_k = 1$ is equivalent to the event $\{X_i = 0 : i, k_i = 0\}$, Theorem 5 shows, using the setting of Remark 1, that $r(A_k)$ is equal to the conditional mutual information between sources $\mathbb{s}_i$, i such that $k_i = 1$, being reliable, given that the sources $\mathbb{s}_i$, i such that $k_i = 0$, are reliable.

Let us conclude by remarking that much as Eq. (54) is the disjunctive counterpart of Eq. (11), it is possible to obtain a disjunctive counterpart of Theorem 3 as follows. In [33], a general approach for the *disjunctive* combination of bodies of evidence is also mentioned: it consists in a simple modification of Step 2 of the general approach to the conjunctive combination of bodies of evidence recalled in Section 4.1, where each joint mass $jm(A_1, \ldots, A_N)$ is now allocated to the subset $\bigcup_{i=1}^N A_i$. As a result, the disjunctive combination of N mass functions $m_1, \ldots, m_N$ on $\mathcal{Y}$ is the mass function $m_\cup$ defined as

$$m_\cup(A) = \sum_{\bigcup_{i=1}^N A_i = A} jm(A_1, \ldots, A_N). \quad \forall A \subseteq \mathcal{Y}.$$

Let us consider the case where the mass functions $m_1, \ldots, m_N$ are binary and such that $m_i = (A_{i_0}, A_{i_1})^{\pi_i}$. In such case, the dependence structure represented by a joint mass function jm, is completely determined as explained in Section 4.1 by vector $\boldsymbol{\sigma}$ associated with the MBD such that $p_k = jm(\mathbf{A}_k)$ with $\mathbf{A}_k := (A_{1_{k_1}}, \ldots, A_{N_{k_N}})$ using $k \leftrightarrow (k_1, \ldots, k_N)$, $1 \leq k \leq 2^N$. As a result, the disjunctive combination of N binary mass functions can be expressed as a parameterised combination rule $\mathbb{O}_{\boldsymbol{\sigma}} : \mathcal{B}^N \to \mathcal{M}$, with parameter $\boldsymbol{\sigma}$ representing the dependence structure, defined as

$$\mathbb{O}_{\boldsymbol{\sigma}}((A_{1_0}, A_{1_1})^{\pi_1}, \ldots, (A_{N_0}, A_{N_1})^{\pi_N}) := m_\cup,$$

with $m_\cup$ the result of the disjunctive combination of mass functions m_i with dependence represented by the joint mass function jm determined by vector $\boldsymbol{\sigma}$ and having marginals m_i, $i = 1, \ldots, N$.

When $A_{i_0} = \emptyset$, $i = 1, \ldots, N$, we may denote $\mathbb{O}_{\boldsymbol{\sigma}}\big((A_{1_0}, A_{1_1})^{\pi_1}, \ldots, (A_{N_0}, A_{N_1})^{\pi_N}\big)$ for short as $\mathbb{O}_{\boldsymbol{\sigma}}\Big(A_{1_{1\,1-\pi_1}}, \ldots, A_{N_{1\,1-\pi_N}}\Big)$, since a binary mass function $(A_{i_0}, A_{i_1})^{\pi_i}$ such that $A_{i_0} = \emptyset$ is nothing but the negative simple mass function $A_{i_{1\,1-\pi_i}}$.

Theorem 6. *Any mass function m defined on a domain $\mathcal{Y} = \{y_1, \ldots, y_n\}$ with Teugels function t satisfies*

$$m = \mathbb{O}_{\boldsymbol{\sigma}}\big(\{y_1\}_{1-t(\{y_1\})}, \ldots, \{y_n\}_{1-t(\{y_n\})}\big), \tag{55}$$

with $\boldsymbol{\sigma}$ the vector such that $\sigma_1 = 1$ and, for $1 < k \leq 2^n$, $\sigma_k = t(A_k)$ if $|A_k| > 1$, and $\sigma_k = 0$ otherwise.

Proof. The MBD in one-to-one correspondence with the joint mass function jm underlying the disjunctive combination (55), is such that its marginals satisfy $P(X_i = 1) = \pi_i = t(\{y_i\})$, $i = 1, \ldots, n$, and its vector of central moments is $\boldsymbol{\sigma}$. Hence, its Teugels' representation τ is such that $\tau_k = t(A_k)$, $1 < k \leq 2^n$. In other words, the MBD associated with jm is the MBD-equivalent of m, and thus $p_k = m(A_k)$, $1 \leq k \leq 2^n$, from which we obtain $jm(\mathbf{A}_k) = m(A_k)$ since $p_k = jm(\mathbf{A}_k)$, with $\mathbf{A}_k := (A_{1_{k_1}}, \ldots, A_{n_{k_n}})$.

Moreover, for $1 \leq k \leq 2^n$, we have

$$\begin{aligned}
\bigcup_{i=1}^n A_{i_{k_i}} &= \Big(\bigcup_{i=1, k_i=0}^n A_{i_{k_i}}\Big) \bigcup \Big(\bigcup_{i=1, k_i=1}^n A_{i_{k_i}}\Big) \\
&= \bigcup_{i=1, k_i=1}^n A_{i_{k_i}} \\
&= \bigcup_{i=1, k_i=1}^n \{y_i\} \\
&= A_k.
\end{aligned}$$

Hence, $jm(\mathbf{A}_k) = m(A_k)$ is allocated to A_k. □

Theorem 6 is the counterpart of Theorem 3. It shows that any mass function results from the disjunctive combination of $|\mathcal{Y}|$ negative simple mass functions having some dependence structure.

6. Conclusions

The problem of decomposing uniquely any belief function into elementary items received a solution from Smets [9], extending previous ideas from Shafer [1]. As argued in this paper, Smets' solution has a major weakness, which is that it involves elementary items whose proposed semantics lacks formal justifications.

In Dempster's seminal work [25], a belief function is induced from a space equipped with a probability measure and a multi-valued mapping from this space to another one. In this paper we have considered a particular case of this framework where the probability measure is multivariate Bernoulli and where the multi-valued mapping is the conjunction of multi-valued mappings associated to the Bernoulli random variables underlying the multivariate Bernoulli distribution. Using Pichon et al.'s general approach to information fusion [26], we have provided a setting in which this particular case

of Dempster's framework receives a concrete meaning: it may be associated to a simple situation where partially reliable sources provide crisp pieces of information. Furthermore, using this latter setting and a representation of the multivariate Bernoulli distribution due to Teugels [17], we have been able to propose a new canonical decomposition of belief functions and an associated new equivalent representation of a belief function which we called the Teugels function. This decomposition resembles Smets' decomposition in that a belief function is also seen as the result of partially reliable pieces of evidence. However, all the elementary items that it involves have well-defined semantics. In a nutshell, a belief function stems according to this decomposition, from as many crisp pieces of information as there are elements in its domain of definition $\mathcal{Y}$, and from simple probabilistic knowledge concerning both the marginal reliability of each of these pieces of information and the dependencies between their reliability. In addition, we showed that computing the Teugels function of a finite random set provides a meaningful canonical decomposition of the random set.

We were also led to consider the weight function associated with Smets' decomposition in light of our framework. We showed that the weight function corresponds to measures of information of the case where the sources involved in our framework are not reliable. This constitutes a completely different perspective on this function than the one proposed by Smets. This new semantics for the weight function is unfortunately less appealing than Smets', but it is well-defined, contrary to Smets'. In addition, we provided a similar semantics for the disjunctive weight function, whose interpretation had never been discussed.

The t-canonical decomposition and its building blocks open a potentially fruitful path. It seems indeed interesting to revisit in light of our framework, the main problems concerning belief functions and their associated solutions, such as combination [24] and specifically combination under ill-known dependency [10,33], distance evaluation [39], building methods [40], measurement of uncertainty [41] and of consistency [42].

On the one hand, revisiting an existing solution or notion, as we have done in this paper for instance for the unnormalised Dempster's rule and the equivalent representations of a belief function that are the commonality, implicability, weight and disjunctive weight functions, may provide a new and useful perspective on the solution or notion, as was the case in particular for the weight functions that we were able to relate to information measures.

On the other hand, tackling these main problems using our framework might lead to new solutions for them. For instance, the problem of combining belief functions under ill-known dependency could be looked at in light of the elementary pieces of evidence underlying a belief function according to the t-decomposition. New distances between belief functions, based on Teugels function t, could be investigated. Uncertainty measures and building methods (such as the one used in [32, Section 2.2.1]) for the MBD could yield new uncertainty measures and building methods for belief functions. The conjunctive combination with dependence $\boldsymbol{\sigma}$ of simple mass functions could present some interest in preference modelling as a means to resolve inconsistency between elementary information items, represented by simple mass functions and assumed so far to be independent [43], since the dependence assumption made between pieces of information is known to have potentially a significant impact on consistency [42].

Acknowledgment

The author thanks the anonymous referees for their valuable comments that helped to improve this paper.

References

[1] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, N.J., 1976.
[2] P. Smets, R. Kennes, The transferable belief model, Artif. Intell. 66 (1994) 191–243.
[3] T. Denœux, S. Sriboonchitta, O. Kanjanatarakul, Evidential clustering of large dissimilarity data, Knowl. Based Syst. 106 (2016) 179–195.
[4] T. Denœux, Z. Younes, F. Abdallah, Representing uncertainty on set-valued variables using belief functions, Artif. Intell. 174 (7) (2010) 479–499.
[5] D. Mercier, E. Lefèvre, D. Jolly, Object association with belief functions, an application with vehicles, Inf. Sci. 181 (24) (2011) 5485–5500.
[6] T. Denœux, N.E. Zoghby, V. Cherfaoui, A. Jouglet, Optimal object association in the Dempster–Shafer framework, IEEE Trans. Cybern. 44 (12) (2014) 2521–2531.
[7] L. Jiao, Q. Pan, T. Denœux, Y. Liang, X. Feng, Belief rule–based classification system: extension of FRBCS in belief functions framework, Inf. Sci. 309 (2015) 26–49.
[8] T. Denœux, 40 years of Dempster–Shafer theory, Int. J. Approx. Reason. 79 (2016) 1–6.
[9] P. Smets, The canonical decomposition of a weighted belief, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95), San Mateo, California, USA, 1995, Morgan Kaufmann, 1995, pp. 1896–1901.
[10] T. Denœux, Conjunctive and disjunctive combination of belief functions induced by non-distinct bodies of evidence, Artif. Intell. 172 (2008) 234–264.
[11] A. Kallel, S.L. Hégarat-Mascle, Combination of partially non-distinct beliefs: the cautious-adaptive rule, Int. J. Approx. Reason. 50 (7) (2009) 1000–1021.
[12] F. Pichon, T. Denœux, The unnormalized Dempster's rule of combination: a new justification from the least commitment principle and some extensions, J. Autom. Reason. 45 (1) (2010) 61–87.
[13] D. Mercier, E. Lefèvre, F. Delmotte, Belief functions contextual discounting and canonical decompositions, Int. J. Approx. Reason. 53 (2) (2012) 146–158.
[14] D. Mercier, F. Pichon, E. Lefèvre, Corrigendum to "Belief functions contextual discounting and canonical decompositions" [International Journal of Approximate Reasoning 53 (2012) 146–158], Int. J. Approx. Reason. 70 (2016) 137–139.
[15] F. Pichon, D. Mercier, E. Lefèvre, F. Delmotte, Proposition and learning of some belief function contextual correction mechanisms, Int. J. Approx. Reason. 72 (2016) 4–42.
[16] J. Schubert, Clustering decomposed belief functions using generalized weights of conflict, Int. J. Approx. Reason. 48 (2) (2008) 466–480.
[17] J.L. Teugels, Some representations of the multivariate Bernoulli and binomial distributions, J. Multivar. Anal. 32 (2) (1990) 256–268.
[18] D.G. Kendall, Foundations of a theory of random sets, in: E.F. Harding, D.G. Kendall (Eds.), Stochastic Geometry, Wiley, New York, 1974, pp. 322–376.
[19] G. Matheron, Random Sets and Integral Geometry, Wiley, New York, 1975.
[20] R. Fano, Transmission of Information: A Statistical Theory of Communications, MIT Press, Cambridge, MA, 1961.
[21] I. Couso, D. Dubois, Statistical reasoning with set-valued information: ontic vs. epistemic views, Int. J. Approx. Reason. 55 (7) (2014) 1502–1518.
[22] D. Dubois, H. Prade, P. Smets, A definition of subjective possibility, Int. J. Approx. Reason. 48 (2) (2008) 352–364.

[23] P. Smets, The application of the matrix calculus to belief functions, Int. J. Approx. Reason. 31 (1–2) (2002) 1–30.
[24] P. Smets, Analyzing the combination of conflicting belief functions, Inf. Fus. 8 (4) (2007) 387–412.
[25] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.
[26] F. Pichon, D. Dubois, T. Denœux, Relevance and truthfulness in information correction and fusion, Int. J. Approx. Reason. 53 (2) (2012) 159–175.
[27] H.T. Nguyen, On random sets and belief functions, J. Math. Anal. Appl. 65 (1978) 531–542.
[28] O. Kanjanatarakul, T. Denœux, S. Sriboonchitta, Prediction of future observations using belief functions: a likelihood-based approach, Int. J. Approx. Reason. 72 (2016) 71–94.
[29] F. Pichon, T. Denœux, On latent belief structures, in: K. Mellouli (Ed.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Computer Science, 4724, Springer, 2007, pp. 368–380.
[30] I. Song, S. Lee, Explicit formulae for product moments of multivariate Gaussian random variables, Stat. Probab. Lett. 100 (2015) 27–34.
[31] C. Rose, M.D. Smith, Mathematical Statistics with Mathematica, Springer-Verlag, New York, 2002.
[32] J.L. Teugels, J. Van Horebeek, Algebraic descriptions of nominal multivariate discrete data, J. Multivar. Anal. 67 (2) (1998) 203–226.
[33] S. Destercke, D. Dubois, Idempotent conjunctive combination of belief functions: extending the minimum rule of possibility theory, Inf. Sci. 181 (18) (2011) 3925–3945.
[34] S. Ferson, R. Nelsen, J. Hajagos, D. Berleant, J. Zhang, W.T. Tucker, L. Ginzburg, W.L. Oberkampf, Dependence in Probabilistic Modeling, Dempster-Shafer Theory, and Probability Bounds Analysis, Technical Report SAND2004-3072, Sandia National Laboratories, Albuquerque, New Mexico, 2004.
[35] P. Smets, The transferable belief model and random sets, Int. J. Intell. Syst. 7 (1992) 37–46.
[36] M. Ha-Duong, Hierarchical fusion of expert opinions in the transferable belief model, application to climate sensitivity, Int. J. Approx. Reason. 49 (3) (2008) 555–574.
[37] D. Dubois, H. Prade, A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets, Int. J. Gen. Syst. 12 (3) (1986) 193–226.
[38] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, Int. J. Approx. Reason. 9 (1) (1993) 1–35.
[39] A.-L. Jousselme, P. Maupin, Distances in evidence theory: comprehensive survey and generalizations, Int. J. Approx. Reason. 53 (2) (2012) 118–145.
[40] N.B. Abdallah, N. Mouhous-Voyneau, T. Denœux, Combining statistical and expert evidence using belief functions: application to centennial sea level estimation taking into account climate change, Int. J. Approx. Reason. 55 (1, Part 3) (2014) 341–354.
[41] G.J. Klir, Uncertainty and Information: Foundations of Generalized Information Theory, Wiley-IEEE Press, 2005.
[42] S. Destercke, T. Burger, Toward an axiomatic definition of conflict between belief functions, IEEE Trans. Syst. Man Cybern.Part B 43 (2) (2013) 585–596.
[43] S. Destercke, A generic framework to include belief functions in preference handling for multi-criteria decision, in: A. Antonucci, L. Cholvy, O. Papini (Eds.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 14th European Conference, ECSQARU 2017, Lecture Notes in Artificial Intelligence, 10369, Springer, 2017, pp. 179–189.

Annexe D

Face pixel detection using evidential calibration and fusion

Article paru dans *International Journal of Approximate Reasoning*, 91 : 202-215, 2017.

International Journal of Approximate Reasoning 91 (2017) 202–215

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Face pixel detection using evidential calibration and fusion ☆

Pauline Minary [a,b,*], Frédéric Pichon [a], David Mercier [a], Eric Lefevre [a], Benjamin Droit [b]

[a] Univ. Artois, EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), Béthune, F-62400, France
[b] SNCF Réseau, Département des Télécommunications, La Plaine Saint Denis, France

ARTICLE INFO

Article history:
Received 28 February 2017
Received in revised form 31 August 2017
Accepted 8 September 2017
Available online 19 September 2017

Keywords:
Belief functions
Information fusion
Evidential calibration
Face blurring

ABSTRACT

Due to legal reasons, faces on a given image may have to be blurred. This may be achieved by combining several information sources, which may provide information at different levels of granularity; for instance face detectors return bounding boxes corresponding to assumed positions of faces, whereas skin detectors may return pixel level information. A general, well-founded and efficient approach to combining box-based information sources was recently proposed in the context of pedestrian detection. This approach relies on evidence theory to calibrate and combine sources. In this paper, we apply this approach to combine face (rather than pedestrian) detectors, in order to obtain a state-of-the-art face blurring system based on multiple detectors. Then, we propose another approach to tackle the blurring problem, which consists essentially in applying at the pixel-level the central idea – combining evidentially calibrated information sources – of the preceding box-based approach. This shift of focus induces several conceptual advantages. In addition, the proposed approach shows better performances on a classical face dataset, as well as on a more challenging one.

1. Introduction

Due to legal reasons, faces on a given image may have to be blurred. Yet, it can rapidly become a tedious task if it is done manually, especially if there is a large amount of images to process. A solution may consist in using a face detection system, which aims to automatically find the positions of the faces in a given image.

Since the early 2000s, there has been significant research on face detection and many algorithms have been proposed, in particular based on machine learning techniques, such as the well-known Viola and Jones approach [34] or the neural network-based approach proposed by Rowley et al. [27]. Recently, more elaborate algorithms based on deep convolutional neural networks [10,37,40] made a major breakthrough in the field. Yet, another path of research consists in merging information given by multiple sources, whether situated at the pixel level or directly on the faces [1,11,23,32]. Indeed, since sources, such as face detectors, generally provide complementary information, using several of them is a means to improve overall performance.

☆ This paper is part of the Virtual special issue on Belief Functions: Theory and Applications, edited by Jirina Vejnarova and Vaclav Kratochvil.

* Corresponding author.
E-mail addresses: pauline.minary@reseau.sncf.fr (P. Minary), frederic.pichon@univ-artois.fr (F. Pichon), david.mercier@univ-artois.fr (D. Mercier), eric.lefevre@univ-artois.fr (E. Lefevre), benjamin.droit@reseau.sncf.fr (B. Droit).

http://dx.doi.org/10.1016/j.ijar.2017.09.002
0888-613X/

There are many different ways to perform the fusion of some given information. Among them, in the context of pedestrian detection, Xu et al. [35] recently proposed a well-founded and general approach. In this approach, for a given image each used detector provides a set of bounding boxes corresponding to the assumed positions of the pedestrians, as well as a confidence score for each of these boxes. The main idea is then to use a step called score calibration [26], in order to be able to combine these calibrated scores afterwards, and to obtain better detection performance. Of particular interest is that the calibration and combination steps of this approach rely on a framework for reasoning under uncertainty called evidence theory [28,29]. This theory is a generalization of probability theory, which enables to account for uncertainties due to randomness and incompleteness. As a matter of fact, Xu et al. [36] subsequently proposed more elaborate calibration procedures than those used in [35], which exploit more fully the expressive power of this theory and as a result, model more precisely the uncertainties inherent to the calibration process. Hence, by replacing the calibration procedure in [35] by one of the most efficient ones studied in [36], and by applying to faces the general detection approach introduced in [35], one obtains what may be considered presently as a state-of-the-art face detection system based on multiple detectors. Nonetheless, despite its appeals, we note that such a system suffers from two main limitations inherited from Xu et al.'s approach [35]. First, it is designed to handle only detectors providing bounding boxes, *i.e.*, it can not integrate directly sources providing information at the pixel level. Second, this approach relies on a parameter (so-called overlap threshold) necessary in the handling of boxes.

Using a face detection system is a natural means to solve the face blurring problem. However, we may remark that this problem is not exactly equivalent to face detection: face blurring amounts merely to deciding whether a given pixel belongs to a face, whereas face detection amounts to determining whether a given set of pixels corresponds to the same face. This remark opens the path for a different approach to reasoning about blurring, which may then be situated at the pixel-level. Within this scope, we propose in this paper a face blurring system, which consists essentially in applying at the pixel-level the central idea and contributions of Xu et al. [35,36], *i.e.*, combining evidentially calibrated information sources. As it will be seen, this pixel-level perspective presents several conceptual advantages over operating at the box-level. In particular, sources providing pixel-level information can be directly integrated and the parameter necessary in the handling of boxes can be avoided. Nonetheless, let us note that while our approach presents some interests over box-based methods for the problem of face blurring, these latter methods provide more information (specifically, they isolate faces) and are thus relevant for other problems, such as face recognition.

This paper is organized as follows. Section 2 recalls necessary background on evidence theory as well as on calibration. Section 3 exposes what may be considered as a state-of-the-art face detection system based on multiple detectors, that is, a system performing face detection using Xu et al.'s evidential box-based detection approach [35], improved using evidential calibration [36]. In Section 4, our proposed pixel-based face blurring system is detailed and its fundamental differences with respect to blurring using Xu et al.'s box-based approach are discussed. The performances of the box-based and pixel-based approaches, given the same input information, are then compared in Section 5 on two datasets (one from the literature and one composed of railway platforms images coming from the French railway company SNCF). The ability of the proposed approach to integrate directly pixel-level information is illustrated in Section 6 on these same two datasets. Finally, conclusions and perspectives are given in Section 7.

2. Background

In this section, necessary concepts of evidence theory, such as combination and decision-making schemes, are recalled. Classical calibration methods based on probability theory are then described, followed by their extensions to the evidential framework.

2.1. Evidence theory

The theory of evidence is a framework for reasoning under uncertainty. Let Ω be a finite set called the frame of discernment, which contains all the possible answers to a given question of interest Q. In this theory, uncertainty with respect to the answer to Q is represented using a *Mass Function* (MF) defined as a mapping $m^\Omega : 2^\Omega \to [0,1]$ that satisfies $m^\Omega(\emptyset) = 0$ and

$$\sum_{A \subseteq \Omega} m^\Omega(A) = 1. \tag{1}$$

The quantity $m^\Omega(A)$ corresponds to the share of belief that supports the claim that the answer is contained in $A \subseteq \Omega$ and nothing more specific. A mass function can be equivalently represented by the *plausibility function*, defined by

$$Pl^\Omega(A) = \sum_{B \cap A \neq \emptyset} m^\Omega(B), \quad \forall A \subseteq \Omega. \tag{2}$$

It represents the amount of evidence which does not contradict the hypothesis $\omega \in A$. The plausibility function restricted to singletons is called the *contour function*, denoted pl^Ω and defined by

$$pl^\Omega(\omega) = Pl^\Omega(\{\omega\}), \quad \forall \omega \in \Omega. \tag{3}$$

Given two independent MFs m_1^Ω and m_2^Ω about the answer to Q, it is possible to combine them using *Dempster's rule of combination*. The result of this combination is a MF $m_{1\oplus 2}^\Omega$ defined by

$$m_{1\oplus 2}^\Omega(\emptyset) = 0, \tag{4}$$

$$m_{1\oplus 2}^\Omega(A) = \frac{1}{1-\kappa} \sum_{B\cap C=A} m_1^\Omega(B) m_2^\Omega(C), \tag{5}$$

where

$$\kappa = \sum_{B\cap C=\emptyset} m_1^\Omega(B) m_2^\Omega(C), \tag{6}$$

represents the *degree of conflict* between m_1^Ω and m_2^Ω. If $\kappa = 1$, there is a total conflict between the two pieces of evidence and they cannot be combined.

Different decision strategies exist to make a decision about the true answer to Q, given a MF m^Ω on this answer [6]. In particular, the answer having the smallest so-called *upper expected cost* may be selected, where the upper expected cost $R^*(\omega)$ of some answer $\omega \in \Omega$ is defined as

$$R^*(\omega) = \sum_{A\subseteq\Omega} m^\Omega(A) \max_{\omega'\in A} c(\omega, \omega'), \tag{7}$$

with $c(\omega, \omega')$ is the cost of deciding ω when the true answer is ω'.

2.2. Probabilistic calibration of sources

Consider an object whose true label Y is such that $Y \in \mathbb{Y} = \{0, 1\}$. Furthermore, suppose that after observing this object, a source returns a piece of information of the form $X \in \mathbb{X}$ for some domain $\mathbb{X}$. To learn how to interpret what this piece of information tells us about Y, a step called calibration may be used, which consists in estimating the probability distribution $p^{\mathbb{Y}}(\cdot|X)$. This step relies on a training set $\mathcal{L}$, which contains n other objects for which the variable Y is known, and for which we observed what the source returned on $\mathbb{X}$, *i.e.*, $\mathcal{L} = \{(X_1, Y_1), ..., (X_n, Y_n)\}$ where $X_i \in \mathbb{X}$ represents the information given by the source for the ith object whose true label is $Y_i \in \mathbb{Y}$. Domain $\mathbb{X}$ can be either discrete or continuous, thus different calibration procedures are necessary depending on the output type of the source. For instance, a face detector may return a score associated to a box, which is a continuous piece of information, while a skin detector may return a binary decision for a pixel, which is a discrete piece of information.

2.2.1. Probabilistic calibration of discrete information

Let $\mathbb{X} = \{1, 2, ..., M\}$. The probability that the label $Y = 1$ given information $X = j \in \mathbb{X}$ may be estimated by

$$P^{\mathbb{Y}}(1|j) = \frac{|\{(X_i, Y_i) \in \mathcal{L} | X_i = j, Y_i = 1\}|}{|\{(X_i, Y_i) \in \mathcal{L} | X_i = j\}|}. \tag{8}$$

2.2.2. Probabilistic calibration of continuous information

Let $\mathbb{X} = \mathbb{R}$. The probability that the label $Y = 1$ given information $X = S \in \mathbb{R}$, *i.e.*, $P^{\mathbb{Y}}(1|S)$, may be estimated using three main approaches defined to calibrate sources returning real-valued confidence scores, which will be the case in this paper. These approaches are binning [38], isotonic regression [39] and logistic regression [26].

The binning approach consists in dividing the score spaces into different bins, for example $]-3; -2]$, $]-2; -1]$, etc. For each bin j, the number k_j of pairs $(X_i, Y_i) \in \mathcal{L}$ such that $Y_i = 1$ and X_i in bin j, and the number n_j of pairs (X_i, Y_i) such that X_i in bin j, can be obtained. Then, for a score $X = S$ such that S belongs to bin j, we have

$$P^{\mathbb{Y}}(1|S) = \frac{k_j}{n_j}. \tag{9}$$

Yet, the accuracy of binning highly depends on the number and size of the bins.

Isotonic regression can be seen as a sort of binning, where the size and the boundaries of the bins are dynamically created. It relies on the pool adjacent violators algorithm [2], which consists in fitting a non-decreasing function to the training data by minimizing the mean-squared error [39].

Logistic regression is a more elaborate and accurate method based on a fitting of a sigmoid function h defined by

$$P^{\mathbb{Y}}(1|S) \approx h_S(\theta) = \frac{1}{1 + e^{(\theta_0 + \theta_1 S)}}, \tag{10}$$

where the parameter $\theta = (\theta_0, \theta_1) \in \Theta = \mathbb{R}^2$ is chosen as the one maximizing the following likelihood function:

$$L(\theta) = \prod_{i=1}^{n} p_i^{Y_i}(1-p_i)^{1-Y_i}, \tag{11}$$

with

$$p_i = \frac{1}{1+e^{(\theta_0+\theta_1 X_i)}}. \tag{12}$$

Note that some score values may be less present than others in the training set, thus some estimated probabilities may be less accurate than others. To address this issue, Xu et al. proposed to refine the above three calibrations using the theory of evidence [36], in order to manage these inaccuracies. The following section deals with the evidential versions of calibration procedures.

2.3. Evidential calibration of sources

As for the probabilistic analysis, evidential calibration procedures can be defined differently depending on the type of outputs returned by the considered source. Any of these evidential calibration procedures yields a MF $m^{\mathbb{Y}}(\cdot|X)$ (rather than a probability distribution) accounting explicitly for uncertainties in the calibration process.

2.3.1. Evidential calibration of discrete information

Evidence theory provides different models to extend a probabilistic approach into an evidential one, as detailed in [36]. Hence, in the discrete case where the received information from the source is such that $X = j$, there are different ways to obtain the MF $m^{\mathbb{Y}}(\cdot|j)$. In particular, one may use the model of Dempster [5], which leads to the following MF:

$$\begin{aligned} m^{\mathbb{Y}}(\{0\}|j) &= \frac{|\{(X_i,Y_i)\in\mathcal{L}|X_i=j, Y_i=0\}|}{|\{(X_i,Y_i)\in\mathcal{L}|X_i=j\}|+1}, \\ m^{\mathbb{Y}}(\{1\}|j) &= \frac{|\{(X_i,Y_i)\in\mathcal{L}|X_i=j, Y_i=1\}|}{|\{(X_i,Y_i)\in\mathcal{L}|X_i=j\}|+1}, \end{aligned} \tag{13}$$

and

$$m^{\mathbb{Y}}(\{0,1\}|j) = \frac{1}{|\{(X_i,Y_i)\in\mathcal{L}|X_i=j\}|+1}.$$

Hereafter, we will refer to this type of calibration as evidential Dempster calibration.

2.3.2. Evidential calibration of continuous information

Xu et al. [36] proposed several evidential extensions of probabilistic calibration methods of scores. This paper focuses on the extension of the logistic regression based on the so-called likelihood model [7,18,17], as Xu et al. showed that this is the one presenting overall the best performances of all methods [36].

In this extension of the logistic regression, one first represents knowledge about parameter $\tau = h_s(\theta) \in T = [0,1]$ after observing $X = S$, in the form of a consonant belief function $Bel^T(.|S)$ with contour function $pl^T(.|S)$ defined by

$$pl^T(\tau|S) = \sup_{\theta_1\in\mathbb{R}} pl^{\Theta}(\ln(\tau^{-1}-1)-\theta_1 S, \theta_1), \forall \tau \in (0,1), \tag{14}$$

with pl^{Θ} the function defined on Θ by

$$pl^{\Theta}(\theta) = \frac{L(\theta)}{L(\hat{\theta})}, \qquad \forall \theta \in \Theta, \tag{15}$$

with L the likelihood function given in Eq. (11), and $\hat{\theta} = (\hat{\theta_0}, \hat{\theta_1})$ the maximum likelihood estimate of θ. Then, viewing the label of an object after observing its score S as the realization of a random variable Y with a Bernoulli distribution $\mathcal{B}(\tau)$, one uses the solution proposed by Kanjanatarakul et al. [18,17] to make statements about Y. In a nutshell, it consists in using the fact that $Bel^T(.|S)$ is equivalent to a random set [24], and in using the sampling model of Dempster [5] to deduce a belief function on $\mathbb{Y}$. As shown by Xu et al. [36], this belief function obtained given information S has associated mass function $m^{\mathbb{Y}}(.|S)$ defined by

$$\begin{aligned} m^{\mathbb{Y}}(\{0\}|S) &= 1-\hat{\tau}-\int_{\hat{\tau}}^{1} pl^T(u|S)du, \\ m^{\mathbb{Y}}(\{1\}|S) &= \hat{\tau}-\int_{0}^{\hat{\tau}} pl^T(u|S)du, \end{aligned} \tag{16}$$

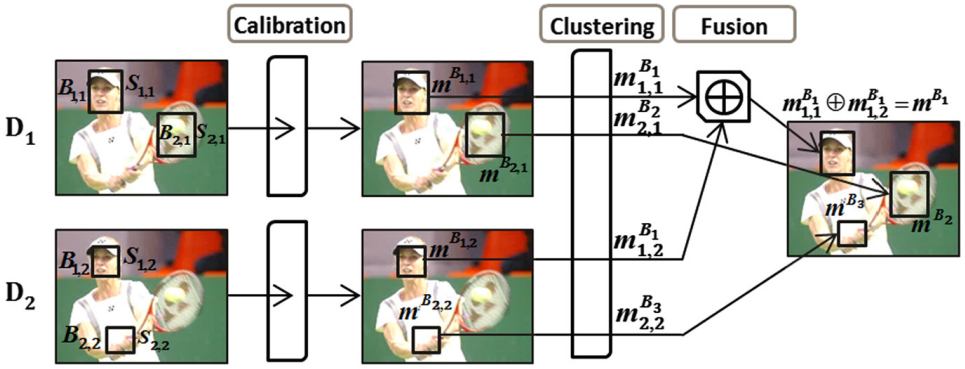


Fig. 1. Illustration of the box-based approach.

and

$$m^{\mathbb{Y}}(\{0,1\}|S) = \int_0^1 pl^T(u|S)du,$$

where $\hat{\tau}$ maximizes the contour function pl^T.

3. An evidential box-based face detection approach

Face blurring may be achieved using simply the boxes returned by a face detection system. In this section, we present such a system, which may be considered as a state-of-the-art system with respect to face detection based on multiple detectors returning box information. In a nutshell, this system is merely Xu et al. [35] evidential box-based detection approach, whose calibration step has been replaced by the evidential likelihood-based logistic regression calibration procedure proposed in [36] and recalled in the previous section. This section first provides an overview of this approach and then details some of its steps.

3.1. Overview of the approach

Let us consider a given image and assume that J face detectors are run on this image. Formally, each detector D_j, $j = 1, ..., J$, provides N_j pairs $(B_{i,j}, S_{i,j})$, where $B_{i,j}$ denotes the ith box, $i = 1, ..., N_j$, returned by the jth detector and $S_{i,j}$ is the confidence score associated to this box.

Through a calibration procedure using a training set that will be described in Section 3.2, score $S_{i,j}$ is transformed into a MF $m^{\mathcal{B}_{i,j}}$ defined over the frame $\mathcal{B}_{i,j} = \{0, 1\}$, where 1 (resp. 0) means that there is a face (resp. no face) in box $B_{i,j}$.

Then, using a clustering procedure detailed in Section 3.3, all the boxes $B_{i,j}$ returned by the J detectors for the considered image, are grouped into K clusters C_k, $k = 1, ..., K$, each of these clusters being represented by a single box B_k.

In addition, for each box $B_{i,j} \in C_k$, its associated MF $m^{\mathcal{B}_{i,j}}$ is assumed to represent a piece of evidence regarding the presence of a face in B_k, that is, $m^{\mathcal{B}_{i,j}}$ is converted into a MF $m_{i,j}^{\mathcal{B}_k}$ on $\mathcal{B}_k = \{0, 1\}$ defined by $m_{i,j}^{\mathcal{B}_k}(A) = m^{\mathcal{B}_{i,j}}(A)$, for all $A \subseteq \{0, 1\}$. These pieces of evidence are then combined using Dempster's rule:

$$m^{\mathcal{B}_k} = \bigoplus_{i,j} m_{i,j}^{\mathcal{B}_k}. \tag{17}$$

The combination results in a MF $m^{\mathcal{B}_k}$ representing the overall system uncertainty with respect to the presence of a face in B_k. We note that the use of Dempster's rule is appropriate when the sources may be considered to be independent and reliable. More complex combination schemes are also considered in [35]. However, only Dempster's rule, which presents good performance in [35], is considered here.

The three main steps of the approach, namely calibration, clustering and fusion, are illustrated in Fig. 1. For the sake of simplicity only two detectors, each returning two boxes, are considered in this example. $B_{i,j}$ corresponds to the ith box, $i = 1, 2$, returned by the jth detector $j = 1, 2$, and which has $S_{i,j}$ as associated score. In this scenario, the boxes $B_{1,1}$ and $B_{1,2}$ are grouped into the same cluster C_1, represented by the box B_1. Their associated scores, transformed into mass functions, are combined and result in the final mass function $m_{1,1}^{\mathcal{B}_1} \oplus m_{1,2}^{\mathcal{B}_1}$, which is denoted by $m^{\mathcal{B}_1}$. The other boxes $B_{2,1}$ and $B_{2,2}$ form their own clusters, respectively represented by B_2 and B_3. Finally, for each resulting box with its associated MF, a decision has to be made whether the box has to be blurred or not; it may be done using the decision strategy given in Section 2.1 and in particular using Eq. (7) for some cost function c.

3.2. Box-based score calibration for a detector

In order to transform the score $S_{i,j}$ associated to a box $B_{i,j}$ into a MF $m^{\mathcal{B}_{i,j}}$, detector D_j needs to be calibrated. In particular, the evidential likelihood-based logistic regression calibration procedure recalled in Section 2.3 may be used instead of the cruder procedures used in [35]. This procedure requires a training set, which we denote by $\mathcal{L}_{cal,j}$. We detail below how $\mathcal{L}_{cal,j}$ is built.

Assume that L images are available. Besides, the positions of the faces really present in each of these images are known in the form of bounding boxes. Formally, this means that for a given image ℓ, a set of M^ℓ boxes $G_r^\ell, r = 1, ..., M^\ell$, is available, with G_r^ℓ the rth bounding (ground truth) box on image ℓ.

Furthermore, detector D_j to be calibrated is run on each of these images, yielding N_j^ℓ pairs $(B_{t,j}^\ell, S_{t,j}^\ell)$ for each image ℓ, where $B_{t,j}^\ell$ denotes the tth box, $t = 1, ..., N_j^\ell$, returned on image ℓ by detector D_j and $S_{t,j}^\ell$ is the confidence score associated to this box.

From these data, training set $\mathcal{L}_{cal,j}$ is defined as the set of pairs $(S_{t,j}^\ell, YB_{t,j}^\ell)$, $\ell = 1, ..., L$, and $t = 1, ..., N_j^\ell$, with $YB_{t,j}^\ell \in \{0, 1\}$ the label obtained by evaluating whether box $B_{t,j}^\ell$ "matches" some face in image ℓ, *i.e.*,

$$YB_{t,j}^\ell = \begin{cases} 1 & \text{if } \exists G_r^\ell, r = 1, ..., M^\ell, \text{ such that } ov(G_r^\ell, B_{t,j}^\ell) \geq \lambda, \\ 0 & \text{otherwise,} \end{cases}$$

where λ is some threshold in $(0, 1)$ and $ov(G_r^\ell, B_{t,j}^\ell)$ is a measure of the overlap between boxes G_r^ℓ and $B_{t,j}^\ell$. It is defined by [9]

$$ov(B_1, B_2) = \frac{area(B_1 \cap B_2)}{area(B_1 \cup B_2)}, \tag{18}$$

for any two boxes B_1 and B_2. Informally, $\mathcal{L}_{cal,j}$ stores the scores associated to all the boxes returned by detector D_j on images where the positions of faces are known, and records for each score whether its associated box is a true or false positive. It is then clear that the MF $m^{\mathcal{B}_{i,j}}$ associated to a new score $S_{i,j}$ and obtained from calibration relying on $\mathcal{L}_{cal,j}$, represents uncertainty toward box $B_{i,j}$ containing a face.

3.3. Clustering of boxes

As several detectors are used, some boxes may be located in the same area of an image, which means that different boxes assume that there is a face in this particular area. The step of clustering allows one to group those boxes and to retain only one per cluster. A greedy approach is used in [35], based on the work of Dollar et al. [8]: the procedure starts by selecting the box $B_{i,j}$ with the highest mass of belief on the face hypothesis, *i.e.*, the box $B_{i,j}$ such that $m^{\mathcal{B}_{i,j}}(\{1\}) > m^{\mathcal{B}_{u,v}}(\{1\}), \forall (u, v) \neq (i, j)$, and this box is considered as the representative of the first cluster. Then, each box $B_{u,v}$, $\forall (u, v) \neq (i, j)$, such that the overlap $ov(B_{i,j}, B_{u,v})$ is above the threshold λ, is grouped into the same cluster as $B_{i,j}$, and is then no longer considered for further associations. Among the remaining boxes, the box $B_{i,j}$ with the highest $m^{\mathcal{B}_{i,j}}(\{1\})$ is selected as representative of the next cluster, and the procedure is repeated until all the boxes are clustered.

4. Proposed evidential pixel-based approach

The approach exposed in the previous section is general and well-founded. It is designed for detectors returning boxes, but it does not allow to directly integrate pixel-based information. Besides, as explained in the introduction, for the purpose of blurring it seems interesting to work at the pixel level rather than box level. Thus, the idea of the approach proposed in this section is to use elements from the previous system, in particular the evidential calibration and fusion, and to apply them at the pixel level. This section first exposes an overview of the proposed approach. Then, in order to be able to compare subsequently the proposed pixel-based approach to the previous system, we detail how the same input information as in the previous section, *i.e.*, boxes and scores returned by detectors, can be used within our pixel-based approach. Finally, fundamental differences between the two approaches are discussed.

4.1. Overview of the approach

To each pixel $p_{x,y}$ in an image, we associate a frame of discernment $\mathcal{P}_{x,y} = \{0, 1\}$, where x and y are the coordinates of the pixel in the image and 1 (resp. 0) means that there is a face (resp. no face) in pixel $p_{x,y}$. For the pixel $p_{x,y}$, N mass functions are obtained on $\mathcal{P}_{x,y}$ from N independent sources. They are then combined using Dempster's rule of combination, resulting in the MF denoted $m^{\mathcal{P}_{x,y}}$, *i.e.*,

$$m^{\mathcal{P}_{x,y}} = \bigoplus_{k=1}^{N} m_k^{\mathcal{P}_{x,y}}, \tag{19}$$

with $m_k^{\mathcal{P}_{x,y}}$ the MF representing the uncertainty with respect to the presence of a face in the pixel $p_{x,y}$ for the kth source. Each MF $m_k^{\mathcal{P}_{x,y}}, k = 1, \ldots, N$, is obtained using the calibration method corresponding to the type of the outputs of the kth source. Specifically, if the source gives a score information, the MF is obtained through the evidential likelihood-based logistic regression calibration, using a training set $\mathcal{L}$ composed of pairs (X_i, Y_i), with X_i the score associated to the ith object which is now a pixel, and Y_i its true label. Otherwise, if the source gives discrete information, the evidential Dempster calibration is used, with a training set $\mathcal{L} = \{(X_1, Y_1), ..., (X_n, Y_n)\}$ where X_i is the discrete information provided by the source for the ith pixel.

4.2. Face detections as inputs to our approach

Consider strictly the same input information as in Section 3, that is J detectors each returning a set of bounding boxes with associated scores corresponding to the assumed positions of the faces. This section exposes how our approach can be applied in that case.

For a given pixel in an image and a given detector, two exclusive situations occur: either the pixel $p_{x,y}$ is contained by one of the box $B_{i,j}$ returned by the detector, or it is not. If it is contained by a box $B_{i,j}$, the score $S_{i,j}$ of the box is associated ("transferred") to the pixel. If the pixel does not belong to any box, no score is associated to it. As a consequence, the considered pixel either has an associated score, or it does not. These two situations are now detailed.

In the first case, when a score is available for the considered pixel, it is transformed into a MF using the evidential logistic regression and a training set, that we denote $\mathcal{L}_{calP,j}$. Let us describe this set $\mathcal{L}_{calP,j}$ underlying the transformation using calibration of a score $S_{i,j}$ associated to a pixel $p_{x,y}$ by a detector D_j, into a MF $m_{i,j}^{\mathcal{P}_{x,y}}$. For a given image ℓ, each pair $(B_{t,j}^{\ell}, S_{t,j}^{\ell})$ introduced in Section 3.2 yields, *via* "transfer", $\left|B_{t,j}^{\ell}\right|$ pairs $(p_{d,t,j}^{\ell}, S_{t,j}^{\ell})$, with $d = 1, \ldots, \left|B_{t,j}^{\ell}\right|$, and $\left|B_{t,j}^{\ell}\right|$ the number of pixels in box $B_{t,j}^{\ell}$, and where $p_{d,t,j}^{\ell}$ denotes the pixel in dth position in box $B_{t,j}^{\ell}$. From these data, we define $\mathcal{L}_{calP,j}$ as the set of pairs $(S_{t,j}^{\ell}, YP_{d,t,j}^{\ell})$, with $\ell = 1, \ldots, L$, $t = 1, \ldots, N_j^{\ell}$, and $d = 1, \ldots, \left|B_{t,j}^{\ell}\right|$, with $YP_{d,t,j}^{\ell} \in \{0, 1\}$ the label simply obtained by checking whether pixel $p_{d,t,j}^{\ell}$ belongs to some ground truth box G_r^{ℓ} in the image ℓ, *i.e.*,

$$YP_{d,t,j}^{\ell} = \begin{cases} 1 & \text{if } \exists\, G_r^{\ell},\ r = 1, \ldots, M^{\ell},\ \text{such that } p_{d,t,j}^{\ell} \in G_r^{\ell}, \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

$\mathcal{L}_{calP,j}$ may pose a complexity issue as $\left|\mathcal{L}_{calP,j}\right| = \sum_{\ell=1}^{L} \sum_{t=1}^{N_j^{\ell}} \left|B_{t,j}^{\ell}\right|$. To avoid this, one may use a smaller set $\mathcal{L}'_{calP,j} \subset \mathcal{L}_{calP,j}$, which represents roughly the same information as $\mathcal{L}_{calP,j}$ and built as follows: for each triple (ℓ, t, j), only 10 pairs among the pairs $(S_{t,j}^{\ell}, YP_{d,t,j}^{\ell})$, $d = 1, \ldots, \left|B_{t,j}^{\ell}\right|$, are selected such that the ratio

$$\frac{\left|\{YP_{d,t,j}^{\ell} | d = 1, ..., \left|B_{t,j}^{\ell}\right|, YP_{d,t,j}^{\ell} = 1\}\right|}{\left|\{YP_{d,t,j}^{\ell} | d = 1, ..., \left|B_{t,j}^{\ell}\right|, YP_{d,t,j}^{\ell} = 0\}\right|} \tag{21}$$

is preserved. $\mathcal{L}'_{calP,j}$ has then a size of $\left|\mathcal{L}'_{calP,j}\right| = 10 \sum_{\ell=1}^{L} N_j^{\ell}$.

Let us now consider the second situation, where a pixel is not contained by any of the boxes and thus does not have an associated score. Since it should be taken into account that detectors do not present the exact same performances (in particular, some may have many more pixels not in boxes than others), it seems interesting to calibrate this kind of outputs from detectors. We propose to do so by viewing this information of score absence as a discrete information, and thus by applying the evidential Dempster calibration. Specifically, the training set, denoted $\mathcal{L}_{*,j}$, necessary for this calibration is obtained using L images on which the detector D_j is applied. The number n_j of pixels of these images, which are not contained by any of the boxes returned by the detector D_j, can be obtained. For the ith of these n_j pixels, the absence of score can be encoded by the discrete information $X_i = 1$. As the ground truth of these L images is known, its associated true label Y_i is available. It is then possible to obtain a MF, denoted $m_{*,j}^{\mathcal{P}_{x,y}}$ and calculated using $\mathcal{L}_{*,j}$ and Eq. (13) with $j = 1$, representing the uncertainty with respect to the presence of a face on pixel $p_{x,y}$, when this pixel is not included in a box of detector D_j.

4.3. Comparison of both approaches

The proposed pixel-based approach presents several advantages over the one of Section 3. First, as can be seen in Section 4.2, the construction of the training set for calibration in case of pixels avoids the use of the parameter λ, whose value needs to be fixed either *a priori* (but then it is arguably arbitrary) or empirically.

Furthermore, our approach avoids the use of the clustering step, which also involves the parameter λ and that may behave non optimally in a multi-object situation, especially when they are close to each other, which may be the case with faces in a crowd.

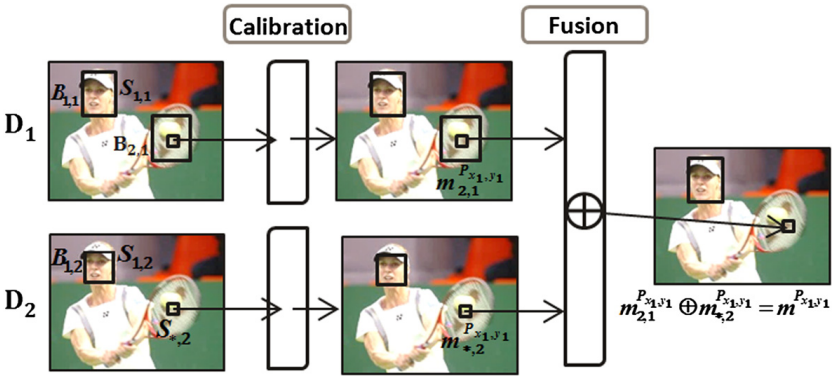


Fig. 2. Illustration of the pixel-based approach.

In addition, it allows us to have an arguably more consistent modeling of box absence than the box-based method. Indeed, in this latter method, for a given area in an image, there are two different modelings of box absence depending on the situation: either none of the detectors has provided a box, in which case the area is considered as non face, which amounts to considering that the detectors know that there is no face; or only a subset of the detectors has provided a box, in which case the other detectors are ignored, which is equivalent (under Dempster's rule) to considering that these detectors know nothing. By contrast, in the proposed method, the use of calibration enables us to take into account in a consistent manner the information of score absence into the fusion process, as when a detector D_j does not return a box for a given pixel $p_{x,y}$, its associated MF $m_{*,j}^{\mathcal{P}_{x,y}}$ is considered regardless of the outputs of the other detectors for this pixel. Thus, all detectors are involved in each fusion. Fig. 2 illustrates this point, highlighting the differences with the previous approach. For the sake of simplicity only one pixel, at the position (x_1, y_1), is considered here. Pixel p_{x_1,y_1} is contained by the box $B_{2,1}$, with $S_{2,1}$ as associated score, so the corresponding mass function is obtained trough the evidential logistic regression. However, there is no box containing p_{x_1,y_1} for the second detector, and thus it does not have an associated score. Yet, the opinion of the second detector is still taken into account *via* the MF $m_{*,2}^{\mathcal{P}_{x_1,y_1}}$ defined in Section 4.2.

As explained before, one of the disadvantages of the box-based approach is that the integration of a pixel-based information is not straightforward. In the proposed system however, a source of information which gives pixel-based information can be integrated into the fusion process as easily as a box-based information. It will be illustrated with an experiment in Section 6.

Finally, we note that locating the approach at the pixel level brings potentially a complexity issue. This will be discussed in the next section.

5. Experimental comparison of both solutions

In this section, the results of the proposed approach are presented and compared to those of the box-based method, when all available inputs are box-based information. The experiment is performed on a literature dataset as well as on another dataset, composed of images coming from cameras filming railway platforms. The experiment is first described, then the results are discussed.

5.1. Description of the experiment

We selected four face detectors based on machine learning techniques for which an open source implementation was available. The first detector is the one proposed by Viola and Jones [34], which is based on a classification algorithm called Adaboost and that uses Haar feature extraction. The second detector is a variant of the previous one: the same classification algorithm is used but with Local Binary Patterns (LBP) feature extraction [13]. The third detector relies on Support Vector Machine (SVM) and uses Histogram of Oriented Gradients (HOG) features [4,25]. It was provided by the DLIB library [21]. Following the current popularity of deep learning techniques, the fourth selected face detector is the deep neural network[1] proposed in [16], which is based on a compact design of a convolutional neural network and a cascade approach.

We used a literature dataset called Face Detection Data Set and Benchmark (FDDB) [14]. It contains the annotations (ground truth) for 5171 faces in a set of 2845 images. We trained both Adaboost-based detector with the same 2000 images of this dataset; the third and fourth detector were already trained. 200 others images were used for the calibration of the four detectors. The performances of the box-based and pixel-based approaches were then evaluated over the remaining 645 images.

Although the FDDB dataset presents various situations, we note that on the whole the images are generally of good quality and the faces of reasonable size. Thus, we also considered a more challenging dataset, composed of low-quality

[1] Available at https://github.com/Bkmz21/CompactCNNCascade.

(a) FDDB image (b) SNCF image

Fig. 3. Example of FDDB image (3a) and SNCF image (3b).

images and with situations where faces are more difficult to detect. These images, which we refer to as SNCF images, are extracted from video footage provided by video-protection cameras filming some railway platforms. We created a dataset of 600 images, containing multiple different conditions such as indoor and outdoor environment, different light settings and low image quality. The true positions of the 1089 faces on these images were manually annotated. Fig. 3 shows an example of images extracted from the two datasets. As there were not enough face examples in the SNCF dataset to train detectors, we used the two Adaboost-based detectors trained with the 2000 images of the FDDB dataset and the other two already trained detectors. Nonetheless, we calibrated these detectors using 100 annotated SNCF images. Performance tests were then conducted over the remaining 500 images.

The box-based approach returns MFs associated to boxes while our approach gives an MF for each pixel. Whatever the approach, to decide if a given pixel or a given box has to be blurred or not, we use the decision procedure relying on upper expected costs recalled in Section 2.1; in a binary case, they are simply defined by

$$R^*(\{0\}) = m^{\Omega}(\{1\})c(0,1) + m^{\Omega}(\{0,1\})c(0,1), \tag{22}$$

$$R^*(\{1\}) = m^{\Omega}(\{0\})c(1,0) + m^{\Omega}(\{0,1\})c(1,0), \tag{23}$$

by considering that the cost is equal to zero when the answer is correct ($c(0,0) = c(1,1) = 0$). As our purpose is to minimize the number of non-blurred faces, it is worse to consider a face as non-face than the opposite. In other words, decisions were made with costs such that $c(1,0) <= c(0,1)$. More specifically, we fixed $c(1,0) = 1$ and gradually increased $c(0,1)$ starting from $c(0,1) = 1$, to obtain different performance points. To quantify performances, we used the recall rate (proportion of pixels correctly blurred among the pixels to be blurred) and the precision rate (proportion of pixels correctly blurred among blurred pixels).

5.2. Results

Fig. 4 compare the results of the four selected detectors taken alone to that of our approach relying on a combination of their outputs, on the FDDB dataset. As it can be seen, the fusion of the four detectors outputs considerably increased the performances, as for example a precision of 80% gives a recall of around 52% for the Haar/Adaboost detector instead of 77% for the combination result. Let us note that the performances of the deep neural network face detector are only represented by a point because all the scores returned by this detector were similar, thus all the boxes have the same associated MF and increasing the cost $c(0,1)$ (the cost of deciding not to blur a pixel while it has to be) does not gradually increase the number of blurred pixels.

Fig. 5 shows the result for the same experiment but this time on the SNCF dataset. The conclusion is the same as the proposed approach has better performances than the detectors taken alone. Let us remark that their performances could be improved by training them with face and non-face images closer to those encountered in the SNCF dataset.

Comparison on the FDDB dataset between the box-based approach used with different values of the overlap threshold λ and our approach is shown in Fig. 6. As it can be noticed, for a same precision rate, the recall of our approach is always the highest. Fig. 7 shows the results of this comparison on the SNCF dataset; the conclusions are the same.

5.3. Discussion

Reasoning at the pixel level rather than with boxes as in the box-based approach may involve a complexity issue. Indeed, as the fusion is performed on every pixel instead of on sets of boxes, the proposed approach has *a priori* a higher complexity. For the pixel approach and for a given image, the number of operations is equal to $J \times a$, where J the number of fusion operations (which is equal to the number of used detectors) and a the number of pixels in the image. By contrast, in the

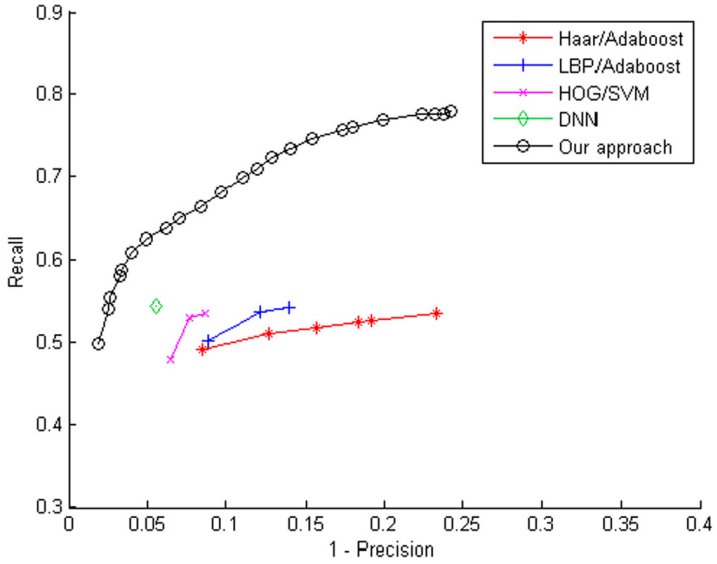


Fig. 4. Pixel-based approach vs detectors on FDDB.

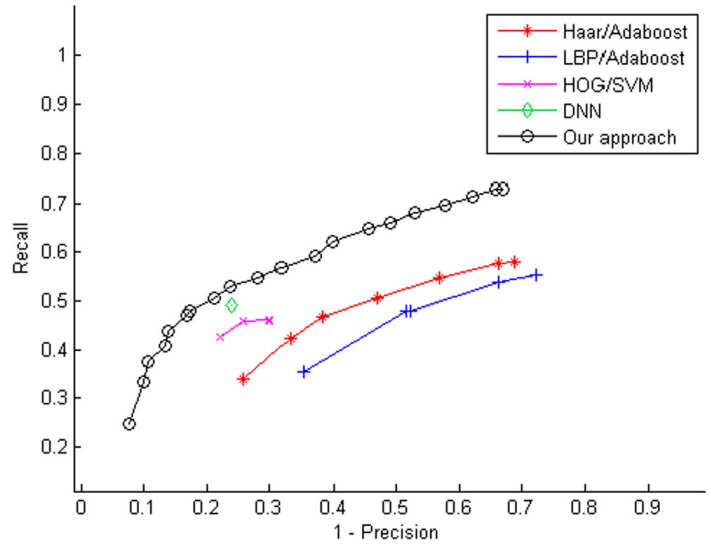


Fig. 5. Pixel-based approach vs detectors on SNCF dataset.

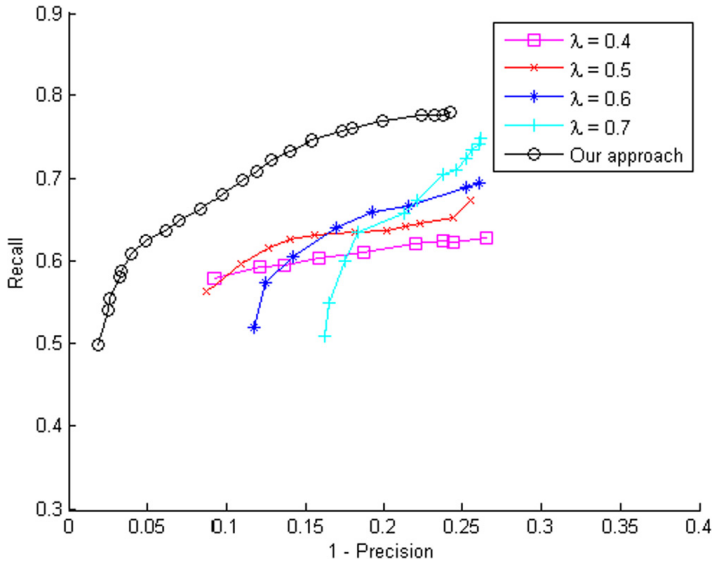


Fig. 6. Pixel-based approach vs box-based approach on FDDB.

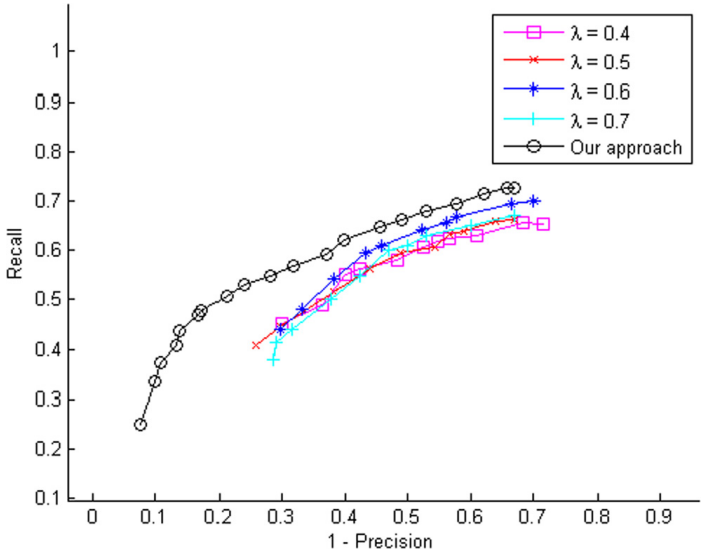


Fig. 7. Pixel-based approach vs box-based approach on SNCF dataset.

box-based approach, the complexity is $O(b^2)$, with b the total number of boxes returned by J detectors. Indeed, at worst the clustering procedure is $O(b^2)$ [8] and this is the most costly step. Thus, at first glance, it seems that the complexity is much higher for the proposed approach as a is generally significantly higher than b^2. However, any two pixels $p_{x,y}$ and $p_{x',y'}$ that do not belong to any box of D_j have associated MFs with the same definitions, *i.e.*, we have $m_{*,j}^{\mathcal{P}_{x,y}}(A) = m_{*,j}^{\mathcal{P}_{x',y'}}(A)$, for all $A \subseteq \{0,1\}$. Thus, pixels that do not belong to any of the returned boxes by the detectors have the same resulting MF. This latter case happens often in practice, hence this allows us to have a common processing. For instance, in a set of 200 images of FDDB, with the four face detectors considered in our experiment, it corresponds on average at around 80% of the pixels of the image. In terms of time processing, an image takes on average around 120 milliseconds to process (including the time of detection of the four detectors) for the box-based approach and 150 milliseconds for the proposed system; we consider that it is a reasonable difference.

This section showed that given the same information, *i.e.*, detectors returning boxes, the proposed approach gives better results than the box-based approach. Our approach is a little more time-consuming but the difference is reasonable. The following section illustrates another advantage of our approach, which is its ability to integrate directly sources providing pixel-based information.

6. Using pixel-based information

Color information can be useful for the face blurring problem as the color of the faces, the skin tone, is very distinct from others colors. It is thus an interesting information that can be used to detect skin, and thus faces, in complex scene images. It is actually a widely studied subject [3,15,33]. We used in this paper the same detector as in [30], with the same parameters, and which gives a classification of pixels as skin or non skin.

In order to combine this color information with the others detectors, a mass function has to be associated to each pixel of the image. As the used skin detector returns a binary decision, either skin or not skin, it returns a discrete information. Thus, this information can be calibrated using the evidential Dempster calibration. When a pixel $p_{x,y}$ is classified as skin by the skin detector, it is possible to obtain a MF representing the uncertainty with respect to the presence of a face on pixel $p_{x,y}$. The necessary training set is obtained using L images on which the skin detector is applied; the process is the same as in Section 4.2. The numbers n of pixels which have been classified as skin can be obtained, and for the ith of these n pixels the classification of this pixel as skin can be encoded by the discrete information $X_i = 1$. As the positions of the faces on these L images are available, its true label Y_i is available. Thus, using this training set and Eq. (13), the MF representing the uncertainty with respect to the presence of a face on pixel $p_{x,y}$ when this pixel is classified as skin can be calculated. In addition, given a pixel classified as non skin, the whole process can be applied to define a MF representing the uncertainty with respect to the presence of a face on pixel $p_{x,y}$.

The same experiment as in Section 5 was performed, including the four face detectors, the two different datasets, and the decision strategy. The repartition of the images for the calibration training and the tests was also the same. The fifth source, *i.e.*, the skin detector which gives information on pixels, was simply added to the global system. Fig. 8 compare the results of the pixel-based approach proposed in Section 5 and the new system now relying on a combination of the outputs of five detectors instead of four. The color detector is only represented by one point in Fig. 8 because all the pixels considered as skin have the same MF, likewise for the pixels indicating non skin. Thus, as for the deep neural network detector, increasing the cost $c(0,1)$ does not gradually increase the number of blurred pixels. Actually, at some value of cost

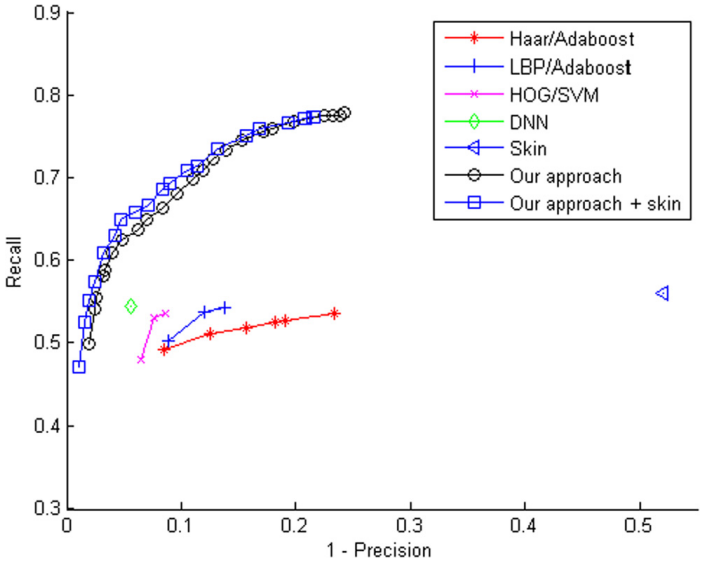


Fig. 8. Integration of skin color information to the proposed approach on FDDB.

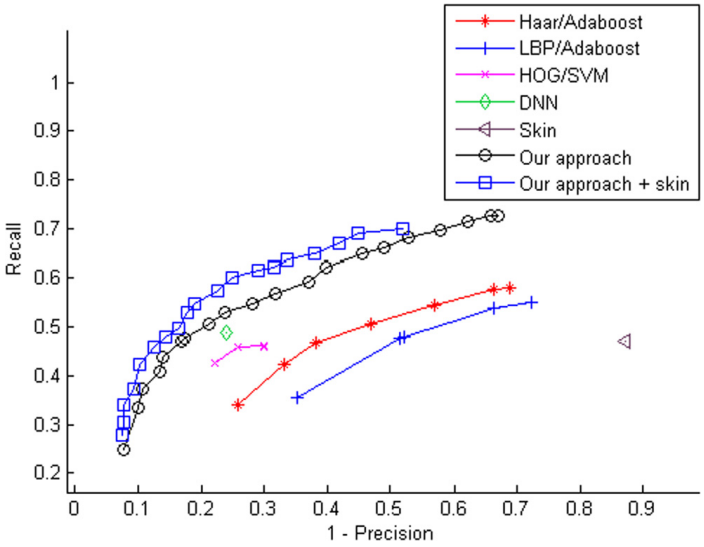


Fig. 9. Integration of skin color information to the proposed approach on SNCF dataset.

$c(0, 1)$, which is not represented in Fig. 8, a second point for the color detector is obtained but it corresponds to a useless point where all the pixels are blurred by the color detector.

As it can be noticed, the addition of the skin color information improves the global combination although the performance of skin detection is not that good. Finally, we conducted the experiment on the SNCF dataset and the results are shown in Fig. 9. It also improves the overall performances.

7. Conclusion

In this paper, a pixel-based face blurring system relying on evidential calibration and fusion of several detector outputs was proposed. This pixel-based approach brings several advantages over a previous box-based proposal. First, an overlap threshold is no longer necessary, as well as a clustering step. Furthermore, it enables to integrate all kind of detectors, either returning discrete or continuous information, as well as pixel-based or box-based information. In particular, in the considered blurring problem, it allows us to model and to integrate to the fusion process the information of score absence for each detector, *i.e.*, a MF is defined for pixels which are not contained by any of the boxes returned by the detector. The proposed system also shown better performances than the box-based approach, either on a literature dataset or on a more challenging one. We also illustrated the ability of natively integrating a detector giving pixel-based outputs by adding a skin color detector to the global system; this latter addition further improved the overall performances.

The proposed approach can be applied with other detectors, which may return discrete or continuous information, and can be based on boxes or pixels. One perspective consists in replacing one of the face detectors, or to add one to the global system.

Another perspective is to make use of the spatio-temporal context of a given pixel. It is reasonable to consider that a pixel is more likely to be blurred if its neighbors have been blurred. Similarly, if videos have to be handled instead of still images, one could use the fact that a pixel is more likely to be blurred if it has been blurred on a previous image. Taking advantage of such contextual information is an important field of application of the Markov random field theory [19,20,22], which has been extended to the evidential framework in [31,12], and it could be an inspiration to extend our approach or, alternatively, a detector using this theory could be added to the global system.

References

[1] P. Aarabi, J.C.L. Lam, A. Keshavarz, Face detection using information fusion, in: Proceedings of the 10th International Conference on Information Fusion, July 2007, pp. 1–8, Quebec, Canada.
[2] M. Ayer, H. Brunk, G. Ewing, W. Reid, E. Silverman, An empirical distribution function for sampling with incomplete information, Ann. Math. Stat. 26 (4) (1955) 641–647.
[3] J. Brand, J.S. Mason, A comparative assessment of three approaches to pixel-level human skin-detection, in: Proceedings of the 15th International Conference on Pattern Recognition, vol. 1, Sept. 2000, pp. 1056–1059, Barcelona, Spain.
[4] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, vol. 1, June 2005, pp. 886–893, San Diego, California.
[5] A.P. Dempster, New methods for reasoning towards posterior distributions based on sample data, Ann. Math. Stat. 37 (2) (1966) 355–374.
[6] T. Denœux, Analysis of evidence-theoretic decision rules for pattern classification, Pattern Recognit. 30 (7) (1997) 1095–1107.
[7] T. Denœux, Likelihood-based belief function: justification and some extensions to low-quality data, Int. J. Approx. Reason. 55 (7) (2014) 1535–1547.
[8] P. Dollár, Z. Tu, P. Perona, S. Belongie, Integral channel features, in: Proceedings of the British Machine Vision Conference, Sept. 2009, pp. 91.1–91.11, London, England.
[9] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The PASCAL visual object classes (VOC) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338.
[10] S.S. Farfade, M. Saberian, L.-J. Li, Multi-view face detection using deep convolutional neural networks, in: Proceedings of the International Conference on Multimedia Retrieval, June 2015, Shanghai, China.
[11] F. Faux, F. Luthon, Theory of evidence for face detection and tracking, Int. J. Approx. Reason. 53 (5) (2012) 728–746.
[12] L. Fouque, A. Appriou, W. Pieczynski, An evidential Markovian model for data fusion and unsupervised image classification, in: Proceedings of the Third International Conference on Information Fusion, vol. 1, July 2000, p. TUB4-25, Paris, France.
[13] A. Hadid, M. Pietikäinen, T. Ahonen, A discriminative feature space for detecting and recognizing faces, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, vol. 2, June 2004, p. II-797, Washington, DC.
[14] V. Jain, E. Learned-Miller, FDDB: a Benchmark for Face Detection in Unconstrained Settings, Technical Report UM-CS-2010-009, Univ. Massachusetts, 2010.
[15] M.J. Jones, J.M. Rehg, Statistical color models with application to skin detection, Int. J. Comput. Vis. 46 (1) (2002) 81–96.
[16] I.A. Kalinovskii, V.G. Spitsyn, Compact convolutional neural network cascade for face detection, in: Proceedings of the 10th Annual International Scientific Conference on Parallel Computing Technologies, PCT, vol. 1576, March 2016, pp. 375–387, Arkhangelsk, Russia.
[17] O. Kanjanatarakul, T. Denœux, S. Sriboonchitta, Prediction of future observations using belief functions: a likelihood-based approach, Int. J. Approx. Reason. 72 (2016) 71–94.
[18] O. Kanjanatarakul, S. Sriboonchitta, T. Denœux, Forecasting using belief functions: an application to marketing econometrics, Int. J. Approx. Reason. 55 (5) (2014) 1113–1128.
[19] S.H. Khatoonabadi, I.V. Bajic, Video object tracking in the compressed domain using spatio-temporal Markov random fields, IEEE Trans. Image Process. 22 (1) (2013) 300–313.
[20] R. Kindermann, L. Snell, Markov Random Fields and Their Applications, American Mathematical Society, 1980.
[21] D.E. King, Dlib-ml: a machine learning toolkit, J. Mach. Learn. Res. 10 (2009) 1755–1758.
[22] S.Z. Li, Markov random field models in computer vision, in: Proceedings of the European Conference on Computer Vision, May 1994, pp. 361–370, Stockholm, Sweden.
[23] G.C. Luh, Face detection using combination of skin color pixel detection and Viola–Jones face detector, in: Proceedings of the International Conference on Machine Learning and Cybernetics, vol. 1, July 2014, pp. 364–370, Lanzhou, China.
[24] H.T. Nguyen, An Introduction to Random Sets, Chapman and Hall/CRC Press, 2006.
[25] E. Osuna, R. Freund, F. Girosi, Training support vector machines: an application to face detection, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, June 1997, pp. 130–136, San Juan, Puerto Rico.
[26] J.C. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, Adv. Larg. Margin Classif. 10 (3) (1999) 61–74.
[27] H.A. Rowley, S. Baluja, T. Kanade, Neural network-based face detection, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1) (1998) 23–38.
[28] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[29] P. Smets, R. Kennes, The transferable belief model, Artif. Intell. 66 (1994) 191–243.
[30] M. Soriano, B. Martinkauppi, S. Huovinen, M. Laaksonen, Using the Skin Locus to Cope with Changing Illumination Conditions in Color-Based Face Tracking, Proceedings of the Nordic Signal Processing Symposium, vol. 38, June 2000, pp. 383–386, Kolmarden, Sweden.
[31] H. Soubaras, On evidential Markov chains, in: Foundations of Reasoning Under Uncertainty, Springer, 2010, pp. 247–264.
[32] Z.S. Tabatabaie, R.W. Rahmat, N.I.B. Udzir, E. Kheirkhah, A hybrid face detection system using combination of appearance-based and feature-based methods, Int. J. Comput. Sci. Netw. Secur. 9 (5) (2009) 181–185.
[33] J-C. Terrillon, M.N. Shirazi, H. Fukamachi, S. Akamatsu, Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images, in: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, March 2000, pp. 54–61, Grenoble, France.
[34] P. Viola, M.J. Jones, Robust real-time face detection, Int. J. Comput. Vis. 57 (2) (2004) 137–154.
[35] P. Xu, F. Davoine, T. Denœux, Evidential combination of pedestrian detectors, in: Proceedings of the 25th British Machine Vision Conference, BMVC, Sept. 2014, pp. 1–14, Nottingham, England.
[36] P. Xu, F. Davoine, H. Zha, T. Denœux, Evidential calibration of binary SVM classifiers, Int. J. Approx. Reason. 72 (2016) 55–70.
[37] S. Yang, P. Luo, C.C. Loy, X. Tang, From facial parts responses to face detection: a deep learning approach, in: Proceedings of International Conference on Computer Vision, ICCV, Dec. 2015, pp. 3676–3684, Santiago, Chile.

[38] B. Zadrozny, C. Elkan, Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers, in: Proceedings of the International Conference on Machine Learning, June 2001, pp. 609–616, Williamstown, Massachusetts.
[39] B. Zadrozny, C. Elkan, Transforming classifier scores into accurate multiclass probability estimates, in: Proceedings of the International Conference on Knowledge Discovery and Data Mining, July 2002, pp. 694–699, Edmonton, Canada.
[40] C. Zhang, Z. Zhang, Improving multiview face detection with multi-task deep convolutional neural networks, in: Proceedings of the Winter Conference on Applications of Computer Vision, March 2014, pp. 1036–1041, Steamboat Springs, Colorado.

Annexe E

Evidential joint calibration of binary SVM classifiers

Article accepté pour publication dans *Soft Computing*, juillet 2018.

Soft Computing
https://doi.org/10.1007/s00500-018-3429-x

FOUNDATIONS

Evidential joint calibration of binary SVM classifiers

Pauline Minary[1,2] · Frédéric Pichon[1] · David Mercier[1] · Eric Lefevre[1] · Benjamin Droit[2]

Abstract
In order to improve overall performance with respect to a classification problem, a path of research consists in using several classifiers and to fuse their outputs. To perform this fusion, some approaches merge the classifier outputs using a rule of combination. This requires that the outputs be made comparable beforehand, which is usually done thanks to a probabilistic calibration of each classifier. The fusion can also be performed by concatenating the classifier outputs into a vector and applying a joint probabilistic calibration to this vector. Recently, extensions of probabilistic calibration techniques of an individual classifier have been proposed using evidence theory, in order to better represent the uncertainties inherent to the calibration process. In this paper, we adapt this latter idea to joint probabilistic calibration techniques, leading to evidential versions of joint calibration techniques. In addition, our proposal was tested on generated and real datasets and the results showed that it either outperforms or is comparable to state-of-the-art approaches.

Keywords Belief functions · Information fusion · Evidential calibration · Classification

1 Introduction

An important path of research in classification consists in using several classifiers, which are trained with different data or based on different training models, instead of relying on a single one (Kuncheva 2004). Since they do not necessarily give the same output after observing a given object, a central issue in this approach consists in figuring out how to exploit these outputs to classify this object.

There are different ways of performing the fusion of some classifier outputs (Kuncheva 2004; Tulyakov et al. 2008). These various fusion methods are usually separated into two categories: the non-trainable and trainable combiners.

In the first category, the outputs returned by the classifiers after observing a given object are combined using a predetermined rule of combination. As the used classifiers are different, their outputs are not scaled with respect to each other and thus have to be made comparable before being combined. A step called calibration (Platt 1999) is thus usually performed to transform each output into a probability. In particular, the three calibration techniques the most commonly used are based on binning (Zadrozny and Elkan 2001), isotonic regression (Zadrozny and Elkan 2002) and logistic regression (Platt 1999). These calibration techniques suffer from an over-fitting problem, especially when only few training data are available. Within this scope, Xu et al. (2016) recently proposed a refinement of the main calibration procedures within a framework for reasoning under uncertainty called evidence theory (Shafer 1976; Smets and Kennes 1994). This theory allows Xu et al. to model more precisely the uncertainties inherent to such calibration process and thus to prevent the over-fitting issue. Xu et al. (2016) used this refinement to propose an approach of the non-trainable kind

Communicated by A. Di Nola.

This paper is an extended and revised version of (Minary et al. 2017).

Pauline Minary
pminary0@gmail.com

Frédéric Pichon
frederic.pichon@univ-artois.fr

David Mercier
david.mercier@univ-artois.fr

Eric Lefevre
eric.lefevre@univ-artois.fr

Benjamin Droit
benjamin.droit@reseau.sncf.fr

[1] EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), Univ. Artois, Bethune 62400, France

[2] Département des Télécommunications, SNCF Réseau, La Plaine Saint Denis, France

Published online: 31 July 2018

for binary classification problems. This latter approach consists in: using several SVM classifiers returning confidence scores, calibrating each of the returned scores using an evidential calibration technique, hence transforming each of the score into a belief function, and finally merging them using Dempster's rule of combination (Shafer 1976).

The second category regroups the approaches using the concatenation of the outputs of the classifiers as an input vector for another classifier. In particular, the approach defined in (Zhong and Kwok 2013) is a member of that category as a vector of scores obtained from an ensemble of classifiers is provided as an input vector to a probabilistic classifier based on multiple isotonic regression. Note that such kind of approach may be regarded as a probabilistic *joint* calibration as it learns how to convert a vector of scores into a probability, that is, it calibrates jointly the classifiers. In addition, as logistic regression can also be defined with multiple inputs (Hosmer et al. 2013), one may envisage to extend this kind of approach to the logistic model.

Both categories present some disadvantages. As already mentioned, the calibration techniques used in the non-trainable combiners are prone to an over-fitting problem. In addition, non-trainable combiners rely on a fixed rule of combination; as explained by Duin (2002), a predetermined rule may be the best combination only under very strict conditions, and an improved result may be obtained using an approach of the trainable combiner category. For the trainable combiners, a training set common to all classifiers is required, and the combiner must be re-learned each time a new classifier is added to the system. Furthermore, trainable combiner approaches corresponding to a probabilistic joint calibration may also be prone to the over-fitting problem inherent to probabilistic calibration.

Within this scope, we propose in this paper to study the application of the appealing element of Xu et al.'s approach (Xu et al. 2016), i.e., the evidential extension of calibration, to joint calibration techniques. As a result, we obtain methods that transform the vector of scores returned by the classifiers for a given object into a belief function.

This paper is organized as follows. First, necessary background on evidence theory is recalled in Sect. 2. In Sect. 3, probabilistic calibration methods of a single classifier are presented, followed by their extension using the evidence theory. Then, probabilistic joint calibrations and their extension to the evidential framework that we propose are exposed in Sect. 4. In Sect. 5, the proposed approach is compared experimentally to other approaches, and in particular to Xu et al. non-trainable combiner approach relying on evidential calibration of individual classifiers and to probabilistic joint calibration. Finally, conclusion and perspectives are given in Sect. 6.

2 Evidence theory

Basic notions of the theory of evidence (Shafer 1976; Smets and Kennes 1994) are first exposed in Sect. 2.1. Applications of this theory to statistical inference and prediction, which are useful to derive calibration in the evidential framework, are addressed in Sects. 2.2 and 2.3.

2.1 Basic notions

Evidence theory, also referred to as belief function theory, is a general framework for modeling uncertainty. Let $\boldsymbol{\omega}$ be a variable whose possible values belong to the finite set $\Omega = \{\omega_1, \ldots, \omega_K\}$. In this theory, uncertainty with respect to the actual value ω_0 taken by $\boldsymbol{\omega}$ is represented using a *Mass Function* (MF) defined as a mapping $m^\Omega : 2^\Omega \rightarrow [0, 1]$ verifying $m^\Omega(\emptyset) = 0$ and

$$\sum_{A \subseteq \Omega} m^\Omega(A) = 1. \tag{1}$$

The quantity $m^\Omega(A)$ corresponds to the belief committed exactly to the hypothesis $\omega_0 \in A$ and nothing more specific. Any subset A of Ω such that $m^\Omega(A) > 0$ is called a focal set of m^Ω. When the focal sets are nested, m^Ω is said to be consonant.

Equivalent representations of a mass function exist. In particular, the belief and plausibility functions are, respectively, defined by

$$Bel^\Omega(A) = \sum_{B \subseteq A} m^\Omega(B), \quad \forall A \subseteq \Omega, \tag{2}$$

$$Pl^\Omega(A) = \sum_{B \cap A \neq \emptyset} m^\Omega(B), \quad \forall A \subseteq \Omega. \tag{3}$$

The degree of belief $Bel^\Omega(A)$ measures the amount of evidence strictly in favor of the hypothesis $\omega_0 \in A$, while the plausibility $Pl^\Omega(A)$ is the amount of evidence not contradicting it. The plausibility function restricted to singletons is called the contour function, denoted pl^Ω and defined by

$$pl^\Omega(\omega) = Pl^\Omega(\{\omega\}), \quad \forall \omega \in \Omega. \tag{4}$$

When a mass function is consonant, the plausibility function can be recovered from its contour function as follows:

$$Pl^\Omega(A) = \sup_{\omega \in A} pl^\Omega(\omega), \quad \forall A \subseteq \Omega. \tag{5}$$

Given two independent MFs m_1^Ω and m_2^Ω, it is possible to combine them using Dempster's rule of combination. The

result of this combination is a MF $m^{\Omega}_{1\oplus 2}$ defined by

$$m^{\Omega}_{1\oplus 2}(A) = \frac{1}{1-\kappa} \sum_{B\cap C=A} m^{\Omega}_1(B) m^{\Omega}_2(C), \qquad \forall A \neq \emptyset, \tag{6}$$

where

$$\kappa = \sum_{B\cap C=\emptyset} m^{\Omega}_1(B) m^{\Omega}_2(C), \tag{7}$$

represents the degree of conflict between m^{Ω}_1 and m^{Ω}_2, and $m^{\Omega}_{1\oplus 2}(\emptyset) = 0$. If $\kappa = 1$, there is a total conflict between the two pieces of evidence and they cannot be combined.

Different decision strategies exist to make a decision about the actual value ω_0 of $\boldsymbol{\omega}$, given a MF m^{Ω} (Denœux 1997). In particular, the value $\omega \in \Omega$ having the smallest so-called *upper* or *lower expected costs* may be selected. The upper and lower expected costs of some value $\omega \in \Omega$, respectively denoted by $R^*(\omega)$ and $R_*(\omega)$, are defined as

$$R^*(\omega) = \sum_{A\subseteq\Omega} m^{\Omega}(A) \max_{\omega'\in A} c(\omega, \omega'), \tag{8}$$

$$R_*(\omega) = \sum_{A\subseteq\Omega} m^{\Omega}(A) \min_{\omega'\in A} c(\omega, \omega'), \tag{9}$$

where $c(\omega, \omega')$ is the cost of deciding ω when the true answer is ω'. When the set of focal elements is reduced to singletons and Ω, and when the costs are taken equal to 0 if $\omega = \omega'$ and 1 otherwise, the upper and lower expected costs are, respectively, defined as

$$\begin{aligned} R^*(\omega) &= 1 - m^{\Omega}(\{\omega\}), \\ &= 1 - Bel^{\Omega}(\{\omega\}). \end{aligned} \tag{10}$$

$$\begin{aligned} R_*(\omega) &= 1 - m^{\Omega}(\{\omega\}) - m^{\Omega}(\Omega), \\ &= 1 - Pl^{\Omega}(\{\omega\}). \end{aligned} \tag{11}$$

Choosing the value ω minimizing the lower (resp. upper) expected costs is called the optimistic (resp. pessimistic) strategy.

To avoid making wrong decisions in the risky cases, i.e., when the expected costs are high, a reject decision may be introduced. Formally, a reject cost $R_{rej} \in [0, 1]$ is introduced and a decision to reject is made when R_{rej} is lower than the other expected costs.

2.2 Statistical inference

The theory of evidence can be used for statistical inference. Consider $\theta \in \Theta$ an unknown parameter, $x \in \mathbb{X}$ some observed data and $f_\theta(x)$ the density function generating the data. Statistical inference consists in making statements about θ after observing the data x. Shafer (1976) proposed to represent knowledge about θ given x by a consonant belief function Bel^{Θ}_x based on the likelihood function $L_x : \theta \to f_\theta(x)$ (see also justifications by Denœux (2014)), whose contour function is the normalized likelihood function:

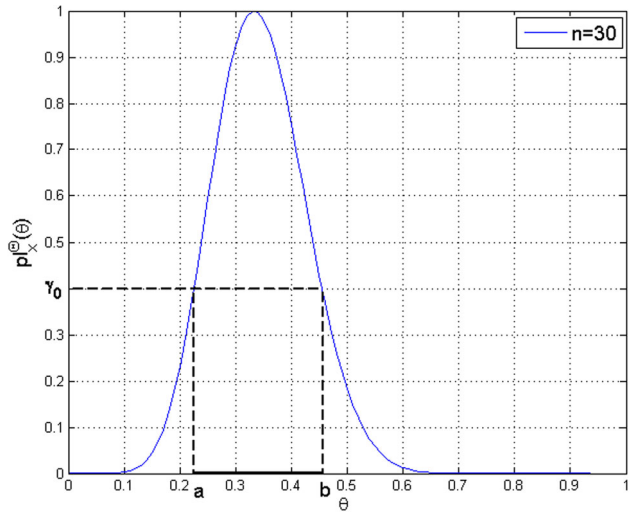


Fig. 1 Contour function of a binomial distribution, with $n = 30$ and $x = 10$

$$pl^{\Theta}_x(\theta) = \frac{L_x(\theta)}{\sup_{\theta'\in\Theta} L_x(\theta')}, \qquad \forall \theta \in \Theta. \tag{12}$$

Let us consider an important particular case. Assume that we observe a random variable X, which has a binomial distribution with parameters $n \in \mathbb{N}$ and $\theta \in [0, 1]$, i.e., $X \sim \mathcal{B}(n, \theta)$. In that case, we have

$$f_\theta(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}. \tag{13}$$

The likelihood-based belief function has the following contour function:

$$pl^{\Theta}_x(\theta) = \frac{\theta^x (1-\theta)^{n-x}}{\hat{\theta}^x (1-\hat{\theta})^{n-x}}, \tag{14}$$

for all $\theta \in \Theta = [0, 1]$, where $\hat{\theta} = \frac{x}{n}$ is the Maximum Likelihood Estimate (MLE) of θ. Figure 1 shows the contour function of the binomial distribution, with $n = 30$ and $x = 10$.

2.3 Forecasting

Let us now suppose that we have some knowledge about $\theta \in \Theta$ after observing some data x, given under a form of a consonant belief function Bel_x^{Θ}. The aim of forecasting is to make statements about a not yet observed data $Y \in \mathbb{Y}$, whose conditional distribution $g_{x,\theta}(y)$ given $X = x$ depends on θ. A solution to this problem, proposed by Kanjanatarakul et al. (2014, 2016), consists in using the fact that Bel_x^{Θ} is equivalent to a random set, and in using the sampling model of Dempster (Dempster 1966) to deduce a belief function on $\mathbb{Y}$. We detail these two points below.

Let us recall that the focal sets of Bel_x^{Θ} are the level sets of pl_x^{Θ}, defined by (Nguyen 2006)

$$\Gamma_x(\gamma) = \left\{\theta \in \Theta | pl_x^{\Theta}(\theta) \geq \gamma\right\}, \qquad \forall \gamma \in [0, 1]. \tag{15}$$

For instance in Fig. 1, for $\gamma = \gamma_0 = 0.4$, the set $\Gamma_x(\gamma_0)$ is defined as the set of all values of $\theta \in \Theta$ such that $pl_x^{\Theta}(\theta) \geq 0.4$, *i.e.*, $\Gamma_x(\gamma_0) = [a, b] \approx [0.225, 0.454]$. Moreover, the belief function Bel_x^{Θ} is equivalent to the random set induced by the Lebesgue measure λ on $[0, 1]$ and the multi-valued mapping $\Gamma_x : [0, 1] \to \Theta$ (Nguyen 2006). Thus, we have

$$Bel_x^{\Theta}(A) = \lambda(\{\gamma \in [0, 1] | \Gamma_x(\gamma) \subseteq A\}), \tag{16}$$
$$Pl_x^{\Theta}(A) = \lambda(\{\gamma \in [0, 1] | \Gamma_x(\gamma) \cap A \neq \emptyset\}), \tag{17}$$

for all $A \subseteq \Theta$.

The sampling model of Dempster proposes to express Y using a function φ depending on the parameter θ and some unobserved variable $Z \in \mathbb{Z}$, whose probability distribution μ is known and independent of θ:

$$Y = \varphi(\theta, Z). \tag{18}$$

From Eqs. (15) and (18), for a given $(\gamma, z) \in [0, 1] \times \mathbb{Z}$, we can assert that $Y \in \varphi(\Gamma_x(\gamma), z)$. This can be represented by a multi-valued mapping $\Gamma'_x : [0, 1] \times \mathbb{Z} \to \mathbb{Y}$ defined by composing Γ_x with φ, i.e., $\Gamma'_x(\gamma, z) = \varphi(\Gamma_x(\gamma), z), \forall (\gamma, z) \in [0, 1] \times \mathbb{Z}$. The product measure $\lambda \otimes \mu$ on $[0, 1] \times \mathbb{Z}$ and the multi-valued mapping Γ'_x induce the belief and plausibility functions on $\mathbb{Y}$, which are defined by

$$Bel_x^{\mathbb{Y}}(A) = (\lambda \otimes \mu)(\{(\gamma, z) | \varphi(\Gamma_x(\gamma), z) \subseteq A\}), \tag{19}$$
$$Pl_x^{\mathbb{Y}}(A) = (\lambda \otimes \mu)(\{(\gamma, z) | \varphi(\Gamma_x(\gamma), z) \cap A \neq \emptyset\}), \tag{20}$$

for all $A \subseteq \mathbb{Y}$.

Let us consider a binary case, which will be useful hereafter. Let $Y \in \mathbb{Y} = \{0, 1\}$ be a random variable with a Bernoulli distribution, i.e., $Y \sim \mathcal{B}(\theta)$. In that case, the function φ can be defined as follows:

$$Y = \varphi(\theta, Z) = \begin{cases} 1, & \text{if } Z \leq \theta, \\ 0, & \text{otherwise,} \end{cases} \tag{21}$$

with Z having a uniform distribution on $[0, 1]$. Assume that the consonant belief function Bel_x^{Θ} has a unimodal and continuous contour function pl_x^{Θ}. In that case, each level set of Bel_x^{Θ} is a closed interval, i.e., $\Gamma_x(\gamma) = [U(\gamma), V(\gamma)]$ (Dempster 1968), and the multi-valued mapping Γ'_x defined by composing Γ_x with φ is given by

$$\Gamma'_x(\gamma, z) = \varphi([U(\gamma), V(\gamma)], z) = \begin{cases} \{1\}, & \text{if } z \leq U(\gamma), \\ \{0\}, & \text{if } z > V(\gamma), \\ \{0, 1\}, & \text{otherwise.} \end{cases} \tag{22}$$

By applying Eq. (19), we get

$$Bel_x^{\mathbb{Y}}(\{1\}) = (\lambda \otimes \mu)(\{(\gamma, z) | z \leq U(\gamma)\}), \tag{23}$$
$$Bel_x^{\mathbb{Y}}(\{0\}) = (\lambda \otimes \mu)(\{(\gamma, z) | z > V(\gamma)\}). \tag{24}$$

Xu et al. (2016) showed that in this situation, the belief function $Bel_x^{\mathbb{Y}}$ and plausibility function $Pl_x^{\mathbb{Y}}$ are defined by

$$Bel_x^{\mathbb{Y}}(\{1\}) = \hat{\theta} - \int_0^{\hat{\theta}} pl_x^{\Theta}(u)du, \tag{25}$$
$$Pl_x^{\mathbb{Y}}(\{1\}) = \hat{\theta} + \int_{\hat{\theta}}^1 pl_x^{\Theta}(v)dv, \tag{26}$$

where $\hat{\theta}$ maximizes pl_x^{Θ}.

Let us consider again the particular case of Sect. 2.2, where $X \sim \mathcal{B}(n, \theta)$. In that case, the contour function on Θ defined in Eq. (14) is unimodal and continuous, as illustrated in Fig. 1. Thus, to represent knowledge about an unobserved data $Y \in \mathbb{Y}$, with $Y \sim \mathcal{B}(\theta)$, we can apply Eqs. (25) and (26) and Xu et al. showed that the obtained belief and plausibility functions boil down in that case to (Xu et al. 2016):

$$Bel_x^{\mathbb{Y}}(\{1\}) = \begin{cases} 0, & \text{if } \hat{\theta} = 0, \\ \hat{\theta} - \frac{\underline{B}(\hat{\theta}; x+1, n-x+1)}{\hat{\theta}^x(1-\hat{\theta})^{n-x}}, & \text{if } 0 < \hat{\theta} < 1, \\ \frac{n}{n+1}, & \text{if } \hat{\theta} = 1, \end{cases} \tag{27}$$

$$Pl_x^{\mathbb{Y}}(\{1\}) = \begin{cases} \frac{1}{n+1}, & \text{if } \hat{\theta} = 0, \\ \hat{\theta} + \frac{\overline{B}(\hat{\theta}; x+1, n-x+1)}{\hat{\theta}^x(1-\hat{\theta})^{n-x}}, & \text{if } 0 < \hat{\theta} < 1, \\ 1, & \text{if } \hat{\theta} = 1, \end{cases} \tag{28}$$

where $\underline{B}$ and $\overline{B}$ are, respectively, the lower and upper incomplete beta functions, defined when a and b are integers and

$0 < z < 1$ by

$$\underline{B}(z; a, b) = \sum_{j=a}^{a+b-1} \frac{(a-1)!(b-1)!}{j!(a+b-1-j)!} z^j (1-z)^{a+b-1-j}, \tag{29}$$

and

$$\overline{B}(z; a, b) = \underline{B}(1-z; b, a). \tag{30}$$

3 Calibration of a single binary SVM classifier

Let us consider an object, whose true label y is such that $y \in \mathbb{Y} = \{0, 1\}$, and a confidence score $s \in \mathbb{R}$ returned by a classifier after observing this object. To learn how to interpret what this score represents with respect to y, a step called calibration may be used. This step relies on a training set $\mathcal{X}$, which contains n other objects for which the label is known, and for which we observed the score that the classifier returned, i.e., $\mathcal{X} = \{(s_1, y_1), \ldots, (s_n, y_n)\}$ where s_i represents the score given by the classifier for the ith object whose true label is y_i. The calibration procedures commonly used are the binning (Zadrozny and Elkan 2001), isotonic regression (Zadrozny and Elkan 2002) and logistic regression (Platt 1999). This paper focuses on binning and logistic regression as the isotonic regression can be seen as an intermediary approach between these two (Zadrozny and Elkan 2002). The probabilistic version of these two calibrations is described in Sect. 3.1, followed by their extension to the evidential framework in Sect. 3.2.

3.1 Probabilistic calibration of a single classifier

Given a score $s \in \mathbb{R}$ returned by a classifier after observing a given object, the aim of the calibration in the probabilistic framework consists in estimating the probability distribution $P^{\mathbb{Y}}(\cdot|s)$.

3.1.1 Binning

The binning approach consists in dividing the score spaces into B_U different bins, for example $(-3; -2]$, $(-2; -1]$, etc. For each bin j, the number k_j of couples $(s_i, y_i) \in \mathcal{X}$ such as $y_i = 1$ and s_i is in bin j, and the number n_j of couples $(s_i, y_i) \in \mathcal{X}$ such as s_i is in bin j can be obtained. Then, for a score s such that s belongs to bin j, we have

$$P^{\mathbb{Y}}(y = 1|s) = \frac{k_j}{n_j}. \tag{31}$$

3.1.2 Logistic regression

The calibration based on logistic regression proposed by Platt (1999) is a more elaborate method, which is based on fitting a sigmoid function h defined by

$$P^{\mathbb{Y}}(y = 1|s) \approx h_s(\sigma) = \frac{1}{1 + e^{(\sigma_0 + \sigma_1 s)}}, \tag{32}$$

where the parameter $\sigma = (\sigma_0, \sigma_1) \in \mathbb{R}^2$ is chosen as the one maximizing the following likelihood function:

$$L_{\mathcal{X}}(\sigma) = \prod_{i=1}^{n} p_i^{t_i} (1 - p_i)^{1-t_i}, \tag{33}$$

with

$$p_i = \frac{1}{1 + e^{(\sigma_0 + \sigma_1 s_i)}}, \tag{34}$$

and

$$t_i = \begin{cases} \frac{N_+ + 1}{N_+ + 2} & \text{if } y_i = 1, \\ \frac{1}{N_- + 2} & \text{if } y_i = 0, \end{cases} \tag{35}$$

where N_+ and N_- are, respectively, the number of positive and negative samples in the training set $\mathcal{X}$.

Yet, it is usually easier to maximize the log-likelihood instead, which is defined by

$$\ell_{\mathcal{X}}(\sigma) = \log L_{\mathcal{X}}(\sigma) \tag{36}$$

$$= \sum_{i}^{n} \big(t_i \log(p_i) + (1 - t_i) \log(1 - p_i)\big). \tag{37}$$

Since the logarithm function is a strictly increasing function, maximizing the logarithm of the likelihood is the same as maximizing the likelihood. The parameter σ maximizing this log-likelihood function can be approximated using iterative methods such as gradient descent. As the log-likelihood function of the logistic regression is concave (Minka 2003), all local maxima are global maxima and thus an unique solution is found for σ.

We may notice that the less training samples are available, the more the estimated probabilities are uncertain. Within this scope, Xu et al. (2016) proposed to refine the above calibrations using the theory of evidence, in order to better handle the uncertainties. The following section recalls the evidential versions of the binning and logistic regression calibration procedures.

3.2 Evidential calibration of a single classifier

Xu et al. have recently extended the probabilistic calibration methods to the evidential framework (Xu et al. 2016). In their approach, the calibration of a given score s is seen as a prediction problem of a Bernoulli variable $Y \in \mathbb{Y} = \{0, 1\}$ with parameter θ, where uncertainty on θ depends on s. They studied different models to estimate the uncertainty on θ and highlighted in particular the benefits of the so-called likelihood-based model. Thus, this paper focuses on the evidential extension of binning and logistic regression calibrations based on this likelihood model. These evidential calibration procedures yields a MF $m^{\mathbb{Y}}(\cdot|s)$ (rather than a probability distribution), equivalently represented by the belief and plausibility functions $Bel^{\mathbb{Y}}(\cdot|s)$ and $Pl^{\mathbb{Y}}(\cdot|s)$.

3.2.1 Binning

For a given bin j, binning can be seen as a binomial experiment, where the number of examples n_j corresponds to the number of trials and the number of positive examples k_j represents the number of successes. Thus, it corresponds to the particular case of estimation considered in Sect. 2.2 and used for forecasting in Sect. 2.3. Considering that the given score s is in bin j, the likelihood-based contour function defined in Eq. (14) becomes

$$pl^{\Theta}_{\mathcal{X}}(\theta|s) = \frac{\theta^{k_j}(1-\theta)^{n_j-k_j}}{\hat{\theta}^{k_j}(1-\hat{\theta})^{n_j-k_j}}, \tag{38}$$

where $\hat{\theta} = \frac{k_j}{n_j}$ is the Maximum Likelihood Estimate (MLE) of θ. The belief and plausibility functions $Bel^{\mathbb{Y}}(\cdot|s)$ and $Pl^{\mathbb{Y}}(\cdot|s)$ are then simply obtained using Eqs. (27) and (28) with $x = k_j$ and $n = n_j$.

3.2.2 Logistic regression

Logistic-based calibration can also be extended in the evidential framework through the likelihood model. Xu et al. (2016) express uncertainty on the parameter $\sigma = (\sigma_0, \sigma_1)$ of the sigmoid function, by a consonant belief function Bel^{Σ}, whose contour function is defined by

$$pl^{\Sigma}_{\mathcal{X}}(\sigma) = \frac{L_{\mathcal{X}}(\sigma)}{L_{\mathcal{X}}(\hat{\sigma})}, \quad \forall \sigma \in \Sigma, \tag{39}$$

where $\hat{\sigma} = (\hat{\sigma}_0, \hat{\sigma}_1)$ is the MLE of σ and $L_{\mathcal{X}}$ is the likelihood function defined in Eq. (33). The corresponding plausibility function is defined as

$$Pl^{\Sigma}_{\mathcal{X}}(A) = \sup_{\sigma \in A} pl^{\Sigma}_{\mathcal{X}}(\sigma), \quad \forall A \subseteq \Sigma. \tag{40}$$

As seen in Sect. 2.3, the belief and plausibility functions on $\mathbb{Y}$ can be deduced from the contour function $pl^{\Theta}_{\mathcal{X}}$ defined on Θ. Xu et al. showed in (Xu et al. 2016) that this function $pl^{\Theta}_{\mathcal{X}}$ can be computed from $Pl^{\Sigma}_{\mathcal{X}}$. Indeed, as θ is defined by $\theta = h_s(\sigma)$, we get

$$pl^{\Theta}_{\mathcal{X}}(\theta|s) = \begin{cases} 0 & \text{if } \theta \in \{0, 1\}, \\ Pl^{\Sigma}_{\mathcal{X}}(h_s^{-1}(\theta)) & \text{otherwise,} \end{cases} \tag{41}$$

with

$$h_s^{-1}(\theta) = \{(\sigma_0, \sigma_1) \in \Sigma | h_s(\sigma) = \theta\}, \tag{42}$$

$$= \left\{(\sigma_0, \sigma_1) \in \Sigma | \frac{1}{1 + \exp(\sigma_0 + \sigma_1 s)} = \theta\right\}, \tag{43}$$

$$= \left\{(\sigma_0, \sigma_1) \in \Sigma | \sigma_0 = \ln(\theta^{-1} - 1) - \sigma_1 s\right\}. \tag{44}$$

Finally, Eqs. (41) and (44) yield the following function

$$pl^{\Theta}_{\mathcal{X}}(\theta|s) = \sup_{\sigma_1 \in \mathbb{R}} pl^{\Sigma}_{\mathcal{X}}(\ln(\theta^{-1} - 1) - \sigma_1 s, \sigma_1), \quad \forall \theta \in [0, 1]. \tag{45}$$

The value $pl^{\Theta}_{\mathcal{X}}(\theta|s)$ can be obtained by an iterative maximization algorithm, for all $\theta \in [0, 1]$. The belief and plausibility functions $Bel^{\mathbb{Y}}(\cdot|s)$ and $Pl^{\mathbb{Y}}(\cdot|s)$ can then be calculated using Eqs. (25) and (26).

4 An evidential joint calibration approach

In a context of multiple classifiers, one may independently calibrate the score given by each classifier after observing an object, and merge them using a predetermined rule of combination. In particular, this kind of process is followed by Xu et al. (2016), where scores provided by binary SVM classifiers are transformed into belief functions using evidential calibration and combined using Dempster's rule of combination. We refer hereafter to this latter approach as the disjoint method.

We propose in this paper to use the multivariable versions of the techniques underlying the calibrations and to apply it to the outputs of multiple classifiers, i.e., to perform a joint calibration of the scores provided by the binary SVM classifiers. More specifically, in order to better handle the uncertainties of the calibration process, we propose to perform the joint calibration in the evidential framework.

For a given object, we take as input the score vector $\mathbf{s} = (s_1, s_2, \ldots, s_J)$, with s_j the score returned by the jth classifier after observing the object. The required training set is defined by $\mathcal{X}' = \{(s_{11}, s_{21}, \ldots, s_{J1}, y_1), \ldots, (s_{1n}, s_{2n}, \ldots, s_{Jn}, y_n)\}$, where s_{ji} corresponds to the score

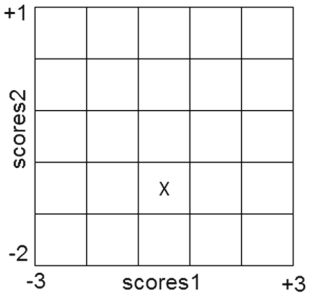


Fig. 2 Example of score space for joint binning, with $J = 2$ and $B_M = 5$

given by the jth classifier for the ith test sample, and y_i the true label of this sample.

We first expose in Sect. 4.1 the multivariable version of binning calibration, followed by the multivariable version of the calibration based on logistic regression in Sect. 4.2.

4.1 Joint binning

The idea consists in dividing the score space into multi-dimensional bins (cells) or more precisely into J-dimensional bins with J the number of classifiers. Let us illustrate the building of these cells with a $2D$ scenario, i.e., when only two classifiers are considered. If the first classifier has score values between -3 and 3 and the second classifier has score values between -2 and 1, the score space is $[-3, 3] \times [-2, 1]$. This score space can be divided in different ways. In particular, a number of bins per classifier can be chosen and the score space can be divided uniformly based on this number. An illustration of this naive scheme is given in Fig. 2, where the number of bins by classifier, denoted B_M, is chosen equal to 5.

Given a cell c, the number k_c of tuples $(s_{1i}, \ldots, s_{Ji}, y_i) \in \mathcal{X}'$ such that $y_i = 1$ and $(s_{1i}, s_{2i}, \ldots, s_{Ji})$ belongs to cell c, and the number n_c of tuples such that $(s_{1i}, \ldots, s_{Ji})$ belongs to cell c can be obtained. For a given input vector $\mathbf{s} = (s_1, s_2, \ldots, s_J)$ such that $\mathbf{s}$ belongs to the cell c, we have

$$P^{\mathbb{Y}}(y = 1|\mathbf{s}) = \frac{k_c}{n_c}. \tag{46}$$

For instance, let us consider that we have $\mathbf{s} = (0.5, -1)$, i.e., after observing a given example the first classifier returns the score 0.5 and the second -1. The probability associated with this object can thus be found by looking into the corresponding cell c, which is the one marked by a cross in Fig. 2.

This probabilistic joint approach of binning can be extended to the evidential framework. Similarly to the single classifier case, the label y of a given score vector $\mathbf{s}$ can be seen as a realization of a random variable with a Bernoulli distribution, and binning can be seen as a binomial experiment for each cell. If the score vector $\mathbf{s}$ is in cell c, the belief and plausibility functions associated with this score vector can be calculated using the following equations:

$$Bel^{\mathbb{Y}}(\{1\}|\mathbf{s}) = \begin{cases} 0, & \text{if } \hat{\theta} = 0, \\ \hat{\theta} - \frac{B(\hat{\theta}; k_c+1, n_c-k_c+1)}{\hat{\theta}^{k_c}(1-\hat{\theta})^{n_c-k_c}}, & \text{if } 0 < \hat{\theta} < 1, \\ \frac{n_c}{n_c+1}, & \text{if } \hat{\theta} = 1, \end{cases} \tag{47}$$

$$Pl^{\mathbb{Y}}(\{1\}|\mathbf{s}) = \begin{cases} \frac{1}{n_c+1}, & \text{if } \hat{\theta} = 0, \\ \hat{\theta} + \frac{\overline{B}(\hat{\theta}; k_c+1, n_c-k_c+1)}{\hat{\theta}^{k_c}(1-\hat{\theta})^{n_c-k_c}}, & \text{if } 0 < \hat{\theta} < 1, \\ 1, & \text{if } \hat{\theta} = 1, \end{cases} \tag{48}$$

with $\hat{\theta} = \frac{k_c}{n_c}$.

4.2 Joint logistic regression

The logistic regression, exposed in Sect. 3, is used to calibrate a score given by a single classifier. Yet, the logistic model works as well when more than one input is available: It is then called a multivariable (or multiple) logistic regression (Hosmer et al. 2013). It has been widely used in many applications, such as for instance in medicine field (Bagley et al. 2001). We propose to use this multiple version of logistic regression and apply it to the vector of scores returned by different classifiers for a given object, in order to calibrate this vector.

Given a vector of scores $\mathbf{s} = (s_1, s_2, \ldots, s_J)$, the probabilistic joint calibration based on multiple logistic regression is defined by

$$P^{\mathbb{Y}}(y = 1|\mathbf{s}) = \frac{1}{1 + \exp(\sigma_0 + \sigma_1 s_1 + \sigma_2 s_2 + \cdots + \sigma_J s_J)}, \tag{49}$$

where the parameter $\boldsymbol{\sigma} = (\sigma_0, \ldots, \sigma_J) \in \mathbb{R}^{J+1}$ is obtained by maximizing the likelihood function $L_{\mathcal{X}'}$ defined by

$$L_{\mathcal{X}'}(\sigma) = \prod_{i=1}^{n} p_i^{t_i}(1 - p_i)^{1-t_i}, \tag{50}$$

with

$$p_i = \frac{1}{1 + \exp(\sigma_0 + \sigma_1 s_{1i} + \cdots + \sigma_J s_{Ji})}, \tag{51}$$

and

$$t_i = \begin{cases} \frac{N_+ + 1}{N_+ + 2} & \text{if } y_i = 1, \\ \frac{1}{N_- + 2} & \text{if } y_i = 0, \end{cases} \tag{52}$$

where N_+ and N_- are, respectively, the number of positive and negative samples in the training set $\mathcal{X}'$. The log-likelihood can be used instead of the likelihood, and an unique solution is found for $\boldsymbol{\sigma}$.

We propose to extend this joint logistic-based calibration to the evidential framework by following the same likelihood-based reasoning as for the single classifier case. The knowledge about $\boldsymbol{\sigma} = (\sigma_0, \ldots, \sigma_J)$ can be represented by a consonant belief function whose contour function is defined by

$$pl^{\Sigma}_{\mathcal{X}'} = \frac{L_{\mathcal{X}'}(\sigma)}{L_{\mathcal{X}'}(\hat{\sigma})}, \quad \forall \sigma \in \Sigma. \tag{53}$$

Furthermore, $pl^{\Theta}_{\mathcal{X}'}$ can be computed from $Pl^{\Sigma}_{\mathcal{X}'}$:

$$pl^{\Theta}_{\mathcal{X}'}(\theta|\mathbf{s}) = \begin{cases} 0 & \text{if } \theta \in \{0, 1\}, \\ Pl^{\Sigma}_{\mathcal{X}'}(h_{\mathbf{s}}^{-1}(\theta)) & \text{otherwise,} \end{cases} \tag{54}$$

with

$$h_{\mathbf{s}}^{-1}(\theta) = \{(\sigma_0, \sigma_1, \ldots, \sigma_J) \in \Sigma | h_s(\sigma) = \theta\}, \tag{55}$$

$$= \left\{(\sigma_0, \ldots, \sigma_J) \in \Sigma | \frac{1}{1 + \exp(\sigma_0 + \sigma_1 s_1 + \cdots + \sigma_J s_J)} = \theta\right\}, \tag{56}$$

$$= \left\{(\sigma_0, \ldots, \sigma_J) \in \Sigma | \sigma_0 = \ln(\theta^{-1} - 1) - \sigma_1 s_1 - \cdots - \sigma_J s_J\right\}. \tag{57}$$

Thus, the contour function $pl^{\Theta}_{\mathcal{X}'}(\theta|\mathbf{s})$ is defined by

$$pl^{\Theta}_{\mathcal{X}'}(\theta|\mathbf{s}) = \sup_{\sigma_1, \ldots, \sigma_J \in \mathbb{R}} pl^{\Sigma}_{\mathcal{X}'}(\ln(\theta^{-1} - 1) - \sigma_1 s_1 - \cdots - \sigma_J s_J, \sigma_1, \ldots, \sigma_J), \tag{58}$$

for all $\theta \in [0, 1]$. The vector of parameters $(\sigma_1, \sigma_2, \ldots, \sigma_J)$ which maximizes $pl^{\Sigma}_{\mathcal{X}'}$ can be approximated using an iterative maximization algorithm (the computational complexity of such algorithm is $O(nJ)$ per iteration). Then, the belief and plausibility functions $Bel^{\mathbb{Y}}(\cdot|\mathbf{s})$ and $Pl^{\mathbb{Y}}(\cdot|\mathbf{s})$ can be obtained through Eqs. (25) and (26).

5 Experiments

In this section, the performance of our proposed evidential joint calibration approach is compared to those of other approaches using different datasets, which are presented in Sect. 5.1. In Sect. 5.2, our approach is compared to the disjoint approach of Xu et al. Both binning and logistic regression calibrations are studied. Then, in Sect. 5.3, these latter two calibrations are compared to a conceptually similar approach, that is a trainable combiner based on an evidential classifier, i.e., a classifier returning a mass function after observing an object. Finally, we focus on the calibration based on multiple logistic regression and we compare the probabilistic and evidential versions of this joint calibration in Sect. 5.4.

Table 1 Number of instance vectors and number of features by vector for different datasets from UCI

Dataset	# Instance vectors	# Features
Australian	690	14
Diabetes	768	8
Heart	270	13
Ionosphere	351	34
Sonar	208	60

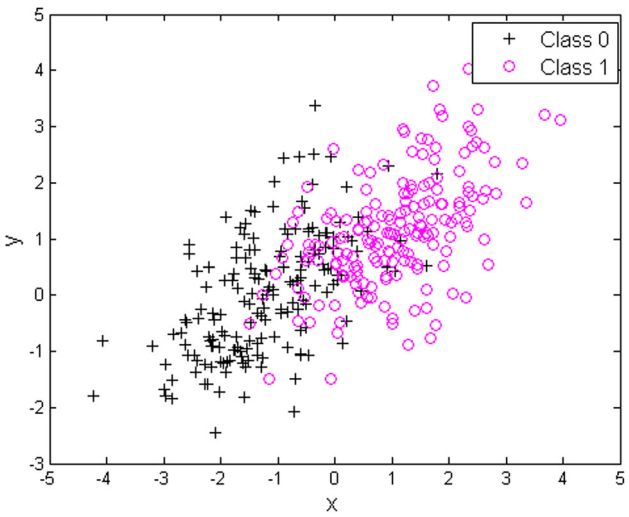


Fig. 3 Illustration of 360 instance vectors of the simulated dataset

5.1 Datasets

The experiments are conducted on five binary classification problems provided by UCI repository (Bache and Lichman 2013). They are all of different sizes, and their sample vectors have various number of features. This is presented in Table 1.

We also simulated a dataset composed of 360 randomly generated instance vectors from a multivariate normal distribution, with means $\mu_0 = (-1, 0)$ in class 0 and $\mu_1 = (1, 1)$ in class 1, and with a covariance matrix equals to $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ for both classes. Each instance vector has two features. An illustration of these data on the feature space is represented in Fig. 3, where x and y represent, respectively, the first and second features of each instance vector.

5.2 Comparison with Xu et al.'s approach (Xu et al. 2016)

The following experiment follows the same protocol as the first experiment detailed in (Xu et al. 2016). For each dataset,

Table 2 Number of examples used for training and testing

Dataset	# Train 1	# Train 2	# Train 3	# Test
Australian	30	70	10–60–190	400
Diabetes	30	70	10–50–200	468
Heart	20	40	10–50–140	70
Ionosphere	20	40	10–80–190	101
Sonar	20	40	10–40–90	58
Simulated	20	40	10–50–200	100

three SVM classifiers are trained on non-overlapping subsets, using the LIBSVM library (Chang and Lin 2011). The numbers of examples used for training and testing for each dataset are described in Table 2. For the first two classifiers, the number of training examples is fixed, while different training set sizes are considered for the third one. The training set of each classifier is partitioned into two equal-sized subsets. One of these subsets is for training the classifier, and in Xu et al.'s approach the second subset is for training the calibration of the classifier. In the proposed approach, the joint calibration is trained using the set composed of the concatenation of each second subset of each classifier.

For each sample belonging to the test set, the three classifiers return a score. In Xu et al.'s approach, each of these scores is calibrated using the trained calibration of its corresponding classifier, and the three obtained mass functions are merged into a final mass function using Dempster's rule. In our proposed approach, the scores are grouped into a score vector and this vector is calibrated using a joint calibration, which directly returns a final mass function. In both cases, the decision corresponds to the singleton with the highest belief, since we use {0, 1} costs without the possibility to reject, in which case upper and lower expected costs lead to the same decision. The error rate is calculated on the test set and corresponds to the number of samples misclassified over the number of tested samples. The whole process is repeated for 100 rounds of random partitioning, and thus the final error rate corresponds to the average of 100 calculated error rates.

For the binning calibration, we may remark that there are in total a number of $B_U \times J$ bins in the disjoint case against $(B_M)^J$ bins for the joint binning. In order to fairly compare our approach to the disjoint one, the number of bins for each classifier is chosen such that each method has the same total number of bins. In particular, as $J = 3$, we chose, respectively, $B_U = 9$ and $B_M = 3$ for disjoint and joint approaches.

Figure 4 shows the results of the experiments for binning and logistic-based approaches, in the evidential framework, and for disjoint and joint cases. Results of the probabilistic version of joint calibrations are also given. As it can be noticed, the approaches based on the logistic regression are always better than those based on binning, as their obtained error rates are lower. For binning approaches, the joint case is not always better than the disjoint case, but it might come from the chosen value for B_M; with a higher value, the results might be better. For logistic regression, the evidential joint approach always presents better results than the evidential disjoint approach. It can also be noticed that the probabilistic and evidential joint versions nearly give the same results in this experiment. Comparison between probabilistic and evidential versions of calibration based on multiple logistic regression will be performed in Sect. 5.4.

5.3 Comparison with evidential trainable combiner approach

In the previous experiment, we compared our approach to its probabilistic version and to the so-called disjoint method, which belongs to the non-trainable combiner category. In this section, we perform the same experiment but with the aim of comparing our results to those of approaches of the same category, i.e., to evidential trainable combiners. Indeed, there exist other approaches similar to ours to be compared to, and in particular some methods which can take a score vector as input and return a belief function on the class of a given observed object.

The first evidential trainable combiner that we consider in this experiment relies on the evidential classifier described in (Denœux and Smets 2006) and based on the Generalized Bayesian Theorem (GBT) (Smets 1993).

Let us consider a classification problem with $\Omega = \{w_k\}_{k=1}^K$ the finite set of classes. After observing the feature vector $\mathbf{x}$ of an object, the aim is to obtain a belief function about the class label of this object, based on a training set $\mathcal{L} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ where $\mathbf{x}_i$ represents the feature vector of the ith object, whose true label is y_i. The application of the GBT gives the following MF on Ω about the class of $\mathbf{x}$ (Denœux and Smets 2006):

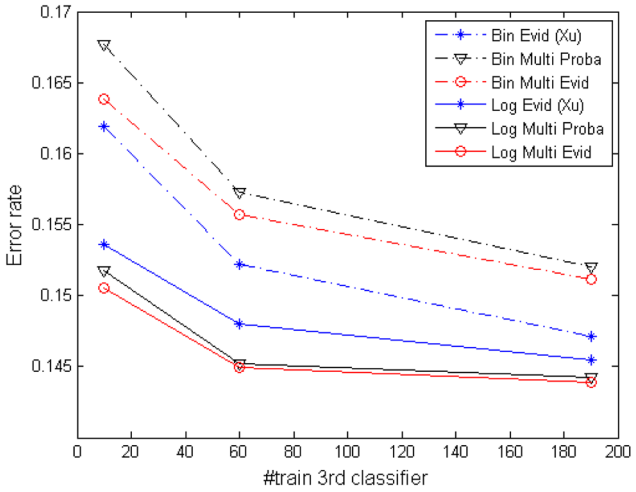
$$m^{\Omega}(A|\mathbf{x}) = \prod_{w_k \in A} Pl[w_k](\mathbf{x}) \prod_{w_k \in \overline{A}} (1 - Pl[w_k](\mathbf{x})), \qquad (59)$$

$\forall A \subseteq \Omega$, where $\overline{A}$ denotes the complement of A, and $Pl[w_k](\mathbf{x})$ represents the plausibility of observing $\mathbf{x}$ under the hypothesis that the true class is ω_k. In particular, Denœux and Smets (2006) have considered a special case, where

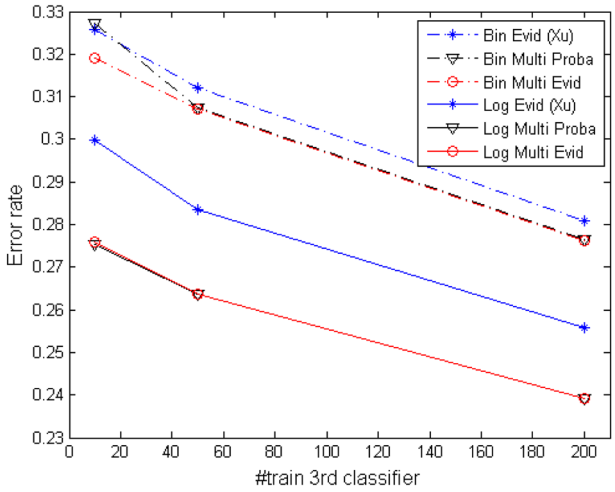
$$Pl[w_k](\mathbf{x}) = \frac{N(\mathbf{x}, k)}{N(k)}, \qquad (60)$$

with $N(\mathbf{x}, k)$ the number of samples in $\mathcal{L}$ from class w_k contained in a ball S_r of radius r and centered on $\mathbf{x}$, and $N(k)$ the total number of samples from class w_k in $\mathcal{L}$.

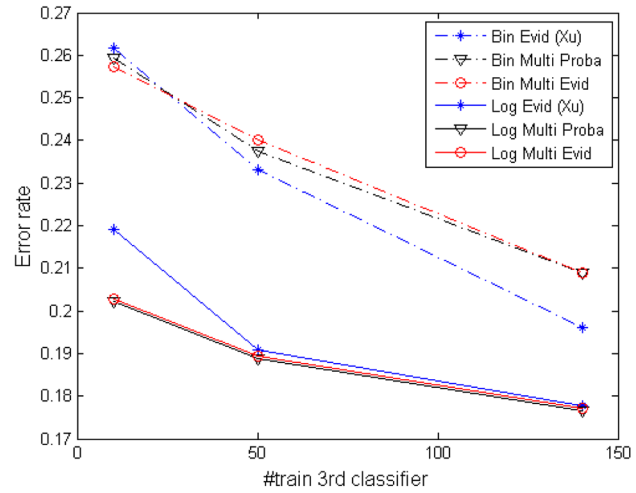
We note that it may happen that $m^{\Omega}(\emptyset|\mathbf{x}) > 0$, and in that case the MF $m^{\Omega}(\cdot|\mathbf{x})$ can be transformed into a normalized MF $M^{\Omega}(\cdot|\mathbf{x})$ using the operation defined by



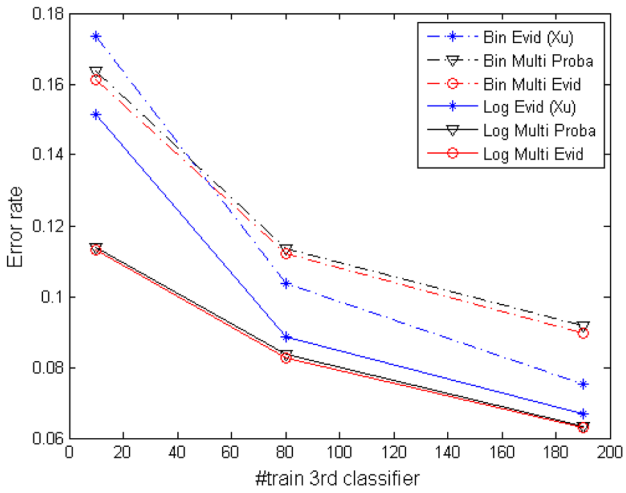
(a)



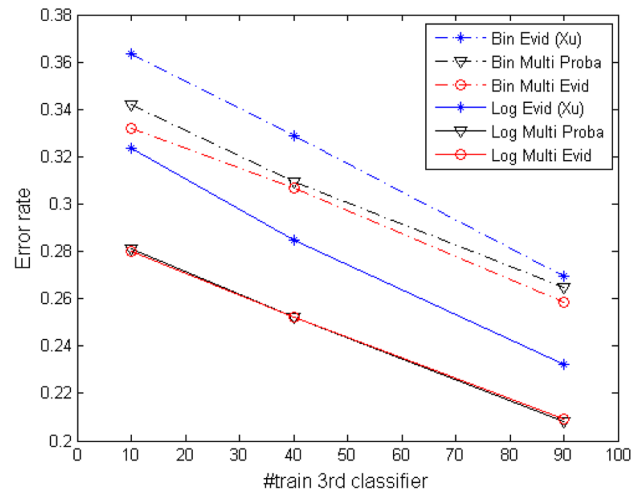
(b)



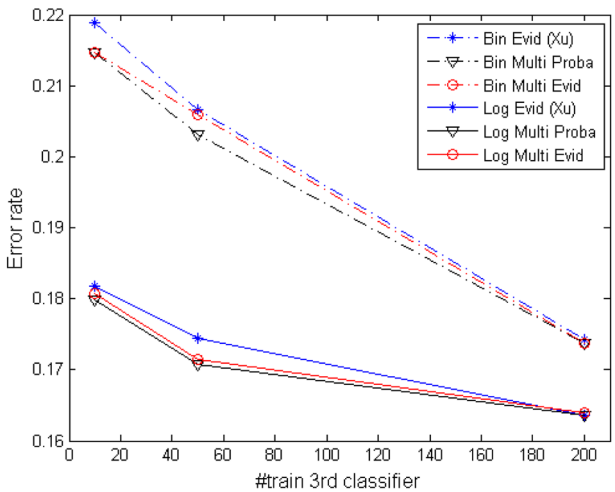
(c)



(d)



(e)



(f)

Fig. 4 Average error rates using binning and logistic regression, with joint (referred to as "multi" in the figures) and disjoint approaches and with both probabilistic and evidential frameworks. The X-axis corresponds to the number of training examples used to train the third classifier. **a** Australian. **b** Diabetes. **c** Heart. **d** Ionosphere. **e** Sonar. **f** Simulated data

$$M^{\Omega}(A|\mathbf{x}) = \frac{m^{\Omega}(A|\mathbf{x})}{1 - m^{\Omega}(\emptyset|\mathbf{x})}, \qquad \forall A \subseteq \Omega, A \neq \emptyset, \tag{61}$$

and $M^{\Omega}(\emptyset|\mathbf{x}) = 0$.

We now apply this classifier to our binary problem, by taking the same inputs as for our approach. In particular, after observing a given object, the feature vector is now the vector of scores $\mathbf{s} = (s_1, \ldots, s_J)$ obtained by J classifiers, and the training set $\mathcal{L}$ is now $\mathcal{X}'$. Using the definition of the MF given in Eq. (59) and the considered particular case of Eq. (60), we obtain the MF $m^{\mathbb{Y}}(\cdot|\mathbf{s})$ defined by

$$m^{\mathbb{Y}}(\{0\}|\mathbf{s}) = \frac{N(\mathbf{s},0)}{N(0)} \times \left(1 - \frac{N(\mathbf{s},1)}{N(1)}\right), \tag{62}$$

$$m^{\mathbb{Y}}(\{1\}|\mathbf{s}) = \frac{N(\mathbf{s},1)}{N(1)} \times \left(1 - \frac{N(\mathbf{s},0)}{N(0)}\right), \tag{63}$$

$$m^{\mathbb{Y}}(\{0,1\}|\mathbf{s}) = \frac{N(\mathbf{s},1)}{N(1)} \times \frac{N(\mathbf{s},0)}{N(0)}, \tag{64}$$

and

$$m^{\mathbb{Y}}(\emptyset|\mathbf{s}) = \left(1 - \frac{N(\mathbf{s},0)}{N(0)}\right) \times \left(1 - \frac{N(\mathbf{s},1)}{N(1)}\right), \tag{65}$$

with $N(\mathbf{s}, k)$ being the number of samples in $\mathcal{X}'$ from class k (equal to 0 or 1), contained in a ball S_r of radius r and centered on $\mathbf{s}$. This MF is then normalized similarly as $m^{\Omega}(\cdot|\mathbf{x})$ is normalized using Eq. (61).

We may notice that using a ball S_r to build the MFs has some similarities with our multivariable version of binning. Let us illustrate this statement with a simple example, using the dataset *Diabetes* and with $J = 2$. Figure 5 shows the scores returned by two trained classifiers for each sample of a given calibration training set. The X-axis corresponds to the scores given by the first classifier, and Y-axis by the second one. A test sample is illustrated by a blue asterisk and corresponds to $\mathbf{s} = (s_1, s_2)$ the values of the scores returned by the two classifiers. The continuous green lines correspond to the bounds of the joint binning, with $B_M = 3$, and the red circle represents the ball S_r of the GBT-based classifier, with $r = 1$ and centered on $\mathbf{s}$. To build the MF $m^{\mathbb{Y}}(\cdot|\mathbf{s})$, the joint binning uses the training samples belonging to the bin containing $\mathbf{s}$, while the GBT-based classifier uses the ones contained by the ball S_r.

The second evidential trainable combiner that we consider is the evidential κ Nearest Neighbor (κNN) classification rule (Denœux 1995), whose parameters are optimized using the procedure described in (Zouhal and Denœux 1998). Let us consider the same classification problem as for GBT-based approach, i.e., obtaining a belief function about the class label of an observed object with feature vector $\mathbf{x}$, based on a training set $\mathcal{L} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ where $\mathbf{x}_i$ represents the feature vector of the ith object, whose true label is y_i. If $\mathbf{x}$

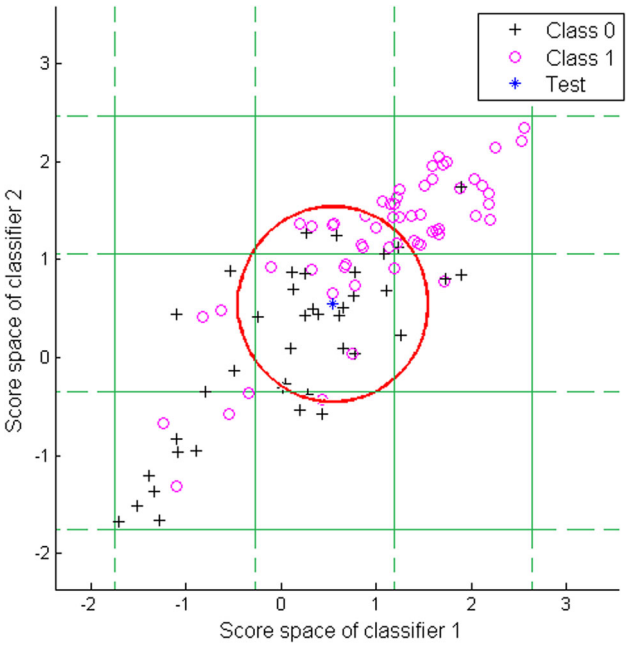


Fig. 5 Illustration of the multi-dimensional bins and the ball S_r, using *Diabetes* data

is close to $\mathbf{x}_i$ according to some distance measure d, then it is reasonable to believe that both vectors belong to the same class. More generally, the closer $\mathbf{x}_i$ is to $\mathbf{x}$, the more reasons to believe that the class of $\mathbf{x}$ is the same as the one of $\mathbf{x}_i$. This piece of information brought by the ith object about the class of the observed object may be represented by a MF m_i^{Ω} defined by (Denœux 1995):

$$m_i^{\Omega}(\{\omega_k\}) = \alpha\phi_k(d_i), \tag{66}$$

$$m_i^{\Omega}(\Omega) = 1 - \alpha\phi_k(d_i), \tag{67}$$

$$m_i^{\Omega}(A) = 0, \quad \forall A \in 2^{\Omega}\backslash\{\Omega, \{\omega_k\}\}, \tag{68}$$

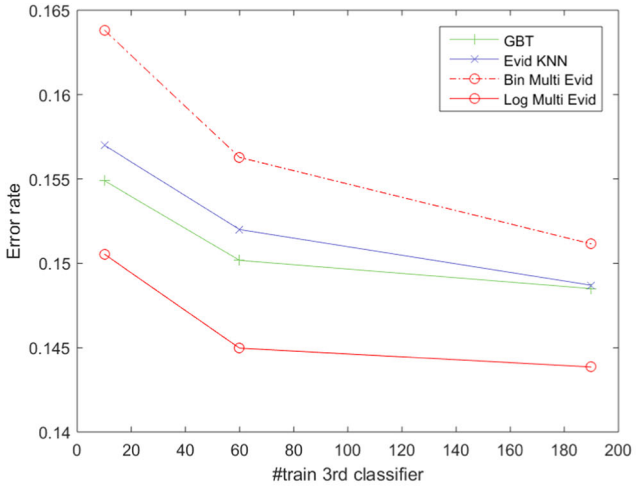
where ω_k is the class y_i of the ith object, $d_i = d(\mathbf{x}, \mathbf{x_i})$ is the distance between the feature vector of the observed object and the feature vector of the ith object, α is a parameter such that $0 < \alpha < 1$ and ϕ_k is a decreasing function verifying $\phi_k(0) = 1$ and $\lim_{d\to\infty}\phi_k(d) = 0$. A common choice for ϕ_k is given by (Denœux 1995):

$$\phi_k(d) = \exp(-\alpha_k d^2), \tag{69}$$

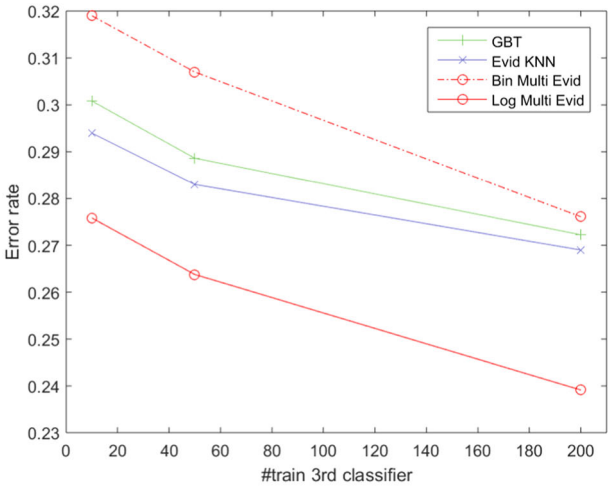
where $\alpha_k > 0$ is a parameter associated with class ω_k.

Thus, a MF may then be obtained for each sample of the training set $\mathcal{L}$. Denœux (Denœux 1995) proposed to pool by Dempster's rule the evidence of the κ nearest neighbors, $1 \leq \kappa \leq n$, of the observed object in order to obtain a MF $m^{\Omega}(\cdot|\mathbf{x})$ about its class. Let $\kappa_{\mathbf{x}}$ denote the set of the κ nearest objects of $\mathbf{x}$ in $\mathcal{L}$. The MF $m^{\Omega}(\cdot|\mathbf{x})$ is then defined as

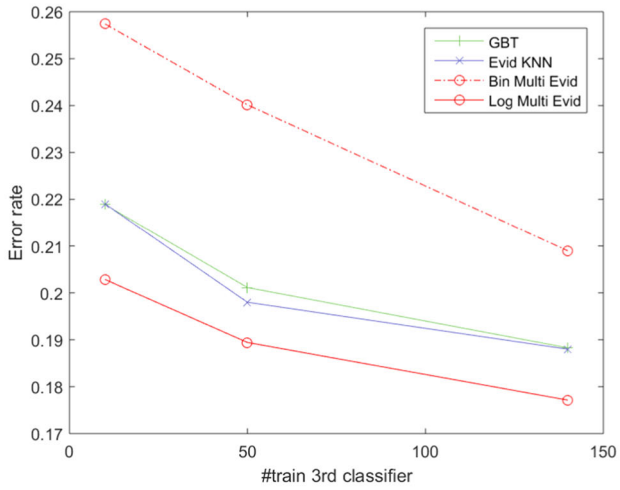
$$m^{\Omega}(A|\mathbf{x}) = \left(\oplus_{\mathbf{x_i}\in\kappa_{\mathbf{x}}} m_i^{\Omega}\right)(A), \quad \forall A \subseteq \Omega. \tag{70}$$



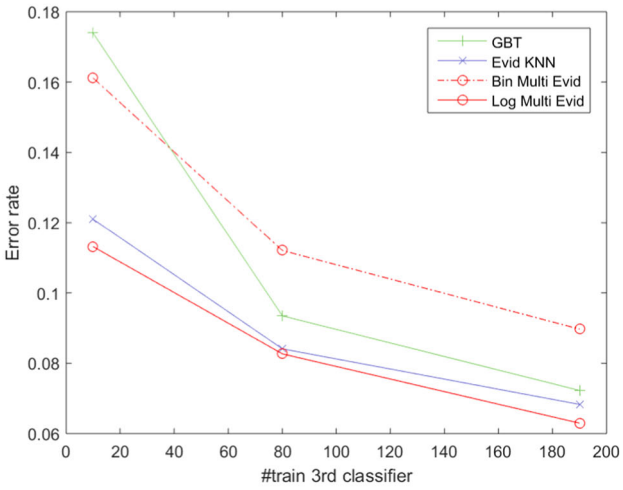
(a)



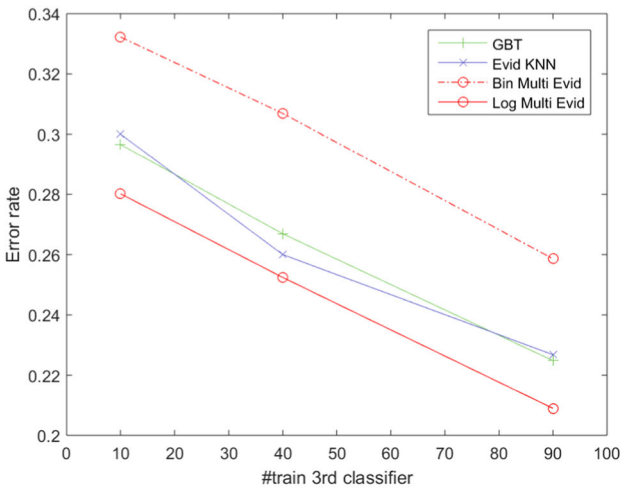
(b)



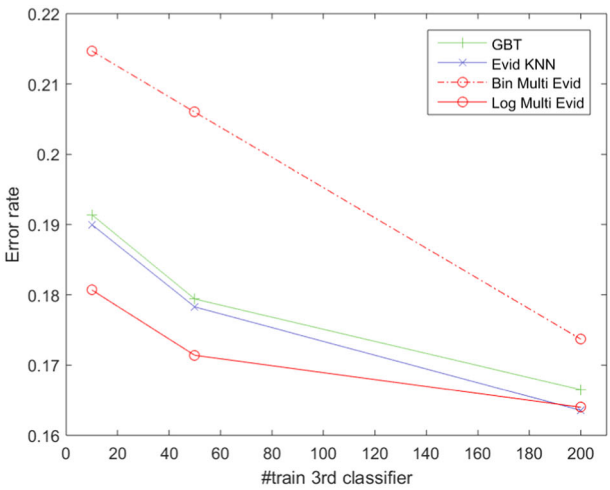
(c)



(d)



(e)



(f)

Fig. 6 Average error rates using the GBT-based and κNN-based approaches, and using the binning and logistic regression with evidential joint approaches. The X-axis corresponds to the number of training examples used to train the third classifier. **a** Australian. **b** Diabetes. **c** Heart. **d** Ionosphere. **e** Sonar. **f** Simulated data.

When applying this classifier to our binary problem, the feature vector $\mathbf{x}$ is now the vector of scores $\mathbf{s} = (s_1, \ldots, s_J)$ obtained by J classifiers, and the training set $\mathcal{L}$ is now $\mathcal{X}'$.

We performed the experiment with $r = 1$ for the GBT-based approach and $\kappa = 15$ for the evidential κNN approach[1] as some preliminary tests showed that the best results were obtained with these values.

Figure 6 shows the error rates for the GBT and κNN-based approaches, compared to those obtained with our evidential multivariable versions of binning and logistic regression. As it can be noticed, the results obtained with the GBT and the κNN-based classifiers are better than those obtained with the binning approach. It can be explained by the fact that in the binning approach the bounds of the multi-dimensional bins are fixed, and any test sample belonging to the same multi-dimensional bin has the same associated MF, no matter where the sample is positioned in the bin. By contrast, for the GBT classifier, the ball is centered on the considered test sample, so the neighborhood of the test sample is taken into account in a better way. A similar explanation can be provided for the κNN classifier. Furthermore, with other values of r and of κ, or with other size and number of our multi-dimensional bins, the obtained results may vary significantly, as these approaches highly rely on these parameters. Finally, we can see that the evidential joint calibration using logistic regression is always better than the GBT and κNN-based approaches in our experiments.

5.4 Comparison between evidential and probabilistic joint calibrations based on logistic regression

As seen in Sects. 5.2 and 5.3, the evidential joint logistic-based calibration always presents the best results. Yet, we also noted (in Sect. 5.2) that the performance of the probabilistic version of this calibration was nearly the same. Thus, in this section, probabilistic and evidential versions of the calibration based on the multiple logistic regression are further compared. To do that, we introduce the possibility of a third decision for the system given a test sample, by allowing a reject option. Hence, for a given test sample, three possible decisions can be rendered: 0, 1, or R. This option R expresses doubt and is used for some examples that are hard to classify. In addition, as recalled in Sect. 2.1, there are different decision-making criteria in the evidential framework and thus the evidential approach has two possible strategies of decision, either pessimistic or optimistic.

Using the simulated dataset previously defined, 290 training examples were generated: three SVM classifiers were trained with three non-overlapping subsets of 30 training examples of this set, and the joint calibration using logistic regression was trained with the remaining 200 examples of this set. Then, the same experiment was performed, but the joint logistic-based calibration was trained with 15 examples instead of 200. The decision frontiers for both the pessimistic and optimistic strategies and for both cases are illustrated in Fig. 7 for $R_{rej} = 0.15$.

As it can be seen, the evidential joint calibration based on the optimistic strategy tends to reject less the test samples than the two others. It is the exact opposite for the evidential joint calibration based on the pessimistic strategy, which decides to reject in more cases. The probabilistic approach

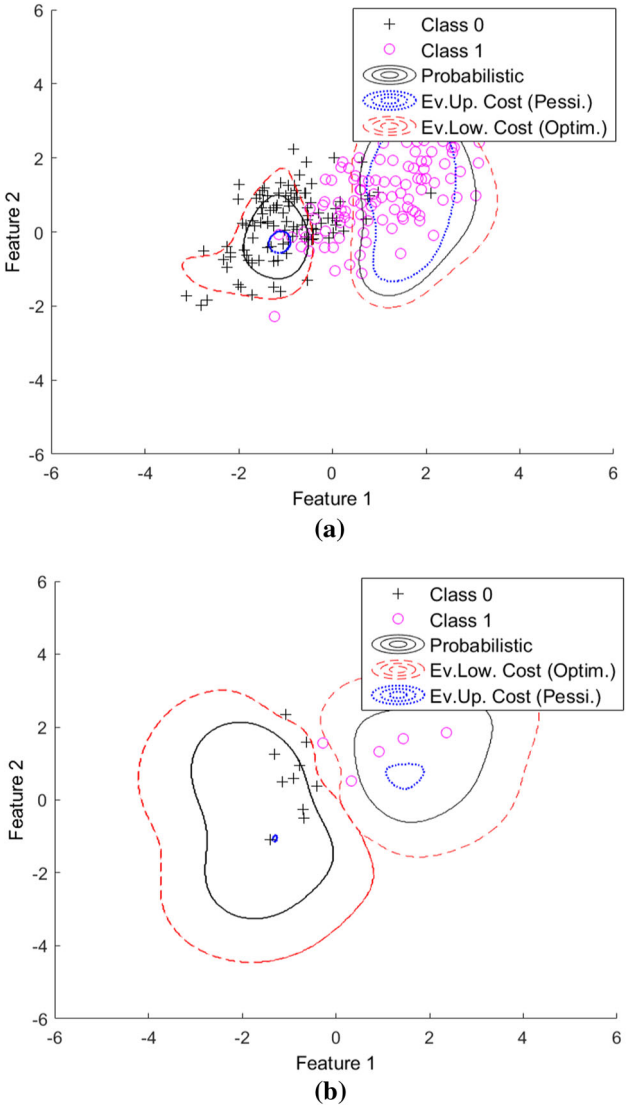


Fig. 7 Decision frontiers in feature space of the probabilistic and evidential joint calibrations based on logistic regression trained with 200 (7a) and 15 training examples (7b), and with $R_{rej} = 0.15$. **a** Joint logistic-based calibration trained with 200 training samples. **b** Joint logistic-based calibration trained with 15 training samples

[1] We used the software for the evidential κNN classifier with parameter optimization available at: https://www.hds.utc.fr/~tdenoeux/dokuwiki/en/software/k-nn

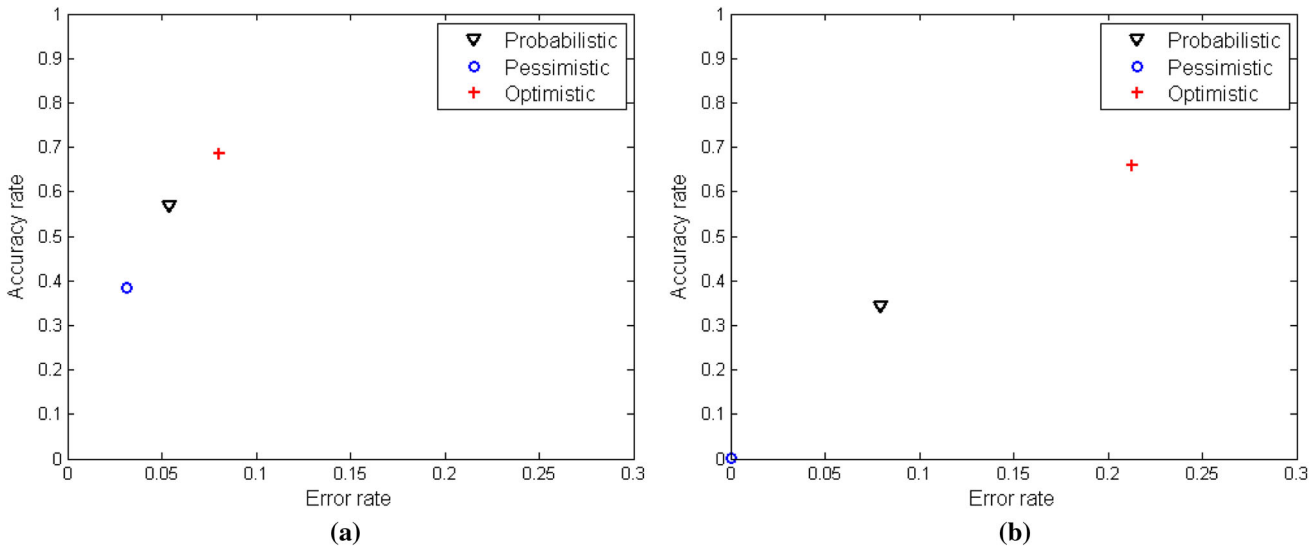


Fig. 8 Obtained error rates for $R_{rej} = 0.15$ and with 200 (8a) and 15 (8b) training examples. **a** 200 training examples. **b** 15 training examples

is between these two. Furthermore, the frontiers associated with the pessimistic and optimistic strategies are a lot more distant from each other in Fig. (7b) than in a), i.e., when there are less examples to train the joint calibration and thus more uncertainties. Probabilistic approach is only represented by one frontier, so the impact of the uncertainties is not visible. Thus, the evidential approach better reflects the uncertainties than the probabilistic one.

Let us illustrate this point further. The three SVM classifiers were still trained with three non-overlapping subsets of 30 training samples, and the calibration with 200 then 15 samples. We calculated the error rate and accuracy rate for 100 test samples and with $R_{rej} = 0.15$. Accuracy rate represents the number of correctly classified objects over the number of classified objects, i.e., not over the total number of test examples as some of them are rejected. The whole process was repeated for 100 rounds of random partitioning.

As it can be seen in Fig. 8, if there are a lot of examples to train the joint calibration, the obtained error rates are almost equal. Yet, when less training examples are available, the two points obtained for the evidential approach are more distant from each other. This interval reflects the uncertainties, as when it is larger the uncertainties are more important. This information cannot be obtained with the probabilistic calibration, as it is represented by only one point. Thus, the joint calibration based on evidence theory better reflects the uncertainties.

Finally, we performed a similar experiment with R_{rej} varying from 0 to 1, on five datasets (*Australian*, *Diabetes*, *Heart*, *Ionosphere*, *Sonar*) of UCI repository (Bache and Lichman 2013) and on the simulated dataset. The only difference with the previous experiment is that the multivariable logistic regression was trained with 45 then 15 samples. Due to the size of *Sonar*, it was tested on 50 sample tests instead of 100 for the other datasets. The whole process was carried out for 100 rounds of random partitioning, and Figs. 9 and 10 show the obtained results.

As it can be noticed, for a given error rate, the results obtained with the pessimistic strategy has a higher (or equal) accuracy rate than the probabilistic calibration when few training examples are available (right column). Let us underline that for a fixed error rate, the accuracy rates obtained with the probabilistic calibration and the pessimistic strategy are obtained for different values of R_{rej} (as seen in the previous experiment, the results of which are given in Fig. 8, a given value of R_{rej} leads in general to different error rates). Furthermore, when the number of training examples is more important (left column of Figs. 9 and 10), the obtained results become similar for the probabilistic and evidential approaches, as should be.

6 Conclusion

In this paper, an evidential joint calibration approach was proposed in order to handle the scores returned by multiple SVM classifiers. This approach belongs to the category of trainable combiners as it takes a score vector as input and does not need a predetermined rule of combination. We used evidence theory to prevent the over-fitting problem and to handle better the uncertainties associated with calibration techniques. Our approach was compared to Xu et al.'s disjoint approach, which independently calibrates the scores of SVM classifiers using the evidence theory and combines the obtained mass functions using Dempster's rule of combination. We compared also our proposed method to two approaches belonging to the trainable combiner category and based on an evidential classifier. In both cases, the obtained results for our evidential

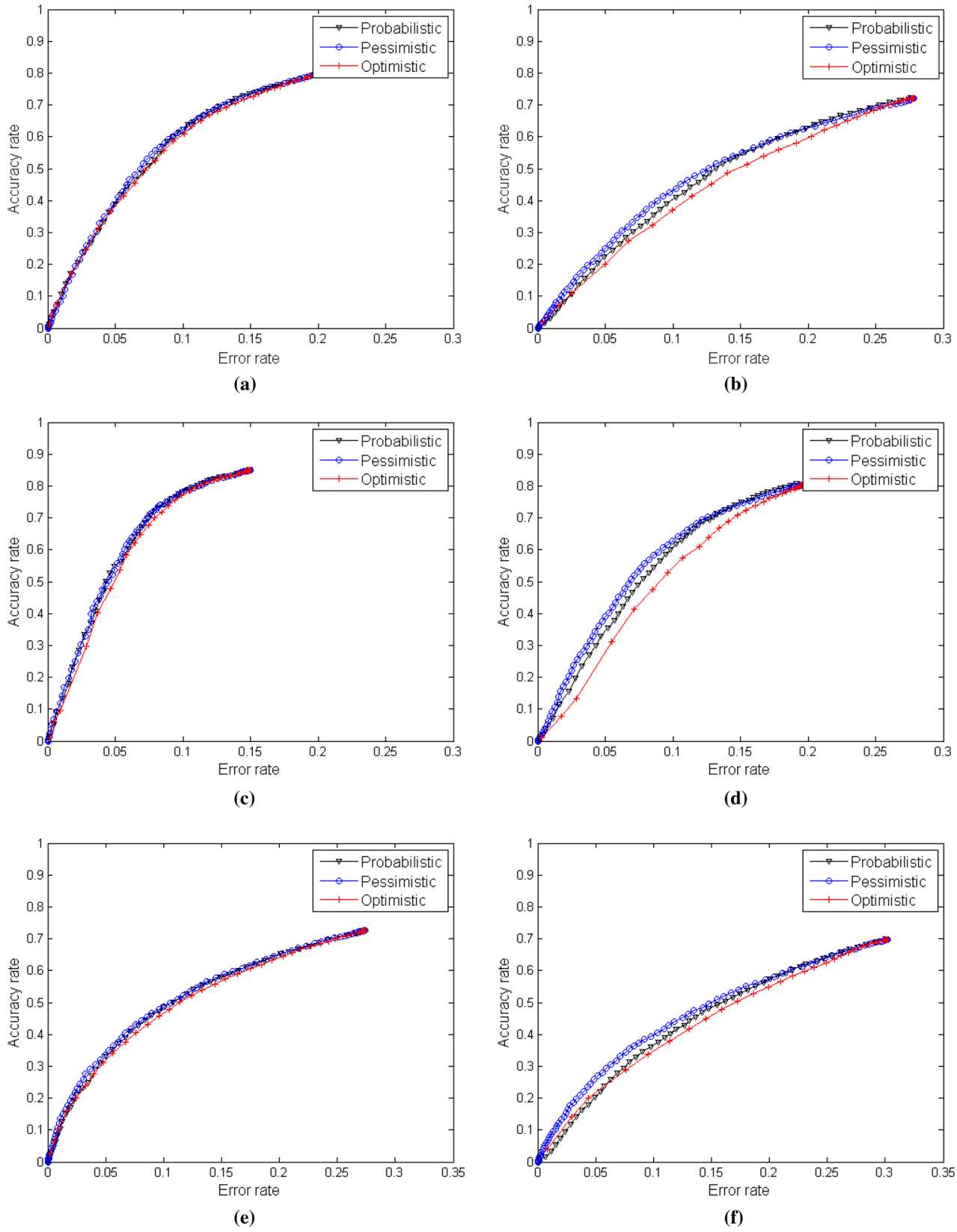


Fig. 9 Obtained error rates with 45 training samples (left) and 15 training samples (right) for the simulated dataset, *Australian* and *Diabetes*. **a** Simulated data—45 training samples. **b** Simulated data—15 training samples. **c** Australian—45 training samples. **d** Australian—15 training samples. **e** Diabetes—45 training samples. **f** Diabetes—15 training samples

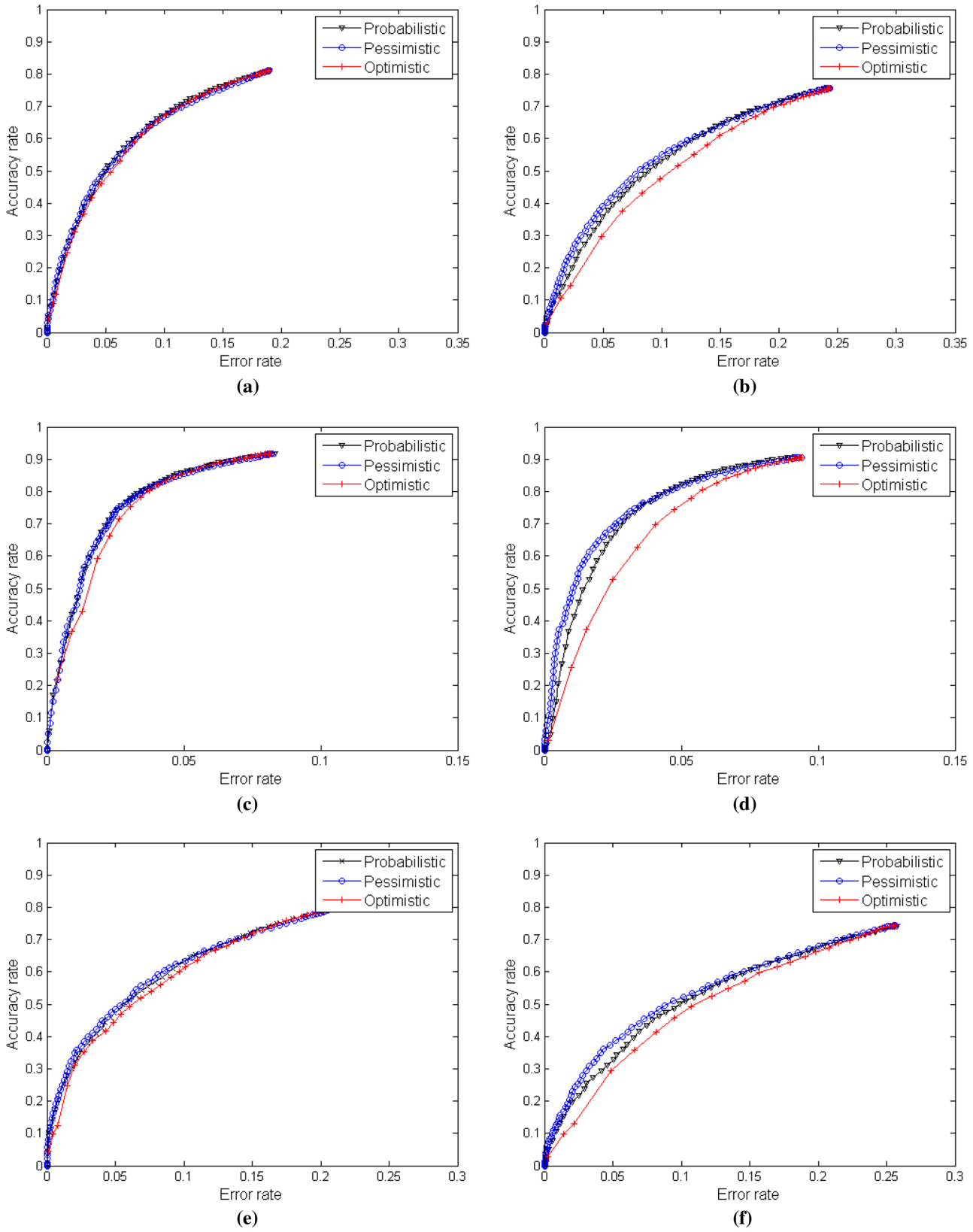


Fig. 10 Obtained error rates with 45 training samples (left) and 15 training samples (right) for *Heart*, *Ionosphere* and *Sonar*. **a** Heart—45 training samples. **b** Heart—15 training samples. **c** Ionosphere—45 training samples. **d** Ionosphere—15 training samples. **e** Sonar–45 training samples. **f** Sonar—15 training samples

joint calibration based on logistic regression either are better or are comparable to those of the other approaches. Furthermore, by introducing the possibility to reject a test sample, we showed the advantages of the evidential multivariable logistic-based calibration over the probabilistic version: It models more precisely the uncertainties, and it exhibits better performances.

The approach presented in this paper was applied to the calibration of binary SVM classifiers, but they may also be applied to any other binary classifiers returning scores. As a matter of fact, future works include applying the evidential multivariable calibration to the face blurring application described in (Minary et al. 2016), which involves four different binary classifiers, and which was solved in (Minary et al. 2016) using the disjoint approach. The extension of the proposed evidential joint calibration to the multi-class problem may also be tackled in future works, following Xu et al. (2015) which addressed this extension in the single classifier case.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Bache K, Lichman M (2013) UCI machine learning repository. University of California, School of Information and Computer Sciences. http://archive.ics.uci.edu/ml

Bagley SC, White H, Golomb BA (2001) Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. J Clin Epidemiol 54(10):979–985

Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2:27:1–27:27. http://www.csie.ntu.edu.tw/~cjlin/libsvm

Dempster AP (1966) New methods for reasoning towards posterior distributions based on sample data. Ann Math Stat 37(2):355–374

Dempster AP (1968) Upper and lower probabilities generated by a random closed interval. Ann Math Stat 39(3):957–966

Denœux T (1995) A k-nearest neighbor classification rule based on dempster-shafer theory. IEEE Trans Syst Man Cybern 25(5):804–813

Denœux T (1997) Analysis of evidence-theoretic decision rules for pattern classification. Pattern Recognit 30(7):1095–1107

Denœux T (2014) Likelihood-based belief function: justification and some extensions to low-quality data. Int J Approx Reason 55(7):1535–1547

Denœux T, Smets P (2006) Classification using belief functions: relationship between case-based and model-based approaches. IEEE Trans Syst Man Cybern B 36(6):1395–1406

Duin RPW (2002) The combining classifier: to train or not to train? In: Proceedings of the 16th International Conference on Pattern Recognition, Quebec City, Quebec, Canada, August, 2002, IEEE, vol 2, pp 765–770

Hosmer DW, Lemeshow S, Sturdivant RX (2013) Applied logistic regression, vol 398. Wiley, Hoboken

Kanjanatarakul O, Sriboonchitta S, Denœux T (2014) Forecasting using belief functions: an application to marketing econometrics. Int J Approx Reason 55(5):1113–1128

Kanjanatarakul O, Denœux T, Sriboonchitta S (2016) Prediction of future observations using belief functions: a likelihood-based approach. Int J Approx Reason 72:71–94

Kuncheva LI (2004) Combining pattern classifiers: methods and algorithms. Wiley, Hoboken

Minary P, Pichon F, Mercier D, Lefevre E, Droit B (2016) An evidential pixel-based face blurring approach. In: Vejnarov J, Kratochvil V (eds) Proceedings of the 4th International Conference on Belief Functions, Prague, Czech Republic, September 21–23, Springer, Lecture Notes in Computer Science, vol 9861, pp 222–230

Minary P, Pichon F, Mercier D, Lefevre E, Droit B (2017) Evidential joint calibration of binary svm classifiers using logistic regression. In: Proceedings of the 11th international conference on scalable uncertainty management, Granada, Spain, October 4–6, 2017, Lecture Notes in Artificial Intelligence, Springer, p 7

Minka TP (2003) Algorithms for maximum-likelihood logistic regression. Technical Report 758, Carnegie Mellon University

Nguyen HT (2006) An Introduction to Random Sets. Chapman and Hall/CRC Press, Boca Raton

Platt JC (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv Large Margin Classif 10(3):61–74

Shafer G (1976) A mathematical theory of evidence, vol 1. Princeton University Press, Princeton

Smets P (1993) Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. Int J Approx Reason 9(1):1–35

Smets P, Kennes R (1994) The transferable belief model. Artif Intell 66:191–243

Tulyakov S, Jaeger S, Govindaraju V, Doermann D (2008) Review of classifier combination methods. In: Marinai S, Fujisawa H (eds) Machine learning in document analysis and recognition. Springer, Berlin, pp 361–386

Xu P, Davoine F, Denœux T (2015) Evidential multinomial logistic regression for multiclass classifier calibration. In: Proceedings of the 18th international conference on information fusion, Washington, DC, USA, July 6–9, 2015, IEEE, pp 1106–1112

Xu P, Davoine F, Zha H, Denœux T (2016) Evidential calibration of binary SVM classifiers. Int J Approx Reason 72:55–70

Zadrozny B, Elkan C (2001) Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers. In: Proceedings of the 18th international conference on machine learning, Morgan Kaufmann, pp 609–616

Zadrozny B, Elkan C (2002) Transforming classifier scores into accurate multiclass probability estimates. In: Proceedings of the 8th international conference on knowledge discovery and data mining, New York, NY, USA, 2002, ACM, pp 694–699

Zhong W, Kwok JT (2013) Accurate probability calibration for multiple classifiers. In: Proceedings of the 23rd international joint conference on artificial intelligence, Beijing, China, August, 2013, pp 1939–1945

Zouhal LM, Denœux T (1998) An evidence-theoretic k-nn rule with parameter optimization. IEEE Trans Syst Man Cybern C 28(2):263–271

Annexe F

A consistency-specificity trade-off to select source behavior in information fusion

---

Article paru dans *IEEE Transactions on Cybernetics*, 45(4) : 598-609, 2015.

A Consistency-Specificity Trade-Off to Select Source Behavior in Information Fusion

Frédéric Pichon, Sébastien Destercke, and Thomas Burger

***Abstract*—Combining pieces of information provided by several sources without or with little prior knowledge about the behavior of the sources is an old yet still important and rather open problem in the belief function theory. In this paper, we propose an approach to select the behavior of sources based on a very general and expressive fusion scheme, that has the important advantage of making clear the assumptions made about the sources. The selection process itself relies on two cornerstones that are the notions of specificity and consistency of a knowledge representation, and that we adapt to the considered fusion scheme. We illustrate our proposal on different examples and show that the proposed approach actually encompasses some important existing fusion strategies.**

***Index Terms*—Conflict, consistency, Dempster–Shafer theory, information fusion, specificity.**

I. Introduction

DETERMINING, from information provided by multiple sources, the actual value taken by an ill-known variable $\mathbf{x}$ defined on a space $\mathcal{X}$ is a central problem in many information systems, commonly known as information fusion. As previously argued in [10], [32], and [41], the task of information fusion necessarily involves making some (possibly uncertain) assumptions about the relation between the sources of information and about their behavior (e.g., their relevance and truthfulness [32]). A main concern in information fusion is thus to find, or to select, assumptions to make about the source behaviors for the fusion result to be sensible.

Two situations can be distinguished with respect to this problem: either one has some strong knowledge of the source behaviors, inherited from past experiences, or one has only vague or no knowledge about such behaviors. In the former case, depending on the form of the prior experience, i.e., data or expert knowledge, one may resort to some learning procedures [9], [14], [27] or multicriteria aggregations [7], [33] to estimate the behavior of the sources. When there is no or little prior experience with the sources, which is the case treated in this paper, then the selection of an appropriate assumption about source behaviors needs to be based on other considerations.

To tackle this problem, we propose a practical scheme that relies on the theoretical framework introduced by Pichon *et al*. [32] in the belief function theory [35], [42]. The problem of information fusion has indeed received a lot of attention within this theory [41] that allows for a flexible modeling of uncertainty. Pichon *et al*. [32] framework proposes to model source possible behaviors by generic sets of hypotheses, and provides a very expressive way of performing information fusion. It also has the advantage of making transparent the assumptions made about the sources, thus resulting in very readable and interpretable rules.

Yet, the framework remains theoretical, and does not propose any practical means to apply it nor ways to select a particular assumption. To make such a selection, we propose to base the selection on two criteria formalizing the two primary features one may seek regarding one's knowledge about $\mathbf{x}$: to make it, first, as specific as possible and, second, as consistent as possible.

Indeed, a fusion result is all the better if it satisfies both criteria. On the one hand, a specific but poorly consistent knowledge is not desirable, as it cannot be trusted: in belief function theory, it is usual to question the fusion results of the (unnormalized) Dempster's rule [4], [35], which assumes all sources to be truthful and relevant [32], when the inconsistency (conflict) [8] resulting from it is too high. How to deal with this inconsistency is itself a hard problem, as shows the literature on conflict management (e.g., [23] and [41]). On the other hand, a consistent but poorly specific result is also not desirable, as it is indecisive: this explains why some rules that makes weaker assumptions about the sources, such as the disjunctive rule [11], [39] that corresponds to assuming that at least one source is relevant, despite ensuring better consistency, are seldom used. Note that the goals of consistency and specificity are somehow antagonists, as the more informative the sources are, the more likely they will be conflicting about $\mathbf{x}$ true value. Let us remark also that the need to balance between consistency and informativeness of knowledge is present in other fields facing similar issues: for instance, to deal with inconsistency in a logical knowledge base, Grant and Hunter [15], [16] propose a stepwise procedure to improve consistency while minimizing information loss.

Manuscript received October 1, 2013; revised February 17, 2014 and June 12, 2014; accepted June 15, 2014. This work was supported in part by Thales Research and Technology France, in part by the ANR funding ANR-11-IDEX-0004-02 (Labex MS2T, "Investissements d'Avenir" call), in part by the ANR funding ANR-10-INBS-08 (ProFI project, "Infrastructures Nationales en Biologie et Santé"; "Investissements d'Avenir" call), and in part by the Prospectom project of the Mastodons 2012 challenge (CNRS). This paper is an extended and revised version of [31] with full proofs. This paper was recommended by Associate Editor S. E. Shimony.

F. Pichon is with Université Lille Nord de France, Lille F-59000, France, and also with the Université d'Artois, Béthune F-62400, France (e-mail: frederic.pichon@univ-artois.fr).

S. Destercke is with Centre National de la Recherche Scientifique, UMR 7253 Heudiasyc, Université de Technologie de Compiègne, Compiègne F-60200, France (e-mail: sebastien.destercke@hds.utc.fr).

T. Burger is with Université Grenoble-Alpes, CEA (iRSTV/BGE), INSERM (U1038), CNRS (FR3425), Grenoble F-38000, France (e-mail: thomas.burger@cea.fr).

Digital Object Identifier 10.1109/TCYB.2014.2331800

Our contribution may be summarized as follows: we define the notion of specificity and consistency by relying on the notion of specialization between belief functions and by extending recent works on conflict [8] to Pichon *et al.* [32] framework. We do this for the cases of single (Section III) and multiple (Section IV) sources. We then propose (Section V) a practical yet generic fusion scheme to find source assumptions that lead to a good consistency-specificity trade-off of the final result. To our knowledge, this scheme is the first to propose a practical way to apply Pichon *et al*. [32] theoretical results, and one of the few that explicitly exploits the consistency-specificity trade-off. The scheme is then illustrated in two ways: we propose (Section VI) various instances of sets of assumptions from which the selection can be done, connecting them to existing fusion strategies and shedding some new light on those strategies; we apply (Section VII) our fusion scheme to a nuclear safety problem, in which rule interpretability is as important as the fusion result. Background material is recalled in Section II.

II. Preliminaries

A. Basics of Belief Functions

Belief functions have been originally introduced by Dempster [4], mainly to model imprecise observations in statistical inferences. In this view, belief functions usually describe some imprecisely known probability distribution of a random variable. Shafer [35] then extended the theory so that belief functions could also model uncertainty related to ill-known but possibly fixed (deterministic) quantities. This interpretation, unrelated to probabilities and statistics, was taken over by Smets and Kennes [42] in the so-called transferable belief model. In this paper, we consider this latter interpretation, called singular in [8], where beliefs concern a fixed quantity. This is most often the case in information fusion problems, where one searches a unique true value of some variable of interest.

Accordingly, we assume the beliefs held by an agent about the actual value taken by a variable $\mathbf{x}$ defined on a finite domain $\mathcal{X}$, to be modeled using belief functions and to be represented using associated mass functions. A mass function $m^{\mathcal{X}}$ on $\mathcal{X}$ is defined as a mapping from the power set $2^{\mathcal{X}}$ to $[0, 1]$ satisfying $\sum_{A\subseteq\mathcal{X}} m^{\mathcal{X}}(A) = 1$. The mass $m^{\mathcal{X}}(A)$ may be understood as the amount of belief given to the assumption that the agent knows that the value of the variable of interest lies somewhere in set A, and nothing more specific [13].

From the mass function are usually defined two uncertainty measures, the belief and plausibility measures, which respectively read for an event $A \subseteq \mathcal{X}$

$$Bel(A) = \sum_{\emptyset\neq B\subseteq A} m^{\mathcal{X}}(B) \text{ and } Pl(A) = \sum_{B\cap A\neq\emptyset} m^{\mathcal{X}}(B).$$

That is, *Bel* is the sum of masses of sets that implies A, *Pl* the sum of masses of sets that are consistent with A. The contour function [35] $pl^{\mathcal{X}}{:}\mathcal{X} \to [0, 1]$ associated to a mass function $m^{\mathcal{X}}$ is defined by $pl^{\mathcal{X}}(x) = Pl^{\mathcal{X}}(\{x\})$. The focal sets of a mass $m^{\mathcal{X}}$ are the subsets A of $\mathcal{X}$ such that $m^{\mathcal{X}}(A) > 0$. We will denote by $\mathcal{F}$ the set of focal sets of $m^{\mathcal{X}}$. The classical notion of set E is modeled by the categorical mass $m(E) = 1$. Besides, a mass function $m^{\mathcal{X}}$ is called Bayesian if its focal sets are singletons.

B. Comparing Informative Contents

The most natural way to compare the informative content of two sets E_1, E_2 is to say that E_1 is more informative than E_2 if $E_1 \subset E_2$. This can be extended in several ways to compare the informative content of mass functions in terms of specificity [13]. In a singular interpretation, the most sensible extension is arguably the notion of specialization, that we use here.

Definition 1 (Specialization): A mass function $m_1^{\mathcal{X}}$ defined on $\mathcal{X}$ with $\mathcal{F}_1 = \{E_1, \ldots, E_q\}$ is said to be a specialization of another mass function $m_2^{\mathcal{X}}$ defined on $\mathcal{X}$ with $\mathcal{F}_2 = \{F_1, \ldots, F_p\}$ if and only if there exists a nonnegative matrix $W = [w_{ij}]$ of size $q \times p$ such that

$$\text{for} \quad i = 1, \ldots, p, \quad \sum_{j=1}^{q} w_{ij} = m_1^{\mathcal{X}}(E_i)$$

$$\text{for} \quad j = 1, \ldots, q, \quad \sum_{i=1}^{p} w_{ij} = m_2^{\mathcal{X}}(F_j)$$

$$w_{ij} > 0 \Rightarrow E_i \subseteq F_j.$$

This relation is denoted by $m_1^{\mathcal{X}} \sqsubseteq m_2^{\mathcal{X}}$ and by $m_1^{\mathcal{X}} \sqsubset m_2^{\mathcal{X}}$ if there is at least a pair i, j such that $w_{ij} > 0$ and $E_i \subset F_j$.

This can be seen as a transfer of mass from each E_i to supersets F_j, w_{ij} denoting the part of $m_1(E_i)$ transferred to F_j. In other words, $m_1^{\mathcal{X}}$ is a specialization of $m_2^{\mathcal{X}}$ if the mass of any focal set F_j of $m_2^{\mathcal{X}}$ can be redistributed among subsets of E_i in $m_1^{\mathcal{X}}$. Let us recall that we have [11]

$$m_1^{\mathcal{X}} \sqsubseteq m_2^{\mathcal{X}} \Rightarrow pl_1^{\mathcal{X}}(x) \leq pl_2^{\mathcal{X}}(x), \quad \forall x \in \mathcal{X}. \tag{1}$$

C. Source Behavioral States

Pichon *et al.* [32] framework integrating source behaviors is the following. Assume an agent wants to know the actual value taken by $\mathbf{x}$ based on testimonies provided by several sources of information identified as $\mathfrak{s}_i$, $1 \leq i \leq K$. These testimonies can be of several forms: a value $x_i \in \mathcal{X}$, a set $A_i \subseteq \mathcal{X}$, a probability distribution p_i on $\mathcal{X}$, or in the most general form a mass function $m_i^{\mathcal{X}}$ on $\mathcal{X}$. To interpret those testimonies, the agent must have some knowledge (or make some assumptions) about the behavioral state (referred to as meta-knowledge in [32]) of the sources.

In the approach of Pichon *et al.* [32], the possible elementary behavioral states of a source $\mathfrak{s}_i$ are formalized as a set $\mathcal{H}^i = \{h_1^i, \ldots, h_N^i\}$. The set of elementary joint states on sources is therefore the Cartesian product $\mathcal{H}^K := \times_{i=1}^K \mathcal{H}^i$. The state space $\mathcal{H}^i$ can be very general and may include being unreliable, lying, etc. Two common assumptions for which we will use specific notations are the assumptions that a source $\mathfrak{s}_i$ is relevant (R^i) or not ($\neg R^i$), and truthful (T^i) or not ($\neg T^i$). Together, they form the space of possible states $\mathcal{H}^i = \{(T^i, R^i), (T^i, \neg R^i), (\neg T^i, R^i), (\neg T^i, \neg R^i)\}$. Like the testimonies provided by the sources, the meta-knowledge of the

agent can be of several forms, the most general one being a mass function defined over $\mathcal{H}^K$.

We can now detail how the notions of consistency introduced in [8] and of specificity (specialization) can be extended to include source behaviors, and how to then use these extensions to select an assumption about the sources.

III. Consistency and Specificity: Single Source

A. Crisp Testimony and Sure Meta-Knowledge

The simplest situation is a source $\mathfrak{s}$ delivering a testimony of the form $\mathbf{x} \in A$ with $A \subseteq \mathcal{X}$, and being known to be in a state $h \in \mathcal{H}$, with $\mathcal{H}$ the state space of the source. The testimony $\mathbf{x} \in A$ should then be modified according to this state [32]. This transformation can be encoded by a multivalued mapping Γ_A:$\mathcal{H} \rightarrow \mathcal{X}$, where $\Gamma_A(h)$ indicates how to interpret the piece of information $\mathbf{x} \in A$ for each possible state h of the source. For instance, if $\mathcal{H} = \{(T, R), (T, \neg R), (\neg T, R), (\neg T, \neg R)\}$ are the possible states of the source, we have for all $A \subseteq \mathcal{X}$

$$\begin{aligned} \Gamma_A(R, T) &= A \\ \Gamma_A(\neg R, T) &= \mathcal{X} \\ \Gamma_A(R, \neg T) &= A^c \\ \Gamma_A(\neg R, \neg T) &= \mathcal{X} \end{aligned} \tag{2}$$

with A^c the complement of A. Equation (2) translates that if $\mathfrak{s}$ is not relevant, it does not bring any information, while if it is not truthful, it declares the opposite of what it knows to be true [29]. If the knowledge about the source state is imprecise and given by $H \subseteq \mathcal{H}$, then the transformation is the image $\Gamma_A(H) := \bigcup_{h \in H} \Gamma_A(h)$ of H by Γ_A.

To measure consistency, we extend the work of Destercke and Burger [8], where any piece of knowledge $\mathbf{x} \in A$ about a variable $\mathbf{x}$ is considered consistent if $A \neq \emptyset$, and inconsistent otherwise. This extends easily to the current framework, a transformed testimony yielding a consistent piece of knowledge on $\mathcal{X}$ when $\Gamma_A(H) \neq \emptyset$, in which case $\mathbf{x} \in A$ is said *H-consistent*, and an inconsistent piece of knowledge when $\Gamma_A(H) = \emptyset$. We can then extend the measure of consistency of $\mathbf{x} \in A$ introduced in [8], to measure H-consistency of a testimony $\mathbf{x} \in A$ as the degree $\phi_H : 2^{\mathcal{X}} \rightarrow \{0, 1\}$ such that

$$\phi_H(A) = \begin{cases} 1 & \text{if } \Gamma_A(H) \neq \emptyset \\ 0 & \text{if } \Gamma_A(H) = \emptyset. \end{cases}$$

In some way, this consistency measure evaluates whether H is a valid assumption on the source when it provides the testimony $\mathbf{x} \in A$. Consider, for instance, the assumption $h = (R, \neg T)$ corresponding to a relevant and lying source. This assumption will be considered invalid only when the source provides the certainly true testimony $\mathbf{x} \in \mathcal{X}$, as $\Gamma_{\mathcal{X}}(h) = \emptyset$ and $\phi_h(\mathcal{X}) = 0$.

Meta-knowledge can also be characterized in terms of specificity: a piece of meta-knowledge $H_1 \subseteq \mathcal{H}$ will be said at least as meta-specific as another piece of meta-knowledge $H_2 \subseteq \mathcal{H}$ when $\Gamma_A(H_1) \subseteq \Gamma_A(H_2)$ for any $A \subseteq \mathcal{X}$, and we will denote it $H_1 \sqsubseteq_{\mathcal{H}} H_2$. For example, the assumption (R, T) is at least as meta-specific as the assumption $(\neg R, T)$. In general, this only induces a partial order over possible states, as for instance none of the two assumptions (R, T) and $(R, \neg T)$ is more meta-specific than another. Note that we have the relation

$$H_1 \subseteq H_2 \Rightarrow H_1 \sqsubseteq_{\mathcal{H}} H_2 \tag{3}$$

since $H_1 \subseteq H_2 \Rightarrow \Gamma_A(H_2) = \Gamma_A(H_1) \cup (\bigcup_{h \in H_2 \setminus H_1} \Gamma_A(h))$ and thus $\Gamma_A(H_2) \supseteq \Gamma_A(H_1)$. We also have

$$H_1 \sqsubseteq_{\mathcal{H}} H_2 \Rightarrow \phi_{H_1}(A) \leq \phi_{H_2}(A) \tag{4}$$

since either $\Gamma_A(H_1) \neq \emptyset$ thus $\Gamma_A(H_2) \neq \emptyset$ (definition of $\sqsubseteq_{\mathcal{H}}$) and then $\phi_{H_1}(A) = \phi_{H_2}(A)$, or $\Gamma_A(H_1) = \emptyset$ and thus $\phi_{H_1}(A) = 0 \Rightarrow \phi_{H_1}(A) \leq \phi_{H_2}(A), \ \forall \ \Gamma_A(H_2)$.

Remark 1: Relation (4) clearly indicates that reaching both consistency and specificity are somewhat opposite goals.

B. Uncertain Testimony and Meta-Knowledge

More generally, both the testimony and the meta-knowledge of the agent may be uncertain. Let $m^{\mathcal{X}}$ be the uncertain testimony and $m^{\mathcal{H}}$ the uncertain meta-knowledge. The knowledge of the agent on $\mathcal{X}$ is then given by the mass function $m[m^{\mathcal{H}}]^{\mathcal{X}}$ defined for all $B \subseteq \mathcal{X}$ as [32]

$$m\left[m^{\mathcal{H}}\right]^{\mathcal{X}}(B) = \sum_{H \subseteq \mathcal{H}} m^{\mathcal{H}}(H) \sum_{A : \Gamma_A(H) = B} m^{\mathcal{X}}(A). \tag{5}$$

This definition is rather general and cover numerous cases, such as Shafer's discounting rule [35], as explained in [32].

The results of the previous section can be extended to this general setting: following [8], the mass function modeling the empty set ($m[m^{\mathcal{H}}]^{\mathcal{X}}(\emptyset) = 1$) can be associated to a complete inconsistent knowledge and a mass function $m[m^{\mathcal{H}}]^{\mathcal{X}}$ whose focal sets have a nonempty intersection can be associated to a totally consistent knowledge. That is, the testimony $m^{\mathcal{X}}$ is totally consistent under meta-knowledge $m^{\mathcal{H}}$ if and only if

$$\bigcap_{\substack{A \in \mathcal{F} \\ H \in \mathcal{F}_{\mathcal{H}}}} \Gamma_A(H) \neq \emptyset \tag{6}$$

where $\mathcal{F}$ and $\mathcal{F}_{\mathcal{H}}$ denote the sets of focal sets of $m^{\mathcal{X}}$ and $m^{\mathcal{H}}$, respectively. A mass function $m^{\mathcal{X}}$ is then said $m^{\mathcal{H}}$-consistent if and only if (6) holds. Lemma 1 characterizes $m^{\mathcal{H}}$-consistent testimonies in terms of the contour function.

Lemma 1: $\bigcap_{\substack{A \in \mathcal{F} \\ H \in \mathcal{F}_{\mathcal{H}}}} \Gamma_A(H) \neq \emptyset \Leftrightarrow \exists x \in \mathcal{X}$ such that $pl[m^{\mathcal{H}}]^{\mathcal{X}}(x) = 1$, where $pl[m^{\mathcal{H}}]^{\mathcal{X}}$ is the contour function associated to $m[m^{\mathcal{H}}]^{\mathcal{X}}$ obtained from (5).

Proof: This follows directly from [8, Lemma 1], when one recognizes that subsets $\Gamma_A(H) \subseteq \mathcal{X}$, such that $A \in \mathcal{F}$ and $H \subseteq \mathcal{F}_{\mathcal{H}}$, are the focal sets of $m[m^{\mathcal{H}}]^{\mathcal{X}}$. ■

A source is thus $m^{\mathcal{H}}$-consistent if it allows us to conclude that at least one value of $\mathbf{x}$ is totally plausible under meta-knowledge $m^{\mathcal{H}}$. Following [8], this characterization of $m^{\mathcal{H}}$-consistency suggests the following definition.

Definition 2: [$m^{\mathcal{H}}$-consistency measure] The measure $\phi_{m^{\mathcal{H}}}$:$\mathcal{M}^{\mathcal{X}} \rightarrow [0, 1]$ of $m^{\mathcal{H}}$-consistency, where $\mathcal{M}^{\mathcal{X}}$ denotes the set of all mass functions on $\mathcal{X}$, reads

$$\phi_{m^{\mathcal{H}}}(m^{\mathcal{X}}) = \max_{x \in \mathcal{X}} pl\left[m^{\mathcal{H}}\right]^{\mathcal{X}}(x). \tag{7}$$

Equation (7) is relatively easy to evaluate. Indeed, in (5), the quantity $\sum_{A:\Gamma_A(H)=B} m^{\mathcal{X}}(A)$ is the mass allocated to $B \subseteq \mathcal{X}$ when the source is assumed to be in some state $H \subseteq \mathcal{H}$, i.e., we have

$$m[H]^{\mathcal{X}}(B) = \sum_{A:\Gamma_A(H)=B} m^{\mathcal{X}}(A).$$

Equation (5) can thus be rewritten

$$m\left[m^{\mathcal{H}}\right]^{\mathcal{X}}(B) = \sum_{H\subseteq\mathcal{H}} m^{\mathcal{H}}(H) m[H]^{\mathcal{X}}(B). \tag{8}$$

As (8) is a convex mixture of mass functions (each $m[H]^{\mathcal{X}}$ is weighted by $m^{\mathcal{H}}(H)$ which is positive and sums up to one), and as the plausibility measure of a convex mixture is the convex mixture of plausibility measures, computing (7) only requires to compute the weighted average of the contour functions of $m[H]^{\mathcal{X}}$ for all $H \in \mathcal{F}_{\mathcal{H}}$.

Remark 2: Many generalizations of Shannon entropy (see [20, Ch. 3] for a review) have been defined for belief functions, such as the degree of dissonance [20, Sec. 6.5], which corresponds to a mean value of conflict or inconsistency between the focal sets of a mass function. Using them instead of Def. 2 is tempting, however, it should be noticed that most of them would not reach extremal values when $\bigcap_{\substack{A\in\mathcal{F}\\ H\in\mathcal{F}_{\mathcal{H}}}} \Gamma_A(H) \neq \emptyset$ nor when $m[m^{\mathcal{H}}]^{\mathcal{X}}(\emptyset) = 1$.

Remark 3: The more classical consistency measure $m(\emptyset)$ is also considered in [8]. We will not consider it here, as: 1) it is argued in [8] that this measure is less adapted to a singular interpretation; 2) estimating $m(\emptyset)$ usually requires heavier computations, hence is of less practical interest; and 3) most results presented here adapt easily to $m(\emptyset)$.

Meta-specificity may also be generalized to this setting.

Definition 3 (Meta-specificity): An uncertain piece of meta-knowledge $m_1^{\mathcal{H}}$ is said to be at least as meta-specific as another uncertain piece $m_2^{\mathcal{H}}$ when $m[m_1^{\mathcal{H}}]^{\mathcal{X}} \sqsubseteq m[m_2^{\mathcal{H}}]^{\mathcal{X}}$ for any $m^{\mathcal{X}} \in \mathcal{M}^{\mathcal{X}}$. This is denoted by $m_1^{\mathcal{H}} \sqsubseteq_{\mathcal{H}} m_2^{\mathcal{H}}$.

We may then show that in the general case we have relations extending (3) and (4); in particular that consistency and specificity are also at odds as shown by Proposition 2.

Proposition 1: Let $m_1^{\mathcal{H}}, m_2^{\mathcal{H}} \in \mathcal{M}^{\mathcal{H}}$ such that $m_1^{\mathcal{H}} \sqsubseteq m_2^{\mathcal{H}}$. We have $m_1^{\mathcal{H}} \sqsubseteq_{\mathcal{H}} m_2^{\mathcal{H}}$.

Proof: Consider a focal set A of a mass $m^{\mathcal{X}}$ and a focal set H_j of $m_2^{\mathcal{H}}$. The mass $m^{\mathcal{X}}(A)m_2^{\mathcal{H}}(H_j)$ is then affected to $\Gamma_A(H_j)$. Now, considering the states H_i of $m_1^{\mathcal{H}}$, we have that a fraction $w_{ij} > 0$ of the mass $m^{\mathcal{X}}(A)m_1^{\mathcal{H}}(H_i)$ affected to $\Gamma_A(H_i)$ will be transferred to $\Gamma_A(H_j)$, and $\Gamma_A(H_i) \subseteq \Gamma_A(H_j)$ since $H_i \subseteq H_j$ by definition. ■

Proposition 2: Let $m_1^{\mathcal{H}}, m_2^{\mathcal{H}} \in \mathcal{M}^{\mathcal{H}}$ such that $m_1^{\mathcal{H}} \sqsubseteq_{\mathcal{H}} m_2^{\mathcal{H}}$. We have $\phi_{m_1^{\mathcal{H}}}(m^{\mathcal{X}}) \leq \phi_{m_2^{\mathcal{H}}}(m^{\mathcal{X}})$ for all $m^{\mathcal{X}} \in \mathcal{M}^{\mathcal{X}}$.

Proof: From the definition of $\sqsubseteq_{\mathcal{H}}$, we have $m[m_1^{\mathcal{H}}]^{\mathcal{X}} \sqsubseteq m[m_2^{\mathcal{H}}]^{\mathcal{X}}$ for all $m^{\mathcal{X}} \in \mathcal{M}^{\mathcal{X}}$, which from (1) implies $pl[m_1^{\mathcal{H}}]^{\mathcal{X}}(x) \leq pl[m_2^{\mathcal{H}}]^{\mathcal{X}}(x)$ for all $x \in \mathcal{X}$. ■

The notions introduced in this section, and in particular Proposition 2, are illustrated in Example 1.

Example 1 (Inspired from Example 1 of [32]): Let $\mathcal{X} = \{x_1, x_2, x_3, x_4, x_5\}$ be an ordered space and consider $m^{\mathcal{X}}$ such that $m^{\mathcal{X}}(\{x_1, x_2\}) = 0.3$, $m^{\mathcal{X}}(\{x_4, x_5\}) = 0.3$ and $m^{\mathcal{X}}(\{x_3\}) = 0.4$. Now consider the following assumptions.

1) h_1 "reliable" such that $\Gamma_A(h_1) = A$.
2) h_3 "unreliable" such that $\Gamma_A(h_3) = \mathcal{X}$.
3) h_2 "approximately reliable" such that if $A = \{x_i, x_{i+1}, \ldots, x_j\}$ then $\Gamma_A(h_2) = \{x_{i-1}\} \cup A \cup \{x_{j+1}\}$ with $x_0 = x_6 = \emptyset$ and meaning that the source is not totally reliable neither totally unreliable, but it is somewhere between these two extremes.

We have $h_1 \sqsubseteq_{\mathcal{H}} h_2 \sqsubseteq_{\mathcal{H}} h_3$ while $\phi_{h_1}(m^{\mathcal{X}}) = 0.4$, $\phi_{h_2}(m^{\mathcal{X}}) = 1$, $\phi_{h_3}(m^{\mathcal{X}}) = 1$.

This example allows us to lay bare some preliminary ideas on source behavior selection with a consistency-specificity trade-off. As observed, assumptions h_2 and h_3 are the most desirable in terms of consistency, since they both yield a totally consistent state of knowledge. However, as the state of knowledge obtained under h_2 is more specific (informative) than the one obtained under h_3, h_2 appears preferable. This will be developed at length in Section V.

IV. Consistency and Specificity: Multiple Sources

Let us now consider the main case where multiple sources $\mathfrak{s}_i$, $i = 1, \ldots, K$ provide information, each as a mass function $m_i^{\mathcal{X}}$. As recalled in the Introduction section, the main problem in such a case is to combine these pieces of information in a sensible way. Since we focus in this paper on the problem of finding appropriate source behaviors, the sources will be assumed to rely on distinct evidences [38], [41], a usual assumption when merging belief functions.

Many combination rules have been proposed for belief functions [41]: the most usual is the unnormalized Dempster's rule (or conjunctive rule), which applies when sources are assumed to rely on distinct evidences and are both relevant and truthful. It is denoted here by $\textcircled{\cap}$. The mass $m_{1\textcircled{\cap}2}^{\mathcal{X}}$ resulting from its application on $m_1^{\mathcal{X}}$ and $m_2^{\mathcal{X}}$ is

$$m_{1\textcircled{\cap}2}^{\mathcal{X}}(A) = \sum_{B\cap C=A} m_1^{\mathcal{X}}(B)\, m_2^{\mathcal{X}}(C), \quad \forall A \subseteq \mathcal{X}. \tag{9}$$

The disjunctive rule $\textcircled{\cup}$ [11], [39] is obtained by simply replacing $\cap$ with $\cup$ in (9). Both the conjunctive and disjunctive rules can be given a clear interpretation in terms of source behavior assumptions [32]. In this section, we extend our characterization of consistency and specificity to the theoretical results of Pichon *et al.* [32].

A. General Case: Uncertain Testimonies and Meta-Knowledge

To simplify notations, we consider that all source $\mathfrak{s}_i$ share the same[1] possible state space $\mathcal{H} = \{h_1, \ldots, h_N\}$. For any state $\mathbf{h} = (h^1, \ldots, h^k) \in \mathcal{H}^K$ we define a mapping for any $\mathbf{A} = (A_1, \ldots, A_K) \subseteq \mathcal{X}^K$ as $\Gamma_{\mathbf{A}}(\mathbf{h}) = \bigcap_{i=1}^K \Gamma_{A_i}(h^i)$. $\Gamma_{\mathbf{A}}(\mathbf{h})$ represents the information on $\mathcal{X}$ deduced from testimonies $(A_1, \ldots, A_K)$ provided by sources $\mathfrak{s}_1, \ldots, \mathfrak{s}_K$ when they are in states $(h^1, \ldots, h^K)$ [32]. For nonelementary hypotheses

[1]The extension to particularized state spaces $\mathcal{H}^i$ is straightforward.

$H \subseteq \mathcal{H}^K$, we keep the previous notation, i.e., $\Gamma_{\mathbf{A}}(H) := \cup_{\mathbf{h}\in H}\Gamma_{\mathbf{A}}(\mathbf{h}),\ \forall H \subseteq \mathcal{H}^K,\ \forall \mathbf{A} \subseteq \mathcal{X}^K$.

If we have some joint meta-knowledge $m^{\mathcal{H}^K}$ over $\mathcal{H}^K$ and if sources $\mathfrak{s}_1, \ldots, \mathfrak{s}_K$ deliver distinct testimonies $m_i^{\mathcal{X}}$, $i = 1, \ldots, K$, then the combined mass function $m[m^{\mathcal{H}^K}]^{\mathcal{X}}$ defined by (10) represents what can be inferred about $\mathbf{x}$ from $\mathbf{m}^{\mathcal{X}} = (m_1^{\mathcal{X}}, \ldots, m_K^{\mathcal{X}})$ [32]

$$m\left[m^{\mathcal{H}^K}\right]^{\mathcal{X}}(B) = \sum_{H\subseteq\mathcal{H}^K} m^{\mathcal{H}^K}(H) \sum_{\substack{\mathbf{A}\subseteq\mathcal{X}^K \\ \Gamma_{\mathbf{A}}(H)=B}} \left[\prod_{i=1}^{K} m_i^{\mathcal{X}}(A_i)\right] \quad (10)$$

the product $\prod_{i=1}^{K} m_i^{\mathcal{X}}(A_i)$ coming from the distinctness assumption.

An interesting feature of using meta-knowledge is that all Boolean operators on sets $\mathbf{A} = (A_1, \ldots, A_K) \subseteq \mathcal{X}^K$ can be obtained through particular assumptions on the behavior of the sources [32]. As a result, (10) covers all combination rules based on Boolean operators. For instance, consider the assumption H_r^K on $\mathcal{H}^K$ meaning the sources are truthful and "r-out-of-K" of them are relevant (an assumption that is common in interval analysis [17]). This amounts to

$$\Gamma_{\mathbf{A}}\left(H_r^K\right) = \bigcup_{\mathcal{A}\subseteq\{A_1,\ldots,A_K\},|\mathcal{A}|=r} (\cap_{A\in\mathcal{A}} A)\,. \quad (11)$$

When applying H_r^K to (10), the conjunctive and disjunctive rules are retrieved when $r = K$ and $r = 1$, respectively.

Keeping the same definition of complete inconsistent and consistent knowledge as in Section III-B, the counterpart of Lemma 1 suggests to use the following equation as a degree of $m^{\mathcal{H}^K}$-consistency for the collection $\mathbf{m}^{\mathcal{X}}$:

$$\phi_{m^{\mathcal{H}^K}}\left(\mathbf{m}^{\mathcal{X}}\right) = \max_{x\in\mathcal{X}} pl\left[m^{\mathcal{H}^K}\right]^{\mathcal{X}}(x) \quad (12)$$

where $pl[m^{\mathcal{H}^K}]$ is the contour function of (10). Equation (12) extends the conflict measure defined by Destercke and Burger [8] to any combination rule that can be obtained from (10), in particular to all rules based on Boolean operators. In addition, we have again that if $m_1^{\mathcal{H}^K} \sqsubseteq_{\mathcal{H}} m_2^{\mathcal{H}^K}$, then $\phi_{m_1^{\mathcal{H}^K}}(\mathbf{m}^{\mathcal{X}}) \leq \phi_{m_2^{\mathcal{H}^K}}(\mathbf{m}^{\mathcal{X}})$ for $\mathbf{m}^{\mathcal{X}}$.

Those powerful results, i.e., the extension of the conflict measure to any combination rule and the duality between specificity and consistency are heavily used in Section V where we introduce a method to select source behaviors.

B. Computation of the Consistency Degree

Equation (10) can be rewritten as

$$m\left[m^{\mathcal{H}^K}\right]^{\mathcal{X}}(B) = \sum_{H\subseteq\mathcal{H}^K} m^{\mathcal{H}^K}(H)\, m[H]^{\mathcal{X}}(B) \quad (13)$$

with $m[H]^{\mathcal{X}}$ the mass function representing the knowledge inferred on $\mathcal{X}$ when the sources are assumed to be in some state $H \subseteq \mathcal{H}^K$, defined by

$$m[H]^{\mathcal{X}}(B) = \sum_{\mathbf{A}\subseteq\mathcal{X}^K;\Gamma_{\mathbf{A}}(H)=B} \left[\prod_{i=1}^{K} m_i^{\mathcal{X}}(A_i)\right].$$

Hence, the computation of (12), which can be resource demanding, may be simplified as in the single source case, using the convexity property of plausibility measures. However, there are cases where it is easier to compute (13) and thus even easier to compute (12) than merely using this convexity property. In particular, when all focal elements of $m^{\mathcal{H}^K}$ are behavior-separable (or b-separable).

Definition 4 (Behavior-separability): A subset $H \subseteq \mathcal{H}^K$ is said b-separable iff $H = H^{\downarrow 1} \times \ldots \times H^{\downarrow K}$ (where $H^{\downarrow i}$ is the projection of $H \subseteq \mathcal{H}^K$ on the ith source state space).

Proposition 3: When each focal set of $m^{\mathcal{H}^K}$ is b-separable, (13) can be rewritten as

$$m\left[m^{\mathcal{H}^K}\right]^{\mathcal{X}}(B) = \sum_{H\subseteq\mathcal{H}^K} m^{\mathcal{H}^K}(H) \left[\textcircled{\cap}_{i=1}^{K} m\left[H^{\downarrow i}\right]^{\mathcal{X}}\right](B) \quad (14)$$

where $m[H^{\downarrow i}]^{\mathcal{X}}$ denotes $m_i^{\mathcal{X}}$ transformed according to $H^{\downarrow i}$.

Proof: Let H be a b-separable assumption on the sources. Such b-separable meta-knowledge H satisfies the property of so-called meta-independence in [32]. Therefore, from [32, Th. 1], we have

$$m[H]^{\mathcal{X}}(B) = \left[\textcircled{\cap}_{i=1}^{K} m\left[H^{\downarrow i}\right]^{\mathcal{X}}\right](B). \quad (15)$$

If each focal set of a piece of meta-knowledge $m^{\mathcal{H}^K}$ is b-separable, it is direct to obtain (14) from (13) and (15). ■

That is, to compute $m[m^{\mathcal{H}^K}]^{\mathcal{X}}$ we first transform each $m_i^{\mathcal{X}}$ according to $H^{\downarrow i}$, apply unnormalized Dempster's rule to them and compute the weighted sum according to $m^{\mathcal{H}^K}$. We can therefore make use of efficient algorithms to compute Dempster's rule result [43].

This property simplifies the computation of the consistency measure (12) as follows. Consider the meta-knowledge $m^{\mathcal{H}^K}(H) = 1$ with H b-separable and let $pl[H]^{\mathcal{X}}$ be the corresponding contour function. Furthermore, let $pl[H^{\downarrow i}]^{\mathcal{X}}$ denote the contour function obtained by transforming $m_i^{\mathcal{X}}$ according to meta-knowledge $H^{\downarrow i}$. We have

$$pl[H]^{\mathcal{X}}(x) = \prod_{i=1}^{K} pl\left[H^{\downarrow i}\right]^{\mathcal{X}}(x). \quad (16)$$

Now, let $m^{\mathcal{H}^K}$ be a piece of meta-knowledge with b-separable focal sets, we have then

$$pl\left[m^{\mathcal{H}^K}\right]^{\mathcal{X}}(x) = \sum_{H\subseteq\mathcal{H}^K} m^{\mathcal{H}^K}(H) \cdot \prod_{i=1}^{K} pl\left[H^{\downarrow i}\right]^{\mathcal{X}}(x).$$

In other words, when each focal set of a piece of meta-knowledge is b-separable, computing consistency measure (12) only requires to compute contour functions and to take their weighted averaged products (hence not necessitating any combination). However, let us stress that in general

$$\phi_{m^{\mathcal{H}^K}}(\mathbf{m}^{\mathcal{X}}) \neq \sum_{H} m^{\mathcal{H}^K}(H)\phi_H\left(\mathbf{m}^{\mathcal{X}}\right) \quad (17)$$

even in the case where the focal sets of $m^{\mathcal{H}^K}$ are b-separable, as shown by Example 2.

Example 2: Let $\mathbf{m}^{\mathcal{X}} = (m_1^{\mathcal{X}}, m_2^{\mathcal{X}})$ with $\mathcal{X} = \{x_1, x_2\}$ and $m_1^{\mathcal{X}}(\{x_1\}) = 0.2, m_1^{\mathcal{X}}(\{x_2\}) = 0.3, m_1^{\mathcal{X}}(\mathcal{X}) = 0.5$ and $m_2^{\mathcal{X}}(\{x_1\}) = 0.4, m_2^{\mathcal{X}}(\{x_2\}) = 0.2, m_2^{\mathcal{X}}(\mathcal{X}) = 0.4$. Furthermore, sources $\mathfrak{s}_1$ and $\mathfrak{s}_2$ are assumed to be truthful and to have the following joint behavior in terms of relevance: with mass 0.7 $\mathfrak{s}_1$ is relevant and $\mathfrak{s}_2$ is not, and with mass 0.3 $\mathfrak{s}_2$ is relevant and $\mathfrak{s}_1$ is not. Let $m^{\mathcal{H}^2}$ denote such meta-knowledge on the sources

$$m^{\mathcal{H}^2}(H_1 = \{((R,T),(\neg R,T))\}) = 0.7$$
$$m^{\mathcal{H}^2}(H_2 = \{((\neg R,T),(R,T))\}) = 0.3.$$

We have [32] $m[m^{\mathcal{H}^2}]^{\mathcal{X}} = 0.7 \cdot m_1^{\mathcal{X}} + 0.3 \cdot m_2^{\mathcal{X}}$. Computing the contour function associated to $m[m^{\mathcal{H}^2}]^{\mathcal{X}}$, we find

$$pl[m^{\mathcal{H}^2}]^{\mathcal{X}}(x_1) = 0.7 \cdot pl_1^{\mathcal{X}}(x_1) + 0.3 \cdot pl_2^{\mathcal{X}}(x_1) = 0.73$$
$$pl[m^{\mathcal{H}^2}]^{\mathcal{X}}(x_2) = 0.7 \cdot pl_1^{\mathcal{X}}(x_2) + 0.3 \cdot pl_2^{\mathcal{X}}(x_2) = 0.74$$

and thus $\phi_{m^{\mathcal{H}^2}}(\mathbf{m}^{\mathcal{X}}) = 0.74$. On the other hand, we have

$$m^{\mathcal{H}^2}(H_1) \cdot \phi_{\{H_1\}}\left(\mathbf{m}^{\mathcal{X}}\right) + m^{\mathcal{H}^2}(H_2) \cdot \phi_{H_2}\left(\mathbf{m}^{\mathcal{X}}\right) = 0.7 \cdot 0.8 + 0.3 \cdot 0.8 = 0.8. \quad (18)$$

Such a behavior can be easily explained by the fact that the sum operator of (17) and the maximum operator of (12) do not distribute over each other.

V. Source Behavior Selection Approach

When only little knowledge about the sources is available, it is not possible to estimate their behavior using learning procedures or multicritera aggregations (see Section I), as information to do so is lacking. In such a case, a strategy is to consider a set of sensible behavior assumptions—hence a set of readable rules—to choose from, together with selection criteria. We provide guidelines to define such a set and selection criterion, based on previous section materials. Roughly speaking, we specify a set of behavior assumptions inducing decreasingly specific results, and take a minimal consistency threshold as a simple, single selection parameter to pick an assumption from this set. This leads to a general, yet practical and sensible, approach to select the behavior of the sources in poorly informed cases. Instances of Section VI illustrate the approach.

A. Initial Meta-Knowledge

We propose to consider a basic initial assumption $m_1^{\mathcal{H}^K}$ such that $m_1^{\mathcal{H}^K}(\mathbf{h}) = 1$, with $\mathbf{h} \in \mathcal{H}^K$ and $\Gamma_A(\mathbf{h}^{\downarrow i}) = A$ for all $A \subseteq \mathcal{X}$ and $i = 1, \ldots, K$, i.e., an assumption that induces no transformation of the testimonies provided by the sources.

This assumption corresponds to not altering in any way the initial information, i.e., we accept the testimonies as they are. Most importantly, this assumption is the most classical one in information fusion in general and in belief function theory in particular, as it corresponds to unnormalized Dempster's rule. Hence, $m_1^{\mathcal{H}^K}$ is a natural default meta-knowledge.

Equation (12) gives an assessment of whether the assumption $m_1^{\mathcal{H}^K}$ applies to the current testimonies. As is classically advocated in belief function theory, we propose that assumption $m_1^{\mathcal{H}^K}$ should be used to combine the testimonies if the consistency derived from (12) induced by $m_1^{\mathcal{H}^K}$ is high enough, that is if it is above some threshold τ, and that $m_1^{\mathcal{H}^K}$ should be rejected as a valid assumption if the consistency is too low, i.e., below τ. In this latter case, other assumptions leading to higher consistency should be sought.

B. Specificity Ordering Approach

To define other assumptions that will result in more consistent results after merging, we use the counterpart of Proposition 2 in the multiple source case: choosing a meta-knowledge $m_2^{\mathcal{H}^K}$ such that $m_1^{\mathcal{H}^K} \sqsubseteq_{\mathcal{H}} m_2^{\mathcal{H}^K}$ will ensure a consistency increment. This leads us to propose the following two-step strategy to select the meta-knowledge to be used.

1) Define a collection of meta-knowledge $\mathbf{m}^{\mathcal{H}^{\mathbf{K}}} = (m_1^{\mathcal{H}^K}, \ldots, m_M^{\mathcal{H}^K})$ such that for any $1 \leq j < M$, $m_j^{\mathcal{H}^K} \sqsubset_{\mathcal{H}} m_{j+1}^{\mathcal{H}^K}$, and with $m_1^{\mathcal{H}^K}$ as defined above.
2) Test each $m_j^{\mathcal{H}^K}$ iteratively with $j = 1, \ldots, M$, until $\phi_{m_j^{\mathcal{H}^K}}(\mathbf{m}^{\mathcal{X}}) \geq \tau$.

This ensures that, at each iteration from j to $j+1$, specificity will decrease since $m_j^{\mathcal{H}^K} \sqsubset_{\mathcal{H}} m_{j+1}^{\mathcal{H}^K}$ and consistency will increase since $\phi_{m_j^{\mathcal{H}^K}}(\mathbf{m}^{\mathcal{X}}) \leq \phi_{m_{j+1}^{\mathcal{H}^K}}(\mathbf{m}^{\mathcal{X}})$, the process stopping whenever one thinks the result is consistent (trustworthy) enough. In other words, this strategy gradually decreases specificity until a satisfactory consistency level is reached. Once a collection is defined (some interesting instances are given in Section VI), the only parameter to set is τ (any value over 0.5 seems reasonable, with 0.8 being a good compromise).

In itself, this idea is not entirely new, as it already appears in Dubois and Prade [12], where the foundations of source behavior assumptions are laid bare. More recently, one can find specificity-based comparisons of combination rules in [37], or strategies relaxing specificity to gain consistency [21], [22]. Yet, to our knowledge this is the first proposal to provide a so generic formal procedure allowing one to select interpretable fusion rules and to rely on consistency measured through contour functions.

Remark 4: The construction of $\mathbf{m}^{\mathcal{H}^{\mathbf{K}}}$ should be based on common-sense and readability: pieces of meta-knowledge $m_j^{\mathcal{H}^K}$ should have a clear semantic given the application, and the space $\mathcal{H}$ should be of reduced size, e.g., $\mathcal{H} = \{(T,R),(T,\neg R),(\neg T,R),(\neg T,\neg R)\}$.

VI. Practical Instances

Before providing a real world case study in Section VII, we provide three peculiar instances illustrating the approach that are of practical and theoretical interest: the first one uses imprecise pieces of meta-knowledge; the second one is based on Bayesian pieces of meta-knowledge; and the third one explores an interesting link between our approach and the α-conjunctions [40]. Whenever possible, we relate to classical fusion strategies that our approach subsumes.

TABLE I
MASS FUNCTIONS RESULTING FROM THE THREE DIFFERENT ASSUMPTIONS

A	$m_1^{\mathcal{X}}$	$m_2^{\mathcal{X}}$	$m_3^{\mathcal{X}}$	$m[H_1^3]^{\mathcal{X}}$	$m[H_2^3]^{\mathcal{X}}$	$m[H_3^3]^{\mathcal{X}}$
$\emptyset$	0	0	0	0	0	0.36
$\{x_1\}$	0.5	0	0	0	0.06	0.2
$\{x_1, x_2\}$	0	0.2	0	0	0.04	0.04
$\{x_3\}$	0	0	0.6	0	0	0.24
$\{x_1, x_3\}$	0	0	0	0	0.24	0
$\mathcal{X}$	0.5	0.8	0.4	1	0.66	0.16

A. r-Out-of-K Relevant Sources

The collection is defined by $m_j^{\mathcal{H}^K}(H_{K-j+1}^K) = 1$, with H_{K-j+1}^K being the assumption that all the sources are truthful and $r = K-j+1$ out of them are relevant (see (11)). This collection satisfies our need.

Proposition 4: If $m_j^{\mathcal{H}^K}(H_{K-j+1}^K) = 1$, then $m_j^{\mathcal{H}^K} \sqsubseteq_{\mathcal{H}} m_{j+1}^{\mathcal{H}^K}$ for $1 \leq j < K$.

Proof: For each $j \in \{1, \ldots, K-1\}$ and each $\mathbf{h} \in H_{K-j+1}^K$ ($\mathbf{h}$ is a joint state assuming that the K sources are truthful and that $K-j+1$ specific sources are relevant), there exists an assumption $\mathbf{h}' \in H_{K-j}^K$, such that

$$\Gamma_{\mathbf{A}}(\mathbf{h}) \subseteq \Gamma_{\mathbf{A}}\left(\mathbf{h}'\right) \tag{19}$$

for all $\mathbf{A} \subseteq \mathcal{X}^K$. Since $\Gamma_{\mathbf{A}}(H_{K-j+1}^K) = \cup_{\mathbf{h} \in H_{K-j+1}^K} \Gamma_{\mathbf{A}}(\mathbf{h})$ and $\Gamma_{\mathbf{A}}(H_{K-j}^K) = \cup_{\mathbf{h}' \in H_{K-j}^K} \Gamma_{\mathbf{A}}(\mathbf{h}')$ for all $\mathbf{A} \subseteq \mathcal{X}^K$, we have, using (19), $\Gamma_{\mathbf{A}}(H_{K-j+1}^K) \subseteq \Gamma_{\mathbf{A}}(H_{K-j}^K)$ for all $\mathbf{A} \subseteq \mathcal{X}^K$, i.e., H_{K-j+1}^K is as least as meta-specific as H_{K-j}^K. ■

Example 3: Consider the mass functions $m_1^{\mathcal{X}}$, $m_2^{\mathcal{X}}$ and $m_3^{\mathcal{X}}$ on $\mathcal{X} = \{x_1, x_2, x_3\}$ in the left part of Table I. Assume they were received from three distinct sources. Let $\mathbf{m}^{\mathcal{H}^3} = (m_1^{\mathcal{H}^3}, m_2^{\mathcal{H}^3}, m_3^{\mathcal{H}^3}) = (H_3^3, H_2^3, H_1^3)$ be three pieces of meta-knowledge we want to test on these sources. $m_1^{\mathcal{H}^3}$ corresponds to the use of the unnormalized Dempster's rule, while $m_3^{\mathcal{H}^3}$ corresponds to the use of the disjunctive rule. $m_2^{\mathcal{H}^3}$ corresponds to the assumption H_2^3 that the three sources are truthful and that two of them are relevant, but we do not know which ones.

The right part of Table I presents the mass functions on $\mathcal{X}$ resulting from the three different assumptions. We have $\phi_{H_1^3}(\mathbf{m}^{\mathcal{X}}) = 1, \phi_{H_2^3}(\mathbf{m}^{\mathcal{X}}) = 1$ and $\phi_{H_3^3}(\mathbf{m}^{\mathcal{X}}) = 0.4$, hence our approach suggests to use H_2^3 to combine the pieces of information in this example.

Note that the assumption "r-out-of-K" is not b-separable in general. However, this assumption treats all sources in the same way, which seems interesting in absence of meta-knowledge about each individual source.

B. Partially Relevant Sources

Another interesting case is when we consider $\mathcal{H} = \{R, \neg R\}$ (relevant or not) and a vector $\mathbf{p} = (p_1, \ldots, p_K)$ such that $m^{\mathcal{H}}(R^i) = p_i$, $m^{\mathcal{H}}(\neg R^i) = 1 - p_i$ and where $m^{\mathcal{H}^K}$ is obtained by considering the stochastic product of probabilities $p_1, \ldots, p_K$. In such case, the assumption $m^{\mathcal{H}^K}$ amounts to discounting each source $\mathfrak{s}_i$ according to reliability rate $1 - p_i$ and then combining the discounted sources using the conjunctive rule [32]. If we define a set $\mathbf{p}^1, \ldots, \mathbf{p}^M$ of such vectors with $p_i^j \geq p_i^{j+1}$ with the inequality strict for at least one i, we get corresponding pieces of meta-knowledge $m_1^{\mathcal{H}^K}, \ldots, m_M^{\mathcal{H}^K}$ that satisfy the property described by Proposition 5 below, the proof of which requires Lemmas 2 and 3.

Lemma 2 ([6] Proposition 2): The conjunctive rule is monotonic with respect to $\sqsubseteq$, i.e., for all mass functions $m_1^{\mathcal{X}}$ and $m_2^{\mathcal{X}}$ on $\mathcal{X}$ such that $m_1^{\mathcal{X}} \sqsubseteq m_2^{\mathcal{X}}$, we have

$$m_1^{\mathcal{X}} \textcircled{\cap} m_3^{\mathcal{X}} \sqsubseteq m_2^{\mathcal{X}} \textcircled{\cap} m_3^{\mathcal{X}}, \quad \forall m_3^{\mathcal{X}}.$$

Lemma 3: Let $m_k^{\mathcal{X}}$, $m_k^{'\mathcal{X}}$ $(k = 1, \ldots, K)$ be $2K$ mass functions on $\mathcal{X}$ such that $m_k^{\mathcal{X}} \sqsubseteq m_k^{'\mathcal{X}}$. We have

$$\textcircled{\cap}_{k=1}^K m_k^{\mathcal{X}} \sqsubseteq \textcircled{\cap}_{k=1}^K m_k^{'\mathcal{X}}. \tag{20}$$

Proof: Since $m_k^{\mathcal{X}} \sqsubseteq m_k^{'\mathcal{X}}$, $k = 1, \ldots, K$, (20) holds for $K = 1$. Assume now that (20) holds for $K = N$. To prove this lemma, it suffices then to show that (20) holds for $K = N+1$. From Lemma 2, we have

$$\begin{aligned} &\textcircled{\cap}_{k=1}^N m_k^{\mathcal{X}} \sqsubseteq \textcircled{\cap}_{k=1}^N m_k^{'\mathcal{X}} \\ \Rightarrow &\left(\textcircled{\cap}_{k=1}^N m_k^{\mathcal{X}}\right) \textcircled{\cap} m_{N+1}^{\mathcal{X}} \sqsubseteq \left(\textcircled{\cap}_{k=1}^N m_k^{'\mathcal{X}}\right) \textcircled{\cap} m_{N+1}^{\mathcal{X}} \end{aligned} \tag{21}$$

as well as

$$\begin{aligned} &m_{N+1}^{\mathcal{X}} \sqsubseteq m_{N+1}^{'\mathcal{X}} \\ \Rightarrow &\left(\textcircled{\cap}_{k=1}^N m_k^{'\mathcal{X}}\right) \textcircled{\cap} m_{N+1}^{\mathcal{X}} \sqsubseteq \left(\textcircled{\cap}_{k=1}^N m_k^{'\mathcal{X}}\right) \textcircled{\cap} m_{N+1}^{'\mathcal{X}}. \end{aligned} \tag{22}$$

Equations (21) and (22) lead to $\textcircled{\cap}_{k=1}^{N+1} m_k^{\mathcal{X}} \sqsubseteq \textcircled{\cap}_{k=1}^{N+1} m_k^{'\mathcal{X}}$. ■

Proposition 5: Let $m_j^{\mathcal{H}^K}$, $j = 1, \ldots, M$, be the mass functions defined using $\mathbf{p}^1, \ldots, \mathbf{p}^M$. We have $m_j^{\mathcal{H}^K} \sqsubseteq_{\mathcal{H}} m_{j+1}^{\mathcal{H}^K}$, for $1 \leq j < M$.

Proof: Let $m[p_i^j]$ denote the mass function $m_i^{\mathcal{X}}$ on $\mathcal{X}$ discounted according to reliability rate $1 - p_i^j$ (see [35] and [39]). If $p_i^j \geq p_i^{j+1}$, then it is immediate that

$$m[p_i^j] \sqsubseteq m[p_i^{j+1}]. \tag{23}$$

From Lemma 3 and (23), we obtain

$$\textcircled{\cap}_{i=1}^K m\left[p_i^j\right] \sqsubseteq \textcircled{\cap}_{i=1}^K m\left[p_i^{j+1}\right]. \tag{24}$$

As shown in [32], we have

$$\textcircled{\cap}_{i=1}^K m\left[p_i^j\right] = m\left[m_j^{\mathcal{H}^K}\right]^{\mathcal{X}} \tag{25}$$

with $m_j^{\mathcal{H}^K}$ the meta-knowledge obtained by taking the stochastic product of probabilities $p_1^j, \ldots, p_K^j$. Finally, from (24) and (25), we have $m[m_j^{\mathcal{H}^K}]^{\mathcal{X}} \sqsubseteq m[m_{j+1}^{\mathcal{H}^K}]^{\mathcal{X}}$. ■

A useful feature of such $\mathbf{m}^{\mathcal{H}^K}$ is that the pieces of meta-knowledge $m_j^{\mathcal{H}^K}$ are b-separable, and therefore $\phi_{m_j^{\mathcal{H}^K}}(\mathbf{m}^{\mathcal{X}})$ can be computed efficiently using the results of Section IV-B.

In practice, we will have $\mathbf{p}^1 = \mathbf{1}$ (a vector of 1 s), and specifying $\mathbf{p}^2, \ldots, \mathbf{p}^M$ can be done in two main ways: by directly giving them, or by specifying their decreasing rate according to, e.g., expert opinions. For example, the statement "$\mathfrak{s}_\ell$ is less reliable than $\mathfrak{s}_k$" implies that p_ℓ should decrease more quickly than p_k.

Example 4: Consider the same three masses $m_i^{\mathcal{X}}$, $i = 1, \ldots, 3$ as in Example 3, and the fact that according to some

TABLE II
MASS FUNCTIONS RESULTING FROM THREE RELIABILITY VECTORS OF EXAMPLE 4

A	$m_1^{\mathcal{X}}$	$m_2^{\mathcal{X}}$	$m_3^{\mathcal{X}}$	$m[\mathbf{p}^3]^{\mathcal{X}}$	$m[\mathbf{p}^2]^{\mathcal{X}}$	$m[\mathbf{p}^1]^{\mathcal{X}}$
$\emptyset$	0	0	0	0.04	0.18	0.36
$\{x_1\}$	0.5	0	0	0.26	0.26	0.2
$\{x_1, x_2\}$	0	0.2	0	0.08	0.06	0.04
$\{x_3\}$	0	0	0.6	0.08	0.18	0.24
$\mathcal{X}$	0.5	0.8	0.54	1	0.32	0.16

experts, the third source is likely to be far less reliable than the two others. This could be modeled by the successive vectors $\mathbf{p}^1 = (1, 1, 1)$, $\mathbf{p}^2 = (0.8, 0.8, 0.6)$, $\mathbf{p}^3 = (0.6, 0.6, 0.4)$, and so on (until $\mathbf{p}^M = (0, 0, 0)$), to which are associated the meta-knowledges $m_i^{\mathcal{H}^K}$, $i = 1, \ldots, 3$.

Table II summarizes the obtained results (showing only the focal elements). We have $\phi_{\mathbf{p}^1}(\mathbf{m}^{\mathcal{X}}) = 0.4$, $\phi_{\mathbf{p}^2}(\mathbf{m}^{\mathcal{X}}) = 0.64$ and $\phi_{\mathbf{p}^3}(\mathbf{m}^{\mathcal{X}}) = 0.88$, hence if we fix $\tau = 0.8$, then $\mathbf{p}^3$ is selected. Note that the resulting masses are quite different from those obtained in Example 3.

As indicate the next remarks, existing fusion schemes already use this idea.

Remark 5: The sequential discounting approaches proposed in [19][2], [34], and [45], and based, respectively, on the degrees of dissent, falsity, and disagreement, can all be included in the present approach. For instance, if we associate p_i^j with the discounting rate used at step j in Schubert's [34] sequential discounting, then $m[m_j^{\mathcal{H}^K}]^{\mathcal{X}}$ is nothing else but the mass function on $\mathcal{X}$ obtained at step j in Schubert's scheme, showing that it is included in our approach.

Remark 6: Consider K input data o_i in $\mathbb{R}^N$, $i = 1, \ldots, K$ with outputs in space $\mathcal{X}$. To predict the output of a new data o, the evidential k-nearest neighbor method [5] first orders $o_i, i = 1, \ldots, K$ by increasing distance to o. Without loss of generality, let us assume that data are indexed such that $d(o, o_i) \leq d(o, o_{i+1})$ with d some distance. Let $f(d(o, o'))$ be some discounting function that increases with d (we refer to [5] for details). Then, if we define p_i^j as

$$p_i^j = (1 - f(d(o, o_i))) \times \delta(\{i \leq K - j + 1\})$$

with $\delta(\{i \leq K-j+1\})$ the indicator function of $\{i \leq K-j+1\}$ ($= 1$ if $i \leq K-j+1$, 0 else), $m[m_j^{\mathcal{H}^N}]^{\mathcal{X}}$ is the mass function obtained by the evidential k-nn rule with $k = K-j+1$. Using Proposition 5, we have that the lower k is, the less specific is the prediction of the output, and therefore the higher its consistency, that is choosing k in the evidential k-nn rule can be seen as choosing a consistency-specificity trade-off.

C. Similarly Untruthful Sources

Smets [40] introduces a family of combination rules, the α-conjunctions, depending on a parameter $\alpha \in [0, 1]$. This family represents the set of associative, commutative, and linear operators for belief functions with the vacuous mass ($m(\mathcal{X}) = 1$) as neutral element. Let $m_1^{\mathcal{X}}$ and $m_2^{\mathcal{X}}$ be two mass functions on $\mathcal{X}$ and let $m_{1\textcircled{}^{\alpha}2}^{\mathcal{X}}$ denote the mass function resulting from the α-conjunction of $m_1^{\mathcal{X}}$ and $m_2^{\mathcal{X}}$. We have, for all $D \subseteq \mathcal{X}$ [30]

$$m_{1\textcircled{}^{\alpha}2}^{\mathcal{X}}(D) = \sum_{(A\cap B)\cup(\overline{A}\cap\overline{B}\cap C)=D} m_1^{\mathcal{X}}(A)\, m_2^{\mathcal{X}}(B)\, m_{\alpha}^{\mathcal{X}}(C) \quad (26)$$

$$\text{where } m_{\alpha}^{\mathcal{X}}(A) = \alpha^{|\overline{A}|}(1-\alpha)^{|A|}, \quad \forall A \subseteq \mathcal{X}. \quad (27)$$

Similarly to the assumption "r-out-of-K" that allows one to go from the unnormalized Dempster's rule to the disjunctive rule as a function of r, the α-junctions also allows one to move between two rules based on Boolean operators as α decreases: it includes the unnormalized Dempster's rule (for $\alpha = 1$) and the so-called equivalence rule [29] (for $\alpha = 0$), based on the operator $A \underline{\cap} B = (A \cap B) \cup (\overline{A} \cap \overline{B})$ expressing logical equivalence.

Smets derived these rules in [40] from axiomatic requirements, but admitted that they lacked a clear interpretation for $\alpha \in (0, 1)$. Recently, in [29], such an interpretation was provided in terms of truthfulness of the sources: it was shown that they correspond to assuming that either both[3] sources tell the truth or they commit the same lie[4] with some particular mass depending on α. Let $m_{\alpha}^{\mathcal{H}^2}$ denote this meta-knowledge on the sources (we refrain from providing the definition of $m_{\alpha}^{\mathcal{H}^2}$ since it is not needed in this paper; see [29] for the definition). We may then show Proposition 6 concerning $m_{\alpha}^{\mathcal{H}^2}$, using Lemmas 4 and 5.

Lemma 4 ([6] Proposition 2): Disjunctive rule is monotonic with respect to $\sqsubseteq$.

Lemma 5: Let $m_k^{\mathcal{X}}$, $m_k'^{\mathcal{X}}$, $k = 1, \ldots, K$ be $2K$ mass functions on $\mathcal{X}$ such that $m_k^{\mathcal{X}} \sqsubseteq m_k'^{\mathcal{X}}$, $k = 1, \ldots, K$. Let $0 \leq v_k \leq 1$, such that $\sum_{k=1}^K v_k = 1$. We have

$$m_{\Sigma}^{\mathcal{X}} = \sum_{k=1}^K v_k \cdot m_k^{\mathcal{X}} \sqsubseteq \sum_{k=1}^K v_k \cdot m_k'^{\mathcal{X}} = m_{\Sigma}'^{\mathcal{X}}. \quad (28)$$

Proof: Let $\mathcal{F}_{\Sigma} = \cup_{i=1}^K \mathcal{F}_i$ with $\mathcal{F}_i$ the focal sets of $m_i^{\mathcal{X}}$ denote the set of focal elements of all $m_i^{\mathcal{X}}$, $i = 1, \ldots, K$. Similarly, let us denote $\mathcal{F}_{\Sigma}' = \cup_{i=1}^K \mathcal{F}_i'$ the set of focal elements of all $m_i'^{\mathcal{X}}$, $i = 1, \ldots, K$. We also know that for each $k \in \{1, \ldots, K\}$, there is a matrix $W^k = [w_{ij}^k]$ satisfying Definition 1. Now, for any $E_i \in \mathcal{F}_{\Sigma}$, we can write

$$m_{\Sigma}^{\mathcal{X}}(E_i) = \sum_{k=1}^K v_k m_k^{\mathcal{X}}(E_i) = \sum_{k=1}^K v_k \sum_{j=1}^q w_{ij}^k = \sum_{j=1}^q \sum_{k=1}^K v_k w_{ij}^k.$$

The last equality follows by distributing the v_k over $\sum_{j=1}^q w_{ij}^k$. Similarly, for any $F_j \in \mathcal{F}_{\Sigma}'$, we can write

$$m_{\Sigma}'^{\mathcal{X}}(F_i) = \sum_{k=1}^K v_k m_k'^{\mathcal{X}}(F_j) = \sum_{k=1}^K v_k \sum_{i=1}^p w_{ij}^k = \sum_{i=1}^p \sum_{k=1}^K v_k w_{ij}^k.$$

From the above equality, the matrix W^{Σ} with elements w_{ij}^{Σ} satisfies the two first conditions of Definition 1, and we also have $w_{ij}^{\Sigma} > 0$ only if $E_i \subseteq F_j$, since $w_{ij}^k = 0$ when

[2]Klein and Colot's [19] approach amounts basically to an iterative use of the Martin *et al.* approach [26], which is one of the methods (see also, e.g., [24]) to estimate the reliability of a source from its dissimilarity from the other ones.

[3]The interpretation in [29] was only provided for the case where $K = 2$.

[4]In [29], different forms of lack of truthfulness are considered besides the crudest one, that is, telling the opposite of what one knows.

$E_i \not\subseteq F_j$ for any $k \in \{1, \ldots, K\}$. This means that W^{Σ} satisfies Definition 1. ■

Proposition 6: $m_{\alpha}^{\mathcal{H}^2} \sqsubseteq_{\mathcal{H}} m_{\alpha'}^{\mathcal{H}^2}$ holds with $1 \geq \alpha \geq \alpha' \geq 0$.

Proof: Equation (26) can be equivalently rewritten

$$m_{1 \textcircled{\cap}^{\alpha} 2}^{\mathcal{X}}(D) = \sum_{A,B} m_1^{\mathcal{X}}(A)\, m_2^{\mathcal{X}}(B)\, m_{\alpha}^{\mathcal{X}}(D|A,B) \tag{29}$$

with $m_{\alpha}^{\mathcal{X}}(\cdot|A,B)$ the mass function defined $\forall\, D \subseteq \mathcal{X}$ as

$$m_{\alpha}^{\mathcal{X}}(D|A,B) = \sum_{(A \cap B) \cup (\overline{A} \cap \overline{B} \cap C) = D} m_{\alpha}^{\mathcal{X}}(C) \tag{30}$$

where $m_{\alpha}^{\mathcal{X}}$ is the mass function defined by (27).

We may remark that $m_{\alpha}^{\mathcal{X}}(\cdot|A,B)$ is actually the mass function obtained on $\mathcal{X}$ after combining by $\textcircled{\cap}^{\alpha}$ the (certain) testimonies $\mathbf{x} \in A$ and $\mathbf{x} \in B$, i.e., we have

$$m_{\alpha}^{\mathcal{X}}(\cdot|A,B) = A \textcircled{\cap}^{\alpha} B.$$

From the definition (30) of $m_{\alpha}^{\mathcal{X}}(\cdot|A,B)$, we can furthermore remark that $A \textcircled{\cap}^{\alpha} B$ can be obtained by combining $m_{\alpha}^{\mathcal{X}}$ with the information $\mathbf{x} \in \overline{A} \cap \overline{B}$ using the conjunctive rule, and then combining the result of this combination with the information $\mathbf{x} \in A \cap B$ using the disjunctive rule, i.e., we have $A \textcircled{\cap}^{\alpha} B = (m_{\alpha}^{\mathcal{X}} \textcircled{\cap} (\overline{A} \cap \overline{B})) \textcircled{\cup} (A \cap B)$. Now, as shown by Lemma F.1 of [28], we have

$$m_{\alpha}^{\mathcal{X}} = \textcircled{\cap}_{x \in \mathcal{X}} m_{\alpha,x}^{\mathcal{X}} \tag{31}$$

with $m_{\alpha,x}^{\mathcal{X}}$ the mass function defined by $m_{\alpha,x}^{\mathcal{X}}(\{\mathcal{X} \backslash x\}) = \alpha$ and $m_{\alpha,x}^{\mathcal{X}}(\mathcal{X}) = 1 - \alpha$.

Let $1 \geq \alpha \geq \alpha' \geq 0$. We have clearly, $\forall x \in \mathcal{X}$, $m_{\alpha,x}^{\mathcal{X}} \sqsubseteq m_{\alpha',x}^{\mathcal{X}}$. From Lemma 3 and (31), we may then conclude that $m_{\alpha}^{\mathcal{X}} \sqsubseteq m_{\alpha'}^{\mathcal{X}}$. Using Lemmas 2 and 4, we find

$$\left(m_{\alpha}^{\mathcal{X}} \textcircled{\cap} (\overline{A} \cap \overline{B})\right) \textcircled{\cup} (A \cap B) \sqsubseteq (m_{\alpha'}^{\mathcal{X}} \textcircled{\cap} (\overline{A} \cap \overline{B})) \textcircled{\cup} (A \cap B)$$

that is, we have

$$A \textcircled{\cap}^{\alpha} B \sqsubseteq A \textcircled{\cap}^{\alpha'} B, \quad \forall A, B. \tag{32}$$

Using Lemma 5 together with (29) and (32), we can prove that

$$m_1^{\mathcal{X}} \textcircled{\cap}^{\alpha} m_2^{\mathcal{X}} \sqsubseteq m_1^{\mathcal{X}} \textcircled{\cap}^{\alpha'} m_2^{\mathcal{X}}, \quad \forall m_1^{\mathcal{X}}, m_2^{\mathcal{X}} \tag{33}$$

and thus $m_{\alpha}^{\mathcal{H}^2} \sqsubseteq_{\mathcal{H}} m_{\alpha'}^{\mathcal{H}^2}$. ■

Thanks to Proposition 6, one may define a collection $\mathbf{m}^{\mathcal{H}^2}$ based on $m_{\alpha}^{\mathcal{H}^2}$, with α decreasing, that satisfies our proposed source behavior selection approach. An example of such a collection is provided in Example 5.

Example 5: Consider the two mass functions $m_1^{\mathcal{X}}$ and $m_2^{\mathcal{X}}$ on $\mathcal{X} = \{x_1, x_2, x_3\}$ in the left part of Table III. Let $\mathbf{m}^{\mathcal{H}^2} = (m_1^{\mathcal{H}^2}, m_{0.5}^{\mathcal{H}^2}, m_0^{\mathcal{H}^K})$ be three meta-knowledge we want to test on these sources, corresponding respectively to $\alpha = 1$, $\alpha = 0.5$, and $\alpha = 0$. The right part of Table III presents the mass functions on $\mathcal{X}$ resulting from the three different assumptions. We have $\phi_{\alpha=0}(\mathbf{m}^{\mathcal{X}}) = 0.62, \phi_{\alpha=0.5}(\mathbf{m}^{\mathcal{X}}) = 0.59$ and $\phi_{\alpha=1}(\mathbf{m}^{\mathcal{X}}) = 0.56$. Depending on the chosen value for τ, different meta-knowledge can be selected.

The above results provide practical means to exploit α-conjunctions. This may be regarded as important, as despite the fact that α-conjunctions represent an important theoretical family of combination rules, they are seldom exploited in practice. As shown by our results, these rules can be used to design new strategies to deal with conflicting situations in a principled and meaningful manner. Yet, we must note that the appeal of these rules with respect to conflict resolution is limited, as Proposition 6 which means that for all $m_1^{\mathcal{X}}, m_2^{\mathcal{X}}$

$$m_1^{\mathcal{X}} \textcircled{\cap}^{\alpha} m_2^{\mathcal{X}} \sqsubseteq m_1^{\mathcal{X}} \textcircled{\cap}^{\alpha'} m_2^{\mathcal{X}}, \quad 1 \geq \alpha \geq \alpha' \geq 0.$$

does not extend to the general case of $K > 2$ sources, i.e., we do not have in general

$$\textcircled{\cap}_{k=1}^{\alpha\, K} m_k^{\mathcal{X}} \sqsubseteq \textcircled{\cap}_{k=1}^{\alpha'\, K} m_k^{\mathcal{X}}, \quad 1 \geq \alpha \geq \alpha' \geq 0$$

as shown by Example 6 below. Hence, it is not possible to use α-conjunction meta-knowledge to more than two sources in the present framework, unless we relax the assumption of the first step that meta-knowledges $m_j^{\mathcal{H}^K}$ of the collection must form a complete order with respect to $\sqsubseteq_{\mathcal{H}}$. In Section VI-D, we suggest some ways to deal with this case, considering partial rather than complete orders.

TABLE III
MASS FUNCTIONS RESULTING FROM THE THREE DIFFERENT HYPOTHESES

A	$m_1^{\mathcal{X}}$	$m_2^{\mathcal{X}}$	$m[\alpha=0]^{\mathcal{X}}$	$m[\alpha=0.5]^{\mathcal{X}}$	$m[\alpha=1]^{\mathcal{X}}$
$\emptyset$	0	0	0	0.09	0.18
$\{x_1\}$	0.2	0	0.2	0.17	0.14
$\{x_2\}$	0.4	0	0.22	0.25	0.28
$\{x_1, x_2\}$	0	0.3	0.12	0.12	0.12
$\{x_3\}$	0	0.3	0.12	0.12	0.12
$\{x_1, x_3\}$	0	0	0.06	0.03	0
$\{x_2, x_3\}$	0	0	0.12	0.06	0
X	0.4	0.4	0.16	0.12	0.16

Example 6: Consider three mass functions $m_i^{\mathcal{X}}, i \in \{1, 2, 3\}$ on $\mathcal{X} = \{x_1, x_2\}$ defined by $m_1^{\mathcal{X}}(\emptyset) = m_2^{\mathcal{X}}(\emptyset) = 1$, and $m_3^{\mathcal{X}}(\{x_1\}) = 1$. Let $1 > \alpha > \alpha' = 0$. We have $(m_1^{\mathcal{X}} \textcircled{\cap}^{\alpha'} m_2^{\mathcal{X}})(\mathcal{X}) = 1$, and thus $((m_1^{\mathcal{X}} \textcircled{\cap}^{\alpha'} m_2^{\mathcal{X}}) \textcircled{\cap}^{\alpha'} m_3^{\mathcal{X}}) = m_3^{\mathcal{X}}$. Besides, we have

$$\left(m_1^{\mathcal{X}} \textcircled{\cap}^{\alpha} m_2^{\mathcal{X}}\right)(A) = m_{\alpha}^{\mathcal{X}}(A), \quad \forall A \subseteq \mathcal{X}.$$

According to (26), the quantity

$$\left(m_1^{\mathcal{X}} \textcircled{\cap}^{\alpha} m_2^{\mathcal{X}}\right)(\{x_1\})\, m_3^{\mathcal{X}}(\{x_1\}) m_{\alpha}^{\mathcal{X}}(\mathcal{X}) = m_{\alpha}^{\mathcal{X}}(\{x_1\})\, m_{\alpha}^{\mathcal{X}}(\mathcal{X})$$

is transfered to $\mathcal{X}$. Therefore, we have

$$\left((m_1^{\mathcal{X}} \textcircled{\cap}^{\alpha} m_2^{\mathcal{X}}) \textcircled{\cap}^{\alpha} m_3^{\mathcal{X}}\right)(\mathcal{X}) \geq m_{\alpha}^{\mathcal{X}}(\{x_1\})\, m_{\alpha}^{\mathcal{X}}(\mathcal{X}).$$

Since $m_{\alpha}^{\mathcal{X}}(\{x_1\})\, m_{\alpha}^{\mathcal{X}}(\mathcal{X}) > 0$ and

$$\left((m_1^{\mathcal{X}} \textcircled{\cap}^{\alpha'} m_2^{\mathcal{X}}) \textcircled{\cap}^{\alpha'} m_3^{\mathcal{X}}\right)(\mathcal{X}) = m_3^{\mathcal{X}}(\mathcal{X}) = 0$$

$m_1^{\mathcal{X}} \textcircled{\cap}^{\alpha} m_2^{\mathcal{X}} \textcircled{\cap}^{\alpha} m_3^{\mathcal{X}}$ cannot be a specialization of $m_1^{\mathcal{X}} \textcircled{\cap}^{\alpha'} m_2^{\mathcal{X}} \textcircled{\cap}^{\alpha'} m_3^{\mathcal{X}}$.

D. Considering Partially Ordered $\mathbf{m}^{\mathcal{H}^K}$

Although general enough to include many existing fusion strategies, as well as to propose new ones, the approach described so far still relies on assumptions that we may want to question or to relax. In this section, we extend the previous

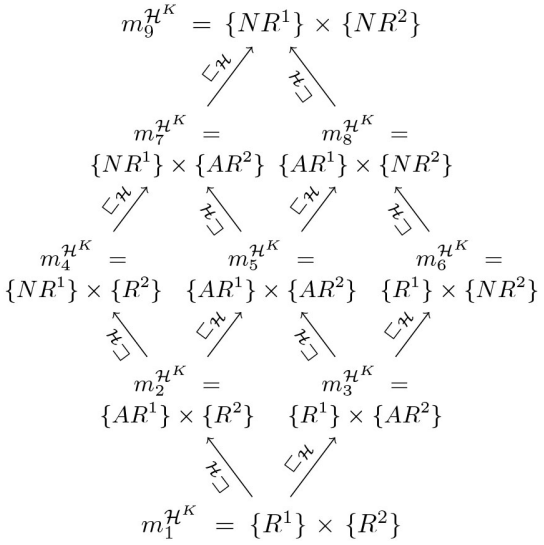


Fig. 1. Illustration of partially ordered collection $\mathbf{m}^{\mathcal{H}^\mathbf{K}}$.

proposal by considering that pieces of meta-knowledge that are of interest may be only partially ordered; hence the mass functions resulting from their applications may also be partially ordered with respect to specialization, and the problem of selecting a meta-knowledge and its associated assumption then becomes more complex.

The approach to select the source behavior proposed in Section V requires to choose a particular collection $\mathbf{m}^{\mathcal{H}^\mathbf{K}} = (m_1^{\mathcal{H}^K}, \ldots, m_M^{\mathcal{H}^K})$ of totally ordered pieces of meta-knowledge. This total order easily allows us getting a unique solution, i.e., the one for which $\phi_{m_j^{\mathcal{H}^K}}(\mathbf{m}^{\mathcal{X}}) \geq \tau$.

However, in some cases, we may want to define a collection $\mathbf{m}^{\mathcal{H}^\mathbf{K}} = (m_1^{\mathcal{H}^K}, \ldots, m_M^{\mathcal{H}^K})$ that is only partially ordered with respect to meta-specificity, that is we may have i, j with $m_i^{\mathcal{H}^K} \not\sqsubset_{\mathcal{H}} m_j^{\mathcal{H}^K}$ and $m_j^{\mathcal{H}^K} \not\sqsubset_{\mathcal{H}} m_i^{\mathcal{H}^K}$. In such a case, we can have $\min\{\phi_{m_j^{\mathcal{H}^K}}(\mathbf{m}^{\mathcal{X}}), \phi_{m_i^{\mathcal{H}^K}}(\mathbf{m}^{\mathcal{X}})\} \geq \tau$ (both assumptions have sufficient consistency) with $m[m_i^{\mathcal{H}^K}]^{\mathcal{X}} \not\sqsubseteq m[m_j^{\mathcal{H}^K}]^{\mathcal{X}}$ and $m[m_j^{\mathcal{H}^K}]^{\mathcal{X}} \not\sqsubseteq m[m_i^{\mathcal{H}^K}]^{\mathcal{X}}$, in which case the consistency-specificity trade-off cannot be used anymore.

As an example, consider two sources $m_1^{\mathcal{X}}$ and $m_2^{\mathcal{X}}$ such that $\mathcal{H} = \{R, AR, NR\}$, where R stands for reliable, AR for approximately reliable (see Example 1) and NR for nonreliable. We can then consider the collection of certain pieces of meta-knowledge where each source can be in one state. Such partial order is illustrated in Fig. 1.

A possible procedure to deal with such partially ordered collection of hypotheses is the following: first retrieve the set $C \subseteq \mathbf{m}^{\mathcal{H}^\mathbf{K}}$ of pieces of meta-knowledge such that

$$C = \left\{ m_i^{\mathcal{H}^K} | \phi_{m_i^{\mathcal{H}^K}}\left(\mathbf{m}^{\mathcal{X}}\right) \geq \tau \wedge \nexists m_j^{\mathcal{H}^K}, j \neq i \right.$$
$$\left. \text{s.t.} \left(\phi_{m_j^{\mathcal{H}^K}}\left(\mathbf{m}^{\mathcal{X}}\right) \geq \tau \wedge m\left[m_j^{\mathcal{H}^K}\right]^{\mathcal{X}} \sqsubset m\left[m_i^{\mathcal{H}^K}\right]^{\mathcal{X}}\right)\right\}.$$

C corresponds to the pieces of meta-knowledge that induce sufficient consistency and that are incomparable with respect to specialization. According to Fig. 1, we could have for example the set $C = \{m_3^{\mathcal{H}^K}, m_4^{\mathcal{H}^K}\}$ but not $M = \{m_3^{\mathcal{H}^K}, m_5^{\mathcal{H}^K}\}$ as $m_3^{\mathcal{H}^K} \sqsubset_{\mathcal{H}} m_5^{\mathcal{H}^K}$. The next step is to select an element from C. We see at least three ways to do so.

First, the most obvious is to select the element in C with the highest degree of $m^{\mathcal{H}}$-consistency. As any element in C has a consistency greater than τ while being minimal with respect to specificity, such selection makes sense.

Second, if the previous strategy is not applicable because at least two elements have very close degree of $m^{\mathcal{H}}$-consistency, one may use some information measures [2], [20] refining the specialization partial ordering (see [1]) to compare the induced specificity of the elements in C.

A last strategy is to consider an additional criteria on top of consistency and specificity, for instance a minimal change principle (see, for example [15], [16], and [25]), favoring the piece of meta-knowledge whose fusion result is the closest to the original testimonies. This change may be measured, e.g., by some distances between belief functions [18].

VII. Application to Case Study

Our case study concerns the analysis of results issued from BEMUSE international exercise [3]. It consisted in comparing results of uncertainty analysis performed on nuclear computer codes to estimate (among other things) peak temperatures during transient conditions of nuclear power plant. Ten participants gave analysis results that were compared to measurements coming from the experiment L2–5 performed on the loss-of-fluid test (LOFT) facility, to test their assessment abilities.

The final results of all these uncertainty analysis are usually difficult to combine and analyze. Moreover, given the lack (and cost) of experimental data together with the complexity of the phenomena involved, there are no reliable means to know the source reliabilities. Also, another aspect of such exercises is that the analysis of the results is as important as the results themselves; it enhances the importance of performing a readable and interpretable merging (if only to be able to explain the results to non-computer scientists).

We focus on the most important of the measured variables, i.e., the second peak clad temperature (PCT2), a critical value of the reactor. The different values are summarized in Table IV, where focal elements are presented as intervals i.e., $[x_i, x_j]$ with $i < j$ represents the set $\{x_i, x_{i+1}, \ldots, x_j\}$, unless it is a singleton. We applied r-out-of-K strategy, with $K = 10$; the results on the contour function (for $r \in \{10, 9, 8\}$) are displayed in Table V. The value $\phi_{H_{10}^{10}} \simeq 0.2$ shows that the sources are globally disagreeing, but the values $\phi_{H_9^{10}} \simeq 0.81$ and $\phi_{H_8^{10}} \simeq 1$ show that assuming nine out of ten sources to be reliable ensures an important agreement, and that assuming eight out of ten sources to be reliable ensures a totally coherent answer. As a conclusion, our method allows delivering results with the three following desirable properties: 1) they are consistent; 2) they are informative (x_4, x_5 are definitely more plausible; and 3) they can be provided through a readable format (for instance, the end-user receiving a message such as "A fully consistent result can be obtained

TABLE IV
FOCAL SETS AND ASSOCIATED MASSES OF BEMUSE PARTICIPANTS FOR PCT2. THE DOMAIN $\mathcal{X} = \{x_1, \ldots, x_6\}$ IS A UNIFORM DISCRETIZATION OF THE INTERVAL [592, 1228), I.E., x_i, $i = 1, \ldots, 6$, CORRESPONDS TO THE INTERVAL $[592 + (i-1) \cdot \frac{1228-592}{6}, 592 + i \cdot \frac{1228-592}{6})$

$\mathfrak{s}_1$		$\mathfrak{s}_2$		$\mathfrak{s}_3$		$\mathfrak{s}_4$		$\mathfrak{s}_5$		$\mathfrak{s}_6$		$\mathfrak{s}_7$		$\mathfrak{s}_8$		$\mathfrak{s}_9$		$\mathfrak{s}_{10}$	
E	m	E	m	E	m	E	m	E	m	E	m	E	m	E	m	E	m	E	m
$[x_4, x_5]$	.25	$[x_5, x_6]$	.5	$[x_4, x_5]$	.25	$[x_4, x_5]$	.25	$[x_4, x_5]$	.25	$[x_4, x_5]$	.5	$[x_3, x_5]$	.25	$\{x_4\}$	.5	$[x_3, x_4]$	.25	$[x_5, x_6]$	.75
$[x_3, x_5]$	.25	$[x_4, x_6]$	.25	$[x_3, x_5]$	.5	$[x_3, x_5]$	.25	$[x_3, x_5]$	.25	$[x_3, x_6]$	.25	$[x_2, x_5]$	.25	$[x_3, x_4]$	.25	$[x_3, x_5]$	.25	$\mathcal{X}$	.25
$[x_2, x_6]$	.25	$\mathcal{X}$	.25	$\mathcal{X}$	.25	$\mathcal{X}$	.5	$[x_2, x_5]$	.25	$\mathcal{X}$	.25	$[x_2, x_6]$	.25	$\mathcal{X}$	.25	$[x_2, x_5]$	.25		
$\mathcal{X}$	.25							$\mathcal{X}$	.25			$\mathcal{X}$	.25			$\mathcal{X}$	.25		

TABLE V
r-OUT-OF-K STRATEGY, WITH $K = 10$ AND $r \in \{10, 9, 8\}$

pl	x_1	x_2	x_3	x_4	x_5	x_6
$r = 8$	0.001	0.008	0.216	1	1	0.031
$r = 9$	0	0.001	0.0511	0.625	0.813	0.004
$r = 10$	0	0	0.005	0.125	0.188	0

by assuming that eight sources out of the ten available are reliable").

VIII. CONCLUSION

When little is known about the sources, that is when one does not have access to observed (training) data or to very accurate expert assessments, then traditional source behavior estimation approaches cannot be used. We have proposed a practical and sensible method to select a source behavior assumption, and thus incidentally an interpretable rule, in such poorly informed environment.

Our approach relies on a general framework for modeling source behavior [32]. It proposes to pick, from a set of sensible behavior assumptions allowed by this latter framwork, the assumption that achieves the best trade-off between specificity and consistency. Of particular interest is the fact that we have extended the notions of inconsistency [8] and specificity to this framework, in order to introduce our approach. It should also be noticed that, up to now, Pichon *et al.* [32] framework has remained mainly theoretical, and this is a first proposal to apply it. We have illustrated our approach by different practical instances, and have provided an illustrative case-study extracted from a real-world problem. In addition, we have related the instances to classical fusion strategies, showing that our framework is quite general and subsumes some existing proposals.

Our method also opens some interesting related questions, which fall outside the scope of the paper.

First, while we allowed behaviors to be dependent, e.g., $\mathfrak{s}_1$ is truthful iff source $\mathfrak{s}_2$ is, information was assumed to be distinct. It would be interesting to address the case of nondistinctness, as in [8]. However, while formally this can be easily done, studying the interplay of meta-knowledge (in)dependency and of source (non)-distinctness is nontrivial; we leave this interesting topic for further researches.

It would also be interesting to take advantage of extra information, if such information is available. One means is to extend the current framework so that it integrates other approaches such as Smets expert system [41] or Mercier *et al.* [27] contextual discounting. Another issue is how to take account of previous experiences or data concerning the sources. Solving such issues would mean bridging the gap between two extremes: no knowledge on the sources (our approach) and a refined knowledge on the sources issued from learning or experts [9], [14], [27].

Finally, the spirit of our approach, which proposes to lower the inconsistency of fusion results by modifying source behavior assumptions, is quite different from techniques redistributing the resulting mass $m(\emptyset)$ to nonempty focal sets [36], [44]. An interesting future work would be to make a general practical or theoretical comparison of the performances of these different approaches, yet it is not entirely clear how such generic and systematic comparisons (i.e., not relying on very specific examples) could be done, as the two kinds of techniques rely on different basic assumptions.

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for their comments that helped to improve this paper. Special thanks also to Dr. M. Florea for his support for some of the programming.

REFERENCES

[1] J. Abellán and S. Gomez, "Measures of divergence on credal sets," *Fuzzy Sets Syst.*, vol. 157, no. 11, pp. 1514–1531, 2006.
[2] J. Abellán and S. Moral, "Difference of entropies as a non-specificity function on credal sets," *Int. J. Gen. Syst.*, vol. 34, pp. 201–214, Jan. 2005.
[3] A. de Crécy *et al.*, "Uncertainty and sensitivity analysis of the loft l2-5 test: Results of the BEMUSE programme," *Nuclear Eng. Design*, vol. 238, no. 12, pp. 3561–3578, 2008.
[4] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Stat.*, vol. 38, no. 1, pp. 325–339, 1967.
[5] T. Denœux, "A *k*-nearest neighbor classification rule based on Dempster–Shafer theory," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 5, pp. 804–813, May 1995.
[6] T. Denœux, "Inner and outer approximation of belief structures using a hierarchical clustering approach," *Int. J. Uncertain. Fuzz. Knowl. Based Syst.*, vol. 9, no. 4, pp. 437–460, 2001.
[7] S. Destercke, P. Buche, and B. Charnomordic, "Evaluating data reliability: An evidential answer with application to a web-enabled data warehouse," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 92–105, Jan. 2013.
[8] S. Destercke and T. Burger, "Toward an axiomatic definition of conflict between belief functions," *IEEE Trans. Cybern*, vol. 43, no. 2, pp. 585–596, Apr. 2013.
[9] S. Destercke and E. Chojnacki, "Methods for the evaluation and synthesis of multiple sources of information applied to nuclear computer codes," *Nuclear Eng. Design*, vol. 238, no. 9, pp. 2484–2493, 2008.
[10] S. Destercke and D. Dubois, "Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory," *Inf. Sci.*, vol. 181, no. 18, pp. 3925–3945, 2011.
[11] D. Dubois and H. Prade, "A set-theoretic view of belief functions: Logical operations and approximations by fuzzy sets," *Int. J. Gen. Syst.*, vol. 12, no. 3, pp. 193–226, 1986.

[12] D. Dubois and H. Prade, "Representation and combination of uncertainty with belief functions and possibility measures," *Comput. Intell.*, vol. 4, no. 3, pp. 244–264, 1988.
[13] D. Dubois, H. Prade, and P. Smets, "A definition of subjective possibility," *Int. J. Approx. Reason.*, vol. 48, no. 2, pp. 352–364, 2008.
[14] Z. Elouedi, E. Lefevre, and D. Mercier, "Discountings of a belief function using a confusion matrix," in *Proc. Int. Conf. Tools Artif. Intell. (ICTAI)*, Arras, France, 2010, pp. 287–294.
[15] J. Grant and A. Hunter, "Measuring consistency gain and information loss in stepwise inconsistency resolution," in *Proc. ECSQARU*, Belfast, U.K., 2011, pp. 362–373.
[16] J. Grant and A. Hunter, "Measuring the good and the bad in inconsistent information," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2011, pp. 2632–2637.
[17] L. Jaulin, "Robust set-membership state estimation; application to underwater robotics," *Automatica*, vol. 45, no. 1, pp. 202–206, 2009.
[18] A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *Int. J. Approx. Reason.*, vol. 53, no. 2, pp. 118–145, 2012.
[19] J. Klein and O. Colot, "Automatic discounting rate computation using a dissent criterion," in *Proc. Workshop Theory Belief Funct.*, Brest, France, 2010, pp. 1–6.
[20] G. J. Klir, *Uncertainty and Information: Foundations of Generalized Information Theory*. Hoboken, NJ, USA: Wiley, 2006.
[21] E. Lefevre, O. Colot, and P. Vannoorenberghe, "Knowledge modeling methods in the framework of evidence theory: An experimental comparison for melanoma detection," in *Proc. Syst. Man Cybern. (SMC)*, vol. 4. Nashville, TN, USA, 2000, pp. 2806–2811.
[22] E. Lefevre, P. Vannoorenberghe, and O. Colot, "About the use of Dempster–Shafer theory for color image segmentation," in *Proc. Color Graph. Image Process. (CGIP)*, 2000, pp. 164–169.
[23] W. Liu, "Analyzing the degree of conflict among belief functions," *Artif. Intell.*, vol. 170, no. 11, pp. 909–924, 2006.
[24] Z.-G. Liu, J. Dezert, Q. Pan, and G. Mercier, "Combination of sources of evidence with different discounting factors based on a new dissimilarity measure," *Decis. Support Syst.*, vol. 52, no. 1, pp. 133–141, 2011.
[25] J. Ma, W. Liu, D. Dubois, and H. Prade, "Bridging Jeffrey's rule, AGM revision and Dempster conditioning in the theory of evidence," *Int. J. Artif. Intell. Tools*, vol. 20, no. 4, pp. 691–720, 2011.
[26] A. Martin, A.-L. Jousselme, and C. Osswald, "Conflict measure for the discounting operation on belief functions," in *Proc. Int. Conf. Inf. Fusion*, Cologne, Germany, 2008, pp. 1003–1010.
[27] D. Mercier, B. Quost, and T. Denoeux, "Refined modeling of sensor reliability in the belief function framework using contextual discounting," *Inf. Fusion*, vol. 9, no. 2, pp. 246–258, 2008.
[28] F. Pichon, "Belief functions: Canonical decompositions and combination rules," Ph.D. dissertation, Université de Technologie de Compiègne, Compiègne, France, 2009.
[29] F. Pichon, "On the α–conjunctions for combining belief functions," in *Proc. 2nd Int. Conf. Belief Funct.*, Compiègne, France, 2012, pp. 285–292.
[30] F. Pichon and T. Denœux, "Interpretation and computation of α-junctions for combining belief functions," in *Proc. Int. Symp. Imprecise Probab. Theor. Appl. (ISIPTA)*, 2009.
[31] F. Pichon, S. Destercke, and T. Burger, "Selecting source behavior in information fusion on the basis of consistency and specificity," in *Proc. ECSQARU*, Utrecht, The Netherlands, 2013, pp. 473–484.
[32] F. Pichon, D. Dubois, and T. Denœux, "Relevance and truthfulness in information correction and fusion," *Int. J. Approx. Reason.*, vol. 53, no. 2, pp. 159–175, 2012.
[33] F. Pichon, C. Labreuche, B. Duqueroie, and T. Delavallade, "Multidimensional approach to reliability evaluation of information sources," in *Information Evaluation*, P. Capet and T. Delavallade, Eds. Hoboken, NJ, USA: Wiley-ISTE, 2014, pp. 129–156.
[34] J. Schubert, "Conflict management in Dempster–Shafer theory using the degree of falsity," *Int. J. Approx. Reason.*, vol. 52, no. 3, pp. 449–460, 2011.
[35] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.
[36] F. Smarandache and J. Dezert, "Information fusion based on new proportional conflict redistribution rules," in *Proc. Int. Conf. Inf. Fusion*, 2005.
[37] F. Smarandache, A. Martin, and C. Osswald, "Contradiction measures and specificity degrees of basic belief assignments," in *Proc. Int. Conf. Inf. Fusion*, Chicago, IL, USA, 2011.
[38] P. Smets, "The concept of distinct evidence," in *Proc. Int. Conf. Inf. Process. Manag. Uncertain. Knowl. Based Syst. (IPMU)*, 1992, pp. 789–794.
[39] P. Smets, "Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem," *Int. J. Approx. Reason.*, vol. 9, no. 1, pp. 1–35, Aug. 1993.
[40] P. Smets, "The α-junctions: Combination operators applicable to belief functions," in *Proc. ECSQARU-FAPR*, Bad Honnef, Germany, 1997, pp. 131–153.
[41] P. Smets, "Analyzing the combination of conflicting belief functions," *Inf. Fusion*, vol. 8, no. 4, pp. 387–412, 2007.
[42] P. Smets and R. Kennes, "The transferable belief model," *Artif. Intell.*, vol. 66, no. 2, pp. 191–243, 1994.
[43] N. Wilson, "Algorithms for Dempster–Shafer theory," in *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, vol. 5. Dordrecht, The Netherlands: Kluwer Academic, 2000, pp. 421–475.
[44] J.-B. Yang and D.-L. Xu, "Evidential reasoning rule for evidence combination," *Artif. Intell.*, vol. 205, pp. 1–29, Dec. 2013.
[45] Y. Yang, D. Han, and C. Han, "Discounted combination of unreliable evidence using degree of disagreement," *Int. J. Approx. Reason.*, vol. 54, no. 8, pp. 1197–1216, 2013.

Authors' photographs and biographies not available at the time of publication.

Annexe G

Several shades of conflict

Version longue soumise à *Fuzzy Sets and Systems*, 2018, d'un article paru dans *Rencontres Francophones sur la Logique Floue et ses Applications, LFA 2016*, La Rochelle, France, 3-4 novembre 2016, pp. 153–159, Cépaduès, 2016.

Several shades of conflict[☆]

Frédéric Pichon[a,*], Anne-Laure Jousselme[b], Nadia Ben Abdallah[b]

[a] *Univ. Artois, EA 3926,*
Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A)
Béthune, F-62400, France
[b] *NATO STO Centre for Maritime Research & Experimentation (CMRE)*
Viale San Bartolomeo, 400, La Spezia, Italy

Abstract

Recently, the measurement of conflict between belief functions has been given an axiomatic foundation by Destercke and Burger, resulting in conflict being measured as the inconsistency of their conjunctive combination and excluding metric distances as suitable candidates for conflict measures. The contribution of this paper is twofold. First, we define a parameterised family of consistency measures which encompasses three existing definitions of consistency of a belief function. An induced family of conflict measures between belief functions is then derived, and each of these conflict measures is shown to satisfy the previously proposed axiomatic. The family of conflict measures defines several shades of conflict as it encompasses the classical measure of conflict, associated to the weakest definition of consistency, as well as two other conflict measures associated, respectively, to a stronger definition of consistency by Yager and to the strongest definition of consistency by Destercke and Burger. The different measures are illustrated on a toy example of vessel destination estimation. Second, we provide a geometric view on consistency measures as well as on the associated conflict measures. In particular, we show that the consistency of a belief function (whatever the considered definition of consistency) is its distance to the belief function representing the state of total inconsistency. This geometric view is then transposed to conflict measures, shedding some new light on the relation

[☆]This paper is an extended and revised version of [1]. Section 4 is an extension of the results of [1] encompassing a parameterised family of conflict measures, itself presented in Section 3.

[*]Corresponding author

Email addresses: frederic.pichon@univ-artois.fr (Frédéric Pichon), anne-laure.jousselme@cmre.nato.int (Anne-Laure Jousselme), nadia.benabdallah@cmre.nato.int (Nadia Ben Abdallah)

between conflict and distances.

Keywords: Belief functions, Evidence theory, Consistency, Inconsistency, Conflict, Norm, Distance.

---

1. Introduction

Intelligent systems need to be able to cope with large amount of information that often displays different dimensions of imperfection. In particular, inconsistent evidence is a challenging problem that may arise when an information source is partially reliable and provides inaccurate or aberrant information. When inconsistent information needs to be merged, a particular attention should be paid to characterising, measuring and understanding the conflict. The conflict can be used as an indicator of a lack of reliability of some source and further used to discount the corresponding piece of information in the overall fusion process or to choose the most appropriate combination rule (see, *e.g.*, [2]). Besides its importance in deriving relevant overall belief, the conflict can be used to support decision making. For instance in maritime security, inconsistency may reveal maritime anomalies such as vessels deviating from normalcy (*e.g.*, "off-route vessels", "too fast vessels") and those possibly spoofing the Automatic Identification System (AIS) signal to hide suspect behaviour [3, 4].

Investigating methods to characterise and measure conflict in information fusion has attracted particular attention in the setting of belief functions [5, 6]. To clarify the semantics of conflict measures for belief functions, axiomatic foundations have been recently provided in [7] and [8] and corresponding measures proposed: Martin's measure combines a degree of inclusion with a distance, whereas Destercke and Burger measure conflict between belief functions as the inconsistency yielded by their conjunctive combination. The uniqueness of these measures with respect to their axiomatic was not addressed though. Besides, although the two sets of axioms of [7] and [8] highly overlap, they differ notably by the identity axiom present in [8] and not in [7]. Whether a conflict measure between two identical belief functions should be null or not has been discussed in several work [9, 10, 8, 7, 11], questioning the appropriateness of the classical conflict measure in belief function theory [6] to quantify conflict in all situations, and led to alternative considerations of conflict. Liu [9] proposes to complement the classical conflict measure with a distance measure, Daniel [10, 12] distinguishes between internal, external and total conflict, Destercke and Burger [7] consider knowledge about source dependence, and Burger [11] investigates the suitability of distance measures and other geometrical objects for conflict measurement.

In this paper, we follow the axiomatic proposed by Destercke and Burger [7] for conflict measures, which encompasses the classical measure of conflict, and revisit and extend some of their results. Specifically, we tackle two research questions: (1) the uniqueness of the proposed measures, and (2) whether geometrical objects may be relevant for this kind of measures of conflict.

Our starting point are the two different definitions (one being stronger than the other one) of the notion of nonconflicting belief functions proposed in [7]. We define a new parameterised family of conflict measures that captures gradual notions ("shades") of conflict. The proposed family satisfies Destercke and Burger's axiomatic approach, and interestingly, subsumes their definitions of nonconflict as extreme cases. In particular, we show that so-called strong nonconflict can be captured by a sound measure which is an alternative to the original contour-based measure proposed in [7]. We also show that this family is compatible with a geometric view and shed some new light on the relation between conflict and distances.

The study in [7] is quite general and considers the whole spectrum from unknown dependence to known dependence between the sources providing the belief functions. We do not tackle these more general situations in this paper, and consider only the case where it can be safely assumed that the sources are independent, although we note that our results can readily be extended to the more general case of known dependence.

This paper is organised as follows. Necessary concepts of belief function theory, as well as Destercke and Burger's axiomatic approach to conflict measurement, are recalled in Section 2. In Section 3, a parameterised family of conflict measures is unveiled and its special cases and properties are discussed. A geometric perspective on the proposed measures is presented in Section 4, before concluding in Section 5.

2. Preliminaries

In this section, necessary concepts of belief function theory are first recalled. Then, the axiomatic approach to conflict measurement of Destercke and Burger [7] is presented.

2.1. Belief function theory

The theory of belief functions is a framework for uncertainty modeling and reasoning. It was originally introduced by Dempster [13, 5] in the context of statistical inference, as a theory of imprecise probabilities. It was extended by Shafer [6] and then by Smets and Kennes [14] to handle subjective uncertainty related to fixed quantities. In this latter interpretation of this framework, called Transferable Belief Model (TBM) and considered in

this paper, the beliefs held by an agent about the actual value $\boldsymbol{x}$ taken by a variable defined on a finite domain $\mathcal{X} = \{x_1, \ldots, x_K\}$ (called *frame*), are modeled by a so-called *mass function* defined as a mapping $m : 2^{\mathcal{X}} \to [0, 1]$ verifying $\sum_{A \subseteq \mathcal{X}} m(A) = 1$ with $m(\emptyset) \geq 0$. The mass $m(A)$ represents the amount of belief allocated to the fact of knowing only that $\boldsymbol{x} \in A$. The set of all mass functions on $\mathcal{X}$ is denoted by $\mathcal{M}$.

Subsets A of $\mathcal{X}$ such that $m(A) > 0$ are called *focal sets* of m, and the set of focal sets of m is denoted by $\mathcal{F}$. In order to simplify some expressions, the number $|\mathcal{F}|$ of focal sets of m is denoted by $\mathfrak{F}$. Several specific cases of mass functions are often distinguished. A mass function m is called:

- *categorical* if $m(A) = 1$ for some $A \subseteq \mathcal{X}$, in which case it defines a classical set and will be denoted by m_A in the following;
- *vacuous* if $m(\mathcal{X}) = 1$ and denoted by $m_{\mathcal{X}}$. It represents total ignorance;
- *empty* if $m(\emptyset) = 1$ and denoted by $m_\emptyset$. It represents total inconsistency in the agent's beliefs about the set of values that are conceivable for $\boldsymbol{x}$ [15];
- *normalised* if $m(\emptyset) = 0$.

Equivalent representations of a mass function m are the *plausibility function* and the *belief function*, defined respectively as, for all $A \subseteq \mathcal{X}$,

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \tag{1}$$

and

$$bel(A) = \sum_{B \subseteq \mathcal{X}, B \neq \emptyset} m(B). \tag{2}$$

That is, $pl(A)$ is the amount of belief consistent with $\boldsymbol{x} \in A$, and $bel(A)$ is the amount of belief implying $\boldsymbol{x} \in A$. The plausibility function restricted to the singletons of $\mathcal{X}$ is the *contour function* $\pi : \mathcal{X} \to [0, 1]$ such that $\pi(x) = pl(\{x\})$, for all $x \in \mathcal{X}$. Due to the one-to-one correspondence between functions m, *bel* and *pl*, any of these functions may be loosely referred to as "belief function" for simplicity – it should nonetheless always be clear from the context what is meant from a technical point of view.

We note that in the imprecise probabilistic interpretation of belief function theory, the belief and plausibility measures *bel* and *pl* represent boundaries on an ill-known probability measure P on $\mathcal{X}$ and m is associated to the

set $\mathcal{P}(m)$ of compatible probability measures defined by $\mathcal{P}(m) = \{P | \forall A \subseteq \mathcal{X}, bel(A) < P(A)\}$.

The informative contents of two mass functions can be compared using the notion of specialisation [16]: a mass function m_1 defined on $\mathcal{X}$ is said to be a *specialisation* of another mass function m_2 defined on $\mathcal{X}$, which is denoted by $m_1 \sqsubseteq m_2$, if and only if there exists a non-negative square matrix $S = [S(A, B)]$, $A, B \in 2^{\mathcal{X}}$, verifying

$$\sum_{A \subseteq \mathcal{X}} S(A, B) = 1, \quad \forall B \subseteq \mathcal{X},$$
$$S(A, B) > 0 \Rightarrow A \subseteq B, \quad A, B \subseteq \mathcal{X},$$
$$m_1(A) = \sum_{B \subseteq \mathcal{X}} S(A, B) m_2(B), \quad \forall A \subseteq \mathcal{X}.$$

The term $S(A, B)$ may be seen as the proportion of the mass $m_2(B)$ which "flows down" to A. The specialisation relation extends the relation of inclusion between classical sets. Let us also recall that we have [16]

$$m_1 \sqsubseteq m_2 \Rightarrow pl_1(A) \leq pl_2(A), \quad \forall A \subseteq \mathcal{X}. \tag{3}$$

Another useful notion is that of *refined* mass function. Recall that a refinement of a space $\mathcal{X}$ to a space $\mathcal{Y}$ is formally defined as a function $\rho : 2^{\mathcal{X}} \rightarrow 2^{\mathcal{Y}}$ such that the set $\{\rho(\{x\}) | x \in \mathcal{X}\}$ is a partition of $\mathcal{Y}$, and $\rho(A) = \bigcup_{x \in A} \rho(\{x\}), \forall A \subseteq \mathcal{X}$. If ρ is a refinement function from $\mathcal{X}$ into $\mathcal{Y}$, the refined mass function $\rho(m_i)$, denoted by $m_{\rho(i)}$ for short, of some mass function m_i is defined such that for any $A \in \mathcal{F}_i$, with $\mathcal{F}_i$ the set of focal sets of m_i, we have $m_i(A) = m_{\rho(i)}(\rho(A))$.

The TBM is appealing because it makes it possible to combine multiple pieces of information about a variable. One of the most classical combination rule of the theory is Dempster's unnormalised rule [13], also known as *conjunctive rule*. Let m_1 and m_2 be two mass functions representing pieces of information about $\boldsymbol{x}$. Their combination by the conjunctive rule, denoted by $\textcircled{\cap}$, results in the mass function $m_{1\textcircled{\cap}2}$ defined by, for all $A \subseteq \mathcal{X}$,

$$m_{1\textcircled{\cap}2}(A) \quad = \quad \sum_{B \cap C = A} m_1(B) m_2(C). \tag{4}$$

This combination is appropriate when m_1 and m_2 have been provided by two independent and reliable sources. Furthermore, the rule $\textcircled{\cap}$ is commutative and associative, and admits the vacuous mass function $m_{\mathcal{X}}$ as neutral element.

In the TBM, conditioning of a mass function m by some $B \subseteq \mathcal{X}$ is equivalent to conjunctive combination of m with the categorical mass function m_B. The result is denoted by $m[B]$, with $m[B] = m \textcircled{\cap} m_B$. The conjunctive rule admits a simple expression using conditioning:

$$m_{1\textcircled{\cap}2}(A) \quad = \quad \sum_{B \subseteq \mathcal{X}} m_1(B) m_2[B](A), \forall A \subseteq \mathcal{X}. \tag{5}$$

2.2. Axiomatic approach to conflict measurement

The axiomatic approach to conflict measurement of [7] relies on the notion of consistency of a mass function, which is recalled first.

2.2.1. Consistency of a mass function

Although a *totally inconsistent* information state is uniquely represented as the empty mass function $m_\emptyset$ [7, 15], *total consistency* of an information state represented by a mass function can be understood differently. Specifically, two different definitions of total consistency are considered in [7]:

Definition 1 (Logical consistency [7])**.** *A mass function m is* logically consistent *iff $\bigcap_{A \in \mathcal{F}} A \neq \emptyset$.*

Definition 2 (Probabilistic consistency [7])**.** *A mass function m is* probabilistically consistent *iff $m(\emptyset) = 0$ (*i.e.*, m is normalised).*

Logical consistency of m corresponds to logical consistency between the focal sets of m, hence its name. Probabilistic consistency takes its name from the fact that $m(\emptyset) = 0$ is equivalent to $\mathcal{P}(m) \neq \emptyset$. We note that a logically consistent mass function is normalised and thus is also probabilistically consistent. Hence, logical consistency is a stronger form of consistency than probabilistic consistency.

Based on these notions of total inconsistency and total consistency, two properties that a measure of consistency ϕ of a mass function should obey have been defined in [7]:

Property 1 (Bounded [7])**.** *A measure of consistency should be bounded,* i.e.*, possess minimal and maximal values.*

Property 2 (Extreme consistent values [7])**.** *A measure of consistency should reach its maximal value if and only if information is totally consistent (according to the considered definition), and its minimal value if and only if information is totally inconsistent.*

Property 2 depends thus on the definition of total consistency considered.

Remarking that $\bigcap_{A \in \mathcal{F}} A \neq \emptyset \Leftrightarrow \exists x \in \mathcal{X}$ s.t. $\pi(x) = 1$ [7, Lemma 1], two consistency measures ϕ_π and ϕ_m from $\mathcal{M}$ to $[0, 1]$ are proposed in [7]:

$$\begin{aligned} \phi_\pi(m) &= \max_{x \in \mathcal{X}} \pi(x), \\ \phi_m(m) &= 1 - m(\emptyset). \end{aligned}$$

The measure $\phi_\pi(m)$ has been proposed originally in [10] to quantify the internal conflict of a belief function. As detailed in [7], measure ϕ_π satisfies Property 1 and Property 2, in the case where total consistency is understood according to Definition 1. Measure ϕ_m on the other hand satisfies Property 1 and Property 2, in the case where total consistency is understood according to Definition 2. Furthermore, we have $\phi_m(m) \geq \phi_\pi(m)$ for any $m \in \mathcal{M}$ [7, Lemma 2].

Measure ϕ_π agrees with the TBM interpretation of belief functions since a mass function is totally logically consistent if and only if it considers that at least one value in $\mathcal{X}$ is totally plausible [7]. Furthermore, it is argued in [7] that ϕ_m is in accordance with the imprecise probabilistic interpretation of belief functions since a mass function is totally probabilistically consistent if and only if its associated set $\mathcal{P}(m)$ is not empty. Yet, as reported in [7, Section VII.A], the correlation between ϕ_m and ϕ_π is high enough in some cases, to consider ϕ_m as a good approximation of ϕ_π.

Despite Destercke and Burger [7] position about measure ϕ_m, we remark that this measure has been justified by Smets [15] as a measure of consistency *in the TBM interpretation* of belief functions, using an argument based on the notion of belief updating. Smets concludes in [15] that ϕ_m quantifies the amount of consistency present in the mass function representing the agent's beliefs about the set of values $\mathcal{X}$ that are conceivable for $\boldsymbol{x}$ by the agent. This can simply be seen by remarking that $pl(\mathcal{X}) = 1 - m(\emptyset) = \phi_m(m)$, that is $\phi_m(m)$ is the amount of belief consistent with the proposition that the actual value $\boldsymbol{x}$ is in $\mathcal{X}$ (we also have $bel(\mathcal{X}) = \phi_m(m)$). In particular, we have $m(\emptyset) = 0 \Leftrightarrow pl(\mathcal{X}) = 1$ so that we could also name the notion of probabilistic consistency as *frame consistency* (in contrast, $\phi_\pi(m)$ is the maximum amount of belief consistent with $\boldsymbol{x} = x$ for some $x \in \mathcal{X}$, and logical consistency corresponds thus to consistency with at least one value in $\mathcal{X}$). This view on ϕ_m is totally in line with the so-called open-world assumption of the TBM, where $m(\emptyset)$ quantifies the belief that $\boldsymbol{x}$ does not lie in $\mathcal{X}$. Overall, ϕ_m and ϕ_π seem thus both relevant in the TBM interpretation of belief functions (ϕ_m being in addition relevant in the imprecise probabilistic interpretation).

Let us finally remark that besides ϕ_π and ϕ_m, another measure of consistency of a mass function has been proposed by Yager in [17]:

$$\phi_Y(m) = \sum_{A \cap B \neq \emptyset} m(A)m(B). \tag{6}$$

This measure reaches its maximum, that is a mass function m is considered totally consistent according to Yager, if and only if the focal sets of m have a pairwise non-empty intersection. This alternative view on total consistency will be called in this paper *pairwise consistency*:

Definition 3 (Pairwise consistency)**.** *A mass function m is* pairwise consistent *iff* $\forall (A, B) \in \mathcal{F}^2, A \cap B \neq \emptyset$.

It is easy to check that measure ϕ_Y satisfies Property 1 and Property 2, in the case where total consistency is understood according to Definition 3.

2.2.2. Conflict between mass functions

As rightfully remarked in [7], two mass functions can be considered as totally conflicting if none of their focal sets intersect. Formally:

Definition 4 (Total conflict [7])**.** *Let m_1 and m_2 be two mass functions with sets of focal sets $\mathcal{F}_1$ and $\mathcal{F}_2$ respectively. Let $\mathcal{D}_i = \cup_{A \in \mathcal{F}_i} A$ denote the disjunction[1] of all focal sets of m_i. m_1 and m_2 are* totally conflicting *when $\mathcal{D}_1 \cap \mathcal{D}_2 = \emptyset$.*

However, there does not seem to be a unique way to define totally nonconflicting mass functions. As a matter of fact, two[2] definitions of nonconflicting mass functions are considered in [7]:

Definition 5 (Strong nonconflict [7])**.** *Two mass functions m_1 and m_2 are said to be* strongly nonconflicting *if and only if*

$$\bigcap_{A \in \{\mathcal{F}_1 \cup \mathcal{F}_2\}} A \neq \emptyset.$$

Definition 6 (nonconflict [7])**.** *m_1 and m_2 are said to be* nonconflicting *if and only if $\forall (A, B)$ such that $A \in \mathcal{F}_1, B \in \mathcal{F}_2$, we have $A \cap B \neq \emptyset$.*

[1] The disjunction of all focal sets is called the *core* of the belief function by Shafer [6].

[2] In [7], three definitions of nonconflicting mass functions are considered (Definitions 4, 5 and 6 in [7]). However, in the case of independent sources, which is assumed in this paper, Definitions 5 and 6 in [7] are equivalent.

Definition 5 requires that all focal sets of m_1 and m_2 have a non-empty intersection, which is stronger than requiring each focal set of m_1 to have a non-empty intersection with each focal set of m_2 as required by Definition 6.

Based on these notions of totally conflicting and totally nonconflicting mass functions, five properties that a measure of conflict κ between two mass functions defined on $\mathcal{X}$ and provided by two independent sources, should satisfy are provided in [7]. The five properties are the following (further details on the motivations of these requirements can be found in [7]):

Property 3 (Extreme conflict values [7]). *$\kappa(m_1, m_2) = 0$ if and only if m_1 and m_2 are nonconflicting (according to the considered definition) and $\kappa(m_1, m_2) = 1$ if and only if m_1 and m_2 are totally conflicting.*

Property 4 (Symmetry [7]). *$\kappa(m_1, m_2) = \kappa(m_2, m_1)$.*

Property 5 (Imprecision monotonicity [7]). *If $m_1 \sqsubseteq m_{1'}$, then $\kappa(m_1, m_2) \geq \kappa(m_{1'}, m_2)$.*

Property 6 (Ignorance is bliss [7]). *If m_2 is vacuous, then $\kappa(m_2, m_1) = 1 - \phi(m_1)$.*

Property 7 (Insensitivity to refinement [7]). *If ρ is a refinement function from $\mathcal{X}$ into $\mathcal{Y}$, then $\kappa(m_1, m_2) = \kappa(m_{\rho(1)}, m_{\rho(2)})$.*

As Property 2 for consistency measures, which depends on the definition of total consistency considered, Property 3 for conflict measures depends on the definition of nonconflict considered (*i.e.*, either Definition 5 or Definition 6). Property 5 states that the conflict should not increase as the imprecision (defined in terms of specialisation) of a mass function increases. Property 6 states that a state of ignorance should not conflict with any other state of information represented by a mass function m_1, whilst accounting for the fact that this state of information m_1 may be partially inconsistent itself – we note however that this property does not specify which kind of consistency (*e.g.*, logical or probabilistic) should consistency measure ϕ satisfy. Property 7 states that the refinement of a mass function should not change its conflict value with any other mass function refined in the same way.

In [7], the authors propose to evaluate the conflict between two mass functions as the inconsistency of their conjunctive combination, where inconsistency is the "inverse" of consistency (*i.e.*, $1 - \phi(\cdot)$. More precisely, they propose two measures of conflict satisfying Properties 3-7, induced by consistency measures ϕ_π and ϕ_m:

$$\begin{aligned} \kappa_\pi(m_1, m_2) &= 1 - \phi_\pi(m_{1 \textcircled{\cap} 2}) = 1 - \max_{x \in \mathcal{X}} \pi_{1 \textcircled{\cap} 2}(x), \\ \kappa_m(m_1, m_2) &= 1 - \phi_m(m_{1 \textcircled{\cap} 2}) = m_{1 \textcircled{\cap} 2}(\emptyset). \end{aligned}$$

Measure κ_π satisfies Properties 3-7, when nonconflict in Prop. 3 is understood in terms of Definition 5, whereas κ_m satisfies these properties when nonconflict is understood in terms of Definition 6. We note that with κ_π, Prop. 6 is satisfied for $\phi = \phi_\pi$, whereas with κ_m this property is satisfied for $\phi = \phi_m$. Let us finally stress that $\kappa_m(m_1, m_2)$ is nothing but the classical measure of conflict $m_{1\textcircled{\cap}2}(\emptyset)$ in the TBM.

3. *N*-consistency and induced conflict

The previous section has reviewed the axiomatic approach to conflict measurement proposed in [7]. It was recalled that the measures κ_π and κ_m satisfy this approach. In particular, measure κ_π, which relies on the measure ϕ_π of logical consistency, is suitable as a measure for strong nonconflict. In this section, it is shown that there exists an alternative measure of logical consistency (*i.e.*, a measure ϕ satisfying Property 2 for Definition 1) and that its induced conflict measure (defined as the inconsistency of the conjunctive combination) is an alternative measure for strong nonconflict (*i.e.*, a measure κ satisfying Property 3 for Definition 5). More generally, a parameterised family of consistency measures capturing different forms of consistency including logical, pairwise and probabilistic consistency is introduced and is shown to induce a family of conflict measures satisfying the axiomatic approach of [7] for a family of definitions of the notion of nonconflicting mass functions subsuming strong nonconflict (Definition 5) and nonconflict (Definition 6).

3.1. N-consistency

Let us start by introducing the following definition:

Definition 7 (N-consistency)**.** *A mass function m is said to be* consistent of order N *(N-*consistent *for short), with $1 \leq N \leq \mathfrak{F}$, iff its focal sets are N-wise consistent,* i.e., *$\forall \mathcal{F}' \subseteq \mathcal{F}$ s.t. $|\mathcal{F}'| = N$, we have*

$$\bigcap_{A \in \mathcal{F}'} A \neq \emptyset.$$

In addition, let ϕ_N denote the measure from $\mathcal{M}$ to $[0, 1]$ such that, for $1 \leq N \leq \mathfrak{F}$ and all $m \in \mathcal{M}$,

$$\phi_N(m) \;:=\; 1 - m^N(\emptyset), \tag{7}$$

where m^N denotes the mass function resulting from the combination of m by itself N times, *i.e.*

$$m^N \;:=\; m^{N-1}\textcircled{\cap}m,$$

with $m^0 := m_{\mathcal{X}}$. Hence, we have $m^1 = m$, $m^2 = m \textcircled{\cap} m$ and more generally $m^N = \bigcirc\!\!\!\!\cap_1^N m$. Note that $m^N(\emptyset)$, *i.e.*, the mass associated to the empty set after N combinations of m by itself, is called *auto-conflict of order N* of mass function m in [18].

We will show in the following that the family ϕ_N measures different "shades" of consistency of m as N varies and in particular encompasses the three forms of consistency already defined in the literature and recalled in Section 2.2.1.

3.2. 1*-consistency*

Let us first remark that probabilistic consistency (Definition 2) of a mass function m is nothing but 1-consistency, since we have

$$m(\emptyset) = 0 \quad \Leftrightarrow \quad A \neq \emptyset, \quad \forall A \in \mathcal{F}.$$

In addition, we have $\phi_m(m) = \phi_1(m)$, for all $m \in \mathcal{M}$, since $\phi_1(m) = 1 - m(\emptyset)$, and thus $\phi_1(m)$ is a measure of probabilistic consistency (or 1-consistency), in the sense that it satisfies Properties 1 and 2 in the case where total consistency is understood according to Definition 2.

Let $\kappa_1(m_1, m_2) := 1 - \phi_1(m_1 \textcircled{\cap} 2)$, for all $m_1, m_2 \in \mathcal{M}$. We have $\kappa_1(m_1, m_2) = \kappa_m(m_1, m_2)$, for all $m_1, m_2 \in \mathcal{M}$, and κ_1 is thus a measure for nonconflict (Definition 6) in the sense that it verifies Properties 3-7, when nonconflict in Prop. 3 is understood in terms of Definition 6,.

It will be useful to remark that the notion of nonconflicting mass functions (Definition 6) can be equivalently presented as follows. Let m_1 and m_2 be any two mass functions and let $\mathcal{F}_{12} := \{A \cap B | A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$. It is clear that: m_1 and m_2 are nonconflicting (Definition 6) $\Leftrightarrow E \neq \emptyset, \forall E \in \mathcal{F}_{12}$. In other words, nonconflict is equivalent to each set in $\mathcal{F}_{12}$ being non-empty, *i.e.*, to the sets in $\mathcal{F}_{12}$ being "1-wise" consistent. Let us note that $\mathcal{F}_{12} = \mathcal{F}_{1\textcircled{\cap}2}$, with $\mathcal{F}_{1\textcircled{\cap}2}$ the set of focal sets of $m_{1\textcircled{\cap}2}$.

3.3. $\mathfrak{F}$*-consistency*

Lemma 1 below shows that the notion of logical consistency of a mass function (Definition 1) is a particular case of that of N-consistency, when N is the number of focal sets of m.

Lemma 1. *m is logically consistent if and only if m is $\mathfrak{F}$-consistent.*

Proof. m is $\mathfrak{F}$-consistent $\Leftrightarrow \forall \mathcal{F}' \subseteq \mathcal{F}$ s.t. $|\mathcal{F}'| = \mathfrak{F}$, we have $\bigcap_{A \in \mathcal{F}'} A \neq \emptyset$. There is only one set $\mathcal{F}'$ s.t. $|\mathcal{F}'| = \mathfrak{F}$, which is $\mathcal{F}' = \mathcal{F}$. Hence m is $\mathfrak{F}$-consistent $\Leftrightarrow \bigcap_{A \in \mathcal{F}} A \neq \emptyset$. □

Moreover, the following results hold for any mass function $m \in \mathcal{M}$:

Lemma 2. *m is $\mathfrak{F}$-consistent if and only if $m^{\mathfrak{F}}(\emptyset) = 0$.*

Proof. Follows from $m^{\mathfrak{F}}(\emptyset) = \sum_{\cap_{i=1}^{\mathfrak{F}} A_i = \emptyset} \prod_{i=1}^{\mathfrak{F}} m(A_i)$. □

Lemma 3. *m is totally inconsistent if and only if $m^{\mathfrak{F}}(\emptyset) = 1$.*

Proof. $\Rightarrow$: $m(\emptyset) = 1$, *i.e.*, m is totally inconsistent, implies clearly $m^{\mathfrak{F}}(\emptyset) = 1$.

$\Leftarrow$: assume this is not true, *i.e.*, $m^{\mathfrak{F}}(\emptyset) = 1$ and $m(\emptyset) \neq 1$. We reach a contradiction since $m(\emptyset) \neq 1$ implies that $\exists A \subseteq \mathcal{X}, A \neq \emptyset$, s.t. $m(A) > 0$, in which case we have $m^{\mathfrak{F}}(A) \geq (m(A))^{\mathfrak{F}} > 0$, and thus $m^{\mathfrak{F}}(\emptyset) \neq 1$. □

Lemmas 1, 2 and 3 suggest to use $\phi_{\mathfrak{F}}(m) = 1 - m^{\mathfrak{F}}(\emptyset)$ as an alternative measure of logical consistency of a mass function m. Indeed, these results show that similarly to ϕ_π, $\phi_{\mathfrak{F}}$ verifies Property 1 and Property 2, in the case where total consistency is understood according to Definition 1 (logical consistency). This measure appears thus as justified as ϕ_π to evaluate logical consistency of a mass function.

Remark 1. *We note that the measures ϕ_π and $\phi_{\mathfrak{F}}$ are not equal. For instance, denoting by*

$$m : (A_1, m(A_1); \ldots; A_{\mathfrak{F}}, m(A_{\mathfrak{F}}))$$

a mass function m with $\mathfrak{F}$ focal elements A_i with associated masses $m(A_i)$, then if m is defined on $\mathcal{X} = \{d_1, d_2, d_3\}$ by

$$m : (\{d_1, d_2\}, 0.8; \{d_3\}, 0.2)\,,$$

we have $\phi_\pi(m) = 0.8 \neq \phi_{\mathfrak{F}}(m) = 0.68$. Further details about the relationships between these measures will be given in Section 3.6.

The measure $\kappa_\pi(m_1, m_2)$ associated to the notion of strong nonconflict (Definition 5) is derived from $\phi_\pi(m_{1\textcircled{\cap}2})$, which is a measure of logical consistency of $m_{1\textcircled{\cap}2}$. Let $\mathfrak{F}_{12}$ denote the cardinality of $\mathcal{F}_{12}$, *i.e.*, $\mathfrak{F}_{12} = |\mathcal{F}_{12}|$. Then $\phi_{\mathfrak{F}_{12}}(m_{1\textcircled{\cap}2})$ is also a measure of logical consistency of $m_{1\textcircled{\cap}2}$. This prompts us to consider the following conflict measure from $\mathcal{M} \times \mathcal{M}$ to $[0, 1]$:

$$\kappa_{\mathfrak{F}_{12}}(m_1, m_2) := 1 - \phi_{\mathfrak{F}_{12}}(m_{1\textcircled{\cap}2}). \tag{8}$$

Proposition 1. *Measure $\kappa_{\mathfrak{F}_{12}}$ satisfies Properties 3-7, when nonconflict in Property 3 is understood in terms of Definition 5 (strong nonconflict).*

Proof. We show each property in turn:

- Property 3: $\kappa_{\mathfrak{F}_{12}}(m_1, m_2) = 0 \Leftrightarrow \phi_{\mathfrak{F}_{12}}(m_{1\textcircled{\cap}2}) = 1 \Leftrightarrow m^{\mathfrak{F}_{12}}_{1\textcircled{\cap}2}(\emptyset) = 0$, which is equivalent using Lemma 2 to $m_{1\textcircled{\cap}2}$ is $\mathfrak{F}_{12}$-consistent, which in turn is equivalent using Lemma 1 to $m_{1\textcircled{\cap}2}$ is logically consistent, *i.e.*, $\cap_{A\in\mathcal{F}_{12}} A \neq \emptyset$ or equivalently $\exists x \in \mathcal{X}$ s.t. $x \in A$, $\forall A \in \mathcal{F}_{12}$. From the definition of $\textcircled{\cap}$, we have $\exists x \in \mathcal{X}$ s.t. $x \in A$, $\forall A \in \mathcal{F}_{12}$ iff m_1 and m_2 are strongly nonconflicting.

  $\kappa_{\mathfrak{F}_{12}}(m_1, m_2) = 1 \Leftrightarrow \phi_{\mathfrak{F}_{12}}(m_{1\textcircled{\cap}2}) = 0 \Leftrightarrow m^{\mathfrak{F}_{12}}_{1\textcircled{\cap}2}(\emptyset) = 1$, which using Lemma 3 is equivalent to $m_{1\textcircled{\cap}2}$ is totally inconsistent, *i.e.*, $m_{1\textcircled{\cap}2}(\emptyset) = 1$. From the definition of $\textcircled{\cap}$, $m_{1\textcircled{\cap}2}(\emptyset) = 1$ iff m_1 and m_2 are totally conflicting.

- Property 4: We have

$$\begin{aligned}
\kappa_{\mathfrak{F}_{12}}(m_1, m_2) &= m^{\mathfrak{F}_{12}}_{1\textcircled{\cap}2}(\emptyset) \\
&= (\textcircled{\bigcap}^{\mathfrak{F}_{12}}_{i=1} m_{1\textcircled{\cap}2})(\emptyset) \\
&= (\textcircled{\bigcap}^{\mathfrak{F}_{12}}_{i=1} (m_1 \textcircled{\cap} m_2))(\emptyset) \\
&= (\textcircled{\bigcap}^{\mathfrak{F}_{12}}_{i=1} (m_2 \textcircled{\cap} m_1))(\emptyset) \\
&= (\textcircled{\bigcap}^{\mathfrak{F}_{12}}_{i=1} m_{2\textcircled{\cap}1})(\emptyset) \\
&= m^{\mathfrak{F}_{12}}_{2\textcircled{\cap}1}(\emptyset) \\
&= \kappa_{\mathfrak{F}_{12}}(m_2, m_1).
\end{aligned}$$

- Property 5: $\textcircled{\cap}$ is monotonic with respect to $\sqsubseteq$ [19, Proposition 2] and thus $m_1 \sqsubseteq m_{1'} \Rightarrow m_1 \textcircled{\cap} m_2 \sqsubseteq m_{1'} \textcircled{\cap} m_2, \forall m_2 \in \mathcal{M}$. Using [2, Lemma 3], we obtain

$$\textcircled{\bigcap}^{\mathfrak{F}_{12}}_{i=1} m_1 \textcircled{\cap} m_2 \;\sqsubseteq\; \textcircled{\bigcap}^{\mathfrak{F}_{12}}_{i=1} m_{1'} \textcircled{\cap} m_2.$$

  That is $m^{\mathfrak{F}_{12}}_{1\textcircled{\cap}2} \sqsubseteq m^{\mathfrak{F}_{12}}_{1'\textcircled{\cap}2}$, which implies using (3) that

$$\begin{aligned}
pl^{\mathfrak{F}_{12}}_{1\textcircled{\cap}2}(\mathcal{X}) &\leq pl^{\mathfrak{F}_{12}}_{1'\textcircled{\cap}2}(\mathcal{X}) \\
\Leftrightarrow 1 - m^{\mathfrak{F}_{12}}_{1\textcircled{\cap}2}(\emptyset) &\leq 1 - m^{\mathfrak{F}_{12}}_{1'\textcircled{\cap}2}(\emptyset) \\
\Leftrightarrow \kappa_{\mathfrak{F}_{12}}(m_1, m_2) &\geq \kappa_{\mathfrak{F}_{12}}(m_{1'}, m_2).
\end{aligned}$$

- Property 6: $m_1(\mathcal{X}) = 1$ means that $m_1 = m_{\mathcal{X}}$. It is thus the neutral

element of $\bigcirc\!\!\!\!\cap$ and we have $\mathfrak{F}_{12} = \mathfrak{F}_2$, from which we obtain

$$\begin{aligned}
\kappa_{\mathfrak{F}_{12}}(m_1, m_2) &= m_{1\bigcirc\!\!\!\!\cap 2}^{\mathfrak{F}_{12}}(\emptyset) \\
&= (\bigcirc\!\!\!\!\bigcap_{i=1}^{\mathfrak{F}_{12}} m_{1\bigcirc\!\!\!\!\cap 2})(\emptyset) \\
&= (\bigcirc\!\!\!\!\bigcap_{i=1}^{\mathfrak{F}_{12}} m_2)(\emptyset) \\
&= m_2^{\mathfrak{F}_{12}}(\emptyset) \\
&= m_2^{\mathfrak{F}_2}(\emptyset) \\
&= 1 - \phi_{\mathfrak{F}_2}(m_2).
\end{aligned}$$

- Property 7: We have

$$\begin{aligned}
\kappa_{\mathfrak{F}_{12}}(m_1, m_2) &= m_{1\bigcirc\!\!\!\!\cap 2}^{\mathfrak{F}_{12}}(\emptyset) \\
&= \sum_{\cap_{i=1}^{\mathfrak{F}_{12}}(A_i \cap B_i) = \emptyset} \prod_{i=1}^{\mathfrak{F}_{12}} m_1(A_i) m_2(B_i),
\end{aligned}$$

and

$$\begin{aligned}
\kappa_{\mathfrak{F}_{12}}(m_{\rho(1)}, m_{\rho(2)}) &= m_{\rho(1)\bigcirc\!\!\!\!\cap \rho(2)}^{\mathfrak{F}_{12}}(\emptyset) \\
&= \sum_{\cap_{i=1}^{\mathfrak{F}_{12}}(\rho(A_i) \cap \rho(B_i)) = \emptyset} \prod_{i=1}^{\mathfrak{F}_{12}} m_{\rho(1)}(\rho(A_i)) m_{\rho(2)}(\rho(B_i)).
\end{aligned}$$

$\forall (A_1, B_1, \ldots, A_{\mathfrak{F}_{12}}, B_{\mathfrak{F}_{12}}) \in \times_{i=1}^{\mathfrak{F}_{12}} \mathcal{F}_1 \times \mathcal{F}_2$, we have:

  - either $\cap_{i=1}^{\mathfrak{F}_{12}}(\rho(A_i) \cap \rho(B_i)) \neq \emptyset$, in which case since $\{\rho(\{x\}) | x \in \mathcal{X}\}$ is a partition of $\mathcal{Y}$, $\exists x \in \mathcal{X}$ s.t. $x \in A_i$ and $x \in B_i$, $i = 1, \ldots, \mathfrak{F}_{12}$, *i.e.*, $\cap_{i=1}^{\mathfrak{F}_{12}}(A_i \cap B_i) \neq \emptyset$;
  - or $\cap_{i=1}^{\mathfrak{F}_{12}}(\rho(A_i) \cap \rho(B_i)) = \emptyset$, in which case $\nexists x \in \mathcal{X}$ s.t. $x \in A_i$ and $x \in B_i$, $i = 1, \ldots, \mathfrak{F}_{12}$, *i.e.*, $\cap_{i=1}^{\mathfrak{F}_{12}}(A_i \cap B_i) = \emptyset$.

Hence, $\forall (A_1, B_1, \ldots, A_{\mathfrak{F}_{12}}, B_{\mathfrak{F}_{12}})$, when the mass $\prod_{i=1}^{\mathfrak{F}_{12}} m_{\rho(1)}(\rho(A_i)) m_{\rho(2)}(\rho(B_i))$ is allocated to $\emptyset$, so does the mass $\prod_{i=1}^{\mathfrak{F}_{12}} m_1(A_i) m_2(B_i)$, and when the mass $\prod_{i=1}^{\mathfrak{F}_{12}} m_{\rho(1)}(\rho(A_i)) m_{\rho(2)}(\rho(B_i))$ is not allocated to $\emptyset$, so does the mass $\prod_{i=1}^{\mathfrak{F}_{12}} m_1(A_i) m_2(B_i)$. Property 7 is then obtained since

$$\prod_{i=1}^{\mathfrak{F}_{12}} m_{\rho(1)}(\rho(A_i)) m_{\rho(2)}(\rho(B_i)) = \prod_{i=1}^{\mathfrak{F}_{12}} m_1(A_i) m_2(B_i),$$

$\forall (A_1, B_1, \ldots, A_{\mathfrak{F}_{12}}, B_{\mathfrak{F}_{12}})$.

□

In the proof of Property 6 of Proposition 1, we remark that if m_1 is vacuous then $\kappa_{\mathfrak{F}_{12}}(m_1, m_2) = 1 - \phi_{\mathfrak{F}_2}(m_2)$, *i.e.*, the consistency of m_2 is evaluated using measure $\phi_{\mathfrak{F}_2}$, which is a measure of the logical consistency of m_2. In other words, we have a similar behaviour to that of κ_π, which satisfies Prop.6 for $\phi = \phi_\pi$ that is another measure of the logical consistency of m_2.

According to Proposition 1, measure $\kappa_{\mathfrak{F}_{12}}(m_1, m_2)$, which is nothing but the auto-conflict of order $\mathfrak{F}_{12}$ of the mass function $m_{1\textcircled{\cap}2}$, constitutes thus a sound alternative to measure $\kappa_\pi(m_1, m_2)$ for evaluating the (strong) conflict between two mass functions m_1 and m_2.

Let us finally note that we have the following equivalence: m_1 and m_2 are strongly nonconflicting (Definition 5) $\Leftrightarrow \bigcap_{E \in \mathcal{F}_{12}} E \neq \emptyset$. In other words, strong nonconflict is equivalent to the intersection of all sets in $\mathcal{F}_{12}$ being non-empty, *i.e.*, to the sets in $\mathcal{F}_{12}$ being $|\mathcal{F}_{12}|$-wise consistent.

3.4. 2-consistency

It is clear that pairwise consistency (Definition 3) of a mass function m is equivalent to m being 2-consistent.

Furthermore, the measure ϕ_Y of consistency proposed by [17] (see Eq. (6)) is obtained with $N = 2$ in (7):

$$\phi_Y(m) = \phi_2(m) = 1 - m^2(\emptyset), \quad \forall m \in \mathcal{M}.$$

Hence, ϕ_2 is a measure of pairwise consistency of m, since it satisfies Properties 1 and 2 in the case where total consistency is understood according to Definition 3.

We established that probabilistic consistency of a mass function m is equivalent to m being 1-consistent and that logical consistency of m is equivalent to m being $\mathfrak{F}$-consistent. Hence, pairwise consistency of m can be situated in between the two extremes that are probabilistic consistency and logical consistency:

- probabilistic consistency is the weakest form – it requires focal sets of m to be "onewise" consistent, *i.e.*, each focal set is non-empty;
- then comes pairwise consistency – focals sets need to be pairwise consistent, *i.e.*, the intersection of any two focal sets must be non-empty;
- and logical consistency is the strongest form – focal sets must be $\mathfrak{F}$-consistent, *i.e.*, the intersection of all focal sets must be non-empty.

Similarly, we have seen that nonconflict (Definition 6) is equivalent to each set in $\mathcal{F}_{12}$ being non-empty, whereas strong nonconflict is equivalent to the intersection of all sets in $\mathcal{F}_{12}$ being non-empty. Considering the above discussion about the three possible definitions of total consistency, nonconflict and strong nonconflict appear thus to be two extreme forms of nonconflict, and Yager's definition of consistency suggests then an alternative definition of nonconflict:

Definition 8. *m_1 and m_2 are said to be* nonconflicting of order 2 *(2-nonconflicting for short) if and only if $\forall E_1 \in \mathcal{F}_{12}, E_2 \in \mathcal{F}_{12}$, we have $E_1 \cap E_2 \neq \emptyset$.*

Let us denote by κ_2 the measure of conflict induced by ϕ_2, *i.e.*, the measure $\kappa_2(m_1, m_2) : \mathcal{M} \times \mathcal{M} \to [0, 1]$ defined by

$$\kappa_2(m_1, m_2) \;:=\; 1 - \phi_2(m_{1 \textcircled{\cap} 2}).$$

Proposition 2. *Measure κ_2 satisfies Properties 3-7, when nonconflict in Property 3 is understood in terms of Definition 8 (2-nonconflict).*

Proof. We only show Property 3 (the other properties can be shown using a similar proof to that of Proposition 1 – for Property 6, we obtain $\kappa_2(m_1, m_2) = 1 - \phi_2(m_2)$):

$\kappa_2(m_1, m_2) = 0 \Leftrightarrow \phi_2(m_{1 \textcircled{\cap} 2}) = 1 \Leftrightarrow m^2_{1 \textcircled{\cap} 2}(\emptyset) = 0 \Leftrightarrow m_{1 \textcircled{\cap} 2}$ is 2-consistent $\Leftrightarrow \forall E_1 \in \mathcal{F}_{12}, E_2 \in \mathcal{F}_{12}$, we have $E_1 \cap E_2 \neq \emptyset$.

$\kappa_2(m_1, m_2) = 1 \Leftrightarrow m^2_{1 \textcircled{\cap} 2}(\emptyset) = 1$, which, using a similar proof to that of Lemma 3, is equivalent to $m_{1 \textcircled{\cap} 2}$ is totally inconsistent, *i.e.*, $m_{1 \textcircled{\cap} 2}(\emptyset) = 1$. From the definition of $\textcircled{\cap}$, $m_{1 \textcircled{\cap} 2}(\emptyset) = 1$ iff m_1 and m_2 are totally conflicting. □

According to Proposition 2, $\kappa_2(m_1, m_2)$ is thus a measure for 2-nonconflict.

3.5. A family of conflict measures

The three definitions of nonconflict encountered so far suggest the following generalisation, obtained by applying the notion of N-consistency to the sets in $\mathcal{F}_{12}$:

Definition 9. *m_1 and m_2 are said to be* nonconflicting of order N *(N-nonconflicting for short), with $1 \leq N \leq \mathfrak{F}_{12}$, if and only if $\forall E_i \in \mathcal{F}_{12}, i = 1, \ldots, N$, we have $\bigcap_{i=1}^{N} E_i \neq \emptyset$.*

N-nonconflict subsumes obviously 2-nonconflict but also nonconflict (Definition 6), which is recovered for $N = 1$, and strong nonconflict (Definition 5), which corresponds to $N = \mathfrak{F}_{12}$.

$\phi(m)$				$\kappa(m_1, m_2) = 1 - \phi(m_1 \textcircled{\cap} 2)$			
Total inconsistency				**Total conflict**			
$m(\emptyset) = 1$				$\mathcal{D}_1 \cap \mathcal{D}_2 = \emptyset$			
Total consistency				**Total nonconflict**			
Probabilistic consistency	$\forall A \in \mathcal{F}, A \neq \emptyset$	$\phi_1 = \phi_m$ [7]	1-consistency	Nonconflict	$\forall E \in \mathcal{F}_{12}, E \neq \emptyset$	$\kappa_1 = \kappa_m$ [7]	1-nonconflict
Pairwise consistency	$\forall (A, B) \in \mathcal{F}^2, A \cap B \neq \emptyset$	ϕ_2 [17]	2-consistency	Pairwise nonconflict	$\forall (E_1, E_2) \in \mathcal{F}_{12}^2, E_1 \cap E_2 \neq \emptyset$	κ_2	2-nonconflict
Logical consistency	$\bigcap_{A \in \mathcal{F}} A \neq \emptyset$	$\phi_{\mathfrak{F}}$ ϕ_π [7]	$\mathfrak{F}$-consistency	Strong nonconflict	$\bigcap_{E \in \mathcal{F}_{12}} E \neq \emptyset$	$\kappa_{\mathfrak{F}_{12}}$ κ_π [7]	$\mathfrak{F}_{12}$-nonconflict

Table 1: Notions of consistency and associated nonconflict.

Let $\kappa_N(m_1, m_2) : \mathcal{M} \times \mathcal{M} \to [0, 1]$ be the measure defined, for $1 \leq N \leq \mathfrak{F}_{12}$, by

$$\kappa_N(m_1, m_2) \;\; := \;\; 1 - \phi_N(m_1 \textcircled{\cap} 2). \tag{9}$$

Proposition 3. *Measure κ_N satisfies Properties 3-7, when nonconflict in Property 3 is understood in terms of Definition 9 (N-nonconflict).*

Proof. The proof is similar to that of Proposition 2. □

Proposition 3 is a generalisation of Propositions 1 and 2. The parameterised family of conflict measures κ_N includes obviously measures $\kappa_{\mathfrak{F}_{12}}$ and κ_2, but also the classical measure of conflict in the TBM, *i.e.*, κ_m, which is equivalent to κ_1.

Remark 2. *If m_1 and m_2 are categorical, then measures $\kappa_N(m_1, m_2)$, $1 \leq N \leq \mathfrak{F}_{12}$, reduce to classical inconsistency between sets.*

Table 1 summarises the notions of consistency and associated conflict together with the associated definitions and measures. The measures defined so far in the literature are mentioned with the proper reference, while the ones proposed in this paper are not assigned any specific reference. Only three values of N are considered for consistency (1, 2 and $\mathfrak{F}$) and for conflict (1, 2 and $\mathfrak{F}_{12}$).

Lemma 4. *$\forall m_1 \in \mathcal{M}, m_2 \in \mathcal{M}$, we have $\kappa_{N-1}(m_1, m_2) \leq \kappa_N(m_1, m_2)$.*

Proof. Follows from the fact that auto-conflict verifies $m^{N-1}(\emptyset) \leq m^N(\emptyset)$ [18]. □

Lemma 4 shows that as N increases, the stronger the conflict measure becomes. In particular, $\kappa_{\mathfrak{F}_{12}}$ is the strongest conflict measure, while κ_1 is

the weakest. An equivalent relationship can straightforwardly be derived between consistency measures ϕ_N.

The following Section 3.6 further studies the relative behaviours of the measures considered so far.

3.6. Correlation analysis

In this section, we conduct some correlation analysis of the different measures, the new ϕ_N family for $N = 1, \ldots, \mathfrak{F}$, the existing logical consistency measure ϕ_π and their counter-part conflict measures, to illustrate and better grasp their relative behaviours. In particular, the experiments show that the measures capture different shades of consistency (and conflict). We use the Spearman coefficient which is a rank correlation coefficient quantifying how much two measures are monotonically correlated. It reaches the value 1 if the two measures are linked through some increasing monotonic function.

3.6.1. Conflict measures κ_N vs κ_π

We first study how the family of conflict measures κ_N behaves relatively to the existing (strong) conflict measure κ_π, using a similar experiment to that in [7, Section VII.A].

For a given size of the frame of discernment $\mathcal{X}$, we have drawn randomly (following Algorithm 3 of [20]) 5000 couples of normalised mass functions (m_1, m_2) having $2^{|\mathcal{X}|} - 1$ focal sets . Table 2 shows the Spearman correlation between κ_π and κ_N for different values of N and of $\mathcal{X}$. The first line of Table 2 corresponds to the first line of Table II of [7, Section VII.A] with very similar values, while the other lines differ in the sense that our N does not correspond to the number of sources combined as in [7]. We can observe that for $|\mathcal{X}| = 2$ and $|\mathcal{X}| = 3$, for which we have $\mathfrak{F}_{12} = 4$ and $\mathfrak{F}_{12} = 8$ respectively, the correlation between $\kappa_{\mathfrak{F}_{12}}$ and κ_π is really high (0.9895 and 0.9941, respectively).

In the experiment of [7, Section VII.A], it was observed that the correlation between κ_1 and κ_π is relatively high, and that this correlation increases as the cardinality of the frame decreases and as the number of sources to be combined increases. In our experiment, a similar observation can be made: the correlation between κ_N and κ_π increases as $|\mathcal{X}|$ decreases and as N increases, for the values of $|\mathcal{X}|$ and N considered here. This latter result can be better observed in Figure 1, which displays the scatter plots for the specific case of $|\mathcal{X}| = 3$, whose corresponding correlation coefficients can be found in column "$|\mathcal{X}| = 3$" of Table 2.

		Size of the frame of discernment $\lvert\mathcal{X}\rvert$					
		2	3	4	5	6	7
N	1	0.8467	0.7747	0.7268	0.6624	0.6116	0.5551
	2	0.9535	0.8960	0.8451	0.7772	0.7187	0.6616
	3	0.9796	0.9452	0.9024	0.8395	0.7713	0.7064
	4	0.9895	0.9689	0.9373	0.8853	0.8168	0.7477
	5	-	0.9811	0.9585	0.9175	0.8546	0.7851
	6	-	0.9877	0.9716	0.9395	0.8847	0.8178
	7	-	0.9916	0.9798	0.9548	0.9081	0.8457
	8	-	0.9941	0.9852	0.9655	0.9260	0.8694

Table 2: Spearman correlation between κ_π and κ_N according to $|\mathcal{X}|$ and $N \leq 8$.

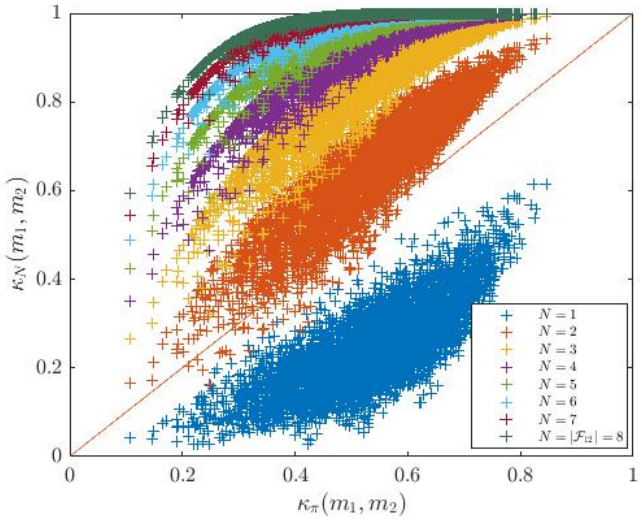


Figure 1: Scatter plots of measures κ_π and κ_N obtained from 5000 couples of randomly generated mass functions over $\mathcal{X}$ such that $|\mathcal{X}| = 3$.

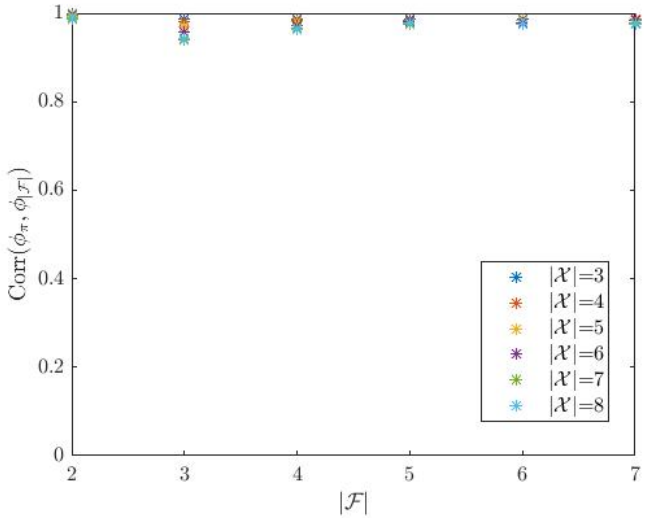


Figure 2: Spearman correlation between ϕ_π and $\phi_{\mathfrak{F}}$ for 5000 randomly generated mass functions.

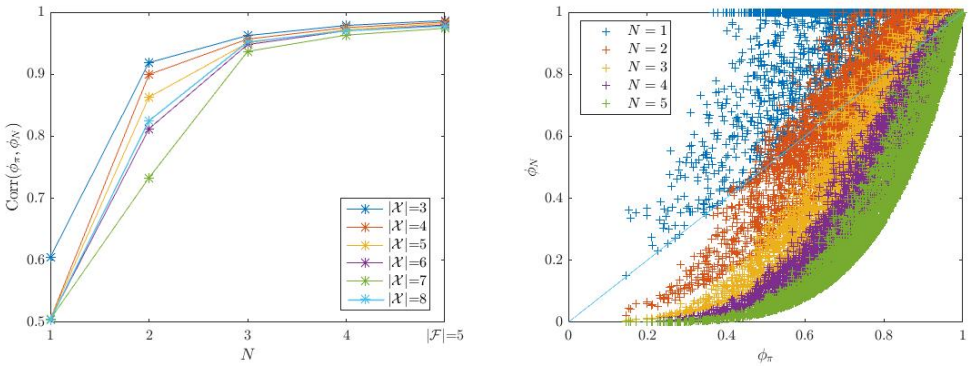
Our experiment shows that while κ_1 and κ_π are quite correlated as both are measures of conflict, the family of measures κ_N captures different shades of conflict and the correlation with κ_π increases with N. However, it shows that the measures $\kappa_{\mathfrak{F}_{12}}$ and κ_π for strong nonconflict do not have a maximal correlation and hence are different.

3.6.2. Consistency measures ϕ_N

Here we study the correlation between ϕ_N and ϕ_π on the one hand and within the family ϕ_N as N varies on the other hand, by analysing the impact of the number of focal elements $\mathfrak{F}$ as well as the size of the frame of discernment $|\mathcal{X}|$. We thus prefer Algorithm 7 to Algorithm 3 of [20] which allows to control the number of focal elements of the mass function generated.

As one can see from Figure 2, the correlation between $\phi_{\mathfrak{F}}$ and ϕ_π, which are two measures of logical consistency, is high (> 0.9) whatever the size of the frame of discernment and the number of focal elements considered here. Furthermore, if we consider a particular value of $\mathfrak{F}$, for instance $\mathfrak{F} = 5$, it appears that ϕ_π is more correlated to $\phi_{\mathfrak{F}}$ than to any ϕ_N, for all $N < \mathfrak{F}$, as shown in Figure 3. Figure 3(a) displays how the correlation between ϕ_π and ϕ_N increases with N for different sizes of $\mathcal{X}$, while Figure 3(b) is the corresponding scatter plot for the specific case of $|\mathcal{X}| = 4$.

We conclude this correlation analysis section with the behaviour of the ϕ_N measures between themselves. Table 3 reports the pairwise correlation



(a) Spearman correlation between ϕ_π and ϕ_N, N varying, for several $|\mathcal{X}|$.

(b) Scatter plots of ϕ_π and ϕ_N, $|\mathcal{X}| = 4$.

Figure 3: Correlation between ϕ_π and ϕ_N, 5000 randomly generated mass functions over $\mathcal{X}$ with $\mathfrak{F} = 5$ focal sets.

coefficients between ϕ_N and ϕ_M, for $1 \leq N, M \leq \mathfrak{F} = 5$ and $|\mathcal{X}| = 4$. We

		N				
		1	2	3	4	5
M	1	1.0000	0.6459	0.5869	0.5591	0.5438
	2	-	1.0000	0.9829	0.9653	0.9520
	3	-	-	1.0000	0.9965	0.9910
	4	-	-	-	1.0000	0.9987
	5	-	-	-	-	1.0000

Table 3: Correlation between ϕ_N and ϕ_M for $1 \leq N, M \leq \mathfrak{F} = 5$ and $|\mathcal{X}| = 4$.

observe that the correlation between successive pairs of measures (ϕ_N, ϕ_{N+1}) increases as N increases. Additionally, we observe that ϕ_N is less correlated to ϕ_{N+2} than to ϕ_{N+1}.

The scatter plots of Figure 4 illustrate the correlation coefficients corresponding to the second diagonal of Table 3. For a given m, the measures ϕ_N are less correlated for low values of N.

We illustrated that the family of consistency measures ϕ_N allows to capture different shades of consistency, and thus that the derived family of conflict measures κ_N allows to capture different shades of conflict between two mass functions m_1 and m_2. In practice, the family of measures κ_N offers alternatives to evaluate the validity of the result of the combination by the conjunctive rule $\textcircled{\cap}$, as it will be illustrated in Section 3.7.

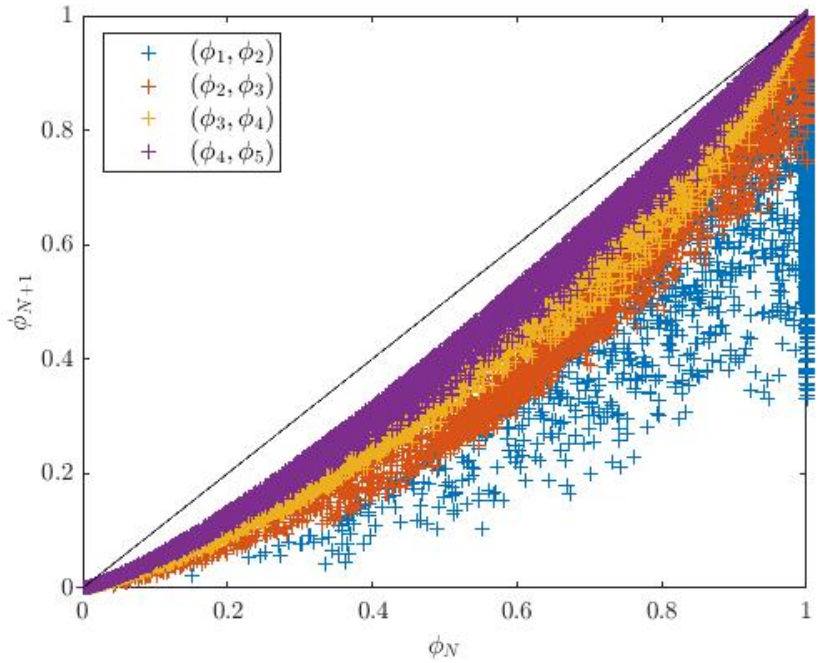


Figure 4: Scatter plots of ϕ_N and ϕ_{N+1} for 5000 random mass functions such that $|\mathcal{X}| = 4$ and $\mathfrak{F} = 5$.

3.7. Toy example

We illustrate some of the consistency and conflict measures, as well as some of their relationships, described previously, on a problem of vessel's destination estimation and associated anomaly detection, in which we would like to estimate a vessel's destination while detecting any inconsistency given different sources of information.

Let $\mathcal{X} = \{d_1, d_2, d_3, d_4\} = \{\text{Savona}, \text{Genova}, \text{La Spezia}, \text{Livorno}\}$ denote the set of possible destinations for the vessel.

We consider a source S_1, as an algorithm analysing the kinematic features of the vessel, compared to some pre-extracted patterns-of-life as maritime routes [21]. Once computed, the routes provide contextual information describing some normalcy against which the current positions of vessels can be compared. In particular, assigning a vessel to a route provides information about its final destination [3]. However, because the routes share some portions of the sea, some indeterminacy may occur and a subset of possible destinations may rather be deduced. Let m_1 be the mass function that encodes the information provided by the source S_1:

$$m_1 : (\{d_1, d_2, d_3\}, 0.6; \{d_1, d_2, d_4\}, 0.3; \{d_3, d_4\}, 0.1).$$

We have:

$$m_1^2: \quad (\{d_1,d_2,d_3\},0.36;\{d_1,d_2,d_4\},0.09;\{d_3,d_4\},0.01;\\ \{d_1,d_2\},0.36;\{d_4\},0.06;\{d_3\},0.12).$$

and

$$m_1^3: \quad (\{d_1,d_2,d_3\},0.216;\{d_1,d_2,d_4\},0.027;\{d_3,d_4\},0.001;\\ \{d_1,d_2\},0.486;\{d_4\},0.036;\{d_3\},0.126;\emptyset,0.108).$$

The mass function m_1 is 1-consistent (probabilistically consistent) and 2-consistent (pairwise consistent), but not 3-consistent (*i.e.*, not logical consistent since $\mathfrak{F}_1 = 3$). Indeed:

$$\begin{cases} \phi_1(m_1) = \phi_2(m_1) = 1, \\ \phi_3(m_1) = 1 - m_1^3(\emptyset) = 0.892. \end{cases} \tag{10}$$

Consider now a second source S_2 which provides a piece of evidence encoded by the following mass function, corresponding to the subjective assessment of a human operator, knowing the maritime traffic of the area and excluding Genova (d_2) as a possible destination for that vessel, but considering as more probable the destination of Savona (d_1) or Livorno (d_4) than that of La Spezia (d_3):

$$m_2 : (\{d_1,d_3,d_4\},0.7;\{d_1,d_4\},0.3)$$

The conjunctive combination of m_1 and m_2 has $\mathfrak{F}_{12} = 5$ focal sets:

$$m_{1\textcircled{\cap}2} : (\{d_1,d_3\},0.42;\{d_1,d_4\},0.3;\{d_3,d_4\},0.07;\{d_1\},0.18;\{d_4\},0.03)$$

The sources S_1 and S_2 are 1-nonconflicting since $\kappa_1(m_1,m_2) = 0$. However, they are not nonconflicting of order 2 (and consequently higher orders) since some focal sets of $m_{1\textcircled{\cap}2}$ are pairwise disjoint (*e.g*, $\{d_3,d_4\} \cap \{d_1\} = \emptyset$):

$$\begin{cases} \kappa_1(m_1,m_2) = 1 - \phi_1(m_{1\textcircled{\cap}2}) = m^1_{1\textcircled{\cap}2}(\emptyset) = m_{1\textcircled{\cap}2}(\emptyset) = 0, \\ \kappa_2(m_1,m_2) = 1 - \phi_2(m_{1\textcircled{\cap}2}) = m^2_{1\textcircled{\cap}2}(\emptyset) = 0.0612, \\ \kappa_3(m_1,m_2) = 1 - \phi_3(m_{1\textcircled{\cap}2}) = m^3_{1\textcircled{\cap}2}(\emptyset) = 0.1908, \\ \kappa_4(m_1,m_2) = 1 - \phi_4(m_{1\textcircled{\cap}2}) = m^4_{1\textcircled{\cap}2}(\emptyset) = 0.2999, \\ \kappa_5(m_1,m_2) = \kappa_{\mathfrak{F}_{12}}(m_1,m_2) = 1 - \phi_{\mathfrak{F}_{12}}(m_{1\textcircled{\cap}2}) = m^{\mathfrak{F}_{12}}_{1\textcircled{\cap}2}(\emptyset) = 0.3865. \end{cases} \tag{11}$$

Note that, as stated in Lemma 4, the conflict does not increase with N.

The (1-)nonconflict of the sources is a valid justification for considering the result of their conjunctive combination for further reasoning [22] and based on some standard decision procedure (such as the pignistic transform [23]), d_1=Savona would be a sensible estimated destination.

We note also that while κ_1 is null, κ_2 is very small (0.0612), and $\kappa_{\mathfrak{F}_{12}}$, the measure for strong nonconflict, is quite higher (0.3865). A criterion for combination based on $\kappa_{\mathfrak{F}_{12}}$ instead of κ_1 would lead then possibly to the decision to not combine the sources.

4. A norm-based view on conflict

Section 3 has brought to light the existence of a family of conflict measures, respecting Destercke and Burger's axiomatics of conflict quantification and encompassing Yager's definition of consistency.

As argued in [10, 7] and further developed in [11], distances between belief functions [24] are questionable to measure their conflict; for instance, they do not satisfy Properties 5 (imprecision monotonicity) and 7 (insensitivity to refinement). Still, this does not mean that geometrical objects are not relevant to conflict quantification as will be shown in this section.

After recalling in Section 4.1 necessary material on norms and distances, we lay bare in Section 4.2 a pseudo-norm based view on consistency measures, which leads us to investigate the relationship between the consistency of a mass function and its distance to the state of total inconsistency. Then, this geometric view with respect to consistency measures is carried over to conflict measures in Section 4.3.

4.1. Norms and distances

Any vector $\boldsymbol{v}$ of the Cartesian space $\mathbb{R}^N$ spanned by the set of vectors $\{\boldsymbol{e}_i, 1 \leq i \leq N\}$, can be written as $\boldsymbol{v} = \sum_{1 \leq i \leq N} v_i \boldsymbol{e}_i$, with $v_i \in \mathbb{R}$ the coordinate of $\boldsymbol{v}$ along dimension $\boldsymbol{e}_i$.

By definition, a *norm* is a function $n : \mathbb{R}^N \to [0, \infty)$ that satisfies the following properties:

(n.1) Definiteness: $n(\boldsymbol{v}) = 0 \Leftrightarrow \boldsymbol{v} = \boldsymbol{0}$ (*i.e.*, $v_i = 0$, $i = 1, \ldots, N$),
(n.2) Homogeneity: $n(b\boldsymbol{v}) = |b| n(\boldsymbol{v})$, $\forall b \in \mathbb{R}$, $\forall \boldsymbol{v} \in \mathbb{R}^N$,
(n.3) Subadditivity (triangle inequality): $n(\boldsymbol{v}_1 + \boldsymbol{v}_2) \leq n(\boldsymbol{v}_1) + n(\boldsymbol{v}_2)$, $\forall \boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^N$.

A *pseudo-norm* corresponds to changing condition (n.1) in the definition of the norm to $n(\boldsymbol{v}) = 0 \Leftarrow \boldsymbol{v} = \boldsymbol{0}$ [25].

The function

$$n^p_{\boldsymbol{w}}(\boldsymbol{v}) := (\sum_i w_i |v_i|^p)^{1/p}, \tag{12}$$

with $p \geq 1$ and where, for $i = 1, \ldots, N$, $w_i > 0$ and finite, is an example of a norm; if we allow some w_i to equal zero, then $n^p_{\boldsymbol{w}}$ is a pseudo-norm [25]. The weights w_i actually distort the space of reference by increasing or reducing the importance of some dimensions. The (pseudo-)norm $n^1_{\boldsymbol{w}}$ will be more simply denoted by $n_{\boldsymbol{w}}$ and we have $n_{\boldsymbol{w}}(\boldsymbol{v}) = \sum_i w_i |v_i|$. Note also that $n^\infty_{\boldsymbol{w}}(\boldsymbol{v}) = \max \boldsymbol{v}$, $\forall \boldsymbol{w}$ such that $\sum_i w_i = 1$ [26]. As a consequence n^∞ will denote any norm $n^\infty_{\boldsymbol{w}}$ with $\boldsymbol{w}$ such that $\sum_i w_i = 1$.

A metric or distance function (called simply *distance*) is a function $d : \mathbb{R}^N \times \mathbb{R}^N \to [0, \infty)$ that satisfies the following properties:

(d.1) Definiteness: $d(\boldsymbol{v}_1, \boldsymbol{v}_2) = 0 \Leftrightarrow \boldsymbol{v}_1 = \boldsymbol{v}_2$,
(d.2) Symmetry: $d(\boldsymbol{v}_1, \boldsymbol{v}_2) = d(\boldsymbol{v}_2, \boldsymbol{v}_1)$, $\forall \boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^N$,
(d.3) Triangle inequality: $d(\boldsymbol{v}_1, \boldsymbol{v}_2) \leq d(\boldsymbol{v}_1, \boldsymbol{v}_3) + d(\boldsymbol{v}_2, \boldsymbol{v}_3)$, $\forall \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \in \mathbb{R}^N$.

A *pseudo-distance* corresponds to changing condition (d.1) in the definition of the distance to $d(\boldsymbol{v}_1, \boldsymbol{v}_2) = 0 \Leftarrow \boldsymbol{v}_1 = \boldsymbol{v}_2$ [25]. If n is a norm (resp. pseudo-norm) then $d_n(\boldsymbol{v}_1, \boldsymbol{v}_2) := n(\boldsymbol{v}_1 - \boldsymbol{v}_2)$ is a distance (resp. pseudo-distance) and is called the induced distance (resp. pseudo-distance) from n. The distance (resp. pseudo-distance) induced by the norm (resp. pseudo-norm) $n^p_{\boldsymbol{w}}$ will be denoted by $d^p_{\boldsymbol{w}}$; $d^1_{\boldsymbol{w}}$ will be more simply denoted by $d_{\boldsymbol{w}}$ and the distance induced by n^∞ will be denoted by d^∞.

4.2. Consistency as distance to total inconsistency

Let $\mathcal{E}_\mathcal{X}$ denote the Cartesian space $\mathbb{R}^{2^K}$ spanned by the set of vectors $\{\boldsymbol{e}_A, A \subseteq \mathcal{X}\}$. Any vector $\boldsymbol{v}$ of $\mathcal{E}_\mathcal{X}$ can then be written as $\boldsymbol{v} = \sum_{A \subseteq \mathcal{X}} v_A \boldsymbol{e}_A$, with $v_A \in \mathbb{R}$ the coordinate of $\boldsymbol{v}$ along dimension $\boldsymbol{e}_A$. A mass function m may then be represented as the vector $\boldsymbol{m}$ of $\mathcal{E}_\mathcal{X}$ such that $v_A = m(A)$. Similarly, a plausibility function pl may be represented by the vector $\boldsymbol{pl} = \sum_{A \subseteq \mathcal{X}} pl(A) \boldsymbol{e}_A$, with its plausibility values $pl(A)$ as coordinates of $\boldsymbol{pl}$. The vector associated to m_A is denoted by $\boldsymbol{m}_A$. Special cases are the empty mass function $\boldsymbol{m}_\emptyset$ and the vacuous mass function $\boldsymbol{m}_\mathcal{X}$.

Let $k : 2^\mathcal{X} \to [0, 1]$ be the function defined by $k(A) := m[A](\emptyset)$, for all $A \subseteq \mathcal{X}$. Function k may be represented by the vector $\boldsymbol{k}$ such that $k_A = k(A)$. k_A measures the weight allocated to the empty set after conditioning m by A, meaning that it represents how much m is inconsistent with A only. Lemma 5 below links function k with other well-known quantities.

Lemma 5. *For all $A \subseteq \mathcal{X}$, we have*

$$\begin{aligned} k(A) &= \kappa_1(m, m_A) && (13)\\ &= 1 - pl(A). && (14) \end{aligned}$$

Proof. By definition

$$m[A](\emptyset) = (m_A \textcircled{\cap} m)(\emptyset), \quad \forall A \subseteq \mathcal{X}.$$

We have $(m_A \textcircled{\cap} m)(\emptyset) = \kappa_1(m, m_A)$ and

$$\begin{aligned} (m_A \textcircled{\cap} m)(\emptyset) &= \sum_{B \cap A = \emptyset} m(B) && (15)\\ &= 1 - \sum_{B \cap A \neq \emptyset} m(B)\\ &= 1 - pl(A). \end{aligned}$$

□

From (15), $k(A)$ is the amount of belief inconsistent with $\boldsymbol{x} \in A$. From (13), $k(A)$ can also be interpreted as the amount of conflict, according to κ_1, between m and A. Hereafter, we will refer to k as the *inconsistency* function associated to m. k is in one-to-one correspondence with m.

We are ready to show the first result of this section:

Proposition 4. *For all $m \in \mathcal{M}$ and $1 \leq N \leq \mathfrak{F}$, we have*

$$\begin{aligned} \phi_N(m) &= n_{\boldsymbol{m}^{N-1}}(\boldsymbol{pl}), && (16)\\ m^N(\emptyset) &= n_{\boldsymbol{m}^{N-1}}(\boldsymbol{k}). && (17) \end{aligned}$$

Proof. We have $m^N = m^{N-1} \textcircled{\cap} m$, which using (5) yields

$$\begin{aligned} m^N(\emptyset) &= \sum_{A \subseteq \mathcal{X}} m^{N-1}(A) m[A](\emptyset)\\ &= \sum_{A \subseteq \mathcal{X}} m^{N-1}(A) k(A)\\ &= n_{\boldsymbol{m}^{N-1}}(\boldsymbol{k}), \end{aligned}$$

and

$$\begin{aligned}
\phi_N(m) &= 1 - m^N(\emptyset) \\
&= 1 - \sum_{A \subseteq \mathcal{X}} m^{N-1}(A) k(A) \\
&= 1 - \sum_{A \subseteq \mathcal{X}} m^{N-1}(A)(1 - pl(A)) \\
&= \sum_{A \subseteq \mathcal{X}} m^{N-1}(A) pl(A) \\
&= n_{\boldsymbol{m}^{N-1}}(\boldsymbol{pl}).
\end{aligned}$$

□

Proposition 4 shows that the N-consistency of a mass function amounts to the $n_{\boldsymbol{m}^{N-1}}$ pseudo-norm of the vector $\boldsymbol{pl}$ and its N-inconsistency amounts to the same pseudo-norm but of the vector $\boldsymbol{k}$. Note that $n_{\boldsymbol{m}^{N-1}}$ is in general a pseudo-norm, and not a norm, since it is possible that $m^{N-1}(A) = 0$ for some $A \subseteq \mathcal{X}$.

Corollary 1. *From $\boldsymbol{k} = \mathbf{1} - \boldsymbol{pl}$ and $\phi_N(m) = 1 - m^N(\emptyset)$, we also have the following equalities:*

$$\begin{aligned}
\phi_N(m) &= 1 - n_{\boldsymbol{m}^{N-1}}(\boldsymbol{k}) \\
&= n_{\boldsymbol{m}^{N-1}}(\mathbf{1} - \boldsymbol{k}),
\end{aligned}$$

and

$$\begin{aligned}
m^N(\emptyset) &= 1 - n_{\boldsymbol{m}^{N-1}}(\boldsymbol{pl}) \\
&= n_{\boldsymbol{m}^{N-1}}(\mathbf{1} - \boldsymbol{pl}).
\end{aligned} \tag{18}$$

Corollary 1 shows that it is equivalent to compute the pseudo-norm of the inverse of the inconsistency function (resp. plausibility function), and to compute the inverse of the pseudo-norm of the inconsistency function (resp. plausibility function). Briefly, the pseudo-norm and inverse commute.

Proposition 4 leads to the main result of this section, which expresses the consistency measures ϕ_N in terms of pseudo-distances to the state of total inconsistency:

Proposition 5. *For all $m \in \mathcal{M}$ and $1 \leq N \leq \mathfrak{F}$, we have*

$$\phi_N(m) = d_{\boldsymbol{m}^{N-1}}(\boldsymbol{pl}, \boldsymbol{pl}_\emptyset). \tag{19}$$

Proof. Eq. (19) follows from (16) and $\boldsymbol{pl}_\emptyset = \mathbf{0}$. □

In short, Proposition (5) shows that the N-consistency of a mass function is nothing but its pseudo-distance, induced from $n_{\boldsymbol{m}^{N-1}}$, to the totally inconsistent knowledge state. Accordingly, the farther a mass function from total inconsistency, the more consistent it is, which makes sense. It may be worth noting that this proposition holds because $\boldsymbol{pl}_\emptyset$ is a null vector in space $\mathcal{E}_\mathcal{X}$, and coincides thus with the origin of this space.

Particular cases $N = 1, 2, \mathfrak{F}$, of Proposition (5) yield

$$\begin{aligned} \phi_1(m) &= d_{\boldsymbol{m}^0}(\boldsymbol{pl}, \boldsymbol{pl}_\emptyset) = d_{\boldsymbol{m}_\mathcal{X}}(\boldsymbol{pl}, \boldsymbol{pl}_\emptyset), \qquad (20) \\ \phi_2(m) &= d_{\boldsymbol{m}^1}(\boldsymbol{pl}, \boldsymbol{pl}_\emptyset) = d_{\boldsymbol{m}}(\boldsymbol{pl}, \boldsymbol{pl}_\emptyset), \\ \phi_\mathfrak{F}(m) &= d_{\boldsymbol{m}^{\mathfrak{F}-1}}(\boldsymbol{pl}, \boldsymbol{pl}_\emptyset). \qquad (21) \end{aligned}$$

Example 1. *Let us illustrate Eq.* (21) *with mass function* m_1 *from Section 3.7. This mass function has* $\mathfrak{F}_1 = 3$ *focal sets and we have seen that* $\phi_{\mathfrak{F}_1}(m_1) = \phi_3(m_1) = 0.892$.

Using (21)*, we find:*

$$\begin{aligned} \phi_{\mathfrak{F}_1}(m_1) &= d_{\boldsymbol{m}_1^{\mathfrak{F}_1-1}}(\boldsymbol{pl}_1, \boldsymbol{pl}_\emptyset) \\ &= d_{\boldsymbol{m}_1^2}(\boldsymbol{pl}_1, \boldsymbol{pl}_\emptyset) \\ &= n_{\boldsymbol{m}_1^2}(\boldsymbol{pl}_1) \\ &= \sum_{A \subseteq \mathcal{X}} m_1^2(A) pl(A) \\ &= m_1^2(\{d_1, d_2, d_3\}) pl_1(\{d_1, d_2, d_3\}) + m_1^2(\{d_1, d_2, d_4\}) pl_1(\{d_1, d_2, d_4\}) \\ &\quad + m_1^2(\{d_3, d_4\}) pl_1(\{d_3, d_4\}) + m_1^2(\{d_1, d_2\}) pl_1(\{d_1, d_2\}) \\ &\quad + m_1^2(\{d_4\}) pl_1(\{d_4\}) + m_1^2(\{d_3\}) pl_1(\{d_3\}) \\ &= 0.892. \end{aligned}$$

Remark 3. *Similar results to Propositions 4 and 5 exist for* ϕ_π*. Indeed, let us denote by* $\mathcal{E}_x$ *the* K*-dimensional subspace of* $\mathcal{E}_\mathcal{X}$ *spanned by the set* $\{\boldsymbol{e}_x, x \in \mathcal{X}\}$ *corresponding to singletons. Then, the contour function* π *associated to a mass function* m *is represented by the vector* $\boldsymbol{\pi} = \sum_{x \in \mathcal{X}} \pi(x) \boldsymbol{e}_x$ *of* $\mathcal{E}_x$*. Since* $\phi_\pi(m) = \max_{x \in \mathcal{X}} \pi(x)$ *and* $\boldsymbol{\pi}_\emptyset$ *is a null vector in space* $\mathcal{E}_x$*, we have*

$$\begin{aligned} \phi_\pi(m) &= n^\infty(\boldsymbol{\pi}) \\ &= d^\infty(\boldsymbol{\pi}, \boldsymbol{\pi}_\emptyset). \qquad (22) \end{aligned}$$

Note also that $\phi_1(m) = 1 - m(\emptyset) = pl(\mathcal{X}) = \max_{A \subseteq \mathcal{X}} pl(A)$ *and thus we have, besides the equality* (20)*,*

$$\begin{aligned} \phi_1(m) &= n^\infty(\boldsymbol{pl}) \\ &= d^\infty(\boldsymbol{pl}, \boldsymbol{pl}_\emptyset). \qquad (23) \end{aligned}$$

Hence, $\phi_\pi(m)$ and $\phi_1(m)$ amount to the n^∞ norms of the contour and plausibility functions, respectively, as well as to the d^∞ distances of these functions to total inconsistency.

Let us denote by d_ϕ the pseudo-distance associated to consistency measure ϕ, via the equalities (19) and (22) (*i.e.*, $d_{\phi_N} = d_{\boldsymbol{m}^{N-1}}$ and $d_{\phi_\pi} = d^\infty$). This section has thus shown that the consistency of a mass function amounts, for any consistency measure ϕ considered in this paper, to its pseudo-distance d_ϕ to total inconsistency.

Note that the above results suggest that measure $\phi_N^p(m) := n^p_{\boldsymbol{m}^{N-1}}(\boldsymbol{pl})$, with $p \neq 1, \infty$, might be interesting to investigate as a candidate consistency measure. This is left for further research.

4.3. Geometric perspective on conflict measures

As a natural extension of the preceding results, based on the fact that a given conflict measure is induced by a consistency measure, it is shown below how conflict measures κ_N can be expressed in terms of distances to the total inconsistency state.

Proposition 6. *For any $m_1, m_2 \in \mathcal{M}$, we have*

$$\begin{aligned} \kappa_N(m_1, m_2) &= 1 - n_{\boldsymbol{m}_{1\bar{\bigcirc}2}^{N-1}}(\boldsymbol{pl}_{1\bigcirc 2}) \\ &= n_{\boldsymbol{m}_{1\bar{\bigcirc}2}^{N-1}}(\boldsymbol{k}_{1\bigcirc 2}) \end{aligned}$$

Proof. Follows from Eq. (9) and Proposition 4. □

Proposition 6 shows that the κ_N conflict between mass functions amounts to the $n_{\boldsymbol{m}_{1\bar{\bigcirc}2}^{N-1}}$ pseudo-norm of the inconsistency function of their combination. Equivalently, it is equal to one minus the $n_{\boldsymbol{m}_{1\bar{\bigcirc}2}^{N-1}}$ pseudo-norm of the plausibility function of their combination.

As the norm is simply the distance to the origin, Proposition 5 leads to the following relation between conflict and pseudo-distance:

Proposition 7. *For any $m_1, m_2 \in \mathcal{M}$ and $1 \leq N \leq \mathfrak{F}_{12}$, we have*

$$\kappa_N(m_1, m_2) = 1 - d_{\boldsymbol{m}_{1\bar{\bigcirc}2}^{N-1}}(\boldsymbol{pl}_{1\bigcirc 2}, \boldsymbol{pl}_\emptyset).$$

Special cases $N = 1, 2, \mathfrak{F}_{12}$, yield

$$\begin{aligned} \kappa_1(m_1, m_2) &= 1 - d_{\boldsymbol{m}_\mathcal{X}}(\boldsymbol{pl}_{1\bigcirc 2}, \boldsymbol{pl}_\emptyset), \\ \kappa_2(m_1, m_2) &= 1 - d_{\boldsymbol{m}_{1\bigcirc 2}}(\boldsymbol{pl}_{1\bigcirc 2}, \boldsymbol{pl}_\emptyset), \\ \kappa_{\mathfrak{F}_{12}}(m_1, m_2) &= 1 - d_{\boldsymbol{m}^{\mathfrak{F}_{12}-1}}(\boldsymbol{pl}_{1\bigcirc 2}, \boldsymbol{pl}_\emptyset). \end{aligned}$$

In particular, we note that the classical conflict measure $m_{1\textcircled{\cap}2}(\emptyset)$ between mass functions in the TBM is equal to one minus the pseudo-distance $d_{\boldsymbol{m}_\mathcal{X}}$ between the plausibility function of their combination and the plausibility function of the empty mass function.

Informally, Proposition 7 shows that the conflict between m_1 and m_2 amounts to one minus the pseudo-distance between their conjunctive combination and the totally inconsistent knowledge state (the counterpart to Remark 3 for conflict measure κ_π yields a similar conclusion: $\kappa_\pi(m_1, m_2) = 1 - d^\infty(\boldsymbol{\pi}_{1\textcircled{\cap}2}, \boldsymbol{\pi}_\emptyset)$). This shows that while a distance between m_1 and m_2 is not an appropriate measure of their conflict [7, 11], distances can be used to express it.

Conversely, conflict measures can be used to express the distance between m_1 and m_2, as detailed in Remark 4 for a special case. Whether such relationships exist for other distances and conflict measures is an open question.

Remark 4. *Let the Euclidean distance between the plausibility functions be denoted d^2 and defined as*

$$d^2(pl_1, pl_2) := \sqrt{\sum_{A\subseteq\mathcal{X}} (pl_1(A) - pl_2(A))^2}.$$

By Lemma 5, we obtain for all $m_1, m_2 \in \mathcal{M}$:

$$d^2(pl_1, pl_2) \;=\; \sqrt{\sum_{A\subseteq\mathcal{X}} (\kappa_1(m_1, m_A) - \kappa_1(m_2, m_A))^2}. \tag{24}$$

As can be seen with (24)*, $d^2(pl_1, pl_2)$ does* not *evaluate how much m_1 and m_2 are in conflict with each other, but rather quantifies how much they are in conflict (according to the classical conflict measure κ_1) with the same categorical mass functions m_A (sets A), $A \subseteq \mathcal{X}$. This gives also some intuition behind the property that $d^2(pl_1, pl_2) = 0$ iff $m_1 = m_2$ whereas we can have $\kappa_N(m_1, m_2) \neq 0$ for $m_1 = m_2$: if $m_1 = m_2 = m$ then m_1 and m_2 have the same conflict with the same sets,* i.e.*, $\kappa_1(m_1, m_A) - \kappa_1(m_2, m_A) = 0$ $\forall A \subseteq \mathcal{X}$, whatever m may be, whereas $\kappa_N(m_1, m_2) = 0$ iff m_1 and m_2 are N-nonconflicting, that is m^2 is N-consistent.*

5. Conclusions

Conflict between belief functions has been defined by Destercke and Burger as the inconsistency of the belief function resulting from their conjunctive combination. The three existing definitions of the consistency of a belief

function were shown in this paper to be specific cases, or *shades*, of a parameterised family of consistency definitions, which we called consistency of order N. We introduced a corresponding family of consistency measures and derived an associated family of conflict measures between belief functions. Each of these conflict measures is shown to verify a list of desirable properties for quantifying conflict. The family of conflict measures encompasses the classical measure of conflict in belief function theory, associated to the weakest ("brightest") definition of consistency, as well as two other conflict measures associated, respectively, to a stronger ("darker") definition of consistency by Yager and to the strongest ("darkest") definition of consistency by Destercke and Burger (called logical consistency). In addition, a geometric view on consistency measures as well as on the associated conflict measures was provided. In particular, we showed that measuring the consistency of a belief function amounts to measuring its distance to the totally inconsistent knowledge state, whatever the definition of consistency considered. A similar result was also obtained for conflict measures.

Our conflict measures are applicable in the case where it can be safely assumed that the sources are independent. However, the measures can readily be extended to the case where the dependence between sources is known and is captured by a joint mass function as in [7]. Handling the case where the dependence structure cannot be uniquely identified would need some further work, but it seems possible (in particular, measure κ_1 for nonconflict has already been adapted to this case in [7]). Besides, similarly to what has been done in [7] for measure κ_π, it would be interesting to find a decomposition of the conflict κ_N between two belief functions, in terms of some internal conflict (inconsistency) of each of the belief functions and of some external conflict between them, such that this decomposition satisfies some sensible relations originally introduced in [10].

Other interesting perspectives include finding theoretical or practical considerations that could help in selecting in a given application, one specific conflict measure among the κ_N measures, or in choosing between the two measures for strong nonconflict $\kappa_{\mathfrak{F}_{12}}$ and κ_π. In addition, it would be interesting to know whether there exist other pseudo-norms of the plausibility vector, besides the family of pseudo-norms we have introduced and presented, which could serve as consistency measures.

Acknowledgements

The authors wish to thank the NATO STO CMRE and the Université d'Artois for their support via their visiting scientist programs.

References

[1] F. Pichon, A.-L. Jousselme, A norm-based view on conflict, in: Rencontres Francophones sur la Logique Floue et ses Applications, LFA 2016, La Rochelle, France, Cépaduès, 2016, pp. 153–159.

[2] F. Pichon, S. Destercke, T. Burger, A consistency-specificity trade-off to select source behavior in information fusion, IEEE Trans. Cybern. 45 (4) (2015) 598–609.

[3] G. Pallotta, A.-L. Jousselme, Data-driven detection and context-based classification of maritime anomalies, in: Proc. of the 18th Int. Conf. on Information Fusion, Washington, D. C., USA, 2015, pp. 1152–1159.

[4] C. Ray, R. Gallen, C. Iphar, A. Napoli, A. Bouju, Deais project: Detection of AIS spoofing and resulting risks, in: Proc. of OCEANS 2015, Genova, Italy, 2015, pp. 1–6.

[5] A. P. Dempster, A generalization of Bayesian inference, J. Roy. Stat. Soc. B Met. 30 (2) (1968) 205–247.

[6] G. Shafer, A mathematical theory of evidence, Princeton University Press, Princeton, N.J., 1976.

[7] S. Destercke, T. Burger, Toward an axiomatic definition of conflict between belief functions, IEEE Trans. Syst. Man Cybern. B 43 (2) (2013) 585–596.

[8] A. Martin, About conflict in the theory of belief functions, in: T. D. ux, M. Masson (Eds.), Belief Functions: Theory and Applications - Proceedings of the 2nd International Conference on Belief Functions, Compiègne, France, 9-11 May 2012, Vol. 164 of Advances in Intelligent and Soft Computing, Springer, 2012, pp. 161–168.

[9] W. Liu, Analyzing the degree of conflict among belief functions, Artificial Intelligence 170 (11) (2006) 909–924.

[10] M. Daniel, Conflicts within and between belief functions, in: E. Hüllermeier, R. Kruse, F. Hoffman (Eds.), Computational Intelligence for Knowledge-Based Systems Design, Lecture Notes in Computer Science, Vol. 6178, Springer Berlin Heidelberg, 2010, pp. 696–705.

[11] T. Burger, Geometric views on conflicting mass functions: From distances to angles, Int. J. Approx. Reason. 70 (2016) 36–50.

[12] M. Daniel, Conflict between belief functions: A new measure based on their non-conflicting parts, in: F. Cuzzolin (Ed.), Belief Functions: Theory and Applications - Third International Conference, BELIEF 2014, Oxford, UK, September 26-28, 2014. Proceedings, Vol. 8764 of Lecture Notes in Computer Science, Springer, 2014, pp. 321–330.

[13] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.

[14] P. Smets, R. Kennes, The transferable belief model, Artif. Intell. 66 (1994) 191–243.

[15] P. Smets, The nature of the unnormalized beliefs encountered in the transferable belief model, in: Proc. of the 8th Int. Conf. on Uncertainty in Artificial Intelligence, UAI'92, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1992, pp. 292–297.

[16] D. Dubois, H. Prade, A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets, Int. J. Gen. Syst. 12 (3) (1986) 193–226.

[17] R. R. Yager, On considerations of credibility of evidence, Int. J. Approx. Reason. 7 (1/2) (1992) 45–72.

[18] C. Osswald, A. Martin, Understanding the large family of Dempster-Shafer theory's fusion operators - a decision-based measure, in: Proc. of the 9th Int. Conf. on Information Fusion, Florence, Italy, 2006, pp. 1–7.

[19] T. Denœux, Inner and outer approximation of belief structures using a hierarchical clustering approach, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems 9 (4) (2001) 437–460.

[20] T. Burger, S. Destercke, How to randomly generate mass functions, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems 21 (05) (2013) 645–673.

[21] G. Pallotta, M. Vespe, K. Bryan, Vessel pattern knowledge discovery from AIS data - A framework for anomaly detection and route prediction, Entropy 5 (6) (2013) 2218–2245.

[22] P. Smets, Analyzing the combination of conflicting belief functions, Information Fusion 8 (2007) 387–412.

[23] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, Int. J. Approx. Reason. 38 (2) (2005) 133 – 147.

[24] A.-L. Jousselme, P. Maupin, Distances in evidence theory: Comprehensive survey and generalizations, Int. J. Approx. Reason. 53 (2) (2012) 118–145.

[25] R. R. Yager, Norms induced from OWA operators, IEEE Trans. on Fuzzy Syst. 18 (1) (2010) 57–66.

[26] P. S. Bullen, Handbook of means and their inequalities, Kluwer, Dordrecht, The Netherlands, 2003.

Annexe H

The capacitated vehicle routing problem with evidential demands

---

Article paru dans *International Journal of Approximate Reasoning*, 95 : 124-151, 2018.

International Journal of Approximate Reasoning 95 (2018) 124–151

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

The capacitated vehicle routing problem with evidential demands ☆

Nathalie Helal [a,*], Frédéric Pichon [a], Daniel Porumbel [b], David Mercier [a], Éric Lefèvre [a]

[a] Univ. Artois, EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), F-62400 Béthune, France
[b] Conservatoire National des Arts et Métiers, EA 4629, Cedric, 75003 Paris, France

ARTICLE INFO

Article history:
Received 12 September 2017
Received in revised form 8 February 2018
Accepted 8 February 2018
Available online 13 February 2018

Keywords:
Vehicle routing problem
Chance constrained programming
Stochastic programming with recourse
Belief function

ABSTRACT

We propose to represent uncertainty on customer demands in the Capacitated Vehicle Routing Problem (CVRP) using the theory of evidence. To tackle this problem, we extend classical stochastic programming modelling approaches. Specifically, we propose two models for this problem. The first model is an extension of the *chance-constrained programming* approach, which imposes certain minimum bounds on the belief and plausibility that the sum of the demands on each route respects the vehicle capacity. The second model extends the *stochastic programming with recourse* approach: for each route, it represents by a belief function the uncertainty on its recourses, *i.e.*, corrective actions performed when the vehicle capacity is exceeded, and defines the cost of a route as its classical cost (without recourse) plus the worst expected cost of its recourses. We solve the proposed models using a metaheuristic algorithm and present experimental results on instances adapted from a well-known CVRP data set.

1. Introduction

In the Capacitated Vehicle Routing Problem (CVRP), we are given a fleet of vehicles with identical capacity located at a depot and a set of customers with known demands located on the vertices of a graph. The goal of this problem is to determine a route for each vehicle, such that the set of routes for all the vehicles has the least total cost, all customer demands are fully serviced, the capacity of each vehicle is always respected and each customer is visited by exactly one route. The CVRP is NP-hard since it contains the traveling salesman problem as a particular case (one route and unbounded capacity). It can be written as an integer linear program. The CVRP has generated a large body of research, since it belongs to the class of local transportation or delivery problems affecting the most expensive component in the distribution network [8].

Yet, many industrial applications are confronted with uncertainty on customer demands in their distribution problems involving the CVRP, and the exact customer demands are mostly revealed when the servicing vehicles arrive at the customers. Accordingly, several authors (see, *e.g.*, [28,29] and the references therein) tackled this issue by assuming that customer demands are random variables and the associated problem is the well-known Capacitated Vehicle Routing Problem with Stochastic Demands (CVRPSD). Two of the most widely-used frameworks for modelling stochastic problems, such as the

☆ This paper is an extended and revised version of [33] and [35].

* Corresponding author.
E-mail addresses: nathalie_helal@ens.univ-artois.fr (N. Helal), frederic.pichon@univ-artois.fr (F. Pichon), daniel.porumbel@cnam.fr (D. Porumbel), david.mercier@univ-artois.fr (D. Mercier), eric.lefevre@univ-artois.fr (É. Lefèvre).

https://doi.org/10.1016/j.ijar.2018.02.003
0888-613X/

CVRPSD, are the Chance-Constrained Programming (CCP) approach and the Stochastic Programming with Recourse (SPR) approach [7]. Modelling the CVRPSD via CCP amounts to using a probabilistic capacity constraint that requires the probability of respecting the capacity constraint to be above a certain threshold. The CCP modelling technique does not consider the additional cost of recourse (or corrective) actions necessary if capacity constraints fail to be satisfied. The SPR approach does consider situations needing recourses and it aims at minimizing the initially-planned travel cost plus the expected cost of the recourses executed along routes, *e.g.*, returning to the depot and unloading in order to bring to feasibility a violated capacity constraint.

The probabilistic approach to modelling uncertainty is not necessarily well-suited to all real-life situations. In particular, its ability to handle epistemic uncertainty (uncertainty arising from lack of knowledge) has been criticised [4,1]. The typical approach to representing basic epistemic uncertainty is the set-valued approach [26]. It is sensible when, *e.g.*, all that is known about the customer demands is that they belong to some intervals. This kind of uncertainty in the CVRP is generally addressed using robust optimisation, where one optimises against the worst-case scenario, that is, one wants to obtain solutions that are robust to all realisations of customer demands that are deemed possible (see, *e.g.*, [49]). However, the set-valued approach to uncertainty representation may be too coarse and may thus lead to solutions that are too conservative, hence not useful.

In the last forty years, the necessity to account for all facets of uncertainty has been recognized and alternative uncertainty frameworks extending both the probabilistic and set-valued ones have appeared [4]. In particular, the theory of evidence introduced by Shafer [47], based on some previous work from Dempster [15], has emerged as a theory offering a compromise between expressivity and complexity, which seems interesting in practice as its successful application in several domains testifies (see [18] for a recent survey of evidence theory applications). This theory, also known as belief function theory, may be used to model various forms of information, such as expert judgements and statistical evidence, and it also offers tools to combine and propagate uncertainty [1].

In the context of the CVRP, the theory of evidence may be used to represent uncertainty on customer demands leading to an optimisation problem, which will be referred to as the CVRP with Evidential Demands (CVRPED). Using the theory of evidence in this problem seems particularly interesting as it allows one to account for imperfect knowledge about customer demands, such as knowing that each customer demand belongs to one or more sets with a given probability allocated to each set – an intermediary situation between probabilistic and set-valued knowledge. In this paper, we propose to address the CVRPED by extending the CCP and SPR modelling approaches into the formalism of evidence theory. Although the focus will be to extend stochastic programming approaches, we will also connect our formulations with robust optimisation.

To our knowledge, evidence theory has not yet been considered to model uncertainty in large-scale instances of an NP-hard optimisation problem like the CVRP. Indeed, it seems that so far, only other non-classical uncertainty theories, and in particular fuzzy set theory [50,9,42,11], have been used in such problems. Besides, modelling uncertainty in optimisation problems using evidence theory has concerned only continuous design optimisation problems[1] [41,48] and continuous linear programs [40]. Specifically in [41], the reliability of the system is optimized, while uncertainty is handled by limiting the plausibility of constraints violation into a small degree; while in [48] the problem was handled differently, and the plausibility of a constraint failure was converted into a second objective to the problem that should be minimized. Of particular interest is the work of Masri and Ben Abdelaziz [40], who extended the CCP and SPR modelling approaches, in order to model continuous linear programs embedding belief functions, which they called the Belief Constrained Programming (BCP) and the recourse approaches, respectively. In comparison, in this work, we generalise CCP and SPR to an *integer* linear program involving uncertainty modelled by evidence theory. Borrowing from [40], we propose to model the CVRPED by methods that may be called the BCP modelling of the CVRPED and the recourse modelling of the CVRPED. For both models, the resolution algorithm is a simulated annealing algorithm; we use a metaheuristic, as the CVRPED derives from the CVRP, which is NP-hard.

The paper is structured as follows. Section 2 summarises the basic preliminaries on the CVRP and on the CVRPSD modelling via CCP and SPR, along with the necessary background on evidence theory. In Section 3, the BCP model and the recourse model for the CVRPED are presented and some of their properties are studied. In Section 4 we solve the BCP model and the recourse model of the CVRPED using a simulated annealing algorithm and perform experiments on instances generated from CVRP benchmarks. In Section 5, we conclude and state the perspectives of the present work.

2. Background

This section recalls necessary background on the CVRP, the CVRPSD and its stochastic programming formulations, as well as some concepts of belief function theory needed in this paper.

[1] Designing physical systems in the engineering field using optimisation techniques, so design costs are minimized, while the system performance is fulfilled [2].

2.1. The CVRP

In the CVRP, a fleet of m identical vehicles with a given capacity limit Q, initially located at a depot, must collect[2] goods from n customers, with d_i such that $0 < d_i \leq Q$ the deterministic collect demand of client i, $i = 1, \ldots, n$. The objective in the CVRP is to find a set of m routes with minimum cost to serve all the customers such that total customer demands on any route must not exceed Q, each route starts and ends at the depot, and each customer is serviced only once.

Formally, it is convenient to represent the depot by an artificial client $i = 0$, whose demand always equals 0, *i.e.*, $d_0 = 0$. The CVRP may be defined on a graph $G = (V, E)$ such that $V = \{0, \ldots, n\}$ is the vertex set and $E = \{(i, j) \,|\, i \neq j;\ i, j \in V\}$ is the arc set. V represents the customers and the depot that corresponds to vertex 0. A travel cost (or travel time or distance – these terms are interchangeable) $c_{i,j}$ is associated with every edge in E. Travel costs are such that $c_{i,j} = c_{j,i},\ \forall\ (i, j) \in E$ and they satisfy the triangle inequality: $c_{i,j} \leq c_{i,l} + c_{l,j},\ \forall i, l, j \in V$. Besides, $c_{i,i} = +\infty,\ \forall i \in V$ [51]. Let R_k be the route associated to vehicle k and $w^k_{i,j}$ a binary variable that equals 1 if vehicle k travels from i to j and serves j (except if j is the depot), and 0 if it does not. A proper formulation for the CVRP [8,38] is:

$$\min \sum_{k=1}^{m} C(R_k), \tag{1}$$

where

$$C(R_k) = \sum_{i=0}^{n} \sum_{j=0}^{n} c_{i,j} w^k_{i,j}, \tag{2}$$

subject to

$$\sum_{i=0}^{n} \sum_{k=1}^{m} w^k_{i,j} = 1, \qquad j = 1, \ldots, n, \tag{3}$$

$$\sum_{i=0}^{n} w^k_{i,\ell} = \sum_{j=0}^{n} w^k_{\ell,j}, \qquad k = 1, \ldots, m,\ \ell = 0, \ldots, n, \tag{4}$$

$$\sum_{j=1}^{n} w^k_{0,j} \leq 1, \qquad k = 1, \ldots, m, \tag{5}$$

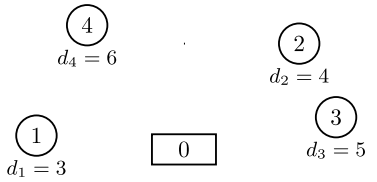
$$\sum_{\substack{i,j \in L \\ i \neq j}} \sum_{k=1}^{m} w^k_{i,j} \leq |L| - 1, \qquad L \subseteq V \setminus \{0\}, \tag{6}$$

$$\sum_{i=1}^{n} d_i \sum_{j=0}^{n} w^k_{i,j} \leq Q, \qquad k = 1, \ldots, m. \tag{7}$$

Constraints (3) make sure that exactly one vehicle arrives at client j, $j = 1, \ldots, n$. Constraints (4) ensure the continuity of the routes (flow): if vehicle k leaves vertex ℓ, vehicle k must also enter vertex ℓ, ensuring that the route is a proper unbroken cycle in the graph. Constraints (5) oblige vehicle k, $k = 1, \ldots, m$, to leave at most one time the depot. The choices of the arcs that are represented by $w^k_{i,j}$ is also restricted by constraints (6), that forbids subtours solutions [8]. Without these latter constraints, we can have a vehicle performing the path $(i_1, i_2, \ldots, i_t)$ with $0 \notin \{i_1, i_2, \ldots, i_t\}$. Constraints (7) state that every vehicle cannot carry more than its capacity limit. We note that constraints (3) and (4) imply $\sum_{i=0}^{n} \sum_{k=1}^{m} w^k_{j,i} = 1,\ j = 1, \ldots, n$, *i.e.*, exactly one vehicle leaves client j. Constraints (5) and (4) imply $\sum_{i=1}^{n} w^k_{i,0} \leq 1,\ k = 1, \ldots, m$, *i.e.*, vehicle k is obliged to return at most one time to the depot. Finally, we remark that this model requires using *at most* m vehicles, since for some k, we might have $w^k_{i,j} = 0,\ i, j = 1, \ldots, n$.

Example 1. Suppose $m = 2$ vehicles with capacity limit $Q = 10$, which must collect the demands of $n = 4$ customers with demands $d_1 = 3, d_2 = 4, d_3 = 5, d_4 = 6$. These customers are illustrated in Fig. 1a (the depot is denoted by "0"), along with their associated travel cost matrix in Fig. 1b. A candidate solution, *i.e.*, a set of routes satisfying constraints (3)–(7), to this problem is shown in Fig. 1c; the total travel cost (the value of the objective function in Equation (1)) of this solution is

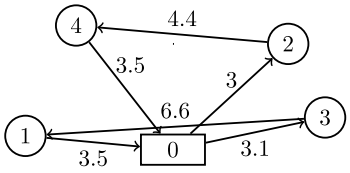
[2] The problem can also be presented in terms of delivery, rather than collection, of goods.



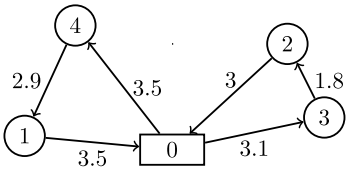
(a) The four clients with their demands

	0	1	2	3	4
0	$+\infty$	3.5	3	3.1	3.5
1	3.5	$+\infty$	6	6.6	2.9
2	3	6	$+\infty$	1.8	4.4
3	3.1	6.6	1.8	$+\infty$	5.7
4	3.5	2.9	4.4	5.7	$+\infty$

(b) The travel cost matrix



(c) A candidate solution



(d) An optimal solution

Fig. 1. A simple CVRP.

24.1. An optimal solution, *i.e.*, a set of routes satisfying constraints (3)–(7) and with minimum cost among the candidate solutions, is provided in Fig. 1d; its total travel cost is 17.8.

2.2. The CVRPSD

The CVRPSD is a variation of the CVRP, which introduces stochastic demands in the CVRP, *i.e.*, $d_i, i = 1, \ldots, n$, are now random variables, such that $P(d_i \leq Q) = 1$ (these random variables are usually assumed to be independent). The CVRPSD is typically handled using the framework of stochastic programming, which models stochastic programs in two stages: an "a priori" solution is established in the first stage, and then in the second stage the realisations of the random variables – the actual demands in the case of the CVRPSD – are revealed and corrective actions are carried out if necessary on the first stage solution [28]. More precisely, the CVRPSD is either modelled as a so-called chance-constrained program [10] or as a stochastic program with recourse [7]; these two models are detailed in the next two sections.

2.2.1. The CVRPSD modelled by CCP

Chance constrained programming consists in finding a first stage solution for which the probability that the total demand on any route exceeds the capacity is constrained to be below a given threshold. Formally, a CCP formulation for the CVRPSD corresponds to the same optimisation problem described for the CVRP in Section 2.1 except that deterministic constraints represented by (7) are replaced by the following so called chance-constraints:

$$P\left(\sum_{i=0}^{n} d_i \sum_{j=0}^{n} w_{i,j}^{k} \leq Q\right) \geq 1 - \beta, \qquad k = 1, \ldots, m, \tag{8}$$

where $1 - \beta$ is the minimum allowable probability that any route respects vehicle capacity and thus succeeds. Note that this model represents a so-called *individual* chance-constrained model, since the inequality must be satisfied for every k separately; see [45,7] for more details.

This model does not consider the cost of corrective actions that may be necessary when the first stage solution is implemented. Indeed, when implementing this solution, it is unlikely yet possible that the vehicle capacity is exceeded, *i.e.*, route failures occur, when the actual demands are revealed and thus corrective actions may have to be carried out in the second stage.

2.2.2. The CVRPSD modelled by SPR

Stochastic programming with recourse deals explicitly with the possibility of a first stage solution failure, by incorporating into the objective of the problem the penalty cost of corrective, or *recourse*, actions such as allowing vehicles to return to the depot to unload. More specifically, in the SPR modelling of the CVRPSD, the *expected* penalty cost of the recourse actions happening in the second stage is considered and the problem is to find a set of routes which has the minimal expected cost defined as the cost of the first stage solution if no failures occur, plus the expected penalty cost of the recourse actions of the second stage. Formally, let $C_{\mathrm{E}}(R_k)$ denote the expected cost of a route R_k defined by

$$C_{\mathrm{E}}(R_k) = C(R_k) + C_{\mathrm{P}}(R_k), \tag{9}$$

with $C(R_k)$ the cost defined by (2) representing the cost of traveling along R_k if no recourse action is performed, and $C_{\mathrm{P}}(R_k)$ the expected penalty cost on R_k – $C_{\mathrm{P}}(R_k)$ may be defined in many different ways depending on the recourse policy used

(see, *e.g.*, [12,27,39,23]). Then, a SPR model for the CVRPSD consists in modifying the CVRP model presented in Section 2.1 as follows. The objective is to find a set of routes minimizing the sum of the expected costs of routes R_k, *i.e.*,

$$\min \sum_{k=1}^{m} C_{\mathrm{E}}(R_k), \tag{10}$$

subject to constraints (3)–(6) excluding constraints (5), which is replaced by

$$\sum_{j=1}^{n} w_{0,j}^{k} = 1, \qquad k = 1, \ldots, m, \tag{11}$$

that is *exactly m* vehicles must be used. Constraints (5) may be considered instead of (11), but then the problem becomes even more difficult to solve. In addition, note that the binary variables $w_{i,j}^{k}$ do not encode recourse actions: they represent only the initially planned solution routes, *i.e.*, the first stage solution.

2.3. Evidence theory

In this section, basic concepts as well as some more advanced notions of evidence theory [47], which are necessary in our study on the CVRPED, are recalled.

2.3.1. Basics of evidence theory

Let x be a variable taking its values in a finite domain $X = \{x_1, \ldots, x_K\}$. In this theory, uncertain knowledge about x may be represented by a *Mass Function* (MF) defined as a mapping $m^X : 2^X \to [0, 1]$ such that $m^X(\emptyset) = 0$ and $\sum_{A \subseteq X} m^X(A) = 1$. The superscript ^{X} can be omitted when there is no risk of confusion. Each mass $m^X(A)$ represents the probability of knowing only that $x \in A$. Subsets $A \subseteq X$ such that $m^X(A) > 0$ are called the *focal sets* of m^X. To be consistent with the stochastic case terminology, a variable x whose true value is known in the form of a MF will be called an *evidential variable*.

Mass functions generalise both probabilistic and set valued representations of uncertainty since:

- a MF whose focal sets are singletons, *i.e.*, $m^X(A) > 0$ iff $|A| = 1$, corresponds to a probability mass function and is called a *Bayesian* MF;
- a MF having only one focal set, *i.e.*, $m^X(A) = 1$ for some $A \subseteq X$, corresponds to a set and is called a *categorical* MF.

Another special case of mass functions are those whose focal sets are nested, in which case they are called *consonant*.

Equivalent representations of a MF m^X are the *belief* and *plausibility* functions defined, respectively, as

$$Bel^X(x \in A) = \sum_{C \subseteq A} m^X(C), \quad \forall A \subseteq X,$$

$$Pl^X(x \in A) = \sum_{C \cap A \neq \emptyset} m^X(C), \quad \forall A \subseteq X.$$

The *degree of belief* $Bel^X(x \in A)$ can be interpreted as the probability that the evidence about x and represented by m^X, supports (implies) $x \in A$, whereas the *degree of plausibility* $Pl^X(x \in A)$ is the probability that the evidence is consistent with $x \in A$. For all $A \subseteq X$, we have $Bel^X(x \in A) \leq Pl^X(x \in A)$ and

$$Pl^X(x \in A) = 1 - Bel^X(x \in A^{\complement}), \tag{12}$$

where $A^{\complement}$ denotes the complement of A. Besides, if m^X is Bayesian, then $Bel^X(x \in A) = Pl^X(x \in A)$, for all $A \subseteq X$, and this function is a probability measure. If m^X is categorical, then $Bel^X(x \in A) \in \{0, 1\}$ and $Pl^X(x \in A) \in \{0, 1\}$, for all $A \subseteq X$, and the plausibility function restricted to the singletons corresponds to the indicator function of the set associated to m^X. If m^X is consonant, then its associated plausibility function is a possibility measure [54]: it verifies $Pl^X(x \in A \cup B) = \max(Pl^X(x \in A), Pl^X(x \in B))$, for all $A, B \subseteq X$.

Given a MF m^X and a function $h : X \to \mathbb{R}^+$, it is possible to compute the *lower expected value* $E_*(h, m^X)$ and *upper expected value* $E^*(h, m^X)$ of h relative to m^X defined, respectively, as [16]

$$E_*(h, m^X) = \sum_{A \subseteq X} m^X(A) \min_{x \in A} h(x), \tag{13}$$

$$E^*(h, m^X) = \sum_{A \subseteq X} m^X(A) \max_{x \in A} h(x). \tag{14}$$

If m^X is Bayesian, then $E_*(h, m^X)$ and $E^*(h, m^X)$ reduce to the classical (probabilistic) expected value of h relative to the probability mass function m^X.

2.3.2. Comparisons of belief functions

The informative content of two set-valued pieces of information $x \in A$ and $x \in B$, $A, B \subseteq X$, about x is naturally compared by saying that $x \in A$ is more informative than $x \in B$ if $A \subset B$. An extension of this to compare the informative content of mass functions in terms of specificity is the notion of specialisation [24]: a MF m_1^X defined on X is said to be at least as informative (or *specific*) as another MF m_2^X defined on X, which is denoted by $m_1^X \sqsubseteq m_2^X$, if and only if there exists a non-negative square matrix $S = [S(A, B)]$, $A, B \in 2^X$, verifying

$$\sum_{A \subseteq X} S(A, B) = 1, \quad \forall B \subseteq X, \tag{15}$$

$$S(A, B) > 0 \Rightarrow A \subseteq B, \quad A, B \subseteq X, \tag{16}$$

$$m_1^X(A) = \sum_{B \subseteq X} S(A, B) m_2^X(B), \quad \forall A \subseteq X. \tag{17}$$

The term $S(A, B)$ may be seen as the proportion of the mass $m_2^X(B)$ which "flows down" to A. Let us also recall that we have [24]

$$m_1^X \sqsubseteq m_2^X \Rightarrow [Bel_1^X(x \in A), Pl_1^X(x \in A)] \subseteq [Bel_2^X(x \in A), Pl_2^X(x \in A)], \forall A \subseteq X. \tag{18}$$

Assume now that an ordering has been defined among the elements of X. By convention, assume that $x_1 < \ldots < x_K$. Let $A_{\underline{a},\overline{a}}$ denote the subset $\{x_{\underline{a}}, \ldots, x_{\overline{a}}\}$, for $1 \leq \underline{a} \leq \overline{a} \leq K$ and let $\mathcal{I}$ denote the set of intervals of X: $\mathcal{I} = \{A_{\underline{a},\overline{a}}, 1 \leq \underline{a} \leq \overline{a} \leq K\}$. Deciding whether an interval $A_{\underline{a},\overline{a}}$, *i.e.* an interval-valued piece of information $x \in A_{\underline{a},\overline{a}}$ about x, is smaller or equal to another interval $B_{\underline{b},\overline{b}}$ can be done in several ways, and in particular the so-called lattice ordering denoted $\leq_{lo}$ is defined as [20]: $A_{\underline{a},\overline{a}} \leq_{lo} B_{\underline{b},\overline{b}}$ if $\underline{a} \leq \underline{b}$ and $\overline{a} \leq \overline{b}$. This ordering can be extended to arbitrary subsets A and B of X as follows: $A \leq_{lo} B$ if $\underline{a} \leq \underline{b}$ and $\overline{a} \leq \overline{b}$, where $\underline{a}$ and $\underline{b}$ (resp. $\overline{a}$ and $\overline{b}$) denote the indices of the lowest (resp. highest) values in A and B. More generally, following the extension above of inclusion between sets to mass functions, the ordering $\leq_{lo}$ of subsets can be extended to compare mass functions in terms of ranking as follows: a MF m_1^X is said to be at least as small as another MF m_2^X, which is denoted by $m_1^X \preceq m_2^X$, if and only if there exists a non-negative square matrix $R = [R(A, B)]$, $A, B \in 2^X$, verifying

$$\sum_{A \subseteq X} R(A, B) = 1, \quad \forall B \subseteq X, \tag{19}$$

$$R(A, B) > 0 \Rightarrow A \leq_{lo} B, \quad A, B \subseteq X, \tag{20}$$

$$m_1^X(A) = \sum_{B \subseteq X} R(A, B) m_2^X(B), \quad \forall A \subseteq X. \tag{21}$$

In other words, the mass $m_2^X(B)$ can be shared among smaller (according to $\leq_{lo}$) subsets than B. We note that extensions of interval rankings to belief functions were already proposed in [17], but in the context of belief functions on the real line and the ranking $\leq_{lo}$ was not considered. To our knowledge, the definition of $\preceq$ appears thus to be new; it will be particularly useful in conjunction with Proposition 1, which is somewhat of a counterpart to Eq. (18) for $\preceq$, to exhibit a property of the BCP model for the CVPRED:

Proposition 1. *For any Q, $1 \leq Q \leq K$, we have*

$$m_1^X \preceq m_2^X \Rightarrow \begin{cases} Bel_1^X(x \in A_{1,Q}) \geq Bel_2^X(x \in A_{1,Q}), \\ Pl_1^X(x \in A_{1,Q}) \geq Pl_2^X(x \in A_{1,Q}). \end{cases} \tag{22}$$

The converse does not hold, i.e., *the implication in* (22) *is strict.*

Proof. See Appendix A. □

Proposition 1 basically says that if a MF m_1^X is at least as small as a MF m_2^X, then the belief, according to the piece of evidence m_1^X, that the value of x is smaller or equal than a value x_Q is at least as great as the belief of the same event according to the piece of evidence m_2^X (and the same goes for the plausibility), as may be expected from the meaning of $\preceq$.

To sum up this section, $m_1^X \sqsubseteq m_2^X$ basically means that m_2^X represents a less precise piece of uncertain knowledge about x than m_1^X, whereas $m_1^X \preceq m_2^X$ means that m_2^X represents a piece of uncertain knowledge telling that x takes a higher value than what m_1^X tells.

2.3.3. Uncertainty propagation

Let $x^1, \ldots, x^N$ be N variables defined on the finite domains $X_1, \ldots, X_N$, respectively. A MF $m^{X_1 \times \cdots \times X_N}$ defined on the Cartesian product $X_1 \times \cdots \times X_N$ represent joint knowledge about the values of these variables.

Similarly as in probability theory, one can obtain joint knowledge about a subset of the evidential variables $x^1, \ldots, x^N$ by *marginalising* MF $m^{X_1 \times \cdots \times X_N}$ on the domains of these variables. For instance, and without lack of generality, the marginalisation of $m^{X_1 \times \cdots \times X_N}$ on $X_1 \times X_2$ is the MF $m^{X_1 \times \cdots \times X_N \downarrow X_1 \times X_2}$ on $X_1 \times X_2$ defined as, $\forall A \subseteq X_1 \times X_2$,

$$m^{X_1 \times \cdots \times X_N \downarrow X_1 \times X_2}(A) = \sum_{\{B \subseteq X_1 \times \cdots \times X_N, B^{\downarrow X_1 \times X_2} = A\}} m^{X_1 \times \cdots \times X_N}(B), \tag{23}$$

where $B^{\downarrow X_1 \times X_2}$ denotes the projection of B onto $X_1 \times X_2$.

If MF $m^{X_1 \times \cdots \times X_N}$ satisfies, for all $A \subseteq X_1 \times \cdots \times X_N$,

$$m^{X_1 \times \cdots \times X_N}(A) = \begin{cases} \prod_{i=1}^{N} m^{X_1 \times \cdots \times X_N \downarrow X_i}(A^{\downarrow X_i}) & \text{if } A = \times_{i=1}^{N} A^{\downarrow X_i}, \\ 0 & \text{otherwise,} \end{cases} \tag{24}$$

then variables $x^1, \ldots, x^N$ are said to be *evidentially independent* (or independent for short) [47]. In practice, this happens when joint knowledge about variables $x^1, \ldots, x^N$, is built from marginal knowledge m^{X_i}, $i = 1, \ldots, N$, on each of these variables, supplied by sources assumed to be independent [43], as illustrated by Example 2 (other reasons for this to happen can also be found in [14,13]).

Example 2. Let $X_1 = \{x_1^1, x_2^1, x_3^1\}$ and $X_2 = \{x_1^2, x_2^2\}$. Furthermore, assume two sources providing the pieces of evidence m^{X_1} and m^{X_2}, respectively, about x_1 and x_2, defined as $m^{X_1}(\{x_1^1, x_2^1\}) = 0.8$, $m^{X_1}(\{x_2^1, x_3^1\}) = 0.2$ and $m^{X_2}(\{x_1^2\}) = 0.7$, $m^{X_2}(X_2) = 0.3$. Assuming that the sources are independent, we obtain

$$\begin{aligned} m^{X_1 \times X_2}(\{x_1^1, x_2^1\} \times \{x_1^2\}) &:= m^{X_1}(\{x_1^1, x_2^1\}) \cdot m^{X_2}(\{x_1^2\}) = 0.56, \\ m^{X_1 \times X_2}(\{x_1^1, x_2^1\} \times X_2) &:= m^{X_1}(\{x_1^1, x_2^1\}) \cdot m^{X_2}(X_2) = 0.24, \\ m^{X_1 \times X_2}(\{x_2^1, x_3^1\} \times \{x_1^2\}) &:= m^{X_1}(\{x_2^1, x_3^1\}) \cdot m^{X_2}(\{x_1^2\}) = 0.14, \\ m^{X_1 \times X_2}(\{x_2^1, x_3^1\} \times X_2) &:= m^{X_1}(\{x_2^1, x_3^1\}) \cdot m^{X_2}(X_2) = 0.06. \end{aligned}$$

$m^{X_1 \times X_2}$ clearly satisfies (24).

Furthermore, let y be a variable with finite domain Y, such that $y = f(x^1, \ldots, x^N)$ for some mapping $f : X_1 \times \cdots \times X_N \to Y$. As shown in [25], uncertain knowledge $m^{X_1 \times \cdots \times X_N}$ about variables $x^1, \ldots, x^N$, induces MF m^Y about the value of y defined as

$$m^Y(B) = \sum_{f(A)=B} m^{X_1 \times \cdots \times X_N}(A), \quad \forall B \subseteq Y, \tag{25}$$

with $f(A) = \{f(x_{k^1}^1, \ldots, x_{k^N}^N) | (x_{k^1}^1, \ldots, x_{k^N}^N) \in A\}$ for all $A \subseteq X_1 \times \cdots \times X_N$.

3. Modelling the CVRPED

This section formalises and studies the CVRPED, which is an integer linear program involving uncertainty represented by belief functions. We obtain this problem when customer demands in the CVRP are no longer deterministic or random, but evidential, *i.e.*, the variables d_i, $i = 1, ..., n$, are evidential. Following what has been done for the case of linear programs involving evidential uncertainty [40], we may extend stochastic programming approaches to this integer linear program embedding belief functions: the CCP modelling of the CVRPSD is generalised into a BCP modelling of the CVRPED in Section 3.1, and the recourse modelling of the CVRPSD is generalised into a recourse modelling of the CVRPED in Section 3.2.

Note that, to simplify the exposition, we assume actual customer demands to be positive *integers*, hence the value of the demand of any customer belongs to the set $\Theta = \{1, 2, \ldots, Q\}$. In addition, since the CVPRED involves n evidential variables d_i, $i = 1, ..., n$, with respective domains $\Theta_i := \Theta$, $i = 1, ..., n$, then formally this means that knowledge about customer demands in this problem is represented by a MF m^{Θ^n} on $\Theta^n := \times_{i=1}^{n} \Theta_i$. In practical situations, it may be the case that only marginal knowledge in the form of a MF m^{Θ_i} may be available about the individual demand of each customer i, $i = 1, \ldots, n$. In such case, as explained in Section 2.3.3, m^{Θ^n} can be derived by assuming that these pieces of knowledge about individual customer demands have been provided by independent sources. In other words, if necessary and justified, evidential variables d_i, $i = 1, ..., n$, may be assumed to be independent, similarly as it may be done in the stochastic case. However, let us underline that the BCP and recourse modellings of the CVRPED proposed in the next two sections, are general in that they do not rely on such independence assumption, *i.e.*, they do not need m^{Θ^n} to satisfy a property of the form (24).

3.1. The CVRPED modelled by BCP

A generalisation of the CCP modelling of the CVRPSD to the case of evidential demands is proposed in this section. The model is provided in Section 3.1.1. Important particular cases of this model are discussed in Section 3.1.2. Influences of model parameters and of customer demand ranking on the optimal solution cost, are studied in Sections 3.1.3 and 3.1.4, respectively.

3.1.1. Formalisation

A BCP modelling of the CVRPED amounts to keeping the same optimisation problem described for the CVRP in Section 2.1 except that capacity constraints (7) are replaced by the following *belief*-constraints:

$$Bel\left(\sum_{i=1}^{n} d_i \sum_{j=0}^{n} w_{i,j}^k \leq Q\right) \geq 1-\underline{\beta}, \qquad k=1,\ldots,m, \tag{26}$$

$$Pl\left(\sum_{i=1}^{n} d_i \sum_{j=0}^{n} w_{i,j}^k \leq Q\right) \geq 1-\overline{\beta}, \qquad k=1,\ldots,m, \tag{27}$$

with $\underline{\beta} \geq \overline{\beta}$ and where $1-\underline{\beta}$ (resp. $1-\overline{\beta}$) is the minimum allowable degree of belief (resp. plausibility) that a vehicle capacity is respected on any route.

Remark 1. From (12), constraints (27) are equivalent to

$$Bel\left(\sum_{i=1}^{n} d_i \sum_{j=0}^{n} w_{i,j}^k > Q\right) \leq \overline{\beta}, \qquad k=1,\ldots,m. \tag{28}$$

Hence, constraints (26) and (27) amount to requiring that for any route there is a lot (at least $1-\underline{\beta}$) of support (belief) of respecting vehicle capacity and not a lot (at most $\overline{\beta}$) of support of violating vehicle capacity.

Note that in order to evaluate the belief-constraints (26) and (27), the total demand on every route must be determined by summing all customer demands on that route. Suppose a route R having N clients, then the sum of customer demands on R is obtained using (25), where f is the addition of integers and where $m^{X_1\times\cdots\times X_N}$ is the marginalisation of m^{Θ^n} on the domains of the evidential variables $dr_1,\ldots,dr_N$ associated with the N clients on the route, with X_i the domain of the evidential variable dr_i associated with the i-th client on R. The computation of the total demand on a route as well as the evaluation of constraints (26) and (27) for that route are illustrated by Example 3.

Example 3. Suppose that $\underline{\beta}=0.1$ and $\overline{\beta}=0.05$ and that we have $n=5$ customers and $m=2$ vehicles with capacity limit $Q=15$. Moreover, suppose knowledge about customer demands is represented by MF m^{Θ^n} defined on $\Theta^n=\Theta^5=\Theta_1\times\Theta_2\times\Theta_3\times\Theta_4\times\Theta_5$ by:

$$\begin{aligned} m^{\Theta^5}(\{(2,3,8,4,5),(3,5,6,7,4),(3,4,7,6,2)\}) &= 0.5,\\ m^{\Theta^5}(\{(5,5,6,4,7),(7,6,5,3,4)\}) &= 0.3,\\ m^{\Theta^5}(\{(4,6,7,4,6),(5,5,6,5,7)\}) &= 0.2. \end{aligned} \tag{29}$$

Consider the two routes represented in Fig. 2. Let us compute the sum of the customer demands on the route $(0,4,1,2,0)$, *i.e.*, the route that collects the demand of customer 4, then the demand of customer 1 and finally the demand of customer 2. Call this route R. On this route, there are $N=3$ clients. The first client is client 4, hence according to the above notation, we have $dr_1=d_4$ and $X_1=\Theta_4$. Similarly, we have

$$\begin{aligned} dr_2 &= d_1,\ X_2=\Theta_1,\\ dr_3 &= d_2,\ X_3=\Theta_2. \end{aligned}$$

The marginalisation of m^{Θ^5} on $X_1\times X_2\times X_3$ is the MF $m^{\Theta^5\downarrow X_1\times X_2\times X_3}$ defined as:

$$\begin{aligned} m^{\Theta^5\downarrow X_1\times X_2\times X_3}(\{(4,2,3),(7,3,5),(6,3,4)\}) &= 0.5,\\ m^{\Theta^5\downarrow X_1\times X_2\times X_3}(\{(4,5,5),(3,7,6)\}) &= 0.3,\\ m^{\Theta^5\downarrow X_1\times X_2\times X_3}(\{(4,4,6),(5,5,5)\}) &= 0.2. \end{aligned} \tag{30}$$

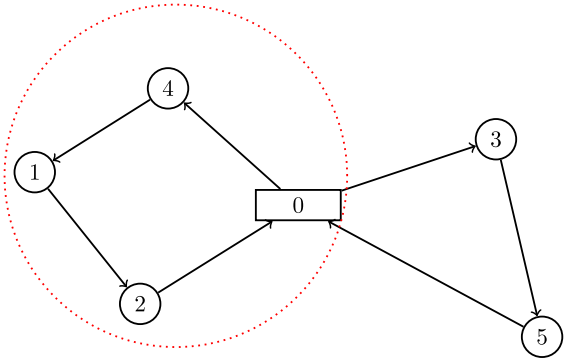


Fig. 2. Evaluation of constraints (26) and (27) on the route $(0, 4, 1, 2, 0)$ (route circled in red).

Now given $m^{\Theta^5 \downarrow X_1 \times X_2 \times X_3}$ and using Equation (25) such that f is the addition of integers, uncertainty on the sum of client demands on route R is represented by a MF denoted $m_{\sum}^{\Theta_R}$ and defined on the domain $\Theta_R := \{1, 2, \ldots, N \cdot Q\} = \{1, \ldots, 45\}$ by:

$$\begin{aligned} m_{\sum}^{\Theta_R}(\{9, 15, 13\}) &= 0.5, \\ m_{\sum}^{\Theta_R}(\{14, 16\}) &= 0.3, \\ m_{\sum}^{\Theta_R}(\{14, 15\}) &= 0.2. \end{aligned} \tag{31}$$

The belief and plausibility that the sum of customer demands on R is smaller or equal than the vehicle capacity Q are then respectively:

$$\begin{aligned} Bel(d_4 + d_1 + d_2 \leq 15) &= m_{\sum}^{\Theta_R}(\{9, 15, 13\}) + m_{\sum}^{\Theta_R}(\{14, 15\}) \\ &= 0.7, \\ Pl(d_4 + d_1 + d_2 \leq 15) &= 1. \end{aligned}$$

Hence, we have

$$\begin{aligned} Bel(dr_1 + dr_2 + dr_3 \leq Q) &< 1 - \underline{\beta} = 0.9, \\ Pl(dr_1 + dr_2 + dr_3 \leq Q) &> 1 - \overline{\beta} = 0.95. \end{aligned}$$

In other words, constraint (27) is satisfied on R but constraint (26) is not, and thus any set of routes containing this route, such as the one shown in Fig. 2, is not a candidate solution.

Suppose further that the number of focal sets of MF $m^{X_1 \times \cdots \times X_N}$ is at most c, then the worst case complexity of evaluating each of the belief constraints (26) and (27) on this route is $\mathcal{O}(N \cdot Q^N \cdot c)$. This latter complexity emerges from the following: the Q^N factor is the maximal number of elements of a focal set of MF $m^{X_1 \times \cdots \times X_N}$. As we have N clients on a route, then for each element of a focal set of MF $m^{X_1 \times \cdots \times X_N}$, the addition of N integers must be performed, this explains $N \cdot Q^N$. The last factor in the complexity which is c, is related to performing the product $N \cdot Q^N$ for the c focal sets of MF $m^{X_1 \times \cdots \times X_N}$. Nonetheless, in a particular case, the worst case complexity drops down significantly:

Remark 2. When the focal sets of $m^{X_1 \times \cdots \times X_N}$ are all Cartesian products of N intervals, *i.e.*, for all $A \subseteq X_1 \times \cdots \times X_N$ such that $m^{X_1 \times \cdots \times X_N}(A) > 0$, we have $A = A^{\downarrow X_1} \times \cdots \times A^{\downarrow X_N}$ with, for $i = 1, \ldots, N$, $A^{\downarrow X_i} = [\![\underline{A_i}; \overline{A_i}]\!]$ for some integers $\underline{A_i}, \overline{A_i} \in X_i$ such that $\underline{A_i} \leq \overline{A_i}$, the worst case complexity is $\mathcal{O}(N \cdot c)$. This complexity to evaluate constraint (26) for route R comes from the fact that (with dr_i the evidential variable associated with the i-th client on R):

$$\begin{aligned} Bel(\sum_{i=1}^{N} dr_i \leq Q) &= \sum \{m^{X_1 \times \cdots \times X_N}(A) | A : \max_{a \in A} f(a) \leq Q\} \\ &= \sum \{m^{X_1 \times \cdots \times X_N}(A) | A : \max_{(a_1, \cdots, a_N) \in A} \sum_{i=1}^{N} a_i \leq Q\} \\ &= \sum \{m^{X_1 \times \cdots \times X_N}(A) | A : \sum_{i=1}^{N} \overline{A_i} \leq Q\}, \end{aligned} \tag{32}$$

that is, at worst for each of the c focal sets of $m^{X_1\times\cdots\times X_N}$, the addition of N integers needs to be performed. The complexity to evaluate constraint (27) is the same since we have

$$Pl(\sum_{i=1}^{N} dr_i \leq Q) = \sum\{m^{X_1\times\cdots\times X_N}(A)|A : \min_{(a_1,\cdots,a_N)\in A}\sum_{i=1}^{N} a_i \leq Q\} \tag{33}$$
$$= \sum\{m^{X_1\times\cdots\times X_N}(A)|A : \sum_{i=1}^{N}\underline{A_i} \leq Q\}.$$

Remark 3. From (32), it is clear that the complexity to evaluate constraint (26) for a given route R, depends on the complexity of finding for each focal set A of $m^{X_1\times\cdots\times X_N}$, the element $(a_1,\cdots,a_N)\in A$ that maximises $\sum_{i=1}^{N} a_i$. Suppose this latter complexity is at worst $\mathcal{O}(M)$, $1\leq M\leq Q^N$, for each focal set. Then, the worst case complexity to evaluate constraint (26) for a route is $\mathcal{O}(N\cdot M\cdot c)$. Remark 2 provides a case, *i.e.*, a particular shape for the focal sets of $m^{X_1\times\cdots\times X_N}$, such that $M=1$. We note that other, more refined yet still leading to tractable values for M, shapes for these focal sets may be considered. For instance, borrowing from robust optimisation [6], suppose that each focal set A of $m^{X_1\times\cdots\times X_N}$ can be written as

$$A = \{(a_1,\cdots,a_N)|a_i\geq\underline{A_i}, a_i\geq\overline{A_i}, \sum_{i=1}^{N}\frac{a_i-\underline{A_i}}{\underline{A_i}}\leq\Gamma\}, \tag{34}$$

for some lower bounds $\underline{A_i}$ and upper bounds $\overline{A_i}$ $i=1,\ldots,N$, and some *uncertainty budget* Γ [6]; budget Γ in (34) limits the sum of the deviations from the minimum demands $\underline{A_i}, i=1,\ldots,N$. Then, maximizing $\sum_{i=1}^{N} a_i$ for each focal set A is done over a more difficult, yet still manageable, shape than in Remark 2. Obviously, similar comments can be made about the complexity of constraint (27).

3.1.2. Particular cases of the BCP modelling of the CVRPED

It is interesting to remark that depending on the values chosen for $\underline{\beta}$ and $\overline{\beta}$ as well as the nature of the evidential demands d_i, $i=1,\ldots,n$, the BCP modelling of the CVRPED may degenerate into simpler or well-known optimisation problems.

In particular, if m^{Θ^n} is Bayesian, *i.e.*, we are dealing really with a CVRPSD, then we have, for $k=1,\ldots,m$,

$$Bel\left(\sum_{i=1}^{n} d_i\sum_{j=0}^{n} w_{i,j}^k\leq Q\right) = Pl\left(\sum_{i=1}^{n} d_i\sum_{j=0}^{n} w_{i,j}^k\leq Q\right), \tag{35}$$

and the BCP modelling of the CVRPED can be converted into an equivalent optimisation problem, which is the CCP modelling of this CVRPSD, with β in (8) set to $\overline{\beta}$.

In contrast, if m^{Θ^n} is categorical and its only focal set is the Cartesian product of n intervals, *i.e.*, we are dealing with a CVRP where each customer demand d_i is only known in the form of an interval $[\![\underline{d_i};\overline{d_i}]\!]$, then the total demand on any given route is also an interval (its endpoints are obtained by summing the endpoints of the interval demands of the customers on the route) and thus for any $k=1,\ldots,m$, $Bel^{\Theta}(\sum_{i=0}^{n} d_i\sum_{j=0}^{n} w_{i,j}^k\leq Q)$ either equals 1 or equals 0, with the former occurring iff $\sum_{i=0}^{n}\overline{d_i}\sum_{j=0}^{n} w_{i,j}^k\leq Q$, and $Pl^{\Theta}(\sum_{i=0}^{n} d_i\sum_{j=0}^{n} w_{i,j}^k\leq Q)$ either equals 1 or equals 0, with the former occurring iff $\sum_{i=0}^{n}\underline{d_i}\sum_{j=0}^{n} w_{i,j}^k\leq Q$. Then, since $\sum_{i=0}^{n}\overline{d_i}\sum_{j=0}^{n} w_{i,j}^k\leq Q\Rightarrow\sum_{i=0}^{n}\underline{d_i}\sum_{j=0}^{n} w_{i,j}^k\leq Q$, the belief-constraints (26) and (27) reduce when $\underline{\beta}<1$ to the following constraints

$$\sum_{i=0}^{n}\overline{d_i}\sum_{j=0}^{n} w_{i,j}^k\leq Q, \qquad k=1,\ldots,m. \tag{36}$$

In other words, in the case of interval demands, the BCP modelling amounts to searching the solution which minimises the overall cost of servicing the customers (1) under constraints (36), *i.e.*, assuming the maximum (worst) possible customer demands, and thus it corresponds to the minimax optimisation procedures encountered in robust optimisation [49].

If m^{Θ^n} is consonant, then we are dealing with a CVRP where uncertainty on customer demand is really of a possibilistic nature, and the BCP modelling may then be connected to fuzzy-based approaches, that is approaches where uncertainty on customer demands is represented by fuzzy sets such as in [50]. In addition, let us remark that if only marginal knowledge in the form of a consonant MF m^{Θ_i} having interval focal sets is available about the individual demand of each customer i, $i=1,\ldots,n$, then, as explained in Section 2.3.3, m^{Θ^n} can be obtained by assuming (if justified) independence of the demands, in which case it will yield a tractable situation thanks to Remark 2, whose conditions are then satisfied (the focal sets of m^{Θ^n} being in this case Cartesian products of intervals). However, m^{Θ^n} may also be derived from these pieces of marginal knowledge by making other assumptions about the demand dependence and in particular by assuming that they are *non-interactive* [25,5] – a more classical independence assumption in the fuzzy setting – in which case the focal sets of m^{Θ^n} will also be Cartesian products of intervals but they will also be nested (m^{Θ^n} will then be consonant).

If $\underline{\beta} = \overline{\beta}$, then constraints (27) can be dropped, that is, only constraints (26) need to be evaluated (if constraints (26) are satisfied then constraints (27) are necessarily satisfied due to the relation between the belief and plausibility functions). As a matter of fact, the BCP approach originally introduced in [40] is of this form (no constraint based on Pl is considered). Most importantly, when $\underline{\beta} = \overline{\beta}$ and the evidential variables d_i, $i = 1 \ldots, n$ are independent, the BCP modelling of the CVRPED can be converted into an equivalent optimisation problem, which is the CCP modelling (with β in (8) set to $\underline{\beta}$) of a CVRPSD where customer demands are represented by independent stochastic variables denoted $d_{\bar{i}}$, $i = 1, \ldots, n$, with associated probability mass function $p_{\bar{i}}$ obtained from $m^{\Theta_i} := m^{\Theta^n \downarrow \Theta_i}$ as follows: for each focal set $A \subseteq \Theta_i$ of m^{Θ_i}, the mass $m^{\Theta_i}(A)$ is transferred to the element $\theta = \max(A)$. Indeed, with such a definition of $p_{\bar{i}}$, it is easy to show that we have, for $k = 1, \ldots, m$,

$$Bel\left(\sum_{i=1}^{n} d_i \sum_{j=0}^{n} w_{i,j}^{k} \leq Q\right) = P\left(\sum_{i=1}^{n} d_{\bar{i}} \sum_{j=0}^{n} w_{i,j}^{k} \leq Q\right). \tag{37}$$

Let us eventually remark that the case $\underline{\beta} = 1 > \overline{\beta}$ is the converse of the case $\underline{\beta} = \overline{\beta}$ in the sense that constraints (26) can be dropped (as they are necessarily satisfied) and only constraints (27) need then to be evaluated. Moreover, in this case, if the evidential variables d_i, $i = 1 \ldots, n$ are independent, the BCP modelling of the CVRPED can be converted into an equivalent optimisation problem, which is the CCP modelling (with β in (8) set to $\overline{\beta}$) of a CVRPSD where customer demands are represented by independent stochastic variables denoted $d_{\underline{i}}$, $i = 1, \ldots, n$, with associated probability mass function $p_{\underline{i}}$ obtained from m^{Θ_i} as follows: for each focal set $A \subseteq \Theta_i$ of m^{Θ_i}, the mass $m^{\Theta_i}(A)$ is transferred to the element $\theta = \min(A)$. Indeed, with such a definition of $p_{\underline{i}}$, it is easy to show that we have, for $k = 1, \ldots, m$,

$$Pl\left(\sum_{i=1}^{n} d_i \sum_{j=0}^{n} w_{i,j}^{k} \leq Q\right) = P\left(\sum_{i=1}^{n} d_{\underline{i}} \sum_{j=0}^{n} w_{i,j}^{k} \leq Q\right). \tag{38}$$

3.1.3. Influence of $\underline{\beta}$, $\overline{\beta}$ and Q on the CVRPED-BCP optimal solution cost

In this section, we study the influence of the parameters $\underline{\beta}$, $\overline{\beta}$ and Q, on the optimal solution cost of the CVRPED modelled via BCP (in the remainder of this article, we will simply say CVRPED-BCP instead of CVRPED modelled via BCP).

The following propositions hold:

Proposition 2. *The optimal solution cost is non-increasing in Q.*

Proof. See Appendix B. □

Proposition 3. *The optimal solution cost is non-increasing in $\underline{\beta}$.*

Proof. See Appendix C. □

Proposition 4. *The optimal solution cost is non-increasing in $\overline{\beta}$.*

Proof. The proof is similar to that of Proposition 3. □

Informally, Propositions 2–4 state that if the decision maker is willing to buy vehicles with a higher capacity or to have vehicle capacity exceeded on any route more often, then he will obtain at least as good (at most as costly) solutions.

3.1.4. Influence of customer demand ranking on the CVRPED-BCP optimal solution cost

In this section, we study the influence on the CVRPED-BCP optimal solution cost, of considering uncertain knowledge representing a more pessimistic estimation of the demand of each customer than currently assumed, that is uncertain knowledge telling that the demand of each customer is higher than currently believed.

Specifically, let m^{Θ^n} and $m_+^{\Theta^n}$ be two MF representing uncertain knowledge about customer demands, such that evidential variables d_i, $i = 1, \ldots, n$, are independent according to both these mass functions. Furthermore, let $m^{\Theta_i} := m^{\Theta^n \downarrow \Theta_i}$ and $m_+^{\Theta_i} := m_+^{\Theta^n \downarrow \Theta_i}$, $i = 1, \ldots, n$. Denote by $\hat{C}_{Q,\underline{\beta},\overline{\beta}}$ and $\hat{C}^{+}_{Q,\underline{\beta},\overline{\beta}}$ the costs of optimal solutions to the CVRPED-BCP when customer demands are known in the form of m^{Θ^n} and $m_+^{\Theta^n}$, respectively, for some $\underline{\beta}$, $\overline{\beta}$ and Q.

The following proposition holds:

Proposition 5. $m^{\Theta_i} \preceq m_+^{\Theta_i}, i = 1, \ldots, n \Rightarrow \hat{C}_{Q,\underline{\beta},\overline{\beta}} \leq \hat{C}^{+}_{Q,\underline{\beta},\overline{\beta}}$.

Proof. See Appendix D. □

Informally, Proposition 5 shows that the more pessimistic knowledge is about customer demands, the greater the cost of the optimal solution.

An immediate consequence of this result is:

Corollary 1. *Assume that the focal sets of MF m^{Θ_i} and $m_+^{\Theta_i}$, $i = 1, \ldots, n$, are all intervals and that $m_+^{\Theta_i}$ can be obtained from m^{Θ_i}, $i = 1, \ldots, n$, as follows: for each interval $A = [\![\underline{A}; \overline{A}]\!]$ such that $m^{\Theta_i}(A) > 0$, the mass $m^{\Theta_i}(A)$ is transferred to the interval $A^+ = [\![\underline{A}; \overline{A} + a^+]\!]$, with $a^+ \in [\![0; Q - \overline{A}]\!]$. Then, $\hat{C}_{Q,\underline{\beta},\overline{\beta}} \leq \hat{C}^+_{Q,\underline{\beta},\overline{\beta}}$.*

Another immediate consequence is:

Corollary 2. *Assume that the focal sets of MF m^{Θ_i} and $m_+^{\Theta_i}$, $i = 1, \ldots, n$, are all intervals and that m^{Θ_i} can be obtained from $m_+^{\Theta_i}$, $i = 1, \ldots, n$, as follows: for each interval $A^+ = [\![\underline{A^+}; \overline{A^+}]\!]$ such that $m_+^{\Theta_i}(A^+) > 0$, the mass $m_+^{\Theta_i}(A^+)$ is transferred to the interval $A = [\![\underline{A^+} - a; \overline{A^+}]\!]$, with $a \in [\![0; \underline{A^+} - 1]\!]$. Then, $\hat{C}_{Q,\underline{\beta},\overline{\beta}} \leq \hat{C}^+_{Q,\underline{\beta},\overline{\beta}}$.*

Remark 4. In both Corollaries 1 and 2, it is easy to show that $m^{\Theta_i} \preceq m_+^{\Theta_i}$, $i = 1, \ldots, n$, which is the reason why these corollaries hold. Note that for Corollary 1, we can also easily show that $m^{\Theta_i} \sqsubseteq m_+^{\Theta_i}$, $i = 1, \ldots, n$, whereas for Corollary 2, we have $m_+^{\Theta_i} \sqsubseteq m^{\Theta_i}$, $i = 1, \ldots, n$. This shows that the CVRPED-BCP optimal solution cost will not necessarily be higher if knowledge about customer demand is less specific. As will be seen in the next section, a different conclusion is reached for the recourse model, and specifically a counterpart to Proposition 5, based on $\sqsubseteq$ rather than $\preceq$, holds.

3.2. The CVRPED modelled by a recourse approach

A recourse approach for the CVRPED is proposed in this section. The general model, extending the one recalled for the CVRPSD in Section 2.2.2, is presented in Section 3.2.1. Then, in Section 3.2.2, we detail how uncertainty on recourse actions is obtained in this model and in Section 3.2.3 we provide a method to compute efficiently this latter uncertainty in an important particular case. Similarly to what has been done for the BCP model, we discuss particular cases of our general model in Section 3.2.4 and study the influence of customer demands specificity on the optimal solution cost in Section 3.2.5.

3.2.1. Formalisation

The CVRPED may be addressed using an extension of the other main approach to modelling stochastic programs, that is the recourse approach. We propose to extend the recourse approach to the CVRPED, for the following policy and assumptions studied for the stochastic case in [12,27,39,23]. Each actual customer demand cannot exceed the vehicle capacity. In addition, when a vehicle arrives at a customer on its planned route, it is loaded with the actual customer demand up to its remaining capacity. If this remaining capacity is sufficient to pick-up the entire customer demand, then the vehicle continues its planned route. However, if it is not sufficient, *i.e.*, there is a failure, then the vehicle returns to the depot, is emptied, goes back to the client to pick-up the remaining customer demand and continues its originally planned route.

Consider a given route R containing N customers and, without lack of generality, that the i-th customer on R is customer i. According to the above setting, a failure cannot occur at the first customer on R. However, it can occur at any other customer on R, and there may even be failure at multiple customers on R (at worst, if the actual demand of each customer is equal to the capacity of the vehicle, failure occurs at each customer except the first one).

Formally, let us introduce a binary variable r_i that equals 1 if failure occurs at the i-th customer on R and 0 otherwise (by problem definition $r_1 = 0$). Then, the possible failure situations that may occur along R may be represented by the vectors $(r_2, r_3, \ldots, r_N) \in \{0, 1\}^{N-1}$. To simplify the exposition, we define the set Ω as the space of binary vectors representing the possible failure situations along R: each failure situation $(r_2, r_3, \ldots, r_N)$ is then a binary vector belonging to $\Omega = \{0, 1\}^{N-1}$. For instance, when R contains only $N = 3$ customers, we have $\Omega = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$, where the binary vectors mean that the vehicle needs to perform a round trip to the depot, respectively, "never", "when it reaches the second customer", "when it reaches the third customer", and "when it reaches both the second and third customers".

Furthermore, let $g : \Omega \to \mathbb{R}^+$ be a function representing the cost of each failure situation $\omega \in \Omega$, with ω being the binary vector $(r_2, r_3, \ldots, r_N)$ representing a failure situation. Since the penalty cost upon failure on customer i is $2c_{0,i}$ (a failure implies a return trip to the depot), the cost associated to failure situation ω is

$$g(\omega) = \sum_{i=2}^{N} r_i 2c_{0,i}. \tag{39}$$

Let m^{Ω} be a MF representing uncertainty towards the actual failure situation occurring on R – as will be shown in the next section, evidential demands may induce such a MF.

Then, adopting a similar pessimistic attitude as in the recourse approach to belief linear programming [40], the upper expected penalty cost $C_{\mathrm{P}}^{*}(R)$ of route R may be obtained using (14) as follows:

$$C_{\mathrm{P}}^{*}(R) = E^{*}(g, m^{\Omega}). \tag{40}$$

Accordingly, the upper expected cost $C_{\mathrm{E}}^{*}(R)$ of route R may be defined as

$$C_{\mathrm{E}}^{*}(R) = C(R) + C_{\mathrm{P}}^{*}(R), \tag{41}$$

with $C(R)$ the cost (2) of travelling along route R when no failure occurs.

The CVRPED under the above recourse policy, may then be modelled using a modified version of the CVRP model of Section 2.1. Specifically, our recourse modelling of the CVRPED aims at

$$\min \sum_{k=1}^{m} C_{\mathrm{E}}^{*}(R_k), \tag{42}$$

subject to constraints (3)–(6), with constraints (5) replaced by constraints (11).

Evaluating the objective function (42) requires the computation for each route, of the MF m^{Ω} representing uncertainty on the actual failure situation occurring on the route. This is detailed in the next section.

3.2.2. Uncertainty on recourses

Consider again a route R containing N customers. In addition, let us first assume that client demands on N are known without any uncertainty, that is we know that the demand of client i, $i = 1, \ldots, N$, is some value $\theta_i \in \Theta$. Then, it is clear that the above recourse policy amounts to the following definition for the binary failure variables r_i:

$$r_i = \begin{cases} 1, & \text{if } q_{i-1} + \theta_i > Q\,, \\ 0, & \text{otherwise,} \end{cases} \qquad \forall i \in \{2, \ldots, N\} \tag{43}$$

where q_j, $j = 1, \ldots, N$, denotes the load in the vehicle after serving the j-th customer such that $q_j = \theta_1$ for $j = 1$ and, for $j = 2, \ldots, N$,

$$q_j = \begin{cases} q_{j-1} + \theta_j - Q\,, & \text{if } q_{j-1} + \theta_j > Q\,, \\ q_{j-1} + \theta_j, & \text{otherwise.} \end{cases} \tag{44}$$

In other words, when it is known that the demand of the i-th customer is θ_i, $i = 1, \ldots, N$, then we have a precise demand vector on R that induces a precise binary failure situation vector $(r_2, r_3, \ldots, r_N)$, with r_i defined by (43). This can be encoded by a function $f : \Theta^N \to \Omega$, s.t. $f(\theta_1, \ldots, \theta_N) = (r_2, r_3, \ldots, r_N)$. For example, suppose we have $N = 3$ customers on route R, with respective demands $\theta_1 = 3, \theta_2 = 3$ and $\theta_3 = 5$, and the vehicle capacity limit is $Q = 5$. In such case, $f(\theta_1, \theta_2, \theta_3)$ implies the failure situation vector $(r_2 = 1, r_3 = 1)$.

In the general case, client demands on R are known in the form of a MF $m^{X_1 \times \cdots \times X_N}$, which is the marginalisation of m^{Θ^n} on the domains of the evidential variables $dr_1, \ldots, dr_N$ associated with the N clients on the route, with X_i the domain of the evidential variable dr_i associated with the i-th client on R. In such case, using (25) with f defined in the preceding paragraph, uncertainty on the actual failure situation on R is represented by a MF m^{Ω} defined as

$$m^{\Omega}(B) = \sum_{f(A)=B} m^{X_1 \times \cdots \times X_N}(A), \quad \forall B \subseteq \Omega. \tag{45}$$

Computing m^{Ω} defined by (45) involves evaluating $f(A)$ for any focal set A of $m^{X_1 \times \cdots \times X_N}$. Evaluating $f(A)$ for some $A \subseteq X_1 \times \cdots \times X_N$, implies $|A|$ (and thus at worst Q^N) times the evaluation of function f at some point $(\theta_1, \ldots, \theta_N) \in \Theta^N$. Hence, computing Equation (45) is generally intractable. Nonetheless, in a particular case, it is possible to compute $f(A)$, and thus Equation (45), with a much more manageable complexity:

Remark 5. When the focal sets of $m^{X_1 \times \cdots \times X_N}$ are all Cartesian products of N intervals, *i.e.*, for all $A \subseteq X_1 \times \cdots \times X_N$ such that $m^{X_1 \times \cdots \times X_N}(A) > 0$, we have $A = A^{\downarrow X_1} \times \cdots \times A^{\downarrow X_N}$ with, for $i = 1, \ldots, N$, $A^{\downarrow X_i} = [\![\underline{A_i}; \overline{A_i}]\!]$, it becomes possible to compute $f(A)$ with a complexity of the order 2^N, as detailed in the next section, and thus in this case if $m^{X_1 \times \cdots \times X_N}$ has at most c focal sets, the worst-case complexity to evaluate Equation (45) is $\mathcal{O}(2^N \cdot c)$.

3.2.3. Interval demands

Let us consider a route R with N customers, such that the demand of customer i, $i = 1, \ldots, N$, is known in the form of an interval of positive integers, which we denote by $[\![\underline{A_i}; \overline{A_i}]\!]$, where $\underline{A_i} \geq 1$ and $\overline{A_i} \leq Q$. In this case, the failure situation on R belongs surely to $f\left([\![\underline{A_1}; \overline{A_1}]\!] \times \cdots \times [\![\underline{A_N}; \overline{A_N}]\!]\right) \subseteq \Omega$. Hereafter, we provide a method to efficiently compute $f\left([\![\underline{A_1}; \overline{A_1}]\!] \times \cdots \times [\![\underline{A_N}; \overline{A_N}]\!]\right)$.

In a nutshell, this method consists in generating a rooted binary tree, which represents synthetically yet exhaustively what can possibly happen on R in terms of failure situations.

More precisely, this tree is based on the following remark. Suppose a vehicle travelling along R and all that is known about its load when it arrives at the i-th customer on R is that its load belongs to an interval $[\![\underline{q};\overline{q}]\!]$. Let us denote by q_i its load after visiting the i-th customer. Then, there are three exclusive cases:

1. either $\overline{q}+\overline{A_i}\leq Q$, hence there will surely be no failure at that customer and all that is known is that $q_i\in[\![\underline{q};\overline{q}]\!]+[\![\underline{A_i};\overline{A_i}]\!]$;
2. or $\underline{q}+\underline{A_i}>Q$, hence there will surely be a failure at that customer and all that is known is that $q_i\in[\![\underline{q};\overline{q}]\!]+[\![\underline{A_i};\overline{A_i}]\!]-Q$;
3. or $\underline{q}+\underline{A_i}\leq Q<\overline{q}+\overline{A_i}$, hence it is not sure whether there will be or not a failure at that customer. However, we can be sure that if there is no failure at that customer, *i.e.*, the sum of the actual vehicle load and of the actual customer demand is lower or equal to Q, then it means that $q_i\in[\![\underline{q}+\underline{A_i};Q]\!]$; and if there is a failure at that customer, then it means that $q_i\in[\![1;\overline{q}+\overline{A_i}-Q]\!]$.

By applying the above reasoning repeatedly, starting from the first customer and ending at the last customer, whilst accounting for and keeping track of all possibilities and their associated failures (or absence thereof) along the way, one obtains a binary tree. The tree levels are associated to the customers according to their order on R. The nodes at a level i represent the different possibilities in terms of imprecise knowledge about the vehicle load after the i-th customer, and they also store whether these imprecise pieces of knowledge about the load were obtained following a failure or an absence of failure at the i-th customer. The pseudo code of the complete tree induction procedure is provided in Algorithm 1 and illustrated afterwards by Example 4.

Algorithm 1 Induction of Recourse Tree (RT).

Input: interval load $[\![\underline{q};\overline{q}]\!]$, Boolean failure variable r, next customer number i
Output: final tree $Tree$

create a root node containing interval load $[\![\underline{q};\overline{q}]\!]$ and Boolean failure r
if $i=N+1$ **then**
  return $Tree=$ {root node}
else if $\overline{q}+\overline{A_i}\leq Q$ **then**
  $[\![\underline{q_L};\overline{q_L}]\!]=[\![\underline{q};\overline{q}]\!]+[\![\underline{A_i};\overline{A_i}]\!]$
  $r_L=0$
  $Tree_L=RT([\![\underline{q_L};\overline{q_L}]\!],r_L,i+1)$
  attach $Tree_L$ as left branch of $Tree$
else if $\underline{q}+\underline{A_i}>Q$ **then**
  $[\![\underline{q_R};\overline{q_R}]\!]=[\![\underline{q};\overline{q}]\!]+[\![\underline{A_i};\overline{A_i}]\!]-Q$
  $r_R=1$
  $Tree_R=RT([\![\underline{q_R};\overline{q_R}]\!],r_R,i+1)$
  attach $Tree_R$ as right branch of $Tree$
else
  $[\![\underline{q_L};\overline{q_L}]\!]=[\![\underline{q}+\underline{A_i};Q]\!]$
  $r_L=0$
  $Tree_L=RT([\![\underline{q_L};\overline{q_L}]\!],r_L,i+1)$
  attach $Tree_L$ as left branch of $Tree$
  $[\![\underline{q_R};\overline{q_R}]\!]=[\![1;\overline{q}+\overline{A_i}-Q]\!]$
  $r_R=1$
  $Tree_R=RT([\![\underline{q_R};\overline{q_R}]\!],r_R,i+1)$
  attach $Tree_R$ as right branch of $Tree$
end if

Example 4. Let us illustrate Algorithm 1 on a route R where $Q=10$ and containing 3 customers, with $[\![4;8]\!]$, $[\![5;7]\!]$ and $[\![7;9]\!]$ the imprecise demands of the first, second and third customers, respectively. Since the demand of the first customer is $[\![4;8]\!]$, and there is no failure by definition at the first customer, and the customer following the first customer is the second customer, the tree is obtained with $RT([\![4;8]\!],0,2)$ and is shown in Fig. 3.

For each leaf of the tree, by concatenating in a vector the Boolean failure variable r_i at level $i=2,\ldots,N$, written on the path from the root to the leaf, we obtain the binary failure situation vector $(r_2,r_3,\ldots,r_N)$. Hence, all the leaves of the tree, yield the subset $B\subseteq\Omega$. For instance, the rightmost leaf of the tree in Fig. 3 yields the failure situation vector $(r_2=1,r_3=1)$, the leftmost leaf yields $(r_2=0,r_3=1)$ and the remaining leaf yields $(r_2=1,r_3=0)$. The tree in this example yields thus the set $B=\{(1,0),(0,1)(1,1)\}$.

Proposition 6. *The set B built using the tree generated by Algorithm 1 verifies $B=f\left([\![\underline{A_1};\overline{A_1}]\!]\times\cdots\times[\![\underline{A_N};\overline{A_N}]\!]\right)$.*

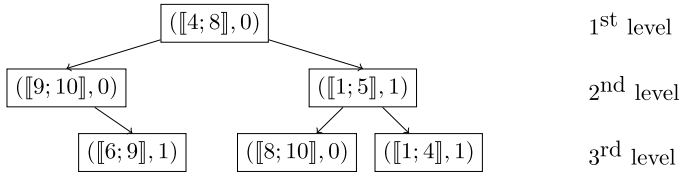


Fig. 3. Recourse tree constructed for Example 4.

Table 1
Travel cost matrix C.

	0	1	2	3
0	$+\infty$	1	1.1	3
1	1	$+\infty$	1	2.1
2	1.1	1	$+\infty$	2.1
3	3	2.1	2.1	$+\infty$

Proof. See Appendix E. □

The maximum number of leaf nodes in the tree is 2^{N-1}. Thus, the algorithmic complexity to obtain set $B \subseteq \Omega$ is of the order 2^N.

3.2.4. Particular cases of the recourse modelling of the CVRPED

In this section, some comments are provided on the behaviour of our recourse modelling, especially with respect to some particular evidential demands.

If m^{Θ^n} is Bayesian, *i.e.*, we are dealing really with a CVRPSD, then m^{Ω} is Bayesian on any given route R. Hence, the upper expected penalty cost $C_{\mathrm{P}}^*(R)$ reduces to the classical (probabilistic) expected value of cost function g with respect to the probability mass function m^{Ω}, and thus our recourse modelling of the CVRPED clearly degenerates into the recourse modelling of the aforementioned CVRPSD.

We showed in Section 3.1.2 that the CVRPED-BCP can be converted, when $\underline{\beta} = \overline{\beta}$ and the evidential variables d_i, $i = 1 \ldots, n$, are independent, into an equivalent CVRPSD modelled via chance constrained programming, by transforming each evidential demand represented by MF m^{Θ_i} into a stochastic demand represented by probability mass function $p_{\overline{i}}$ obtained from m^{Θ_i} by transferring the mass $m^{\Theta_i}(A)$ to the element $\theta = \max(A)$. Example 5 shows that under the recourse approach, this latter transformation cannot be used in general to convert a CVRPED into an equivalent CVRPSD.

Example 5. Suppose we have one available vehicle with capacity limit $Q = 14$, $n = 3$ clients with $[\![2; 8]\!]$, $[\![3; 8]\!]$ and $[\![3; 8]\!]$ the imprecise demands of clients 1, 2 and 3, respectively. The depot is denoted by 0 and the travel cost matrix $C = (c_{i,j})$ where $i, j \in \{0, 1, 2, 3\}$ is shown in Table 1. Under the recourse approach, the optimal solution to this CVRPED instance is the route defined by the path $(0, 3, 2, 1, 0)$ (its upper expected cost is 9.3). Using the above-mentioned transformation to transform the evidential demands into stochastic demands, we obtain $p_{\overline{1}}(8) = 1$, $p_{\overline{2}}(8) = 1$, and $p_{\overline{3}}(8) = 1$, and under the recourse approach the optimal solution to this CVRPSD instance is either the route defined by the path $(0, 2, 1, 3, 0)$ or $(0, 3, 1, 2, 0)$ (the expected cost of each one of these routes being 9.2), which are different from the optimum found for the CVRPED.

Furthermore, let us remark that for a given route R containing N clients whose demands are known in the form of a MF $m^{X_1 \times \cdots \times X_N}$, its upper expected cost $C_{\mathrm{E}}^*(R)$ is necessarily reached for some probability measure $P^{X_1 \times \cdots \times X_N}$ belonging to the set $\mathcal{P}(m^{X_1 \times \cdots \times X_N})$ of probability measures compatible with $m^{X_1 \times \cdots \times X_N}$ and defined by

$$\mathcal{P}(m^{X_1 \times \cdots \times X_N}) = \{P | \forall A \subseteq X_1 \times \cdots \times X_N, Bel^{X_1 \times \cdots \times X_N}(A) < P(A)\}. \tag{46}$$

However, this measure cannot be determined easily in advance; it can be exhibited once possible recourses on the route and their associated costs are known. Moreover, we note that it is impossible to find a sensible transformation of evidential demands into stochastic demands such that each solution has the same costs in both the evidential and stochastic approaches, as shown by Example 6.

Example 6. Suppose a problem where we have $m = 2$ available vehicles and $n = 5$ clients with demands represented by MF m^{Θ^n} defined as

$$m^{\Theta^n}([\![7]\!] \times [\![6]\!] \times [\![1; 2]\!] \times [\![1]\!] \times [\![1]\!]) = 1. \tag{47}$$

Consider the two following possible solutions, containing each two routes:

$$S_1 = \{R_1^1, R_2^1\},$$

$$S_2 = \{R_1^2, R_2^2\},$$

with $R_1^1 = (0, 1, 2, 3, 0)$, $R_2^1 = (0, 4, 5, 0)$ and $R_1^2 = (0, 1, 2, 3, 5, 0)$, $R_2^2 = (0, 4, 0)$. The upper expected cost of S_i, $i = 1, 2$, is $C_E^*(S_i) := C_E^*(R_1^i) + C_E^*(R_2^i)$. Consider now the possible transformations of the evidential demands m^{Θ^n} into stochastic demands p^{Θ^n}: it seems sensible that p^{Θ^n} be chosen in the set of probability distributions compatible with m^{Θ^n}, that is in

$$\mathcal{P}(m^{\Theta^n}) = \{(p^{\Theta^n}(7, 6, 1, 1, 1) = \alpha, p^{\Theta^n}(7, 6, 2, 1, 1) = 1 - \alpha) | \alpha \in [0, 1]\}. \tag{48}$$

Let $C_E^\alpha(S_i)$ denote the expected cost of solution S_i, $i = 1, 2$, under stochastic demands represented by probability distribution $p_\alpha^{\Theta^n}$ defined by $p_\alpha^{\Theta^n}(7, 6, 1, 1, 1) = 1 - \alpha$, $p_\alpha^{\Theta^n}(7, 6, 2, 1, 1) = \alpha$, for some $\alpha \in [0, 1]$. Assume that customer 5 is further away from the depot than customer 3, that is $c_{0,5} > c_{0,3}$. Then, we obtain

$$C_E^*(S_1) - C_E^\alpha(S_1) = 2c_{0,3} \cdot (1 - \alpha) \tag{49}$$

and

$$C_E^*(S_2) - C_E^\alpha(S_2) = 2c_{0,5} \cdot \alpha + 2c_{0,3} \cdot \alpha. \tag{50}$$

The upper expected cost $C_E^*(S_1)$ of S_1 is thus reached for probability distribution $p_1^{\Theta^n}$, whereas the upper expected cost $C_E^*(S_2)$ of S_2 is reached for probability distribution $p_0^{\Theta^n}$. Hence, for S_1 to have the same costs in both the evidential and stochastic approaches and for S_2 to also have the same costs in both the evidential and stochastic approaches, different transformations of evidential demands into stochastic demands must be used.

The objective function (42) of our recourse model relies on optimising against $C_{\mathrm{E}}^*(R)$ which is the upper, *i.e.*, worst, *expected* cost of a route. In particular, if m^{Θ^n} is categorical, then $C_{\mathrm{E}}^*(R)$ is the worst *possible* cost of R. Hence, optimising (42) has some similarities with the protection against the worst case popular in robust optimisation [49].

Though, another approach could be followed [52], where one optimises against the lower, *i.e.*, best, expected cost $C_{*\mathrm{E}}(R) = C(R) + C_{*\mathrm{P}}(R)$, where $C_{*\mathrm{P}}(R)$ is evaluated using (13) such that $C_{*\mathrm{P}}(R) = E_*(g, m^\Omega)$. This approach is appropriate when we are interested in the most optimistic solution. More complex decision schemes could also be considered, such as interval dominance [52], which would rely on $C_{\mathrm{E}}^*(R)$ and $C_{*\mathrm{E}}(R)$ and yield in general a set of optimal (non-dominated) solutions. Borrowing from what is done in label ranking [22], an interesting study would then be to identify from the set of non-dominated solutions, some parts of routes that would be more relevant (or preferred) to be included in a solution, over some irrelevant ones. This is left for future work.

Finally, another interesting particular case is when m^{Θ^n} is obtained from marginal knowledge about individual customer demands in the form of consonant MF m^{Θ_i}, $i = 1, \ldots, n$, having interval focal sets and the assumption that the demands are independent or non-interactive. As recalled in Section 3.1.2, the focal sets of m^{Θ^n} are then Cartesian products of intervals under both assumptions, in which case the recourse modelling of the CVRPED becomes tractable according to Remark 5.

3.2.5. Influence of customer demand specificity on the CVRPED-recourse optimal solution cost

In this section, we study the behaviour of the optimal solution cost of the CVPRED modelled via the recourse approach (in the remainder of this article, we will simply say CVRPED-recourse rather than the CVRPED modelled via the recourse approach), when knowledge specificity about customer demands decreases.

Specifically, let m^{Θ^n} and $m_\star^{\Theta^n}$ be two MF representing uncertain knowledge about customer demands, such that evidential variables d_i, $i = 1, \ldots, n$, are independent according to both these mass functions. Furthermore, let $m^{\Theta_i} := m^{\Theta^n \downarrow \Theta_i}$ and $m_\star^{\Theta_i} := m_\star^{\Theta^n \downarrow \Theta_i}$, $i = 1, \ldots, n$. Denote by $\hat{C}_{Rec}$ and $\hat{C}_{Rec}^\star$ the costs of optimal solutions to the CVRPED-recourse when customer demands are known in the form of m^{Θ^n} and $m_\star^{\Theta^n}$, respectively.

The following proposition holds:

Proposition 7. $m^{\Theta_i} \sqsubseteq m_\star^{\Theta_i}, i = 1, \ldots, n \Rightarrow \hat{C}_{Rec} \leq \hat{C}_{Rec}^\star$.

Proof. See Appendix F. □

Informally, Proposition 7 shows that the less specific knowledge is about customer demands, the greater the cost of the optimal solution.

An immediate consequence of this result is:

Corollary 3. *Assume that for $i = 1, \ldots, n$, $m_\star^{\Theta_i}$ is built from m^{Θ_i} as follows: for each $A \subseteq \Theta_i$ such that $m^{\Theta_i}(A) > 0$, the mass $m^{\Theta_i}(A)$ is transferred to a subset $A^\star$ such that $A \subseteq A^\star \subseteq \Theta_i$. Then, we have $\hat{C}_{Rec} \leq \hat{C}_{Rec}^\star$.*

Proposition 7 and Propositions 2–5 provide theoretical properties of the CVRPED solutions obtained under the recourse and the BCP approaches, respectively, when using exact optimisation methods. For now, such methods can not solve large instances of the CVRP, from which the CVRPED derives. As a matter of fact, Section 4 reports solution strategies to these two CVRPED models using a metaheuristic algorithm.

4. Solving the CVRPED

This section presents a metaheuristic algorithm to solve the two proposed CVRPED models and reports some experimental tests of this algorithm. More precisely, a simulated annealing algorithm for both the BCP and the recourse models is first described in Section 4.1. Next, benchmarks for the CVRPED are presented in Section 4.2. Finally, experimental tests using these benchmarks are provided in Sections 4.3 and 4.4 for the BCP and the recourse models, respectively.

4.1. A simulated annealing algorithm for the CVRPED

As for many NP-hard problems like the CVRP, exact solution methods might need a prohibitively large time to solve large instances. When uncertainty is introduced into the CVRP, the problem can easily become even more difficult. Metaheuristics are algorithms built on general (meta) concepts that search the solution space in a reasonable time and thus may be employed as a successful alternative to solve such a combinatorial optimisation problem. Various metaheuristic algorithms exist like: simulated annealing [37], genetic algorithms [36], tabu search [30,31], etc. In this study, we will use simulated annealing, which is a well-known local search optimisation method. Generally speaking, a local search algorithm moves iteratively from solution to solution in the space of candidate solutions (the search space) by applying local changes, until a satisfying near-optimal solution is found.

Indeed, to try to find the global minimum of the cost function, the simulated annealing moves from solution to solution using either descent (improving) moves or deterioration moves, hoping that these non-improving moves will eventually help the process escape local optima [37]. One can say that it starts from a known initial configuration of a system (candidate solution) with a high temperature and then uses general neighbourhood search strategies to explore other candidate solutions by following neighbourhood transitions (moves). For every temperature, the configuration of the system (candidate solution) is rearranged (transformed) by a series of neighbourhood moves. A rearranged configuration becomes the new candidate solution with a probability depending on the current temperature, *i.e.*, the lower the temperature, the lower the probability to accept non-improving moves. The temperature of the system is gradually lowered, and the process continues until reaching the freezing temperature of the system.

Our instantiation of simulated annealing for the CVRPED is provided by Appendix G (see in particular its pseudo-code given by Algorithm 2). The same pseudo-code is used for both the BCP and the recourse approaches. One can find in Algorithm 2 a modelling technique parameter MD that takes the value $MD =$ "BCP" when the CVRPED-BCP is solved. In this case, the initial configuration is generated either randomly or using a first-fit greedy approach[3] while respecting the constraints of the BCP model from Section 3.1.1. The `neighbourhood_configuration(...)` routine generates a neighbourhood configuration that respects the CVRPED-BCP constraints, and is based on the following two operators that are applied consecutively on the current configuration $\mathcal{C}$ at each iteration. These operators are called `fix_minimum` and `replace_highest_average`. They are described below and illustrated using examples in Appendix H.

- `Fix_minimum`: This operator is applied on 80% of the iterations. It is based on selecting and freezing the positions in routes of the five customers, with the shortest distances to their right side customer.[4] This is done by computing distances between each pair of consecutive customers on all routes, including distances to the depot. Accordingly, `fix_minimum` selects the five smallest distances values and fixes their corresponding left side customers. Next, `fix_minimum` selects five random customers that exclude the depot and the customers fixed before, and removes them from their route. Subsequently, every customer removed will be inserted in a random route, while satisfying the problem constraints. The insert position of each customer on the selected route is determined based on the shortest distance separating it from its new left side customer.
- `Replace_highest_average`: This neighbourhood operator calculates the average distance separating every customer from its neighbours in a current route configuration. Computing the average distance for a client i reduces to calculating $\frac{c_{i-1,i}+c_{i,i+1}}{2}$, assuming clients $i-1$, i and $i+1$ are consecutive in the route.[5] Afterwards, `replace_highest_average` selects five customers having the five highest average distances and removes them from their routes. The removed clients are then randomly inserted in the available routes, as long as the problem constraints are respected. Furthermore, every removed customer i is inserted into the route position leading to the smallest average distance that will separate this customer from its new $(i-1)$ th and $(i+1)$ th neighbours.

[3] A blind algorithm that inserts clients in routes in turn: each client is inserted in the first route that can serve it.

[4] Given the pair of customer $< i-1, i >$, the right side customer is the i-th customer.

[5] The notation $c_{i-1,i}$ indicates the travel cost between $i-1$ and i, as defined in Section 2.1, which in our algorithm is assumed to be the euclidean distance separating customers.

If the above operators do not lead to a new neighbourhood configuration that satisfies the CVRPED-BCP constraints, we apply the operators a second time. If the second attempt fails as well, then the configuration $\mathcal{C}$ is not modified, *i.e.*, $C^* = \mathcal{C}$.

The complexity of an iteration for the BCP model emerges from evaluating the CVRPED-BCP constraints for the neighboring configuration, in particular the belief constraints (26) and (27), the complexity of which is provided in Section 3.1.1.

We solve the CVRPED-recourse with the same simulated annealing, by setting the modelling technique parameter to $MD =$ "recourse" in Algorithm 2. The `initial_config(...)` method proceeds by generating initial configurations that are subject to the CVRPED-recourse constraints mentioned in Section 3.2.1, using either a first fit greedy approach or randomly. The routine `neighbourhood_configuration(...)` applies three consecutive neighbourhood operators to each iteration: `fix_minimum`, `replace_highest_average` and `flip_ route`.

- `Fix_minimum` and `replace_highest_average` operate similarly as in the BCP case, except that the problem constraints are now the CVRPED-recourse constraints. Recall that the recourse model lifts the capacity constraints, in the sense that any capacity excess (overflow) is addressed by the objective function using recourse decisions. This means that as `fix_minimum` and `replace_highest_average` are iteratively applied by the simulated annealing, we may end up with routes holding excessive total demands. Consequently, these routes will systematically fail, while other routes will hold limited total customer demands. Therefore, we incorporate to each of the `fix_minimum` and `replace_highest_average` operators, a method that maintains relatively balanced total demands on routes, during the process of moving customers from a route to another. More specifically, before a customer is inserted into a new route, total customer demands on each route yields a probability, indicating if a route is favourable for servicing an additional customer. The probability associated to each route is inversely proportional to the total customer demands on a route, *i.e.*, the smaller the total customer demands is on a route, the more it is probable to choose this route to include an additional customer.
- The operator `flip_route` is applied on 25% of the iterations. It reverses the order of a route if this improves its upper expected cost. Indeed, a route R having the path $(0, 1, \ldots, N, 0)$ and its reverse R^{-1} with the path $(0, N, \ldots, 1, 0)$ do not have necessarily the same upper expected penalty cost.

The complexity of an iteration for the recourse model corresponds to evaluating the uncertainty on the recourses of each one of the m routes of a configuration $\mathcal{C}$, the complexity of which is given in Section 3.2.2.

Let us finally emphasize that the evidential approaches proposed in this paper are not limited to using simulated annealing or metaheuristics. Some of the exact methods used for the CVRPSD could be extended to these approaches. This is the case, for instance, of the Column Generation (CG) algorithm proposed in [12] for the recourse modelling of the CVRPSD. A CG algorithm relies on a linear program with prohibitively many variables (columns) associated to feasible routes. One could use our evidential approach to evaluate the total cost (base cost plus recourse cost) of each generated column (route), especially when the columns are constructed by Dynamic Programming (DP).

More exactly, each column (route) can be constructed by generating a cycle in a graph of DP states associated to subproblems satisfying the Bellman principle of optimality. Assuming joint focal sets about customer demands are Cartesian products of intervals, one could define a DP state $\left(v, \left[\underline{q_v}, \overline{q_v}\right]\right)$ for each vertex $v \in V$ and for each interval of residual supply $[\underline{q_v}, \overline{q_v}]$ at v (quantity remaining in a vehicle after servicing clients up to v and performing recourses). The total cost of a route in a given state can be determined, for instance, as the sum of the traversed edges plus the upper expected cost of the recourse actions. In a loose sense, this is similar to the DP approach from [12] that constructs a state for each vertex v and for each possible value of the cumulative expected demand (of all the clients in the route up to v) and of the variation of this cumulative demand.

4.2. The CVRPED benchmarks

We generated two instance sets CVRPED and CVRPED$^+$, based on the set A of the Augerat test bed for the CVRP [46]. Each instance in these two sets corresponds to an instance in Augerat set A and has the same customer coordinates and capacity limit as this instance.

For each instance of the first CVRPED set, the knowledge on customer demands m^{Θ^n} is obtained by assuming that the evidential client demands $d_i, i = 1, \ldots, n$ of this instance are independent. Moreover, each d_i is associated to the mass function m^{Θ_i} defined by

$$m^{\Theta_i}(\{d_i^{det}\}) = 0.8, \tag{51}$$

$$m^{\Theta_i}([\underline{z_i}, \overline{z_i}]) = 0.2, \tag{52}$$

with d_i^{det} the original deterministic demand of client i in the corresponding instance of Augerat set A, and with $\underline{z_i}$ and $\overline{z_i}$ drawn at random in $(d_i^{det}, Q\,]$ and $[\underline{z_i}, Q\,]$, respectively.

Table 2
Results of the simulated annealing algorithm for the CVRPED-BCP using the CVRPED instances.

Instance l-A-nn-mm:	$\underline{\beta} = 0.4, \overline{\beta} = 0.25$			$\underline{\beta} = 0.2, \overline{\beta} = 0.15$		
l instance id, n clients, m vehicles	Best cost	Std. dev.	Avg. runtime	Best cost	Std. dev.	Avg. runtime
1-A-n32-m12	1418,3	3,7	3881s.	1850,9	5,3	3733s.
2-A-n33-m13	1055,3	0	4199s.	1491,6	17,5	4496s.
3-A-n33-m13	1073,1	6	4495s.	1480,2	0,6	4549s.
4-A-n34-m14	1320,6	0,1	3818s.	1749,1	0	3852s.
5-A-n36-m12	1318,9	2,7	5316s.	1718,6	0,2	4914s.
6-A-n37-m13	1110,6	5	4918s.	1358,8	34,5	6158s.
7-A-n37-m14	1597,9	0,5	4135s.	2113,9	2,8	3756s.
8-A-n38-m13	1154,5	0,9	5041s.	1571,1	5,1	5002s.
9-A-n39-m15	1485	8,1	4654s.	1944,8	0,9	4622s.
10-A-n39-m14	1403,9	8,6	4894s.	1906,9	0,2	5108s.
11-A-n44-m17	1693,4	10,1	4956s.	2158,2	1,2	4951s.
12-A-n45-m17	1660,3	0,1	5093s.	2184,8	5,1	5169s.
13-A-n45-m18	1890,1	5,7	4991s.	2573,2	0,7	5211s.
14-A-n46-m17	1552,2	5,4	5323s.	1980,3	38	5707s.
15-A-n48-m17	1872,4	10,8	5996s.	2397,5	3,6	6395s.
16-A-n53-m19	1806,1	11,9	7405s.	2358,2	10	6928s.
17-A-n54-m19	2052,6	13,8	7578s.	2636,8	60,8	6747s.
18-A-n55-m22	1755,3	9,5	6310s.	2352,6	9,8	6172s.
19-A-n60-m22	2263,9	17	8169s.	2969,1	34,8	7449s.
20-A-n61-m24	1793,8	9,9	6965s.	2345	42,3	7319s.
21-A-n62-m22	2532,3	19,1	8212s.	3207,2	32,2	7752s.
22-A-n63-m24	2946,9	14,6	7164s.	3918,9	22	7216s.
23-A-n63-m25	2179	8,9	6969s.	2881,1	6,9	7238s.
24-A-n64-m23	2629,1	16,8	7979s.	3261,5	13,6	7848s.
25-A-n65-m25	2214,7	16,4	7537s.	3070,9	6	7601s.
26-A-n69-m25	2056,5	11,1	8876s.	2668,1	52,9	8377s.
27-A-n80-m27	3507,2	21,5	11110s.	4524,9	21,5	9751s.

For each instance of the second set CVRPED$^+$, the evidential client demands d_i are also assumed to be independent, and their associated mass function is denoted by $m_+^{\Theta_i}$ and defined from m^{Θ_i} as follows:

$$m_+^{\Theta_i}([d_i^{det}, d_i^{det} + a_i^+]) = 0.8, \tag{53}$$

$$m_+^{\Theta_i}([\underline{z_i}, \overline{z_i}]) = 0.2, \tag{54}$$

with a_i^+ drawn randomly in $[0, \underline{z_i} - d_i^{det} - 1]$. Note that $m^{\Theta_i} \sqsubseteq m_+^{\Theta_i}$ and $m^{\Theta_i} \preceq m_+^{\Theta_i}$, $i = 1, \ldots, n$.

In the next sections, an experimental study based on the CVRPED and CVRPED$^+$ instances[6] is presented for the BCP and recourse models solved using the algorithm described in Section 4.1. The programs were written in Java and the experiments were conducted on the 5 nodes of a cluster. The configuration of each node is as follows: 2 processors Intel R Xeon R E5-2630 v3 with 8 cores per processor having a 48GB memory shared between the 2×8 cores of the node. Each instance was executed on one core that has a memory of 2.8GB.

4.3. Experimental study for the CVRPED-BCP

Our set of experiments on the CVRPED-BCP involve varying the values $\underline{\beta}$ and $\overline{\beta}$ involved in the constraints (26) and (27) for the CVRPED and the CVRPED$^+$ instances separately, where for each variation each instance was solved 30 times.

4.3.1. The CVRPED-BCP cost variation based on $\underline{\beta}$ and $\overline{\beta}$

In this first part, we show results of our experiments for the CVRPED and the CVRPED$^+$ instances in Tables 2 and 3, respectively. Indeed, we solved the CVRPED-BCP for two different values that we chose for the pair $(\underline{\beta}, \overline{\beta})$, such that $\underline{\beta} \geq \overline{\beta}$, and both values were employed for the CVRPED and the CVRPED$^+$ instances, separately. The columns figuring in each of these tables, are explained in the following. The first column is the name of each instance. The first field in this column (l in Table 2 and l^+ in Table 3), exposes the identification number of an instance and the second field "A" stands for aleatory indicating that the coordinates of the problem graph vertices were generated randomly in the original Augerat set A [3]. The third field designates the number n of vertices, while the last field provides the number m of vehicles. The "Best cost", "Std. dev." and "Avg. runtime" columns show respectively, the best solution, the standard deviation and the average running time that we obtained for each indicated value of the pair $(\underline{\beta}, \overline{\beta})$.

[6] Our data sets can be found on the web site of the LGI2A laboratory [34].

Table 3
Results of the simulated annealing algorithm for the CVRPED-BCP using the CVRPED$^+$ instances.

Instance l^+-A-nn-mm:	$\underline{\beta}=0.4, \overline{\beta}=0.25$			$\underline{\beta}=0.2, \overline{\beta}=0.15$		
l instance id, n clients, m vehicles	Best cost	Std. dev.	Avg. runtime	Best cost	Std. dev.	Avg. runtime
1^+-A-n32-m16	1830,8	28,3	3087s.	2225,4	0	3565s.
2^+-A-n33-m16	1428,8	0,7	4048s.	1676,9	0,4	3790s.
3^+-A-n33-m14	1196	0	4330s.	1502	24,2	4604s.
4^+-A-n34-m16	1596	0,01	3389s.	1993	0	3719s.
5^+-A-n36-m16	1755,3	4	3928s.	2145,7	0	4178s.
6^+-A-n37-m18	1379,2	8,7	5179s.	1761,2	0	4347s.
7^+-A-n37-m19	1957	0	3344s.	2542,6	0	4112s.
8^+-A-n38-m16	1437,4	2,2	4017s.	1846,4	2,2	4479s.
9^+-A-n39-m19	1915,5	11,2	3915s.	2203,5	0	4089s.
10^+-A-n39-m18	1759	4,2	4255s.	2148,1	0	4140s.
11^+-A-n44-m26	2234,2	1,5	4218s.	2796,6	0,4	5698s.
12^+-A-n45-m22	2165,7	3.1	4360s.	2690,8	0	4601s.
13^+-A-n45-m22	2287,7	0,9	4339s.	3099,4	0	4783s.
14^+-A-n46-m22	1950,4	6,7	4603s.	2690,1	0	4795s.
15^+-A-n48-m22	2359,8	5,9	5742s.	2956,3	1,8	5500s.
16^+-A-n53-m25	2411,2	9,4	5747s.	3199,8	0	5883s.
17^+-A-n54-m24	2591	11,4	6025s.	3165,7	0	5746s.
18^+-A-n55-m27	2237,1	5,1	5716s.	2803,1	0	5493s.
19^+-A-n60-m27	2744,8	9,5	6224s.	3415,5	6,4	6548s.
20^+-A-n61-m30	2313,4	9,1	6125s.	3059,2	0	5877s.
21^+-A-n62-m31	3217,7	8,9	6814s.	4291	3,1	6276s.
22^+-A-n63-m30	3833	9,7	6057s.	4942,1	4,9	6257s.
23^+-A-n63-m31	2755	5,8	5757s.	3638,6	0,4	6127s.
24^+-A-n64-m29	3311,2	27,9	6656s.	4123,1	1,7	6848s.
25^+-A-n65-m34	2748,1	7,7	6762s.	3509,1	9	6842s.
26^+-A-n69-m33	2573,4	8,5	7135s.	3300,5	41,1	7125s.
27^+-A-n80-m38	4995,8	14,7	7444s.	5986,3	13,5	7612s.

We notice the costs of the best solutions obtained with $\underline{\beta}=0.4$, $\overline{\beta}=0.25$ are lower than the costs of the best solutions obtained with $\underline{\beta}=0.2$, $\overline{\beta}=0.15$ in Table 2 as well as in Table 3, that is, the most constraining pair $(\underline{\beta}, \overline{\beta})$ induces the worst costs, as can be expected from Propositions 3 and 4. This shows that while our solving algorithm is not an exact optimisation method, it does exhibit experimentally a sound behaviour with respect to parameters $\underline{\beta}$ and $\overline{\beta}$.

4.3.2. The CVRPED-BCP cost variation based on client demand ranking

This section compares the BCP results on the CVRPED instances with the BCP results on the CVRPED$^+$ instances. Specifically, Table 4 compares the best costs from Table 2 (CVRPED instances) with the bests costs from Table 3 (CVRPED+ instances), for the same instance id.

Recall that for each client i in an l instance, its MF m^{Θ_i} is at least as small as the associated MF to client i in the l^+ instance, *i.e.*, $m^{\Theta_i} \preceq m_+^{\Theta_i}$, $i=1,\ldots,n$. Proposition 5 (and more specifically Corollary 1) predicted an increase in the cost of an optimal solution to the CVRPED-BCP when knowledge about clients demands is more pessimistic. We can observe this behaviour in the results presented in Table 4: for each pair $(\underline{\beta}, \overline{\beta})$ the best cost obtained with the CVRPED$^+$ instances is higher than the one obtained with the CVRPED instances. This constitutes another experimental validation of the behaviour of our algorithm.

4.4. Experimental study for the CVRPED-recourse

In the experiments conducted for the CVRPED-recourse, we used the same generated CVRPED and CVRPED$^+$ instances used for the CVRPED-BCP experiments. Our results are reported in Table 5.

The columns "Instance id l" and "Instance id l^+" in this table represent the CVRPED and the CVRPED$^+$ instances id, respectively. Each one of these instances was solved 30 times and the best, average and standard deviation of the costs along with the average running times are reported in the respective columns "Best cost", "Avg cost", "Stand. dev." and "Avg. runtime" for the CVRPED and the CVRPED$^+$ instances, separately. In the "Penalty cost" column for the CVRPED instances (respectively the "Penalty cost" column for the CVRPED$^+$ instances), the contribution of the expected penalty costs to the overall costs of the best solutions to the CVRPED instances (respectively the CVRPED$^+$ instances) is provided as percentages. In the case of the CVRPED instances, it varies between 16% to 25%. As for the CVRPED$^+$ instances, it varies between 11% and 23%.

Recall that for each client i in an l instance, its MF m^{Θ_i} is at least as specific as the associated MF to client i in the l^+ instance, *i.e.*, $m^{\Theta_i} \sqsubseteq m_+^{\Theta_i}$, $i=1,\ldots,n$. As expected from Proposition 7 (and more specifically from Corollary 3), best costs obtained with the CVRPED$^+$ instances are higher than those obtained with the CVRPED instances, which shows that our algorithm for the recourse model exhibits experimentally also a sound behaviour.

Table 4
Results of the simulated annealing algorithm for the CVRPED-BCP for the CVRPED and CVRPED$^+$ instances.

CVRPED instances			CVRPED$^+$ instances		
Instance id l	Best cost $\underline{\beta} = 0.4, \overline{\beta} = 0.25$	Best cost $\underline{\beta} = 0.2, \overline{\beta} = 0.15$	Instance id l^+	Best cost $\underline{\beta} = 0.4, \overline{\beta} = 0.25$	Best cost $\underline{\beta} = 0.2, \overline{\beta} = 0.15$
1	1418,3	1850,9	1^+	1830,8	2225,4
2	1055,3	1491,6	2^+	1428,8	1676,9
3	1073,1	1480,2	3^+	1196	1502
4	1320,6	1749,1	4^+	1596	1993
5	1318,9	1718,6	5^+	1755,3	2145,7
6	1110,6	1358,8	6^+	1379,2	1761,2
7	1597,9	2113,9	7^+	1957	2542,6
8	1154,5	1571,1	8^+	1437,4	1846,4
9	1485	1944,8	9^+	1915,5	2203,5
10	1403,9	1906,9	10^+	1759	2148,1
11	1693,4	2158,2	11^+	2234,2	2796,6
12	1660,3	2184,8	12^+	2165,7	2690,8
13	1890,1	2573,2	13^+	2287,7	3099,4
14	1552,2	1980,3	14^+	1950,4	2690,1
15	1872,4	2397,5	15^+	2359,8	2956,3
16	1806,1	2358,2	16^+	2411,2	3199,8
17	2052,6	2636,8	17^+	2591	3165,7
18	1755,3	2352,6	18^+	2237,1	2803,1
19	2263,9	2969,1	19^+	2744,8	3415,5
20	1793,8	2345	20^+	2313,4	3059,2
21	2532,3	3207,2	21^+	3217,7	4291
22	2946,9	3918,9	22^+	3833	4942,1
23	2179	2881,1	23^+	2755	3638,6
24	2629,1	3261,5	24^+	3311,2	4123,1
25	2214,7	3070,9	25^+	2748,1	3509,1
26	2056,5	2668,1	26^+	2573,4	3300,5
27	3507,2	4524,9	27^+	4995,8	5986,3

Table 5
Results of the simulated annealing algorithm for the CVRPED-recourse for the CVRPED and CVRPED$^+$ instances.

CVRPED instances						CVRPED$^+$ instances					
Instance id l	Best cost	Penalty cost	Avg cost	Stand. dev.	Avg. runtime	Instance id l^+	Best cost	Penalty cost	Avg cost	Stand. dev.	Avg. runtime
1	1750,3	16,8%	1783,9	16,1	1958s.	1^+	2252,6	18,3%	2283,5	13,2	1272s.
2	1327,5	16,2%	1353,2	13,6	1704s.	2^+	1650,6	17,7%	1676	9,6	1329s.
3	1296,1	18%	1338,8	16,4	1642s.	3^+	1490,3	16,6%	1510,6	11,3	1540s.
4	1661,9	19,8%	1698,7	24,6	1728s.	4^+	1999,6	19,7%	2044,9	18,7	1428s.
5	1670,1	24,2%	1741,9	29	2673s.	5^+	2205,3	16,9%	2247,3	18,6	1554s.
6	1391,2	20,1%	1425,8	12,5	3586s.	6^+	1697,5	14,9%	1737	13,2	1612s.
7	1895,6	24,4%	1947,3	21,2	2286s.	7^+	2561,5	17%	2593,8	17,5	1382s.
8	1493,8	16,1%	1525,6	15,9	2450s.	8^+	1769,7	16,6%	1802,9	18,9	1686s.
9	1851	21,1%	1897,6	27,4	2580s.	9^+	2319,9	19,9%	2355	20,5	1783s.
10	1715,2	22,2%	1755,6	22,9	3264s.	10^+	2099,3	20,7%	2146,3	20,8	1863s.
11	2127,8	20,7%	2216,7	25,3	2349s.	11^+	2858,5	11,8%	2889	15,5	1491s.
12	2147,1	17%	2193,7	21,1	2344s.	12^+	2667,9	18 %	2705,2	22	1808s.
13	2530,2	22,7%	2629,7	33,6	2427s.	13^+	3084,7	15,7%	3145,9	29,5	1755s.
14	1994,9	24,9%	2089,4	32,6	2948s.	14^+	2483,1	17,2%	2524,8	23,3	1950s.
15	2499,5	21,2%	2559,1	30,2	3348s.	15^+	3135,5	15,8%	3168,4	21,3	2293s.
16	2420,4	18,3%	2499,7	35	4715s.	16^+	3100	14,9%	3132,5	17,7	2294s.
17	2709,6	19,8%	2792,8	39,1	4129s.	17^+	3366,9	16,6%	3427	29,9	2572s.
18	2301,7	16,3%	2348,4	27,3	2844s.	18^+	2788,3	13,5%	2837,6	24,8	2110s.
19	3083	23,4%	3190,1	45,9	4193s.	19^+	3696,4	16%	3759,3	32,8	2706s.
20	2322,7	15,7%	2378	30,7	3398s.	20^+	2960,8	15,4%	3000,5	19,2	2484s.
21	3317,8	23,6%	3426,4	54,9	6604s.	21^+	4437,5	17,5%	4517,8	46,8	2382s.
22	4158,8	21%	4261,4	51,9	4148s.	22^+	5249,4	22,2%	5395,1	47,3	2692s.
23	2966,7	19,6%	3043,6	39,9	4217s.	23^+	3578,1	19,4%	3648,8	40,6	2727s.
24	3528,3	23,9%	3631,1	54,3	6365s.	24^+	4435,1	17,2%	4548,4	42,8	3162s.
25	2889,7	22,4%	3040	55,1	3857s.	25^+	3665,9	15,4%	3712,6	23,4	2573s.
26	2712,5	19,2%	2847,3	49,1	4461s.	26^+	3466,7	14,2%	3525,2	32,7	2748s.
27	5016,2	24,2%	5137,7	69,2	10401s.	27^+	6790,5	16,5%	6953,5	51,5	3357s.

5. Conclusions

In this article, we proposed to represent uncertainty on customer demands in the capacitated vehicle routing problem via the theory of evidence. We tackled this problem by generalising the most popular approaches to stochastic programming: chance constrained programming and stochastic programming with recourse. We obtained belief constrained programming and evidential recourse approaches. We studied the optimal solution cost behaviour with respect to the model parameters and customer demand ranking in the case of the belief constrained programming model, and with respect to customer demand specificity in the case of the recourse model. In addition, by considering particular cases of evidential demands, we were able to connect our models not only to stochastic programming but also to robust optimisation. In the last part of this article, we solved both models by a simulated annealing metaheuristic algorithm that uses a combination of operators that aim at minimising the objective function of the problem. We reported the results of our experiments on instances of this difficult optimisation problem. Our experiments showed that our algorithm behave accordingly to the theoretical results studied in this paper.

Future work will include i) comparing our evidential models to the other models, and particularly the stochastic ones, using historical data on customer demands in order to show empirically the advantages of our evidential models; ii) extending to the evidential framework other stochastic variations of the CVRP, such as the CVRP with stochastic customers [28]; iii) extending our evidential models to the case of incomplete knowledge about the dependency between the evidential variables [21]; iv) performing a sensitivity analysis that would allow us to identify certain "key" customers, such that better knowledge about their demands leads to better solutions in each one of these models; and v) considering more general uncertainty frameworks than evidence theory to model uncertainty on customer demands. Lower previsions [53], whose interest to model uncertainty about constraint parameters in optimisation problems has been investigated in [44], may be such a framework. In particular, in the case of independent demands, 2-monotone lower probabilities would offer more generality while being still tractable, thanks to the results of [19].

Acknowledgements

The authors thank the anonymous referees for their valuable comments that helped to improve this paper.

Appendix A. Proof of Proposition 1

Suppose $m_1^X \preceq m_2^X$. For any Q, $1 \leq Q \leq K$, we have then

$$\begin{aligned}
Bel_1^X(x \in A_{1,Q}) &= \sum_{\overline{a} \leq Q} m_1^X(A) \\
&= \sum_{\overline{a} \leq Q} \sum \{R(A,B) m_2^X(B) | A \leq_{lo} B\} \\
&= \sum \{R(A,B) m_2^X(B) | A \leq_{lo} B, \overline{a} \leq Q\} \\
&= \sum \{R(A,B) m_2^X(B) | A \leq_{lo} B, \overline{a} \leq Q, \overline{b} \leq Q\} \\
&\quad + \sum \{R(A,B) m_2^X(B) | A \leq_{lo} B, \overline{a} \leq Q, \overline{b} > Q\} \\
&= \sum \{R(A,B) m_2^X(B) | A \leq_{lo} B, \overline{b} \leq Q\} \\
&\quad + \sum \{R(A,B) m_2^X(B) | A \leq_{lo} B, \overline{a} \leq Q, \overline{b} > Q\} \\
&= \sum_{\overline{b} \leq Q} m_2^X(B) \sum_{A \leq_{lo} B} R(A,B) + \sum_{\overline{b} > Q} m_2^X(B) \sum_{A \leq_{lo} B, \overline{a} \leq Q} R(A,B) \\
&= Bel_2^X(x \in A_{1,Q}) + \sum_{\overline{b} > Q} m_2^X(B) \sum_{A \leq_{lo} B, \overline{a} \leq Q} R(A,B) \\
&\geq Bel_2^X(x \in A_{1,Q}).
\end{aligned} \tag{A.1}$$

The proof is similar to show that $Pl_1^X(x \in A_{1,Q}) \geq Pl_2^X(x \in A_{1,Q})$.

That the implication in Equation (22) is strict is indicated by the following counter-example. Let $X = \{x_1, \ldots, x_8\}$, $Q = 5$ and m_1^X and m_2^X be two mass functions defined as

$$m_1^X(\{x_3, x_4, x_5\}) = 0.8,\ m_1^X(\{x_1, x_3, x_4\}) = 0.2,$$
$$m_2^X(\{x_2, x_3, x_5\}) = 0.7,\ m_2^X(\{x_6, x_8\}) = 0.3.$$

We have

$$Bel_1^X(x \in A_{1,5}) = 1 \geq Bel_2^X(x \in A_{1,5}) = 0.7,$$
$$Pl_1^X(x \in A_{1,5}) = 1 \geq Pl_2^X(x \in A_{1,5}) = 0.7.$$

However, to have $m_1^X \preceq m_2^X$, it must be the case that the mass $m_2^X(\{x_2, x_3, x_5\}) = 0.7$ can be shared among the focal sets of m_1^X that are smaller (according to $\leq_{lo}$) than $\{x_2, x_3, x_5\}$, which is impossible since m_1^X has only one focal set ($\{x_1, x_3, x_4\}$) that is smaller than $\{x_2, x_3, x_5\}$ and this focal set has mass 0.2.

Appendix B. Proof of Proposition 2

Let us consider a set $\mathcal{C} = \{R_1, \ldots, R_m\}$ composed of m routes R_k, $k = 1, ..., m$, such that it is not known whether this set respects the belief-constraints (26) and (27), but it is known that it respects all the other constraints of the CVRPED-BCP.

It is clear that for any $\underline{\beta}$ and $\overline{\beta}$, as Q increases (starting from 1), it reaches necessarily a value at which constraints (26) and (27) are satisfied, and thus at which $\mathcal{C}$ becomes a solution to the CVRPED-BCP. Hence, for $Q' \geq Q$, the set of solutions to the CVRPED-BCP associated with value Q is included in or equal to the set of solutions to the CVRPED-BCP associated with value Q'.

Appendix C. Proof of Proposition 3

Let us consider a set $\mathcal{C} = \{R_1, \ldots, R_m\}$ composed of m routes R_k, $k = 1, ..., m$, such that it is not known whether this set respects the belief-constraints (26), but it is known that it respects all the other constraints of the CVRPED-BCP, in particular constraints (27).

It is clear that for any Q, as $\underline{\beta}$ increases from $\overline{\beta}$ to 1, it reaches necessarily a value at which constraints (26) are satisfied, and thus at which $\mathcal{C}$ becomes a solution to the CVRPED-BCP. Hence, for $\underline{\beta}' \geq \underline{\beta}$, the set of solutions to the CVRPED-BCP associated with value $\underline{\beta}$ is included in or equal to the set of solutions to the CVRPED-BCP associated with value $\underline{\beta}'$.

Appendix D. Proof of Proposition 5

Let R denote a route containing N clients. Without lack of generality, assume that the i-th client on R is the client i. Let $m_{\sum}^{\Theta_R}$ denote the MF defined on $\Theta_R := \{1, 2, \ldots, N \cdot Q\}$ and representing the sum of the customer demands on R when the demand of client i is known in the form of MF m^{Θ_i}, and let $m_{\sum^+}^{\Theta_R}$ denote the MF representing the sum of the customer demands on R when the demand of client i is known in the form of MF $m_+^{\Theta_i}$.

Using a similar proof to that of [25, Proposition 3] (with operation $*$ instantiated to addition $+$, specialisation $\sqsubseteq$ replaced by ranking $\preceq$ and set inclusion $\subseteq$ replaced by lattice ordering $\leq_{lo}$), it is direct to show that $m^{\Theta_i} \preceq m_+^{\Theta_i}, i = 1, \ldots, N \Rightarrow m_{\sum}^{\Theta_R} \preceq m_{\sum^+}^{\Theta_R}$.

Let Bel and Bel_+ (resp. Pl and Pl_+) denote the belief functions (resp. plausibility functions) associated to $m_{\sum}^{\Theta_R}$ and $m_{\sum^+}^{\Theta_R}$, respectively. From Proposition 1, we have then

$$Bel(\sum_{i=1}^{N} d_i \leq Q) \geq Bel_+(\sum_{i=1}^{N} d_i \leq Q), \tag{D.1}$$
$$Pl(\sum_{i=1}^{N} d_i \leq Q) \geq Pl_+(\sum_{i=1}^{N} d_i \leq Q). \tag{D.2}$$

The proposition follows from the fact that Equations (D.1) and (D.2) hold for any route.

Appendix E. Proof of Proposition 6

Let $B_i \subseteq \Omega_i := \{0, 1\}^{i-1}$, $i = 2, \ldots, N$, denote the set of possible failure situations that may occur at the i-th customer on route R, *i.e.*,

$$B_i = f(A_1 \times \cdots \times A_i), \tag{E.1}$$

with $A_\ell := [\![\underline{A_\ell}; \overline{A_\ell}]\!]$, $\ell = 1, \ldots, i$.

Let h_i be the function from Θ^i to $\mathbb{N}^*$ defined by $h_i(\theta_1, \ldots, \theta_i) = q_i$, with q_i defined by (44). In other words, h_i provides the load in the vehicle after serving the i-th customer given that customer demands are $(\theta_1, \ldots, \theta_i)$.

Remark that any $\omega^i \in B_i$ may be obtained by several vectors $(\theta_1, \ldots, \theta_i) \in A_1 \times \cdots \times A_i$. As a consequence, when it is known that the failure situation ω^i has occurred at the i-th customer, then the load in the vehicle after serving the i-th customer is known only in the form of a set L_{ω^i} such that

$$L_{\omega^i} = \left\{ h_i(\theta_1, \ldots, \theta_i) | \quad \forall (\theta_1, \ldots, \theta_i) \in A_1 \times \cdots \times A_i, f(\theta_1, \ldots, \theta_i) = \omega^i \right\}. \tag{E.2}$$

Consider the tree built according to Algorithm 1 and remove all its nodes below level i. Call $Tree_i$ the resulting tree. Then, for a given leaf of $Tree_i$, by concatenating in a vector the Boolean failure variable r_ℓ at level ℓ, $\ell = 2, \ldots, i$, written on the path from the root to the leaf, we obtain the binary failure situation vector $t^i = (r_2, r_3, \ldots, r_i) \in \Omega_i$ and this leaf contains also an interval LT_{t^i} of integers representing imprecise knowledge about the vehicle load after serving the i-th customer when t^i has occurred. Besides, all the leaves of $Tree_i$ yield the subset $BT_i \subseteq \Omega_i$.

We will now show by induction that for $i = 2, \ldots, N$, we have: $B_i = BT_i$ and $\forall \omega^i \in B_i$, $L_{\omega^i} = LT_{t^i}$ for $t^i \in BT_i$ such that $t^i = \omega^i$. Note that from the definition of the addition of two intervals of integers I_1 and I_2, *i.e.*, $I_1 + I_2 = \{x_1 + x_2 | x_1 \in I_1, x_2 \in I_2\}$, we have $\forall x \in I_1 + I_2$, $\exists x_1 \in I_1, x_2 \in I_2$ such that $x_1 + x_2 = x$.

- Consider first the case $i = 2$, hence $\Omega_2 = \{\omega_1^2, \omega_2^2\}$ with $\omega_1^2 = (0)$ and $\omega_2^2 = (1)$. In such case, either $B_2 = \{\omega_1^2\}$ or $B_2 = \{\omega_2^2\}$ or $B_2 = \{\omega_1^2, \omega_2^2\}$.
  - If $B_2 = \{\omega_1^2\}$, then it implies that $\overline{A_1} + \overline{A_2} \leq Q$ and clearly $L_{\omega_1^2} = A_1 + A_2$. Besides, if $\overline{A_1} + \overline{A_2} \leq Q$, then according to Algorithm 1 we have $BT_2 = \{\omega_1^2\}$ and $LT_{\omega_1^2} = A_1 + A_2$.
  - If $B_2 = \{\omega_2^2\}$, then it implies that $\underline{A_1} + \underline{A_2} > Q$ and clearly $L_{\omega_2^2} = A_1 + A_2 - Q$. Besides, if $\underline{A_1} + \underline{A_2} > Q$, then according to Algorithm 1 we have $BT_2 = \{\omega_2^2\}$ and $LT_{\omega_2^2} = A_1 + A_2 - Q$.
  - If $B_2 = \{\omega_1^2, \omega_2^2\}$, then it implies that $\exists (\theta_1, \theta_2) \in A_1 \times A_2$ such that $f(\theta_1, \theta_2) = \omega_1^2$, and thus $\exists (\theta_1, \theta_2) \in A_1 \times A_2$ such that $\theta_1 + \theta_2 \leq Q$, and it also implies $\exists (\theta_1, \theta_2) \in A_1 \times A_2$ such that $f(\theta_1, \theta_2) = \omega_2^2$, and thus $\exists (\theta_1, \theta_2) \in A_1 \times A_2$ such that $\theta_1 + \theta_2 > Q$. In particular, it implies that $\underline{A_1} + \underline{A_2} \leq Q < \overline{A_1} + \overline{A_2}$. Hence, since for $\theta_1 + \theta_2 \leq Q$, we have $q_2 = \theta_1 + \theta_2$, and for $\theta_1 + \theta_2 > Q$, we have $q_2 = \theta_1 + \theta_2 - Q$, we obtain that $L_{\omega_1^2} = [\![\underline{A_1} + \underline{A_2}; Q]\!]$ and $L_{\omega_2^2} = [\![1; \overline{A_1} + \overline{A_2} - Q]\!]$. Besides, if $\underline{A_1} + \underline{A_2} \leq Q < \overline{A_1} + \overline{A_2}$, then according to Algorithm 1 we have $BT_2 = \{\omega_1^2, \omega_2^2\}$ and $LT_{\omega_1^2} = [\![\underline{A_1} + \underline{A_2}; Q]\!]$ and $LT_{\omega_2^2} = [\![1; \overline{A_1} + \overline{A_2} - Q]\!]$.
- Suppose that for $i < N$ we have: $B_i = BT_i$ and $\forall \omega^i \in B_i$, $L_{\omega^i} = LT_{t^i}$ for $t^i \in BT_i$ such that $t^i = \omega^i$. Let us show that it holds for $i + 1$.
  From the preceding assumption, we have $\forall \omega^i = (r_1^i, \ldots, r_i^i) \in B_i$ that L_{ω^i} is the interval LT_{t^i}, *i.e.*, $L_{\omega^i} = [\![\underline{L_{\omega^i}}; \overline{L_{\omega^i}}]\!] = [\![\underline{LT_{t^i}}; \overline{LT_{t^i}}]\!]$ for $t^i \in BT_i$ such that $t^i = \omega^i$. In addition, we have $\forall \omega^i = (r_1^i, \ldots, r_i^i) \in B_i$:
  - Either $\overline{L_{\omega^i}} + \overline{A_{i+1}} \leq Q$, in which case the failure situation ω^i at the i-th customer will induce a failure situation $\omega^{i+1} \in B_{i+1}$ at the $i+1$-th customer such that $\omega^{i+1} = (r_1^i, \ldots, r_i^i, 0)$ and $L_{\omega^{i+1}} = L_{\omega^i} + A_{i+1}$. In addition, $\overline{L_{\omega^i}} + \overline{A_{i+1}} \leq Q$ is equivalent to $\overline{LT_{t^i}} + \overline{A_{i+1}} \leq Q$, in which case the leaf of $Tree_i$ associated to t^i will induce according to Algorithm 1 the leaf of $Tree_{i+1}$ with associated vector $t^{i+1} = \omega^{i+1}$ and interval $LT_{t^{i+1}} = LT_{t^i} + A_{i+1}$.
  - Or $\underline{L_{\omega^i}} + \underline{A_{i+1}} > Q$, in which case the failure situation ω^i at the i-th customer will induce a failure situation $\omega^{i+1} \in B_{i+1}$ at the $i+1$-th customer such that $\omega^{i+1} = (r_1^i, \ldots, r_i^i, 1)$ and $L_{\omega^{i+1}} = L_{\omega^i} + A_{i+1} - Q$. In addition, $\underline{L_{\omega^i}} + \underline{A_{i+1}} > Q$ is equivalent to $\underline{LT_{t^i}} + \underline{A_{i+1}} > Q$, in which case the leaf of $Tree_i$ associated to t^i will induce according to Algorithm 1 the leaf of $Tree_{i+1}$ with associated vector $t^{i+1} = \omega^{i+1}$ and interval $LT_{t^{i+1}} = LT_{t^i} + A_{i+1} - Q$.
  - Or $\underline{L_{\omega^i}} + \underline{A_{i+1}} \leq Q < \overline{L_{\omega^i}} + \overline{A_{i+1}}$, in which case the failure situation ω^i at the i-th customer will induce a failure situation $\omega_L^{i+1} \in B_{i+1}$ at the $i+1$-th customer such that $\omega_L^{i+1} = (r_1^i, \ldots, r_i^i, 0)$ and $L_{\omega_L^{i+1}} = [\![\underline{L_{\omega^i}} + \underline{A_{i+1}}; Q]\!]$ since for $q_i + \theta_{i+1} \leq Q$ we have $q_{i+1} = q_i + \theta_{i+1}$. It will also induce a failure situation $\omega_R^{i+1} \in B_{i+1}$ at the $i+1$-th customer such that $\omega_R^{i+1} = (r_1^i, \ldots, r_i^i, 1)$ and $L_{\omega_R^{i+1}} = [\![1; \overline{L_{\omega^i}} + \overline{A_{i+1}} - Q]\!]$ since for $q_i + \theta_{i+1} > Q$ we have $q_{i+1} = q_i + \theta_{i+1} - Q$. In addition, $\underline{L_{\omega^i}} + \underline{A_{i+1}} \leq Q < \overline{L_{\omega^i}} + \overline{A_{i+1}}$ is equivalent to $\underline{L_{t^i}} + \underline{A_{i+1}} \leq Q < \overline{L_{t^i}} + \overline{A_{i+1}}$, in which case the leaf of $Tree_i$ associated to t^i will induce according to Algorithm 1 the leaf of $Tree_{i+1}$ with associated vector $t_L^{i+1} = \omega_L^{i+1}$ and interval $LT_{t_L^{i+1}} = [\![\underline{LT_{t^i}} + \underline{A_{i+1}}; Q]\!]$. It will also induce the leaf of $Tree_{i+1}$ with associated vector $t_R^{i+1} = \omega_R^{i+1}$ and interval $LT_{t_R^{i+1}} = [\![1; \overline{LT_{t^i}} + \overline{A_{i+1}} - Q]\!]$.

Appendix F. Proof of Proposition 7

The proof of Proposition 7 relies on the following lemma.

Lemma 1. *Let m^Ω and m'^Ω be two MF representing uncertainty about the recourses on a given route R, such that $m^\Omega \sqsubseteq m'^\Omega$. Let $C_E^*(R)$ and $C_E'^*(R)$ denote the upper expected costs of R under m^Ω and m'^Ω, respectively. We have $C_E^*(R) \leq C_E'^*(R)$.*

Proof. Let $C_{\mathrm{P}}^{*}(R)$ and $C_{\mathrm{P}}'^{*}(R)$ denote the upper expected penalty costs of some route R, under m^{Ω} and m'^{Ω} (denoted for simplicity m and m' in this proof), respectively.

We have

$$C_{\mathrm{P}}'^{*}(R) = \sum_{B\subseteq\Omega} m'(B) \max_{\omega\in B} g(\omega), \tag{F.1}$$

and

$$C_{\mathrm{P}}^{*}(R) = \sum_{A\subseteq\Omega} m(A) \max_{\omega\in A} g(\omega). \tag{F.2}$$

Since $m \sqsubseteq m'$,

$$\begin{aligned} C_{\mathrm{P}}^{*}(R) &= \sum_{A\subseteq\Omega} \left(\left(\sum_{B\subseteq\Omega} S(A,B) m'(B) \right) \max_{\omega\in A} g(\omega) \right) \\ &= \sum_{B\subseteq\Omega} m'(B) \left(\sum_{A\subseteq\Omega} S(A,B) \max_{\omega\in A} g(\omega) \right). \end{aligned} \tag{F.3}$$

Since $S(A,B) = 0, \forall A \not\subseteq B$, we can replace the condition of the second sum from $A \subseteq \Omega$ to $A \subseteq B$:

$$C_{\mathrm{P}}^{*}(R) = \sum_{B\subseteq\Omega} m'(B) \left(\sum_{A\subseteq B} S(A,B) \max_{\omega\in A} g(\omega) \right). \tag{F.4}$$

In addition, for any $A, B \subseteq \Omega$ such that $A \subseteq B$, we have

$$\max_{\omega\in A} g(\omega) \leq \max_{\omega\in B} g(\omega),$$

hence

$$\begin{aligned} C_{\mathrm{P}}^{*}(R) &= \sum_{B\subseteq\Omega} m'(B) \left(\sum_{A\subseteq B} S(A,B) \max_{\omega\in A} g(\omega) \right) \\ &\leq \sum_{B\subseteq\Omega} m'(B) \left(\sum_{A\subseteq B} S(A,B) \max_{\omega\in B} g(\omega) \right) \\ &= \sum_{B\subseteq\Omega} m'(B) \max_{\omega\in B} g(\omega) \\ &= C_{\mathrm{P}}'^{*}(R), \end{aligned} \tag{F.5}$$

where we used the fact that S is stochastic so that $\max_{\omega\in B} g(\omega) = \sum_{A\subseteq B} S(A,B) \max_{\omega\in B} g(\omega)$ for any $B \subseteq \Omega$. Using $C_{\mathrm{P}}'^{*}(R) \geq C_{\mathrm{P}}^{*}(R)$ we obtain

$$C_{\mathrm{P}}'^{*}(R) + C(R) \geq C_{\mathrm{P}}^{*}(R) + C(R)\,,$$

which means

$$C_{\mathrm{E}}'^{*}(R) \geq C_{\mathrm{E}}^{*}(R)\,. \qquad \square$$

Proposition 7 may then be proved as follows.

Let R denote a route containing N clients. Without lack of generality, assume that the i-th client on R is the client i. Let m^{Ω} denote the MF representing uncertainty about recourses on R when the demand of client i, $i = 1, \ldots, N$ is known in the form of MF m^{Θ_i}, and let $m_{\star}^{\Omega}$ denote the MF representing uncertainty about recourses on R when the demand of client i is known in the form of MF $m_{\star}^{\Theta_i}$.

Using a similar proof to that of [25, Proposition 3] (with operation $*$ replaced by function f defined in Section 3.2.2), it is direct to show that $m^{\Theta_i} \sqsubseteq m_{\star}^{\Theta_i}, i = 1, \ldots, n \Rightarrow m^{\Omega} \sqsubseteq m_{\star}^{\Omega}$. From Lemma 1, we obtain then $C_{\mathrm{E}}^{*}(R) \leq C_{\mathrm{E}}^{\star *}(R)$.

Considering that the optimal solution with mass functions $m_{\star}^{\Theta_i}$ consists of a set S of routes $\{R_1, \ldots, R_m\}$, we have then $C_{\mathrm{E}}^{\star *}(R_k) \geq C_{\mathrm{E}}^{*}(R_k)$ for each $k \in \{1, \ldots, m\}$, which yields

$$\hat{C}_{Rec}^{\star} = \sum_{k=1}^{m} C_{\mathrm{E}}^{\star *}(R_k) \geq \sum_{k=1}^{m} C_{\mathrm{E}}^{*}(R_k) \geq \hat{C}_{Rec}. \tag{F.6}$$

Appendix G. The simulated annealing algorithm for the CVRPED

The pseudo-code of our simulated annealing algorithm is provided by Algorithm 2. It starts by generating an initial con-

Algorithm 2 Simulated annealing algorithm.

Input: initial temperature T, temperature reduction multiplier κ, freezing temperature $freez$, total number of iterations tot_iter, total number of trials tot_tr, modelling technique MD (BCP or recourse)
Output: Best solution ever visited $Best_C$

1: $Best_{cost} = \infty$
2: **for** $tr = 0$ **to** tot_tr **do**
3: **if** $tr == 0$ **then**
4: $\mathcal{C} =$ initial_config(greedy, MD) ▷ greedy generation
5: **else**
6: $\mathcal{C} =$ initial_config(random, MD) ▷ random generation
7: **end if**
8: $\mathcal{C}_{cost} =$cost($\mathcal{C}$, MD)
9: $TBest_C = \mathcal{C}$
10: $TBest_{cost} = \mathcal{C}_{cost}$
11: **repeat**
12: **for** $iter = 0$ **to** tot_iter **do**
13: $C^* =$ neighbourhood_configuration($\mathcal{C}$, MD)
14: $C^*_{cost} =$ cost(C^*, MD)
15: $\Delta_{cost} = C^*_{cost} - \mathcal{C}_{cost}$
16: **if** ($\Delta_{cost} < 0$) **then**
17: $\mathcal{C} = C^*$
18: $\mathcal{C}_{cost} = C^*_{cost}$
19: **if** $C^*_{cost} < TBest_{cost}$ **then**
20: $TBest_C = C^*$
21: $TBest_{cost} = C^*_{cost}$
22: **end if**
23: **else if** $rnd \leq e^{-\frac{\Delta_{cost}}{T}}$ **then** ▷ rnd is a random number in $[0, 1]$
24: $\mathcal{C} = C^*$
25: $\mathcal{C}_{cost} = C^*_{cost}$
26: **end if**
27: **end for**
28: $T = \kappa \cdot T$
29: **until** ($T == freez$)
30: **if** ($TBest_{cost} < Best_{cost}$) **then**
31: $Best_C = TBest_C$
32: $Best_{cost} = TBest_{cost}$
33: **end if**
34: **end for**

figuration (candidate solution) $\mathcal{C}$ using the `initial_config(...)` routine, when the initial temperature of the system T is at its highest value. Afterwards, T is progressively decreased until reaching the freezing temperature $freez$, while a sequence of iterations tot_iter are performed for each T. Throughout each iteration $iter$, a neighbourhood configuration C^* of the current configuration $\mathcal{C}$ is generated, and the variation in the cost Δ_{cost} is computed. In other words, each configuration represents an intermediate solution that has a different cost which is computed using the cost method, and Δ_{cost} is equal to the difference between the new cost of the neighbourhood configuration C^*_{cost} and the current cost of the current configuration $\mathcal{C}_{cost}$. If the cost decreases then the move to the new cost is accepted (lines 16–18).

However, if Δ_{cost} is positive then the move is accepted or rejected with a probability that equals $e^{-\frac{\Delta_{cost}}{T}}$. Effectively, the probability of accepting inferior solutions is a function of the temperature T and the change in cost Δ_{cost}. We repeat this whole process for a total number of trials tot_tr and the algorithm finally returns the best configuration ever visited.

The set of parameters controlling Algorithm 2 were experimentally determined using Augerat set A instances of the CVRP [46], in which vertices coordinates were randomly constructed [3]. Specifically, the initial temperature T was set to 5000 and was decreased by a temperature reduction multiplier κ that was set to 0.82 until reaching a freezing temperature $freez$ that equals 1. The total number of iterations tot_iter was regulated to 30000, while the total number of trials of the algorithm tot_tr was determined to be 5. The results with our algorithm varied between 1% and 12% from the optimal solutions of the CVRP, with an average running time under 30 minutes, for all instances. This algorithm is an adaptation of the algorithm introduced in [32] for the CVRP.

Let us remark that a configuration in Algorithm 2 is a set of routes that can be generated either by the `initial_config(...)` or the `neighbourhood _configuration(...)` routines. Besides, the `cost(...)` routine evaluates the objective value of the configuration. All these routines depend on a "modelling technique parameter" MD that specifies whether we are solving the CVPRED-BCP or the CVRPED-recourse model. In particular, if $MD =$ "BCP" the `cost(...)` routine evaluates the objective value of the configuration in the BCP model. The objective function is thus the (classical) total travelled distance of the routes (1). If $MD =$ "recourse", the `cost(...)` routine corresponds to using the recourse objective function, that aims at minimising the total upper expected cost (42). The descriptions of the routines `initial_config(...)` and `neighbourhood _configuration(...)` given the value of parameter MD, are provided in Section 4.1.

Appendix H. Examples for the neighbourhood operators `Fix_minimum` and `Replace_highest_average`

Example 7. (`Fix_minimum`) Suppose configuration $\mathcal{C}$ consists of the set of three routes $\mathcal{C} = \{(0, 3, 6, 10, 0), (0, 1, 5, 8, 4, 9, 7, 0), (0, 2, 0)\}$. Assume that the smallest distances between consecutive clients are those between clients $< 3, 6 >$, $< 6, 10 >$, $< 1, 5 >$, $< 4, 9 >$ and $< 7, 0 >$. This means that customers $3, 6, 1, 4$ and 7 cannot be removed from their routes by the `fix_minimum` operator. Consequently, `fix_minimum` selects five random customers excluding $3, 6, 1, 4, 7$ and the depot. For instance customer 5 is selected and removed from the second route in $\mathcal{C}$ and inserted randomly in one of the three available routes in $\mathcal{C}$, while respecting all problem constraints. After selecting the new route for client 5, it is inserted at the position with the resulting smallest distance to client 5. We repeat the same process to move the four other random customers which in this example cannot be other than customers $2, 5, 8, 9$ and 10.

Example 8. (`Replace_highest_average`) Suppose a configuration $\mathcal{C}$ that consists of the set of routes $\mathcal{C} = \{(0, 1, 3, 8, 0), (0, 2, 5, 4, 9, 10, 0), (0, 6, 7, 0)\}$. Suppose 1, 8, 5, 9 and 7 are the clients having the five highest average distances (separating each one of these clients from its neighbours). Then, these clients are removed from their routes, and we will have $\mathcal{C} = \{(0, 3, 0), (0, 2, 4, 10, 0), (0, 6, 0)\}$. Afterwards, each removed client is inserted randomly in one of the available routes. The position where to insert a customer is chosen, such that: i) the problem constraints are respected; and ii) the new position of that customer on the chosen route has the smallest average distance, if compared to all other possible positions of this client on the chosen route. For instance, suppose the first route in $\mathcal{C}$ was chosen randomly for client 1. We know that inserting client 1 on that route does not violate the belief-constraints, and it can be inserted either before client 3 or right after client 3. Suppose also that inserting client 1 right after client 3 has a smallest average distance, than if client 1 was inserted right before client 3. Then, client 1 is inserted right after client 3. This same process is repeated for the remaining clients 8, 5, 9 and 7.

References

[1] N. Ben Abdallah, N. Mouhous-Voyneau, T. Denoeux, Combining statistical and expert evidence using belief functions: application to centennial sea level estimation taking into account climate change, Int. J. Approx. Reason. 55 (1) (2014) 341–354.
[2] H. Agarwal, Reliability Based Design Optimization: Formulations and Methodologies, PhD thesis, University of Notre Dame, Indiana, United States of America, 2004.
[3] P. Augerat, Approche polyédrale du problème de tournées de vehicules, PhD thesis, Institut National Polytechnique de Grenoble, 1995.
[4] C. Baudrit, I. Couso, D. Dubois, Joint propagation of probability and possibility in risk analysis: towards a formal framework, Int. J. Approx. Reason. 45 (1) (2007) 82–105.
[5] C. Baudrit, D. Dubois, D. Guyonnet, Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment, IEEE Trans. Fuzzy Syst. 14 (5) (2006) 593–608.
[6] D. Bertsimas, D.B. Brown, C. Caramanis, Theory and applications of robust optimization, SIAM Rev. 53 (3) (2011) 464–501.
[7] J.R. Birge, F. Louveaux, Introduction to Stochastic Programming, Springer-Verlag, New York, 1997.
[8] L.D. Bodin, B.L. Golden, A.A. Assad, M.O. Ball, Routing and scheduling of vehicles and crews: the state of the art, Comput. Oper. Res. 10 (2) (1983) 63–212.
[9] J. Brito, J.A. Moreno, J.L. Verdegay, Fuzzy optimization in vehicle routing problems, in: Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference, Lisbon, Portugal, 2009.
[10] A. Charnes, W.W. Cooper, G.H. Symonds, Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil, Manag. Sci. 4 (3) (1958) 235–263.
[11] J.Q. Chen, W.L. Li, T. Murata, Particle swarm optimization for vehicle routing problem with uncertain demand, in: 4th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, IEEE, 2013.
[12] C.H. Christiansen, J. Lysgaard, A branch-and-price algorithm for the capacitated vehicle routing problem with stochastic demands, Oper. Res. Lett. 35 (2007) 773–781.
[13] I. Couso, S. Moral, Independence concepts in evidence theory, Int. J. Approx. Reason. 51 (7) (2010) 748–758.
[14] I. Couso, S. Moral, P. Walley, A survey of concepts of independence for imprecise probabilities, Risk Decis. Policy 5 (2) (2000) 165–181.
[15] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (2) (1967) 325–339.
[16] T. Denoeux, Analysis of evidence-theoretic decision rules for pattern classification, Pattern Recognit. 30 (7) (1997) 1095–1107.
[17] T. Denoeux, Extending stochastic ordering to belief functions on the real line, Inf. Sci. 179 (9) (2009) 1362–1376.
[18] T. Denoeux, 40 years of Dempster–Shafer theory, Int. J. Approx. Reason. 79 (C) (2016) 1–6.
[19] S. Destercke, Independence and 2-monotonicity: nice to have, hard to keep, Int. J. Approx. Reason. 54 (2013) 478–490.
[20] S. Destercke, I. Couso, Ranking of fuzzy intervals seen through the imprecise probabilistic lens, Fuzzy Sets Syst. 278 (Supplement C) (2015) 20–39.
[21] S. Destercke, D. Dubois, Idempotent conjunctive combination of belief functions: extending the minimum rule of possibility theory, Inf. Sci. 181 (18) (2011) 3925–3945.

[22] S. Destercke, M-H. Masson, M. Poss, Cautious label ranking with label-wise decomposition, Eur. J. Oper. Res. 246 (3) (2015) 927–935.
[23] M. Dror, G. Laporte, P. Trudeau, Vehicle routing with stochastic demands: properties and solution frameworks, Transp. Sci. 23 (3) (1989) 166–176.
[24] D. Dubois, H. Prade, A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets, Int. J. Gen. Syst. 12 (3) (1986) 193–226.
[25] D. Dubois, H. Prade, Random sets and fuzzy interval analysis, Fuzzy Sets Syst. 42 (1991) 87–101.
[26] D. Dubois, H. Prade, Formal representations of uncertainty, in: Decision-Making Process: Concepts and Methods, ISTE, London, UK, 2009, pp. 85–156.
[27] C. Gauvin, G. Desaulniers, M. Gendreau, A branch-cut-and-price algorithm for the vehicle routing problem with stochastic demands, Comput. Oper. Res. 50 (2014) 141–153.
[28] M. Gendreau, G. Laporte, R. Séguin, Stochastic vehicle routing, Eur. J. Oper. Res. 88 (1996) 3–12.
[29] M. Gendreau, G. Laporte, R. Séguin, A tabu search heuristic for the vehicle routing problem with stochastic demands and customers, Oper. Res. 44 (1996) 469–477.
[30] F. Glover, Tabu search-part 1, ORSA J. Comput. 1 (3) (1989) 190–206.
[31] F. Glover, Tabu search-part 2, ORSA J. Comput. 2 (1) (1990) 4–32.
[32] H. Harmanani, D. Azar, N. Helal, W. Keirouz, A simulated annealing algorithm for the capacitated vehicle routing problem, in: 26th International Conference on Computers and Their Applications, New Orleans, USA, 2011.
[33] N. Helal, F. Pichon, D. Porumbel, D. Mercier, É. Lefèvre, The capacitated vehicle routing problem with evidential demands: a belief-constrained programming approach, in: Jiřina Vejnarová, Václav Kratochvíl (Eds.), Belief Functions: Theory and Applications, Proc. of the Fourth International Conference, BELIEF 2016, in: Lecture Notes in Computer Science, vol. 9861, Springer, Prague, Czech Republic, 2016, pp. 212–221.
[34] N. Helal, F. Pichon, D. Porumbel, D. Mercier, É. Lefèvre, CVRPED benchmarks, https://www.lgi2a.univ-artois.fr/spip/en/benchmarks/the-capacitated-vehicle-routing-problem-with-evidential-demands, 2017. (Accessed 25 August 2017).
[35] N. Helal, F. Pichon, D. Porumbel, D. Mercier, É. Lefèvre, A recourse approach for the capacitated vehicle routing problem with evidential demands, in: Antonucci, et al. (Eds.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, in: Lecture Notes in Artificial Intelligence, vol. 10369, Springer, Lugano, Switzerland, 2017, pp. 190–200.
[36] J.H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, Michigan, USA, 1975, MIT Press, Cambridge, Massachusetts, 1992, 2nd edn.
[37] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science 220 (4598) (1983) 671–680.
[38] G. Laporte, The vehicle routing problem: an overview of exact and approximate algorithms, Eur. J. Oper. Res. 59 (3) (1992) 345–358.
[39] G. Laporte, F. Louveaux, L. van Hamme, An integer l-shaped algorithm for the capacitated vehicle routing problem with stochastic demands, Oper. Res. 50 (2002) 415–423.
[40] H. Masri, F. Ben Abdelaziz, Belief linear programming, Int. J. Approx. Reason. 51 (2010) 973–983.
[41] Z.P. Mourelatos, J. Zhou, A design optimization method using evidence theory, J. Mech. Des. 128 (2006) 901–908.
[42] Y. Peng, J. Chen, Vehicle routing problem with fuzzy demands and the particle swarm optimization solution, in: 2010 International Conference on Management and Service Science (MASS), Wuhan, China, IEEE, 2010.
[43] F. Pichon, D. Dubois, T. Denoeux, Relevance and truthfulness in information correction and fusion, Int. J. Approx. Reason. 53 (2) (2012) 159–175.
[44] E. Quaeghebeur, K. Shariatmadar, G. De Cooman, Constrained optimization problems under uncertainty with coherent lower previsions, Fuzzy Sets Syst. 206 (2012) 74–88.
[45] J.K. Sengupta, A generalization of some distribution aspects of chance-constrained linear programming, Int. Econ. Rev. 11 (2) (1970) 287–304.
[46] Vehicle routing data sets, http://www.coin-or.org/SYMPHONY/branchandcut/VRP/data/index.htm. (Accessed 20 March 2016).
[47] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[48] R.K. Srivastava, K. Deb, R. Tulshyan, An evolutionary algorithm based approach to design optimization using evidence theory, J. Mech. Des. 135 (8) (2013) 081003.
[49] I. Sungur, F. Ordónez, M. Dessouky, A robust optimization approach for the capacitated vehicle routing problem with demand uncertainty, IIE Trans. 40 (2008) 509–523.
[50] D. Teodorovic, G. Pavkovic, The fuzzy set theory approach to the vehicle routing problem when demand at nodes is uncertain, Fuzzy Sets Syst. 82 (3) (1996) 307–317.
[51] P. Toth, D. Vigo, An overview of vehicle routing problems, in: The Vehicle Routing Problem, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2002, pp. 1–26.
[52] M.C.M. Troffaes, Decision making under uncertainty using imprecise probabilities, Int. J. Approx. Reason. 45 (1) (2007) 17–29.
[53] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall/CRC Monographs on Statistics and Applied Probability, Taylor and Francis, 1991.
[54] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets Syst. 100 (Supplement 1) (1999) 9–34.

Bibliographie

[1] N. Ben Abdallah, A.-L. Jousselme et F. Pichon : Une famille de mesures de conflit entre fonctions de croyance. Dans *Rencontres Francophones sur la Logique Floue et ses Applications, LFA 2018, Arras, France, 8-9 novembre 2018*. Cépaduès, à paraître.

[2] N. Ben Abdallah, A.-L. Jousselme et F. Pichon : An ordered family of consistency measures of belief functions. Dans F. Cuzzolin, S. Destercke, T. Denœux et A. Martin, éditeurs : *Belief Functions : Theory and Applications, Proc. of the 5th International Conference, BELIEF 2018, Compiègne, France, 17-21 septembre 2018*, volume 11069 de *Lecture Notes in Computer Science*, pages 199–207. Springer, 2018.

[3] M. Banerjee et D. Dubois : A simple modal logic for reasoning about revealed beliefs. Dans C. Sossai et G. Chemelle, éditeurs : *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Proc. of the 10th European Conférence, ECSQARU 2009, Vérone, Italie, 1-3 juillet 2009*, numéro 5590 de *Lecture Notes in Artificial Intelligence*, pages 805–816. Springer, 2009.

[4] C. Baudrit, D. Dubois et D. Guyonnet : Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. *IEEE Transactions on Fuzzy Systems*, 14(5):593–608, 2006.

[5] D. Bertsimas, D. B. Brown et C. Caramanis : Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[6] J. R. Birge et F. Louveaux : *Introduction to stochastic programming*. Springer-Verlag, New York, 1997.

[7] I. Bloch : Fusion of image information under imprecision and uncertainty : Numerical methods. Dans G. Della Riccia, HJ. Lenz et R. Kruse, éditeurs : *Data Fusion and Perception*, volume 431 de *CISM Courses and Lectures*, page 135–168. Springer, 2001.

[8] L. D. Bodin, B. L. Golden, A. A. Assad et M. O. Ball : Routing and scheduling of vehicles and crews : the state of the art. *Computers and Operations Research*, 10(2):63–212, 1983.

[9] A. Charnes, W. W. Cooper et G. H. Symonds : Cost horizons and certainty equivalents : an approach to stochastic programming of heating oil. *Management Science*, 4(3):235 – 263, 1958.

[10] C. H. Christiansen et J. Lysgaard : A branch-and-price algorithm for the capacitated vehicle routing problem with stochastic demands. *Operations Research Letters*, 35:773–781, 2007.

[11] B. R. Cobb et P. P. Shenoy : On the plausibility transformation method for translating belief function models to probability models. *International Journal of Approximate Reasoning*, 41(3):314–330, 2006.

[12] I. Couso et D. Dubois : Statistical reasoning with set-valued information : Ontic vs. epistemic views. *International Journal of Approximate Reasoning*, 55(7):1502–1518, 2014.

[13] I. Couso et D. Dubois : A general framework for maximizing likelihood under incomplete data. *International Journal of Approximate Reasoning*, 93:238–260, 2018.

[14] A. P. Dempster : Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[15] A.P. Dempster : New methods for reasoning towards posterior distributions based on sample data. *Annals of Mathematical Statistics*, 37:355–374, 1966.

[16] T. Denœux : A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(5): 804–213, 1995.

[17] T. Denœux : Analysis of evidence-theoretic decision rules for pattern classification. *Pattern recognition*, 30(7):1095–1107, 1997.

[18] T. Denœux : Conjunctive and disjunctive combination of belief functions induced by non-distinct bodies of evidence. *Artificial Intelligence*, 172:234–264, 2008.

[19] T. Denœux : Maximum likelihood estimation from uncertain data in the belief function framework. *IEEE Transactions on Knowledge and Data Engineering*, 25(1):119–130, 2013.

[20] T. Denœux : Likelihood-based belief function : Justification and some extensions to low-quality data. *International Journal of Approximate Reasoning*, 55(7):1535–1547, 2014.

[21] T. Denœux : 40 years of Dempster–Shafer theory. *International Journal of Approximate Reasoning*, 79:1 – 6, 2016.

[22] T. Denœux : Logistic regression, neural networks and Dempster-Shafer theory : a new perspective. IEEE (soumis), juillet 2018.

[23] T. Denœux : Quantifying predictive uncertainty using belief functions : different approaches and practical construction. Dans V. Kreinovich, S. Sriboonchitta et N. Chakpitak, éditeurs : *Predictive Econometrics and Big Data*, pages 157–176. Springer International Publishing, 2018.

[24] T. Denœux : Logistic regression revisited : belief function analysis. Dans F. Cuzzolin, S. Destercke, T. Denœux et A. Martin, éditeurs : *Belief Functions : Theory and Applications, Proc. of the 5th International Conference, BELIEF 2018, Compiègne, France, 17-21 septembre 2018*, volume 11069 de *Lecture Notes in Computer Science*, pages 57–64. Springer, 2018.

[25] T. Denœux et P. Smets : Classification using belief functions : relationship between case-based and model-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 36(6):1395–1406, 2006.

[26] T. Denœux, S. Sriboonchitta et O. Kanjanatarakul : Evidential clustering of large dissimilarity data. *Knowledge-Based Systems*, 106:179 – 195, 2016.

[27] T. Denœux, Z. Younes et F. Abdallah : Representing uncertainty on set-valued variables using belief functions. *Artificial Intelligence*, 174(7):479 – 499, 2010.

[28] T. Denœux, N. El Zoghby, V. Cherfaoui et A. Jouglet : Optimal object association in the Dempster-Shafer framework. *IEEE Transactions on Cybernetics*, 44(12):2521 – 2531, 2014.

[29] S. Destercke : Independence and 2-monotonicity : Nice to have, hard to keep. *International Journal of Approximate Reasoning*, 54(4):478 – 490, 2013.

[30] S. Destercke et T. Burger : Toward an axiomatic definition of conflict between belief functions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 43(2):585–596, 2013.

[31] S. Destercke et I. Couso : Ranking of fuzzy intervals seen through the imprecise probabilistic lens. *Fuzzy Sets and Systems*, 278(Supplement C):20 – 39, 2015.

[32] S. Destercke et D. Dubois : Idempotent conjunctive combination of belief functions : extending the minimum rule of possibility theory. *Information Sciences*, 181(18):3925 – 3945, 2011.

[33] S. Destercke et D. Dubois : Other uncertainty theories based on capacities. Dans T. Augustin, F. P. A. Coolen, G. de Cooman et M. C. M. Troffaes, éditeurs : *Introduction to Imprecise Probabilities*, pages 93–113. John Wiley & Sons, 2014.

[34] S. Destercke, F. Pichon et J. Klein : From relations between sets to relations between belief functions. Dans F. Cuzzolin, S. Destercke, T. Denœux et A. Martin, éditeurs : *Belief Functions : Theory and Applications, Proc. of the 5th International Conference, BELIEF 2018, Compiègne, France, 17-21 septembre 2018*, volume 11069 de *Lecture Notes in Computer Science*, pages 65–72. Springer, 2018.

[35] D. Diderot et J. le Rond d'Alembert, éditeurs. *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers.* University of Chicago : ARTFL Encyclopédie Project (Autumn 2017 Edition), R. Morrissey and G. Roe (eds), http://encyclopedie.uchicago.edu/, XVIII^{e} siècle.

[36] M. Dror, G. Laporte et P. Trudeau : Vehicle routing with stochastic demands : properties and solution frameworks. *Transportation science*, 23(3): 166–176, 1989.

[37] D. Dubois et T. Denœux : Pertinence et sincérité en fusion d'informations. Dans *Rencontres Francophones sur la Logique Floue et ses Applications, LFA 2009, Annecy, France, 5-6 septembre 2009*, pages 23–30. Cépaduès, 2009.

[38] D. Dubois, F. Faux et H. Prade : Prejudiced information fusion using belief functions. Dans F. Cuzzolin, S. Destercke, T. Denœux et A. Martin, éditeurs : *Belief Functions : Theory and Applications, Proc. of the 5th International Conference, BELIEF 2018, Compiègne, France, 17-21 septembre 2018*, volume 11069 de *Lecture Notes in Computer Science*, pages 77–85. Springer, 2018.

[39] D. Dubois, W. Liu, J. Ma et H. Prade : The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks. *Information Fusion*, 32:12–39, 2016.

[40] D. Dubois et H. Prade : A set-theoretic view of belief functions : logical operations and approximations by fuzzy sets. *International Journal of General Systems*, 12(3):193–226, 1986.

[41] D. Dubois et H. Prade : *Possibility theory : an approach to computerized processing of uncertainty*. Plenum Press, New-York, 1988.

[42] D. Dubois et H. Prade : Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4(3):244–264, 1988.

[43] D. Dubois et H. Prade : Random sets and fuzzy interval analysis. *Fuzzy Sets and Systems*, 42:87 – 101, 1991.

[44] D. Dubois et H. Prade : Formal representations of uncertainty. Dans D. Bouyssou, D. Dubois, M. Pirlot et H. Prade, éditeurs : *Decision-making Process*, chapitre 3, pages 85–156. Wiley-Blackwell, 2010.

[45] D. Dubois, H. Prade et F. Faux : Raisons de croire, raisons de douter. Une relecture de quelques travaux en théorie des fonctions de croyance. Dans *Rencontres Francophones sur la Logique Floue et ses Applications, LFA 2017, Amiens, France, 19-20 octobre 2017*. Cépaduès, 2017.

[46] D. Dubois, H. Prade et P. Smets : A definition of subjective possibility. *International Journal of Approximate Reasoning*, 48(2):352 – 364, 2008.

[47] R. Duda et P. Hart : *Pattern recognition and scene analysis*. Wiley, New York, 1973.

[48] I. N. Durbach et T. J. Stewart : Modeling uncertainty in multi-criteria decision analysis. *European Journal of Operational Research*, 223(1):1 – 14, 2012.

[49] Z. Elouedi, E. Lefevre et D. Mercier : Discountings of a belief function using a confusion matrix. Dans *Proc. of the 22th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2010, Arras, France, octobre 2010*, pages 287–294, 2010.

[50] Z. Elouedi, K. Mellouli et P. Smets : Assessing sensor reliability for multisensor data fusion within the transferable belief model. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 34(1):782–787, 2004.

[51] R. Fano : *Transmission of information : a statistical theory of communications*. MIT Press, Cambridge, MA, 1961.

[52] S. Ferson, C. A. Joslyn, J. C. Helton, W. L. Oberkampf et K. Sentz : Summary from the epistemic uncertainty workshop : consensus amid diversity. *Reliability Engineering & System Safety*, 85(1):355–369, 2004.

[53] C. Gauvin, G. Desaulniers et M. Gendreau : A branch-cut-and-price algorithm for the vehicle routing problem with stochastic demands. *Computers and Operations Research*, 50:141–153, 2014.

[54] M. Gendreau, G. Laporte et R. Séguin : Stochastic vehicle routing. *European Journal of Operations Research*, 88:3–12, 1996.

[55] M. Gendreau, G. Laporte et R. Séguin : A tabu search heuristic for the vehicle routing problem with stochastic demands and customers. *Operations Research*, 44:469–477, 1996.

[56] M. Grabisch et C. Labreuche : A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *Annals of Operations Research*, 175:247–286, 2010.

[57] J. Grant et A. Hunter : Measuring consistency gain and information loss in stepwise inconsistency resolution. Dans W. Liu, éditeur : *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Proc. of the 11th European Conférence, ECSQARU 2011, Belfast, Royaume-Uni, 29 juin - 1 juillet 2011*, volume 6717 de *Lecture Notes in Computer Science*, pages 362–373. Springer Berlin Heidelberg, 2011.

[58] J. Grant et A. Hunter : Measuring the good and the bad in inconsistent information. Dans T. Walsh, éditeur : *Proc. of the International Joint Conference on Artificial Intelligence, IJCAI 2011, Barcelone, Espagne, 16-22 juillet 2011*, pages 2632–2637. IJCAI/AAAI, 2011.

[59] M. Ha-Duong : Hierarchical fusion of expert opinions in the transferable belief model, application to climate sensitivity. *International Journal of Approximate Reasoning*, 49(3):555–574, 2008.

[60] R. Haenni : Uncover Dempster's rule where it is hidden. Dans *Proc. of the 9th International Conference on Information Fusion, FUSION 2006, Florence, Italy*, 2006.

[61] R. Haenni et S. Hartmann : Modeling partially reliable information sources : a general approach based on Dempster-Shafer theory. *Information Fusion*, 7(4):361–379, 2006.

[62] H. Harmanani, D. Azar, N. Helal et W. Keirouz : A simulated annealing algorithm for the capacitated vehicle routing problem. Dans *Proc of the 26th International Conference on Computers and their Applications, New Orleans, USA*, 2011.

[63] N. Helal, F. Pichon, D. Porumbel, D. Mercier et E. Lefèvre : The capacitated vehicle routing problem with evidential demands. *International Journal of Approximate Reasoning*, 95:124–151, 2018.

[64] R. D. Huddleston et G. K. Pullum : *A student's introduction to English grammar*. Cambridge University Press, 2005.

[65] A. Hunter et S. Konieczny : Measuring inconsistency through minimal inconsistent sets. Dans *Principles of Knowledge Representation and Reasoning, Proc. of the 11th International Conference, KR 2008, Sydney, Australie*, volume 8, pages 358–366. AAAI Press, 2008.

[66] S. Le Hégarat-Mascle et R. Seltz : Automatic change detection by evidential fusion of change indices. *Remote Sensing of Environment*, 91:390–404, 2004.

[67] V. Jain et E. Learned-Miller : FDDB : A benchmark for face detection in unconstrained settings. Rapport technique UM-CS-2010-009, Univ. Massachusetts, 2010.

[68] F. Janez et A. Appriou : Théorie de l'évidence et cadres de discernement non exhaustifs. *Traitement du Signal*, 13(3):237–250, 1996.

[69] F. Janez et A. Appriou : Theory of evidence and non-exhaustive frames of discernment : plausibilities correction methods. *International Journal of Approximate Reasoning*, 18:1–19, 1998.

[70] L. Jaulin : Robust set-membership state estimation : application to underwater robotics. *Automatica*, 45(1):202–206, 2009.

[71] L. Jiao, Q. Pan, T. Denœux, Y. Liang et X. Feng : Belief rule–based classification system : extension of FRBCS in belief functions framework. *Information Sciences*, 309:26 – 49, 2015.

[72] A.-L. Jousselme et P. Maupin : Distances in evidence theory : Comprehensive survey and generalizations. *International Journal of Approximate Reasoning*, 53(2):118–145, 2012.

[73] I. A. Kalinovskii et V. G. Spitsyn : Compact convolutional neural network cascade for face detection. Dans *Proc. of the 10th Annual International Scientific Conference on Parallel Computing Technologies, PCT 2016, Arkhangelsk, Russie, mars 2016*, pages 375–387, 2016.

[74] A. Kallel et S. Le Hégarat-Mascle : Combination of partially non-distinct beliefs : the cautious-adaptive rule. *International Journal of Approximate Reasoning*, 50(7):1000 – 1021, 2009.

[75] O. Kanjanatarakul, T. Denœux et S. Sriboonchitta : Prediction of future observations using belief functions : a likelihood-based approach. *International Journal of Approximate Reasoning*, 72:71–94, 2016.

[76] O. Kanjanatarakul, S. Kuson et T. Denœux : An evidential k-nearest neighbor classifier based on contextual discounting and likelihood maximization. Dans F. Cuzzolin, S. Destercke, T. Denœux et A. Martin, éditeurs : *Belief Functions : Theory and Applications, Proc. of the 5th International Conference, BELIEF 2018, Compiègne, France, 17-21 septembre 2018*, volume 11069 de *Lecture Notes in Computer Science*, pages 155–162. Springer, 2018.

[77] O. Kanjanatarakul, S. Sriboonchitta et T. Denœux : Forecasting using belief functions : an application to marketing econometrics. *International Journal of Approximate Reasoning*, 55(5):1113–1128, 2014.

[78] S. Kirkpatrick, C. D. Gelatt et M. P. Vecchi : Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[79] J. Klein : *Structures algébriques et métriques pour les fonctions de croyance.* Habilitation à Diriger les Recherches, Université de Lille, décembre 2017.

[80] J. Klein et O. Colot : Automatic discounting rate computation using a dissent criterion. Dans *1st Workshop on the Theory of Belief Functions, BELIEF 2010, Brest, France*, 2010.

[81] J. Kohlas et P. P. Shenoy : Computation in valuation algebras. Dans D. M. Gabbay et Ph. Smets, éditeurs : *Handbook of Defeasible Reasoning and Uncertainty Management Systems : Algorithms for Uncertainty and Defeasible Reasoning*, volume 5, pages 5–39. Kluwer, Dordrecht, 2000.

[82] D. Koller et N. Friedman : *Probabilistic graphical models : principles and techniques - adaptive computation and machine learning.* The MIT Press, 2009.

[83] M. Kurdej, J. Moras, V. Cherfaoui et P. Bonnifait : Map-aided evidential grids for driving scene understanding. *IEEE Intelligent Transportation Systems Magazine*, 7(1):30–41, 2015.

[84] M. Lachaize, S. Le Hégarat-Mascle, E. Aldea, A. Maitrot et R. Reynaud : Evidential framework for error correcting output code classification. *Engineering Applications of Artificial Intelligence*, 73:10–21, 2018.

[85] G. Laporte : The vehicle routing problem : An overview of exact and approximate algorithms. *European Journal of Operational Research*, 59(3):345–358, 1992.

[86] G. Laporte, F. Louveaux et L. van Hamme : An integer l-shaped algorithm for the capacitated vehicle routing problem with stochastic demands. *Operations Research*, 50:415–423, 2002.

[87] E. Lefevre, O. Colot et P. Vannoorenberghe : Knowledge modeling methods in the framework of evidence theory : an experimental comparison for melanoma detection. Dans *Proc. of the International Conference on Systems, Man, Cybernetics, SMC 2000, Nashville, États-Unis, 8-11 octobre 2000*, volume 4, pages 2806–2811. IEEE, 2000.

[88] E. Lefèvre, F. Pichon, D. Mercier, Z. Elouedi et B. Quost : Estimation de sincérité et pertinence à partir de matrices de confusion pour la correction de fonctions de croyance. Dans *Rencontres Francophones sur la Logique Floue et ses Applications, LFA 2014, Cargèse, France, 22-24 octobre 2014*, pages 287–294. Cépaduès, 2014.

[89] E. Lefevre, P. Vannoorenberghe et O. Colot : About the use of Dempster-Shafer theory for color image segmentation. Dans *Proc. of the 1st*

International Conference on Color in Graphics and Image Processing, CGIP 2000, Saint-Étienne, France, pages 164–169, 2000.

[90] M.-J. Lesot, F. Pichon et T. Delavallade : Quantitative information evaluation : modeling and experimental evaluation. Dans P. Capet et T. Delavallade, éditeurs : *Information Evaluation*, page 187–228. Wiley-ISTE, 2014.

[91] W. Liu : Analyzing the degree of conflict among belief functions. *Artificial Intelligence*, 170(11):909–924, 2006.

[92] L. Ma, S. Destercke et Y. Wang : Online active learning of decision trees with evidential data. *Pattern Recognition*, 52:33–45, 2016.

[93] H. Masri et F. Ben Abdelaziz : Belief linear programming. *International Journal of Approximate Reasoning*, 51:973–983, 2010.

[94] D. Mercier, T. Denœux et M.-H. Masson : Belief function correction mechanisms. Dans B. Bouchon-Meunier, R. R. Yager, J.-L. Verdegay, M. Ojeda-Aciego et L. Magdalena, éditeurs : *Foundations of Reasoning under Uncertainty*, volume 249 de *Studies in Fuzziness and Soft Computing*, pages 203–222. Springer-Verlag, Berlin, 2010.

[95] D. Mercier, E. Lefèvre et F. Delmotte : Belief functions contextual discounting and canonical decompositions. *International Journal of Approximate Reasoning*, 53(2):146–158, 2012.

[96] D. Mercier, F. Pichon et E. Lefèvre : Corrigendum to "Belief functions contextual discounting and canonical decompositions" [International Journal of Approximate Reasoning 53 (2012) 146-158]. *International Journal of Approximate Reasoning*, 70:137–139, 2016.

[97] D. Mercier, B. Quost et T. Denœux : Refined modeling of sensor reliability in the belief function framework using contextual discounting. *Information Fusion*, 9(2):246 – 258, 2008.

[98] P. Minary, F. Pichon, D. Mercier, E. Lefevre et B. Droit : Face pixel detection using evidential calibration and fusion. *International Journal of Approximate Reasoning*, 91:202–215, 2017.

[99] P. Minary, F. Pichon, D. Mercier, E. Lefevre et B. Droit : Evidential joint calibration of binary SVM classifiers. *Soft Computing*, à paraître.

[100] Z. P. Mourelatos et J. Zhou : A design optimization method using evidence theory. *Journal of Mechanical Design*, 128:901–908, 2006.

[101] H. T. Nguyen : On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65:531–542, 1978.

[102] A. Niculescu-Mizil et R. Caruana : Predicting good probabilities with supervised learning. Dans *Proc. of the 22nd International Conference on Machine Learning, ICML 2005, Bonn, Allemagne, août 2005*, pages 625–632, 2005.

[103] F. Pichon : Equivalence of correction and fusion schemes under meta-independence of sources. Dans *1st Workshop on the Theory of Belief Functions, BELIEF 2010, Brest, France*, 2010.

[104] F. Pichon : On the α–conjunctions for combining belief functions. Dans T. Denœux et M.-H. Masson, éditeurs : *Belief Functions : Theory and Applications, Proc. of the 2nd International Conference, BELIEF 2012, Compiègne, France, 9-11 mai 2012*, volume 164 de *Advances in Intelligent and Soft Computing*, pages 285–292. Springer Berlin / Heidelberg, 2012.

[105] F. Pichon : Canonical decomposition of belief functions based on Teugels' representation of the multivariate bernoulli distribution. *Information Sciences*, 428:76–104, 2018.

[106] F. Pichon et T. Denœux : The unnormalized Dempster's rule of combination : a new justification from the least commitment principle and some extensions. *Journal of Automated Reasoning*, 45(1):61–87, 2010.

[107] F. Pichon, S. Destercke et T. Burger : A consistency-specificity trade-off to select source behavior in information fusion. *IEEE Transactions on Cybernetics*, 45(4):598–609, 2015.

[108] F. Pichon, D. Dubois et T. Denœux : Relevance and truthfulness in information correction and fusion. *International Journal of Approximate Reasoning*, 53(2):159–175, 2012.

[109] F. Pichon, D. Dubois et T. Denœux : Quality of information sources in information fusion. Dans E. Bossé et G. Rogova, éditeurs : *Information Quality in Information Fusion and Decision Making*. Springer, à paraître.

[110] F. Pichon et A.-L. Jousselme : A norm-based view on conflict. Dans *Rencontres Francophones sur la Logique Floue et ses Applications, LFA 2016, La Rochelle, France, 3-4 novembre 2016*, pages 153–159. Cépaduès, 2016.

[111] F. Pichon, A.-L. Jousselme et N. Ben Abdallah : Several shades of conflict. *Fuzzy Sets and Systems (soumis)*, janvier 2018.

[112] F. Pichon, C. Labreuche, B. Duqueroie et T. Delavallade : Multidimensional approach to reliability evaluation of information sources. Dans P. Capet et T. Delavallade, éditeurs : *Information Evaluation*, page 129–156. Wiley-ISTE, 2014.

[113] F. Pichon, D. Mercier, E. Lefèvre et F. Delmotte : Proposition and learning of some belief function contextual correction mechanisms. *International Journal of Approximate Reasoning*, 72:4–42, 2016.

[114] B. Quost, T. Denœux et S. Li : Parametric classification with soft labels using the evidential EM algorithm : linear discriminant analysis versus logistic regression. *Advances in Data Analysis and Classification*, 11(4):659–690, 2017.

[115] S. Ramel, F. Pichon et F. Delmotte : Active evidential calibration of binary SVM classifiers. Dans F. Cuzzolin, S. Destercke, T. Denœux et A. Martin, éditeurs : *Belief Functions : Theory and Applications, Proc. of*

the 5th International Conference, BELIEF 2018, Compiègne, France, 17-21 septembre 2018, volume 11069 de *Lecture Notes in Computer Science*, pages 208–216. Springer, 2018.

[116] T. Reineking : Active classification using belief functions and information gain maximization. *International Journal of Approximate Reasoning*, 72:43–54, 2016.

[117] C. Rose et M. D. Smith : *Mathematical statistics with Mathematica*. Springer-Verlag, New York, 2002.

[118] J. Schubert : Clustering decomposed belief functions using generalized weights of conflict. *International Journal of Approximate Reasoning*, 48(2): 466–480, 2008.

[119] J. Schubert : Conflict management in Dempster-Shafer theory using the degree of falsity. *International Journal of Approximate Reasoning*, 52(3):449–460, 2011.

[120] R. Senge, S. Bösner, K. Dembczynski, J. Haasenritter, O. Hirsch, N. Donner-Banzhoff et E. Hüllermeier : Reliable classification : Learning classifiers that distinguish aleatoric and epistemic uncertainty. *Information Sciences*, 255:16–29, 2014.

[121] Vehicle Routing Data Sets : http://www.coin-or.org/SYMPHONY/branchandcut/VRP/data/index.htm. Date de téléchargement : 20 mars 2016.

[122] Burr Settles : *Active learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2012.

[123] G. Shafer : *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.

[124] P. P. Shenoy : A valuation-based language for expert systems. *International Journal of Approximate Reasoning*, 3(5):383 – 411, 1989.

[125] P. P. Shenoy et G. Shafer : Axioms for probability and belief-function propagation. Dans R. D. Shachter, T. Levitt, J. F. Lemmer et L. N. Kanal, éditeurs : *Uncertainty in Artificial Intelligence 4*, pages 169–198. North-Holland, 1990.

[126] P. Smets : Belief functions : the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9(1):1–35, 1993.

[127] P. Smets : The canonical decomposition of a weighted belief. Dans *Proc. of the 14th International Joint Conference on Artificial Intelligence, IJCAI 1995, San Mateo, Californie, États-Unis, 1995*, pages 1896–1901. Morgan Kaufmann, 1995.

[128] P. Smets : The α-junctions : combination operators applicable to belief functions. Dans D. M. Gabbay, R. Kruse, A. Nonnengart et H. J. Ohlbach, éditeurs : *First International Joint Conference on Qualitative and Quantitative Practical Reasoning, ECSQARU-FAPR 1997, Bad Honnef, Allemagne, 1997*,

volume 1244 de *Lecture Notes in Computer Science*, pages 131–153. Springer, 1997.

[129] P. Smets : The application of the matrix calculus to belief functions. *International Journal of Approximate Reasoning*, 31(1–2):1–30, 2002.

[130] P. Smets : Analyzing the combination of conflicting belief functions. *Information Fusion*, 8(4):387–412, 2007.

[131] P. Smets et R. Kennes : The transferable belief model. *Artificial Intelligence*, 66:191–243, 1994.

[132] Ph. Smets : Managing deceitful reports with the transferable belief model. Dans *Proc. of the 8th International Conference on Information Fusion, FUSION 2005, Philadelphie, États-Unis*, juillet 2005.

[133] I. Song et S. Lee : Explicit formulae for product moments of multivariate gaussian random variables. *Statistics and Probability Letters*, 100:27–34, 2015.

[134] R. K. Srivastava, K. Deb et R. Tulshyan : An evolutionary algorithm based approach to design optimization using evidence theory. *Journal of Mechanical Design*, 135(8):081003 – 081003–12, 2013.

[135] I. Sungur, F. Ordónez et M. Dessouky : A robust optimization approach for the capacitated vehicle routing problem with demand uncertainty. *IIE Transactions*, 40:509–523, 2008.

[136] D. Teodorovic et G. Pavkovic : The fuzzy set theory approach to the vehicle routing problem when demand at nodes is uncertain. *Fuzzy Sets and Systems*, 82(3):307 – 317, 1996.

[137] J. L. Teugels : Some representations of the multivariate Bernoulli and binomial distributions. *Journal of Multivariate Analysis*, 32(2):256 – 268, 1990.

[138] J. L. Teugels et J. Van Horebeek : Algebraic descriptions of nominal multivariate discrete data. *Journal of Multivariate Analysis*, 67(2):203 – 226, 1998.

[139] P. Viola et M. J. Jones : Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[140] F. Voorbraak : A computationally efficient approximation of Dempster-Shafer theory. *International Journal of Man-Machine Studies*, 30:525–536, 1989.

[141] P. Walley : *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London, 1991.

[142] P. Xu, F. Davoine et T. Denœux : Evidential combination of pedestrian detectors. Dans *Proc. of the 25th British Machine Vision Conference, BMVC 2014, Nottingham, Angleterre*, pages 1–14, septembre 2014.

[143] P. Xu, F. Davoine, H. Zha et T. Denœux : Evidential calibration of binary SVM classifiers. *International Journal of Approximate Reasoning*, 72:55–70, 2016.

[144] Y. Yang, D. Han et C. Han : Discounted combination of unreliable evidence using degree of disagreement. *International Journal of Approximate Reasoning*, 54(8):1197–1216, 2013.

[145] S. Zair et S. Le Hégarat-Mascle : Evidential framework for robust localization using raw gnss data. *Engineering Applications of Artificial Intelligence*, 61:126–135, 2017.

[146] W. Zhong et J. T. Kwok : Accurate probability calibration for multiple classifiers. Dans *Proc. of the 23rd International Joint Conference on Artificial Intelligence, IJCAI 2013, Pékin, Chine*, pages 1939–1945, août 2013.

Qualité des sources en fusion d'informations dans le cadre de la théorie des fonctions de croyance.

Résumé : La fusion d'informations est un processus dont l'objectif est d'extraire une connaissance véridique et aussi fine que possible à propos d'une entité d'intérêt, étant donné des témoignages incertains provenant de sources de qualité variable. Un ensemble de résultats synthétisés dans ce document indique que la théorie des fonctions de croyance est particulièrement adaptée à ce problème. Il est montré que tout ensemble de témoignages élémentaires issus de sources partiellement fiables induit une fonction de croyance et, surtout, que toute fonction de croyance résulte d'un tel ensemble. Une approche générale est proposée pour la fusion de fonctions de croyance émises par des sources, où les connaissances utilisées quant à leur qualité sont explicites et où cette qualité peut concerner d'autres aspects que la fiabilité tels que la sincérité. Des moyens sont fournis pour établir ces connaissances à partir d'une expérience préalable des sources ou, à défaut, pour les dériver de principes généraux. L'intérêt pratique de ces propositions est montré à travers quelques problèmes de classification supervisée. L'intérêt plus général de la théorie des fonctions de croyance en tant que théorie de l'incertain, est exhibé dans le cas du problème de tournées de véhicules avec demandes incertaines. Un projet de recherches est finalement présenté ; bien que la vitalité et la maturité de la théorie des fonctions de croyance et de sa communauté soient déjà reconnues, ce projet illustre incidemment une partie de ce que ce formalisme peut encore apporter.

Mots clés : Théorie des fonctions de croyance, Théorie de Dempster-Shafer, Fusion d'informations, Correction contextuelle, Décomposition canonique, Prédiction, Classification supervisée, Étalonnage, Conflit, Problème de tournées de véhicules.