Articles written on the occasion of the 50th anniversary of fuzzy set theory

Didier Dubois Henri Prade

(with Davide Ciucci & Jim Bezdek)

Rapport Interne IRIT
RR--2015--11--FR

Décembre 2015

Articles written on the occasion of the 50^{th} anniversary of fuzzy set theory

Didier Dubois Henri Prade
(with Davide Ciucci & Jim Bezdek)

Summary:
The first paper by Lotfi A. Zadeh on fuzzy sets appeared 50 years ago. On the occasion of this anniversary, the authors of this report have been led to contribute a series of papers in relation with this event. The report gathers 8 papers, and thus covers many issues in relation with fuzzy sets. The first two provide overviews about the historical emergence of fuzzy sets, and the first steps of the fuzzy set research in France. The third one discusses the scientific legacy of fuzzy sets after 50 years. The next two survey some developments of possibility theory, focusing on two specific issues: the elicitation of qualitative or quantitative possibility distributions, and the forms of inconsistency representable in a possibilistic logic setting. The 6th paper (with Davide Ciucci) surveys the different forms of hybridation between fuzzy set and rough set theories using squares and cubes of opposition. The 7th paper discusses granular computing from different theoretical viewpoints including extensional fuzzy sets, formal concept analysis and rough sets. The last paper (with James C. Bezdek) presents a selected, annotated bibliography of fuzzy set contributions based on representative papers chosen by IEEE CIS Fuzzy Systems pioneers.

Contents:

The emergence of fuzzy sets:
a historical perspective *

Didier Dubois and Henri Prade
IRIT-CNRS, Université Paul Sabatier, 31062 Toulouse Cedex 09, France

December 15, 2015

Abstract: This paper tries to suggest some reasons why fuzzy set theory came to life 50 years ago by pointing out the existence of streams of thought in the first half of the XXth century in logic, linguistics and philosophy, that paved the way to the idea of moving away from the Boolean framework, through the proposal of many-valued logics and the study of the vagueness phenomenon in natural languages. The founding paper in fuzzy set theory can be viewed as the crystallization of such ideas inside the engineering arena. Then we stress the point that this publication in 1965 was followed by several other seminal papers in the subsequent 15 years, regarding classification, ordering and similarity, systems science, decision-making, uncertainty management and approximate reasoning. The continued effort by Zadeh to apply fuzzy sets to the basic notions of a number of disciplines in computer and information sciences proved crucial in the diffusion of this concept from mathematical sciences to industrial applications.

Key-words Fuzzy sets, many-valued logics, vagueness, possibility theory, approximate reasoning

1 Introduction

The notion of a fuzzy set stems from the observation made by Zadeh [60] fifty years ago in his seminal paper that

*To appear in FUZZY LOGIC IN ITS 50TH YEAR A perspective of New Developments, Directions and Challenges, C. KAHRAMAN, U. KAYMAK & A. YAZICI, Eds., Springer, 2016.

> *"more often than not, the classes of objects encountered in the real physical world do not have precisely defined criteria of membership."*

By "precisely defined", Zadeh means all-or-nothing, thus emphasizing the continuous nature of many categories used in natural language. This observation emphasizes the gap existing between mental representations of reality and usual mathematical representations thereof, which are traditionally based on binary logic, precise numbers, differential equations and the like. Classes of objects referred to in Zadeh's quotation exist only through such mental representations, e.g., through natural language terms such as *high* temperature, *young* man, *big* size, etc., and also with nouns such as *bird, chair*, etc. Classical logic is too rigid to account for such categories where it appears that membership is a gradual notion rather than an all-or-nothing matter.

The ambition of representing human knowledge in a human-friendly, yet rigorous way might have appeared like a futile exercice not worth spending time on, and even ridiculous from a scientific standpoint, only one hundred years ago. However in the meantime the emergence of computers has significantly affected the landscape of science, and we have now entered the era of information management. The development of sound theories and efficient technology for knowledge representation and automated reasoning has become a major challenge, now that many people possess computers and communicate with them in order to find information that helps them when making decisions. An important issue is to store and exploit human knowledge in various domains where objective and precise data are seldom available. Fuzzy set theory participates to this trend, and, as such, has close connection with Artificial Intelligence. This chapter is meant to account for the history of how the notion of fuzzy set could come to light, and what are the main landmark papers by its founder that stand as noticeable steps towards the construction of the fuzzy set approach to classification, decision, human knowledge representation and uncertainty. Besides, the reader is invited to consult a recent personal account, written by Zadeh [85], of the circumstances in which the founding paper on fuzzy sets was written.

2 A prehistory of fuzzy sets

This section gives some hints to works what can be considered as forerunners of fuzzy sets. Some aspects of the early developments are described in more details by Gottwald [28] and Ostasiewicz [45, 46]. This section freely borrows from [17], previously written with the later author.

2.1 Graded membership to sets before Zadeh

In spite of the considerable interest for multiple-valued logics raised in the early 1900s by Jan Łukasiewicz and his school who developed logics with intermediary truth value(s), it was the American philosopher Max Black [7] who first proposed so-called "consistency profiles" (the ancestors of fuzzy membership functions) in order to "characterize vague symbols."

As early as in 1946, the philosopher Abraham Kaplan argued in favor of the usefulness of the classical calculus of sets for practical applications. The essential novelty he introduces with respect to the Boolean calculus consists in entities which have a degree of vagueness characteristic of actual (empirical) classes (see Kaplan [33]). The generalization of the traditional characteristic function has been first considered by H. Weyl [55] in the same year; he explicitly replaces it by a continuous characteristic function. They both suggested calculi for generalized characteristic functions of vague predicates, and the basic fuzzy set connectives already appeared in these works.

Such calculus has been presented by A. Kaplan and H. Schott [34] in more detail, and has been called the calculus of empirical classes (CEC). Instead of notion of "property", Kaplan and Schott prefer to use the term "profile" defined as a type of quality. This means that a profile could refer to a simple property like *red, green*, etc. or to a complex property like *red and 20 cm long, green and 2 years old*, etc. They have replaced the classical characteristic function by an indicator which takes on values in the unit interval. These values are called the weight from a given profile to a specified class. In the work of Kaplan and Schott, the notion of "empirical class" corresponds to the actual notion of "fuzzy set", and a value in the range of the generalized characteristic function (indicator, in their terminology) is already called by Kaplan and Schott a "degree of membership" (Zadehian grade of membership). Indicators of profiles are now called membership functions of fuzzy sets. Strangely enough it is the mathematician of probabilistic metric spaces, Karl Menger, who, in 1951, was the first to use the term "ensemble flou" (the French counterpart of "fuzzy set") in the title of a paper [40] of his.

2.2 Many-valued logics

The Polish logician Jan Łukasiewicz (1878-1956) is considered as the main founder of multi-valued logic. This is an important point as multi-valued logic is to fuzzy set theory what classical logic is to set theory. The new system he proposed has been published for the first time in Polish in 1920. However, the meaning of truth-values other than "true" and "false" remained rather unclear until Zadeh introduced fuzzy sets. For instance, Łukasiewicz [38] interpreted the third truth-value of his 3-valued logic as "possible", which refers to a modality rather than a truth-value. Kleene [35] suggests that the third truth-value means "unknown" or "undefined". See Ciucci and Dubois [9] for a overview of such epistemic interpretations of three-valued logics. On the contrary, Zadeh [60] considered intermediate truth-degrees of fuzzy propositions as ontic, that is, being part of the definition of a gradual predicate. Zadeh observes that the case where the unit interval is used as a membership scale "corresponds to a multivalued logic with a continuum of truth values in the interval $[0, 1]$", acknowledging the link between fuzzy sets and many-valued logics. Clearly, for Zadeh, such degrees of truth do not refer to any kind of uncertainty, contrary to what is often found in more recent texts about fuzzy sets by various authors. Later on, Zadeh[70] would not consider fuzzy logic to be another name for many-valued logic. He soon considered that fuzzy truth-values should be considered as fuzzy sets of the unit interval, and that fuzzy logic should be viewed as a theory of approximate reasoning whereby fuzzy truth-values act as modifiers of the fuzzy statement they apply to.

2.3 The issue of vagueness

More than one hundred years ago, the American philosopher Charles Peirce [47] was one of the first scholars in the modern age to point out, and to regret, that

> *"Logicians have too much neglected the study of vagueness, not suspecting the important part it plays in mathematical thought."*

This point of view was also expressed some time later by Bertrand Russell [49]. Even Wittgenstein [57] pointed out that concepts in natural language do not possess a clear collection of properties defining them, but have extendable boundaries, and that there are central and less central members in a category.

The claim that fuzzy sets are a basic tool for addressing vagueness of linguistic terms has been around for a long time. For instance, Novák [44] insists that fuzzy logic is tailored for vagueness and he opposes vagueness to uncertainty.

Nevertheless, in the last thirty years, the literature dealing with vagueness has grown significantly, and much of it is far from agreeing on the central role played by fuzzy sets in this phenomenon. Following Keefe and Smith [53], vague concepts in natural language display at least one among three features:

- **The existence of borderline cases**: That is, there are some objects such that neither a concept nor its negation can be applied to them. For a borderline object, it is difficult to make a firm decision as to the truth or the falsity of a proposition containing a vague predicate applied to this object, even if a precise description of the latter is available. The existence of borderline cases is sometimes seen as a violation of the law of excluded middle.

- **Unsharp boundaries**: The extent to which a vague concept applies to an object is supposed to be a matter of degree, not an all-or-nothing decision. It is relevant for predicates referring to continuous scales, like *tall, old*, etc. This idea can be viewed as a specialization of the former, if we regard as borderline cases objects for which a proposition is neither totally true nor totally false. In the following we shall speak of "gradualness" to describe such a feature. Using degrees of appropriateness of concepts to objects as truth degrees of statements involving these concepts goes against the Boolean tradition of classical logic.

- **Susceptibility to Sorites paradoxes.** This is the idea that the presence of vague propositions make long inference chains inappropriate, yielding debatable results. The well-known examples deal with heaps of sand (whereby, since adding a grain of sand to a small heap keeps its small, all heaps of sand should be considered small), young persons getting older by one day, bald persons that are added one hair, etc.

Since their inception, fuzzy sets have been controversial for philosophers, many of whom are reluctant to consider the possibility of non-Boolean predicates, as it questions the usual view of truth as an absolute entity. A disagreement opposes those who, like Williamson, claim a vague predicate has a standard, though ill-known, extension [56], to those who, like Kit Fine, deny the existence of a decision threshold and just speak of a truth value gap [24]. However, the two latter views reject the concept of gradual truth, and concur on the point that fuzzy sets do not propose a good model for vague predicates. One of the reasons for the misunderstanding between fuzzy sets and the philosophy of vagueness may lie in the fact that Zadeh was trained in engineering mathematics, not in the area of philosophy. In particular, vagueness is often understood as a defect of natural language (since it is not appropriate for devising formal proofs, it questions usual rational forms of reasoning). Actually, vagueness of linguistic terms was considered as a logical nightmare for early 20th century philosophers. In contrast, for Zadeh, going from Boolean logic to fuzzy logic is viewed as a positive move: it captures tolerance to errors (softening blunt threshold effects in algorithms) and may account for the flexible use of words by people [73]. It also allows for information summarization: detailed descriptions are sometimes hard to make sense of, while summaries, even if imprecise, are easier to grasp [69].

However, the epistemological situation of fuzzy set theory itself may appear kind of unclear. Fuzzy sets and their extensions have been understood in various ways in the literature: there are several notions that are appealed to in connection with fuzzy sets, like similarity, uncertainty and preference [19]. The concept of similarity to prototypes has been central in the development of fuzzy sets as testified by numerous works on fuzzy clustering. It is also natural to represent incomplete knowledge by fuzzy sets (of possible models of a fuzzy knowledge base, or fuzzy error intervals, for instance), in connection to possibility theory [74, 18]. Utility functions in decision theory also appear as describing fuzzy sets of good options. These topics are not really related to the issue of vagueness.

Indeed, in his works, Zadeh insists that, even when applied to natural language, fuzziness is not vagueness. The term fuzzy is restricted to sets where the transition between membership and non-membership is gradual rather than abrupt, not when it is crisp but unknown. Zadeh [73] argues as follows:

> *"Although the terms fuzzy and vague are frequently used interchangeably in the literature, there is, in fact, a significant difference between them. Specifically, a proposition, p, is fuzzy if it contains words which are labels of fuzzy sets; and p is vague if it is both fuzzy and insufficiently specific for a particular purpose. For example, "Bob will be back in a few minutes" is fuzzy, while "Bob will be back sometime" is vague if it is insufficiently informative as a basis for a decision. Thus, the vagueness of a proposition is a decision-dependent characteristic whereas its fuzziness is not. "*

Of course, the distinction made by Zadeh may not be so strict as he claims. While "in a few minutes" is more specific than "sometime" and sounds less vague, one may argue that there is some residual vagueness in the former, and that the latter does not sound very crisp after all. Actually, one may argue that the notion of non-Boolean linguistic categories proposed by Zadeh from 1965 on is capturing the idea of gradualness, not vagueness in its philosophical understanding. Zadeh repetitively claims that gradualness is pervasive in the representation of information, especially human-originated.

The connection from gradualness to vagueness does exist in the sense that, insofar as vagueness refers to uncertainty about meaning of natural language categories, gradual predicates tend to be more often vague than Boolean ones: indeed, it is more difficult to precisely measure the membership function of a fuzzy set representing a gradual category than to define the characteristic function of a set representing the extension of a Boolean predicate [13]. In fact, the power of expressiveness of real numbers is far beyond the limited level of precision perceived by the human mind. Humans basically handle meaningful summaries. Analytical representations of physical phenomena can be faithful as models of reality, but remain esoteric to lay people; the same may hold real-valued membership grades. Indeed, mental representations are tainted with vagueness, which encompasses at the same time the lack of specificity of linguistic terms, and the lack of well-defined boundaries of the class of objects they refer to, as much as the lack of precision of membership grades. So moving from binary membership to continuous is a bold step, and real-valued membership grades often used in fuzzy sets are just another kind of idealization of human perception, that leaves vagueness aside.

3 The development of fuzzy sets and systems

Having discussed the various streams of ideas that led to the invention of fuzzy sets, we now outline the basic building blocks of fuzzy set theory, as they emerged from 1965 all the way to the early 1980's, under the impulse of the founding father, via several landmark papers, with no pretense to exhaustiveness. Before discussing the landmark papers that founded the field, it is of interest to briefly summary how L. A. Zadeh apparently came to the idea of developing fuzzy sets and more generally fuzzy logic. See also [50] for historical details. First, it is worth mentioning that, already in 1950, after commenting the first steps towards building thinking machines (a recently hot topic at the time), he indicated in his conclusion [58]:

> *"Through their association with mathematicians, the electronic engineers working on thinking machines have become familiar with such hitherto remote subjects as Boolean algebra, multivalued logic, and so forth."*,

which shows an early concern for logic and many-valued calculi. Twelve years later, when providing "a brief survey of the evolution of system theory' [59] he wrote (p. 857)

> "*There are some who feel that this gap reflects the fundamental inadequacy of the conventional mathematics - the mathematics of precisely-defined points, functions, sets, probability measures, etc. - for coping with the analysis of biological systems, and that to deal effectively with such systems, which are generally orders of magnitude more complex than man-made systems, we need a radically different kind of mathematics, the mathematics of fuzzy or cloudy quantities which are not describable in terms of probability distributions."*

This quotation shows that Zadeh was first motivated by an attempt at dealing with complex systems rather than with man-made systems, in relation with the current trends of interest in neuro-cybernetics in that time (in that respect, he pursued the idea of applying fuzzy sets to biological systems at least until 1969 [64]).

3.1 Fuzzy sets: the founding paper and its motivations

The introduction of the notion of a fuzzy set by L. A. Zadeh was motivated by the fact that, quoting the founding paper [60]:

> *"imprecisely defined "classes" play an important role in human thinking, particularly in the domains of pattern recognition, communication of information, and abstraction".*

This seems to have been a recurring concern in all of Zadeh's fuzzy set papers since the beginning, as well as the need to develop a sound mathematical framework for handling this kind of "classes". This purpose required an effort to go beyond classical binary-valued logic, the usual setting for classes. Although many-valued logics had been around for a while, what is really remarkable is that due to this concern, Zadeh started to think in terms of *sets* rather than only in terms of degrees of truth, in accordance with intuitions formalized by Kaplan but not pursued further. Since a set is a very basic notion, it was opening the road to the introduction of the fuzzification of any set-based notions such as relations, events, or intervals, while sticking with the many-valued logic point of view only does not lead you to consider such generalized notions. In other words, while Boolean algebras are underlying both propositional logic and naive set theory, the set point of view may be found richer in terms of mathematical modeling, and the same thing takes place when moving from many-valued logics to fuzzy sets.

A fuzzy set can be understood as a class equipped with an ordering of elements expressing that some objects are more inside the class than others. However, in order to extend the Boolean connectives, we need more than a mere relation in order to extend intersection, union and complement of sets, let alone implication. The set of possible membership grades has to be a complete lattice [27] so as to capture union and intersection, and either the concept of residuation or an order-reversing function in order to express some kind of negation and implication. The study of set operations on fuzzy sets has in return strongly contributed to a renewal of many-valued logics under the impulse of Petr Hájek [29] (see [14] for an introductory overview).

Besides, from the beginning, it was made clear that fuzzy sets were not meant as probabilities in disguise, since one can read [60] that

> *"the notion of a fuzzy set is completely non-statistical in nature"*

and that it provides

> *"a natural way of dealing with problems where the source of imprecision is the absence of sharply defined criteria of membership rather than the presence of random variables."*

Presented as such, fuzzy sets are prima facie not related to the notion of uncertainty. The point that typicality notions underlie the use of gradual membership functions of linguistic terms is more connected to similarity than to uncertainty. As a consequence,

- originally, fuzzy sets were designed to formalize the idea of soft classification, which is more in agreement with the way people use categories in natural language.
- fuzziness is just implementing the concept of gradation in all forms of reasoning and problem-solving, as for Zadeh, everything is a matter of degree.
- a degree of membership is an abstract notion to be interpreted in practice.

Important definitions appear in the funding papers such as cuts (a fuzzy set can be viewed as a family of nested crisp sets called its cuts), the basic fuzzy set-theoretic connectives (e.g. minimum and product as candidates for intersection, inclusion via inequality of membership functions) and the extension principle whereby the domain of a function is extended to fuzzy set-valued arguments. As pointed out earlier, according to the area of application, several interpretations can be found such as degree of similarity (to a prototype in a class), degree of plausibility, or degree of preference [19]. However, in the founding paper, membership functions are considered in connection with the representation of human categories only. The three kinds of interpretation would become patent in subsequent papers.

3.2 Fuzzy sets and classification

A popular part of the fuzzy set literature deals with fuzzy clustering where gradual transitions between classes and their use in interpolation are the basic contribution of fuzzy sets. The idea that fuzzy sets would be instrumental to avoid too rough classifications was provided very early by Zadeh, along with Bellman and Kalaba [3]. They outline how to construct membership functions of classes from examples thereof. Intuitively speaking, a cluster gathers elements that are rather close to each other (or close to some core element(s)), while they are well-separated from the elements in the other cluster(s). Thus, the notions of graded proximity, similarity (dissimilarity) are at work in fuzzy clustering. With gradual clusters, the key issue is to define fuzzy partitions. The most widely used definition of a fuzzy partition, originally due to Ruspini [48], where the sum of membership grades of one element to the various classes is 1. This was enough to trigger the fuzzy clustering literature, that culminated with the numerous works by Bezdek and colleagues, with applications to image processing for instance [5].

3.3 Fuzzy events

The idea of replacing sets by fuzzy sets was quickly applied by Zadeh to the notion of event in probability theory [63]. The probability of a fuzzy event is just the expectation of its membership function. Beyond the mathematical exercise, this definition has the merit of showing the complementarity between fuzzy set theory and probability theory: while the latter models uncertainty of events, the former modifies the notion of an event admitting it can occur to some degree when observing a precise outcome. This membership degree is ontic as it is part of the definition of the event, as opposed to probability that has an epistemic flavor, a point made very early by De Finetti [12], commenting Łukasiewicz logic. However, it took 35 years before a generalization of De Finetti's theory of subjective probability to fuzzy events was devised by Mundici[42]. Since then, there is an active mathematical area studying probability theory on algebras of fuzzy events.

3.4 Decision-making with fuzzy sets

Fuzzy sets can be useful in decision sciences. This is not surprising since decision analysis is a field where human-originated information is pervasive. While the suggestion of modelling fuzzy optimization as the (product-based) aggregation of an objective function with a fuzzy constraint first appeared in the last section of [61], the full-fledged seminal paper in this area was written by Bellman and Zadeh [4] in 1970, highlighting the role of fuzzy set connectives in criteria aggregation. That pioneering paper makes three main points:

1. Membership functions can be viewed as a variant of utility functions or rescaled objective functions, and optimized as such.

2. Combining membership functions, especially using the minimum, can be one approach to criteria aggregation.

3. Multiple-stage decision-making problems based on the minimum aggregation connective can then be stated and solved by means of dynamic programming.

This view was taken over by Tanaka et al.[54] and Zimmermann [86] who developed popular multicriteria linear optimisation techniques in the seventies. The idea is that constraints are soft. Their satisfaction is thus a matter of degree. They can thus be viewed as criteria. The use of the minimum operation instead of the sum for aggregating partial degrees of satisfaction preserves the semantics of constraints since it enforces all of them to be satisfied to some degree. Then any multi-objective linear programming problem becomes a max-min fuzzy linear programming problem.

3.5 Fuzzy relations

Relations being subsets of a Cartesian product of sets, it was natural to make them fuzzy. There are two landmark papers by Zadeh on this topic in the early 1970's. One published in 1971[65] makes the notions of equivalence and ordering fuzzy. A similarity relation extends the notion of equivalence, preserving the properties of reflexivity and symmetry and turning transitivity into maxmin transitivity. Similarity relations come close to the notion of ultrametrics and correspond to nested equivalence relations. As to fuzzy counterparts of order relations, Zadeh introduces first definitions of what will later be mathematical models of a form of fuzzy preference relations [25]. In his attempt the difficult part is the extension of antisymmetry that will be shown to be problematic. The notion of fuzzy preorder, involving a similarity relation turns out to be more natural than the one of a fuzzy order [6].

The other paper on fuzzy relations dates back to 1975 [71] and consists in a fuzzy generalization of the relational algebra. This seminal paper paved the way to the application of fuzzy sets to databases (see [8] for a survey), and to flexible constraint satisfaction in artificial intelligence [16, 41]

3.6 Fuzzy systems

The application of fuzzy sets to systems was not obvious at all, as traditionally systems were described by numerical equations. Zadeh seems to have tried several solutions to come up with a notion of fuzzy system. The idea of fuzzy systems [61] was initially viewed as systems whose state equations involve fuzzy variables or parameters, giving birth to fuzzy classes of systems [61]. Another idea, idea hinted in 1965 [61], was that a system is fuzzy if either its input, its output or its states would range over a family of fuzzy sets. Later in 1971 [67], he suggested that a fuzzy system could be a generalisation of a non-deterministic system, that moves from a state to a fuzzy set of states. So the transition function is a fuzzy mapping, and the transition equation can be captured by means of fuzzy relations, and the sup-min combination of fuzzy sets and fuzzy relations. These early attempts were outlined before the emergence of the idea of fuzzy control [68]. In 1973, though, there was what looks like a significant change, since for the first time it was suggested that fuzziness lies in the description of approximate rules to make the system work. That view, developed first in [69], was the result of a convergence between the idea of system with the ones of fuzzy algorithms introduced earlier [62], and his increased focus of attention on the representation of natural language statements via linguistic variables. In this very seminal paper, systems of fuzzy if-then rules were first described, which paved the way to fuzzy controllers, built from human information, with the tremendous success met by such line of research in the early 1980's. To-day, fuzzy rule-based systems are extracted from data and serve as models of systems more than as controllers. However the linguistic connection is often lost, and such fuzzy systems are rather standard precise systems using membership function for interpolation, than approaches to the handling of poor knowledge in system descriptions.

3.7 Linguistic variables and natural language issues

When introducing fuzzy sets, Zadeh seems to have been chiefly motivated by the representation of human information in natural language. This focus explains why many of Zadeh's papers concern fuzzy languages and linguistic variables. He tried to combine results on formal languages and the idea that the term sets that contain the atoms from which this language was built contain fuzzy sets representing the meaning of elementary concepts [66]. This is the topic where the most numerous and extended papers by Zadeh can be found, especially the large treatise devoted to linguistic variables in 1975 [72] and the papers on the PRUF language [73, 78]. Basically, these papers led to the "computing with words" paradigm, which takes the opposite view of say, logic-based artificial intelligence, by putting the main emphasis, including the calculation method, on the semantics rather than the syntax. Stating with natural language sentences, fuzzy words are precisely modelled by fuzzy sets, and inference comes down to some form of non-linear optimisation. In a later step, numerical results are translated into verbal terms, using the so-called linguistic approximation. While this way of reasoning seems to have been at the core of Zadeh's approach, it is clear that most applications of fuzzy sets only use terms sets and linguistic variables in rather elementary ways. For instance, quite a number of authors define linguistic terms in the form of trapezoidal fuzzy sets on an abstract universe which is not measurable and where addition makes no sense, nor linear membership functions.

In [72], Zadeh makes it clear that linguistic variables refer to objective measurable scales: only the linguistic term has a subjective meaning, while the universe of discourse contains a measurable quantity like height, age, etc.

3.8 Fuzzy intervals

In his longest 3-part paper in 1975 [72], Zadeh points out that a trapezoidal fuzzy set of the real line can model imprecise quantities. These trapezoidal fuzzy sets, called fuzzy numbers, generalize intervals on the real line and appear as the mathematical rendering of fuzzy terms that are the values of linguistic variables on numerical scales. The systematic use of the extension principle to such fuzzy numbers and the idea of applying it to basic operations of arithmetics is a key-idea that will also turn out to be seminal. Since then, numerous papers have developed methods for the calculation with fuzzy numbers. It has been shown that the extension principle enable a generalization of interval calculations, hence opening a whole area of fuzzy sensitivity analysis that can cope with incomplete information in a gradual way. The calculus of fuzzy intervals is instrumental in various areas including:

- systems of linear equations with fuzzy coefficients (a critical survey is [36]) and differential equations with fuzzy initial values, and fuzzy set functions [37];
- fuzzy random variables for the handling of linguistic or imprecise statistical data [26, 10];
- fuzzy regression methods [43];
- operations research and optimisation under uncertainty [31, 15, 32].

3.9 Fuzzy sets and uncertainty: possibility theory

Fuzzy sets can represent uncertainty not because they are fuzzy but because crisp sets are already often used to represent ill-known values or situations, albeit in a crisp way like in interval analysis or in propositional logic. Viewed as representing uncertainty, a set just distinguishes between values that are considered possible and values that are impossible, and fuzzy sets just introduce grades to soften boundaries of an uncertainty set. So in a fuzzy set it is the set that captures uncertainty [20, 21]. This point of view echoes an important distinction made by Zadeh himself [83] between

- conjunctive (fuzzy) sets, where the set is viewed as the conjunction of its elements. This is the case for clusters discussed in the previous section. But also with set-valued attributes like the languages spoken more or less fluently by an individual.
- and disjunctive fuzzy sets which corresponds to mutually exclusive possible values of an ill-known single-valued attribute, like the ill-known birth nationality of an individual.

In the latter case, membership functions of fuzzy sets are called possibility distributions [74] and act as elastic constraints on a precise value. Possibility distributions have an epistemic flavor, since they represent the information we have at our disposal about what values remain more or less possible for the variable under consideration, and what values are (already) known as impossible. Associated with a possibility distribution, is a possibility measure [74], which is a max-decomposable set function. Thus, one can evaluate the possibility of a crisp, or fuzzy, statement of interest, given the available information supposed to be represented by a possibility distribution. It is also important to notice that the introduction of possibility theory by Zadeh was part of the modeling of fuzzy information expressed in natural language [77]. This view contrasts with the motivations of the English economist Shackle [51, 52], interested in a non-probabilistic view of expectation, who had already designed a formally similar theory in the 1940's, but rather based on the idea of degree of impossibility understood as a degree of surprise (using profiles of the form $1 - \mu$, where μ is a membership function). Shackle can also appear as a forerunner of fuzzy sets, of possibility theory, actually.

Curiously, apart from a brief mention in [76], Zadeh does not explicitly use of the notion of necessity (the natural dual of the modal notion of possibility) in his work on possibility theory and approximate reasoning. Still, it is important to distinguish between statements that are *necessarily* true (to some extent), i.e. whose negation is almost impossible, from the statements that are only *possibly* true (to some extent) depending on the way the fuzzy knowledge would be made precise. The simultaneous use of the two notions is often required in applications of possibility theory [23].

3.10 Approximate reasoning

The first illustration of the power of possibility theory proposed by Zadeh was an original theory of approximate reasoning [75, 79, 80], later reworked for emphasizing new points [82, 84], where pieces of knowledge are represented by possibility distributions that fuzzily restrict the possible values of variables, or tuples of variables. These possibility distributions are combined conjunctively, and then projected in order to compute the fuzzy restrictions acting on the variables of interest. This view is the one at work in his calculus of fuzzy relations in 1975. One research direction, quite in the spirit of the objective of "computing with words" [82], would be to further explore the possibility of a syntactic (or symbolic) computation of the inference step (at least for some noticeable fragments of this general approximate reasoning theory), where the obtained results are parameterized by fuzzy set membership functions that would be used only for the final interpretation of the results. An illustration of this idea is at work in possibilistic logic [22], a very elementary formalism handling pairs made of a Boolean formula and a certainty weight, that captures a tractable form of non-monotonic reasoning. Such pairs syntactically encode simple possibility distributions, combined and reasoned from in agreement with Zadeh's theory of approximate reasoning, but more in the tradition of symbolic artificial intelligence than in conformity with the semantic-based methodology of computing with words.

4 Conclusion

The seminal paper on fuzzy sets by Lotfi Zadeh spewed out a large literature, despite its obvious marginality at the time it appeared. There exist many forgotten original papers without off-springs. Why has Zadeh's paper encountered an eventual dramatic success? One reason is certainly that Zadeh, at the time when the fuzzy set paper was released, was already a reknowned scientist in systems engineering. So, his paper, published in a good journal, was visible. However another reason for success lies in the tremendous efforts made by Zadeh in the seventies and the eighties to develop his intuitions in various directions covering many topics from theoretical computer sciences to computational linguistics, from system sciences to decision sciences.

It is not clear that the major successes of fuzzy set theory in applications fully meet the expectations of its founder. Especially there was almost no enduring impact of fuzzy sets on natural language processing and computational linguistics, despite the original motivation and continued effort about the computing with words paradigm. The contribution of fuzzy sets and fuzzy logic was to be found elsewere, in systems engineering, data analysis, multifactorial evaluation, uncertainty modeling, operations research and optimisation, and even mathematics. The notion of fuzzy rule-based system (Takagi-Sugeno form, not Mamdani's, nor even the view developed in [81]) has now been integrated in both the neural net literature and the non-linear control one. These fields, just like in clustering, only use the notion of fuzzy partition and possibly interpolation between subclasses, and bear almost no connection to the issue of fuzzy modeling of natural language. These fields have come of age, and almost no progress can be observed that concern their fuzzy set ingredients. In optimisation, the fuzzy linear programming method is now used in applications, with only minor variants (if we set apart the handling of uncertainty proper, via fuzzy intervals).

However there are some topics where basic research seems to still be active, with high potential. See [11] for a collection of position papers highlighting various perspectives for the future of fuzzy sets. Let us cite two such topics. The notion of fuzzy interval or fuzzy number, introduced in 1975 by Zadeh [72], and considered with the possibility theory lenses, seems to be more promising in terms of further developments because of its connection with non-Bayesian statistics and imprecise probability [39] and its potential for handling uncertainty in risk analysis [1]. Likewise, the study of fuzzy set connectives initiated by Zadeh in 1965, has given rise to a large literature, and significant developments bridging the gap between many-valued logics and multicriteria evaluation, with promising applications (for instance [2]).

Last but not least, we can emphasize the influence of fuzzy sets on some even more mathematically oriented areas, like the strong impact of fuzzy logic on many-valued logics (triggered by P. Hajek [29]) and topological and categorical studies of lattice-valued sets [30]. There are very few papers in the literature that could influence such various areas of scientific investigation to that extent.

References

[1] C. Baudrit D. Guyonnet, and D. Dubois Joint Propagation and Exploitation of Probabilistic and Possibilistic Information in Risk Assessment, IEEE Trans. Fuzzy Systems, 14, 593-608, 2006

[2] G. Beliakov, A. Pradera, T. Calvo. Aggregation Functions: A Guide for Practitioners. Studies in Fuzziness and Soft Computing 221, Springer, 2007.

[3] R. E. Bellman, R. Kalaba, L. A. Zadeh. Abstraction and pattern classification. J. of Mathematical Analysis and Applications, 13,1-7, 1966.

[4] R. E. Bellman and L. A. Zadeh (1970) Decision making in a fuzzy environment, Management Science, 17, B141-B164.

[5] J. Bezdek, J. Keller, R. Krishnapuram, N. Pal , Fuzzy Models for Pattern Recognition and Image processing, Kluwer, 1999.

[6] U. Bodenhofer, B. De Baets, J. C. Fodor A compendium of fuzzy weak orders: Representations and constructions. Fuzzy Sets and Systems 158(8): 811-829, 2007.

[7] M. Black M. Vagueness, Phil. of Science, 4, 427-455,1937.Reprinted in Language and Philosophy: Studies in Method, Cornell University Press, Ithaca and London, 1949, 23-58. Also in Int. J. of General Systems, 17, 1990, 107-128.

[8] P. Bosc, O. Pivert Modeling and querying uncertain relational databases: A survey of approaches based on the possible world semantics, International Journal Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems 18(5), 565-603, 2010.

[9] D. Ciucci, D. Dubois: A map of dependencies among three-valued logics. Inf. Sci. 250: 162-177,2013.

[10] I. Couso, D. Dubois, L. Sanchez. Random Sets and Random Fuzzy Sets as Ill-Perceived Random Variables, Springer, SpringerBriefs in Computational Intelligence, 2014

[11] B. De Baets, D. Dubois, E. Hllermeier, Special Issue Celebrating the 50th Anniversary of Fuzzy Sets. Fuzzy Sets and Systems 281: 1-308 (2015)

[12] B. De Finetti . La logique de la probabilit, Actes Congrs Int. de Philos. Scient., Paris 1935, Hermann et Cie Editions, Paris, pp. IV1- IV9, 1936.

[13] D. Dubois Have Fuzzy Sets Anything to Do with Vagueness ? (with discussion) In : Understanding Vagueness -Logical, Philosophical and Linguistic Perspectives" (Petr Cintula, Chris Fermller, Eds) vol. 36 of Studies in Logic, College Publications, pp. 317-346, 2012.

[14] D. Dubois, F. Esteva, L. Godo, H. Prade. Fuzzy-set based logics - An history-oriented presentation of their main developments. In : Handbook of The history of logic. Dov M. Gabbay, John Woods (Eds.), V. 8, The many valued and nonmonotonic turn in logic, Elsevier, 2007, p. 325-449.

[15] D Dubois, H Fargier, P Fortemps, Fuzzy scheduling: Modelling flexible constraints vs. coping with incomplete knowledge, European Journal of Operational Research 147 (2), 231-252, 2003.

[16] D. Dubois, H. Fargier, H. Prade. Possibility theory in constraint satisfaction problems: Handling priority, preference and uncertainty. Applied Intelligence, 6: 287-309, 1996.

[17] D. Dubois, W. Ostasiewicz, H. Prade . Fuzzy sets: history and basic notions. In: Fundamentals of Fuzzy Sets, (Dubois,D. Prade,H., Eds.), Kluwer, The Handbooks of Fuzzy Sets Series, 21-124, 2000.

[18] D. Dubois, H. Prade Possibility Theory: An Approach to Computerized Processing of Uncertainty, Plenum Press, New York, 1988.

[19] D. Dubois, H. Prade: The three semantics of fuzzy sets. Fuzzy Sets and Systems, 90, 141-150.

[20] D. Dubois, H. Prade. Gradual elements in a fuzzy set. Soft Computing, V. 12, p. 165-175, 2008.

[21] D. Dubois, H. Prade, Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets. Fuzzy Sets and Systems, 192, 3-24, 2012.

[22] D. Dubois, H. Prade. Possibilistic logic - An overview. In: Handbook of the History of Logic. Volume 9: Computational Logic. (J. Siekmann, vol. ed.; D. M. Gabbay, J. Woods, series eds.), 283-342, 2014.

[23] D. Dubois, H. Prade. Possibility theory and its applications: where do we stand? In : Handbook of Computational Intelligence. Janusz Kacprzyk, Witold Pedrycz (Eds.), Springer, p. 31-60, 2015.

[24] Fine K. (1975). Vagueness, truth and logic, Synthese, 30:265-300.

[25] J. Fodor, M. Roubens. Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer Acad. Pub., 1994.

[26] M. A. Gil, G. González-Rodrǵuez, R. Kruse, Eds. Statistics with Imperfect Data, special issue of Inf. Sci. 245: 1-3 (2013)

[27] J. A. Goguen. L-fuzzy sets, J. Math. Anal. Appl., , 8:145-174, 1967.

[28] S. Gottwald. Fuzzy sets theory: Some aspects of the early development, Aspects of vagueness (Skala H.J., Termini S. and Trillas E., eds.), D. Reidel, 13-29, 1984.

[29] P. Hájek. Metamathematics of Fuzzy Logic, Trends in Logic, vol. 4, Kluwer, Dordercht, 1998.

[30] U. Hoehle and S. E. Rodabaugh, eds, Mathematics of Fuzzy Sets: Logic, Topology, and Measure Theory, The Handbooks of Fuzzy Sets Series, Volume 3 (1999), Kluwer Academic Publishers (Dordrecht).

[31] M Inuiguchi, J Ramik Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem Fuzzy sets and systems 111 (1), 3-28, 2000

[32] M. Inuiguchi, J. Watada, D. Dubois, Eds., Special issue on Fuzzy modeling for optimisation and decision support Fuzzy Sets and Systems, Vol. 274, 2015.

[33] A. Kaplan. Definition and specification of meanings, J. Phil., 43, 281-288,1946.

[34] A. Kaplan and H.F. Schott. A calculus for empirical classes, Methods, III, 165-188, 1951.

[35] S.C. Kleene, Introduction to Metamathematics, NorthHolland Pub. Co, Amsterdam, 1952.

[36] W. A. Lodwick and D. Dubois, Interval linear systems as a necessary step in fuzzy linear systems, Fuzzy Sets and Systems 281: 227-251 (2015).

[37] W. A. Lodwick and M. Oberguggenberger, Eds, Differential Equations Over Fuzzy Spaces - Theory, Applications, and Algorithms, special issue of Fuzzy Sets and Systems 230: 1-162, 2013.

[38] J. Łukasiewicz. Philosophical remarks on many-valued systems of propositional logic, 1930. Reprinted in Selected Works (Borkowski, ed.), Studies in Logic and the Foundations of Mathematics, North-Holland, Amsterdam, 1970, 153-179.

[39] G. Mauris. Possibility distributions: A unified representation of usual direct-probability-based parameter estimation methods. Int. J. Approx. Reasoning, 52 (9),1232-1242, 2011.

[40] K. Menger. Ensembles flous et fonctions aleatoires, Comptes Rendus de lAcadmie des Sciences de Paris, 232, 2001-2003, 1951.

[41] P. Meseguer, F. Rossi, T. Schiex, Soft constraints, Chapter 9 in Foundations of Artificial Intelligence, Volume 2, 2006, 281-328.

[42] D. Mundici Bookmaking over infinite-valued events, International Journal of Approximate Reasoning, Volume 43, Issue 3, December 2006, Pages 223-240

[43] S. Muzzioli, A. Ruggieri, B. De Baets A comparison of fuzzy regression methods for the estimation of the implied volatility smile function, Fuzzy Sets and Systems, 266:131-143, 2015

[44] V. Novák Are fuzzy sets a reasonable tool for modeling vague phenomena?Fuzzy Sets and Systems, Volume 156, Issue 3, 16 December 2005, Pages 341-348

[45] W. Ostasiewicz. Pioneers of fuzziness, Busefal, 46, 4-15, 1991.

[46] W. Ostasiewicz. Half a century of fuzzy sets. Supplement to Kybernetika 28:17-20, 1992 (Proc. of the Inter. Symp. on Fuzzy Approach to Reasoning and Decision Making, Bechyne, Czechoslovakia, June 25-29, 1990).

[47] C. S.Peirce. Collected Papers of Charles Sanders Peirce, C. Hartshorne and P. Weiss, eds., Harvard University Press, Cambridge, MA, 1931.

[48] E. H. Ruspini. A new approach to clustering. Inform. and Control, 15, 22-32, 1969.

[49] B. Russell B. Vagueness, Austr. J. of Philosophy, 1, 84-92, 1923.

[50] R. Seising. Not, or, and, - Not an end and not no end! The "Enric-Trillas-path" in fuzzy logic. In: Accuracy and fuzziness. A Life in Science and Politics. A festschrift book to Enric Trillas Ruiz, (R. Seising, L. Argüelles Méndez, eds.), 1-59, Springer, 2015.

[51] G. L. S. Shackle. Expectation in Economics. Cambridge University Press, UK, 1949. 2nd edition, 1952.

[52] G. L. S. Shackle. Decision, Order and Time in Human Affairs. (2nd edition), Cambridge University Press, UK, 1961.

[53] P. Smith and R. Keefe. Vagueness: A Reader. MIT Press, Cambridge, MA, 1997.

[54] H. Tanaka, T. Okuda and K. Asai. On fuzzy-mathematical programming, Journal of Cybernetics, 3(4): 37-46, 1974.

[55] H. Weyl H. Mathematic and logic, Amer. Math. Month., 53, 2-13, 1946.

[56] T. Williamson. Vagueness. Routledge, London, 1994.

[57] Wittgenstein L. (1953). Philosophical Investigations, Macmillan, New York.

[58] L. A. Zadeh. Thinking machines. A new field in electrical engineering. Columbia Engineering Quarterly, 3, 12-13 & 30-31, 1950.

[59] L. A. Zadeh. From circuit theory to system theory. Proc. I.R.E., 50, 856-865, 1962.

[60] L. A. Zadeh. Fuzzy sets. Information and Control, 8 (3), 338-353,1965.

[61] L. A. Zadeh. Fuzzy sets and systems. In: Systems Theory (J. Fox, Ed.) (Proc. Simp. on System Theory, New York, April 20-22, 1965), Polytechnic Press, Brooklyn, N.Y., 29-37, 1966 (reprinted in Int. J. General Systems, 17: 129-138, 1990).

[62] L. A. Zadeh. Fuzzy algorithms. Information and Control, 12, 94-102,1968.

[63] L. A. Zadeh, Probability measures of fuzzy events, J. Math. Anal. Appl. 23 (1968), 421-427

[64] L. A. Zadeh. Biological applications of the theory of fuzzy sets and systems. In: Biocybernetics of the Central Nervous System (with a discussion by W. L. Kilmer, (L. D. Proctor, ed.), Little, Brown & Co., , Boston, 199-206 (discussion pp. 207-212), 1969.

[65] L.A. Zadeh, Similarity relations and fuzzy orderings, Inf. Sci. 3 (1971) 177-200.

[66] L.A. Zadeh, Quantitative fuzzy semantics, Inf. Sci. 3 (1971) 159-176.

[67] L.A. Zadeh,Toward a theory of fuzzy systems, Aspects of Network and System Theory, R.E. Kalman and N. De Claris, Eds. New York: Holt, Rinehart and Winston pp. 469-490, 1971 (originally NASA Contractor Report 1432, Sept. 1969).

[68] L.A. Zadeh, A rationale for fuzzy control. J. of Dynamic Systems, Measurement, and Control, Trans. of the ASME, March issue, 3-4, 1972.

[69] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. on Systems, Man, and Cybernetics, 3 (1), 28-44, 1973.

[70] L.A. Zadeh. Fuzzy Logic and approximate reasoning, Synthese, 30, 407-428, 1975.

[71] L. A. Zadeh, Calculus of fuzzy restrictions. In: Fuzzy sets and Their Applications to Cognitive and Decision Processes, (L. A. Zadeh, K. S. Fu, K. Tanaka, M. Shimura, eds.), Proc. U.S.-Japan Seminar on Fuzzy Sets and Their Applications, Berkeley, July 1-4, 1974, Academic Press, 1-39, 1975.

[72] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. Inf. Sci., Part I, 8 (3), 199-249; Part II, 8 (4), 301-357; Part III, 9 (1), 43-80,1975.

[73] L. A. Zadeh, PRUF - a meaning representation language for natural languages. Int. J. Man-Machine Studies, 10, 395-460, 1978.

[74] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1, 3-28, 1978.

[75] L. A. Zadeh. A theory of approximate reasoning. In: Machine Intelligence, Vol. 9, (J. E. Hayes, D. Mitchie, and L. I. Mikulich, eds.), 149-194, 1979.

[76] L.A. Zadeh, Fuzzy sets and information granularity. In: Advances in Fuzzy Set Theory and Applications, (M. M. Gupta, R. K. Ragade, R. R. Yager, eds.), North-Holland, 3-18, 1979.

[77] L. A. Zadeh: Possibility theory and soft data analysis. In: L. Cobb, R. Thrall, eds., Mathematical Frontiers of Social and Policy Sciences, Boulder, Co.: Westview Press, pages 69-129, 1982.

[78] L. A. Zadeh. Precisiation of meaning via translation into PRUF. In: Cognitive Constraints on Communication, (L. Vaina, J. Hintikka, eds.), Reidel, Dordrecht, 373-402, 1984.

[79] L. A. Zadeh. Syllogistic reasoning in fuzzy logic and its application to usuality and reasoning with dispositions. IEEE Transactions on Systems, Man, and Cybernetics, 15 (6), 754-763,1985.

[80] L. A. Zadeh. Knowledge representation in fuzzy logic. IEEE Trans. Knowl. Data Eng., 1 (1), 89-100, 1989.

[81] L. A. Zadeh. The calculus of fuzzy if-then rules. AI Expert, 7 (3), 23-27,1992.

[82] L. A. Zadeh. Fuzzy logic = computing with words. IEEE Trans. Fuzzy Systems 4(2): 103-111,1996.

[83] L. A. Zadeh. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems, 90, 111-128, 1997.

[84] L. A. Zadeh, Generalized theory of uncertainty (GTU) – principal concepts and ideas. Computational Statistics & Data Analysis, 51, 15-46, 2006.

[85] L. A. Zadeh Fuzzy logic - a personal perspective. Fuzzy Sets and Systems 281: 4-20 (2015)

[86] H. -J. Zimmermann. Fuzzy programming and linear programming with several objective functions Fuzzy Sets and Systems, 1: 45-55, 1978.

The first steps in fuzzy set theory in France forty years ago (and before) *

Didier Dubois and Henri Prade
IRIT-CNRS, Université Paul Sabatier, 31062 Toulouse Cedex 09, France

Abstract

At the occasion of the fiftieth anniversary of the founding article "Fuzzy sets" by L. A. Zadeh, we briefly outline the beginnings of fuzzy set research in France, taking place some ten years later, pointing out the pioneering role of Arnold Kaufmann and few others in this emergence. Moreover, whe also point out that the French counterpart of the name "fuzzy set" had appeared some 15 years before Zadeh's paper, in a paper written in French by the very person who also invented triangular norms in the 1940's. [1]

Keywords: fuzzy set; fuzzy logic; history.

1 Before the beginning

Strangely enough, the phrase "ensembles flous" (the French translation of "fuzzy sets") first appeared in a paper published in French in the Compte-Rendus of the French Academy of Sciences in 1951 [68] by Karl Menger (1902-1985). He was an Austrian mathematician [90] who emigrated to the USA before the second World War. He is the son of the well-known economist Carl Menger (1840-1921) himself one of the fathers of the theory of subjective utility value. Karl Menger was an active member of the Vienna Circle; later he was at the origin of triangular norms with the paper "Statistical metrics" [67], where triangular norms emerge in stochastic geometry from the generalization of the classical triangle inequality when distances between two elements of a metric space are represented by probability distributions rather than by numbers. His spectrum of interest, which was very wide [70], included logic. He especially proposed a "logic of the doubtful" where (italics are from Menger himself):

*To appear in the special issue on "50 Years of Fuzzy Sets" in the *Archives for the Philosophy and History of Soft Computing*, an Online Journal.

[1] This paper is a translated and expanded version of a conference paper [29].

> we divide the propositions into three mutually exclusive classes of modality: μ_+ consisting of the *asserted*, μ_0 consisting of the *doubtful*, μ_- consisting of the *negated* propositions. $[\cdots]$ In contrast to the traditional 3-valued logic, *the modality of a compound is not determined by the modalities of the components*." [66]

He was thus making it clear very early that uncertainty is not compositional.

What is truly remarkable in Menger's 1951 paper is not only the use of the French counterpart to "fuzzy sets", but also its application to a notion closely related to Zadeh's idea: in fuzzy set terminology, the paper is about max-product transitive fuzzy relations! However, in Menger's paper, we can read (p. 2002):

> Nous appellerons cette fonction même un *ensemble flou* et nous interpréterons $\Pi_F(x)$ comme la probabilité que x appartienne à cet ensemble. Si Π_F ne prend que les valeurs 1 et 0, il s'agit essentiellement d'un sous-ensemble de U au sens classique et nous parlerons d'ensemble *rigide*. [2]

Moreover, Menger was soon aware of the emergence of fuzzy sets since one year after the publication of Zadeh's seminal paper [98], he wrote [69]:

> In 1951, I suggested that, besides studying well-defined sets, it might be necessary to develop a theory in which the element-set-relation is replaced by *the probability an element belonging to a set.* In a Paris note [Ref] [3], I called such an object, in contrast to an ordinary or rigid set, *ensemble flou* (= hazy set)." In a slightly different terminology, this idea was recently expressed by Bellman, Kalaba and Zadeh [Ref][4] under the name of *fuzzy set.* (These authors speak of the degree rather than than the probability an element belonging to a set.)

Thus, the distinction between probability and degree of membership was very clear for Menger from the beginning; see [23, 93] for further discussions.

A worth noticing coincidence took place on May 28, 1951, the day when the French mathematician Arnaud Denjoy (1884-1974) transmitted Karl Menger's communication on "ensembles flous" to the French Academy of Sciences. Indeed, Arnaud Denjoy also transmitted on the same day a communication by Gustave Choquet [15] whose abstract was

> En vue d'une théorie des fonctions non additives d'ensembles, on définit et l'on étudie la classe des sous-ensembles des espaces séparés

[2] In English: "We shall call such a function a *fuzzy set* and we shall interpret $\Pi_F(x)$ as the probability that x belongs to this set. If Π_F only takes the values 1 and 0, it is essentially a subset of U in the classical sense and we shall speak of *crisp* set."

[3] Reference [68].

[4] R. Bellman, R. Kalaba, L. Zadeh. Abstraction and pattern classification. J. of Mathematical Analysis and Applications, 13 (1), 1-7, January 1966.

engendrée à partir des compacts par réunions ou intersections dénombrables et par applications continues"[5]

Thus on the same day where the phrase "ensemble flou" appeared, elements towards the theory of capacities (i.e., "fuzzy measures" in Sugeno's terminology [94]) and Choquet integrals [16] (which would appear later as the quantitative counterpart of Sugeno integrals) were also presented. The fuzzy future thus began on May 28, 1951, even if several decades and a significant research effort would still be necessary before the full landscape could be put together.

Besides, another piece of early work, apparently written independently from Zadeh's pioneering paper, is also worth mentioning. It is an article in French published in 1968 by a French linguist, Yves Gentilhomme (b. in 1920) in a Romanian journal [38]. In this paper, Gentilhomme calls "ensemble flou" a nested pair of subsets, one gathering what he regards as "the central elements", while the second larger subset also includes "peripheral elements". Gentilhomme motivates his proposal by an example of "hypergrammaticality" in texts (illustrated by a poem by Alphonse Allais where the author intentionally makes an abusive use of the imperfect tense of the subjunctive mood in order to produce a comical effect) and by an example of more or less credible words in French built from the same root. Then Gentilhomme provides a formal set-theoretic apparatus for combining his "ensembles flous", and he proposes to assign a degree of membership equal to 1/2 to peripheral elements (those that are not central). It is in the 1974 pioneering research monograph by Negoita and Ralescu [75] (who were working in Romania at that time) that Gentilhomme's paper is first reported and put in relation with Zadeh's work; "ensembles flous" are translated by "flou sets" in the English version of the book the year after [76].

2 Arnold Kaufmann

Edwin Diday (b. 1940) seems to have published the first journal paper in France influenced by the fuzzy set idea [22] in 1972. The paper presents a new approach to (fuzzy) clustering, called the dynamical cloud method (in French, "méthode des nuées[6] dynamiques"), and cites Zadeh [98] and Ruspini [86].

However, it is Arnold Kaufmann (1911-1994) [27, 28] who unquestionably introduced fuzzy set theory in France. He was an applied mathematician, author, or co-author of a long series of books covering many areas in engineering mathematics, including automatic control and operations research. His books, many of which were translated into English, were not only covering standard applied mathematics, but also many advanced topics in relation with current research

[5]In English: "In view of a theory of non additive set functions, one defines the class of the subsets of separated spaces, generated from compacts by denumerable unions or intersections and by continuous mappings."

[6]the use of this word , which means "clouds" reminds us that at the beginning Zadeh was hesitating between the words "fuzzy" and "cloudy": Indeed, he wrote in 1962: "we need a radically different kind of mathematics, the mathematics of fuzzy or cloudy quantities which are not describable in terms of probability distributions." in [97]

at that time[19, 20, 21, 39, 59, 40, 41, 42, 43, 57, 58, 60, 44, 46, 47, 48, 61][7]. As he was in contact with Lotfi Zadeh, he heard about fuzzy sets very early, and was quickly enthusiastic about this challenging way of thinking. He was the first in the world to publish a monograph on the theory of fuzzy (sub)sets in 1973 [49]. It was translated into English two years later [52][8]. The first 1973 volume was soon followed by three other ones [50, 51, 53], and by a book of exercises [56] (with Michel Cools and Thierry Dubois). Altogether, this series of five volumes mainly cover fuzzy set theoretic operations, fuzzy relations, and their applications to many fields: classification and pattern recognition, automata and systems, multicriteria decision, as well as linguistics, logic, topology, matroïds, etc. These books, and in particular the first one, had a great impact on the dissemination of fuzzy set theory in France. Continuing to write books on fuzzy logic-related topics, his strong interest for the topic never waned until the end of his life.

Let us quote the last paragraph of the conclusion of his first fuzzy set book [49]:

> Je voudrais exprimer un souhait très sincère. Je voudrais que mes lecteurs, initiés et intéressés par mon modeste travail, puissent aller plus loin, beaucoup plus loin, encore plus loin. Les sciences humaines ont besoin d'une mathématique appropriée à notre nature, à nos attitudes floues, à notre comportement nuancé, à nos dosages, à nos critères multiples. Si ce premier livre est suffisamment stimulant, de nombreux articles, concernant les aspects théoriques ou les applications, seront publiés par des lecteurs; des livres concurrents verront le jour. Tout ceci pour l'amélioration rapide de nos méthodes en vue d'aborder les sciences humaines."[9]

Beyond the fact that enthusiasm and generosity permeate this text, and the correctness of this prediction, it is also worth noticing that Kaufmann was considering that the applications of fuzzy set theory would be human-oriented sciences, while this is not so clear as of to-day. We have to remember that Kaufmann was writing this text at a time where information processing and artificial intelligence were still in infancy.

At the time when he got acquainted with fuzzy sets, Kaufmann was deeply interested in methods for helping creativeness[45], a topic on which he published a book later in 1979 [54]. So, it was one of the first areas where he considered

[7] This list by no means claims to exhaustiveness!

[8] This more theoretical book, first written in Romanian in 1974 [75] would appear in English also in 1975 [76].

[9] "I would like to express a very sincere wish. I would like that my readers, taught and interested by my modest work, go farther, much farther and farther. Human-oriented sciences need a kind of mathematics that fits our nature, our fuzzy attitudes, our nuanced behavior, our balanced judgments, our multiple criteria. If this first book is sufficiently stimulating, many articles, dealing with theoretical or practical aspects, will be published by my readers; concurrent monographs will appear. All this, for the fast improvement of our methods for coping with human-oriented sciences."

applying fuzzy sets [1, 55, 53], for which he proposed a lattice and fuzzy relation-based approach, in the spirit of ideas and methods preiously advocated by Abraham Moles (1920-1992), an engineer by training, then a philosopher, working on the sociology and psychology of information and communication sciences [71, 72, 73] [74]. Kaufmann's approach to creativeness were also developed by his co-authors Michel Cools and Monique Peteau [17].

3 Elie Sanchez - Claude Ponsard - Robert Féron

The first three main followers of Arnold Kaufmann in France in the mid-1970's are Elie Sanchez, Claude Ponsard, and Robert Féron .

Elie Sanchez (1944-2014) [12, 96, 92] was the first in France, in 1974 in Marseilles, to defend a thesis on fuzzy set methods. His thesis is landmark piece of work on fuzzy relation equations, which contains important results on the solving of these equations [88]. This research was motivated by an attempt at a mathematical formalization of medical diagnosis. It had a very strong influence on the development of fuzzy set methods worldwide.

Claude Ponsard (1927-1990) [9, 26, 37], working in Dijon, started to propose in 1975 to apply fuzzy sets to various problems in economics [81, 82]. He then soon after led a small group of researchers on these questions, including Bernard Fustier [36] and Régis Deloche [18].

Robert Féron (b. 1921) [5] is a statistician working in econometrics in Lyon. He was the first to properly provide a theoretical basis for the study of fuzzy random sets and to advocate their interest [31, 32, 33, 34]. His writings, mostly in French, and mostly published in a journal having a limited circulation, would remain unfortunately largely ignored in the English American world. Let us also point out on the same kind of topic and published in the same place a paper by Robert Fortet and Mehri Kambouzia [35].

We should also mention the two pioneering papers by Jean-Pierre Aubin (b. 1939) on game theory with fuzzy cores (the set of multistrategies that are not rejected by any coalition) published at that time [3, 4].

4 1974-1976: The pivotal years

From 1974-1976 on, the number of young French researchers interested in fuzzy sets started to increase (even if the topic was not very popular and remained highly controversial especially in France (many people were considering that it was not a "serious" topic to work on - partly because of the name -, and whose connection / difference with probability was unclear). Let us provide a list of persons in France who began to use fuzzy sets in their works at that time:

- Bernadette Bouchon, a member of the team headed by Claude-François Picard (1926-1979) [79] started working on fuzzy questionaries [10],

- Jacques Brémont (1938-2014) defended his thesis on the use of fuzzy sets on speech recognition in 1975 [13] under the supervision of Michel Lamotte

[14, 64]. Altogether with Gérard Hirsch (b. 1938) [7] they later on formed an active research group on fuzzy set methods in Nancy for about two decades.

- The first French works on fuzzy systems were initiated by Pierre Vidal [63] in Lille in 1974, together with Noël Malvache (1943-2007) who then continued to work on fuzzy rule-based controllers with Didier Willaeys in Valenciennes [95].

- the year 1975 sees the publication of the first fuzzy set papers by two French pure mathematicians, Daniel Ponasse [80] and S. Ribeyre [85]. The former, once back in France, launched a seminar on “Fuzzy Mathematics” in Lyon, which would become very productive in the late 1970’s and the 1980’s. This group included Achille Achache (b. 1934), Nicole Blanchard, Odile Botta, Josette and Jean-Louis Coulon, Marianne Delorme, and Christiane Dujet. Let us also mention Michel Eytan [30] on this mathematical side.

Although the following people have been more briefly involved with fuzzy sets, one may still mention:

- in automation of production processes, the thesis of Moncef Ben Salem (1953-2015) [8] where a fuzzy multicriteria automatic decision-making procedure is proposed for determining the sequencing of operations accomplished by a machine tool. Later, Ben Salem became minister of Higher Education and Scientific Research (2011-2014), after the Tunisian revolution. Lucas Pun [84] was one of the very first in France to foresee the potential interest of fuzzy sets in the modeling of production processes.

- Jean-Marc Adamo (b. 1943) [2] started working in the second half of the 1970’s for some years on dynamical systems and then on fuzzy programming languages.

- Jean-Philippe Massonie [65] in Besançon was the first in France to foresee the potential interest of fuzzy sets in geographical modeling, thus initiating a line of research that still exists in France.

What is more unexpected is that fuzzy sets were also a source of inspiration in the French avant-garde literature. The French novelist Claude Ollier [77], a writer close to the “Nouveau Roman” movement, seems to have met Arnold Kaufmann in September 1970, at a meeting about creativeness in art and science [1], where Kaufmann already spoke about fuzzy sets. In this French novel with an English title “Fuzzy sets”, the author plays with the roles of the protagonists and the reader in the story, as well as with the display of the text on the pages. The novel was reprinted with a slightly less exotic page display two decades later [78]. Let us mention a more classical writer, Jacques Laurent, who published a novel also with the title “Les sous-ensembles flous” (this time in French) a bit later in 1981 [62]. In this book, the fuzzy set idea applies at several levels to the links between the characters and the forces that drive them.

Lastly, it is in 1976 that the authors of this note produced their first (hand-written !) research report of [24] (now indexed by Google Books). This was mainly a survey and a status report. Our first published contributions only appeared one year later [83, 25, **?**]

5 Conclusion

This note is an attempt at offering a short overview of the first years of research in France regarding fuzzy sets and their applications. Only references from mid-seventies and before have been reported, without mentioning further developments of the works of the authors cited. In fact, some authors have encountered fuzzy sets very briefly in their research in this time period, while others have continued to contribute to fuzzy sets for several decades.

As shown by this brief overview, fuzzy set research in France (see [11, **?**] for general overviews), starts in the years 1973-1976, and immediately deals with very different issues. They are led by researchers relatively isolated from one another, who often face suspicion, negative critique, and sometimes disparagement and bashing from their academic colleagues. Nevertheless, those times were more open-minded than the present period that seems to be under the tyranny and normalization of citation rates and impact factors!

It is also worth noticing that the first works rely mainly on fuzzy set operations and on fuzzy relations, and that many important notions have no role in these works, even if they already exist as the extension principle [98], or the notions of fuzzy measures and fuzzy integrals in the sense of Sugeno [94][10]. It is a matter of facts that some crucial developments would only appear a bit later, such as possibility theory [99], or the linkage between fuzzy set connectives and triangular norms, originally introduced in the study of probabilistic metric spaces, whose father was precisely Karl Menger, the man who first used the phrase "ensemble flou"!

References

[1] Collective (J. Bertrand, A. Flocon, M. Fustier, J. Jacques, A. Kaufmann, R. Leclercq, C. Mathieu-Batsch, C. Ollier, J. Ricardou, J.-C. Risset). Art et Science: De la Créativité. Colloquium held on Sept. 11-16, 1970 at the Centre Culturel International de Cerisy-La-Salle, France, UGE, 10-18 series, n°697, 1972.

[2] J.-M. Adamo. Towards the introduction of fuzzy concepts in dynamic modeling. Proc. Conf. on Dynamic Modelling and Control of National Economies, Vienna, North-Holland, 1977.

[3] J.-P. Aubin. Cœur et valeur des jeux flous à paiements latéraux. CRAS, 279, 891-894, 1974.

[4] J.-P. Aubin. Cœur et valeur des jeux flous sans paiements latéraux. CRAS, 279, 963-966, 1974.

[5] J.-P. Auray, H. Prade. Robert Féron: A pioneer in soft methods for probability and statistics. In: Soft Methods for Handling Variability and Imprecision,

[10] Interestingly enough, Michio Sugeno spent the academic year 1976-1977 in Toulouse, at the LAAS laboratory, in the research group of José Aguilar Martin [91] and Gérald Banon [6]. The authors of this note were lucky enough to meet Sugeno and learn about his research by the end of his stay in Toulouse.

Selected papers from the 4th International Conference on Soft Methods in Probability and Statistics, (SMPS'08), (D. Dubois, M. Asunción Lubiano, H. Prade, M. Angeles Gil, P. Grzegorzewski, O. Hryniewicz, eds.), Toulouse, Sept. 8-10, Advances in Soft Computing 48, Springer, 27-32, 2008.

[6] G. Banon. Distinction entre plusieurs sous-ensembles de mesures floues. Note interne 78.1. 11, LAAS-AS, Toulouse, France, 1978.

[7] D. Benlahcen, G. Hirsch, M. Lamotte. Codage et minimization des fonctions floues dans une algèbre floue. R.A.I.R.O. Automatique / Syst. Anal. & Control, 11, 17-31, 1977.

[8] M. Ben Salem. Sur l'extension de la théorie des sous-ensembles flous à l'automatique industrielle. Mise au point d'une nouvelle méthode d'étude des gammes d'usinage. Thèse Univ. Université Paris 7, 1976.

[9] A. Billot, J.-F. Thisse. Claude Ponsard (1927-1990): A biographical essay. Fuzzy Sets and Systems, 49 (1), 3-8, 1992. & The Annals of Regional Science, 26 (3), 191-198, 1992.

[10] B. Bouchon. Du flou dans les questionnaires. In : Information, Questionnaires, et Reconnaissance, (C.- F. Picard, ed.) (Actes Journées de la Société Mathématique de France, Bonas, 20-23 Sept., rencontre organisée par M. Terrenoire), Structures de l'Information publications, Paris, n°2, 111-122, 1976.

[11] B. Bouchon-Meunier. Fuzzy logic in France. Mathware & Soft Computing Magazine, 20 (2), 11-14, 2013.

[12] B. Bouchon-Meunier. Foreword. Int. J. Unc. Fuzz. Knowl. Based Syst., 22 (2), 179-180, 2014.

[13] J. Brémont, Contribution à la reconnaissance automatique de la parole par les sous-ensembles flous. Thèse de l'Université de Nancy, 1975.

[14] J. Brémont, M. Lamotte. Contribution à la reconnaissance automatique de la parole en temps réél par la considération des sous-ensembles flous. CRAS Paris, July 15,1974.

[15] Ensembles boréliens et analytiques dans les espaces topologiques. G. Choquet. C. R. Acad. Sci., Paris, 232, 2174-2176, 1951.

[16] G. Choquet. Theory of capacities. Annales de l'Institut Fourier, 5, 131-295, 1953.

[17] M. Cools, M. Peteau. Un programme de stimulation inventive: STIM 5. R.A.I.R.O. Oper. Res., 8 (3), 5-19, 1974.

[18] R. Deloche. Théorie des sous-ensembles flous et classification en analyse économique spatiale. Working paper IME, n°11, Fac. Sci. Econom. Gestion, Dijon, 1975.

[19] M. Denis-Papin, A. Kaufmann. Cours de calcul matriciel appliqué. Cours de mathématiques supérieures appliquées (Mathématiques modernes): Albin Michel, 1961.

[20] M. Denis-Papin, A. Kaufmann. Cours de calcul tensoriel appliqué (géométrie différentielle absolue). Cours de mathématiques supérieures appliquées (Mathématiques modernes): Albin Michel, 1964.

[21] M. Denis-Papin, A. Kaufmann, R. Faure. Cours de calcul booléien appliqué (notions sur les ensembles et les treillis, algèbres booléiennes, algèbre binaire), Cours de mathématiques supérieures appliquées (Mathématiques modernes),vol. 5, foreword by René de Possel, Albin Michel, 1963.

[22] E. Diday. Optimisation en classification automatique et reconnaissance des formes. Revue franaise d'automatique, d'informatique et de recherche opérationnelle (R.A.I.R.O.) - Recherche opérationnelle, 6 (3), 61-95, 1972.

[23] D. Dubois, W. Ostasiewicz, H. Prade. Fuzzy sets: History and basic notions. In : Fundamentals of Fuzzy Sets, (D. Dubois, H.Prade, eds.), The Handbooks of Fuzzy Sets Series, Kluwer Acad. Publ., Boston, 21-124, 2000.

[24] D. Dubois, H. Prade. Le flou, kouacksexa ? Tech. Rep. C.E.R.T.-D.E.R.A., Toulouse, 170 p., Oct. 1976.

[25] D. Dubois, H. Prade. Algorithmes de plus courts chemins pour traiter des données floues. R.A.I.R.O. - Oper. Res., 12 (2), 213-227, 1978.

[26] D. Dubois, H. Prade. Obituary - Claude Ponsard. BUSEFAL, (LSI, Univ. P. Sabatier), n°42, 5-6, 1990.

[27] D. Dubois, H. Prade. Obituary - Arnold Kaufmann. Fuzzy Sets and Systems, 69 (2), 1995, 103.

[28] D. Dubois, H. Prade. In memoriam: Arnold Kaufmann. Int. J. of General Systems, 25 (1), 1996, 3-6.

[29] D. Dubois, H. Prade. Les débuts des ensembles flous en France il y a quarante ans. Actes 24eme Conférence sur la Logique Floue et ses Applications, Poitiers, Nov. 5-6, Cépaduès, Toulouse, 12-15, 2015.

[30] M. Eytan. Sémantique préordonnée des ensembles flous. Actes Congrès AFCET 'Modélisation et Maîtrise des Systèèmes', Editions Hommes & Techniques, Vol. 2, 601-608, Versailles, Nov. 1977.

[31] R. Féron. Ensembles aléatoires flous. Note CRAS, t. 282, Série A, 26 Avril 1976, 903-906.

[32] R. Féron. Economie d'échange aléatoire floue. Note CRAS, t. 282, Série A, 1976, 1379-1382.

[33] R. Féron. Ensembles flous, ensembles aléatoires flous, et économie aléatoire floue. Publications Econométriques, vol. IX, Fasc. 1, 25-64, 1976.

[34] R. Féron. Ensembles flous attachés à un ensemble aléatoire flou Publications Econométriques, vol. IX, Fasc. 2, 51-65, 1976.

[35] R. Fortet, M. Kambouzia. Ensembles aléatoires et ensembles flous. Publications Econométriques, vol. IX, Fasc. 1, 1-23, 1976.

[36] B. Fustier. L'attraction des points de vente dans des espaces précis et imprécis. Working paper IME, n°10, Fac. Sci. Econom. Gestion, Dijon, 1975.

[37] B. Fustier. Les apports des sous-ensembles flous à l'économie : les travaux des économistes dijonnais durant les années 80. Actes 23^e Rencontres Francophones sur la Logique Floue est ses Applications (LFA'14), Cargèse, Oct. 22-24, Cépaduès, 3-4, 2014.

[38] Y. Gentilhomme. Les ensembles flous en linguistique. Cahiers de Linguistique Théorique et Appliquée (Bucarest), 5, 47-63, 1968.

[39] A. Kaufmann, R. Cruon. Les Phénomènes d'attente. Théorie et applications. Dunod, Paris, 1961.

[40] A. Kaufmann. Methods and Models of Operations Research. Prentice-Hall, Englewood Cliffs, 1963.

[41] A. Kaufmann. A. Kaufmann, R. Cruon. Dynamic Programming: Sequential Scientific Management. Academic Press, 1967.

[42] A. Kaufmann. L'Homme d'Action et la Science. Coll. L'Univers des Connaissances, Hachette, 1968.

[43] A. Kaufmann. The Science of Decision-Making. An introduction to praxeology Weidenfeld & Nicolson London, and McGraw-Hill, 1968.

[44] A. Kaufmann. Graphs, Dynamic Programming and finite games. Mathematics in Science and Engineering Series, Vol. 36, Academic Press, New York, 1969.

[45] A. Kaufmann. L'imagination artificielle - (Heuristique automatique). Revue franaise d'automatique, d'informatique et de recherche opérationnelle (R.A.I.R.O.) - Recherche opérationnelle, 3 (3), 5-24, 1969.

[46] A. Kaufmann. Méthodes et Modèles de la Recherche Operationnelle (Les Mathématiques de l'Entreprise). Dunod Paris, vol. I, 1970; vol II,1972; vol. III, 1974.

[47] A. Kaufmann. Points and Arrows: Theory of Graphs (Student Library). Corgi Childrens, 1972.

[48] A. Kaufmann. Reliability: A Mathematical Approach (Student Library) Corgi Childrens,1972.

[49] A. Kaufmann. Introduction à la Théorie des Sous-Ensembles Flous à l'usage des ingénieurs. Vol. 1 Eléments Théoriques de Base. Masson, Paris, 1973.

[50] A. Kaufmann. Introduction à la Théorie des Sous-Ensembles Flous. Vol. 2. Applications à la Linguistique, à la Logique, et à la Sémantique. Masson, Paris, 1975.

[51] A. Kaufmann. Introduction à la Théorie des Sous-Ensembles Flous. Vol. 3. Applications à la Classification et à la Reconnaissance des Formes, aux Automates et aux Systèmes, et au Choix des Critères. Masson, Paris, 1975.

[52] A. Kaufmann. Introduction to the Theory of Fuzzy Subsets. Vol. 1 Academic Press, New York, 1975.

[53] A. Kaufmann. Introduction à la Théorie des Sous-Ensembles Flous. Vol. 4. Compléments et Nouvelles Applications. Masson, Paris, 1977.

[54] A. Kaufmann. Modèles Mathématiques pour la Stimulation Inventive. Albin-Michel, 1979.

[55] A. Kaufmann, M. Cools, T. Dubois. Stimulation inventive danss un dialogue homme-machine utilisant la méthode des morphologies et la théorie des sous-ensembles flous. IMAGO Discussion Paper n°6, Université Catholique de Louvain, Louvain, Belgium.

[56] A. Kaufmann, T. Dubois, M. Cools. Exercices avec Solutions sur la Théorie des Sous-Ensembles Flous. Masson, Paris, 1975.

[57] A. Kaufmann, R. Faure. Invitation à la recherche opérationnelle. Coll. Initiation aux Nouveautés de la Science, Dunod, 1968.

[58] A. Kaufmann, R. Faure. Introduction to operations research, Mathematics in Science and Engineering Series, vol. 47, Academic Press, 1968.

[59] A. Kaufmann, B. Grabowski, J. Thouzery. Analyse des réseaux électriques à tubes et à transistors. Eyrolles, 1963.

[60] A. Kaufmann, R. Douriaux, G. Cullmann. Exercices de calcul des probabilités, Eyrolles, 1968.

[61] A. Kaufmann, A. Henry-Labordère. Integer and Mixed Programming Theory and Applications, Mathematics in Science & Engineering series, vol. 137, Academic Press,1977.

[62] J. Laurent. Les Sous-Ensembles Flous. Grasset, Paris, 1981.

[63] N. Malvache, P. Vidal. Application des systèmes flous à la modélisation des phénomènes de prise de décision et d'appréhension des informations visuelles chez l'homme. Final report of the CNRS ATP ("Action Thématique Programmée), Université de Lille 1, 1974.

[64] M. T. Mas, M. Lamotte. Commande vocale de processus industriels. Proc. 8th Int. Cong. Cybern., Namur, 725-731, 1976.

[65] J. Massonie. L'utilisation des sous-ensembles flous en géographie. Cahiers de Géographie de l'Université de Besançon, 1975.

[66] K. Menger. A logic of the doubtful on optative and imperative logic. Reports of a mathematical colloquium (Notre Dame, Ind.), ser. 2, no. 1, 53-64, 1939. Reprinted in "Karl Menger. Selected Papers in Logic and Foundations, Didactics, Economics", D. Reidel Publ. Comp., 91-102, 1979.

[67] K. Menger. Statistical metrics. Proc. Nat. Acad. Sci. USA, 28: 535-537, 1942.

[68] K. Menger. Ensembles flous et fonctions aléatoires. C. R. Acad. Sci., Paris, 232, 2001-2003, 1951.

[69] K. Menger. Mathematical implications of Mach's ideas: Positivistic geometry, the clarification of functional connections. In: Ernst Mach: Physicist and Philosopher, (Cohen, R. S., Seeger, R. J., eds.), Vol. 6 of the series Boston Studies in the Philosophy of Science, Springer, 107-125,1970. Read at the Symp. of the American Assoc. for the Advancement of Science, organized for the 50th anniversary of Ernst Machs' death, on December 27, 1966, in Washington, D.C. First half entitled "Positivistic geometry" reprinted as "Geometry and positivism. A probabilistic microgeometry" in "Karl Menger. Selected Papers in Logic and Foundations, Didactics, Economics", D. Reidel Publ. Comp., 225-234, 1979.

[70] K. Menger. Selected Papers in Logic and Foundations, Didactics, Economics. D. Reidel Publ. Comp. 1979.

[71] A. A. Moles. La Création Scientifique. René Kister, Geneva, 1957.

[72] A. A. Moles. Sociodynamique de la Culture. Mouton, Paris & La Haye, 1967.

[73] Créativité et Méthodes d'innovation. Paris, Fayard, 1970.

[74] A. A. Moles (with the collaboration of E. Rohmer-Moles). Les Sciences de l'Imprécis. Seuil, Paris, coll. Science ouverte, 1990.

[75] C. V. Negoita, D. A. Ralescu. Multimi vagi si aplicatiile lor. (Ensembles flous et leurs applications). Editura Tehnica, Bucuresti, 1974.

[76] C. V. Negoita, D. Ralescu. Application of Fuzzy Sets to Systems Analysis. Birkaüser Verlag, Basel, 1975.

[77] C. Ollier. Fuzzy Sets. (in French) Union générale d'édition, collection 10-18, 1975.

[78] C. Ollier. Fuzzy Sets. (in French) Revised version, P.O.L, Paris, 1997.

[79] C.-F. Picard. Théorie des questionnaires. Foreword by J. A. Ville. Paris, Gauthier-Villars, 1965.

[80] D. Ponasse. Sur la notion de distance dans une structure floue régulière. Ann. de la Fac. des Sciences de Yaoundé, n°19, 3-9, 1975.

[81] C. Ponsard. Contribution à une théorie des espaces économiques imprécis. Publications Econométriques, vol. VIII, Fasc. 2, 1-43, 1975.

[82] C. Ponsard. L'imprécision et son traitement en analyse économique. Rev. Econ. Polit., n°1, 17-37, 1975.

[83] H. Prade. Exemple d'approche heuristique, interactive, floue pour un problème d'ordonnancement. Actes Congrès AFCET 'Modélisation et Maîtrise des Systèèmes', Editions Hommes & Techniques, Vol. 2, 347-355, Versailles, Nov. 1977.

[84] L. Pun. Use of fuzzy formalism in problems with various degree of subjectivity. In: Fuzzy Automata and Decision Processes, (M. M. Gupta, G. N. Saridis, B. R. Gaines, eds.), North-Holland, Amsterdam, 357-378, 1977.

[85] S. Ribeyre. Etude des automorphismes de l'algèbre de De Morgan des parties floues d'un ensemble. Ann. de la Fac. des Sciences de Yaoundé, n°19, 11-28, 1975.

[86] E. H. Ruspini. Numerical methods for fuzzy clustering. Information Sciences, 2, 319-350, 1970.

[87] E. Sanchez. Equations de relations floues. Thèse de Biologie Humaine, Faculté de Médecine de Marseille, 1974.

[88] E. Sanchez. Resolution of composite fuzzy relation equations. Information and Control, 30 (1), 38-48, 1976.

[89] R. Sambuc. Fonctions Φ-floues. Application à l'aide au diagnostic en pathologie thyroidienne. Thèse Université de Marseille, 1975.

[90] B. Schweizer, A. Sklar, K. Sigmund, P. Gruber, E. Hlawka, L. Reich, L. Schmetterer (eds.) Karl Menger. Selecta Mathematica. Volumes 1 & 2, Springer Veralg, 2002 & 2013.

[91] A. Seif, J. Aguilar-Martin. Multi-group classification using fuzzy correlation. LAAS, Toulouse France, 1977.

[92] R. Seising. Elie Sanchez, 1944-2014. Artificial Intelligence in Medicine, 62 (2), 73–77, 2014.

[93] R. Seising. The genesis of fuzzy sets and systems - Aspects in science and philosophy. In: Fifty Years of Fuzzy Logic and its Applications, (Tamir, D. E., Rishe, N. D., Kandel, A., eds.), Studies in Fuzziness and Soft Computing Series, vol. 326, Springer, 537-580, 2015.

[94] M. Sugeno. Theory of Fuzzy Integrals and its Applications. Dr. of Engineering, Tokyo Institute of Technology, 1974.

[95] D. Willaeys, N. Malvache. 1976. Utilisation d'un référentiel de sous-ensembles flous. Application à un algorithme flou. Int. Conf. Syst. Sci., Wroclaw, Poland, 1976.

[96] T. Yamakawa. Elie Sanchez (1944-2014) An obituary. Fuzzy Sets and Systems, 258: 134-138, 2015.

[97] L. A. Zadeh. From circuit theory to system theory. Proc. I.R.E., 50, 856-865, 1962.

[98] L. A. Zadeh. Fuzzy sets. Information and Control, 8, 338-353, 1965.

[99] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. Memo UCB / ERL M77/12, Berkeley, 1977.

The legacy of 50 years of fuzzy sets: A discussion *

Didier Dubois and Henri Prade
IRIT, CNRS and Université Paul Sabatier, 31062 Toulouse, France

December 15, 2015

Abstract: This note provides a brief overview of the main ideas and notions underlying fifty years of research in fuzzy set and possibility theory, two important settings introduced by L. A. Zadeh for representing sets with unsharp boundaries and uncertainty induced by granules of information expressed with words. The discussion is organized on the basis of three potential understanding of the grades of membership to a fuzzy set, depending on what the fuzzy set intends to represent: a group of elements with borderline members, a plausibility distribution, or a preference profile. It also questions the motivations for some existing generalized fuzzy sets. This note clearly reflects the shared personal views of its authors.
keywords: fuzzy set, possibility theory, similarity, uncertainty, preference.

1 Introduction

The founding paper on fuzzy sets [72], written by Lotfi Zadeh, is 50 years old. This seminal paper, sometimes ill-regarded at the beginning, has given rise to a huge literature, several dedicated journals, and many conferences each year for several decades now. It has affected many areas of scientific research (sometimes marginally, sometimes significantly) ranging from mathematics (especially many-valued logics, topology, algebra and category theory) to engineering practice especially in modeling, control, optimization, and data processing, but also with some clear impact on techniques devoted to pattern recognition and image processing, operations research, artificial intelligence, databases and information

*This paper is published in Fuzzy Sets and Systems, vol. 281: 21-31 (2015).

systems. Besides, fuzzy sets have influenced uncertainty analysis through the introduction of possibility theory [80] based on the use of membership functions for representing incomplete information, a bit more than a decade after the publication of the founding paper. The latter issue has contributed to a clarification of the confusion, pervading early years, between fuzzy sets and probability.

This discussion paper tries to organize the legacy of fuzzy sets in an orderly way, highlighting the main ideas, sometimes misunderstood, and pointing out what seem to be promising trends and barren areas as well as indicating some neglected views of interest. The paper will briefly situate various subfields of fuzzy sets in the light of the various interpretations of membership functions in terms of distance, preference or uncertainty [20], and suggest potentially fruitful fuzzy set-inspired topics for future research and applications.

2 Basic ideas behind fuzzy sets

The introduction of the notion of a fuzzy set by L. A. Zadeh was motivated by the fact that, quoting the founding paper [72]:

> "imprecisely defined "classes" play an important role in human thinking, particularly in the domains of pattern recognition, communication of information, and abstraction".

This seems to have been a continuous concern in all Zadeh's papers since the beginning, as well as the need to develop a sound mathematical framework for handling these "classes". This purpose required an effort to go beyond classical binary-valued logic, the usual setting for classes. Although many-valued logics had been there for a while, what is really remarkable is that due to this concern, Zadeh started to think in terms of *sets* rather than only in terms of degrees of truth. Since a set is a very basic notion, it was opening the road to the introduction of the fuzzification of any set-based notions such as relations, events, or intervals, while sticking with the many-valued logic point of view only does not lead you to consider such generalized notions. In other words, while Boolean algebras are underlying both propositional logic and naive set theory, the set point of view may be found richer in terms of mathematical modeling, and the same thing takes place when moving from many-valued logics to fuzzy sets. Moreover, the study of set operations on fuzzy sets has in return strongly contributed to a renewal of many-valued logics (see [14] for an introductory overview).

According to the founding paper [72], a fuzzy set represents

> "a class of objects with a continuum of grades of membership",

but in a footnote, Zadeh acknowledges that "the range of the membership function can be taken to be a suitable partially ordered set". This is an important remark, which opens the road to to more abstract constructs and to type-n fuzzy sets as well. Zadeh also observes that the case where the unit interval is used as a membership scale "corresponds to a multivalued logic with a continuum of truth values in the interval $[0, 1]$", acknowledging the link with many-valued logics.

So, a fuzzy set can be understood as a class equipped with an ordering of elements expressing that some objects are more in the class than others. However, in order to generalise the Boolean connectives, we need more than a mere relation between elements if one is to extend intersection, union and complement of sets, let alone implication to fuzzy sets. The set of possible membership grades has to be a complete lattice [33] so as to capture union and intersection, and either the concept of residuation or an order-reversing function are needed in order to express some kind of negation and implication.

Moreover, from the beginning, it was made clear that fuzzy sets were not meant as probabilities in disguise, since one can read [72] that

> "the notion of a fuzzy set is completely non-statistical in nature"

and that it provides

> "a natural way of dealing with problems where the source of imprecision is the absence of sharply defined criteria of membership rather than the presence of random variables."

In a nutshell, the main idea behind fuzzy sets is to make membership to sets gradual rather than abrupt. For instance, in the case of totally ordered universes, changing sharp membership thresholds into soft ones. It leads to extending the usual notions from set theory, logic, and inference, replacing Boolean algebra by many-valued ones, as well as all forms of set-valued mathematics to fuzzy set-valued mathematics. Presented as such, note that this extension is prima facie not related to the idea of uncertainty.

Fuzziness should also not be confused with vagueness [79], which is exclusively a concept pertaining to natural language. Indeed, the representation of gradual properties is not the unique information processing scenario that gives rise to borderline cases, one of the features of vagueness [13]. Vagueness refers to uncertainty of meaning (the membership function is ill-known), which is distinct from gradualness (membership is a matter of degree) [12]. The idea of typicality underlying linguistic terms is more connected to the one of similarity than to uncertainty [58].

As a consequence,

- originally, fuzzy sets were designed to formalize the idea of soft classification, which is more in agreement with the way people use categories in natural language.
- fuzziness is just implementing the idea of gradation in all forms of reasoning and problem-solving, as for Zadeh, everything is a matter of degree.
- a degree of membership is an abstract notion to be interpreted in practice.

According to the area of application, several interpretations can be found such as degree of similarity (to a prototype in a class), degree of plausibility, or degree of preference [20]. We now survey the use of fuzzy sets with respect to these three semantics.

3 Membership grades related to distance

The idea of representing a class by a fuzzy set [2], and later [79, 83] the fuzzy set representation of linguistic terms naming classes, is underlain by the idea of gradual transition between a set of elements that fully belong to the class, or that are fully representative of the term, and a set of elements that do not belong at all to the class, or that are definitely excluded by the meaning of the term. The first set, whose elements have membership 1, may be understood as the typical elements of the fuzzy set. More generally, the membership degree of an element to a fuzzy set is a degree of typicality of this element with respect to the class or the term represented, which is all the greater as the element is closer to the set of typical elements. Thus, in this view, membership grades can be naturally related to the idea of distance.

The most popular part of the fuzzy set literature deals with clustering, modeling and control, where gradual transitions between classes and their use in interpolation are the basic contribution of fuzzy sets. Intuitively speaking, a cluster gathers elements that are rather close to each other (or close to some core element(s)), while they are well-separated from the elements in the other cluster(s). Thus, the notions of graded proximity, similarity (dissimilarity) are at work in fuzzy clustering. With gradual clusters, the key issue is to define fuzzy partitions. The most widely used definition of a fuzzy partition, originally due to Ruspini [59], where the sum of membership grades of one element to the various classes is 1, suggests a connection (present in Ruspini's paper) between membership grades and probabilities, according to which a degree of membership of an element in a class can be identified with the probability that the element will be assigned to the class. Even though this view seems to be at odds with Zadeh's non-statistical intuitions, it is not surprising at all, as the closer an object to the prototypes of a class, the more often it will be assigned to this class. Measuring probability by distance is already present in the early times of statistics, when Gauss discovered the normal distribution as the only error function compatible with the least squares method (minimizing the Euclidean distance to observations), as explained by Stigler [63]. The statistical point of view on clustering is just a reversal of perspective with respect to the one of fuzzy sets, whereby the more often an object is assigned to a class, the closer is this object to the prototypes of the class. The question is then whether we measure strength of membership by observing frequencies in a training set or by computing distances to class prototypes. Using Gaussian-shaped distributions the two points of view are formally equivalent. However, fuzzy sets theory offers a more flexible mathematical framework for error-minimizing estimation methods, that cover distances other than Euclidean, as in the case of data reconciliation methods [15].

The view of a fuzzy set as a fuzzy cluster of elements, clusters forming a partition, has led Zadeh to emphasize the idea of granulation as a core concept supporting fuzzy logic [88], while in the crisp case, granulation and the notion of partition are basic in the theory of rough sets [54]. Interestingly enough, some bridges can be established between, (fuzzy) clusters, extensional fuzzy sets [40, 42], granulation, graded indistinguishability, and formal concept analysis [30] [25]. One application of fuzzy granulation is the notion of fuzzy transform [55] of a real-valued function with respect to a Ruspini's fuzzy partition [59] where the coefficients representing the transform are obtained in terms of the integral of the product of the function with each of the basic functions defining the partition (which evaluate the degree of adequacy of the value of the variable with the corresponding element of the fuzzy partition). From these numbers, an approximation of the function can be recovered. The same concept of fuzzy granulation is at work in density estimation methods based on fuzzy histograms [65], where the analogy with kernel-based methods is striking: again a kernel expresses similarity and plays the same interpolation role as a membership function but it is couched in the phraseology of, and formalised inside, probability theory.

The role of fuzzy sets in modeling and control originated in the idea of fuzzy algorithms and programs [73], where fuzzy instructions are instructions involving fuzzy labels. In that respect, the role of fuzzy if-then rules was soon recognized [76], while their interpolation power was emphasized later [86]. Basically, it offers a reconciliation between logical notions such as Boolean categories and inference, and numerical modeling techniques in engineering, that extensively exploit the notion of (linear) interpolation. A fuzzy model (e. g., fuzzy rules in the sense of Takagi-Sugeno [67]) is typically a collection of local usual mathematical models, each defined on gradual overlapping domains forming a partition of the input space. These mathematical models are related via an interpolation scheme taking advantage of soft boundaries and membership grades to neighboring domains. Such interpolation schemes are similar to the ones of neural nets. The bridge between neural nets and fuzzy sets leads to a useful trade-off between model accuracy (thanks to universal approximation capabilities) and model interpretability, provided that the fuzzy sets appearing in the rules remain meaningful for the expert, as in the first fuzzy controllers (e.g., [48]). It is also worth mentioning that the inference mechanism underlying Takagi-Sugeno fuzzy rules is close in spirit to case-based decision theory, later axiomatized by Gilboa and Schmeidler [32]: in the latter, the decision to apply should maximize a counterpart of expected utility where probabilities are replaced by similarities to previous cases where the decision led to results whose utility is known, while in the former, since the potential decisions belong to a continuum, the similarity based weighting is directly applied to the (linear) models of actions given in the conclusion part of the fuzzy rules.

4 Membership grades related to uncertainty

Fuzziness is also often interpreted as a form of uncertainty. However, this view is sometimes based on a misunderstanding. In its original understanding a grade of membership is not considered as a degree of (un)certainty [22]: asserting that a man is almost bald (implying he has almost no hairs) differs from saying that this man is almost certainly bald (leaving the possibility that finally he is not bald at all).

Fuzzy sets can represent uncertainty (in a gradual way) because crisp sets are often used to represent an ill-known value (in a crisp way like in interval analysis or in propositional logic). Viewed as representing uncertainty a set just distinguishes between values that are considered possible and values that are not, and fuzzy sets just introduce grades to soften boundaries of an uncertainty set. So in a fuzzy set it is the set that captures uncertainty [23]. This point of view echoes an important distinction made by Zadeh himself [79] between

- conjunctive (fuzzy) sets, where the set is viewed as the conjunction of its elements. This is the case for clusters discussed in the previous section. But also with set-valued attributes like the languages spoken more or less fluently by an individual.

- and disjunctive fuzzy sets which corresponds to mutually exclusive possible values of an ill-known single-valued attribute, like the ill-known birth nationality of an individual.

In the latter case, fuzzy sets are called possibility distributions [80] and act as elastic constraints on a precise value [77, 78]. Possibility distributions have an epistemic flavor, since they represent the information we have at our disposal about what values remain more or less possible for the variable under consideration, and what values are (already) known as impossible. This epistemic view is in complete contrast with the ontic view underlying conjunctive fuzzy sets [26].

The first illustration of the power of possibility theory proposed by Zadeh was an original theory of approximate reasoning [81, 84, 85], later reworked for emphasizing new points [87, 90], where pieces of knowledge are represented by possibility distributions that fuzzily restrict the possible values of variables, or tuples of variables. These possibility distributions are combined conjunctively, and then projected in order to compute fuzzy restrictions acting on the variables of interest. This view is the one at work in constraint satisfaction problems (CSP), and anticipates weighted CSP [4] by many years, although without any algorithmic concerns. One research direction, quite in the spirit of the objective of "computing with words" [87], would be to further explore the possibility of a syntactic (or symbolic) computation of the inference step (at least for some noticeable fragments of this general approximate reasoning theory), where the obtained results are parameterized by fuzzy set membership functions that would be used only for the final interpretation of the results. An illustration of this idea can be found in an approach to reasoning with relative orders of magnitude [39]. It is also at work in possibilistic logic [27], a very elementary formalism handling pairs made of a Boolean formula and a certainty weight. Such pairs encode simple possibility distributions.

Associated with a possibility distribution, is a possibility measure [80], which is a max-decomposable set function. Thus, one can evaluate the possibility of a crisp, or fuzzy, statement of interest, given the available information supposed to be represented by a possibility distribution. It is also important to notice that the introduction of possibility theory by Zadeh was closely related to the modeling of information expressed in natural language. This view contrasts with the motivations of the English economist Schackle [61, 62], interested in a non-probabilistic view of expectation, who designed a formally similar theory, but rather based on the idea of degree of impossibility understood as a degree of surprise.

However, apart from a brief mention in [82], Zadeh does not explicitly use of the notion of necessity (the natural dual of the modal notion of possibility) in his work on possibility theory and approximate reasoning. Still, it is important to distinguish between statements that are *necessarily* true (to some extent), i.e. whose negation is almost impossible, from the statements that are only *possibly* true (to some extent) depending on the way the fuzzy knowledge would be made precise. The simultaneous use of the two notions is often required in applications of possibility theory. Besides, in possibility theory, apart from a (strong) necessity measure reflecting, by complementation, what is known as being more or less impossible, and the dual (weak) possibility measure reflecting consistency with the available information, there exist two other set functions of interest (see, e.g., [21]): a strong possibility measure which guarantees that *all* interpretations in a subset are possible to some extent, and a dual weak necessity measure. The joint use of these four set functions provides the proper setting for representing bipolar information, when one distinguishes between positive information, e.g., what has been observed, and negative information, corresponding, e.g., to what is not ruled out by generic knowledge [24].

Interestingly, necessity (resp. possibility) measures are formally special case of belief (resp. plausibility) functions (already pointed out in Shafer's book [64]) and special cases of coherent lower probabilities in the sense of Walley [70]. This direction is at odds with Zadeh's motivations for possibility theory but it opens the way to the systematic use of membership functions (as possibility distributions) to represent incomplete information in statistics, from likelihood functions to confidence intervals and probabilistic inequalities [10] as well as dispersion measures and parameter estimation methods [50].

Necessity degrees are also the building blocks of possibilistic logic (see [27] for an up-to-date overview), where classical logic formulas are associated with lower bounds of a necessity measure for assessing their certainty. The semantics of a (conjunctive) set of possibilistic logic formulas is expressed by a possibility distribution over a set of interpretations (or possible worlds). This is an example of possibility distribution defined on an abstract, non-ordered referential, where the possibility distribution is no longer the precisiation of a linguistic term on an ordered often continuous universe of discourse as it is the case in Zadeh's approximate reasoning approach. Possibilistic logic has found a number of developments in artificial intelligence including the handling of inconsistency, the modeling of exception-tolerant reasoning, the fusion of logical knowledge bases, and the design of possibilistic counterparts to Bayesian networks, which are semantically equivalent to possibilistic logic bases, but exhibit a graphical representation. See [27] for an introduction to these issues and for references. The definition of possibilistic networks requires the introduction of the notion of conditioning in possibility theory. It turns out that two forms of conditioning make sense, one defined with minimum, the other with product. These two forms of conditioning differentiate qualitative and quantitative possibility theory [21]. Let us also mention a recent application of a possibilistic logic-like handling of uncertain functional dependencies to the design of databases containing dubious tuples [44].

The links and differences between modal logic and possibility theory have been a matter of debates, not always well-focused, for a long time; see [91] for a recent position paper by Zadeh. Still, the recent development of generalized possibilistic logic (GPL) (see [27] for a brief account and references), where one can reason both in terms of possibility and necessity, and whose axiomatics is as the one of a (graded) epistemic modal logic, sheds some light on the question: the semantics of GPL is in terms of *sets* of possibility distributions (rather than a unique possibility distribution as in basic possibilistic logic), while the semantics of general modal logics require accessibility relations.

Apart from the reasoning side, another important area of the application of the possibility theory-based understanding of fuzzy sets is the computation with ill-known quantities represented by fuzzy intervals. The possibility of performing arithmetic operations on fuzzy numbers was also pointed out by Zadeh [78], then developed by other scholars (see [19] for a survey in the XXth century). The calculus of fuzzy intervals is a gradual extension of *set-valued* mathematics and the extension principle underlying it can be expressed in possibility theory. The calculus of fuzzy intervals is instrumental in various areas including:

- systems of linear equations with fuzzy coefficients (see the paper by Lodwick and Dubois in this special issue) and differential equations with fuzzy initial values, and fuzzy set functions [45];

- fuzzy random variables for the handling of linguistic or imprecise statistical data [31, 8];

- fuzzy regression methods [52];

- operations research and optimisation under uncertainty [93, 16, 46].

This research trend contrasts with the mainstream fuzzy modeling approach, discussed in the previous section, which promotes mathematical models in the usual sense, albeit constructed by means of fuzzy sets, and does not express uncertainty. Systems analysis based on fuzzy intervals and fuzzy differential equations has been less developed than fuzzy modeling partly because it leads to complex calculations, but also because the epistemic nature of the fuzzy approach under the uncertainty interpretation is sometimes ill-understood at the practical level, running the risk of posing mathematical problems that are not always reflecting their intended meaning. For instance, the equality of membership functions on each side of a fuzzy linear equation with fuzzy numbers is very demanding and is not equivalent to the identity between actual values of the terms on each side of the equality. So there is sometimes a gap between mathematical results and the actual problem they are supposed to model in this area.

5 Membership grades related to preference

Another natural semantics for the grades of membership of a fuzzy set is in terms of degrees of satisfaction, when the fuzzy set represents a value function. For instance, the probability of a fuzzy event [74] has exactly the same form as the expected utility of an act after Savage, interpreting the utility function as the membership function of the fuzzy set of good consequences of this act. Likewise, in multicriteria decision evaluation, the rating profile of an object according to various criteria is easily viewed as the fuzzy set of satisfied criteria.

In this respect the invention of fuzzy set connectives by Zadeh has triggered a large literature on

- aggregation operations both from a mathematical point of view (see [43]) and for multicriteria evaluation (see [36] [1] [69] for recent books).

- maxmin approaches to optimization, in multicriteria linear programming, initiated in [68, 93] and more recently seen as a special case of valued constraint satisfaction [4].

This framework for multicriteria evaluation, constraint-based reasoning, and optimization is clearly part of the legacy of Zadeh's 1965 paper, as well as the one he published with Richard Bellman in 1970 [3]. Originally rather elementary (using max and min), it is characterized by:

- the assumption of a common value scale for the various factors, constraints or criteria. This is a strong assumption that nevertheless includes both quantitative and qualitative scales.

- A unified view of possible aggregation modes ranging from conjunctions and disjunctions to generalized means.

- Sophisticated criteria weighting schemes that allow for dependent criteria. On this issue, the fuzzy set literature has met the economic literature on Choquet integrals [35]. In the qualitative framework, the counterpart to Choquet integral is Sugeno integral [66, 35], also called fuzzy integral by Sugeno, because it uses maximum and minimum, respectively the basic disjunction and conjunction in fuzzy set theory. From its inception, it was construed as a tool for multiple criteria evaluation. Sugeno integral acted as a bridge between fuzzy set theory and Choquet integrals.

- Extensions to bipolar decision analysis methods that measure pros and cons of decisions, losses and gains in a separate way [34].

Another important offspring of Zadeh's early papers is fuzzy preference modeling based on the gradual extension of equivalence relations and orderings proposed in [75], generalizing transitivity to maxmin transitivity. Preference modeling is the first step in multicriteria evaluation, whereby it is more natural for a person to represent his or her preference on a set of objects by an ordering relation than by a utility function. The basic tool for analyzing human-originated preference relations is their decomposition into strict preference, indifference and incomparability [6]. Using fuzzy relations, it is possible to express how much an object is preferred to another. A number of publications starting by an early paper of Orlowski [53] and later on triggered by the book by Fodor and Roubens [29] address the issue of decomposing a fuzzy relation into graded strict preference, indifference and incomparability. Bodenhofer [5] has shown fuzzy ordering relations should be envisaged conjointly with a fuzzy similarity relation expressing indifference.

Besides, here again the meaning of the values in a fuzzy relation matters, as it may either reflect intensity of preference, or uncertainty about all-or-nothing preference. Unless this is clarified, it is very difficult to apply fuzzy preference modeling in concrete decision-making problems. Indeed, one may argue that the measurement issue present in utility theory is for the most part obviated by the use of order relations, while this advantage is lost when making such relations fuzzy.

Note that the use of a fuzzy ordering relation implies that it makes sense to say: object 1 is preferred to object 2 to the same extent as object 3 is preferred to object 4, which suggests to consider preference difference measurement methods. A lot of work needs to be done to let such mathematical models of preference be used as a basis for multicriteria evaluation, in place of questionable methods based on triangular fuzzy numbers defined on arbitrary value scales [11].

Finally, it is not clear all valued relations can be put under the umbrella of fuzzy logic. In so-called reciprocal relations, the sum of the preference values of one object against the other and of the latter against the former is 1, so that it is more natural to interpret them in terms of probability of preference. Such reciprocal relations exist for a long time and differ from the valued relations introduced in Zadeh's paper. As a consequence, calling reciprocal relations fuzzy relations may be misleading [9].

In many practical applications, preference and uncertainty are conjointly present. For instance, fuzzy databases [56, 57] may be both a matter of

1. expressing preferences in the queries by specifying flexible restrictions on the desired attribute values or by handling priorities between attributes,

2. handling uncertainty when the database contains imprecise or fuzzy pieces of information.

When both issues are present, we are led to compute possibility and necessity measures for fuzzy events, in order to distinguish between answers that are certain (to some extent) to reach highly satisfactory values, and answers for which this is only possible (to some extent). At work here are pessimistic and optimistic decision criteria, which have been axiomatized [28], thus providing formal foundations for qualitative possibility theory.

Databases may be viewed as a repertory of cases. This leads to substitute similarity to uncertainty in possibilistic decision criteria, as in fuzzy case-based reasoning methods [18, 41], coming close to the idea of similarity-based possibility [92].

6 Higher-order membership grades

Ten years after the invention of fuzzy sets, Zadeh [78] considered the possibility of iterating the process of changing membership grades into (fuzzy) sets, giving birth to interval-valued fuzzy sets, fuzzy set-valued (type 2) fuzzy sets, and more generally type n fuzzy sets. While this idea is philosophically tempting and reflects the problematic issue of measuring membership grades of linguistic concepts, it has given birth to a high number of publications pertaining to variants of higher-order fuzzy sets, often reinventing the same notions under different names. To name a few:[1] intuitionistic fuzzy sets, vague sets, hesitant fuzzy sets, soft sets, arbitrary combinations of the above notions (for instance, interval-valued intuitionistic fuzzy sets, fuzzy soft sets, etc.).

There are several concerns to be pointed out with these complexifications (rather than generalizations) of fuzzy sets. They shed some doubt on the theoretical or applied merits of such developments of fuzzy set theory:

- Each new kind of fuzzy sets gives birth to a plethora of routine theoretical papers, redefining basic concepts of fuzzy sets in the new setting, irrespective of what the new membership grades mean, and presenting no motivations;

- Several of these constructions reinvent existing notions under different sometimes questionable names. For instance vague sets are the same as intuitionistic fuzzy sets, and formally they are just a different encoding of interval-valued fuzzy sets (a pair of nested sets, versus an orthopair of disjoint sets [7]). Moreover the link between intuitionistic fuzzy sets and intuitionism hardly exists[17]. Hesitant fuzzy sets were proposed already in 1975 under a different name [37]; soft sets are set-valued mappings, a notion that has been well-known for a long time in the theory of random sets.

- Some generalizations of fuzzy sets often underly a misunderstanding [26]: are they special kinds of L-fuzzy sets or an approach to handling uncertainty about membership grades? The latter motivation is often put forward in the introduction of such papers, while the main text adopts an algebraic structure derived from L-fuzzy sets. For instance, many authors speak of "type-2-connectives". However, if a type 2 fuzzy set is understood as a fuzzy set of (a possibility distribution over) membership functions, it is enough to apply the extension principle to the type 1 fuzzy logic expression of interest, in order to compute the resulting fuzzy set-valued membership grade; there is no need to appeal to a specific algebraic structure on some set of higher-order membership grades different from the unit interval, all the more so as

[1] We omit references here, for the sake of conciseness; readers can find a lot of them by searching for the corresponding key-words.

compositionality of connectives is then lost [22]. There is no such thing as type-2 connectives in this case.

- Algebraic structures for complex membership grades stemming from higher-order fuzzy sets are in general special kind of lattices. Actually, these higher-order fuzzy sets are special cases of L-fuzzy sets, hence often redundant from a mathematical point of view [71, 38].

- The incurred complexity of higher-order fuzzy set-based approaches is sometimes not justified. Practical applications of type 2 fuzzy sets in the modeling area sometimes come down to models with more tuning parameters than usual fuzzy systems, but they are often standard input-output models after due defuzzification steps. So it is difficult to understand why they would outperform usual fuzzy systems without being subject to overfitting effects. In practice, many authors restrict to interval-valued fuzzy sets, under the strange name "interval type 2 fuzzy sets", to simplify calculations. Likewise applications of intuitionistic to multicriteria decision making seem to artificially increase the burden of collecting preference ratings in the form of pairs of numbers (or even of intervals) whose meaning may be more unclear to the user than mere membership values.

- Many methods using type 2 fuzzy sets revisit calculations with fuzzy intervals, without referring to the corresponding state of the art.

The point is not to claim that such variants of fuzzy sets are necessarily misleading or useless. They often try to capture convincing intuitions but are too often developed for their own sake, sometimes at odds with these intuitions. See [17] for a full-fledged discussion on intuitionistic fuzzy sets and interval-valued fuzzy sets, and [26] for the clash of intuitions between notions of bipolarity and uncertainty pervading intuitionistic fuzzy sets. Regarding soft sets, only outlined in the founding paper [51], they were originally meant as an extension of the alpha-cut mapping to non-nested sets, a concept more recently considered by several authors [23, 60, 49] in a more applied perspective. However, followers of the soft set trend often adopt the set-valued mapping point of view without reference to cuts of fuzzy sets. For instance, the highly cited paper of Maji *et al.* [47] seem to consider soft sets in the algebraic framework of formal concept analysis [30] only.

As a consequence, there is an effort to be pursued in terms of motivation, mathematical rigor, and convincing applications, in order to make this part of the fuzzy set legacy worth developing further.

7 Conclusion

The intention of this note was to overview research topics that stemmed from Zadeh's founding paper and early subsequent publications of his, and that seem to have a promising future. However, our discussion has no pretense to provide an exhaustive coverage of all the potential application fields of fuzzy set and possibility theory. Several noticeable ones have not been cited (e.g., information retrieval, machine learning), and those that have been mentioned are mainly there for illustrating various usages of fuzzy set notions, rather than for advocating their merits with respect to other approaches.

Note that fuzzy set research now reaches a point where the corpus of basic tools has been already considerably developed, and very few new basic concepts seem to have emerged in the last 10 years. These basic tools become more and more accepted in various established disciplines (for example, fuzzy systems in non-linear control engineering, fuzzy clustering in data analysis, fuzzy interval computations in risk analysis, etc.). In this sense, fuzzy set theory has come of age. These numerous achievements contrast with various attempts to fuzzify mathematical notions or complexify existing fuzzy set concepts, which can be called "fuzzification for its own sake", that seems to be driven, as in many fields nowadays, by the pressure to publish papers in the academic world. This increases the number of publications without always contributing much to science, while at a higher level in the society "the pursuit of knowledge for his own sake is increasingly being replaced by a quest for education as a ticket to a better-paying job", as denounced by Zadeh himself [89] and deplored by the scientific community.

References

[1] G. Beliakov, A. Pradera, T. Calvo. Aggregation Functions: A Guide for Practitioners. Studies in Fuzziness and Soft Computing 221, Springer, 2007.

[2] R. E. Bellman, R. Kalaba, L. A. Zadeh. Abstraction and pattern classification. J. of Mathematical Analysis and Applications, 13,1-7, 1966.

[3] R. E. Bellman and L. A. Zadeh. Decision making in a fuzzy environment. Management Science, 17:B141-B164, 1970.

[4] S. Bistarelli, H. Fargier, U. Montanari, F. Rossi, Th. Schiex, G. Verfaillie. Semiring-based CSPs and valued CSPs: Basic properties and compari-

son. In: Over-Constrained Systems, (M. Jampel, E. C. Freuder, M.J. Maher, eds.), Springer, LNCS 1106, 111-150, 1996.

[5] U. Bodenhofer, Representations and constructions of similarity-based fuzzy orderings, Fuzzy Sets Syst. 137 (2003) 113-136.

[6] D. Bouyssou, P. Vincke Binary Relations and Preference Modeling. In Decision-making Process- Concepts and Methods. D. Bouyssou *et al.* (Eds.), ISTE London & Wiley, Chap. 2, 49-74, 2009.

[7] G. Cattaneo, D. Ciucci Basic intuitionistic principles in fuzzy set theories and its extensions (A terminological debate on Atanassov IFS) Fuzzy Sets and Systems, 157 (24), 3198-3219, 2006.

[8] I. Couso, D. Dubois, L. Sanchez. Random Sets and Random Fuzzy Sets as Ill-Perceived Random Variables, Springer, SpringerBriefs in Computational Intelligence, 2014.

[9] B. De Baets, H. De Meyer, Transitivity frameworks for reciprocal relations: cycle transitivity versus FG-transitivity, Fuzzy Sets Syst., 152, 2005. 249-270.

[10] D. Dubois Possibility theory and statistical reasoning. Computational Statistics & Data Analysis, 51, 47-69, 2006

[11] D. Dubois, The role of fuzzy sets in decision sciences: Old techniques and new directions. Fuzzy Sets and Systems 184 (1), 3-28, 2011.

[12] D. Dubois. Have fuzzy sets anything to do with vagueness ? (with discussion) In: Understanding Vagueness - Logical, Philosophical and Linguistic Perspectives, (P. Cintula, Ch. Fermüller, eds.), vol. 36 of Studies in Logic, College Publications, 317-346, 2012.

[13] D. Dubois, F. Esteva, L. Godo, H. Prade. An information-based discussion of vagueness. In: Handbook of Categorization in Cognitive Science, (H. Cohen, C. Lefebvre, eds.) Chap. 40, 892-913, 2005.

[14] D. Dubois, F. Esteva, L. Godo, H. Prade. Fuzzy-set based logics - An history-oriented presentation of their main developments. In : Handbook of The history of Logic (D. M. Gabbay, J. Woods, eds.), Vol. 8, The Many Valued and Nonmonotonic Turn in Logic, 325-449, 2007.

[15] D. Dubois, H. Fargier, M. Ababou, D. Guyonnet. A fuzzy constraint-based approach to data reconciliation in material flow analysis. International Journal of General Systems, 43 (8), 787-809, 2014.

[16] D Dubois, H Fargier, P Fortemps, Fuzzy scheduling: Modelling flexible constraints vs. coping with incomplete knowledge, European Journal of Operational Research 147 (2), 231-252, 2003

[17] D. Dubois, S. Gottwald, P. Hájek, J. Kacprzyk, H. Prade. Terminological difficulties in fuzzy set theory - The case of "Intuitionistic Fuzzy Set". Fuzzy Sets and Systems, 156 (3), 485-491, 2005.

[18] D. Dubois, E. Hüllermeier, H. Prade. Fuzzy methods for case-based recommendation and decision support. J. Intell. Inf. Syst., 27 (2), 95-115, 2006.

[19] D. Dubois, E. Kerre, R. Mesiar, H. Prade. Fuzzy interval analysis. In: Fundamentals of Fuzzy Sets, (D. Dubois, H. Prade, Eds.), Kluwer, Boston, The Handbooks of Fuzzy Sets Series, 483-581, 2000.

[20] D. Dubois, H. Prade. The three semantics of fuzzy sets. Fuzzy Sets and Systems, 90, 141-150, 1997.

[21] D. Dubois and H. Prade. Possibility theory: Qualitative and quantitative aspects. In: Quantified Representation of Uncertainty and Imprecision, vol. 1 of Handbook of Defeasible Reasoning and Uncertainty Management Systems (D. M. Gabbay, Ph. Smets, eds.), Kluwer Acad. Publ., 169-226, 1998.

[22] D. Dubois, H. Prade. Possibility theory, probability theory and multiple valued logics: A clarification. Annals of Mathematics and Artificial Intelligence. 32, 35-66, 2001.

[23] D. Dubois, H. Prade. Gradual elements in a fuzzy set. Soft Computing, 12, 165-175, 2008.

[24] D. Dubois, H. Prade. An overview of the asymmetric bipolar representation of positive and negative information in possibility theory. Fuzzy Sets and Systems, 160 (10), 1355-1366, 2009.

[25] D. Dubois, H. Prade. Bridging gaps between several frameworks for the idea of granulation. Proc. Symp. on Foundations of Computational Intelligence (FOCI'11), (in Symposium Series on Computational Intelligence (SSCI'11))Paris, April 11-15, IEEE, 59-65, 2011.

[26] D. Dubois, H. Prade, Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets. Fuzzy Sets and Systems, 192, 3-24, 2012.

[27] D. Dubois, H. Prade. Possibilistic logic - An overview. In: Handbook of the History of Logic. Volume 9: Computational Logic. (J. Siekmann, vol. ed.; D. M. Gabbay, J. Woods, series eds.), 283-342, 2015.

[28] D. Dubois, H. Prade, R. Sabbadin. Decision-theoretic foundations of qualitative possibility theory. Europ. J. of Operational Research, 128, 459-478, 2001.

[29] J. Fodor, M. Roubens. Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer Acad. Pub., 1994.

[30] B. Ganter, R. Wille. Formal Concept Analysis. Springer-Verlag, 1999.

[31] M. A. Gil, G. Gonzlez-Rodrguez, R. Kruse, Eds. Statistics with Imperfect Data, special issue of Inf. Sci. 245: 1-3 (2013)

[32] I. Gilboa, D. Schmeidler. Case-based decision theory. The Quarterly Journal of Economics, 110, 605-639, 1995.

[33] J. A. Goguen. L-fuzzy sets. J. Math. Anal. Appl. 18(1967), 145-174.

[34] M. Grabisch, S. Greco, M. Pirlot, Bipolar and bivariate models in multicriteria decision analysis: descriptive and constructive approaches, Int. J. Intell. Syst. 23 (9) (2008) 930-969.

[35] M. Grabisch, C. Labreuche. A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. Ann. Oper. Res., 175, 247-286, 2010.

[36] M. Grabisch, J.-L. Marichal, R. Mesiar, E. Pap, Aggregation Functions, Cambridge University Press, 2009.

[37] I. Grattan-Guiness, Fuzzy membership mapped onto interval and many-valued quantities, Z. Math. Logik. Grundladen Math. 22 (1975) 149-160.

[38] J. Gutiérrez García, S.E. Rodabaugh Order-theoretic, topological, categorical redundancies of interval-valued sets, grey sets, vague sets, interval-valued intuitionistic sets, intuitionistic fuzzy sets and topologies Fuzzy Sets and Systems, 156(3), 2005, 445-484

[39] A. Hadj Ali, D. Dubois, H. Prade. Qualitative reasoning based on fuzzy relative orders of magnitude. IEEE Trans. on Fuzzy Systems,11 (1), 9-23, 2003.

[40] U. Höhle. Quotients with respect to similarity relations. Fuzzy Sets and Systems, 27, 31-44, 1988.

[41] E. Hüllermeier. Case-based Approximate Reasoning. Springer, Berlin, 2007.

[42] F. Klawonn. Fuzzy points, fuzzy relations and fuzzy functions. In: Discovering the World with Fuzzy Logic, (V. Novák and I. Perfilieva, eds.), Physica-Verlag, Heidelberg, 431-453, 2000.

[43] E. P. Klement, R. Mesiar, and E. Pap. Triangular Norms. Springer, 2000.

[44] S. Link, H. Prade. Relational database schema design for uncertain data. Centre for Discrete Mathematics and Theoretical Computer Science, University of Auckland, Research Report 469, Aug. 2014.

[45] W. A. Lodwick and M. Oberguggenberger, Eds, Differential Equations Over Fuzzy Spaces - Theory, Applications, and Algorithms, special issue of Fuzzy Sets and Systems 230: 1-162, 2013.

[46] M.K. Luhandjula Fuzzy optimization: Milestones and perspectives Fuzzy Sets and Systems, 274: 4-11, 2015.

[47] P. K. Maji, R. Biswas, A. R. Roy. Soft set theory. Comput. Math. Appl., 45 (4-5), 555-562, 2003.

[48] E. H. Mamdani. Application of fuzzy algorithms for control of simple dynamic plant. Proc. IEEE, 121 (12), 1585-1588, 1976.

[49] T. P. Martin and B. Azvine. The X-mu Approach: Fuzzy quantities, fuzzy arithmetic and fuzzy association rules. *IEEE Symp. on Foundations of Computational Intelligence (FOCI)*, 2013.

[50] G. Mauris. Possibility distributions: A unified representation of usual direct-probability-based parameter estimation methods. Int. J. Approx. Reasoning, 52 (9),1232-1242, 2011.

[51] D. Molodtsov. Soft set theory - First results. Comput. Math. Appl., 37 (4/5), 19-31, 1999.

[52] S. Muzzioli, A. Ruggieri, B. De Baets A comparison of fuzzy regression methods for the estimation of the implied volatility smile function, Fuzzy Sets and Systems, 266:131-143, 2015

[53] S. Orlovsky. Decision-making with a fuzzy preference relation, Fuzzy Sets Syst., 1, 155-168, 1978.

[54] Z. Pawlak. Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Acad. Publ., Dordrecht, 1991.

[55] I. Perfilieva. Fuzzy transforms: Theory and applications. Fuzzy Sets and Systems, 157, 993-1023, 2005.

[56] F. E. Petry. Fuzzy Databases: Principles and Applications (with a chapter by P. Bosc). International Series in Intelligent Technologies, 1995.

[57] O. Pivert, P. Bosc. Fuzzy Preference Queries to Relational Databases. Imperial College Press, 2012.

[58] H. Prade and S. Schockaert. Handling borderline cases using degrees: An information processing perspective. In: (with discussion) In: Understanding Vagueness - Logical, Philosophical and Linguistic Perspectives, (P. Cintula, Ch. Fermüller, eds.), vol. 36 of Studies in Logic, College Publications, 291-309, 2012.

[59] E. H. Ruspini. A new approach to clustering. Inform. and Control, 15, 22-32, 1969.

[60] D. Sanchez, M. Delgado, M.A. Villa, J. Chamorro-Martinez. On a non-nested level-based representation of fuzziness *Fuzzy Sets and Systems* 192: 159-175, 2012.

[61] G. L. S. Shackle. Expectation in Economics. Cambridge University Press, UK, 1949. 2nd edition, 1952.

[62] G. L. S. Shackle. Decision, Order and Time in Human Affairs (2nd edition), Cambridge University Press, UK, 1961.

[63] S. M. Stigler 1990. The History of Statistics: The Measurement of Uncertainty before 1900. Cambridge, MA: Belknap Press.

[64] G. Shafer. A Mathematical Theory of Evidence. Princeton University Press, 1976.

[65] O. Strauss: Quasi-continuous histograms. Fuzzy Sets and Systems 160(17): 2442-2465 (2009)

[66] M. Sugeno. Fuzzy measures and fuzzy integrals: a survey. In: Fuzzy Automata and Decision Processes, (M. M. Gupta, G. N. Saridis, and B. R. Gaines, eds.), 89-102, North-Holland, 1977.

[67] T. Takagi, M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. IEEE Trans. on Systems, Man, and Cybernetics, 15 (1), 116-132, 1985.

[68] H. Tanaka, T. Okuda and K. Asai, On fuzzy mathematical programming, J. Cybernetics, 3, 37-46, 1974.

[69] V. Torra, Y. Narukawa. Modeling Decisions - Information Fusion and Aggregation Operators. Springer, 2007.

[70] P. Walley. Measures of uncertainty in expert systems. Artif. Intell. 83 (1), 1-58; 1996.

[71] G.-J. Wang, Y.-H. He, Intuitionistic fuzzy sets and L-fuzzy sets, Fuzzy Sets and Systems 110, 271-274, 2000.

[72] L. A. Zadeh. Fuzzy sets. Information and Control, 8 (3), 338-353,1965.

[73] L. A. Zadeh. Fuzzy algorithms. Information and Control, 12, 94-102,1968.

[74] L. A. Zadeh, Probability measures of fuzzy events, J. Math. Anal. Appl. 23 (1968), 421-427

[75] L.A. Zadeh, Similarity relations and fuzzy orderings, Inf. Sci. 3 (1971) 177-200.

[76] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. on Systems, Man, and Cybernetics, 3 (1), 28-44, 1973.

[77] L. A. Zadeh, Calculus of fuzzy restrictions. In: Fuzzy sets and Their Applications to Cognitive and Decision Processes, (L. A. Zadeh, K. S. Fu, K. Tanaka, M. Shimura, eds.), Proc. U.S.-Japan Seminar on Fuzzy Sets and Their Applications, Berkeley, July 1-4, 1974, Academic Press, 1-39, 1975.

[78] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. Inf. Sci., Part I, 8 (3), 199-249; Part II, 8 (4), 301-357; Part III, 9 (1), 43-80,1975.

[79] L. A. Zadeh, PRUF - a meaning representation language for natural languages. Int. J. Man-Machine Studies, 10, 395-460, 1978.

[80] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1, 3-28, 1978.

[81] L. A. Zadeh. A theory of approximate reasoning. In: Machine Intelligence, Vol. 9, (J. E. Hayes, D. Mitchie, and L. I. Mikulich, eds.), 149-194, 1979.

[82] L.A. Zadeh, Fuzzy sets and information granularity. In: Advances in Fuzzy Set Theory and Applications, (M. M. Gupta, R. K. Ragade, R. R. Yager, eds.), North-Holland, 3-18, 1979.

[83] L. A. Zadeh. Precisiation of meaning via translation into PRUF. In: Cognitive Constraints on Communication, (L. Vaina, J. Hintikka, eds.), Reidel, Dordrecht, 373-402, 1984.

[84] L. A. Zadeh. Syllogistic reasoning in fuzzy logic and its application to usuality and reasoning with dispositions. IEEE Transactions on Systems, Man, and Cybernetics, 15 (6), 754-763,1985.

[85] L. A. Zadeh. Knowledge representation in fuzzy logic. IEEE Trans. Knowl. Data Eng., 1 (1), 89-100, 1989.

[86] L. A. Zadeh. The calculus of fuzzy if-then rules. AI Expert, 7 (3), 23-27,1992.

[87] L. A. Zadeh. Fuzzy logic = computing with words. IEEE Trans. Fuzzy Systems 4(2): 103-111,1996.

[88] L. A. Zadeh. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems, 90, 111-128, 1997.

[89] L. A. Zadeh. Editorial. UC-Berkeley Computer Science Commencement Address. IEEE Trans. on Systems, Man and Cybernetics, Part C: Applications and Reviews, 28 (1), 7-8, 1998.

[90] L. A. Zadeh, Generalized theory of uncertainty (GTU) – principal concepts and ideas. Computational Statistics & Data Analysis, 51, 15-46, 2006.

[91] L. A. Zadeh. A note on modal logic and possibility theory. Inf. Sci. 279: 908-913 (2014)

[92] L. A. Zadeh. A note on similarity-based definitions of possibility and probability. Inf. Sci., 267, 334-336, 2014.

[93] H.-J. Zimmermann, Fuzzy programming and linear programming with several objective functions. Fuzzy Sets Syst., 1, 45-55, 1978.

Practical methods for constructing possibility distributions *

Didier Dubois and Henri Prade
IRIT, CNRS and Université de Toulouse,
31062 Toulouse Cedex 09, France

December 11, 2015

Abstract

This survey paper provides an overview of existing methods for building possibility distributions. We both consider the case of qualitative possibility theory, where the scale remains ordinal, and the case of quantitative possibility theory, where the scale is the real interval $[0, 1]$. Methods may be order-based or similarity-based for qualitative possibility distributions, while statistical methods apply in the quantitative case, and then possibilities encode nested random epistemic sets or upper bounds of probabilities. But distance-based approaches, or expert estimates may be also exploited in the quantitative case.

1 Introduction

One of the key questions often raised by scientists when considering fuzzy sets is how to measure membership degrees. However, this question is hardly meaningful if no interpretive context for membership functions is provided. One such context is possibility theory, first outlined by Lotfi Zadeh in 1977 [86]. Possibility distributions are the basic building blocks of possibility theory. Zadeh proposes to consider them as fuzzy set membership functions interpreted in a *disjunctive* way [87], namely, serving as elastic constraints restricting the possible values of a *single-valued* variable. Different kinds of possibility distributions may be encountered in a variety of applications

*To appear in Int. J. Intelligent Systems.

ranging from information systems and databases [5] to operations research [54] and artificial intelligence [35], from computation with ill-known quantities represented by fuzzy intervals [24], to the set of possible models of a possibilistic logic base [23] (see [38] for more references). Whatever the situation, having faithful elicitation or estimation methods for possibility distributions is clearly an important issue.

The idea of graded possibility was thus advocated by Zadeh in the late seventies. But before him, the economist G. L. S. Shackle [67, 68, 69], the philosopher David Lewis [53] did the same, albeit on the basis of concerns very different from Zadeh's. Indeed, Zadeh was mainly motivated by the representation of linguistic terms as a way of expressing uncertain and imprecise information held by humans, referring to some appropriate distance to prototypical examples; in contrast, Shackle was interested in modeling expectations in terms of degrees of potential surprise (which turn out to be degrees of impossibility); and Lewis advocated a comparative possibility-based view of counterfactual conditionals, where the possibility of a world depends on its similarity (or closeness) to a reference world, and is represented in terms of so-called "systems of nested spheres" around this world.

Depending on the situations and the views, the concept of possibility may refer to ideas of feasibility ("it is possible *to* ...") or epistemic consistency ("it is possible *that* ..."), and its evaluation is in practice either a matter of similarity (or distance) – a view recently revived by Zadeh [89], or in terms of cost, or yet of frequency (viewing possibility as upper probability). We shall encounter these different interpretations in the following survey of techniques for constructing possibility distributions.

The paper is organized as follows. In Section 2, we first provide a refresher on possibility theory, distinguishing the qualitative and the quantitative views, emphasizing the role of information principles in the specification of possibility distributions. Section 3 is devoted to methods for generating qualitative possibility distributions as in possibilistic logic, or when dealing with default conditionals. Section 4 provides an overview of elicitation methods for quantitative possibility distributions, based on distances, frequencies, or expert knowledge.

2 Possibility theory: a refresher

This brief overview focuses on the possible meanings of a possibility distribution. We first review the relation between possibility distributions and fuzzy sets, before introducing possibility distributions as a representation tool for imprecise or uncertain information, together with the associated set functions for assessing the plausibility or the certainty of events. We then discuss different qualitative and quantitative scales for grading possibility, and finally address the relations between possibility and probability.

2.1 Possibility distribution and fuzzy set

In his paper introducing possibility theory, Zadeh [86] starts with the representation of pieces of information of the form 'X is A', where X is a parameter or attribute of interest and A is a fuzzy set on the domain of X, often representing a linguistic category (e.g., *John is Tall*, where $X =$ *height(John)*, and A is the fuzzy set of *Tall* heights for humans). The question is then, knowing that 'X is A', to determine what is the possibility distribution π_X restricting the possible values of X (also assuming we know the meaning of A, given by a $[0, 1]$-valued membership function μ_A). Then Zadeh represents the piece of information 'X is A' by the elastic restriction

$$\forall u \in U, \pi_X(u) = \mu_A(u)$$

where U is the universe of discourse on which X ranges. Thus, μ_A is turned into a kind of likelihood function for X. In the above example, U is the set of human heights. Note however that π_X acts as a *disjunctive* restriction (X takes a single value in U), while, prior to using it as above, A is a *conjunctive* fuzzy set [87], the fuzzy set of all values more or less compatible with the meaning of A. Thus the degree of possibility that $X = u$ is evaluated as the degree of compatibility $\mu_A(u)$ of the value u with the fuzzy set A.

2.2 Representation of imprecise information and specificity

In more abstract terms, π_X is a mapping from a referential U (understood as a set of mutually exclusive values for the attribute X) to a totally ordered scale L, with top denoted by 1 and bottom by 0, such as the unit interval $[0, 1]$. Thus any mapping from a set of elements, viewed as a mutually exclusive set of alternatives, to $[0, 1]$ (and more generally to any totally ordered scale) can be seen as acting as an elastic restriction on the value of a single-valued variable, i.e., can be seen as a possibility distribution. Apart from the representation of ill-known numerical quantities defined on continuums, as in the human height example above, another "natural" and simple use of possibility distributions is the representation of ill-known states of affairs (or worlds, according to logicians), a concern of interest for Shackle [68] from a decision perspective.

Then U more generally stands for a (mutually exclusive) set of states of affairs (or descriptions thereof), or states, for short. The function π represents the state of knowledge of an agent (about the actual state of affairs) distinguishing what is plausible from what is less plausible, what is the normal course of things from what is not, what is surprising from what is expected. It represents a flexible restriction on what is the actual state with the following conventions:[1]

- $\pi(u) = 0$ means that state u is rejected as impossible;
- $\pi(u) = 1$ means that state u is totally possible.

If U is exhaustive, at least one of the elements of U should be the actual world, so that $\exists u, \pi(u) = 1$ (normalization). Different values may simultaneously have a degree of possibility equal to 1. In particular, extreme forms of epistemic states can be captured, namely: *complete knowledge*, where for some $u_0, \pi(u_0) = 1$ and $\pi(u) = 0, \forall u \neq u_0$ (only u_0 is possible), and *complete ignorance* where $\pi(u) = 1, \forall u \in U$ (all states are possible).

A possibility distribution π is said to be *at least as specific as* another π' if and only if for each state of affairs u, we have $\pi(u) \leq \pi'(u)$ [84]. Then, π is at least as restrictive and informative as π'. This agrees with Zadeh's entailment principle that 'X is A' entails 'X is B', as soon as $A \subseteq B$. In the presence of pieces of knowledge coming from humans and acting as constraints, possibility theory is driven by the principle of least commitment called *minimal specificity principle* [30]. It states that any hypothesis not known to be impossible cannot be ruled out. In other words, if all we know is that 'X is A', any possibility distribution for which $\pi_X \leq \mu_A$ and $\exists u, \pi_X(u) < \mu_A(u)$ would be too restrictive, since we have no further information that could support the latter strict inequality. Hence, $\pi_X = \mu_A$ is the right representation, if we have no further information. The minimal specificity principle justifies the use of the *minimum*-based combination principle of n pieces of information of the form 'X is A_i', in approximate reasoning [88], since $\pi_X = \min_{i=1}^n \mu_{A_i}$ is the largest possibility distribution such that we have $\pi \leq \mu_{A_i}, \forall i = 1, \ldots, n$.

Sometimes, the opposite principle must be used. This is when we possess statistical information that represents data and not knowledge. In this case we consider the most specific possibility distribution enclosing the data, assuming, like in probability density estimation, that what has not been observed is impossible [40]. This is similar to the closed-world assumption.

[1] The interpretation for 0 is similar to the case of probability, but Shackle's potential surprise scale is stated the other way around: 0 means possible, and the more impossible an event, the more surprising it is.

2.3 Possibilistic set functions

Given a simple query of the form 'does event A occur?' where A is a subset of states, the response to the query can be obtained by computing degrees of possibility and necessity, respectively (assuming the possibility scale $L = [0, 1]$):

$$\Pi(A) = \sup_{u \in A} \pi(u); \ \ N(A) = \inf_{u \notin A} (1 - \pi(u)).$$

$\Pi(A)$ evaluates to what extent A is logically consistent with π, while $N(A)$ evaluates to what extent A is certainly implied by π. The possibility-necessity duality says that a proposition is certain if its opposite is impossible, and this is expressed by

$$N(A) = 1 - \Pi(A^c),$$

where A^c is the complement of A. Generally, $\Pi(U) = N(U) = 1$ and $\Pi(\emptyset) = N(\emptyset) = 0$. Possibility measures satisfy the basic "maxitivity" property

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)).$$

Necessity measures satisfy a "minitivity axiom" dual to that of possibility measures, namely

$$N(A \cap B) = \min(N(A), N(B)),$$

expressing that being certain that $A \cap B$ is the same as being certain of A and of B.

Human knowledge is often expressed in a declarative way, using statements to which belief degrees are attached. This format corresponds to expressing constraints with which the world is supposed to comply. Certainty-qualified pieces of uncertain information of the form '(X is A) is certain to degree α' can then be modeled by the constraint $N(A) \geq \alpha$. The least specific possibility distribution reflecting this information is [30]:

$$\pi_{(A,\alpha)}(u) = \begin{cases} 1, \text{ if } u \in A \\ 1 - \alpha \text{ otherwise} \end{cases} \tag{1}$$

Acquiring further pieces of knowledge consistent with the former leads to updating $\pi_{(A,\alpha)}$ into some $\pi < \pi_{(A,\alpha)}$. Another example where the principle of minimal specificity is useful is when defining the notion of conditioning in possibility theory. The most usual form respects an equation of the form

$$\Pi(A \cap B) = \Pi(A|B) \star \Pi(B), \quad N(A|B) : 1 - \Pi(A^c|B), \tag{2}$$

where $\star$ is a t-norm and $B \neq \emptyset$. The most justified choices of $\star$ are min and product [32]. In the case of product, it looks like probabilistic conditioning applied to possibility measures and corresponds to Dempster conditioning [14]. Using min, the above definition (3) does not yield a unique conditional possibility. Then the idea is to use the least specific possibility measure respecting (2), i.e.,

$$\Pi(A|B) = \begin{cases} \Pi(A \cap B) \text{ if } \Pi(A \cap B) < \Pi(B), \\ 1 \text{ otherwise.} \end{cases} \tag{3}$$

Apart from Π and N, a measure of *guaranteed possibility* or *sufficiency* can be defined [21, 36] : $\Delta(A) = \inf_{u \in A} \pi(u)$. It estimates to what extent *all* states in A are actually possible according to evidence. $\Delta(A)$ can be used as a degree of evidential support for A. In contrast, Π appears to be a measure of *potential* possibility. Uncertain statements of the form "B is possible to degree β" often mean that all realizations of B are possible to degree β. They can then be modeled by the constraint $\Delta(B) \geq \beta$. It corresponds to the idea of observed evidence. This type of information is better exploited by an informational principle opposite to the one discussed above (minimal specificity would give nothing). The most specific distribution $\delta_{(B,\beta)}$ in agreement with $\Delta(B) \geq \beta$ is :

$$\delta_{(B,\beta)}(u) = \left\{ \begin{array}{cc} \beta, & \text{if } u \in B \\ 0 & \text{otherwise.} \end{array} \right\}$$

Acquiring further pieces of evidence leads to updating $\delta_{(B,\beta)}$ into some wider distribution $\delta > \delta_{(B,\beta)}$[36].

2.4 Different scales for graded possibility

There are several representations of epistemic states that are in agreement with the above setting such as: well-ordered partitions [77], Lewis' systems of spheres [53, 51], Spohn's 'Ordinal Conditional Functions' (OCF) [76, 77] (also called ranking functions [78]), and possibilities viewed as upper probabilities. But all these representations of epistemic states do not have the same expressive power. They range from purely qualitative to quantitative possibility distributions, using weak orders, qualitative scales, integers, and reals. In fact we can distinguish several representation settings according to the expressiveness of the scale used [4]:

1. The purely ordinal setting, where an epistemic state on a set of possible worlds is simply encoded by means of a total preorder $\succeq$, telling which worlds are more normal, less surprising than other ones. The quotient set $U/\sim$, built from the equivalence relation $\sim$ extracted from $\succeq$, forms a *well-ordered partition* $E_1, \ldots, E_k$ such that the greater the index i, the less plausible or the less likely

the possible states in E_i. In that case the comparative possibility relation $\succeq_\Pi$ is such that $A \succeq_\Pi B$ if and only if $\exists u_1 \in A, \forall u_2 \in B, u_1 \succeq u_2$. This is the setting used by Lewis [53] and by Grove [51], and Gärdenfors [46] when modeling belief revision. Only possibility measures can account for such relations [16].

2. The qualitative finite setting, with possibility degrees in a finite totally ordered scale: $L = \{\alpha_0 = 1 > \alpha_1 > \cdots > \alpha_{m-1} > 0\}$. This setting has a classificatory flavor, as we assign each event to a class in a finite totally ordered set thereof, corresponding to the finite scale of possibility levels. It is used in possibilistic logic [23]. However, note that the previous purely ordinal representation is less expressive than the qualitative encoding of a possibility distribution on a totally ordered scale, as the former cannot express absolute impossibility.

3. The denumerable setting, using a scale of *powers* $L = \{\alpha^0 = 1 > \alpha^1 > \cdots > \alpha^i > \ldots, 0\}$, for some $\alpha \in (0, 1)$. This is isomorphic to the use of integers in ranking functions by Spohn [78], where the set of natural integers is used as a disbelief scale.

4. The dense ordinal scale setting using $L = [0, 1]$, seen as an ordinal scale. In this case, the possibility distribution Π is defined up to any monotone increasing transformation $f : [0, 1] \to [0, 1], f(0) = 0, f(1) = 1$. This setting is also used in possibilistic logic [23].

5. The dense absolute setting, where $L = [0, 1]$, seen as a genuine numerical scale equipped with product. In this case, a possibility measure can be viewed as a special case of Shafer's plausibility function [71], actually a consonant one, and $1 - \pi$ as a potential surprise function in the sense of Shackle [69].

2.5 Quantitative possibilities and their links with probabilities

The idea of a link between graded possibility and probability is natural since both act as modalities for expressing some form of uncertainty. This link may be stated under the form of a consistency principle [86] stating that "what is possible may not be probable and what is improbable need not be impossible". Proceeding further, we may consider that what is probable should be possible, and what is necessarily (certainly) the case should be probable as well. This amounts to writing $N \leq P \leq \Pi$ where N, P, and Π are, respectively, a necessity, a probability, and a possibility measure ([26] page 138).

Let π be a possibility distribution where $\pi(u) \in [0,1]$. Let $\mathbf{P}(\pi)$ be the never empty set of probability measures P such that $P \leq \Pi$, i.e. $\forall A \subseteq U, P(A) \leq \Pi(A)$ (equivalently, $P \geq N$). Then the possibility measure Π coincides with the upper probability function P^* such that $P^*(A) = \sup\{P(A), P \in \mathbf{P}(\pi)\}$, while the necessity measure N is the lower probability function P_* such that $P_*(A) = \inf\{P(A), P \in \mathbf{P}(\pi)\}$; see [34, 12] for details. P and π are said to be compatible if $P \in \mathbf{P}(\pi)$. So, Π and N are coherent upper and lower probabilities in the sense of Walley [82], as already pointed out very early by Giles [47]. The connection between possibility measures and imprecise probabilistic reasoning is especially interesting for the efficient representation of non-parametric families of probability functions, and it makes sense even in the scope of modeling linguistic information [83].

A possibility measure can thus be computed from a set of nested confidence subsets $\{A_1, A_2, \ldots, A_k\}$ where $A_i \subset A_{i+1}, i = 1 \ldots, k-1$. To each confidence subset A_i is attached a positive confidence level λ_i interpreted as a lower bound of $P(A_i)$, hence a necessity degree. The pair (A_i, λ_i) can be viewed as a certainty-qualified statement that generates a possibility distribution π_i, as recalled above. The corresponding possibility distribution is obtained by intersecting fuzzy sets like those in Equation (1):

$$\pi(u) = \min_{i=1,\ldots,k} \pi_i(u) = \begin{cases} 1 \text{ if } u \in A_1 \\ 1 - \lambda_{j-1} \text{ if } j = \max\{i : u \notin A_i\} > 1. \end{cases} \tag{4}$$

The information modeled by π can also be viewed as a nested random set

$$\{(A_i, m(A_i)), i = 1, \ldots, k\},$$

associated to a belief function [70], letting $m(A_i) = \lambda_i - \lambda_{i-1}$ [27]. This framework allows for imprecision (reflected by the size of the A_i's) and uncertainty (the $m(A_i)$'s). And $m(A_i)$ is the probability that the agent only knows that A_i contains the actual state (it is not $P(A_i)$). The random set view of possibility theory is well adapted to the idea of imprecise statistical data, as developed in Section 4. Conversely, if a belief function is consonant then its contour function $\pi(u) = \sum_{i:u \in A_i} m(A_i)$ is sufficient to recover the belief function, where m is its basic probability assignment ($\sum_i m(A_i) = 1$), and the A_i's are both the nested focal elements associated with m, and the level cuts of π.

Remark 1 *Let us mention another possible kind of link between very small probabilities and possibilities. This interpretation has been pointed out by Spohn [76] for his integer-valued ranking functions κ ranging from 0 to $+\infty$ (0 meaning full possibility,* *and $+\infty$ full impossibility), where $\kappa(A)$ may be thought of as a degree of disbelief modelled by a kind of cost. Namely, $\kappa(A) = k$ is interpreted as a small probability of the form ϵ^k with $\epsilon \ll 1$ (e.g., $P(A) = 10^{-7}$, when $\epsilon = 0.1$, and $k = 7$), i.e.,the probability of a rare event. Indeed if A has a small probability with order of magnitude ϵ^k, and B is another event with a small probability with order of magnitude ϵ^n, the order of magnitude of the probability $P(A \cup B)$ is $\epsilon^{\min(k,n)}$, which mirrors the maxitivity decomposition property of possibility measures, up to a rescaling from $[0, +\infty)$ to $[0, 1]$ [33]. It suggests an interpretation of possibility (and necessity) measures in terms of probabilities of rare events.*

3 Construction methods for qualitative possibility distributions

The elicitation of qualitative possibility distributions is made easier by the qualitative nature of possibility degrees. Indeed, even in a dense ordinal scale $L = [0, 1]$, the precise values of the degrees do not matter, only their relative values are important as expressing strict inequalities between possibility levels. In fact, it basically amounts to determining a well-ordered partition.

In a purely ordinal setting, a possibility ordering is a complete pre-order of states denoted by $\geq_\pi$, which determines a well-ordered partition $\{E_1, \cdots, E_k\}$ of U. It is the comparative counterpart of a possibility distribution π, i.e., $u \geq_\pi u'$ if and only if $\pi(u) \geq \pi(u')$. By convention E_1 contains the most plausible (or normal), or the most satisfactory (or acceptable) states, E_k the least plausible (or most surprising), or the least satisfactory ones, depending if we are modeling knowledge, or preferences. Ordinal counterparts of possibility and necessity measures [16] are defined as follows: $\{u\} \geq_\Pi \emptyset$ for all $u \in U$ and

$$A \geq_\Pi B \text{ if and only if } \max(A) \geq_\pi \max(B)$$

$$A \geq_N B \text{ if and only if } \max(B^c) \geq_\pi \max(A^c).$$

Possibility relations $\geq_\Pi$ are those of Lewis [53]. They satisfy the characteristic property

$$A \geq_\Pi B \text{ implies } C \cup A \geq_\Pi C \cup B,$$

while necessity relations can also be defined as $A \geq_N B$ if and only if $B^c \geq_\Pi A^c$, and satisfy a similar property:

$$A \geq_N B \text{ implies } C \cap A \geq_N C \cap B.$$

Necessity relations coincide with epistemic entrenchment relations in the sense of belief revision theory [46, 33]. In particular the assertion $A >_{\Pi} A^c$ expresses the acceptance of A [19] and is the qualitative counterpart of $N(A) > 0$. This qualitative setting enables qualitative possibility distributions to be derived either from a set of certainty-qualified propositions, or from a set of conditional statements.

3.1 Certainty-qualified propositions

When an agent states beliefs with their (relative) strengths, it is more natural to expect that ordinal information, rather than truly numerical information, is supplied. This gives birth to a knowledge base in the sense of possibilistic logic [23], i.e., a set of weighted statements $K = \{(A_i, \alpha_i) : i = 1, ..., m\}$, each of them representing a constraint $N(A_i) \geq \alpha_i$, where A_i represents a subset of possible states or interpretations, and α_i is the associated certainty level (or priority level) belonging to a denumerable ordinal scale. Such a base K is semantically associated with the possibility distribution in (4), where we no longer assume nested events:

$$\pi_K(u) = \min_{i=1,\ldots,m} \pi_{(A_i,\alpha_i)}(u) = \min_{i=1,\ldots,m} \max(\mu_{A_i}(u), 1 - \alpha_i)$$

and μ_{A_i} is the characteristic function of the subset A_i. Besides, the α_i's may also have a similarity flavor when some pair (A_i, α_i) correspond to the level-cuts of fuzzy subsets [18, 66].

Let us mention that a similar construction can be made in an additive setting where each formula is associated with a cost (in $\mathbb{N} \cup \{+\infty\}$), the weight (cost) attached to an interpretation being the sum of the costs of the formulas in the base violated by the interpretation, as in penalty logic [42]. The so-called "cost of consistency" of a formula is then defined as the minimum of the weights of its models. It is nothing but a ranking function (OCF) in the sense of Spohn [76], the counterpart of a possibility measure defined on $\mathbb{N} \cup \{+\infty\}$, where now 0 expresses full possibility (free violation), and $+\infty$ complete impossibility (a price that cannot be paid). However, this view gives a more quantitative flavor to the construction, thus moving from a qualitative setting to a numerical one.

The construction of π_K from the collection of statements in K clearly relies on the application of the minimal specificity principle. As mentioned in the previous section, a dual principle may be more appropriate when we start from data, rather than constraints excluding impossible states. Assume that we have a collection of weighted data $D = \{(B_j, \beta_j), j = 1, ..., n\}$, understood as $\Delta(B_j) \geq \beta_j$, where the β_j's belong to an ordinal scale and reflect, e.g., some similarity-based relevance of the data. Then by virtue of maximal specificity, we get the lower possibility distribution (which needs not to be normalized):

$$\delta_D(u) = \max_{j=1,\ldots,n} \delta_{(B_j,\beta_j)}(u) = \max_{j=1,\ldots,n} \min(\mu_{B_i}(u), \beta_j).$$

Note that this expression takes the form of the kind of fuzzy conclusions (prior to defuzzification) obtained from Mamdani fuzzy rule-based systems [56].

3.2 Indicative conditionals

Besides, there exists yet another method to obtain a qualitative possibility distribution, starting from a set of *conditionals*, rather than from a set of lower bounds on the necessity, or the guaranteed possibility, of a collection of subsets. This method was originally invented for stratifying a set of default rules in order to design proper methods for handling exception-tolerant reasoning about incompletely described cases; see, e.g., [3]. A default rule "if A then B, generally", denoted $A \rightsquigarrow B$, is then understood formally as the conditional constraint

$$\Pi(A \cap B) > \Pi(A \cap B^c)$$

on a possibility measure Π, expressing that the examples of the rule (the situations where A and B hold) are more plausible than its counter-examples (the situations where A holds and B does not). It is equivalent to the conditional statement $N(B|A) > 0$. Remember that, in contrast, the probabilistic interpretation is such that $P(A \cap B) > P(A \cap B^c)$ if and only if $P(B|A) > 1/2$.

The above possibilistic constraint can be equivalently expressed in terms of a mere comparative possibility relation, namely $A \cap B >_\Pi A \cap B^c$. Any finite *consistent* set of constraints of the form $A_k \cap B_k >_\Pi A_k \cap B_k^c$, representing a set of defaults $\Delta = \{A_k \rightsquigarrow B_k, k = 1, \cdots, r\}$, is compatible with a non-empty family of relations $>_\Pi$, and determines a partially defined ranking $>_\pi$ on U, that can be completed according to the principle of minimal specificity. This principle assigns to each state u the highest possibility level (in forming a well-ordered partition of U) without violating the constraints. It defines a unique complete preorder [3]. Let $E_1, \ldots, E_k$ be the obtained partition. Then $u >_\pi u'$ if $u \in E_i$ and $u' \in E_j$ with $i < j$, while $u \sim_\pi u'$ if $u \in E_i$ and $u' \in E_i$ (where $\sim_\pi$ means $\geq_\pi$ and $\leq_\pi$).

A numerical counterpart to $>_\pi$ on a denumerable finite scale can be defined by $\pi(u) = \frac{k+1-j}{k}$ if $u \in E_j, j = 1, \ldots, k$ [3]. Note that it is purely a matter of convenience to use a numerical scale, and any other numerical counterpart such that $\pi(u) > \pi(u')$ iff $u >_\pi u'$ will work as well. Namely, the range of π is used as an ordinal scale. This approach has an infinitesimal probability counterpart, namely, a procedure called system Z [64]. It has been refined by the numerical system Z^+ [48], whose possibilistic counterpart corresponds to the handling of "strengthened" constraints of the form $\Pi(A_j \cap B_j) > \rho_j \cdot \Pi(A_j \cap B_j^c)$, where $\rho_j \geq 1$. This approach can also be expressed in terms of conditioning in the setting of Spohn's ranking functions. Note that the latter methods were intended to stratify default knowledge bases rather than to explicitly derive possibility distributions.

4 Construction methods for quantitative possibility distributions

The construction of possibility distributions in the quantitative setting either rely on numerical similarity or exploit the connection between probability and possibility inspired by Zadeh [86] according to whom what is probable must be possible, which is understood here by the inequality $\Pi(A) \geq P(A)$, for all measurable subsets A. In the first case, possibility is viewed as a form of renormalized distance to most plausible values. In the second case, it means that we can derive possibility distributions from statistical data or from subjective probability elicitation methods.

4.1 Possibility as similarity

In his approach to the non-Boolean representation of natural language categories, Zadeh [87] uses membership functions representing the extensions of fuzzy predicates in order to derive possibility distributions, as recalled in Section 2.1. If we know the membership function μ_{Tall} of *Tall* on the scale of human heights, then the piece of information *John is Tall*, accepted as being true, can be represented by a possibility distribution $\pi_{hgt(John)}$ equated with μ_{Tall}:

$$\pi_{hgt(John)}(h) = \mu_{Tall}(h).$$

In other words, the measurement of possibility degrees comes down to the measurement of membership functions of linguistic terms. However, in such a situation, $\mu_{Tall}(h)$ is often constructed as a function of the distance between the value a and the closest height $\hat{h}$ that can be considered prototypical for *Tall*, i.e., $\mu_{Tall}(\hat{h}) = 1$, for instance,

$$\mu_{Tall}(h) = f(d(h, \hat{h})) \qquad (5)$$

where f is a non-negative, decreasing function such that $f(0) = 1$, for instance $f(u) = \frac{1}{1+u}$, and $d(h, \hat{h}) = \min\{d(h, x) : \mu_{Tall}(x) = 1\}$, where d is a distance.

Sudkamp [79] points out that conversely, given a possibility distribution π, the two-place function $\delta(x, y) = |\pi(x) - \pi(y)|$ is a pseudo distance indeed.

Results of fuzzy clustering methods can be interpreted as distance-based membership functions. Alternatively one may define a fuzzy set F from a crisp set A of prototypes of μ_{Tall} and a similarity relation $S(x, y)$ on the height scale, such that $S(x, x) = 1$ (then $1 - S(x, y)$ is akin to a distance). Ruspini [65] proposes to define the membership function as a kind of upper approximation of A:

$$\mu_F(h) = \max_{u \in A} S(u, h).$$

Then A stands as the core of the fuzzy set F. We refer the reader to the survey by Türksen and Bilgic [80] for membership degree elicitation using measurement methods outside the possibility theory view, and more recently to papers by Marchant [57, 58].

Besides, the idea of relating plausibility and distance also pervades the probabilistic literature: the use of normal distributions as likelihood functions can be viewed as a way to define degrees of likelihood via the Euclidean distance between a given number and the most likely value (which in that case coincides with the mean value of the distribution). In the neurofuzzy literature, one often uses Gaussian membership functions of the form (5) with $f = e^{-x^2}$.

4.2 Statistical interpretations of possibility distributions

The use of possibility distributions seems to range far beyond the linguistic point of view advocated by Zadeh [87]. Namely, the use of (normalized) membership functions interpreted as ruling out the more or less impossible values of an ill-known quantity X, as well as the maxitivity axiom of possibility measures, are actually often found in the statistical literature, in connection with the non-Kolmogorovian aspects of statistics, namely the maximum likelihood principle, the comparison of probability distributions in terms of dispersion, and the notion of confidence interval; see [17, 61, 62] for surveys of such connections between probability and possibility. In this section, we focus on the derivation of possibility distributions from a (finite) set of statistical data.

4.2.1 Interval data

It is useful to cast the problem in a more general setting, namely the one of set-valued data, and the theory of random sets [8, 45, 50]. Consider a random variable X and a (multi)-set of data reporting the results of some experiments under the form of intervals $\mathcal{D} = \{I_i : i = 1, \ldots, n\}$ subsets of a real interval $U = [a, b]$. In general, due to randomness, one cannot expect this set of intervals to be nested. Representing it by a possibility distribution will result in an approximation to this information. Strictly speaking what is needed to represent this data set exactly is a random set defined by a mass function $m : 2^{[a,b]} \to [0, 1]$ such that

$$m(E) = \frac{|\{I_i : E = I_i\}|}{n}, \forall E \subseteq [a, b] \tag{6}$$

Note that this expression is formally related to a belief function $Bel(A) = \sum_{E \subseteq A} m(E)$ of Shafer [70]. In particular, each focal set E with $m(E) > 0$ represents incomplete information, namely that some $x_i \in I_i$ should have been observed as the result of the ith experiment, but only an imprecise representation of this observation could be obtained in the form of I_i. However, in the theory of evidence, Shafer assumes that $m(E)$ is a subjective probability (the probability that the set E is a faithful representation of an agent's knowledge about X). The interval data is more in conformity with Dempster [14] view, since $m(E)$ is the frequency of observing E.

In fact $\mathcal{D} = \{I_i : i = 1, \ldots, n\}$ is interpreted as an *epistemic random set* [8], i.e., it describes an ill-known standard random variable. It represents the (finite, hence non-convex) set of probabilities obtained by all selections of values in the intervals of $\mathcal{D}$. Let $d^k = \{x_1^k, \ldots, x_n^k\}$ represent a precise data set compatible with $\mathcal{D}$ in the sense that $x_i^k \in I_i, i = 1, \ldots, n$. This is denoted by $d^k \in \mathcal{D}$. Moreover, the belief function $Bel(A)$ is a lower frequency of A, while the plausibility degree $Pl(A) = \sum_{E \cap A \neq \emptyset} m(E)$ is an upper frequency. Let $f^k(a)$ be the frequency of $u = x_i^k$ in d^k. Then :

$$Bel(A) = \min_{d^k \in \mathcal{D}} \sum_{u \in A} f^k(u); \quad Pl(A) = \max_{d^k \in \mathcal{D}} \sum_{u \in A} f^k(u).$$

See [10, 44, 45] for more on statistics with interval data.

A straightforward way of deriving a possibility distribution from such statistical data is to consider what Shafer [70] called the *contour function* of m (actually, the one-point coverage function of the random set):

$$\pi_*(a) = \sum_{a \in E} m(E).$$

Note that this is only a partial view of the data, as it is generally not possible to reconstruct m from π_*. This view of possibility distributions and fuzzy sets as random sets was very early pointed out by Kampé de Feriet [52] and Goodman [49]. From a possibility theory point of view, it has some drawbacks:

- π_* is generally not normalized, hence not a proper possibility distribution (unless the data are not conflicting : $\bigcap_{i=1}^{n} I_i \neq \emptyset$). For instance, $\pi_* = m$ is a probability distribution when data are precise.

- Even when it is normalized, the interval $[N_*(A), \Pi_*(A)]$ determined by π_* is the widest interval of this form contained in $[Bel(A), Pl(A)]$ [31].

One may be more interested to get the narrowest ranges $[N(A), \Pi(A)]$ *containing* intervals $[Bel(A), Pl(A)]$, as being safer; see [31] for an extensive discussion of this difficult problem whose solution is not unique. The idea, first suggested in [29] is to choose a family $\mathcal{F} = \{E_1 \subseteq \cdots \subseteq E_q\}$ of nested intervals such that $I_i \subseteq E_q$ for all intervals I_i, and $I_i \subseteq E_1$ for at least one I_i. Then it is easy to compute a nested random set $m_\mathcal{F}$, as follows: for each interval I_i let $\alpha(i) = \min\{j : I_i \subseteq E_j\}$, such that $E_{\alpha(i)}$ is the most narrow interval in $\mathcal{F}$ containing I_i. Then let $m_\mathcal{F}(E_j) = \sum_{E:E=I_i, j=\alpha(i)} m(E)$, where m is the original mass function given by (6). An upper possibility distribution $\pi_\mathcal{F}$ is derived such that:

$$\pi_\mathcal{F}(a) = \sum_{a \in E_i} m_\mathcal{F}(E_j)$$

in the sense that $[Bel(A), Pl(A)] \subseteq [N_\mathcal{F}(A), \Pi_\mathcal{F}(A)]$. The difficult point is to choose a proper family of nested set $\mathcal{F}$. Clearly, the intervals in $\mathcal{F}$ should be as narrow as possible. One may, for instance, choose $\mathcal{F}$ in the family of cuts of π_*.

Interestingly the random set $\{(E_j, m_\mathcal{F}(E_j)) : j = 1, \ldots, q\}$ can be viewed as a nested histogram, which is what is expected with empirical possibility distributions (while building a standard histogram comes down to choosing a partition of $[a, b]$).

4.2.2 From large precise datasets to possibility distributions

If we consider the special case of a standard point-valued data set, there does not exist a lower possibility distribution, but it is possible to derive an upper possibility distribution using a nested histogram. Of course, we lose much information, as we replace precise values by sets containing them. However, the problem of finding an optimal upper distribution has a solution known for a long time [27, 13]. Consider a histogram $\mathcal{H}$ made of a partition $\{H_1, \ldots, H_n\}$ of $[a, b]$ with corresponding probabilities $p_1 > p_2 > \cdots > p_n$. Note that it is, strictly speaking, a special case of random set with disjoint realizations. Then, there is a most specific possibility distribution π^* dominating the probability distribution, called *optimal transformation*, namely

$$\forall a \in H_i, \pi^*(a) = \sum_{j \geq i} p_j \tag{7}$$

Indeed one can check that $P(A) \in [N^*(A), \Pi^*(A)]$ and $\Pi^*(\bigcup_{i=1}^{j} H_j) = P(\bigcup_{i=1}^{j} H_j)$. The distribution π^* is known as the *Lorentz curve* of the vector $(p_1, p_2, \ldots, p_n)$. In fact, the main reason why this transformation is interesting is that it provides a systematic method for comparing probability distributions in terms of their relative peakedness (or dispersion). Namely, it has been shown that if π_p^* and π_q^* are optimal transformations of distributions p and q (sharing the same order of elements), and $\pi_p^* < \pi_q^*$ (the former is more informative than the latter), then $-\sum_{i=1}^{n} p_i \ln p_i < -\sum_{i=1}^{n} q_i \ln q_i$, and this property holds for all entropies [22].

Note that many authors suggest another transformation consisting in a mere renormalisation of the probability distribution in the style of possibility theory, namely

$$\pi^r(a) = \frac{p_i}{p_1}, \text{ if } a \in H_i. \tag{8}$$

However, it was already indicated in [26], page 259, that the inequality $\Pi^r(A) \geq P(A)$ may fail to hold for some events A. In fact, for $n = 3$, one can prove the following:

Proposition 1 *Consider a probability distribution $p_1 \geq p_2 \geq p_3$ on a 3-element set $\{1, 2, 3\}$. Then $\Pi^r(A) < P(A)$ for some A if and only if $p_1 > 0.5$ and $p_2 < p_1(1-p_1)$.*

Proof The only problematic event is $\{2, 3\}$ as $\Pi^r(A) \geq P(A)$ obviously for other events. Noticing that $p_1 = 1 - p_2 - p_3$, the condition $\Pi^r(\{2, 3\}) = \frac{p_2}{p_1} < P(\{2, 3\})$ boils down to the inequality $p_2 < p_1(1 - p_1)$. Moreover, the condition $p_2 \geq p_3$ is actually $p_2 \geq 1 - p_1 - p_2$, i.e., $p_2 \geq \frac{1-p_1}{2}$. So we need $\frac{1-p_1}{2} < p_1(1 - p_1)$, i.e., $p_1 > 0.5$. $\square$

For instance, take $p_1 = 0.6, p_2 = p_3 = 0.2$; then $\Pi^r(\{2, 3\}) = 1/3 < P(\{2, 3\}) = 0.4$. In the case of more than 3 elements one may find probability values $p_1 \geq \cdots \geq p_n$, such that $\frac{p_i}{p_1} < P(\{i, \ldots, n\})$, for all $i = 2, \ldots, n-1$. It is sufficient to have $p_1 > 0.5$ and then to choose $0 < p_i < p_1(1 - \sum_{j=1}^{i-1} p_j), i = 2, \ldots, n-1$ in this order, making sure that $p_n \leq p_{n-1}$.

4.2.3 Scarce precise data

Another case when a possibilistic representation can be envisaged is when the data set $\mathcal{D} = \{x_i : i = 1, \ldots, n\}$ is too small. Applying estimation methods to compute the probability distribution leads to large confidence intervals. Namely, if $p(x|\theta)$ is the density to be estimated via a parameter θ, then we get confidence intervals J_β for θ with confidence level $\beta \in [0, 1]$. Usually, $\beta = 0.95$ is selected. The interval J_β is random and contains θ with probability at least β. As the confidence intervals are nested, this family of confidence intervals can be modeled by a possibility distribution over the values of θ, which comes down to a possibility distribution over probabilistic models $p(x|\theta)$. This result is similar to the one we get from fuzzy probability qualification of a linguistic statement of the form *X is F' is* $\tilde{p}$ where $\tilde{p}$ is a fuzzy interval on the probability scale. According to Zadeh [87], this piece of information comes down to computing the possibility distribution π over probability measures P (on the range of X) for which $\pi(P) = \mu_{\tilde{p}}(P(F))$ where $P(F)$ is the scalar probability of the fuzzy event F.

Finite setting In the case of a multinomial setting with n states, the identification of the probabilities p_i of states i based on observation frequencies f_i also yields confidence intervals. Fixing the confidence level, one gets probability intervals $[l_i, u_i]$ likely to contain the true probabilities p_i. Such probability intervals lead to upper (and lower) probabilities of events that are submodular (and supermodular), a property far weaker than the property of possibility and necessity measures [11]. They can be approximated by possibility and necessity measures as done by de Campos and Huete [6], Masson and Denoeux [59]; see also Destercke *et al.* [15].

De Campos and Huete consider a finite set of n possibilities, and a small sample of N observations, where N_i is the number of observations of class i. Maximum likelihood gives probabilities $p_i = \frac{N_i}{N}$, and the statistical literature enables bounds $l_i \leq p_i \leq u_i$ to be computed as $p_i \pm c_\epsilon\sqrt{\frac{p_i(1-p_i)}{N}}$ (if inside $[0, 1]$), where c_ϵ is the appropriate percentile of the standard normal distribution. These bounds have the peculiarity that the rankings of the lower bounds, of the upper bounds and of the p_i's are the same. Based on this ranking, the authors consider extending possibility-probability transformations (7) and (8) to probability intervals (as well as the converse of the pignistic transform (11) presented later in this paper) in such a way as to verify a number of expected properties:

1. The obtained possibility degrees for each class should be in agreement with the ranking provided by the sample sizes N_i;
2. The wider the intervals $[l_i, u_i]$, the less specific the possibility distribution;
3. The larger the sample size N, the more specific the possibility distribution;
4. The possibility distribution obtained from any probability assignment in the intervals and in agreement with the sample size should be more specific than the possibility distribution obtained from the intervals.

These transformations are simple to compute. In contrast, Masson and Denoeux [59] consider the probability intervals as being partially ordered and consider the transforms of all probability distributions consistent with these intervals according to all rankings extending the partial order. The obtained possibility distribution is covering all of them. This method is combinatorially more demanding.

Continuous setting An extreme case of scarce data is when a single observation $x = x_0$ on the real line has been obtained. Mauris [60] has shown that if we assume that the generation process is based on a unimodal distribution with mode $M = x_0$, it is possible to compute a possibility distribution whose associated necessity functions bounds the probability of events from below. This perhaps surprising fact comes from the following result [55] used by Mauris: For any value $t > 1$ and any interval $I_t = [x - |x|t, x + |x|t]$ containing the mode M of the distribution, it holds that $P(I_t) \geq 1 - \frac{2}{1-t}, \forall t > 1$. Then if the observed value $x_0 > 0$ is supposed to coincide with the mode of the distribution, we can derive a possibility distribution

$$\pi(x_0(1-t)) = \pi(x_0(1+t)) = \begin{cases} \frac{2}{1-t} \text{ if } t > 1 \\ 1 \text{ otherwise.} \end{cases}$$

This is done by interpreting $1 - \frac{2}{1-t}$ as a degree of necessity and by applying the minimal specificity principle to all such inequality constraints. Then, we know that whatever the underlying probability measure with mode x_0, we get $P(A) \geq N(A)$, where N is constructed from π. The above result of Mauris [60] can be improved if more assumptions are made (symmetry, shape of the distribution) or if several observations obtained. Also, if the variable of interest is known to be bounded, i.e., to lie inside an interval $[a, b]$, Dubois et al. [20] have shown that the triangular possibility distribution with mode x_0 and support $[a, b]$ also dominates the probability of any event A for all unimodal probability distributions with mode x_0 and support in $[a, b]$ (including uniform ones); see Mauris [61, 62] for a more extensive view of the role of possibility distributions in statistics (evaluation of dispersion, estimation methods, etc.).

4.2.4 Possibility measures and cumulative distributions

Possibility distributions, when related to probability measures, are closely related to cumulative distributions, as already suggested by expression (7). Namely, given a family $I_t = [a_t, b_t], t \in [0, 1]$ of nested intervals, such that $t < s$ implies $I_s \subset I_t$, $I_1 = \{\hat{x}\}$, and a probability measure P whose support lies in $[a_0, b_0]$, letting

$$\pi(a_t) = \pi(b_t) = 1 - P(I_t), t \in [0, 1]$$

yields a possibility distribution (it is the membership function of a fuzzy interval) that is compatible with P. Now, $1 - P(I_t) = P((-\infty, a_t)) + P((b_t, +\infty))$ making it clear that the possibility distribution coincides with a two-sided cumulative distribution function. Choosing $I_t = \{x : p(x) \geq t\}$ for $t \in [0, supp]$, where p is the density of P, one gets the most specific possibility distribution compatible with P [39]. It has the same shape as p and $\hat{x}$ is the mode of p. It is the continuous counterpart of equation (7). It provides a faithful description of the dispersion of P.

Conversely, given a possibility distribution in π the form of a fuzzy interval, then the set of probability measures $\mathcal{P}(\pi)$ dominated by its possibility measure Π is equal to $\{P : P(\pi_\alpha) \geq 1 - \alpha, \forall \alpha \in (0, 1]\}$, where $\pi_\alpha = \{x : \pi(x) \geq \alpha\}$, the α-cut of π, is a closed interval $[a_\alpha, b_\alpha]$[9, 20].

When π is an increasing function, it is generally the cumulative distribution of a unique probability measure such that $P((-\infty, x)) = \pi(x)$. Otherwise, a possibility distribution π does not determine a unique probability distribution P, contrary to the situation with usual continuous cumulative distributions. Namely, there is not a unique probability measure such that $\alpha = 1 - P(\pi_\alpha), \forall \alpha \in (0, 1]$. To show there are many probability measures such that $\alpha = 1 - P(\pi_\alpha)$, first consider the upper and lower distributions functions F^+ and F^- determined by π as follows:

$$F^+(x) = \Pi((-\infty, x]), \quad F^-(x) = N((-\infty, x]) \tag{9}$$

It should be clear that if P^+ and P^- are the probability measures associated with cumulative distributions F^+ and F^-, we do have that $\alpha = 1 - P^+(\pi_\alpha)$, and $\alpha = 1 - P^-(\pi_\alpha), \forall \alpha \in (0, 1]$. Indeed, $1 - P^+(\pi_\alpha) = P^+((-\infty, a_\alpha)) + P^+((b_\alpha, +\infty))$. However, $P^+((b_\alpha, +\infty)) = 0$ since the support of P^+ lies at the right-hand side of the core of π. Hence $1 - P^+(\pi_\alpha) = \Pi((-\infty, a_\alpha)) = \alpha$. A similar reasoning holds for P^-, if we notice that $P^-((-\infty, a_\alpha)) = 0$. In fact, we have a more general result:

Proposition 2 *Consider the cumulative distribution function* $F_\lambda = \lambda F^+ + (1-\lambda)F^-$ *with* $\lambda \in [0, 1]$*, and* P_λ *the associated probability measure. Then* $\forall \lambda \in [0, 1], P_\lambda(\pi_\alpha) = 1 - \alpha$.

Proof: Note that

$$F_\lambda(x) = \begin{cases} \lambda\pi(x) \text{ if } x \leq a_1 \\ \lambda \text{ if } x \in [a_1, b_1] \\ \lambda + (1-\lambda)(1-\pi(x)) \text{ if } x \geq b_1 \end{cases}$$

Now: $P_\lambda(\pi_\alpha) = F_\lambda(b_\alpha) - F_\lambda(a_\alpha) = \lambda + (1-\lambda)(1-\alpha) - \lambda\alpha = 1 - \alpha$ $\square$

We also have the following result, laying bare the connection between possibility distributions and the thin clouds of Neumaier [63], already discussed by Destercke et al. [15]:

Proposition 3 *The set of probability measures for which* $\forall\alpha \in [0,1], P(\pi_\alpha) = 1-\alpha$, *where* π *is the membership function of a fuzzy interval, is* $\mathcal{P}(\pi) \cap \mathcal{P}(1-\pi)$.

Proof: We already know that $\mathcal{P}(\pi) = \{P : \forall\alpha \in [0,1], P(\pi_\alpha) \geq 1-\alpha\}$. Now consider the other inequality $P(\pi_\alpha) \leq 1-\alpha$. Let $\bar{\pi} = 1-\pi$ and note that for continuous membership functions we have that $(\bar{\pi})_\alpha = \overline{\pi_{1-\alpha}}$. Now, $P(\pi_\alpha) \leq 1-\alpha$ is equivalent to $P(\overline{\pi_\alpha}) \geq \alpha$, i.e., $P((\bar{\pi})_{1-\alpha}) \geq \alpha$, or, equivalently, $P((\bar{\pi})_\alpha) \geq 1-\alpha$. So, $\{P : \forall\alpha \in [0,1], P(\pi_\alpha) \leq 1-\alpha\} = \mathcal{P}(1-\pi)$. $\square$

See [2] for examples of probabibility measures whose cumulative distributions lie between F^- and F^+ but are not in the credal set $\mathcal{P}(\pi)$. Providing a precise description of the content of $\mathcal{P}(\pi)$ is an interesting topic of research.

4.2.5 Possibility distributions as likelihood functions

Another interpretation of numerical possibility distributions is the likelihood function in non-Bayesian statistics (Smets [73], Dubois et al. [25]). In the framework of an estimation problem, the problem is to determine the value of some parameter $\theta \in \Theta$ that characterizes a probability distribution $P(\cdot \mid \theta)$ over U. Suppose that our observations are summarized by the data set $\hat{d}$. The function $P(\hat{d} \mid \theta), \theta \in \Theta$ is not a probability distribution, but a likelihood function $\mathcal{L}(\theta)$: A value a of θ is considered as being all the more plausible as $P(\hat{d} \mid a)$ is higher, and the hypothesis $\theta = a$ will be rejected if $P(\hat{d} \mid a) = 0$ (or is below some relevance threshold). If we extend the likelihood of elementary hypotheses $\lambda(\theta) = cP(\hat{d}|\theta)$ (it is defined up to a positive multiplicative constant c [43]), viewed as a representation of uncertainty about θ, to disjunctions of hypotheses, the corresponding set-function Λ should obey the laws of possibility measures [7, 17] in the absence of a probabilistic prior, namely, the following properties look reasonable for such a set-function Λ:

- The properties of probability theory enforce $\forall T \subseteq \Theta, \Lambda(T) \leq \max_{\theta \in T} \lambda(\theta)$;
- A set-function representing likelihood should be monotonic with respect to inclusion: If $\theta \in T, \Lambda(T) \geq \lambda(\theta)$;
- Keeping the same scale as probability functions, we assume $\Lambda(\Theta) = 1$.

Then it is clear that

$$\lambda(\theta) = \frac{P(\hat{d}|\theta)}{\max_{\theta\in\Theta} P(\hat{d}|\theta)},$$

and $\Lambda(T) = \max_{\theta\in T}\lambda(\theta)$, i.e., the extended likelihood function is a possibility measure, and the coefficient c is then fixed. We recover Shafer's proposal of a consonant belief function derived from likelihood information [70], more recently studied by Aickin [1]. What is interesting to notice is that a conditional probability $P(A \mid B)$ conveys two meanings. It generally represents frequentist information about the frequency of randomly generated objects having property A in class B; conversely it represents epistemic (non-frequentist) uncertainty about the class B for an object having property A. It is a bifaced notion with one side that is probabilistic and another side possibilistic. Clearly, acquiring likelihood functions is one way of constructing possibility distributions.

4.3 Possibility distributions induced by human-originated estimates

Another source of information for building possibility distributions consists in estimates supplied by human experts on the value of an unknown quantity X of interest, for instance, a failure rate.

4.3.1 Intervals with confidence levels

In the most elementary case, such information from a witness or an expert will most naturally take the form of an interval $I = [a, b]$, since we cannot expect precise knowledge generally. A confidence level λ will be attached to this interval, either because the expert expresses some doubts about the estimate, or because the receiver does not fully trust the competence of the expert. This information can be modeled, following Shafer [70], by a *simple support belief function* with mass $m([a, b]) = \lambda$, while the mass $1-\lambda$ will be allocated to the widest possible range U for the unknown quantity X, expressing ignorance. Clearly, this procedure yields the hat-shaped possibility distribution π, presented in Eq. (1), of the form $\pi(u) = 1$ if $u \in [a, b]$, and $1 - \lambda$ otherwise.

Now the receiver may sometimes find the interval $[a, b]$ too wide to be informative, or, on the contrary, too narrow to be safe enough. It is natural to collect several such human-originated intervals of various sizes and levels of confidence. In contrast with intervals obtained from the imperfect observation of random experiments, intervals coming from one expert will generally be nested, if the latter displays self-consistency.

Considering that there is full dependency between these information items (they come from the same person), the collection of nested intervals $I_1 \subseteq \cdots \subseteq I_n$ with confidence levels λ_i can be viewed as a kind of possibilistic knowledge base and correspond to the "double-staircase-shaped" possibility distribution of Equation (4)

$$\pi(u) = \min_{i=1}^{n} \max(I_i(u), 1 - \lambda_i) = \sum_{i:u \in I_i} m(I_i)$$

where $m(I_i) = \lambda_i - \lambda_{i-1}$. Should the pieces of information (I_i, λ_i) come from independent sources, one would be led to replace min by product in this expression (which would be in full agreement with Dempster's rule of combination). However the intervals would have less chance to be nested.

One may be inspired by the way probability distributions are elicited from experts. In this case information is requested in the form of quantiles of the distributions, typically, the interval $[a, b]$ is such that $P((-\infty, a]) = 0.05$ and $P([b, +\infty)) = 0.05$. Clearly, the hat-shaped possibility distribution induced by the piece of information $[a, b]$ with confidence 0.1 is a weak form of the information supplied by the two quantiles. This information is sometimes augmented by the 0.5 quantile (the median). In that case a more faithful representation of this information is in the form of a belief function with disjoint focal sets.

4.3.2 Expert-originated statistical parameters

Another kind of information experts may supply consists of parameters of an otherwise unknown distribution when the unknown quantity is a random variable. In this case one may use probabilistic inequalities to derive a possibility distribution. For instance, if the expert has a clear idea of the mean $\hat{x}$ of the probability measure P, and of its standard deviation σ, Chebychev inequality gives us a family of inequalities $P(\overline{A_\lambda}) \leq \min(1, \frac{1}{\lambda^2})$ where $A_\lambda = [\hat{x} - \lambda \cdot \sigma, \hat{x} + \lambda \cdot \sigma]$. This nested family corresponds to the possibility distribution $\pi(\hat{x} - \lambda \cdot \sigma) = \pi(\hat{x} + \lambda \cdot \sigma) = \min(1, \frac{1}{\lambda^2})$ [20]. It is consistent with any probability measure with mean $\hat{x}$ and standard deviation σ. The work of Mauris [60] presented above allows to derive a non-trivial possibility distribution from the mere knowledge of the mode of a distribution. Note that the mode corresponds to the idea of most frequently observed values and sounds like a more likely information to be supplied by one expert than for instance the mean value, or even the median. The mode is generally not unique but corresponds to the idea of usual value while the mean value may correspond to seldom observed values e.g. located between modes. If the information about the mode is supplemented by a safe range for the unknown quantity, the triangular fuzzy number with such mode and support is a faithful representation of this information [20, 60], and it a special case of Gauss inequality [81], which dates back to 1823; see Baudrit and Dubois [2] for more details on possibility distributions induced by the knowledge of statistical parameters.

4.3.3 From subjective probabilities to subjective possibilities

One traditional approach to elicitate probability distributions is via fair betting rates. Namely, the subjective probability $P(A)$ of a singular event A, as per an agent, is viewed as the fair price of a lottery ticket that provides one dollar to this agent if this event occurs. Fairness means that the buyer would accept to sell the lottery ticket at the same price. It is clear that for any k mutually exclusive and exhaustive events $A_1, \ldots, A_k$, we must have that $\sum_{i=1}^{k} P(A_i) = 1$ by fear of losing money otherwise. If there is no reason to consider one event more likely than another then $P(A_i) = 1/k$ for all such events.

The legitimacy of this representation of the epistemic state of an agent has been questioned [70, 82, 37]. In particular, it can be considered ambiguous. It presupposes a one-to-one function between epistemic states and probability distributions. However, the subjective distribution would be uniform in both cases where the agent is fully ignorant and when he perfectly knows that the stochastic process generating the events is pure randomness. So it is actually a many-to-one mapping, and given a subjective probability assignment provided by an expert following the betting rate protocol, there is no clue about the precise epistemic state that led to those betting rates.

If we stick to the Bayesian methodology of eliciting fair betting rates from the agent, but we reject the assumption that degrees of beliefs coincide with these betting rates, it follows that the subjective probability distribution supplied by an agent is only a trace of this agent's beliefs. While, in the presence of partial information, beliefs can be more faithfully represented by a set of probabilities, the agent is forced to be additive by the postulates of exchangeable bets. In the Transferable Belief Model [75], the agent's epistemic state is supposed to be represented by a random epistemic set with mass m, and the subjective probability provided by the Bayesian protocol is called the *pignistic probability* [74] (also known as Shapley value in the game-theoretic literature [72]):

$$pp(u_i) = \sum_{j : u_i \in E_j} \frac{m(E_j)}{|E_j|}. \tag{10}$$

This is an extension of the Laplace principle of insufficient reason, whereby uniform betting rates are assumed inside each focal set. Then, given a subjective probability, the problem consists in reconstructing the underlying belief function.

There are clearly several random sets $\{(E_i, m(E_i)) : i = 1 \ldots n\}$ corresponding to a given pignistic probability. It is in agreement with the minimal specificity principle to consider, by default, the least informative among those. It means adopting a pessimistic view on the agent's knowledge. This is in contrast with the case of statistical probability distributions where the available information consists of observed data. Here, the available information being provided by an agent, it is not assumed that the epistemic state is a unique probability distribution. The most elementary way of comparing belief functions in terms of informativeness consists in comparing contour functions in terms of the specificity ordering of possibility distributions. Dubois et al. [41] proved that the least informative random set with a prescribed pignistic probability $p_i = pp(u_i), i = 1, \ldots, n$ is unique and consonant. It is based on a possibility distribution π^{sub}, previously suggested in [28] with a totally different rationale:

$$\pi^{sub}(u_i) = \sum_{j=1}^{n} \min(p_j, p_i). \tag{11}$$

More precisely, let $\mathcal{F}(p)$ be the set of random sets R with pignistic probability p. Let π_R be the possibility distribution induced by R using the one-point coverage Equation (6). Define R_1 to be at least as informative a random set as R_2 whenever $\pi_{R_1} \leq \pi_{R_2}$. Then, the least informative R in $\mathcal{F}(p)$ is precisely the consonant one such that $\pi_R = \pi^{sub}$. Note that, mathematically, Equation (10), when restricted to consonant masses of possibility measures, defines the converse function of Equation (11), i.e., they define a bijection between possibility and probability distributions. Namely, starting from $\pi_1 \geq \cdots \geq \pi_n$ defining the possibility distribution π, computing its associated pignistic probability pp, we have that $\pi^{sub}(u_i) = \sum_{j=1}^{n} \min(pp(u_j), pp(u_i)) = \pi_i$.

By construction, π^{sub} is a subjective possibility distribution. Its merit is that it does not assume human knowledge is precise, like in the subjective probability school. The subjective possibility distribution (11) is less specific than the optimal transformation (7), as expected, i.e., $\pi^{sub} > \pi_p$, generally. The transformation (11) was first proposed in [28] for objective probability, interpreting the empirical necessity of an event as the sum of excesses of probability of realizations of this event with respect to the probability of the most likely realization of the opposite event.

5 Conclusion

One of the most promising seminal off-spring of fuzzy sets introduced in Zadeh's 1965 paper is possibility theory. Possibility theory bridges the gap between artificial intelligence and statistics. The above survey of methods for deriving possibility distributions from data or human knowledge suggests that this framework is one way to go in the problem of membership function assessment. Of course, not all fuzzy sets are possibility distributions, especially those representing utility functions, or those fuzzy sets with a conjunctive interpretation [87], like a vector of ratings in multifactorial evaluations. However, possibility theory clarifies the role of fuzzy sets in uncertainty management and explains why probability degrees, viewed as frequency or betting rates, can be used to derive membership functions.

References

[1] M. Aickin. Connecting Dempster-Shafer belief functions with likelihood-based inference, Synthese, 123(3): 347-364, 2000.

[2] C. Baudrit and D. Dubois. Practical representations of incomplete probabilistic knowledge Computational Statistics & Data Analysis, 51, 86-108, 2006.

[3] S. Benferhat, D. Dubois, H. Prade. Practical handling of exception-tainted rules and independence information in possibilistic logic. Applied Intelligence, 9(2), 101-127, 1998.

[4] S. Benferhat, D. Dubois, H. Prade, M.-A. Williams. A framework for iterated belief revision using possibilistic counterparts to Jeffrey's rule. Fundamenta Informaticae, 99 (2), 147-168, 2010.

[5] P. Bosc, O. Pivert. Fuzzy Preference Queries to Relational Databases, Imperial College Press, 2012

[6] L. de Campos, J. Huete. Measurement of possibility distributions, Int. J. General Systems, 30(3), 309-346, 2001.

[7] G. Coletti and R. Scozzafava. Coherent conditional probability as a measure of uncertainty of the relevant conditioning events. Proc. of ECSQARU03, Aalborg, LNAI 2711, Springer Verlag, 407-418. 2003.

[8] I. Couso, D. Dubois. Statistical reasoning with set-valued information: Ontic vs. epistemic views. International Journal of Approximate Reasoning, Special Issue Harnessing the Information Contained in Low-Quality Data Sources, 55(7),1502-1518, 2014.

[9] I. Couso, S. Montes, P. Gil. The necessity of the strong α-cuts of a fuzzy set. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9(2), 249-262, 2001.

[10] I. Couso, D. Dubois, L. Sanchez. Random Sets and Random Fuzzy Sets as Ill-Perceived Random Variables, Springer Briefs in Computational Intelligence, Springer, 2014.

[11] L. de Campos, J. Huete, S. Moral, Probability intervals: a tool for uncertain reasoning, Int. J. Uncertainty Fuzziness Knowledge-Based Syst. 2, 167-196, 1994.

[12] G. De Cooman, D. Aeyels. Supremum-preserving upper probabilities. Information Sciences, 118, 173-212, 1999.

[13] M. Delgado, S. Moral, On the concept of possibility-probability consistency. Fuzzy Sets and Systems 21, 311-318, 1987.

[14] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat., 38, 325-339,1967.

[15] S. Destercke, D. Dubois, E. Chojnacki. Unifying practical uncertainty representations. Part II: Clouds. International Journal of Approximate Reasoning, 49, 664-677, 2008.

[16] D. Dubois. Belief structures, possibility theory and decomposable confidence measures on finite sets, Computers and Artificial Intelligence (Bratislava), 5(5), 403-416, 1986.

[17] D. Dubois. Possibility theory and statistical reasoning. Computational Statistics & Data Analysis, 51, 47-69, 2006.

[18] D. Dubois, H. Fargier, H. Prade: Possibility theory in constraint satisfaction problems: Handling priority, preference and uncertainty. Applied Intelligence, 6, 1996, 287-309.

[19] D. Dubois, H. Fargier, H. Prade Ordinal and probabilistic representations of acceptance. J. Artificial Intelligence Research, 22, 23-56, 2004.

[20] D. Dubois, L. Foulloy, G. Mauris, H. Prade. Possibility/probability transformations, triangular fuzzy sets, and probabilistic inequalities. Reliable Computing, 10, 273-297, 2004.

[21] D. Dubois, P. Hajek, H. Prade. Knowledge-driven versus data-driven logics. J. Logic, Lang. and Inform. 9, 65–89, 2000.

[22] D. Dubois, E. Huellermeier. Comparing probability measures using possibility theory: A notion of relative peakedness. International Journal of Approximate Reasoning, 45, 364-385, 2007.

[23] D. Dubois, J. Lang, H. Prade, Possibilistic logic. In: Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3, (D. M. Gabbay, C. J. Hogger, J. A. Robinson, and D. Nute, eds.), Oxford University Press, 439-513, 1994.

[24] D. Dubois, E. Kerre, R. Mesiar, H. Prade. Fuzzy interval analysis. In: Fundamentals of Fuzzy Sets, Dubois,D. Prade,H., Eds: Kluwer, Boston, Mass, The Handbooks of Fuzzy Sets Series, 483-581, 2000.

[25] D. Dubois, S. Moral, H. Prade. A semantics for possibility theory based on likelihoods, J. of Mathematical Analysis and Applications, 205, 359-380, 1997.

[26] D. Dubois, H. Prade. Fuzzy Sets and Systems: Theory and Applications. Academic Press 1980.

[27] D. Dubois, H. Prade. On several representations of an uncertain body of evidence. In: Fuzzy Information and Decision Processes, (M. Gupta, E. Sanchez, eds.), North-Holland, Amsterdam, 167-181,1982.

[28] D. Dubois, H. Prade. Unfair coins and necessity measures: towards a possibilistic interpretation of histograms. Fuzzy Sets and Systems, 10, 15-20, 1983.

[29] D. Dubois, H. Prade. Fuzzy sets and statistical data, Europ. J. Operations Research, 25, 345-356, 1986.

[30] D. Dubois, H. Prade. Possibility Theory. Plenum, New York, 1988.

[31] D. Dubois, H. Prade. Consonant approximations of belief functions. Int. J. Approximate Reasoning, 4, 419-449, 1990.

[32] D. Dubois, H. Prade. The logical view of conditioning and its application to possibility and evidence theories, Int. J. of Approximate Reasoning, 4(1), 23-46, 1990.

[33] D. Dubois, H. Prade. Epistemic entrenchment and possibilistic logic. Artificial Intelligence, 50, 223-239, 1991.

[34] D. Dubois, H. Prade. When upper probabilities are possibility measures, Fuzzy Sets and Systems, 49, 65-74, 1992.

[35] D. Dubois, H. Prade. Fuzzy set and possibility theory-based methods in artificial intelligence. Artif. Intell. 148(1-2): 1-9 (2003).

[36] D. Dubois, H. Prade. An overview of the asymmetric bipolar representation of positive and negative information in possibility theory. Fuzzy Sets and Systems, 160(10) 1355-1366, 2009.

[37] D. Dubois, H. Prade. Formal representations of uncertainty. Decision-making Process- Concepts and Methods. In: D. Bouyssou,et al. (Eds.), ISTE London & Wiley, Chap. 3, 85-156, 2009.

[38] D. Dubois, H. Prade. Possibility theory and its applications: Where do we stand? Springer Handbook of Computational Intelligence (J. Kacprzyk, W. Pedrycz, Eds), Springer, p. 31-60, 2015

[39] D. Dubois, H. Prade and Sandri S. On possibility/probability transformations. In: Fuzzy Logic. State of the Art, In: R. Lowen, M. Roubens (eds.), Kluwer Acad. Publ., Dordrecht, 103-112, 1993.

[40] D. Dubois, H. Prade, P. Smets. New semantics for quantitative possibility theory. In: S. Benferhat, A. Hunter, Eds., Symbolic and Quantitative Approaches to Reasoning with Uncertainty, (ECSQARU 2001), LNCS 2143 Springer-Verlag, Berlin, 410-421, 2001.

[41] D. Dubois H. Prade, P. Smets. A definition of subjective possibility. Int. J. of Approximate Reasoning, 48, 352-364, 2008.

[42] F. Dupin de Saint Cyr, J. Lang, and Th. Schiex. Penalty logic and its link with Dempster-Shafer theory. Proc. Annual Conf. on Uncertainty in Artificial Intelligence (UAI'94), (R. Lopez de Mantaras and D. Poole, eds.), Seattle, July 29-31, 204-211. Morgan Kaufmann, 1994.

[43] W. F. Edwards. Likelihood, Cambridge University Press, Cambridge, U.K., 1972.

[44] S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002-4015. Albuquerque, NM, 2003.

[45] S. Ferson, L. Ginzburg, V. Kreinovich, L. Longpré, M. Aviles. Exact Bounds on Finite Populations of Interval Data. Reliable Computing 11(3): 207-233, 2005

[46] P. Gärdenfors. Knowledge in Flux, MIT Press, Cambridge, MA., 1988.

[47] R. Giles. Foundations for a theory of possibility. In: Fuzzy Information and Decision Processes, (M. Gupta, E. Sanchez, eds.), North-Holland, Amsterdam, 183-196,1982.

[48] M. Goldszmidt, J. Pearl. Qualitative probability for default reasoning, belief revision and causal modeling. Artificial Intelligence, 84, 52-112, 1996.

[49] I. R. Goodman. Fuzzy sets as equivalence classes of random sets. In R. Yager, ed. Fuzzy Sets and Possibility Theory: Recent Developments. Pergamon Press, Oxford, 327-342, 1981.

[50] C. Joslyn. Measurement of possibilistic histograms from interval data. Int. J. of General Systems, 26, 9-33, 1997.

[51] A. Grove. Two modellings for theory change. J. Philos. Logic, 17, 157-170, 1988.

[52] J. Kampé de Fériet, Interpretation of membership functions of fuzzy sets in terms of plausibility and belief. In: Fuzzy Information and Decision Processes, (M. Gupta, E. Sanchez, eds.), North-Holland, Amsterdam, 93-98,1982.

[53] D. K. Lewis. Counterfactuals and comparative possibility. J. of Philosophical Logic, 2, 418-446,1973. Reprinted in *Ifs*, (W. L. Harper, R. Stalnaker, G. Pearce, eds.), D. Reidel, Dordrecht, 57-85, 1981.

[54] W. Lodwick, J. Kacprzyk (Eds.) Fuzzy Optimisation: Recent Advances and applications. Suties in Fuzziness and Soft Computing, Vol. 254. Springer, 2010.

[55] R.E. Machol, J. Rosenblatt Confidence intervals based on a single observation. Proc. of the IEEE, 54, 1087-1088, 1966.

[56] E.H. Mamdani. Advances in the linguistic synthesis of fuzzy controllers, International Journal of Man-Machine Studies 8, 669-678, 1976.

[57] T. Marchant. The measurement of membership by comparisons. Fuzzy Sets and Systems, 148: 157-177, 2004.

[58] T. Marchant. The measurement of membership by subjective ratio estimation Fuzzy Sets and Systems, 148: 179-199, 2004.

[59] M. Masson, T. Denoeux, Inferring a possibility distribution from empirical data, Fuzzy sets and Systems, 157, 319-340, 2006.

[60] G. Mauris: Inferring a Possibility Distribution from Very Few Measurements. In: D. Dubois et al., Eds., Soft Methods for Handling Variability and Imprecision (SMPS 2008), Advances in Soft Computing 48, Springer, 92-99, 2008.

[61] G. Mauris: Possibility distributions: A unified representation of usual direct-probability-based parameter estimation methods. Int. J. Approx. Reasoning 52(9): 1232-1242, 2011.

[62] G. Mauris: A Review of Relationships Between Possibility and Probability Representations of Uncertainty in Measurement. IEEE trans. Instrumentation and Measurement 62(3): 622-632, 2013.

[63] A. Neumaier,. Clouds, fuzzy sets and probability intervals. Reliable Computing, 10, 249-272, 2004.

[64] J. Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In Proc. 3rd Conf. on Theoretical Aspects of Reasoning about Knowledge, Pacific Grove, (R. Parikh, ed.), Morgan Kaufmann, 121-135,1990.

[65] E.H. Ruspini. On the semantics of fuzzy logic, Internat. J. Approx. Reasoning 5(1) 45-88, 1991.

[66] S. Schockaert, H. Prade. Solving conflicts in information merging by a flexible interpretation of atomic propositions. Artif. Intell. 175(11): 1815-1855, 2011.

[67] G. L. S. Shackle. The expectational dynamics of the individual. Economica, 10 (38), 99-129, 1943.

[68] G. L. S. Shackle. Expectation in Economics. Cambridge University Press, UK, 1949. 2nd edition, 1952.

[69] G. L. S. Shackle. Decision, Order and Time in Human Affairs (2nd edition), Cambridge University Press, UK, 1961.

[70] G. Shafer. A Mathematical Theory of Evidence. Princeton University Press, 1976.

[71] G. Shafer. Belief functions and possibility measures. In: Analysis of Fuzzy Information, Vol. I: Mathematics and Logic, (J. C. Bezdek, ed.), CRC Press, Boca Raton, 51-84, 1987.

[72] S. Shapley. A value for n-person games. In Kuhn and Tucker, eds., Contributions to the Theory of Games, II, Princeton University Press, 307-317, 1953.

[73] P. Smets, Possibilistic Inference from statistical data. Proceedings of the Second World Conference on Mathematics at the Service of Man (A. Ballester, ed.), Las Palmas (Spain) 611-613, 1982.

[74] Smets P.. Constructing the pignistic probability function in a context of uncertainty, In Henrion M. et al., (Eds.) Uncertainty in Artificial Intelligence, vol. 5, North-Holland, Amsterdam, 29-39, 1990.

[75] P. Smets, R. Kennes. The transferable belief model, Artificial Intelligence, 66, 191-234, 1994.

[76] W. Spohn. Ordinal conditional functions: a dynamic theory of epistemic states. In: Causation in Decision, Belief Change, and Statistics, vol. 2, (W. L. Harper, B. Skyrms, eds.), Kluwer, 105-134, 1988.

[77] W. Spohn. A general, nonprobabilistic theory of inductive reasoning. In: R. D. Shachter, et al., eds., *Uncertainty in Artificial Intelligence*, Vol. 4. Amsterdam: North Holland. pages 149-158, 1990.

[78] W. Spohn. The Laws of Belief: Ranking Theory and its Philosophical Applications, Oxford University Press, UK, 2012.

[79] T. Sudkamp. Similarity and the measurement of possibility. Actes Rencontres Francophones sur la Logique Floue et ses Applications, Cepadues Editions, Toulouse, France,13-26, 2002.

[80] I. B. Türksen, T. Bilgic. Measurement of membership functions: Theoretical and empirical work. In: Fundamentals of Fuzzy Sets (D. Dubois, H. Prade, eds.), The Handbooks of Fuzzy Sets, Kluwer Publ. Comp., 195-230, 2000.

[81] G. Upton, I. Cook. Gauss inequality. A Dictionary of Statistics. Oxford University Press, 2008.

[82] P. Walley . Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, 1991.

[83] P. Walley, G. De Cooman. A behavioural model for linguistic uncertainty. Information Sciences, 134, 1-37, 1999.

[84] R. R. Yager, An introduction to applications of possibility theory. Human Systems Management, 3, 246-269, 1983.

[85] L. A. Zadeh. Fuzzy sets, Information and Control, 8, 338-353, 1965.

[86] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1, 3-28, 1978 (originally, Memo UCB/ERL M77/12, 1977, University of California, Berkeley).

[87] L. A. Zadeh, PRUF - a meaning representation language for natural languages. Int. J. Man-Machine Studies, 10, 395-460, 1978.

[88] L. A. Zadeh. A theory of approximate reasoning. In: Machine Intelligence, Vol. 9, (J. E. Hayes, D. Mitchie, and L. I. Mikulich, eds.), Elsevier, 149-194, 1979.

[89] L. A. Zadeh. A note on similarity-based definitions of possibility and probability. Inf. Sci., 267, 334-336, 2014.

INCONSISTENCY MANAGEMENT FROM THE STANDPOINT OF POSSIBILISTIC LOGIC*

Didier Dubois and Henri Prade
IRIT-CNRS, Université Paul Sabatier, 31062 Toulouse Cedex 09, France

December 15, 2015

Abstract:Uncertainty and inconsistency pervade human knowledge. Possibilistic logic, where propositional logic formulas are associated with lower bounds of a necessity measure, handles uncertainty in the setting of possibility theory. Moreover, central in standard possibilistic logic is the notion of inconsistency level of a possibilistic logic base, closely related to the notion of consistency degree of two fuzzy sets introduced by L. A. Zadeh. Formulas whose weight is strictly above this inconsistency level constitute a sub-base free of any inconsistency. However, several extensions, allowing for a paraconsistent form of reasoning, or associating possibilistic logic formulas with information sources or subsets of agents, or extensions involving other possibility theory measures, provide other forms of inconsistency, while enlarging the representation capabilities of possibilistic logic. The paper offers a structured overview of the various forms of inconsistency that can be accommodated in possibilistic logic. This overview echoes the rich representation power of the possibility theory framework.

Keywords : inconsistency; fuzzy set; possibility theory; possibilistic logic.

1 Introduction

The intersection of two fuzzy sets may not be normalized. This state of fact may have several readings. In his founding paper, Lotfi Zadeh[35] already introduces the

*Published in Int. J. Uncertainty, Fuzziness and Knowledge-based Systems, vol. 23, Suppl. 1, 2015, pp. 15-30

notion of degree of *separation* of two (convex) fuzzy sets A and B, as the complement to 1 of the height[1] of their min-based intersection. At that time, a fuzzy set was understood *conjunctively* as a collection of elements gathered into a class having an unsharp (gradual) boundary. Later, when Zadeh proposes to use fuzzy sets as a basis for representing possibility distributions[37], the understanding of a fuzzy set becomes *disjunctive* as an elastic restriction on the possible values of a *single*-valued variable. Then in this latter perspective, Zadeh[38] defines the *consistency* of two fuzzy sets as the height of their min-based intersection, without making an explicit use of it. However, in the same year, he defines the possibility of a fuzzy event[37], which is nothing but the degree of consistency between the fuzzy event and the possibility distribution representing what is known.

This idea of consistency is in full agreement with the classical logic notion of consistency between two propositions, which requires the existence of at least one common model for the two propositions. In such a case, Zadeh's consistency degree is equal to 1, but his proposal extends the classical view by making it a matter of degree as soon as at least one of the two propositions is associated with a fuzzy set of models. Such a situation is encountered in possibilistic logic[20, 18, 24] where possibilistic logic formulas are semantically represented by means of a special type of possibility distributions. Moreover, the introduction of a consistency degree was also mentioned in relation with Zadeh's theory of approximate reasoning[39] based on the combination / projection of possibility distributions, which encompasses possibilistic logic inference[15, 16]. As it turned out, the degree of *inconsistency*, which is the complement to 1 of the degree of consistency, of a possibilistic logic knowledge base (made of a conjunction of possibilistic logic formulas) plays a great rôle in the possibilistic logic capability to handle inconsistency.

It can be observed that, in practice, a set of pieces of information is often inconsistent. Inconsistency often comes from the fact that the information is provided by different sources, but information provided by a person can be inconsistent as well. It may plainly take place between two opposite statements that are simultaneously held as certain. There are two basic ways to get around this problem: either one may restore consistency by isolating and mending parts of the information base judged to be responsible for the inconsistency; or one may alter the standard inference notion so as to make it more cautious, in order to preserve a consistent set of derived

[1] The notion of height of a fuzzy set (defined as the supremum of the membership degrees) was first introduced by Zadeh in his work on similarity relations[36] where he observes that for a $\max - \min$ transitive fuzzy relation, the height of the intersection of the fuzzy equivalence classes of two elements x_i and x_j is less or equal to their degree of similarity. This expresses that the more the equivalence classes overlap, the more similar their elements.

conclusions.

However more subtle situations of inconsistency exist; for instance, if the representation setting is rich enough, an inconsistency may occur between the fact that something is known, while at the same time it is believed that it is not possible to know it. One may also have situations where apparently contradictory statements are in fact consistent once they are properly represented, as the statements “the museum is open in the morning”, and the “the museum is open from 2 to 5 p.m”. In any case, two contradictory statements cannot be simultaneously accepted as true. Then, rather than concluding, like in mathematics, that anything follows from a contradiction (the famous *ex falso quodlibet sequitur*), it is more useful to understand the origin and the nature of the inconsistency, and try to derive safe conclusions that overcome it.

Possibilistic logic[20, 18, 24] associates classical propositional formulas (and more generally first order logic formulas) with weights which may be lower bounds of different types of confidence evaluations making sense in possibility theory [37, 19]. The fact that possibility and necessity are graded provides additional power for handling inconsistency. This framework is expressive enough to represent various types of information, and may account for different situations of inconsistency. The paper surveys the existing works in possibilistic logic from an inconsistency-handling point of view. We first restate standard possibilistic logic where formulas are associated with lower bounds of necessity measures, before considering its extension to formulas having a graded paraconsistency level, or coming from different sources. We then additionally introduce formulas associated with lower bounds of weak or strong variant of possibility measures. This paper borrows some material from two conference papers[22, 26], and then merges it in an expanded overview.

2 Possibility Theory and Possibilistic Logic

In the following, formulas of a finite propositional language $\mathcal{L}$ will be denoted by Greek letters such as φ, or ψ. $\top$ and $\bot$ stand for tautology and contradiction respectively. For simplicity, we denote by Ω the set of interpretations of $\mathcal{L}$ that describe possible worlds; $\omega \models \varphi$ denotes the satisfaction of φ by interpretation ω, then called a model of φ. The set of models of φ is denoted by $[\varphi]$. The negation of φ is $\neg\varphi$. We also use conjunction and disjunction symbols $\wedge, \vee$. Finally classical syntactic inference is denoted by $\vdash_{CL}$.

2.1 Necessity and possibility measures

A possibility distribution is a mapping π from a set of possible worlds Ω to the interval $[0, 1]$, which is viewed as a totally ordered bounded ordinal scale. Given a possible world $\omega \in \Omega$, $\pi(\omega)$ represents the degree of compatibility of ω with the available information (or beliefs) about the real world. $\pi(\omega) = 0$ means that ω is impossible, and $\pi(\omega) = 1$ means that nothing prevents ω from being the real world. When $\pi(\omega_1) > \pi(\omega_2)$, ω_1 is preferred to ω_2 as a candidate for being the real state of the world. The less $\pi(\omega)$, the less plausible ω, or the less likely it is the real world. A possibility distribution π is said to be *normalized* if $\exists\omega \in \Omega$, such that $\pi(\omega) = 1$, in other words, if at least one possible world is a fully plausible candidate for being the actual world. In that case, the knowledge represented by π is considered to be consistent. Interpretations ω where $\pi(\omega) = 1$ are considered to be normal (they are not at all surprising). A sub-normalized possibility distribution π (such as $height(\pi) = \max_{\omega\in\Omega} \pi(\omega) < 1$) is considered self-conflicting to some extent (since the existence of at least one fully possible interpretation is not acknowledged). The case where $\forall\omega, \pi(\omega) = 0$ encodes a full contradiction. A consistent epistemic state is thus always encoded by a normalized possibility distribution.

Given a possibility distribution π, the possibility degree of proposition φ is defined as:

$$\Pi(\varphi) = \max\{\pi(\omega) : \omega \models \varphi\}.$$

It evaluates to what extent φ is consistent with the possibility distribution π. Note that by definition if $\varphi \equiv \psi$ then $\Pi(\varphi) = \Pi(\psi)$, since $[\varphi] = [\psi]$. A necessity measure N is always associated by duality with a possibility measure Π, namely

$$N(\varphi) = 1 - \Pi(\neg\varphi)$$

where $1 - (.)$ is the order-reversing map of the scale. The necessity measure $N(\varphi) = \min\{1 - \pi(\omega) : \omega \not\models \varphi\}$ evaluates to what extent there does not exist a highly plausible interpretation that violates φ, in other words to what extent φ can be deduced from the underlying possibility distribution π. Hence $N(\varphi)$ is a measure of the certainty of φ.

The duality between possibility and necessity extends the one in modal logic: it expresses that the impossibility of $\neg\varphi$ entails the certainty of φ. A necessity measure N is a function from the set of logical formulas to the totally ordered bounded scale $[0, 1]$, which is characterized by the axioms:

i) $N(\top) = 1$,

ii) $N(\bot) = 0$,

iii) if $\varphi \equiv \psi$ then $N(\varphi) = N(\psi)$,

iv) $N(\varphi \wedge \psi) = \min(N(\varphi), N(\psi))$.

Axiom (iv) expresses that $\varphi \wedge \psi$ is as certain as the least certain of φ and ψ. It follows from the axioms of necessity measures that having both $N(\varphi) > 0$ and $N(\neg\varphi) > 0$ forms a contradiction. In other words, one cannot be both somewhat certain of a proposition and of its negation. Moreover, one may have both $\Pi(\varphi) = 1$ and $\Pi(\neg\varphi) = 1$ without contradiction; it just acknowledges a state of (complete) ignorance about the truth value of φ.

2.2 Possibilistic logic: Syntax

We now recall the main features of possibilistic logic, before discussing some paraconsistent and multiple source extensions thereof in Sections 3 and 4. An atomic possibilistic logic formula[20] is a pair (φ, a) made of a classical logic formula φ and a positive real in $(0, 1]$. The weight a is interpreted as a lower bound for a necessity degree, i.e., the possibilistic logic formula (φ, a) is semantically understood as the constraint $N(\varphi) \geq a$, where N is a necessity measure. Note that formulas of the form $(\varphi, 0)$ do not contain any information (since for all $\varphi, N(\varphi) \geq 0$ always holds) and are not part of the language of possibilistic logic.

The min-decomposability of necessity measures allows us to work with weighted clauses without lack of generality, since $N(\bigwedge_{i=1,k} \varphi_i) \geq a \Leftrightarrow \forall i, N(\varphi_i) \geq a$, i.e., $(\bigwedge_{i=1,k} \varphi_i, a) \Leftrightarrow \bigwedge_{i=1,k} (\varphi_i, a)$.

The proof system of propositional possibilistic logic consists of axioms of propositional logic with weight 1, and the following weighted modus ponens rule of inference:

$$(\varphi, a), (\neg\varphi \vee \psi, b) \vdash (\psi, \min(a, b)).$$

where $\vdash$ denotes the syntactic inference in possibilistic logic. The following derived inference rules are valid in possibilistic logic:

- $(\neg\varphi \vee \psi, a)$, $(\varphi \vee \rho, b) \vdash (\psi \vee \rho, \min(a, b))$ *(resolution)*
- $\forall\, b \leq a$ $(\varphi, a) \vdash (\varphi, b)$ *(weight weakening)*
- if $\varphi \vdash_{CL} \psi$, then $(\varphi, a) \vdash (\psi, a)$ *(logical weakening)*
- (φ, a), $(\varphi, b) \vdash (\varphi, \max(a, b))$ *(weight fusion)*

where $\vdash_{CL}$ denotes the classical logic entailment. Classical inference is retrieved when the weights are equal to 1. Moreover $K \vdash (\varphi, a)$ if and only if $K_a \vdash_{CL} \varphi$, where K_a is a classical logic base that is the a-level cut of the possibilistic logic base K, defined by $K_a = \{\varphi \mid (\varphi, b) \in K \text{ with } b \geq a\}$.

Finally, proving (φ, a) from a possibilistic logic base K also amounts to adding $(\neg\varphi, 1)$, put in clausal form, to K, and using the resolution rule repeatedly in order to show that $K \cup \{(\neg\varphi, 1)\} \vdash (\perp, a)$.

2.3 Possibilistic logic: Semantics

From a semantic viewpoint, a possibilistic logic base $K = \{(\varphi_i, a_i)\}_{i=1,\ldots,m}$ is associated with a possibility distribution π_K representing the fuzzy set of models ω of K:

$$\pi_K(\omega) = \min_{i=1,\ldots,m} \max(\mu_{[\varphi_i]}(\omega), 1 - a_i) \tag{1}$$

where $\mu_{[\varphi_i]}$ is the characteristic function of the sets of models of φ_i. It can be shown that π_K is the largest possibility distribution such that $N_K(\varphi_i) \geq a_i, \forall i = 1, m$, i.e., the possibility distribution that allocates the greatest possible possibility degree to each interpretation in agreement with the constraints induced by K (where N_K is the necessity measure associated with π_K), namely

$$N_K(\varphi) = \min_{\omega \in [\neg\varphi]} (1 - \pi_K(\omega))).$$

Thus, a possibilistic logic base is associated with a fuzzy set of models. It represents the set of more or less plausible states of the world (according to the available information), when dealing with uncertainty. A possibility distribution which rank-orders possible states is thus semantically equivalent to a possibilistic logic base. The semantic entailment is then defined by

$$K \models (\varphi, a) \text{ if and only if } N_K(\varphi) \geq a.$$

It is also equivalent to $\forall\omega\ \pi_K(\omega) \leq \pi_{\{(\varphi,a)\}}(\omega) = \max(\mu_{[\varphi]}(\omega), 1 - a)$. Indeed, $N_K(\varphi) \geq a$ is easily rewritten as $\pi_K(\omega) \leq 1 - a$ if $\omega \models \neg\varphi$. It is worth noticing that $\pi_K \leq \pi_{\{(\varphi,a)\}}$ is nothing but the entailment principle in Zadeh's approach to approximate reasoning[39].

The syntactic inference machinery of possibilistic logic, using resolution and refutation, has been proved to be sound and complete with respect to the semantics[18]. Soundness and completeness are expressed by:

$$K \vdash (\varphi, a) \Leftrightarrow K \models (\varphi, a)$$

2.4 Inconsistency level

An important feature of possibilistic logic is its ability to deal with inconsistency. The level of inconsistency of a possibilistic logic base is defined as

$$inc(K) = \max\{a \mid K \vdash (\bot, a)\}$$

(by convention $\max \emptyset = 0$). We can explain this inconsistency level with the a-cuts: the inconsistency level of a base is the greatest value a such that the corresponding a-cut is classically inconsistent. Clearly, any entailment $K \vdash (\varphi, a)$ with $a > inc(K)$ can be rewritten as

$$K^{cons,a} \vdash (\varphi, a),$$

where

$$K^{cons,a} = \{(\varphi_i, a_i) \in K^{cons} \text{ with } a_i \geq a\}$$

and

$$K^{cons} = K \setminus \{(\varphi_i, a_i) \text{ with } a_i \leq inc(K)\}.$$

K^{cons} is the set of formulas whose weights are above the level of inconsistency. Thus they are not affected by the inconsistency, since more entrenched. Indeed, $inc(K^{cons}) = 0$, and more generally, $inc(K) = 0$ if and only if the *skeleton* $K^* = \{\varphi_i \mid (\varphi_i, a_i) \in K)\}$ of K is consistent in the usual sense. Moreover, it can be shown that

$$inc(K) = 1 - \max_{\omega} \pi_K(\omega) = 1 - height(\pi_K). \tag{2}$$

It is important to observe that formulas φ derived from K with a level at most $inc(K)$ are drowned in the sense that $(\neg\varphi, inc(K))$ can be derived as well. They cannot be inferred nor be used in a valid proof. It includes formulas φ in K whose declared certainty level is smaller or equal to $inc(K)$, which cannot be sufficiently increased by deduction from K (even if these formulas φ do not belong to any minimal inconsistent subset of K^*). A way to partially escape the drowning effect is presented in the next section.

Lastly, let us also observe that if a possibilistic logic base K contains two (fully) *consistent* sub-bases C and C' (i.e., $C \subset K$, $C' \subset K$, $inc(C) = inc(C') = 0$)), such that $C \vdash (\varphi, a)$ and $C' \vdash (\neg\varphi, a')$, then $inc(K) \geq 1 - \max(1 - a, 1 - a') = \min(a, a')$. Thus, $inc(K) > 0$ reveals the existence of consistent arguments in K in favor of contradictory statements with certainty levels at least equal to $inc(K)$.

3 Handling Inconsistency in Possibilistic Logic

One may take advantage of the certainty weights for handling inconsistency in inferences, while avoiding the drowning effect (at least partially). We briefly survey two ways to cope with this problem.

3.1 Degree of paraconsistency and safely supported-consequences

An extension of the possibilistic inference has been proposed for handling inconsistent information and getting safely supported consequences[8] only. It requires the definition of a "paraconsistent completion"[15] of the considered possibilistic logic base K, as a first step. For each formula φ such that (φ, a) is in K, we extend the language and compute triples (φ, b, c) where b (resp. c) is the highest degree with which φ (resp. $\neg\varphi$) is supported in K. More precisely, φ is said to be *supported in K at least at degree b* if there is a *consistent* sub-base of $(K_b)^*$ that entails φ, where $K_b = \{(\varphi_i, a_i) | a_i \geq b\}$. Let K^o denote the set of bi-weighted formulas thus obtained. K^o is called the paraconsistent completion of K.
We call *paraconsistency degree* of a bi-weighted formula (φ, b, c) the value $\min(b, c)$. In particular, the formulas of interest are such that $b \geq c$, i.e. the formula is at least as certain as it is paraconsistent. In particular, formulas such as $c = 0$ are safe from any inconsistency in K. They are said to be *free*[8] in K.

Example 1 *Take*
$K = \{(\varphi, 0.8), (\neg\varphi \vee \psi, 0.6), (\neg\varphi, 0.5), (\neg\xi, 0.3), (\xi, 0.2), (\neg\xi \vee \psi, 0.1)\}$.
Note that $inc(K) = 0.5$.
Then, K^o *is the set of bi-weighted formulas:*
$\{(\varphi, 0.8, 0.5), (\neg\varphi, 0.5, 0.8), (\neg\xi, 0.3, 0.2), (\xi, 0.2, 0.3), (\neg\varphi\vee\psi, 0.6, 0), (\neg\xi\vee\psi, 0.6, 0)\}$.
Consider for instance $(\neg\xi \vee \psi, 0.6, 0)$*. From* $(\varphi, 0.8)$ *and* $(\neg\varphi \vee \psi, 0.6)$ *we infer* $(\psi, 0.6)$ *(by modus ponens), which implies* $(\neg\xi\vee\psi, 0.6, 0)$ *(by logical weakening); note that in this case this inference only uses formulas above the level of inconsistency (0.5). Besides, there is no way to derive* $\neg\psi$ *(nor* $\varphi \wedge \neg\psi$ *consequently) from any consistent subset of* K^**; so* $c = 0$ *for* $\neg\xi \vee \psi$*.* □

Remark 1 *One may think of extending the paraconsistent completion* $\mathcal{L}^o$ *to the whole language* $\mathcal{L}$ *of* K*, in the spirit of the proposal made by Arieli[3] in the "flat" case (where the only certainty degrees are 1 and 0):* $\forall\phi \in \mathcal{L}$*:*

- $\phi \in \mathcal{L}_T$ *if and only if there is a consistent subset of* K *that entails* ϕ *and none that entails* $\neg\phi$*; and we can write* $(\phi, 1, 0) \in \mathcal{L}^o$*.*

- *$\phi \in \mathcal{L}_F$ if and only if there is a consistent subset of K that entails $\neg\phi$ and none that entails ϕ; and we can write $(\phi, 0, 1) \in \mathcal{L}^o$.*
- *$\phi \in \mathcal{L}_U$ if and only if there is no consistent subset of K that entails ϕ nor any that entails $\neg\phi$; and we can write $(\phi, 0, 0) \in \mathcal{L}^o$.*
- *$\phi \in \mathcal{L}_I$ if and only if there is a consistent subset of K that entails ϕ and another one that entails $\neg\phi$; and we can write $(\phi, 1, 1) \in \mathcal{L}^o$.*

In the above definition one can restrict to maximal consistent subbases of K. These four sets of formulas $\mathcal{L}_T, \mathcal{L}_F, \mathcal{L}_U, \mathcal{L}_I$ partition the language. It can be checked that $K^o \subset \mathcal{L}^o$. One can view the four annotations by pairs of Boolean values as akin to Belnap[7] epistemic truth-values, TRUE, FALSE, NONE *and* BOTH *respectively. However, Belnap logic comes down to computing epistemic statuses of atomic propositions based on information from various sources, then obtaining the epistemic status of other formulas via truth-tables extending the usual ones to four values. See Dubois[12] and Dubois and Prade[25] for further discussions.*

Clearly the formulas of the form $(\varphi, b, 0)$ in K^o have an inconsistency level equal to 0, and thus lead to safe conclusions. However, one may obtain a set of *consistent* conclusions from K^o, which is larger than the one that can be obtained from $K^{cons} \cup K_{free}$ (where K^{cons} denotes the set of formulas strictly above the inconsistency level, and K_{free} the set of free formulas), as explained now.

Defining an inference relation from K^o requires two evaluations:
- the *undefeasibility* degree of a consistent set A of formulas:

$$UD(A) = \min\{b \mid (\varphi, b, c) \in K^o \text{ and } \varphi \in A\}$$

- the *unsafeness* degree of a consistent set A of formulas:

$$US(A) = \max\{c|(\varphi, b, c) \in K^o \text{ and } \varphi \in A\}$$

We say that A is a reason for ψ if A is a minimal (for set inclusion) consistent subset of K that implies ψ, i.e.,

- $A \subseteq K$
- $A^* \nvdash_{CL} \bot$
- $A^* \vdash_{CL} \psi$
- $\forall B \subset A, B^* \nvdash_{CL} \psi$

Then, let

$$UD(\phi) = \max\{UD(A) : A \text{ is a reason for } \phi\};$$

$$US(\phi) = \min\{US(A) : A \text{ is a reason for } \phi, UD(A) = UD(\phi)\}.$$

The set of triples $(A, UD(A), US(A))$ such that A is a reason for ψ is denoted by $label(\psi)$. Then, $(\psi, UD(\phi), US(\phi))$ is said to be a DS-consequence of K^o (or K), denoted by $K^o \vdash_{DS} (\psi, UD(\phi), US(\phi))$, if and only if $UD(\phi) > US(\phi)$ [8]. It can be shown that $\vdash_{DS}$ extends the entailment in possibilistic logic.

Example 2 *(Example 1 continued): In the above example,* $label(\psi) = \{(A, 0.6, 0.5), (B, 0.2, 0.3)\}$ *with* $A = \{(\varphi, 0.8, 0.5), (\neg\varphi \vee \psi, 0.6, 0)\}$ *and* $B = \{(\xi, 0.2, 0.3), (\neg\xi \vee \psi, 0.6, 0)\}$*. Then,* $K^o \vdash_{DS} (\psi, 0.6, 0.5)$*.*

If we first minimize $US(A)$ *and then maximize* $UD(A')$*, the entailment would not extend the possibilistic entailment. Indeed in the above example, we would select* $(B, 0.2, 0.3)$ *but* $0.2 > 0.3$ *does not hold, while* $K \vdash (\psi, 0.6)$ *since* $0.6 > inc(K) = 0.5$*. Note that* $\vdash_{DS}$ *is more productive than the possibilistic entailment, as seen on the example, e.g.,* $K^o \vdash_{DS} (\neg\xi, 0.3, 0.2)$*, while* $K \vdash (\neg\xi, 0.3)$ *does not hold since* $0.3 < inc(K) = 0.5$*.*

An entailment denoted by $\vdash_{SS}$, named *safely supported*-consequence relation, less demanding than $\vdash_{DS}$, is defined by $K^o \vdash_{SS} \psi$ if and only $\exists A \in label(\psi)$ such that $UD(A) > US(A)$. It can be shown that the set $\{\psi \mid K^o \vdash_{SS} \psi\}$ is classically consistent[8].

This kind of inference can be also understood in terms of minimal inconsistent subsets[26]. Let S be a minimal inconsistent subset in K^*, and let $inc(S) = \min\{a_j \mid (p_j, a_j) \in K, p_j \in S\}$ be the level of inconsistency of S. Then, observe that

$$inc(K) = \max\{a \mid K \vdash (\bot, a)\} = \max_{S, \text{minimal inconsistent subset of } K} inc(S),$$

Moreover, it turns out that if $(p_i, \pi_i, \gamma_i) \in K^o$, we have

$$\gamma_i = \max_{k\ : p_i \in C_k, C_k \text{minimal inconsistent subset of } K} inc(C_k)$$

with $inc(C_k) = \min\{a_j \mid (p_j, a_j) \in K, p_j \in C_k\}$.

In fact, we have the following result: The safely supported entailment from K coincides with the possibilistic entailment from the *consistent* possibilistic logic base

$K_{\max}^{cons}$ obtained from K by deleting, in all minimal inconsistent subsets S of K, the formulas with a certainty level equal to $inc(S)$. Namely

$$K_{\max}^{cons} = K \setminus \{(p_i, a_i) \in S : S \text{ minimal inconsistent subset of } K, a_i = inc(S)\}.$$

and we have

$$K \vdash_{SS} \phi \iff (K_{\max}^{cons})^* \vdash_{CL} \phi.$$

3.2 From quasi-classical logic to quasi-possibilistic logic

Besnard and Hunter[10, 29] have defined a kind of paraconsistent logic, called quasi-classical logic. This logic has several nice features, in particular the connectives behave classically, and when the knowledge base is classically consistent, then quasi-classical logic gives almost the same conclusions as classical logic.[2] Moreover, the inference in quasi-classical logic has a low computational complexity.

The basic ideas behind this logic is to use all rules of classical logic proof theory, but to forbid the use of resolution after the introduction of a disjunction (it allows us to get rid of the ex falso quodlibet sequitur). So the rules of quasi-classical logic are split into two classes: composition and decomposition rules, and the proofs cannot use decomposition rules once a composition rule has been used. Intuitively speaking, this means that we may have resolution-based proofs both for φ and $\neg\varphi$. We also have as additional valid consequences the disjunctions build from the previous consequences (e.g. $\neg\varphi \vee \psi$). But it is forbidden to reuse such additional consequences for building further proofs[29].

It is clear that while possibilistic logic takes advantage of its weights for handling inconsistency, there are situations where possibilistic logic offers no useful answers, while quasi-classical logic does. This is when formulas involved in inconsistency have the same weight, especially the highest one, 1. For instance, consider the example $K = \{(\varphi, 1), (\neg\varphi \vee \psi, 1), (\neg\varphi, 1)\}$, where quasi-classical logic infers φ, $\neg\varphi$, ψ from K^*, while everything is drowned in possibilistic logic, and nothing is obtained by the safely supported-consequence relation. This has led to propose a quasi-possibilistic logic[14] which has still to be further developed.

It would also have to be related to the simple generalized inference rule, applicable to formulas in K^o,

$$(\neg\varphi \vee \psi, b, c)(\psi \vee \xi, b', c') \vdash (\psi \vee \xi, \min(b, b'), \max(c, c')),$$

proposed by Dubois *et al.*[15]. Note that in the above example, we would obtain $(\varphi, 1, 1)$, $(\neg\varphi, 1, 1)$ and $(\psi, 1, 1)$, as expected, by applying this rule. This rule can be

[2] In fact only tautologies or formulas containing tautologies cannot be recovered.

viewed as the counterpart of the fact that in approximate reasoning the combination / projection principle provides, as consequences, fuzzy subsets whose height is the minimum of the heights of the fuzzy sets involved in the inference.

4 Inconsistency Handling in Multiple Source Information

In multiple source possibilistic logic[17], each formula is associated with a set (a fuzzy set more generally) which gathers the labels of sources according to which the formula is (more or less) certainly true. This leads to a simple extension of possibilistic logic, where propositions are associated not only with certainty levels, but also with the corresponding sources.

Consider, for instance, the following multi-source knowledge base where the information comes from sources s_1, s_2, s_3.

Example 3 $K = \{(\neg\varphi \vee \psi, \{1/s_1, 1/s_2\}), (\neg\varphi \vee \xi, \{0.7/s_1, 0.2/s_2\}),$
$(\neg\psi \vee \xi, \{0.4/s_1, 0.8/s_2, 0.4/s_3\}), (\neg\varphi \vee \neg\xi, \{0.3/s_3\}),$
$(\varphi, \{0.5/s_1, 0.8/s_2, 0.5/s_3\}), (\psi, \{0.8/s_1, 0.9/s_2\}), (\xi, \{0.6/s_2\})\}.$

Then by resolution and combination applied for each source, we can compute the multi-source certainty attached to ξ, for example. We obtain $N(\xi) \supseteq \{0.5/s_1, 0.8/s_2\}$ (where $\supseteq$ denotes fuzzy set inclusion, i.e. it means $N_1(\xi) \geq 0.5$, $N_2(\xi) \geq 0.8$), where N_i is the ordinary necessity measure associated with source i, while N is now a fuzzy set-valued extended necessity measure[17]. We can also prove $N(\neg\xi) \supseteq \{0.3/s_3\}$, i.e. $N_3(\neg\xi) \geq 0.3$. Thus, the source s_3 is in conflict with $\{s_1, s_2\}$ with respect to ξ. But, by distinguishing between the sources, we avoid a global inconsistency problem. This idea can be further elaborated in connection with formal concept analysis[4] in order to associate subsets of sources to combination results obtainable from consistent subsets of pieces of information in an information merging process.

The idea of associating formulas with the sources that support them to some degree has been more systematically investigated in recent papers[21, 5], where formulas of the form $(\varphi, a/A)$ express that at least all agents in subset A believe that φ is true at least with certainty level a. Such formulas can be handled in a multi agent possibilistic logic where both the certainty levels a and the subsets A of agents are combined in the inference process. This enables us to distinguish between inconsistencies shared by some subsets of agents, and inconsistencies between beliefs held by disjoint subsets of agents.

5 Inconsistency with respect to Ignorance

Standard possibilistic logic handles constraints of the form $N(\varphi) \geq a$. Constraints of the form $\Pi(\varphi) \geq a$ can be also considered, although they represent poorer pieces of information. Indeed $N(\varphi) \geq a \Leftrightarrow \Pi(\neg\varphi) \leq 1-a$ expresses partial certainty about φ, hence partial impossibility of $\neg\varphi$, while $\Pi(\varphi) \geq a$ only expresses that φ true is somewhat possible. In particular, the state of (complete) ignorance about the truth value of φ can be represented by $\Pi(\varphi) = 1 = \Pi(\neg\varphi)$, which states that both φ and $\neg\varphi$ are fully possible.

Here appears another form of inconsistency between a statement of the form $N(\varphi) \geq a$ expressing that a proposition is somewhat certain, and a statement of the form $\Pi(\neg\varphi) \geq b$ (equivalently, $N(\varphi) < b$) expressing that the opposite proposition is somewhat possible, when the strict inequality $b > 1 - a$ holds between the degrees.

This situation is at work in the following cut rule[20], which mixes the two types of lower bound constraints on Π and N, namely

$$N(\neg\varphi \vee \psi) \geq a, \Pi(\varphi \vee \xi) \geq b \vdash \Pi(\psi \vee \xi) \geq a\&b$$

with $a\&b = 0$ if $a \leq 1 - b$ and $a\&b = b$ if $a > 1 - b$. This type of inconsistency is of a higher level. It is a statement not dealing with the real world (e.g. claiming that one is sure that something is and is not), but a statement about epistemic states of external agents (agent 1 having reasons to believe that agent 2 is sure of something, and having reasons to believe that agent 2 is ignorant about this thing). This kind of knowledge can be expressed in generalized possibilistic logic[27], since one handles negations and disjunctions of standard possibilistic formulas, which allows contradictions of the form $N(\varphi) \geq a$ and $a \geq b > N(\varphi)$.

6 Inconsistency in Bipolar Information

The representation capabilities of possibilistic logic can be also enlarged in the bipolar possibilistic setting[13, 9]. It allows the separate representation of both negative and positive information. Negative information reflects what is not (fully) impossible and remains potentially possible. It induces (prioritized) constraints on where the real world is (when expressing knowledge), which can be encoded by necessity-based possibilistic logic formulas. Positive information expressing what is actually possible, is encoded by another type of formula based on a set function called guaranteed (or actual) possibility measure (which is to be distinguished from "standard "possibility measures that rather express potential possibility (as a matter of consistency with the available information). This bipolar setting is of interest for representing knowledge and observations, and also for representing positive and negative preferences.

Positive information is represented by formulas denoted by $[\varphi, d]$, which expresses the constraint $\Delta(\varphi) \geq d$, where Δ denotes a measure of strong (actual) possibility[19] defined from a possibility distribution δ by $\Delta(\varphi) = min_{\omega \models \varphi}\delta(\omega)$. This contrasts with a measure of (weak) possibility Π which is *max*-decomposable, rather than *min*-decomposable (as Δ is) for disjunction.

Thus, the piece of positive information $[\varphi, d]$ expresses that any model of φ is at least possible with degree d.

Let $D = \{[\varphi_j, d_j] | j = 1,k\}$ be a positive possibilistic logic base. Its semantics is given by the possibility distribution

$$\delta_D(\omega) = \max_{j=1,k} \delta_{[\varphi_j, d_j]}(\omega)$$

with $\delta_{[\varphi_j, d_j]}(\omega) = 0$ if $\omega \models \neg\varphi_j$, and $\delta_{[\varphi_j, d_j]}(\omega) = d_j$ if $\omega \models \varphi_j$. Thus, δ_D is obtained as the max-based *disjunctive* combination of the representation of each formula in D. This is in agreement with the idea that observations accumulate and are never in conflict with each other. Such a situation was already encountered in Mamdani and Assilian's fuzzy controllers[31, 23], where a weighted *union* of the contributions of each fuzzy rule that is fired, is performed.

A positive possibilistic knowledge base $D = \{[\varphi_j, d_j] | j = 1, k\}$ is inconsistent with a negative possibilistic knowledge base $K = \{(\varphi_i, a_i) | i = 1, m\}$ as soon as the following fuzzy set inclusion is violated:

$$\forall\omega,\ \delta_D(\omega) \leq \pi_K(\omega).$$

This violation occurs when something is observed while one is somewhat certain that the opposite should be true. Such an inconsistency should be handled by giving priority either to the positive or to the negative information[34].

7 Concluding Remarks

This overview has outlined the different forms of inconsistency that are expressible in possibility theory, when representing different types of information in a logical format. It is important to notice that the inconsistency, more precisely the contradictions here (see [11] on this point) may take place between different graded modalities. First, the same source cannot be certain at a positive degree of both φ and $\neg\varphi$, i.e. contradictions between formulas is mirrored at the epistemic level in terms of necessity degrees. Two other forms of contradiction, either between asserted ignorance and certainty, or between what is reputed as being not possible and what is observed, involve two types of modalities. Contradictions can also take place in generalized possibilistic logic[27] at another level, since one handles negations and disjunctions of standard possibilistic formulas, as already seen above. Thus, the way inconsistency has to be managed depends not only from the application perspective (artificial intelligence inference systems vs. handling of dirty data in information systems[32, 30]), but also of the nature of the inconsistency.

The paper has also pointed out the filiation existing between Zadeh's approximate reasoning theory and possibilistic logic. Although the two settings highly rely on possibility theory, it is interesting to notice that they have been developed in different directions and to try to understand why. Approximate reasoning theory mainly exploits the notion of possibility distribution and anticipates the representation of reasoning problems in terms of constraints (soft, in this case) which is at the basis of the constraint satisfaction problems that started to be investigated in artificial intelligence a decade later. The purpose of approximate reasoning following Zadeh was to represent and reason with pieces of fuzzy knowledge expressed in natural language, and encoded by possibility distributions on proper universes. Approximate reasoning is also closely related to fuzzy rule-based systems, but not so much to Mamdani's approach, since in this latter work, information is no longer viewed as constraints to be combined conjunctively, but rather as clues to be combined disjunctively[28]. But this important difference has remained almost unnoticed for a long time. Thus, approximate reasoning theory was based on possibility distributions and to some extent on possibility measures, while the other set functions of possibility theory (necessity, guaranteed possibility) were absent.[3]

On its side, possibilistic logic, while keeping a semantics in terms of possibility distributions (now defined on a set of interpretations, rather than on the domain of a linguistic variable) is much closer to classical logic; it is based on necessity measures and makes an extensive use of the degree of (in)consistency, whose expression formally appears in the first paper on fuzzy sets under the form of a separation degree between fuzzy sets, while generalized possibilistic logic accommodates all the modalities expressed by the set functions of possibility theory. Possibilistic logic is usually restricted to classical logic formulas, although there exist extensions to fuzzy propositions such as the one developed by Alsina and Godo[1, 2]. As in any logical

[3]This was partially counterbalanced by the introduction of the sophisticated notion of *compatibility*[6, 38, 39] of a fuzzy set G with respect to a fuzzy set F, defined as the fuzzy set of the possible values of the membership degree to G of an element fuzzily restricted by F, which gives birth to the ideas of fuzzy truth values and fuzzy truth qualification. In fact, the compatibility both encompasses the consistency of F and G (or if we prefer the possibility of G given F), and the necessity of G given the fuzzy restriction expressed by F[33].

setting, inconsistency is a key notion in standard or generalized possibilistic logics, and may take various forms here due to the richness of the representation setting.

References

References

[1] T. Alsinet and L. Godo, "A complete calculus for possibilistic logic programming with fuzzy propositional variables", in *Proc. 16th Conf. on Uncertainty in Artificial Intelligence (UAI'00)*, Stanford, (Morgan Kaufmann, San Francisco, 2000), pp. 1–10.

[2] T. Alsinet, L. Godo, and S. Sandri "Two formalisms of extended possibilistic logic programming with context-dependent fuzzy unification: a comparative description", *Elec. Notes in Theor. Computer Sci.*, **66** (5) (2002) 1–21.

[3] O. Arieli, "Conflict-tolerant semantics for argumentation frameworks", *Proc. 13th Euro. Conf. Logics in Artificial Intelligence (JELIA12)*, eds. L. Fariñas del Cerro, A. Herzig and J. Mengin, Toulouse, Sept. 26-28, (Springer, LNCS 7519, 2012) pp. 28–40.

[4] Z. Assaghir, A. Napoli, M. Kaytoue, D. Dubois and H. Prade, "Numerical information fusion: Lattice of answers with supporting arguments", *Proc. 23rd Inter. Conf. on Tools with Artificial Intelligence (ICTAI'11)*, Boca Raton, Nov.7-9, (IEEE, 2011) pp. 621–628.

[5] A. Belhadi, D. Dubois, F. Khellaf-Haned and H. Prade, "Multiple agent possibilistic logic", *J. of Applied Non-Classical Logics*, **23** (4) (2013) 299–320.

[6] R. E. Bellman and L. A. Zadeh, "Local and fuzzy logics", in *Modern Uses of Multiple-Valued Logic*, eds. J. M. Dunn and G. Epstein (D. Reidel, Dordrecht, 1977) pp. 103–165.

[7] N. D. Belnap, "A useful four-valued logic", in *Modern Uses of Multiple-Valued Logic*, eds. J. M. Dunn and G. Epstein (D. Reidel, Dordrecht, 1977) pp. 7–37.

[8] S. Benferhat, D. Dubois, and H. Prade, "An overview of inconsistency-tolerant inferences in prioritized knowledge bases", in *Fuzzy Sets, Logic and Reasoning about Knowledge*, vol.15 in Applied Logic Series, (Kluwer, 1999) pp. 395–417.

[9] S. Benferhat, D. Dubois, S. Kaci and H. Prade, "Modeling positive and negative information in possibility theory", *Int. J. of Intelligent Systems* **23** (2008) 1094–1118.

[10] P. Besnard and A. Hunter, "Quasi-classical logic: Non-trivializable classical reasoning from inconsistent information", *Proc. of the 3rd European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (EC-SQARU95)*, LNAI 946, (Springer Verlag, 1995) pp. 44–51.

[11] W. A. Carnielli, M. E. Coniglio and J. Marcos, "Logics of formal inconsistency", *Handbook of Philosophical Logic, vol. 14, 2nd edition*, eds. D. Gabbay and F. Guenthner, (Springer, 2007) pp. 1-93.

[12] D. Dubois, "On ignorance and contradiction considered as truth-values", *Logic J. of the IGPL* **16** (2): (2008)195-216.

[13] D. Dubois, P. Hajek and H. Prade, "Knowledge-driven versus data-driven logics", *J. of Logic, Language, and Information* **9** (2000) 65–89.

[14] D. Dubois, S. Konieczny and H. Prade, "Quasi-possibilistic logic and its measures of information and conflict", *Fundamenta Informaticae* **57** (2-4) (2003) 101–125.

[15] D. Dubois, J. Lang and H. Prade, "Handling uncertainty, context, vague predicates, and partial inconsistency in possibilistic logic", in *Fuzzy Logic and Fuzzy Control (Proc. of the IJCAI'91 Workshop on Fuzzy Logic and Fuzzy Control, Aug. 1991)*, eds. D. Driankov, P. W. Eklund and A. L. Ralescu, LNCS 833, (Springer-Verlag, 1994) pp. 45–55.

[16] D. Dubois, J. Lang and H. Prade, "Fuzzy sets in approximate reasoning Part 2: Logical approaches", *Fuzzy Sets and Systems* **40** (1991) 203–244.

[17] D. Dubois, J. Lang and H. Prade, "Dealing with multi-source information in possibilistic logic", *Proc. of the 10th Europ. Conf. on Artificial Intelligence (ECAI'92)* Vienna, Aug. 3-7, (Wiley, New York, 1992) pp. 38–42.

[18] D. Dubois, J. Lang and H. Prade, "Possibilistic logic", *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3*, eds. D. M. Gabbay, C. J. Hogger, J. A. Robinson, D. Nute, (Oxford University Press, 1994) pp. 439-513.

[19] D. Dubois and H. Prade, "Possibility theory: qualitative and quantitative aspects", in *Quantified Representation of Uncertainty and Imprecision*, eds. D. Gabbay, P. Smets, eds., Handbook of Defeasible Reasoning and Uncertainty Management Systems, (Kluwer Acad. Publ., 1998) vol.1, pp. 169–226.

[20] D. Dubois and H. Prade, "Possibilistic logic: a retrospective and prospective view", *Fuzzy Sets and Systems* **144** (2004) 3–23.

[21] D. Dubois and H. Prade, "Toward multiple-agent extensions of possibilistic logic", *Proc. IEEE Inter. Conf. on Fuzzy Systems (FUZZ-IEEE 2007)* London, July 23-26, 2007, pp. 187–192.

[22] D. Dubois and H. Prade, "Handling various forms of inconsistency in possibilistic logic", *Proc. 1st Int. Workshop on Data, Logic and Inconsistency (DALI11), in Database and Expert Systems Applications (DEXA) International Workshops* (F. Morvan, A Min Tjoa, R. Wagner, eds.), Toulouse, Aug. 29 - Sept. 2, (IEEE Computer Society, 2011), pp. 327–331.

[23] D. Dubois and H. Prade, "Abe Mamdani: A pioneer of soft artificial intelligence", in *Combining Experimentation and Theory - A Hommage to Abe Mamdani*, eds. E. Trillas, P. P. Bonissone, L. Magdalena, and J. Kacprzyk, (Springer, Studies in Fuzziness and Soft Computing, vol. 271, 2012) pp. 49–60.

[24] D. Dubois and H. Prade, "Possibilistic logic. An overview", in *Handbook of The History of Logic. Vol. 9 Computational logic*, eds. D. M. Gabbay, J. H. Siekmann, J. Woods, (North-Holland, 2014) pp. 283–342.

[25] D. Dubois and H. Prade, "Being consistent about inconsistency: Toward the rational fusing of inconsistent propositional logic bases", in *The Road to Universal Logic. Festschrift for the 50th Birthday of Jean-Yves Béziau. Vol.II*, eds. A. Koslow, A. Buchsbaum, (Birkhäuser, 2015) pp. 565–571.

[26] D. Dubois and H. Prade, "A possibilistic analysis of inconsistency", *Proc. 9th Int. Conf. on Scalable Uncertainty Management (SUM'15)* (Ch. Beierle, A. Dekhtyar, eds.), Québec City, Sept. 16-18, (Springer, LNCS 9310, 2015) pp. 347–353.

[27] D. Dubois, H. Prade and S. Schockaert, "Stable models in generalized possibilistic logic", *Proc. 13th Int. Conf Principles of Knowledge Representation and Reasoning (KR'12)* eds. G. Brewka, Th. Eiter, S. A. McIlraith, Rome, June 10-14, (AAAI Press, 2012) pp. 519-529.

[28] D. Dubois, H. Prade, L. Ughetto, “Fuzzy logic, control engineering and artificial intelligence’, in *Fuzzy Algorithms for Control*, eds. H.B. Verbruggen, H.-J. Zimmermann and R. Babuska, (Kluwer Acad., 1999) pp. 17–57.

[29] A. Hunter, “Reasoning with conflicting information using quasi-classical logic”, *J. of Logic and Computation* **10** (2000) 677–703.

[30] S. Link and H. Prade , “Relational database schema design for uncertain data”, Res. Rep. 469, Centre for Discrete Math. and Theoretical Comp. Sci., Univ. of Auckland, August 2014.

[31] E. H. Mamdani and S. Assilian, “An experiment in linguistic synthesis with a fuzzy logic controller”, *Int. J. of Man-Machine Studies* **7** (1975) 1–13.

[32] O. Pivert and H. Prade, “Detecting Suspect Answers in the Presence of Inconsistent Information”, *Proc. 7th Int. Symp. on Foundations of Information and Knowledge Systems (FoIKS’12)*, eds. Th. Lukasiewicz and A. Sali, Kiel, March 5-9, (Springer, LNCS 7153, 2012) pp. 278–297.

[33] H. Prade, “A computational approach to approximate and plausible reasoning with applications to expert systems”, *IEEE Trans. Pattern Anal. Mach. Intell.* **7**(3) (1985) 260–283; “Corrections”, *IEEE Trans. Pattern Anal. Mach. Intell.* **7**(6) (1985) 747–748.

[34] H. Prade and M. Serrurier, “Bipolar version space learning”, *Int. J. Intell. Syst.* **23**(10) (2008) pp. 1135-1152.

[35] L. A. Zadeh, “Fuzzy sets”, *Information and Control* **8** (1965) 338–353.

[36] L. A. Zadeh, “Similarity relations and fuzzy orderings”, *Information Sciences* **3** (1971) 177–200.

[37] L. A. Zadeh, “Fuzzy sets as a basis for a theory of possibility”, *Fuzzy Sets and Systems* **1** (1978) 3–28.

[38] L. A. Zadeh, “PRUF - A meaning representation language for natural languages”, *Int. J. Man-Machine Studies* **10** (1978) 395–460.

[39] L. A. Zadeh, “A theory of approximate reasoning”, in *Machine Intelligence, vol. 9*, eds. J. E. Hayes, D. Mitchie, and L. I. Mikulich (Ellis Horwood, 1979) pp. 149–194.

Structures of Opposition in Fuzzy Rough Sets *

Davide Ciucci[1] Didier Dubois[2] and Henri Prade[2]
1. DISCo, Università di Milano – Bicocca
Viale Sarca 336/14, 20126 Milano, Italia
ciucci@disco.unimib.it
2. IRIT, Université Paul Sabatier
118 route de Narbonne, 31062 Toulouse cedex 9, France
dubois@irit.fr

December 15, 2015

Abstract The square of opposition is as old as logic. There has been a recent renewal of interest on this topic, due to the emergence of new structures (hexagonal and cubic) extending the square. They apply to a large variety of representation frameworks, all based on the notions of sets and relations. After a reminder about the structures of opposition, and an introduction to their gradual extensions (exemplified on fuzzy sets), the paper more particularly studies fuzzy rough sets and rough fuzzy sets in the setting of gradual structures of opposition.

Keywords square of opposition; fuzzy set; fuzzy relation; rough set.

1 Introduction

Fuzzy set theory [41, 42, 44, 45, 46] and rough set theory [29, 31, 34, 33, 32] are two important frameworks which have been introduced and developed in the second half of the previous century, and which proved to be very successful in information processing. They are both mathematically based on the notions of sets and relations, but are motivated by quite different concerns,

*To appear in *Fundamenta Informaticae*

although they can be (somewhat artificially) related [30], and the idea of granulation [43] can be encountered in both settings. While fuzzy set theory makes the notion of membership to a class gradual and softens equivalence relations into similarity relations, rough set theory bounds, from above and below, any subset of elements in terms of equivalence classes of indiscernible elements (having the same attribute values). Since their respective concerns are orthogonal rather than competing, it makes sense to consider different forms of hybridizations of the two theories, as pointed out quite early [17, 18]; see also [15].

Due to their mathematical nature based on sets and relations, the two theories have established connections with logic [26, 16, 14]. Thus it should not come as a surprise that they can be considered in the perspective of the square of opposition. The square of opposition is a representation of different forms of opposition arising among four logical statements. It has been introduced by Aristotle and then studied throughout the centuries, in particular by Middle-Age logicians. Then, it has been forgotten by modern logic, until its interest was rediscovered by Robert Blanché in relation with cognitive modeling concerns [8], in the second half of XXth century. In the last past years, it has raised again a lot of interest [3, 5, 6, 7] and it has been extended in several ways, generating new structures of opposition, which can be displayed on hexagons, or cubes, in particular. Two generic instantiations of the cube of opposition are in terms of intersections of sets and of compositions of relations respectively [21], which explains the universality of this structure in knowledge representation.

These structures can indeed be encountered in different fields including artificial intelligence-related areas [19, 1, 21]. In particular, oppositions in rough sets have been studied, which can be described in terms of approximations, relations, attributes [10, 40, 11]. Recently, a gradual extension of the square, of the hexagon and of the cube of oppositions has been proposed [20, 21]. So, it seems natural to apply these new structures to fuzzy rough sets and rough fuzzy sets [17, 18]. This is the purpose of this paper.

The paper is organized as follows. Sections 2 and 3 provide an introduction to structures of opposition, and their gradual extensions, then exemplified by the case of fuzzy sets. Section 4 studies oppositions in fuzzy rough sets and rough fuzzy sets.

2 Structures of opposition: The Boolean case

In this section we introduce the basic structures of opposition, and then their gradual extensions. For an overview of the square of opposition and generalized geometric representation of opposition we refer to [3, 4, 19].

2.1 Square, hexagon, and cube of opposition

The traditional square of opposition involves four related logical statements with different quantifiers and the classical negation operation $\neg$. Given a statement $p(x)$, the four corners read as $\mathbf{A} : \forall x\ p(x)$, $\mathbf{E} : \forall x\ \neg p(x)$, $\mathbf{I} : \exists x\ p(x)$, $\mathbf{O} : \exists x\ \neg p(x)$. Let us notice that we suppose the existence of some x such that $p(x)$ holds, for avoiding existential import problems. A usual graphical representation of the square is given in Figure 1.

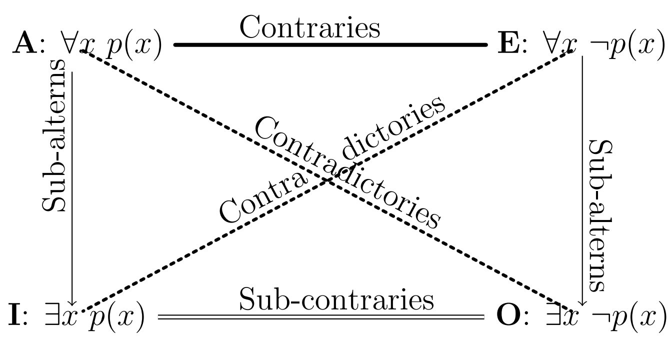

Figure 1: Square of opposition

Clearly, these four corners are not independent from each other. The links among them can be highlighted by interpreting **A**, **I**, **E**, and **O** as the truth values of the statements, that is as Boolean variables. So, we have (see, e.g., [27]):

(a) **A** and **O** are the negation of each other, as well as **E** and **I**. In a logical reading $\mathbf{A} \equiv \neg\mathbf{O}$ and $\mathbf{E} \equiv \neg\mathbf{I}$.

(b) **A** entails **I**, and **E** entails **O**, i.e., vertical arrows represent *implication* relations $\mathbf{A} \to \mathbf{I}$ and $\mathbf{E} \to \mathbf{O}$.

(c) **A** and **E** cannot be true together, but may be false together: $\neg\mathbf{A} \vee \neg\mathbf{E}$ should hold (they are in a *contrariety* relation).

(d) **I** and **O** cannot be false together, but may be true together: $\mathbf{I} \vee \mathbf{O}$ should hold (they are in a *subcontrariety* relation).

Moreover, the above conditions are not independent. Several links can be established among them, which have to be considered when generalizing the square to the gradual case:

(Dep 1) Conditions **(a)(b) imply condition (c).** That is, $\neg\mathbf{A} \vee \neg\mathbf{E}$ is a consequence of $\mathbf{A} \equiv \neg\mathbf{O}$ and $\mathbf{E} \to \mathbf{O}$ (or of $\mathbf{E} \equiv \neg\mathbf{I}$ and $\mathbf{A} \to \mathbf{I}$) in the square.

(Dep 2) **Conditions (a)(b) imply condition (d).** That is, $\mathbf{I} \vee \mathbf{O}$ is a consequence of $\mathbf{A} \equiv \neg\mathbf{O}$ and $\mathbf{A} \to \mathbf{I}$ (or of $\mathbf{E} \equiv \neg\mathbf{I}$ and $\mathbf{E} \to \mathbf{O}$).

(Dep 3) **Conditions (a)(c) imply conditions (b)(d)** and **conditions (a)(d) imply conditions (b)(c):** $\mathbf{A} \equiv \neg\mathbf{O}$, $\mathbf{E} \equiv \neg\mathbf{I}$, together with $\neg\mathbf{A} \vee \neg\mathbf{E}$ entail $\mathbf{A} \to \mathbf{I}$, $\mathbf{E} \to \mathbf{O}$ and $\mathbf{I} \vee \mathbf{O}$. Similarly, $\mathbf{A} \equiv \neg\mathbf{O}$, $\mathbf{E} \equiv \neg\mathbf{I}$, together with $\mathbf{I} \vee \mathbf{O}$ entail $\mathbf{A} \to \mathbf{I}$, $\mathbf{E} \to \mathbf{O}$ and $\neg\mathbf{A} \vee \neg\mathbf{E}$.

The hexagon of opposition [8, 4] is built on the square by considering the union of **A**, **I** obtaining **U**, and the conjunction of **E**, **O** obtaining **Y** (see Figure 2). It was then noticed that the six corners define three squares of opposition: the one we start with **AIEO**, but also **AYOU** and **EYIU**.

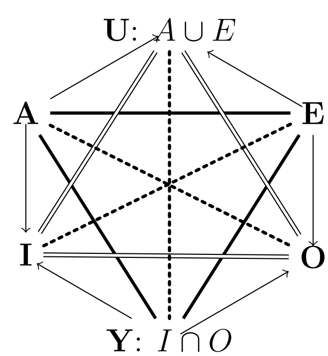

Figure 2: Hexagon of opposition

Besides, the square of opposition can be generalized to a cube of opposition. The cube of opposition has the four corners **A**, **I**, **E**, and **O** in the front facet and four other corners in the back, namely, **a**, **i**, **e**, and **o**. In Figure 3, a Boolean cube is represented with the statements corresponding to the new back corners. This cube was first introduced by Reichenbach [37] in a systematic discussion of syllogisms, and rediscovered in [19]. It is worth mentioning that the vertices of the diagonal squares **AaoO** and **EeiI** are related by a Klein group of transformations applied to logical statements, first identified by Piaget [35]. For instance, $R(\mathbf{A}) = C(N(\mathbf{A})) = \mathbf{a}$, $C(\mathbf{O}) = N(R(\mathbf{O})) = \mathbf{a}$, or $N(R(C(\mathbf{E}))) = N(R(\mathbf{i})) = N(\mathbf{I}) = \mathbf{E}$, where i) $I(\phi) = \phi$ (identity), $N(\phi) = \neg\phi$ (negation), $R(\phi) = f(\neg p, \neg q, ...)$ (reciprocation), and $C(\phi) = \neg f(\neg p, \neg q, ...)$ (correlation). It can be easily checked that $N = RC$, $R = NC$, $C = NR$, and $I = NRC$. See Figure 3 and [19].

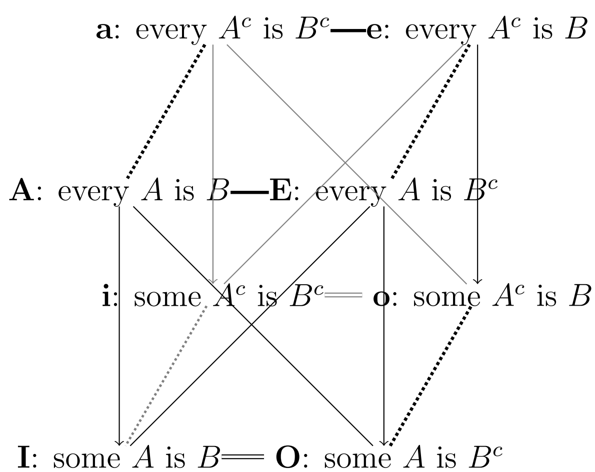

Figure 3: Cube of opposition

2.2 The Cube of Rough Sets

Several kinds of opposition structures can be defined in the rough set context: based on relations, approximations, or attributes [10, 40, 11]. Here we are interested in the cube based on upper and lower approximations. It is well known that a rough set is a pair of lower $L_R(A)$ and upper $U_R(A)$ approximations of a subset A of a set X defined according to a relation R and such that $L_R(A) \subseteq U_R(A)$. More precisely, given an approximation space (X, R)

with R a binary relation on R, the two approximations are defined as [38]:

$$L_R(A) = \{x \in X | xR \subseteq A\}$$
$$U_R(A) = \{x \in X | xR \cap A \neq \emptyset\}$$

where xR is the neighborhood of x with respect to R, that is $xR = \{y|xRy\}$. These two sets are at the basis of a square of oppositions: $L_R(A)$ is the corner **A** and $U_R(A)$ the corner **I**. The other two corners are obtained by complementation: $L^c_R(A)$ is corner **O** and $U^c_R(A)$ corner **E** (this last set is also known as the exterior of A). The usual interpretation attached to these sets is that the lower approximation contains the objects surely belonging to A, the exterior contains objects surely not belonging to A and the remaining objects form the boundary. So, once we extend the square into a hexagon, the top corner contains the totality of objects on which we are certain, namely, $L_R(A) \cup U^c_R(A)$ whereas the bottom one contains the objects on which we are totally undecided: $U_R(A) \setminus L_R(A)$.

Now, when moving to the cube, we distinguish two cases, depending on whether L and U are dual to each other, that is $L(A) = U^c(A^c)$, or not. In this last case, a cube can be defined using as back square the approximations applied to A^c: $L_R(A^c)$, $U_R(A^c)$, $L^c_R(A^c)$ and $U^c_R(A^c)$ are respectively the corners **a, i, o, e**. On the other hand, if the lower and upper approximations are dual, the front and back squares collapse.

However, another kind of cube can be defined by considering a so-called sufficiency operator:

$$[[A]]_R := \{x \in X | A \subseteq xR\}$$

and the dual operator: $\ll A \gg_R = \{x \in X | A \cup xR \neq \emptyset\}$. The whole cube arising from lower, upper and sufficiency approximation is drawn in Figure 4. Note that $[[A]]_R$ can be equivalently written as $L_{R^c}(A^c)$. In other words, the back facet of this cube is the same as the front facet where the relation R is replaced by its complement. However R has a lot of properties, usually, while its complement does not have them. So even if $L_{R^c}(A^c)$ is formally the lower approximation of A^c with respect to R^c, it often hardly stands as a genuine lower approximation.

If R is an equivalence relation with equivalence classes C_i, $i = 1, \dots p$, then $[[A]]_R := C_i$ if $A \subseteq C_i$, and $\emptyset$ otherwise, which indicates that this notion is not very fruitful in that case. If R is only symmetric and reflexive, then R can still be written as $\bigcup_{i=1,\dots,p} C_i \times C_i$, where C_i is maximal such that $C_i \times C_i \subseteq R$, but the C_i's may overlap. This amounts to saying that an undirected graph

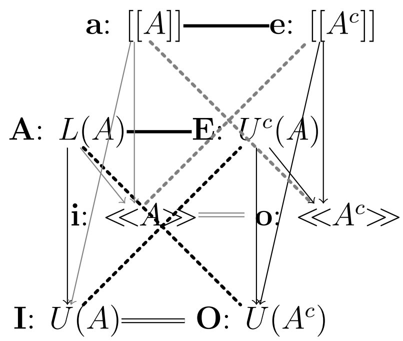

Figure 4: Cube of opposition induced by rough approximations

is the union of its maximal cliques. Then, $xR = \bigcup_{i:x\in C_i} C_i$ and $[[A]]_R = \bigcap_{x\in A}\bigcup_{i:x\in C_i} C_i$. Interestingly we may have that $[[A]]_R \cap A = \emptyset$. For instance, assume that $p = 2$, C_1 and C_2 overlap, and $A = (C_1 \setminus C_2) \cup (C_2 \setminus C_1)$, then $[[A]]_R = C_1 \cap C_2$. Note that $[[A]]_R$ contains all those elements related to all elements in A. So, $[[A]]_R$ can be viewed as all bridges that make all elements in A communicate.

Remark 1 *In modal logic, the standard necessity operator expresses the fact that a property is a necessary condition for some other properties to hold. Moreover, Kripke semantics is given through possible worlds and a binary relation R connecting them. In this standard environment, the idea of a sufficient condition has no place and, further, Kripke semantics cannot account for irreflexive relations. The sufficiency operator is introduced in order to overcome these deficiencies [25, 23]. These ideas were then borrowed by data analysis and the above operator $[[\cdot]]$ introduced as an approximation operator [28, 22].*

3 Gradual Structures of Opposition

In this section, we try to extend the Boolean structure of opposition such as the one in Figure 3 to the case where sets are fuzzy, so that statements appearing on the vertices are true to a degree between 0 and 1. As we shall see, one difficulty we shall meet is due to the fact that the strong link between entailment (relating vertices **A** and **I**, or yet **E** and **O**, for instance) and negation may be lost. In particular, we have that $p \to q \equiv \neg p \vee q$, and negation is $\neg p = p \to \perp$ (where $\perp$ denotes the contradiction) and is involutive. However, in the gradual setting the negation as defined above is generally not involutive. As the square of opposition heavily relies on the involutivity property[12], the design of gradual squares, hexagons and cubes of opposition becomes more tricky.

3.1 The gradual square

The gradual square of opposition associates a degree in $[0, 1]$ to each corner. Let us name the degree of corners **A**, **I**, **E**, and **O** respectively as $\alpha, \iota, \epsilon, o$. Then, we outline two possibilities for generalizing the square: the weak and the strong one. In order to do it, we need an involutive negation n, a commutative conjunction $*$, the dual disjunction $\oplus$ and an implication denoted by $s \Rightarrow t$.

The connectives we are going to consider are based on standard operations on $[0, 1]$:

- *Negations* n are unary functions such that $n(0) = 1$ and $n(1) = 0$. A negation is said *involutive* if $\forall x$, $n(n(x)) = x$.

- *Commutative conjunctions*, i.e., binary operations $* : [0, 1]^2 \mapsto [0, 1]$ such that $x * y = y * x$; $0 * x = 0$; $1 * x = x$. In particular, *triangular norms* (*t–norms*) $*$, are associative and monotonic commutative conjunctions. Given a conjunction $*$ and an involutive negation n, the dual disjunction is defined by De Morgan properties as $x \oplus y = n(n(x) * n(y))$. The dual of a t-norm is named *triangular conorm* (t–conorm).

- *Implications* $\to$, i.e. a binary function on $[0, 1]$ such that $1 \to 0 = 0$ and $1 \to 1 = 0 \to 1 = 0 \to 0 = 1$. It is said to be a *border implication* if $\forall x \in [0, 1]$, $1 \to x = x$. Particular border implications are the *residual* of a left-continuous t-norm, defined as $x \to_* y := \sup\{x \in [0, 1] : x * z \leq y\}$. Another important class is the one of *strong implications* (S-implications): given a conjunction $*$ and an involutive negation n, a strong implication is defined as $x \Rightarrow_S y := n(x * n(y)) = n(x) \oplus y$ where $\oplus$ is the dual of $*$.

The *strong form of the gradual square of opposition* requires that the above constraints (a)–(b) are encoded as follows:

(a) **A** and **O** are the negation of each other, as well as **E** and **I**: $\alpha = n(o)$ and $\epsilon = n(\iota)$

(b) The implication is assumed to be a strong one, i.e., $s \Rightarrow t = n(s * n(t)) = n(s) \oplus t$. Then, **A** entails **I**, and **E** entails **O** is modeled as $\alpha \Rightarrow \iota = 1$ and $\epsilon \Rightarrow o = 1$, i.e., $\alpha * n(\iota) = 0$ and $\epsilon * n(o) = 0$;

(c) **A** and **E** cannot be true together, but may be false together. It can be encoded by $\alpha * \epsilon = 0$ or equivalently $n(\alpha * \epsilon) = 1$;

(d) **I** and **O** cannot be false together, but may be true together. It can be encoded by $n(\iota) * n(o) = 0$ or equivalently $n(n(\iota) * n(o)) = 1$, i.e. $\iota \oplus o = 1$.

The *weak form* of the gradual square differs on condition (b), requiring only that $\alpha \leq \iota$ and $\epsilon \leq o$.

In case of the *strong form*, dependencies (Dep1)–(Dep3) still hold given the four conditions (a)–(d). On the other hand, this is not the case for the weak form, so further contraints have to be considered if we desire to have a complete faithful extension of the square to the gradual case. For instance, we can require the conjunction $*$ to be a nilpotent t-norm and n to be the standard involutive negation $n(x) = 1 - x$.

3.2 Gradual cube

In case of the gradual cube, degrees $\alpha', \iota', \epsilon', o'$ are also associated to corners **a**, **i**, **e**, and **o** of the cube, with the requirement to form a weak/strong gradual square of opposition. That is, the conditions on the front and back squares (strong form) are:

(a) $\alpha = n(o)$, $\epsilon = n(\iota)$ and $\alpha' = n(o')$, $\epsilon' = n(\iota')$;

(b) $\alpha * n(\iota) = 0$, $\epsilon * n(o) = 0$ and $\alpha' * n(\iota') = 0$, $\epsilon' * n(o') = 0$;

(c) $\alpha * \epsilon = 0$ and $\alpha' * \epsilon' = 0$;

(d) $n(\iota) * n(o) = 0$ and $n(\iota') * n(o') = 0$.

Moreover, we have some constraints on the side facets (these conditions derive from analogous ones holding in the Boolean cube, see [12]):

(e) $\alpha * n(\iota') = 0$, that is **A** entails **i**;

(f) $\alpha' * n(\iota) = 0$, **a** entails **I**;

(g) $\epsilon' * n(o) = 0$, **e** entails **O**;

(h) $\epsilon * n(o') = 0$, **E** entails **o**.

which are equivalent to the conditions that we have to require on the top and bottom facets:

(i) $\alpha' * \epsilon = 0$, which means that **a** and **E** cannot be true together;

(j) $\alpha * \epsilon' = 0$, **A** and **e** cannot be true together;

(k) $n(\iota') * n(o) = 0$, that is **i** and **O** cannot be false together;

(l) $n(\iota) * n(o') = 0$, **I** and **o** cannot be false together.

In case of the weak form of the square, while conditions (a), (c) and (d) are left unchanged, the conditions (b) become $\alpha \leq \iota$, $\epsilon \leq o$ and $\alpha' \leq \iota'$, $\epsilon' \leq o'$, whereas, the side (top/bottom) facets conditions read as:

(e') $\alpha \leq \iota'$;

(f') $\alpha' \leq \iota$;

(g') $\epsilon' \leq o$;

(h') $\epsilon \leq o'$.

3.3 Gradual hexagon

Finally, the gradual hexagon of opposition is built from the square by considering the union of **A**, **I** obtaining **U** with degree ν and the conjunction of **E**, **O** obtaining **Y** with degree γ. That is, we define $\nu = \alpha \oplus \epsilon$ and $\gamma = \iota * o$.

Since the six corners define three squares of opposition: the standard one **AIEO**, then **AYOU** and **EYIU**, we have to impose the conditions (a)–(d) on them. In the case of the hexagon, we are going to consider only the weak form of the square since the strong form would require that $*$ is a nilpotent t-norm, hence satisfying all the weak form constraints plus the dependency ones (Dip1)–(Dip3). So, the four constraints on the squares **AYOU** and **EYIU** imply the following:

(a) $\nu = n(\epsilon)$. This is true by definition of ν. Indeed, $\nu = \alpha \oplus \epsilon = n(\iota) \oplus n(o) = n(\iota * o) = n(\gamma)$.

(b) **A** entails **U** and **Y** entails **O**, that is $\alpha \leq \nu$ and $\gamma \leq o$. Again by definition, this means $\alpha \leq \alpha \bot \epsilon$ and $\iota * a \leq o$, which is true for any choice of monotonic conjunction $*$ and disjunction $\oplus$, and in particular for all triangular norms and triangular co-norms.
Similarly, we have to require that **Y** entails **I**, **E** entails **U**, i.e., $\gamma \leq \iota$ and $\epsilon \leq \nu$.

(c) $\alpha * \gamma = 0$ and $\epsilon * \gamma = 0$. This condition, generally, does not follow from the previous ones, so we should impose it.

(d) $n(o) * n(\nu) = 0$ and $n(i) * n(\nu) = 0$. This condition is equivalent to the previous one.

As discussed in [12], sufficient conditions for all these constraints to hold are that condition (c) hold and $*$,$\oplus$ are dual norm and co-norm or that $*$ is a nilpotent triangular norm, such as $\alpha * \beta = \max(0, \alpha + \beta - 1)$.

3.4 Example: The cube of fuzzy sets

Going back to the cube of Figure 3, the entailments of the top facet may be rewritten in terms of empty intersections of sets of objects **A**, B, and their complements A^c, B^c, while the bottom facets refer to non empty intersections, as pointed out in [21]. See Figure 5. Note that we assume $A \neq \emptyset$, $A^c \neq \emptyset$, $B \neq \emptyset$, and $B^c \neq \emptyset$ here, for avoiding the counterpart of the existential import problems, since now the sets **A** and B play symmetric roles in the statements associated to the vertices of the cube.

This cube extends to the case where **A** and B are normalized fuzzy subsets of X, e.g., $A : X \mapsto [0, 1]$. We denote degrees of membership by $A(x), B(x), \ldots$ Suppose we use the min-based and $1-(\cdot)$-based definitions of intersection and complementation respectively. Then $\iota = \sup_x \min(A(x), B(x))$, and $o = \sup_x \min(A(x), 1 - B(x))$; $\alpha = 1 - o$ and $\epsilon = 1 - \iota$. Then, it can be checked that $n(\iota) * n(o) = 0$, or equivalently $\iota \oplus o = 1$, namely,

$$\sup_x \min(A(x), B(x)) + \sup_x \min(A(x), 1 - B(x)) \geq B(x_0) + 1 - B(x_0) = 1,$$

where $A(x_0) = 1$ (normalization of **A**). From which it follows by duality that $\alpha * \epsilon = 0$, and we have $\alpha = \inf_x \max(1 - A(x), B(x)) \leq \iota =$

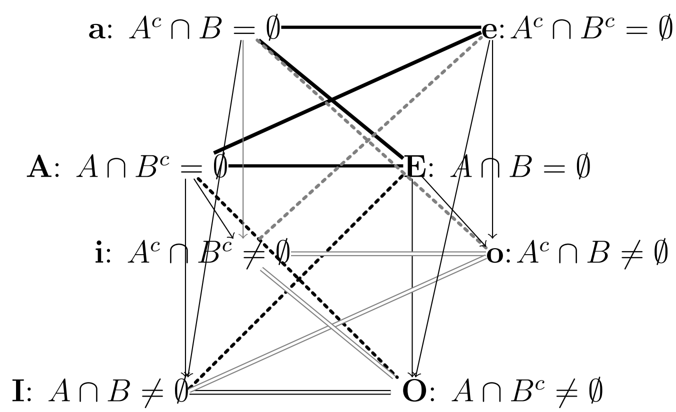

Figure 5: Cube of opposition of set intersection indicators

$\sup_x \min(A(x), B(x))$ if **A** is normalized. The other conditions of the cube can be checked as well (provided that A^c, B, B^c are also normalized).

4 Opposition in Fuzzy Rough Sets

In this section we first recall basic notions of fuzzy rough sets, i.e., approximations of fuzzy sets induced by a fuzzy relation.

4.1 Fuzzy rough sets

As basic definition of fuzzy rough set, we consider the one given in [36] generalized to any kind of fuzzy relation. At first, we need some definitions on fuzzy sets. Let X be the universe of investigation. A fuzzy binary relation is a mapping $R : X \times X \mapsto [0, 1]$ and R is said

serial	iff	$\forall x \in X, \exists y \in X : R(x, y) = 1$
reflexive	iff	$\forall x \in X : R(x, x) = 1$
symmetric	iff	$\forall x, y \in X : R(x, y) = R(y, x)$

Then, we can define fuzzy rough sets.

Definition 1 *[36] Given a t-norm $*$, a fuzzy binary relation R on a universe X, an implication $\rightarrow$, then the lower and upper approximations of a fuzzy* *set A are:*

$$L_R(A)(x) := \inf_{y \in X} \{R(x,y) \to A(y)\} \tag{1}$$

$$U_R(A)(x) := \sup_{y \in X} \{R(x,y) * A(y)\} \tag{2}$$

A fuzzy rough set is the pair $(L_R(A), U_R(A))$.

To qualify as a genuine rough set, this pair must obey some requirements

- $L_R(A) \subseteq U_R(A)$, a sufficient condition being that $R(x,y) \to A(y) \leq R(x,y) * A(y)$, for some y. If we assume R is serial, then take y s.t. $R(x,y) = 1$, which yields $1 \to A(y) \leq A(y)$, which holds if we use a border implication [9]. The stronger natural condition $L_R(A) \subseteq A \subseteq U_R(A)$ also requires a reflexive fuzzy relation.

- the duality condition $U_R(A) = n(L_R(n(A)))$ holds if there is an involutive negation n such that $n(\sup_{y \in X} R(x,y) * n(A(y))) = \inf_{y \in X} R(x,y) \to A(y)$, which means that the implication verifies $a \to b = n(a * n(b))$ so that $n(a) = a \to 0$. These conditions restrict the choice of the pair $(*, \to)$ (for instance Łukasiewicz conjunction and implication connectives, or yet minimum and Kleene-Dienes implication).

In the next subsection, we study the gradual square, cube and hexagon that the approximations in fuzzy rough sets originate.

4.2 Square from approximations

Given a fuzzy set A, its lower and upper approximations with their complement (with respect to an involutive negation n) can generate the standard square of opposition **A**: $L_R(A)$, **I**: $U_R(A)$, **E**: $n(U_R(A))$, **O**: $n(L_R(A))$, where $n(A)$ is the membership function of the complement of fuzzy set A.

Of course, conditions (a)–(d) have to be satisfied and they read as :

(a) $\alpha = n(o) \equiv L_R(A)(x) = n(n(L_R(A)(x))$ and $\epsilon = n(\iota) \equiv n(U_R(A)(x)) = n(U_R(A)(x))$.

(b) $\alpha * n(\iota) = L_R(A)(x) * n(U_R(A)(x) = 0$ in case of the strong form of the square
and $\alpha \leq \iota \equiv L_R(A)(x) \leq U_R(A)(x)$, $\epsilon \leq o \equiv n(U_R(A)(x)) \leq n(L_R(A)(x))$ in case of the weak form.

(c) $\alpha * \epsilon = 0 \equiv L_R(A)(x) * n(U_R(A)(x)) = 0.$

(d) $n(\iota) * n(o) = 0$

Proposition 1 *1. Condition (a) is always true whenever n is involutive.*

2. In case of the strong form, condition $\alpha * n(\iota) = L_R(A)(x) * n(U_R(A)(x)) = 0$ is sufficient to derive the other conditions.

3. In case of the weak form, a sufficient condition for (b) is to have R serial and $\rightarrow$ a border implication and n order reversing.

4. Condition (d) is an immediate consequence of (a) and (c).

Proof 1 *1. It follows by definition.*

2. By construction we have $\alpha = n(o)$ from which we can derive the other conditions (all dependencies Dep1–Dep3 hold in case of the strong form of the square).

3. From seriality of R and the fact that $\rightarrow$ is a border implication we get $L_R(A)(x) \leq U_R(A)(x)$ [9]. Then, if n is order reversing, we easily get $n(U_R(A)(x)) \leq n(L_R(A)(x))$.

We notice that the seriality of R is a standard condition in order to obtain a square of opposition by a relation [11, 12].

Condition (c) is usually neglected in fuzzy rough set approaches. However, it seems quite natural and important to require that the lower approximation and the exterior region ($n(U_R(A))$) are disjoint. Moreover, from point (2) of the above proposition, it plays an important role and it imposes some constraints on the definition of the fuzzy set A and the fuzzy relation R. For instance, it straightforwardly holds that:

Proposition 2 *A sufficient condition for $\alpha * n(\iota)$ to be zero is that either $L_R(A)(x)$ or $n(U_R(A)(x))$ are equal to zero. That is:*

$$\forall x \exists y : \quad R(x,y) \rightarrow A(y) = 0 \quad \textit{or} \quad R(x,y) * A(y) = 1$$

However, these conditions are seldom applicable since they rarely occur. For instance, the second one, due to the properties of the t-norm, comes down to requiring that $R(x,y) = A(y) = 1$, while $R(x,y)$ and $A(y)$ are independent quantities. In the general case, it is not so obvious to impose some constraints on the t-norm $*$ and the implication $\to$ to make $\alpha * n(\iota) = 0$ hold for all possible values x. Further investigations both in theory and on case studies are needed in this direction.

A similar square of opposition can be obtained with other kinds of fuzzy rough approximations, for instance, the *loose* and *tight* ones defined as follows [13].

Definition 2 *Let R be a fuzzy binary relation on X and f a fuzzy set on X. The* tight *approximation of f is defined as*

$$\forall y \in X \quad L_t(A)(y) = inf_{z\in X}\{R_z(y) \to inf_{x\in X}\{R_z(x) \to A(x)\}\}$$
$$\forall y \in X \quad U_t(A)(y) = sup_{z\in X}\{R_z(y) * sup_{x\in X}\{R_z(x) * A(x)\}\}$$

The loose *approximation of f is defined as*

$$\forall y \in X \quad L_l(A)(y) = sup_{z\in X}\{R_z(y) * inf_{x\in X}\{R_z(x) \to A(x)\}\}$$
$$\forall y \in X \quad U_l(A)(y) = inf_{z\in X}\{R_z(y) \to sup_{x\in X}\{R_z(x) * A(x)\}\}$$

Assuming that R is a similarity relation, i.e., it is reflexive and symmetric, we can prove the following relationship with the standard lower and upper approximations:

$$\begin{array}{lll} \text{(loose)} & L_l(A) = U_R(L_R(A)) & U_l(A) = U_R(U_R(A)) \\ \text{(tight)} & L_t(A) = L_R(L_R(A)) & U_t(A) = L_R(U_R(A)) \end{array}$$

Hence, due to the monotonicity of L_R it easily follows that, provided R is reflexive and symmetric, $L_l(A) \subseteq U_l(A)$ and $L_t(A) \subseteq U_t(A)$. So, both the tight and loose approximations, together with their complement with respect to an order reversing negation, can build a weak form of gradual square of opposition. Of course, the further constraints $L_l(A) * n(U_l(A)) = 0$ and $L_t(A) * n(U_t(A)) = 0$ have to be satisfied.

4.3 The gradual cube of approximations

Extending the square of fuzzy rough approximations to a cube leads to several possibilities to explore. As in the Boolean setting, the front and back of the cube coincide in case of dual approximations. If the lower and upper approximations are not dual (that is $L(A) \neq n(U(n(A)))$), we can define a

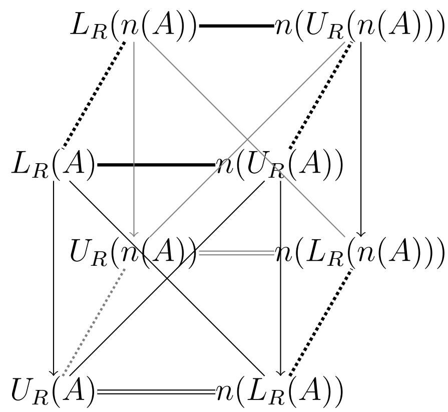

Figure 6: Cube of opposition from non-dual fuzzy rough sets

cube considering the approximations applied to the complement of A. This possibility has been also discussed in [10] with respect to Boolean rough sets (see Section 2.2).

The consideration on conditions (a)–(d) on the back square are the same as before, since it is the same square of the front applied to a different set. If we wish to respect also the side and top conditions, once a order reversing negation is considered, they reduce (weak form) to the following two: $L_R(A) \subseteq U_R(n(A))$ and $L_R(n(A)) \subseteq U_R(A)$. It is not an easy task to give general conditions under which these two conditions hold together. Indeed, we can give examples that do not satisfy them even if in presence of crisp relations with strong properties.

Example 1 *Let R be an equivalence relation on two objects x, y such that it always assumes the value 1 and define the set $A(x) = A(y) = 0.6$. Then, $L_R(A)(x) = L_R(A)(y) = 0.6 \geq U(n(A))(x) = U(n(A))(y) = 0.4$. On the other hand, considering the same relation and the set $B(x) = B(y) = 0.4$, we get $L_R(n(B))(x) = L_R(n(B))(y) = 0.6 \geq U(A)(x) = U(A)(y) = 0.4$. However note that in this example A and B are not normal, which may create existential import problems.*

Finally, if we consider the Klein group of the four Piaget transformations already mentioned, namely: identity $I(\phi) = \phi$; negation $N(\phi) = \neg\phi$; reciprocation $R(\phi) = f(\neg p, \neg q, \ldots)$ and correlation $C(\phi) = \neg f(\neg p, \neg q, \ldots)$, and if we consider the two squares (visualized in Figure 6) obtained from the diagonals of the cube, i.e., those with vertices $(L_R(A), L_R(n(A)), n(L_R(A)), n(L_R(n(A)))$

and with vertices $(U_R(A), U_R(n(A)), n(U_R(A)),\ n(U_R(n(A)))$, we see that these vertices are still exchanged by this Klein group as in the Boolean case, provided that n is involutive.

4.4 Cube from approximations and sufficiency operator

Whether upper and lower approximations are dual or not, another kind of cube can be defined as an extension of the cube of relations defined in [11]. In this case, the back square is built starting from a *sufficiency* operator and its dual. So, we have to introduce a new kind of "approximation" in fuzzy rough sets (as well as its dual) based on a fuzzy sufficiency operator.

Definition 3 *Let R be a fuzzy relation and A a fuzzy set, the fuzzy set of* bridge points *of A and its dual, are respectively defined as:*

$$[[A]]_R(x) := \inf_y \{A(y) \to R(x, y)\} \tag{3}$$

$$\ll A \gg_R(x) := n[[n(A)]] \tag{4}$$

The set $[[A]]_R$ corresponds to the corner (**a**) of the cube, whereas $\ll A \gg_R$ to (**i**). The value $[[A]](x)$ can be interpreted as the degree to which x is related to the set A. If $[[A]](x) = 1$ then $A \subseteq xR$. More precisely, $[[A]]_R(x)$ may be understood as the extent to which x is connecting all elements in A, since it estimates if any y in A is (highly) related to x in the sense of R. In other words, to what extent any element in A can communicate through x. In $[[A]]_R(x)$, the implication is reversed with respect to $L_R(A)(x)$. In case we take the conjunction of both, namely, $L_R(A)(x) \wedge [[A]]_R(x)$, we get an estimate that may represent how much x is R-similar to A, namely $A \sim_R x$. Note that $\sim_R$ is not transitive, but serial. On the other hand, $\ll A \gg_R(x)$ is the degree of non-relationship of x with elements in $n(A)$. The fact that $\ll A \gg_R(x) = 0$ can be interpreted as x is in relation with all the elements in $n(A)$. However, another option for defining $\ll A \gg_R(x)$ in the spirit of Equation (2), such as $\ll A \gg = \sup_y n(A(u)) * n(R(x, y))$, might be worth investigating.

Conditions (a)–(d) on the back square read as:

(a) $[[A]]_R = n(\ll n(A) \gg_R)$ and $[[n(A)]]_R = (\ll A \gg_R)^c$ hold by definition;

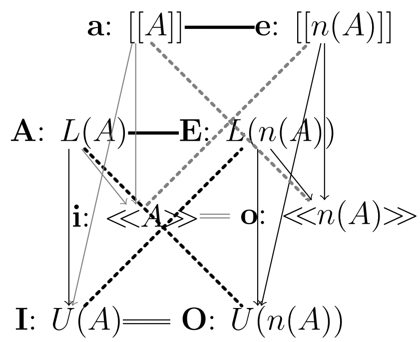

Figure 7: Cube of opposition induced by fuzzy-rough sets and sufficiency operator

(b) The two conditions $\alpha' \leq \iota'$ and $\epsilon' \leq o'$ are equivalent and require that the sufficiency operator implies its dual: $[[A]]_R \subseteq \ll A \gg_R$. A standard requirement in the analogous Boolean case is to ask for seriality of the fuzzy relation $n(R)$. In this case, it means to require that for all x there exists an element y such that $R(x, y) = 0$. So the condition is satisfied, if $n(\neg(n(A))) \subseteq n(A)$ where $\neg$ is the negation obtained by the implication used to define the sufficiency operator $[[\cdot]]$. For instance, this holds for a residual implication induced by a t-norm without non-trivial zero divisors [24] and any involutive negation n, indeed in this case it holds $\neg x \leq n(x)$.

(c) $[[A]]_R * [[n(A)]]_R = 0$. Similarly to the front square, it is not easy to give general conditions for this constraint to hold.

(d) due to duality of $\ll n(A) \gg_R$ and $[[A]]_R$, it is the same as condition (c).

Now, let us consider conditions on side/bottom facets. They read as

- $L_R(A) \subseteq \ll n(A) \gg_R$. With a similar reasoning as in point (b) above, a sufficient condition for this constraint to hold is to have at least one element such that $A(y) = 0$ and $\rightarrow$ to be a residual implication induced by a t-norm without non-trivial zero divisors.

- $[[A]] \subseteq U_R(A)$. In this case, it is sufficient to have a border implication and a normalized fuzzy set, i.e., there should exist a value y such that $A(y) = 1$.

4.5 Hexagon

As discussed in section 3 a hexagon is defined considering the conjunction of $\mathbf{A}, \mathbf{E}$ and the disjunction of $\mathbf{I}, \mathbf{O}$. In this case, they read as

$$\begin{array}{ll} (\mathbf{U}) & L_R(A) \oplus n(U_R(A)) \\ (\mathbf{Y}) & U_R(A) * n(L_R(A)) \end{array}$$

We stress that it is not necessary that $*$ and $\oplus$ are the operators used to define the approximations L and U.

By analogy with the Boolean case [10], $(\mathbf{U})$ represents what we know with certainty on A, indeed $L_R(A)$ are the elements surely belonging to A (to a certain degree in this fuzzy case) whereas $n(U_R(A))$ are those surely not belonging to A. On the other hand, $(\mathbf{Y})$ represents the total uncertainty region, i.e., the elements belonging to the possibility region $U_R(A)$ but not to the certainty one $L_R(A)$, that is, to the boundary.

Following section 3, in order to get a hexagon, besides the conditions on the square, one of the two conditions is sufficient

- $*$ is a nilpotent t-norm;
- $*, \oplus$ are dual t-norm and t-conorm and conditions $\alpha * \gamma = \epsilon * \gamma = 0$ hold. In this case, they read as $L_R(A) * U_R(A) * n(L_R(A)) = 0$ and $n(U_R(A)) * U_R(A) * n(L_R(A)) = 0$.

4.6 Special cases

Up to now, we have considered an extension of classical rough sets using a fuzzy relation and a fuzzy set (see Definition 1). We now investigate what happens when either the relation or the set are fuzzy and the other is crisp.

4.6.1 Crisp set and fuzzy relation

Let A be a subset of the universe X and R a fuzzy relation on X. In order to approximate A given the knowledge expressed by R, a first and immediate solution is to apply the same definitions of fuzzy rough sets to a crisp set.

So, equations in Definition 1 become now:

$$L_R(A)(x) := \begin{cases} 1 & A = X \\ \inf_{y\in A^c}\{\neg R(x,y)\} & A \neq X \end{cases} \tag{5}$$

$$U_R(A)(x) := \begin{cases} 0 & A = \emptyset \\ \sup_{y\in A}\{R(x,y)\} & A \neq \emptyset \end{cases} \tag{6}$$

where $\neg$ is the negation operator induced by the implication. Being a particular case of fuzzy rough sets, all the considerations formulated in the previous section, apply also here. Moreover, some constraints are here always (or more often) satisfied. At first let us notice that the following result holds by definition of L_R and U_R:

Lemma 1 *If R is*

- Serial *then for all x either $L_R(A)(x) = 0$ or $U_R(A)(x) = 1$;*
- Reflexive *then for all $x \notin A$ we have $L_R(A)(x) = 0$ and for all $x \in A$, $U_R(A)(x) = 1$.*

So, considering that R should be serial, we have that

Proposition 3 *Condition (c) of the square holds: for all x, $L_R(A)(x) * n(U_R(A)(x)) = 0$.*

In case of the cube of opposition and non-dual approximations[1], for the condition on the side and bottom faces we can state that

Proposition 4 *If R is reflexive then* $\inf_{y\in A^c}\{\neg R(x,y)\} \leq \sup_{y\in A^c}\{R(x,y)\}$.

That is, reflexivity of R is a sufficient condition to make the condition on side and bottom face holds.

Now, in this context, the sufficiency operator becomes:

$$[[A]]_R(x) := \begin{cases} 1 & A = \emptyset \\ \inf_{y\in A}\{R(x,y)\} & A \neq \emptyset \end{cases}$$

So, also for the sufficiency operator and its dual, we have that

[1]Let us notice that it can happen more frequently than in the general case that the approximations are dual, that is: $L_R(A) = nU_R(nA)$. For instance if $\neg x = n(x) = 1 - x$.

Proposition 5 *The seriality condition on $n(R)$ implies the side condition $[[A]]_R \subseteq \ll A \gg_R$. Moreover, if A is not empty also the other side condition $[[A]]_R \subseteq U_R(A)$ holds.*

Proof 2 *By definition of the sufficiency operator and due to the fact that A is Boolean, either $[[A]]_R(x) = 0$ or $\ll A \gg_R (x) = 1$, and then trivially $[[A]]_R \subseteq \ll A \gg_R$. Also, $[[A]]_R \subseteq U_R(A)$ follows easily by definition.*

We remark that, in general, we suppose that A is not empty since, in the classical square, the existence of some x such that $p(x)$ holds is assumed (see Section 3).

Finally, in case of the hexagon, we see that

Proposition 6 *If the relation R is reflexive, then conditions $\alpha * \gamma = 0$ and $\epsilon * \gamma = 0$ hold.*

So, in order to have a full realization of the hexagon, it is sufficient to consider a reflexive relation and a pair of dual t-norm and t-conorm.

Another possible way to define approximations in case of a crisp set and a fuzzy relation is to consider α-cuts of the fuzzy relation [39]. That is, given R we consider the family of relations

$$R_\alpha(x, y) = \begin{cases} 0 & R(x, y) < \alpha \\ 1 & R(x, y) \geq \alpha \end{cases}$$

In this case, we obtain a family of classical approximation spaces and if R is a fuzzy equivalence (max-min transitive) relation, R_α are all (Boolean) equivalence relations [39]. So, for each R_α, we can compute the standard approximations and obtain a family of classical square/cube/hexagon of opposition [10].

4.7 Fuzzy set and crisp equivalence relation

In case of R crisp and A fuzzy, *rough fuzzy sets* [17, 18] can be defined as

$$L_R(A)(x) = \inf\{A(y)|y \in [x]_R\}$$
$$U_R(A)(x) = \sup\{A(y)|y \in [x]_R\}$$

where $[x]_R$ is the equivalence class of x relatively to the relation R. It can be easily seen that these two equations are special cases of equations in

Definition 1 whenever R can assume only values $0,1$ and we use a border implication. So, the general conditions of Section 4 apply also here and for some constraints can be simplified as follows.

In case of the square, we have that $L_R(A)(x) \leq U_R(A)(x)$ is always true. Then, when extending the square to the cube we have that L and U are dual operators: given an involutive negation n, then $L_R(A) = n(U(n(A)))$. So, we can only consider the cube built from the sufficiency operator, which reads as: $[[A]]_R(x) = \inf\{\neg A(y) | y \notin [x]_R\} = L_{R^c}(\neg A)(x)$. Conditions on the back and side square do not simplify further in this case with respect to what described in section 4.4. The same can be said in the case of the hexagon.

As in the previous case, another approach is to use α-cuts, in this case to build a family of sets approximating the given fuzzy set [39]. Let A be a membership function of a fuzzy set, then for any α-cut A_α of A we can define its classical approximations $(L(A_\alpha), U(A_\alpha))$ and so obtain a family of structures of opposition, based on (Boolean) rough sets [10].

5 Conclusion

Opposition structures are a powerful tool to express all properties of rough sets and fuzzy rough sets with respect to negation in a synthetic way. After having studied the structure of opposition in Boolean rough sets [10] and extended the notion of square, cube and hexagon of opposition to the graded case [20, 12], we studied here the geometric representation of oppositions in the setting of fuzzy rough sets, that is when the basic elements of the approximations, the relation and the sets, are fuzzy. As particular cases also the situation where either the relation or the sets are crisp have been investigated. In all these situations we describe how to obtain at first a square of opposition and then extended structures, such as the cube and the hexagon. This study has stressed the importance of the relation between the inner and exterior regions of a set: they should be disjoint, a constraint always neglected. As an open problem, we leave it to the future a deeper study on the conditions on the fuzzy relation R and on the operations $*, \to$ to obtain the satisfaction of this constraint. We also introduced the sufficiency operator (and its dual) in fuzzy rough sets. The usefulness of this new operator in applications is yet to be explored. Finally, results in this study extend beyond the field of fuzzy rough sets and could be useful in fuzzy formal concept analysis [2].

References

[1] Amgoud, L., Prade, H.: A formal concept view of abstract argumentation, *Proc. 12th Eur. Conf. Symb. and Quant. Appr. to Reas. with Uncert. (ECSQARU'13), Utrecht, July 8-10* (L. C. van der Gaag, Ed.), LNCS 7958, Springer, 2013.

[2] Belohlavek, R.: *Fuzzy Relational Systems: Foundations and Principles*, Kluwer Academic Publishers, 2002.

[3] Béziau, J.-Y.: New light on the square of oppositions and its nameless corner, *Logical Investigations*, **10**, 2003, 218–233.

[4] Béziau, J.-Y.: The power of the hexagon, *Logica Universalis*, **6**(1-2), 2012, 1–43.

[5] Béziau, J.-Y., Gan-Krzywoszyńska, K.: Handbook of abstracts of the 2nd World Congress on the Square of Opposition, Corte, Corsica, June 17-20, 2010.

[6] Béziau, J.-Y., Gan-Krzywoszyńska, K.: Handbook of abstracts of the 3rd World Congress on the Square of Opposition, Beirut, Lebanon, June 26-30, 2010.

[7] Béziau, J.-Y., Gan-Krzywoszyńska, K.: Handbook of abstracts of the 4th World Congress on the Square of Opposition, Roma, Vatican, May 5-9, 2014.

[8] Blanché, R.: *Structures Intellectuelles. Essai sur l'Organisation Systématique des Concepts*, Vrin, Paris, 1966.

[9] Ciucci, D.: Approximation algebra and framework, *Fundamenta Informaticae*, **94**(2), 2009, 147–161.

[10] Ciucci, D., Dubois, D., Prade, H.: Oppositions in rough set theory, *RSKT Proceedings*, LNCS – 7414, 2012.

[11] Ciucci, D., Dubois, D., Prade, H.: The structure of oppositions in rough set theory and formal concept analysis - Toward a new bridge between the two settings, *Proc. 8th Int. Symp. on Foundations of Information and Knowledge Systems (FoIKS'14), Bordeaux, Mar. 3-7* (C. Beierle, C. Meghini, Eds.), 8367, Springer, 2014.

[12] Ciucci, D., Dubois, D., Prade, H.: Structure of opposition induced by relations, *Annals of Mathematics and Artificial Intelligence*, 2015, DOI:10.1007/s10472-015-9480-8.

[13] De Cock, M., Cornelis, C., Kerre, E.: Fuzzy rough sets: The forgotten step, *IEEE Transactions on Fuzzy Systems*, **15**, 2007, 121–130.

[14] Demri, S. P., Orlowska, E.: *Incomplete Information: Structure, Inference, Complexity*, Springer, 2002.

[15] Dubois, D., Esteva, F., Godo, L., Prade, H.: An information-based discussion of vagueness: six scenarios leading to vagueness, in: *Handbook of Categorization in Cognitive Science* (H. Cohen, C. Lefebvre, Eds.), chapter 40, Elsevier, 2005, 891–909.

[16] Dubois, D., Esteva, F., Godo, L., Prade, H.: Fuzzy-set based logics - An history-oriented presentation of their main developments, in: *Handbook of the History of Logic, Vol. 8, The Many-Valued and Nonmonotonic Turn in Logic* (D. M. Gabbay, J. Woods, Eds.), Elsevier, 2007, 325–449.

[17] Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets, *Int. J. of General Systems*, **17**(2-3), 1990, 191–209.

[18] Dubois, D., Prade, H.: Putting rough sets and fuzzy sets together, in: *Intelligent Decision Support - Handbook of Applications and Advances of the Rough Sets Theory* (R. Slowinski, Ed.), Kluwer Acad. Publ., 1992, 203–232.

[19] Dubois, D., Prade, H.: From Blanché's hexagonal organization of concepts to formal concept analysis and possibility theory, *Logica Univers.*, **6**, 2012, 149–169.

[20] Dubois, D., Prade, H.: Gradual structures of oppositions, in: *Enric Trillas: Passion for Fuzzy Sets* (F. Esteva, L. Magdalena, J. L. Verdegay, Eds.), Studies in Fuzziness and Soft Computing, Springer, 2015, 79–91.

[21] Dubois, D., Prade, H., Rico, A.: The cube of opposition: A structure underlying many knowledge representation formalisms, in: *Proc. 24th Int. Joint Conf. on Artificial Intelligence (IJCAI'15), Buenos Aires, July 25-31* (Q. Yang, M. Wooldridge, Eds.), AAAI Press, 2015, 2933–2939.

[22] Düntsch, I., Gediga, G.: Modal-style operators in qualitative data analysis, *Proc. IEEE Int. Conf. on Data Mining*, 2002.

[23] Düntsch, I., Orlowska, E.: Mixing modal and sufficiency operators, *Bulletin of the Section of Logic*, **28**, 1999, 99–107.

[24] Esteva, F., Godo, L., Hájek, P., Navara, M.: Residuated Fuzzy logics with an involutive negation, *Archive for Mathematical Logic*, **39**, 2000, 103–124.

[25] Gargov, G., Passy, S., Tinchev, T.: Modal environment for Boolean speculations, in: *Mathematical Logic and Applications* (D. Skordev, Ed.), Springer, 1987, 253–263.

[26] Hájek, P.: *Metamathematics of Fuzzy Logic*, vol. 4 of *Trends in Logic*, Kluwer Acad. Publ., 1998.

[27] Miclet, L., Prade, H.: Analogical proportions and square of oppositions, in: *Proc. 15th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, July 15-19, Montpellier* (A. Laurent *et al.*, Ed.), vol. 443 of *CCIS*, Springer, 2014, 324–334.

[28] Orlowska, E.: Introduction: What You Always Wanted to Know about Rough Sets, in: *Incomplete Information: Rough Set Analysis* (E. Orlowska, Ed.), Physica–Verlag, 1998, 1–20.

[29] Pawlak, Z.: Rough Sets, *Int. J. of Computer and Infor. Sci.*, **11**, 1982, 341–356.

[30] Pawlak, Z.: Rough sets and fuzzy sets, *Fuzzy Sets and Systems*, **17**, 1985, 99–102.

[31] Pawlak, Z.: *Rough Sets. Theoretical Aspects of Reasoning about Data*, Kluwer, 1991.

[32] Pawlak, Z., Skowron, A.: Rough sets and Boolean reasoning, *Information Sciences*, **177**, 2007, 41–73.

[33] Pawlak, Z., Skowron, A.: Rough sets: Some extensions, *Information Sciences*, **177**, 2007, 28–40.

[34] Pawlak, Z., Skowron, A.: Rudiments of rough sets, *Information Sciences*, **177**, 2007, 3–27.

[35] Piaget, J.: *Traité de Logique. Essai de logistique opératoire*, Armand Colin, Paris, 1949.

[36] Radzikowska, A. M., Kerre, E.: A comparative study of fuzzy rough sets, *Fuzzy Sets and Systems*, **126**, 2002, 137–155.

[37] Reichenbach, H.: The syllogism revised, *Philosophy of Science*, **19**(1), 1952, 1–16.

[38] Yao, J., Ciucci, D., Zhang, Y.: Generalized Rough Sets, in: *Handbook of Computational Intelligence*, Springer, 2015, 413–424.

[39] Yao, Y.: Combination of Rough and Fuzzy Sets based on α-level sets, in: *Rough Sets and Data Mining: Analysis of Imprecise Data*, Kluwer Academic Press, Boston, 1997, 301–321.

[40] Yao, Y. Y.: Duality in rough set theory based on the square of opposition, *Fundamenta Informaticae*, **127**, 2013, 49–64.

[41] Zadeh, L. A.: Fuzzy sets, *Information and Control*, **8**, 1965, 338–353.

[42] Zadeh, L. A.: Similarity relations and fuzzy orderings, *Information Sciences*, **3**, 1971, 177–200.

[43] Zadeh, L. A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, **90**, 1971, 111–127.

[44] Zadeh, L. A.: Theory of fuzzy sets, in: *Encyclopedia of Computer Science and Technology* (J. Belzer, A. Holzmann, A. Kent, Eds.), Marcel Dekker, New York, 1977.

[45] Zadeh, L. A.: Fuzzy logic, *Computer*, **21**(4), 1988, 83–93.

[46] Zadeh, L. A.: Fuzzy logic, in: *Encyclopedia of Complexity and Systems Science* (R. A. Meyers, Ed.), Springer, 2009, 3985–4009.

Bridging Gaps Between Several Forms of Granular Computing *

Didier Dubois, Henri Prade
IRIT, CNRS & Université de Toulouse

December 15, 2015

Abstract

Two important ideas at the core of Zadeh's seminal contributions to fuzzy logic and approximate reasoning are the notions of granulation and that of possibilistic uncertainty. In this paper, elaborating on the basis of some formal analogy, recently laid bare by the authors, between possibility theory and formal concept analysis, we suggest other bridges between theories for which the concept of granulation is central. We highlight the common features between the notion of extensional fuzzy set with respect to a similarity relation and the notion of formal concept. We also discuss the case of fuzzy rough sets. Thus, we point out some fruitful cross-fertilizations between the possibilistic representation of information and several views of granulation emphasizing the idea of clusters of points that can be identified respectively on the basis of their closeness, or of their common labeling in terms of properties.

Key-words possibility theory, formal concept analysis, extensional fuzzy set, rough set, granulation.

1 Introduction

The issue of how to describe items is at the basis of any representation framework and naturally involves notions of similarity and uncertainty. Similarity

*To appear in Vol. 1, issue 1 of the new journal *Granular Computing*, Springer.

is instrumental for grouping items having close or common features on the one hand. On the other hand, there is a need for coping with the fact that information may be incomplete or not precise enough, which is a source of uncertainty. In the non-Boolean setting these notions can be couched in the setting of fuzzy sets [50]. It has been already emphasized in [18] that fuzzy set membership functions can be interpreted diversely, in terms of similarity [2, 51], uncertainty [52, 54] and even preferences [3]. These different views have generally led to distinct families of important developments in data analysis and learning, in approximate reasoning, and in decision making, respectively.

The idea of granulation is at the heart of any knowledge representation system, as it points out that mathematical universes of discourse must be partitioned in agreement with the limitations of human perception. Generally we work with more or less well-defined partitions of idealized measurement scales; for instance the real line is too refined for human limited perception of closeness. Zadeh [53] has emphasized the importance of granulation and granular computing and the need to cast them in a non-Boolean setting, introducing the idea of a fuzzy granules: indeed indistinguishability between two quantities gradually takes place when they get closer to each other, so that the threshold under which they become indistinguishable is fuzzy. Moreover, he makes it clear that uncertainty due to granular descriptions is possibilistic rather than probabilistic, generally.

This discussion paper intends to illustrate the idea that some links can be established at the theoretical level between different concerns related to granular computing, on the basis of formal analogies that can be laid bare between the corresponding formal settings. In the following, we successively consider four settings: possibility theory [52, 14], formal concept analysis (FCA) [29], extensional fuzzy sets [33] and rough sets [41].

The first one, possibility theory, aims at providing a representation setting for epistemic uncertainty where partial ignorance can be encoded, and where a distinction can be made between what is somewhat certain and what is just possible to some extent. Possibility theory uses maximum and minimum operations rather than addition and product like probability theory and involves 4 set functions according to whether, for each event, one focuses on the maximum possibility value reflecting the event or its opposite, or yet the minimum possibility value.

The other three settings, sometimes apparently very different and developed completely independently, are concerned with the ideas of grouping items either because they can be gathered under the umbrella of the same formal concept, or because they are geometrically close enough to constitute fuzzy singletons, or yet because they share the same description in a database. The connection between extensional fuzzy sets and FCA was already discussed by Bělohlávek [5], and the connection between extensional fuzzy sets and fuzzy rough sets was noticed by Boixader et al. [7] (see also the monograph by Recassens [45]).

Here, we first illustrate the interest of the parallel between possibility theory and formal concept analysis that we initiated in [24] and further developed in [10, 23]. We recall the links between FCA and the formalism of rough sets in the special case of equivalence relations. Then, we indicate that this worth-noticing parallel carries over to the theory of extensional fuzzy sets and fuzzy rough sets, relying on previous technical studies. Interestingly enough, such links echo concerns often expressed by Zadeh in the last decade about the need for developing the ideas of granulation and granular computing in the setting of fuzzy sets [53]. The aim of this position paper is to encourage cooperation between schools of research that handle similar notions in various fields around the idea of granular computing.

The paper is organized as follows. Section 2 considers possibility theory and formal concept analysis in the crisp case. It shows that the four set-functions naturally associated with the possibility theory setting have counterparts in the formal concept analysis framework. The benefit of introducing more operators in the latter theory is exemplified by recalling a connection, not considered in standard formal concept analysis, which granulates a formal context into independent formal sub-contexts. Finally, the bridge with rough sets is obtained by restricting to relations between objects, and it shows the formal analogy between concepts and clusters, independent subcontexts and granules. Then, Section 3 considers the non-Boolean case. It recalls the theory of extensional fuzzy sets and the representation of fuzzy extensions of equivalence relations. Then it parallels two views of granulation, namely the one at work in formal concept analysis and the one underlying the theory of extensional fuzzy sets. Finally, we bring fuzzy rough sets into the picture.

2 From Formal Concept Analysis to Possibility Theory

Formal concept analysis associates any considered object with the set of its properties, via a formal context modelled by a binary relation R, a subset of the Cartesian product of the set of objects $\mathcal{O}$ and the set of properties $\mathcal{P}$.

An object is denoted by x, or x_i in case we consider several ones at the same time. It is interesting to notice that in fact, an object may either refer to a particular, unique item, or to a generic item representative of a class of items *sharing the same description*. A subset of objects will be denoted by a capital letter X, and we shall write $X = \{x_1, \ldots, x_i, \ldots, x_m\}$. A set of objects associated with their respective sets of properties defines a formal context $R \subseteq \mathcal{O} \times \mathcal{P}$ [29]. An object x is associated with its description, denoted by $\partial(x)$. In the following, we only consider simple descriptions, expressible in terms of a subset Y of properties y_j, namely, $Y = \{y_1, \ldots, y_j, \ldots, y_n\}$. Let $R(x) = \{y \in \mathcal{P} | (x, y) \in R\}$ be the set of properties of object x, and $R^{-1}(y) = \{x \in \mathcal{O} | (x, y) \in R\}$ is the set of objects having property y. In such a case, we shall write $\partial(x) = R(x) = Y$.

The classical setting of formal concept analysis defined from a formal context relies on a single operator $R^{\triangle}$ that associates a subset of objects with the set of properties they share.

$$R^{\triangle}(X) = \{y \in \mathcal{P} | R^{-1}(y) \supseteq X\} = \cap_{x \in X} R(x). \tag{1}$$

$R^{\triangle}(X)$ is a partial conceptual characterization of objects in X. Objects in X have all properties in $R^{\triangle}(X)$, but they may have some others (that are not shared by all objects in X). Conversely, $R^{-1\triangle}(Y) = \{x \in \mathcal{O} | R(x) \supseteq Y\} = \cap_{y \in Y} R^{-1}(y)$ is the set X of objects having all properties in Y.

In the setting of FCA, a formal concept [29] is defined as a pair $(X, Y) \in \mathcal{O} \times \mathcal{P}$ such that

$$R^{\triangle}(X) = Y \text{ and } R^{-1\triangle}(Y) = X. \tag{2}$$

In this case Y is also the maximal set of properties shared by all objects in X. It forms a Galois connection, and we have:

Proposition 1 *[29]. The following properties of pairs (X, Y) are equivalent*

1. $R^{\triangle}(X) = Y$ *and* $R^{-1\triangle}(Y) = X$
2. (X, Y) *is maximal such that* $X \times Y \subseteq R$

A formal concept (X, Y) is thus a maximal sub-rectangle in the formal context R. Let R_* be the set-union of all formal concepts extracted from R. Then $R_* = R$, by construction.

2.1 Describing imprecise objects using possibility distributions

In contrast with formal contexts, a useful kind of structured description of objects is in terms of attributes [40]. Let a, and $A = \{a_1, \ldots, a_k, \ldots, a_r\}$, respectively denote an attribute, and a set of attributes. The value of attribute a for x is denoted by $a(x) = u$, where u belongs to the attribute domain U_a. In this case, we shall write $\partial(x) = (a_1(x), \ldots, a_k(x), \ldots, a_r(x)) = (u_1, \ldots, u_k, \ldots, u_r)$. This corresponds to a completely informed situation where all the considered attribute values are known for x. When this is not the case, the precise value $a_k(x)$ will be replaced by the possibility distribution $\pi_{a_k(x)}$. Such a possibility distribution [52] is a mapping from U_{a_k} to $[0, 1]$, or more generally any linearly ordered scale. Then $\pi_{a_k(x)}(u) \in [0, 1]$ estimates to what extent it is possible that the value of a_k for x is u. 0 means impossibility; several distinct values may be fully possible (i.e. at degree 1). The characteristic function of an ordinary subset is a particular case of a possibility distribution. Precise information corresponds to the characteristic function of singletons.

An elementary property y can then be viewed as a subset A_y of a single attribute domain U_a, i.e. $y \subseteq U_a$. Note that while Y is a *conjunctive* set of properties (for instance an object possesses all properties in Y), property y, is a *disjunctive* set A_y of mutually exclusive values, one of which is *the* value of a single-valued attribute that is ill-known for some object x).

Four set functions in possibility theory are now recalled [19], emphasizing the symmetrical roles played by the object x and the attribute value u, a point of view unusual in possibility theory, but echoing the symmetrical role played by objects and properties in formal concept analysis. See [21] for a more complete introduction to the use of the four set functions in possibility theory.

2.2 Set-Functions in Possibility Theory

Let $\pi_{a(x)}(u)$ denote the possibility that object x has value $u \in U$ according to attribute a. For simplicity, we only consider the single-valued attribute case here (the actual value of x is not a set). The function $\pi_{a(\cdot)}(\cdot)$ defines a fuzzy set over $\mathcal{O} \times U$ (objects vs.attribute domain). We assume that π_a is bi-normalized: $\forall x \in \mathcal{O}, \exists u \in U, \pi_{a(x)}(u) = 1$ and $\forall u \in U, \exists x \in \mathcal{O},\ \pi_{a(x)}(u) = 1$. This means that for any object x, there is some fully possible value for attribute a, and that for any value u there is an object x that takes this value. Let X be a set of objects, and $y \subseteq U$ be a property. Then, one can define four set-functions, each defined in two domains, respectively the set of objects and the attribute domain:

1. Possibility measures [52], denoted by Π:

$$\begin{aligned} \Pi_u(X) &= \max_{x \in X} \pi_{a(x)}(u) \\ \Pi_x(y) &= \max_{u \in y} \pi_{a(x)}(u). \end{aligned}$$

   $\Pi_u(X)$ estimates to what extent it is possible that there is an object in X having value u, while $\Pi_x(y)$ is the possibility that object x has property y. Function Π is an indicator of non-empty intersection of the fuzzy set, whose membership function is the possibility distribution, with an ordinary subset. They are measures of "*potential possibility*". Clearly, Π is max-decomposable with respect to set union.

2. the dual measures of necessity N (or "actual necessity") [12]:

$$\begin{aligned} N_u(X) &= \min_{x \notin X} 1 - \pi_{a(x)}(u) \\ N_x(y) &= \min_{u \notin y} 1 - \pi_{a(x)}(u). \end{aligned}$$

   $N_u(X)$ estimates to what extent it is certain (necessarily true) that all objects that have value u lie in X, while $N_x(y)$ is the certainty that object x has property y. Note that $N_x(y) = 1 - \Pi_x(\overline{y})$ where $\overline{y} = U \setminus y$. Function N may be viewed as an indicator of inclusion of the fuzzy set whose membership function is the possibility distribution into an ordinary subset. And N is min-decomposable with respect to set intersection.

3. the measures of "actual (or guaranteed) possibility" [16]

$$\begin{aligned} \Delta_u(X) &= \min_{x \in X} \pi_{a(x)}(u) \\ \Delta_x(y) &= \min_{u \in y} \pi_{a(x)}(u) \end{aligned}$$

$\Delta_u(X)$ estimates to what extent it is possible that *all* objects in X have value u, while $\Delta_x(y)$ estimates the possibility that object x may take any value in y. Δ may be viewed as a degree of inclusion of an ordinary subset into the fuzzy set whose membership function is the possibility distribution. Δ is min-decomposable with respect to set union.

4. the dual measures of "potential necessity or certainty" [16]

$$\begin{aligned}\nabla_u(X) &= 1 - \min_{x \notin X} \pi_{a(x)}(u) \\ \nabla_x(y) &= 1 - \min_{u \notin y} \pi_{a(x)}(u)\end{aligned}$$

$\nabla_u(X)$ estimates to what extent there exists at least one object outside X that has a low degree of possibility of having value u, while $\nabla_x(y)$ is the degree to which there is an impossible value for $a(x)$ outside y. Note that $\nabla_x(y) = 1 - \Delta_x(\overline{y})$. ∇ is an indicator of non-full coverage of the considered universe by the fuzzy set whose membership function is the possibility distribution together with an ordinary subset. ∇ is max-decomposable with respect to set intersection.

2.3 Application to the Formal Context Setting

In [24], the setting of formal concept analysis has been enlarged with the introduction of three other operators. We now recall these four operators. They are counterpart, in the setting of a formal context, of the above set-functions from possibility theory.

Namely, let R be the formal context (a Boolean table). Then knowing only that an object x has some property y, the set $R^{-1}(y) = \{x \in \mathcal{O} | (x, y) \in R\}$ is the set of the *possible* objects corresponding to the elementary piece of knowledge "the object has property y" (in the context R). This suggests a possibilistic reading of formal concept analysis: a formal counterpart of possibility theory set-functions can be laid bare in this framework.Then, four remarkable sets can be associated with a subset X of objects (the notations have been chosen here in order to emphasize the parallel with possibility theory) [24, 22]:

- the set $R^{\Pi}(X)$ of properties that are possessed by *at least one* object in X:
$$R^{\Pi}(X) = \{y \in \mathcal{P} | R^{-1}(y) \cap X \neq \emptyset\} = \cup_{x \in X} R(x).$$

Clearly, we have $R^{\Pi}(X_1 \cup X_2) = R^{\Pi}(X_1) \cup R^{\Pi}(X_2)$.

- the set $R^N(X)$ of properties s. t. any object that satisfies *one* of them is necessarily in X:

$$R^N(X) = \{y \in \mathcal{P} | R^{-1}(y) \subseteq X\} = \cap_{x \notin X} \overline{R(x)},$$

  where the overbar dnotes complementation. In other words, possessing *any* property in $R^N(X)$ is a sufficient condition for belonging to X. Moreover, we have $R^N(X_1 \cap X_2) = R^N(X_1) \cap R^N(X_2)$ and $R^N(X) = \overline{R^{\Pi}\overline{X})} = \mathcal{P} \setminus R^{\Pi}(\overline{X})$.

- the set $R^{\triangle}(X)$ of properties shared by *all* objects in X:

$$R^{\triangle}(X) = \{y \in \mathcal{P} | R^{-1}(y) \supseteq X\} = \cap_{x \in X} R(x).$$

  In other words, satisfying *all* properties in $R^{\triangle}(X)$ is a necessary condition for an object to belong to X. Clearly, $R^{\triangle}(X_1 \cup X_2) = R^{\triangle}(X_1) \cap R^{\triangle}(X_2)$.

- the set $R^{\nabla}(X)$ of properties that are not satisfied by at least one object in $\overline{X}$.

$$R^{\nabla}(X) = \{y \in \mathcal{P} | R^{-1}(y) \cup X \neq \mathcal{O}\} = \cup_{x \notin X} \overline{R(x)}.$$

  Note that $R^{\nabla}(X) = \overline{R^{\triangle}(\overline{X})} = \mathcal{P} \setminus R^{\triangle}(\overline{X})$. In other words, in context R, for any property in $R^{\nabla}(X)$, there exists at least one object outside X that misses it. Moreover, we have $R^{\nabla}(X_1 \cap X_2) = R^{\nabla}(X_1) \cup R^{\nabla}(X_2)$.

A number of remarks are worth noticing:

- In negative similarity to $R^{\triangle}(X)$, $\overline{R^{\Pi}(X)}$ provides a negative conceptual characterization of objects in X since it gathers all the properties that are never satisfied by any object in X.

- $R^N(X) \cap R^{\triangle}(X)$ is the set of properties possessed by all objects in X and only by them.

- $R^{\Pi}(X)$ and $R^N(X)$ are isotonic (they become larger when X increases), while $R^{\triangle}(X)$ and $R^{\nabla}(X)$ are antitonic (they become smaller when X increases).

The four subsets $R^{\Pi}(X)$, $R^{N}(X)$, $R^{\triangle}(X)$, and $R^{\triangledown}(X)$ have been considered (with different notations) without any mention of possibility theory by different authors. The standard operator in FCA is $R^{\triangle}$. Düntsch *et al.* [25, 26] calls $R^{\triangle}$ a *sufficiency* operator, and its representation capabilities are studied in the theory of Boolean algebras. Taking inspiration as the previous authors from rough sets [41], Yao [48, 49] also considers these four subsets. In both cases, the four operators were introduced. See also [43, 31]. The interest of the bridge between possibility theory and FCA is that it enables a systematic investigation of alternative connections between objects and properties to be carried out; they differ from the standard Galois connection of FCA.

2.4 Application to Formal Context Decomposition

It can be checked that $R^{\triangledown}$ defines the same Galois connection as the one defined from $R^{\triangle}$, while R^{N} (or equivalently R^{Π}) induces another kind of connection, which is now described.

The connection defined from R^{N} proceeds in a similar formal way as when defining formal concepts [22, 10]. Namely, let us consider pairs (X, Y) s.t. $R^{N}(X) = Y$ and $R^{-1N}(Y) = X$. We can show these pairs also satisfy $R^{\Pi}(X) = Y$ and $R^{-1\Pi}(Y) = X$. Moreover, the pairs (X, Y) s.t. $R^{N}(X) = Y$ and $R^{-1N}(Y) = X$ allow us to characterize independent sub-contexts (i.e. that have no common objects and no common properties), and are thus of interest for the decomposition of a formal context into smaller independent ones. These results are expressed through the following:

Proposition 2 *[23]. The following properties of pairs (X, Y) are equivalent*

1. $R^{N}(X) = Y$ *and* $R^{-1N}(Y) = X$
2. $R^{N}(\overline{X}) = \overline{Y}$ *and* $R^{-1N}(\overline{Y}) = \overline{X}$
3. $R^{\Pi}(X) = Y$ *and* $R^{-1\Pi}(Y) = X$
4. $R \subseteq (X \times Y) \cup (\overline{X} \times \overline{Y})$

Thus, (X, Y) and $(\overline{X}, \overline{Y})$ are two independent sub-context in R, in the sense that there is no object / property pair (x, y) from context R in $X \times \overline{Y}$ nor in $\overline{X} \times Y$. There is no minimality requirement in the inclusion property objects

p		1	2	3	4	5	6	7	8
r	a							$\times$	
o	b					$\times$	$\times$		
p	c						$\times$	$\times$	$\times$
e	d					$\times$	$\times$	$\times$	$\times$
r	e					$\times$	$\times$	$\times$	$\times$
t	f		$\times$		$\times$				
i	g		$\times$	$\times$	$\times$				
e	h		$\times$	$\times$	$\times$				
s	i	$\times$							

Figure 1: Formal Concepts and Sub-contexts

4 of the above proposition. In particular, the pair $(\mathcal{O}, \mathcal{P})$ trivially satisfies it. However, this result leads to a decomposition of R into a disjoint union of *minimal* independent sub-contexts. Indeed, suppose two pairs (X_1, Y_1), (X_2, Y_2) satisfy the above proposition. It implies that for instance, the pair $(X_1 \cap X_2, Y_1 \cap Y_2)$ satisfies it (it can be checked that $R^N(X_1 \cap X_2) = Y_1 \cap Y_2)$, and likewise with any element of the partition refining both partitions $(X_1, \overline{X_1})$ and $(X_2, \overline{X_2})$. Due to point 4 of the proposition, it yields

$$R \subseteq ((X_1 \times Y_1) \cup (\overline{X_1} \times \overline{Y_1})) \cap ((X_2 \times Y_2) \cup (\overline{X_2} \times \overline{Y_2})), \tag{3}$$

where the intersection on the right-hand side comes down to the union of subcontexts $(X_1 \cap X_2) \times (Y_1 \cap Y_2)$, $(X_1 \cap \overline{X_2}) \times (Y_1 \cap \overline{Y_2})$, $(\overline{X_1} \cap X_2) \times (\overline{Y_1} \cap Y_2)$, $(\overline{X_1} \cap \overline{X_2}) \times (\overline{Y_1} \cap \overline{Y_2})$. The decomposition of R into minimal subcontexts is achieved by taking the following intersection [23]

$$R^* = \bigcap_{(X,Y): R^N(X)=Y, R^{-1N}(Y)=X} (X \times Y) \cup (\overline{X} \times \overline{Y}). \tag{4}$$

In general, $R \subset R^*$.

Example [22]. Fig. 1 presents a formal context. Pairs $(\{6,7,8\}, \{c,d,e\})$, $(\{5,6,7,8\}, \{d,e\})$, $(\{2,3,4\}, \{g,h\})$ are examples of formal concepts, while pairs $(\{5,6,7,8\}, \{a,b,c,d,e\})$, $(\{2,3,4\}, (\{f,g,h\})$, $(\{1\}, \{i\})$ are minimal subcontexts. And it can be checked that

$$R \subset \{5,6,7,8\} \times \{a,b,c,d,e\} \cup \{2,3,4\} \times \{f,g,h\} \cup \{1\} \times \{i\}.$$

The connection $(R^{\Pi}, R^{-1\Pi})$ has been originally introduced by Georgescu and Popescu [31] and studied in the framework of multivalued data tables with entries in a residuated lattice, but its practical significance for Boolean data tables was not really discussed. These authors call a pair of operators (f, g), where $f : 2^{Obj} \to 2^{Prop}$, $g : 2^{Prop} \to 2^{Obj}$ relating the subsets of objects and properties, a *conjugated pair of operators* if and only if

$$X \cap g(Y) = \emptyset \iff f(X) \cap Y = \emptyset.$$

It is easy to see that $(R^{\Pi}, R^{-1\Pi})$ is a conjugated pair of operators. To see it note that $R^{\Pi}(X) \cap Y = \emptyset$ also writes $\cup_{x \in X}(R(x) \cap Y) = \emptyset$. It holds if and only if $R \cap (X \times Y) = \emptyset$. So, by symmetry, it is equivalent to $R^{-1\Pi}(Y) \cap X = \emptyset$.

In terms of the dual operator N, the conjugation property reads $Y \subseteq R^N(\overline{X}) \iff X \subseteq R^{-1N}(\overline{Y})$. However, this connection is not a Galois connection. One reason is that iterating R^N and R^{-1N} does not yield an idempotent operation. Of course the same holds for $R^{-1\Pi}(R^{\Pi}(X))$, $R^{\Pi}(R^{-1\Pi}(Y))$, $R^N(R^{-1N}(Y))$. For instance, on the data table of Figure1, $R^{-1N}(\{a, c, d, e\}) = \{7, 8\}$, $R^N(\{7, 8\}) = \{a\}$ and $R^{-1N}(\{a\}) = \emptyset$.

Through the notions of formal sub-contexts and of formal concepts, one sees two aspects of granulation at work. Namely, on the one hand independent sub-contexts are separated granules, while *inside* each sub-context, formal concepts (X, Y) are identified where each object in X is associated with each property in Y, which can be viewed as a cluster. Note that in the special case when a formal context can be decomposed into independent formal concepts (i.e. each minimal sub-context is a formal concept), we have a perfect granulation: two objects are either identical in terms of properties, or they do not have any property in common. However, in the general case, objects in the extension of a formal concept may not be fully similar since they may also possess properties outside the intension of the concept. They are only similar with respect to the properties associated to the formal concept. In practice, it may be interesting to introduce some tolerance in the definition of formal sub-contexts and concepts [23, 30], leading to a more permissive and approximate view of granules or clusters.

Besides, the above results can be also expressed in terms of bipartite graph clustering, where

- There are two kinds of nodes corresponding to objects and properties.
- Formal concepts correspond to sets of object-nodes connected to all nodes in subsets of property-nodes.

- The decomposition into independent subcontexts corresponds to connected components of the bipartite graph (each node of one set being related to at least one node of another set of the opposite type).

One can then take advantage of this exact parallel between formal concept analysis and bipartite graph analysis [30].

2.5 From formal concept analysis to rough sets

The concept of granulation is even more central in rough set theory [41]. Rough set theory focuses on the impossibility to precisely describe any set of objects when the properties used to describe them are not enough discriminant. One connection between FCA and rough sets is that the latter also start from a data table like a formal context (we assume Boolean attributes in the following). Let X_y be the set of objects satisfying the property y. Then there exists a partition generated on $\mathcal{O}$ by the family of subsets $\{X_y : y \in \mathcal{P}\}$, each element of which is an interpretation of the propositional language induced by properties in $\mathcal{P}$, i.e. it is of the form $\times_{y \in \mathcal{P}} X_y^{e_y}$, $e_y \in \{-1, 1\}$, with $X_y^{e_y} = X_y$ if $e_y = 1$, and $X_y^{e_y} = \overline{X}_y$ if $e_y = -1$. If R is the formal context, then two objects x and x' are said to be indiscernible (they are in the same element of the partition) if they share the same properties (which writes $R(x) = R(x')$). It enables the data table to be reduced to the case where no two lines in R are equal.

The rough set approach considers the above partition of the universe $\mathcal{O}$ of objects, say $X_1, \ldots, X_k$ induced by the properties via the equivalence relation E defined by $E(x, x') = 1$ if and only if $R(x) = R(x')$ and 0 otherwise. So, all that is known about any object in $\mathcal{O}$ is which subset of the partition it belongs to. So each subset X of objects is only known in terms of its upper and lower approximations, a pair (X_*, X^*) such that

$$X^* = \bigcup \{X_i, X_i \cap X \neq \emptyset\} \text{ and } X_* = \bigcup \{X_i, X_i \subseteq X\}. \tag{5}$$

It is clear that $(A \cap B)^* \subseteq A^* \cap B^*$ and $A_* \cup B_* \subseteq (A \cup B)_*$. Note that an equivalence class of relation E corresponds to a specialisation of both a formal concept and a formal independent subcontext.

To summarize the links between rough sets and FCA, a formal concept can be viewed as a 2-dimensional extension of an equivalence class. A formal context is a 2-dimensional extension of equivalence relation if it can be decomposed into a disjoint union of elementary sub-contexts, each of which forms a single formal concept. In that case, the context we start with is the perfect extension of the equivalence relation to the 2-dimensional setting.

Another way of putting together FCA and rough sets consist in putting both on a cube of oppositions, whereby their connections to possibility theory functions can be highlighted; see [9].

2.6 Clusters and granules

Assume now a general relation S between objects, that is $S \subseteq \mathcal{O} \times \mathcal{O}$. It can be viewed as a directed graph whose nodes form the set $\mathcal{O}$. We assume the relation is serial, that is $\forall x \in \mathcal{O}, S(x) \neq \emptyset$, and its converse S^{-1} is serial too; we say that S is biserial. The definition of a formal concept then is a maximal Cartesian product $A \times B \subseteq \mathcal{O} \times \mathcal{O}$ contained in S. We can still define it as satisfying the two equalities $S^{\triangle}(A) = B$ and $S^{-1\triangle}(B) = A$. Suppose the relation S is symmetrical, in order to capture some idea of proximity. Then, the maximal Cartesian products $A \times B$ contained in S are of the form $C \times C \subseteq S$, i.e., they are maximal cliques in the non-directed graph associated to S: the two equalities defining formal concepts then boil down to a single one:

$$S^{\triangle}(C) = \cap_{x \in C} S(x) = C, \tag{6}$$

which expresses the fact that each node in C is related to all nodes in C, and corresponds to one major feature of a cluster. We call the set C a *tight cluster*, because each element in C is close to all other elements in C. Note that S must be reflexive (an element is close to itself), otherwise there is no such tight cluster. Then it is enough to require that $C \subseteq S^{\triangle}(C)$ since the other inclusion trivially holds.

Alternatively we can consider minimal Cartesian products $A \times B$ such that $S \subseteq (A \times B) \cup (\overline{A} \times \overline{B})$, which satisfy the two equalities $S^{\Pi}(A) = B$ and $S^{-1\Pi}(B) = A$. If the relation S is symmetrical, it corresponds to the minimal Cartesian products $B \times B$ such that $S \subseteq (B \times B) \cup (\overline{B} \times \overline{B})$. They satisfy the equality

$$S^{\Pi}(B) = \cup_{x \in B} S(x) = B, \tag{7}$$

This is because the identity (7) is equivalent to

$$S \subseteq (B \times B) \cup (\overline{B} \times \overline{B}). \tag{8}$$

If S is reflexive, it is enough to require that $S^{\Pi}(A) \subseteq A$ instead of (7) since the other inclusion trivially holds.

Minimal subsets G that satisfy (8) are such that each element of G is related to at least one element of B and to none outside G. This is the other expected property of a cluster, but we can call it a *loose granule*. Loose granules of S form the set $\mathcal{G}(S)$ and correspond to maximal connected components in the non-directed graph associated to S. Note that tight clusters can only be found inside loose granules: for any tight cluster A, there exists a loose granule containing it. Tight clusters and loose granules cannot be told apart if the relation S is moreover transitive.

Proposition 3 *Consider a symmetric serial relation S. Then $S = E$ is an equivalence relation if and only if its loose granules and tight clusters coincide.*

Proof If $S = E$ is an equivalence relation, it is easy to check that loose granules and tight clusters coincide. Conversely, if loose granules and tight clusters in S coincide then an element in a loose granule is connected to all elements in this granule and to none outside. So S corresponds to a partition, and is an equivalence relation. QED

It is also clear that the relation $S^* = \cap_{G\in\mathcal{G}(S)}(G\times G)\cup(\overline{G}\times\overline{G})$ is transitive, and is actually the transitive closure $cl(S)$ of S. As the transitive closure of S is reflexive, it is thus be an equivalence relation. So loose granules form a partition of $\mathcal{O}$. More precisely:

Proposition 4 *Consider a symmetric serial relation S. The tight clusters of $cl(S)$ are the loose granules of S*

Proof Let B be a loose granule of S. Since the graph with nodes in B is connected, all nodes in B will be related to all nodes in B in the graph of the transitive closure of S, but not to any node outside B. Hence B is a tight cluster of $cl(S)$. If B is not contained in a loose granule of S, then it is made of more than one connected component, hence they remain disconnected via transitive closure. So, B will not be a loose granule of $cl(S)$, a fortiori not a tight one. QED

So it can be seen that a reflexive and symmetric relation represents a partition of separated loose granules, each possibly containing several tight clusters (that may overlap), which makes it very similar to a formal context.

3 Extensional Fuzzy Sets and Fuzzy Contexts

The concept of extensional fuzzy set with respect to a fuzzy equality, proposed in [33, 47], further developed by Jacas and colleagues [7], Klawonn [36], and Recassens [45] also embeds ideas of granulation. It is a multivalued extension of the decomposition of a relation into tight clusters and loose granules recalled above. This approach has mathematical roots in category theory and Heyting algebras [34], whereby a multivalued notion of equality is used. As we are going to see, although defined in a different algebraic setting and on the basis of a completely different intuition, it is also closely related to the gradual version of formal concept analysis [4, 5, 43, 31].

3.1 Fuzzy Singletons and Extensional Hulls

Let E be a fuzzy similarity relation defined on a universe U. For simplicity, we assume the use of the scale $[0,1]$. E is supposed to be

- reflexive ($E(u,u) = 1$),
- symmetric ($E(u,v) = E(v,u)$),
- $*$-transitive ($E(u,v) * E(v,w) \leq E(u,w)$),

where $*$ is a triangular norm [37] (i.e., $*$ is increasing in the broad sense, associative, commutative and such that $0 * 0 = 0$, $1 * a = a$). It was first proposed by Zadeh [51] when $* = \min$.

Such a fuzzy relation models a form of proximity between elements of the set U. Relation E is sometimes also called “fuzzy equivalence” [7], “(fuzzy) equality relation” [36], or ”(fuzzy) indistinguishability relation” [47], or yet “indiscernibility relation” [41]. Note that the terms “indistinguishability” and “equality” refer to quite different intuitions, only the former being naturally understood as the weak version of an equivalence relation [20]. Indeed, one may argue that the 1-cut of a fuzzy equality should be the standard equality (i.e. $E(u,v) \neq 1$ if $u \neq v$), i.e. separability holds. On the contrary, the name *indistinguishability relation* is denying separability. In the following, we do not require separability.

Interesting choices for operation $*$ are min, product or the Łukasiewicz t-norm $a *_{\text{Ł}} b = \max(0, a + b - 1)$. Fuzzy similarity relation are the negative of distances or metrics [7, 45]. The min-transitivity makes a fuzzy similarity closely related to an ultrametric. The $*_{\text{Ł}}$-transitivity corresponds to the triangular inequality.

A fuzzy set F is said to be *extensional* with respect to E [33, 36] iff

$$\forall u, v, F(u) * E(u, v) \leq F(v) \tag{9}$$

Let $F \circ E$ be obtained as $F \circ E(v) = \max_{u \in U} F(u) * E(u, v)$. It is clear that due to the properties of E, it always holds that $F \subseteq F \circ E$. Moreover $F \circ E$ can be written as $E^{\Pi}(F)$ as it is the fuzzy set contains all elements in the vicinity of F. So, Equation (9) can be written as $F \circ E(v) = F$. Equation (9) generalizes the condition $S^{\Pi}(B) = B$ in equation (7), so that we can also write it as $E^{\Pi}(B) = B$.

Consider now the implication connective $\rightarrow$ associated to $*$ by residuation, i.e. we assume $a * b \leq c \Leftrightarrow a \leq b \rightarrow c$. The extensionality of F is obviously equivalent to

$$\forall u, v, F(u) \leftrightarrow F(v) \geq E(u, v) \tag{10}$$

where $a \leftrightarrow b = \min(a \rightarrow b, b \rightarrow a)$, using residuation and the symmetry of E. Equation (10) generalizes the property (8) $S \subseteq (B \times B) \cup (\overline{B} \times \overline{B})$ to multivalued relations.

The *extensional hull* $\hat{F}$ of a fuzzy set F (w.r.t. E) is then defined as

$$\hat{F} = \inf\{G | F \subseteq G \text{ and } G \text{ is extensional w.r.t. } E\}.$$

It is obvious that $F \circ E$ is extensional ($E^{\Pi}(F \circ E) = (F \circ E) \circ E = F \circ (E \circ E) = F \circ E$, since E is $*$-transitive) and is the extensional hull of F.

An important example of extensional fuzzy set is obtained by considering an element u and the fuzzy set F_u of elements similar to it, that is $F_u(v) = E(u, v)$ (it is a line of matrix E). F_u is clearly the extensional hull of the singleton $\{u\}$. Note that $F_v(u) = F_u(v)$, and that if $F_v(u) = 1$ then $F_v = F_u$. F_u is the fuzzy counterpart of an equivalence class. Klawonn [36] calls it a "fuzzy point", understood as the largest cluster of indiscernible entities around u, as per the fuzzy similarity relation E.

Each fuzzy set F_v can be seen as a fuzzy loose granule. It is an atomic entity inside U that cannot be split, if an observer whose myopic eyesight is modelled by the fuzzy similarity E. If E is an equivalence relation (for instance, the 1-cut of a fuzzy similarity is clearly an equivalence relation), F_u is just the equivalence class of u. The extensional hull of a crisp subset $A \subseteq U$ is the union of extensional hulls of all elements in the set:

$$\mu_{\hat{A}}(u) = \sup_{v \in A} E(u, v) \tag{11}$$

An interesting question whether any extensional fuzzy set takes this form. An extensional fuzzy set would then always consist of the fuzzy union of fuzzy extensional hulls of singletons, as in the crisp case. It would hold if an extensional fuzzy set coincides with the extensional hull of its core. But the latter property is not true. For instance, consider a fuzzy set F containing strictly F_u but with the same core A. Clearly, its extensional hull $E^{\Pi}(F)$ strictly contains F_u but also has the same core A (an equivalence class of the 1-cut of E). Hence it is not of the form $\cup_{u \in A} F_u$.

Höhle and Klawonn call a *fuzzy singleton* F (w.r.t. E) a non-empty fuzzy set (i.e., $\max_u F(u) = 1$) such that

$$F(u) * F(v) \leq E(u, v) \tag{12}$$

In particular we equivalently have $F(u) \leq F(v) \to E(u, v), \forall v \in U$, that is,

$$F(u) \leq \min_{v \in U} F(v) \to E(u, v).$$

Considering maximal fuzzy singletons, we generalize the FCA operator: they are such that $F = E^{\Delta}(F)$, since the composition on the right-hand side of the above inequality extends operation Δ. Clearly, the union of two such fuzzy singletons is not a fuzzy singleton. In fact, a fuzzy singleton is a greatest fuzzy set satisfying (12). Maximal fuzzy singletons are the multivalued version of the notion of tight cluster, i.e. the specialization of a formal concept to relations over a set. [1]

Using a $*$-transitive similarity relation we can prove that extensional hulls of singletons are maximal fuzzy singletons.

Proposition 5 *If E is a $*$-transitive similarity relation, and $w \in U$ a singleton, then $F_w(u) * F_w(v) \leq E(u, v)$*

Proof Note that letting $F = F_w$ in (12), we again get the expression of the transitivity of E. Hence F_w satisfies (12). QED

What this result shows is that fuzzy versions of tight clusters and loose granules in the sense of a fuzzy similarity relation coincide with equivalence classes F_u, just like in the classical case for equivalence relations. Due to $*$-transitivity, it holds that $E^{\Delta}(F_u) = E^{\Pi}(F_u) = F_u, \forall u \in U$. One question to

[1] The term "singleton" here means that fuzzy singletons are atomic entities as per the indistinguishability relation E.

be solved is whether there are other fuzzy sets that are at the same time extensional and are fuzzy singletons, that is whether $E^{\Delta}(F) = E^{\Pi}(F) = F$ implies that F is just the extensional hull of a singleton (a fuzzy similarity class). Note that extensional hulls of crisp subsets other than singletons do not qualify as candidates as $E^{\Delta}(A) = \cap_{u\in A} E^{\Delta}(\{u\})$ and $E^{\Pi}(A) = \cup_{u\in A} E^{\Pi}(\{u\})$.

Valverde [47] (see also [7, 36, 45]) considers the converse problem of generating a fuzzy relation from a family of subsets. Given a family $\mathcal{F}$ of fuzzy sets F the coarsest equivalence relation $E^{\mathcal{F}}$ such that all fuzzy sets $F \in \mathcal{F}$ are extensional is

$$E^{\mathcal{F}}(u, v) = \bigwedge_{F\in\mathcal{F}} F(u) \leftrightarrow F(v). \tag{13}$$

In the crisp case, take $\mathcal{F}$ as $\{A_i : y_i \in \mathcal{P}\}$. Then it simply says that two elements are related if and only if they belong to the same sets A_i (they share the same properties). This equation is extended to the case where the properties are more or less important by Bělohlávek [5].

While the coarsest fuzzy similarity relation $E^{\mathcal{F}}$ such that all fuzzy sets $F \in \mathcal{F}$ are extensional is provided above by Valverde result (13), the finest such fuzzy similarity relation $E_{\mathcal{F}}$ is of the form

$$\begin{aligned} E_{\mathcal{F}}(u, v) &= 1 \text{ if } u = v \\ &= \bigvee_{F\in\mathcal{F}} F(u) * F(v) \text{ otherwise.} \end{aligned} \tag{14}$$

Moreover, Klawonn [36] addresses the case when a collection of normalized fuzzy sets can be viewed as forming a family of fuzzy points. If $\forall F_i \in \mathcal{F}, \exists u_i$, such that $F_i(u_i) = 1$, then the fact that $\mathcal{F}$ is a family of fuzzy points with respect to E is equivalent to the following inequality: $\forall F_i, F_j \in \mathcal{F}$,

$$\bigvee_{u\in U} F_i(u) * F_j(u) \leq \bigwedge_{v\in U} F_i(v) \leftrightarrow F_j(v) \tag{15}$$

This condition is a fuzzy counterpart of the fact that equivalence classes (here generalized to fuzzy points) are disjoint.

3.2 Extensional Fuzzy sets and FCA: Analogies

In the Boolean case, the mathematical expressions (6) and (7) are special cases of formal concept analysis expressions. Similarly, in the multivalued case, we can generalize identities (9, 10 12) to the setting of FCA. First, the counterpart to (9) using a formal multivalued context is: $\forall x, y$,

$$\begin{aligned} X(x) * R(x,y) &\leq Y(y) \\ Y(y) * R^{-1}(y,x) &\leq X(x) \end{aligned} \tag{16}$$

It is the multivalued version of the third point of Proposition 2 that operates a decomposition into disjoint subcontexts. It is equivalent to the counterpart of (10) and point 4 of Proposition2 from Section2.4), namely:

$$\forall x, y, X(x) \leftrightarrow Y(y) \geq R(x,y) \tag{17}$$

As already said, in the fuzzy similarity setting, there is only one equation (9) instead of two in FCA because the fuzzy similarity relation is symmetric. This indicates that the idea of extensional fuzzy set bears a strong analogy with the notion of formal sub-context Indeed, (16) expresses that if an object x of X has property y then this property is in Y, and conversely if a property y of Y applies to an object x then this object is in X, i.e. (X, Y) is an independent subcontext; so an independent subcontext is extensional. Moreover, we can deal with a fuzzy extension of the notion of formal sub-context [22] since Equations (16) and (17) make sense in $[0, 1]$, and not only in $\{0, 1\}$.

In fact, the decomposition of R into minimal contexts (forming relation R^* in Equation (4) above the example of Section 2.4) corresponds to the construction of the coarsest fuzzy similarity relation induced by a family of fuzzy sets as per Eq. (13). To see it, just consider instead of the family $\mathcal{F}$ the set of conjugated pairs obtained from the context R. More generally, a fuzzy relation R on U generates a family $\mathcal{F}(R)$ of fuzzy sets F_u such that $F_u(v) = R(u, v), \forall u \in U$. Considering the coarsest fuzzy similarity relation $E^{\mathcal{F}(R)}$, it is clear that $R \subseteq E^{\mathcal{F}(R)}$ just like $R \subseteq R^*$ in the context decomposition framework.

Likewise, multivalued counterparts of formal concepts, as per Proposition 19 can be defined:

$$X(x) * Y(y) \leq R(x,y), \tag{18}$$

which is equivalent to [4, 5] $\forall x, y$,

$$\begin{aligned} X(x) \to R(x,y) &\geq Y(y) \\ Y(y) \to R^{-1}(y,x) &\geq X(x). \end{aligned} \tag{19}$$

one can see a parallel between the idea of a fuzzy point (a maximal fuzzy singleton in the sense of (12)) and the notion of formal concept. Indeed, equation (19) expresses that if a property y is in Y, any object x of X should possess it, and conversely if an object x is in X, any property y in Y should be possessed by it. And equation (12) of fuzzy singletons can also be expressed as $F(u) \leq E(u,v) \to F(v)$, from residuation, so that we do have that $F = E^{\Delta}(F)$ and a pair of fuzzy points (F,F) is like a formal concept. So a concept (X,Y) is similar to a fuzzy point. Equations (18) and (19) in fact provide a fuzzy extension of the notion of formal concept in the sense developed in [4, 5], whose similarity with the extensional fuzzy set construction is thus laid bare.

It is clear that forming the union of fuzzy formal concepts in a context R yields a relation $R_* \subseteq R$ (with equality in the crisp case). It is the counterpart of the finest fuzzy similarity relation in Equation (14) induced by a family of fuzzy sets, while decomposing R into formal contexts yields a relation R^*, defined by Equation (4), that contains R, and reminds us of the coarsest relation induced by a family of fuzzy sets (13). The obvious inclusion $R_* \subseteq R^*$ is clearly the counterpart of Equation (15).

Thus, we have exhibited a formal resemblance between two quite different views of a granulation process. There is a big difference between them, though. One is induced by an approximate equality relation, while the other is based on a binary relation defined on the Cartesian product of two different sets. In the former case, due to the properties of the fuzzy similarity relation what corresponds to concepts in FCA, and what corresponds to minimal independent sub-contexts are the same (they are fuzzy points). Moreover, the fuzzy extensionality problem is to derive an fuzzy similarity relation from any family of fuzzy sets, while in FCA the issue is to find “maximal singletons” and minimal independent subrelations induced by any binary relation. However the common algebraic setting for both problems is a building block of fuzzy FCA as developed by Bělohlávek [5]. This algebraic setting, also used by Klawonn [36] in his approach to extensional fuzzy sets, is the one of residuated lattices.

Lastly, the first part of expression (16) and the expression (18) are also the starting points respectively of the implication-based and of the conjunction-based views of a fuzzy rule “if x is in $\tilde{X}$ then y is in $\tilde{Y}$” [17]. Fuzzy rules defined via these two equations indeed correspond to two different ways of granulating a relation or function defined from the universe containing the (fuzzy) subset $\tilde{X}$ to the universe containing the (fuzzy) subset $\tilde{Y}$. Klawonn [36] shows that the counterpart to inequality (15) is instrumental in the solution of fuzzy relational equations induced by the specification of fuzzy rules, especially if the fuzzy relation must be constructed using the conjunction-based view. In some sense the modelling of fuzzy rules and fuzzy formal concept analysis rely on the same basic algebraic setting and the same basic equations but have opposite programs. While fuzzy FCA tries to extract concepts from fuzzy relations modeling many-valued contexts, with a view to derive interpretable association rules, the other program is to synthetize fuzzy relations between input to output spaces from fuzzy rules expressed in natural language. The formal relations between the two areas are thus worth studying further. For instance, Bělohlávek [6] tries to derive implicative rules from fuzzy formal contexts, using the same equation (inf $\rightarrow$ composition) as the one that turns a set of implicative rules into a fuzzy relation [17].

3.3 Fuzzy Rough sets and Similarity Relations

Rough sets can be extended by replacing an equivalence relation by a fuzzy similarity relation [15], thus introducing degrees of possibility and necessity that an element belongs to a given crisp set, due to the fuzzy granulation of the referential. There is an extensive literature on fuzzy rough sets [44] that seems to be unrelated to the Höhle-Klawonn view of extensional fuzzy sets recalled above, that also relies on similarity relations, and induces a form of granulation of the referential. The bridge between fuzzy rough sets and extensional fuzzy sets is however made in [45].

The notion of extensional fuzzy set with respect to a similarity relation clearly generalises the notion of exact set in rough set theory, that is formed by the union of equivalence classes. The so-called extensional hull of a fuzzy set, viewed as the smallest extensional fuzzy set containing it, is formally the same as the upper approximation of this fuzzy set by means of the partition formed by the fuzzy singletons. In particular the extensional hull $\hat{X}$ of a set X (of the form (11)) does coincide with the upper fuzzy approximation of set A in the sense of fuzzy rough sets [15]. In the theory of extensional fuzzy sets, the lower approximation of a fuzzy set F takes the following form [7, 45]:

$$F_E(u) = \inf_{u \in U} E(u, v) \rightarrow F(v) \tag{20}$$

with a residuated implication $\rightarrow$ with respect to a t-norm $*$. F_E is the largest extensional fuzzy set included in F, namely it is such that $F_E(u) * E(u, v) \leq F(v), \forall u \in U$. In other words, F_E is of the form $E^N(F)$ in the sense of necessity functions. However, we do not have that $F_E = \overline{\overline{F} \circ E}$ in general, which suggests that such approximation pairs may fail to have all properties of usual rough sets. This approach thus differs from [15] where the chosen implication in (20) is Kleene's, so that the lower approximation is precisely defined by $\overline{\overline{F} \circ E}$, respecting the duality between upper and lower approximations, but possibly failing the extensionality property. The connection between extensionality and rough sets has been very recently discussed by Chakraborty [8] in the setting originally described by Higgs [34], that inspired Höhle and Klawonn, and in the fuzzy set setting in [45], Chapter 3.

So, pairs $(F \circ E, F_E)$ can be viewed as fuzzy rough sets. They provide the approximate description of fuzzy sets by means of fuzzy points in the sense of a fuzzy similarity relation, just like rough sets in the more elementary setting of a crisp equivalence relation. In Ruspini [46], and the literature on similarity-based reasoning [32], a fuzzy set is always understood as the extensional hull of a crisp set. The connections and difference of points of view between fuzzy rough sets and similarity-based reasoning after Ruspini, have already been emphasised [20]. While rough sets and granulation insist on the idea that elements of the referential cannot be distinguished, the idea of similarity, often then termed fuzzy equality, and viewed as the negative of a distance, insists on making a difference between elements however close they can be. If obeying separability, fuzzy similarity relations are then more tailored to interpolation purposes [11, 42] than to classification.

4 Concluding Remarks

The idea of granulation [53] is based on the notion of cluster whereby

1. any pair of members of a cluster should be closely related in some sense;

2. any member of a cluster should be sufficiently separated from any member from outside the cluster.

The paper has provided a discussion of several areas, where the idea of granulation [53] is central, and notions of closeness and separation can be defined. On this ground, similarities between different settings like possibility theory, formal concept analysis, extensional fuzzy sets, and rough sets have been laid bare. Similar structures were found to be at work in such settings. This kind of attempt may lead to mutual enrichments between theories, as in the parallel between possibility theory and formal concept analysis.

It is clear that such formal links should be further investigated in more general representation frameworks such as pattern structures [28, 1], but also using algebraic structures beyond residuated lattices exploited in [5]. Indeed, the many-valued FCA suffers from two limitations. First, one may object to the fact that most of the time, the negation in residuated lattice is not involutive, which may make the decomposition of fuzzy contexts into independent subcontexts more difficult: it may be difficult to write Equation (17) in the form of Point 4 of Proposition 2. One way to do so is to interpret implication as $a \to b = n(a * n(b))$ in (17) for an involutive negation n. But then the underlying conjunction associated to $\to$ through residuation will no longer be associative nor commutative [13, 27]. A study of multivalued FCA using non-associative, non-commutative conjunctions is carried out by Medina et al. [39], using so-called multi-adjoint concept lattices. Lastly, it seems to be idealistic to assume that the degrees of satisfaction of all properties of objects can be measured on the same non-Boolean scale. This assumption may be problematic when processing real non-Boolean data. This issue is taken up at the theoretical level by Medina and Ojeda-Aciego [38] using multi-adjoint concept lattices.

References

[1] Z. Assaghir, M. Kaytoue, and H. Prade, "A possibility theory-oriented discussion of conceptual pattern structures," in *Scalable Uncertainty Management (SUM'10), Toulouse*, ser. LNAI, A. Deshpande and A. Hunter, Eds., vol. 6379. Springer, 2010, pp. 70–83.

[2] R. E. Bellman, R. Kalaba, and L. Zadeh, "Abstraction and pattern classication," *J. of Mathematical Analysis and Applications*, vol. 13, pp. 1–7, 1966.

[3] R. E. Bellman and L. Zadeh, "Decision making in a fuzzy environment," *Management Science*, vol. 17, pp. B141–B164, 1970.

[4] R. Bělohlávek, "Fuzzy Galois connections," *Math. Logic Quart*, vol. 45, pp. 497–504, 1999.

[5] R. Bělohlávek, *Fuzzy Relational Systems*. Foundations and Principles. Kluwer Academic/Plenum Publishers, New York, 2002.

[6] R. Bělohlávek, Optimal triangular decompositions of matrices with entries from residuated lattices, *International Journal of Approximate Reasoning* 50, 12501258, 2009.

[7] D. Boixader, J. Jacas, J. Recassens, Fuzzy equivalence relations: advanced material. In *Fundamentals of Fuzzy Sets*, (Dubois, D. Prade,H., Eds.), Kluwer, Boston, Mass., The Handbooks of Fuzzy Sets Series, 261-290, 2000.

[8] M. Chakraborty, "On fuzzy sets and rough sets from the perspective of indiscernibility," in *Proc. 4th Indian Conference on Logic and its Applications, Delhi*, ser. LNAI, vol. 6521. Springer, 2011, pp. 22–37.

[9] D. Ciucci, D. Dubois, H. Prade: The Structure of Oppositions in Rough Set Theory and Formal Concept Analysis - Toward a New Bridge between the Two Settings. In *Foundations of Intelligent Knowledge Systems* (FoIKS 2014), Lecture Notes in Computer Science, Vol. 8367, Springer, 154-173, 2014.

[10] Y. Djouadi, D. Dubois, and H. Prade, "Possibility theory and formal concept analysis: Context decomposition and uncertainty handling," in *Computational Intelligence for Knowledge-Based Systems Design, Proc.13th Inter. Conf. on Information Processing and Management of Uncertainty (IPMU 2010), Dortmund*, ser. LNCS, E. Hüllermeier, R. Kruse, and F. Hoffmann, Eds., vol. 6178. Springer, 2010, pp. 260–269.

[11] D. Dubois, F. Esteva, P. Garcia, and L. Godo, H. Prade,"A logical approach to interpolation based on similarity relations," *Int. J. Approx. Reasoning*, vol. 17, no. 1, pp. 1–36, 1997.

[12] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. New-York: Academic Press, 1980.

[13] D. Dubois, H. Prade. A theorem on implication functions de

ned from triangular norms. Stochastica, 8: 267-279, 1984.

[14] D. Dubois and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, 1988.

[15] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Int. J. of General Systems*, vol. 17, pp. 191–209, 1990.

[16] D. Dubois and H. Prade, "Possibility theory as a basis for preference propagation in automated reasoning," in *Proc. 1st IEEE Inter. Conf. on Fuzzy Systems 1992 (FUZZ-IEEE'92), San Diego, Ca., March 8-12*, 1992, pp. 821–832.

[17] D. Dubois and H. Prade, "What are fuzzy rules and how to use them," *Fuzzy Sets and Systems*, vol. 84, pp. 169–185, 1996.

[18] D. Dubois and H. Prade, "The three semantics of fuzzy sets," *Fuzzy Sets and Systems*, vol. 90, pp. 141–150, 1997.

[19] D. Dubois and H. Prade, "Possibility theory: qualitative and quantitative aspects," in *Quantified Representation of Uncertainty and Imprecision*, ser. Handbook of Defeasible Reasoning and Uncertainty Management Systems, D. Gabbay and P. Smets, Eds. Kluwer Acad. Publ., 1998, vol. 1, pp. 169–226.

[20] D. Dubois and H. Prade, "Similarity versus preference in fuzzy set-based logics ," in *Modelling Incomplete Information: Rough Set Analysis* , ser. Studies in Fuzziness and Soft Computing, E. Orlowska, Ed. Heidelberg: Physica Verlag, 1998, pp. 441–461.

[21] D. Dubois and H. Prade, "Possibility theory and its applications: Where do we stand ?" in *Springer Handbook of Computational Intelligence* (J. Kacprzyk, W. Pedrycz, Eds), Springer, p. 31-60, 2015

[22] D. Dubois and H. Prade, "Possibility theory and formal concept analysis in information systems," in *Proc. Inter. Fuzzy Systems Assoc. World Congress and Conf. of the Europ. Soc. for Fuzzy Logic and Technology (IFSA-EUSFLAT'09), Lisbon, July 20-24*, 2009, pp. 1021–1026.

[23] D. Dubois and H. Prade, "Possibility theory and formal concept analysis: Characterizing independent sub-contexts and handling approximations," *Fuzzy Sets and Systems*, 196, 4-16, 2012.

[24] D. Dubois, F. D. de Saint Cyr, and H. Prade, "A possibilty-theoretic view of formal concept analysis," *Fundamenta Informaticae*, vol. 75, no. 1-4, pp. 195–213, 2007.

[25] I. Düntsch and E. Orlowska, "Mixing modal and sufficiency operators," *Bulletin of the Section of Logic, Polish Academy of Sciences*, vol. 28, no. 2, pp. 99–106, 1999.

[26] I. Düntsch and G. Gediga, "Approximation operators in qualitative data analysis," *In: Theory and Application of Relational Structures as Knowledge Instruments*, pp. 216–233, 2003.

[27] J. Fodor, On fuzzy implication operators, *Fuzzy Sets and Systems*, 42(3),1991, 293-300.

[28] B. Ganter and S. O. Kuznetsov, "Pattern structures and their projections," in *ICCS '01: Proceedings of the 9th International Conference on Conceptual Structures*. Springer-Verlag, 2001, pp. 129–142.

[29] B. Ganter and R. Wille, *Formal Concept Analysis*. Springer-Verlag, 1999.

[30] B. Gaume, E. Navarro, and H. Prade, "A parallel between extended formal concept analysis and bipartite graphs analysis," in *Computational Intelligence for Knowledge-Based Systems Design, Proc.13th Inter. Conf. on Information Processing and Management of Uncertainty (IPMU 2010), Dortmund*, ser. LNCS, E. Hüllermeier, R. Kruse, and F. Hoffmann, Eds., vol. 6178. Springer, 2010, pp. 270–280.

[31] G. Georgescu and A. Popescu, "Non-dual fuzzy connections," *Arch. Math. Log.*, vol. 43, no. 8, pp. 1009–1039, 2004.

[32] L. Godo and R. O. Rodriguez, "Logical approaches to fuzzy similarity-based reasoning: an overview," in *Preferences and Similarities*, ser. CISM Courses and Lectures. Springer, 2008, vol. 504, pp. 75–128.

[33] U. Höhle, "Quotients with respect to similarity relations," *Fuzzy Sets and Systems*, vol. 27, pp. 31–44, 1988.

[34] D. Higgs, "A category approach to Boolean valued set theory," Tech. Rep., University of Waterloo, Canada, 1973.

[35] J. Jacas, On the generators of T-indistinguishability operators, *Stochastica*, 12, 49-63, 1990.

[36] F. Klawonn, "Fuzzy points, fuzzy relations and fuzzy functions," in *Discovering the World with Fuzzy Logic*, V. Novák and I. Perfilieva, Eds. Heidelberg: Physica-Verlag, 2000, pp. 431–453.

[37] E. P. Klement, R. Mesiar, and E. Pap, *Triangular Norms*. Dordrecht: Kluwer Academic, 2000.

[38] J. Medina, M. Ojeda-Aciego, On multi-adjoint concept lattices based on heterogeneous conjunctors. emphFuzzy Sets and Systems 208: 95-110, 2012.

[39] J. Medina, M. Ojeda-Aciego, J. Ruiz-Calviño: Formal concept analysis via multi-adjoint concept lattices. *Fuzzy Sets and Systems* 160(2): 130-144, 2009.

[40] Z. Pawlak, Information systems Theoretical foundations. *Information Systems*, 6(3):205-218, 1981.

[41] Z. Pawlak, *Rough Sets. Theoretical Aspects of. Reasoning about Data.* Dordrecht: Kluwer Acad. Publ., 1991.

[42] I. Perfilieva, D. Dubois, H. Prade, F. Esteva, L. Godo, and P. Hodáková, "Interpolation of fuzzy data: Analytical approach and overview," *Fuzzy Sets and Systems*, 192, 134-158, 2012.

[43] A. Popescu, "A general approach to fuzzy concepts," *Mathematical Logic Quarterly*, vol. 50, pp. 265 – 280, 2004.

[44] A. M. Radzikowska and E. Kerre, "A comparative study of fuzzy rough sets," *Fuzzy Sets and Systems*, vol. 126, no. 2, pp. 137–155, 2002.

[45] J. Recassens, *Indistinguishability Operators*, STUDFUZZ 260, Berlin: Springer-Verlag, 2010

[46] E. H. Ruspini, "On the semantics of fuzzy logic," *Int. J. Approx. Reasoning*, vol. 5, no. 1, pp. 45–88, 1991.

[47] L. Valverde, "On the structure of F-indistinguishability operators," *Fuzzy Sets and Systems*, vol. 17, pp. 313–328, 1985.

[48] Y. Yao, “A comparative study of formal concept analysis and rough set theory in data analysis,” in *Rough Sets and Current Trends in Computing, 4th International Conference, RSCTC 2004*. Uppsala, Sweden: LNCS 3066, Springer, June 1-5 2004, pp. 59–68.

[49] Y. Y. Yao and Y. Chen, “Rough set approximations in formal concept analysis,” *Transactions on Rough Sets V, LNCS 4100,*, pp. 285–305, 2006.

[50] L. A. Zadeh, “Fuzzy sets,” *Inform Control*, vol. 8, pp. 338–353, 1965.

[51] L. A. Zadeh, “Similarity relations and fuzzy orderings,” *Information Sciences*, vol. 3, pp. 177–200, 1971.

[52] L. A. Zadeh,“Fuzzy sets as a basis for a theory of possibility,” *Fuzzy Sets and Systems*, vol. 1, pp. 3–28, 1978.

[53] L. A. Zadeh, “Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic,” *Fuzzy Sets and Systems*, vol. 90, pp. 111–128, 1997.

[54] L. A. Zadeh, “Toward a generalized theory of uncertainty (GTU) – an outline,” *Information Sciences*, vol. 172, pp. 1–40, 2005.

The posterity of Zadeh's 50-year-old paper:

A retrospective in 101 Easy Pieces – and a Few More

James C. Bezdek
Dept. of Computing and Information Systems
University of Melbourne
Melbourne, Australia
jcbezdek@gmail.com

Didier Dubois, Henri Prade
Institut de Recherche en Informatique de Toulouse
Universite Paul Sabatier IRIT
Toulouse, France
dubois@irit.fr, prade@irit.fr

Abstract—This article was commissioned by the 22nd IEEE International Conference of Fuzzy Systems (FUZZ-IEEE) to celebrate the 50th Anniversary of Lotfi Zadeh's seminal 1965 paper on fuzzy sets. In addition to Lotfi's original paper, this note itemizes 100 citations of books and papers deemed "important (significant, seminal, etc.)" by 20 of the 21 living IEEE CIS Fuzzy Systems pioneers. Each of the 20 contributors supplied 5 citations, and Lotfi's paper makes the overall list a tidy 101, as in "Fuzzy Sets 101".

This note is *not* a survey in any real sense of the word, but the contributors did offer short remarks to indicate the reason for inclusion (e.g., historical, topical, seminal, etc.) of each citation. Citation statistics are easy to find and notoriously erroneous, so we refrain from reporting them – almost. The exception is that according to Google scholar on April 9, 2015, Lotfi's 1965 paper has been cited 55,479 times.

Keywords—fuzzy pattern recognition, fuzzy control fuzzy systems, fuzzy models, list of 101 fuzzy citations

I. WHAT WE TRIED TO DO

We begin with a number of disclaimers about what this article is, and is not. First of all, we recognize that any list such as this is completely arbitrary, probably biased, certainly subjective, and open to argument for any number of valid reasons. Our 101 list is presented in the same spirit as lists such as " the 10 best retirement cities in Europe," the "5 greatest guitar players of all time,", "the 20 best Australian beers," and so on, that are easily found in popular newspapers, magazines and websites. For example, a Google search for "10 best vacations" returned *About 128,000,000 results (0.51 seconds)* on October 15, 2014. The travel channel lists Cancun, London, Miami, Myrtle Beach, New York, Orlando, Paris, Rome, San Francisco as the top 10. National Geographic publishes a book titled "The 10 best of everything: The Ultimate Guide for Travellers." And so on. Recognizing the obvious, we have added some supplemental citations and additional remarks in the last section of this article to compensate for the obvious deficiencies of this – or indeed *any* - such list.

You may ask "the 101 best books and papers according to whom?" And in fact one of our pioneers *did* ask this very question, and refused to participate because he felt such lists were completely arbitrary and therefore entirely useless. Well, perhaps they are - *is there* any value to such a list at all? If you believe that history is important – that the way forward is in some sense better understood if presaged by an understanding of the road already travelled, our list may be helpful.

In the age of internet search, we know that all these references are at your fingertips – as long as you ask the right question or know what to look for. Our hope is that the citations given here encourage you to move in a direction you may not have been interested in before seeing them.

II. HOW THE LIST WAS BUILT

Table I lists the 23 pioneers, arranged in the chronological order in which they received the award.

2000	Lotfi Zadeh
2000	Michio Sugeno
2001	Jim Bezdek
2002	Didier Dubois
2002	Henri Prade
2003	Ebrahim J. Mamdani (D)
2004	Ronald Yager
2005	Enric Trillas
2006	Janusz Kacprzyk
2007	James M. Keller
2007	George Klir
2008	Jerry M. Mendel
2008	Takeshi Yamakawa
2009	Enrique H. Ruspini
2009	Tomohiro Takagi
2010	Hideo Tanaka (D)
2011	Hans J. Zimmermann
2012	Piero P. Bonissone
2012	Abraham Kandel
2013	Witold Pedrycz
2014	Masaharu Mizumoto
2015	Nikhil R. Pal
2015	Dimitar Filev

TABLE I. THE IEEE CIS FUZZY SYSTEMS PIONEERS: D~DECEASED

Here is our collection algorithm. Each of the 21 living pioneers was invited to submit up to five citations subject to these constraints: (i) no more than three self-citations; (ii) no more than one citation involving another pioneer; and (iii) at least one citation for a non-pioneer. The response was hardly uniform! Indeed, Abe Kandel, one of the 2012 pioneers, refused to participate at all. Here is his statement of declination, reproduced verbatim from his email to us dated September 7. 2014: Abe wrote:

"I am very sorry but I will decline this invitation due to the following reasons: 1) The concept of an Important publication is not really well defined. Important to whom, the author? His friends ? His students ? World peace ? Applications to improve society ? Etc. 2) why as fuzzy logicians we select a binary number of 100 ? What about the paper in location 101 ? And why not 1000 or just 3 "most important"? 3) who made US [eds: "US" is not the USA here]- the fuzzy pioneers the "God of Fuzziness " to make these kinds of decisions ? Why not to include also other very good and promising researchers in the field. Just because we were on this bus does not imply anything as evaluators in this entirely fuzzy process. I think that we should all consider this Idea and not just spend 5 minutes as recently suggested."

And who's to say Abe is wrong? Some pioneers supplied five citations, some supplied less than five citations, and of course the two deceased pioneers supplied none. We exercised our editorial prerogative to fill in the empty slots. Some of the explanatory comments supplied to us were too long or seemed confusing, so in a few cases, we edited them for brevity and/or clarity. Finally, there is little value in knowing which pioneer suggested which citations, so that information is not reported here.

Section III contains the 101 citations and remarks, ordered alphabetically and within author, chronologically, by the last name of the first author. The references are given in a modified form of the standard IEEE format which we think is self-explanatory, brief, and enables alphabetization. Abbreviations for commonly occuring journals in the citations are listed in Table II.

TABLE II. ABBREVIATIONS USED IN THE 101 LIST

FSS	Fuzzy Sets and Systems
IJAR	International Jo. of Approximate Reasoning
IJGS	International Jo. of General Systems
IJIS	International Jo. of Intelligent Systems
IJMMS	International Jo. of Man-Machine Studies
JMAA	Jo. Math Analysis and Applications
TC	IEEE Transactions on Computers
TCS	IEEE Transactions on Circuits and Systems
TEC	IEEE Transactions on Evolutionary Computation
TFS	IEEE Transactions on Fuzzy Systems
TNN	IEEE Transactions on Neural Networks
TPAMI	IEEE Transactions on Pattern Analysis and Machine Intelligence
TSMC	IEEE Transactions on Systems, Man and Cybernetics

III. THE 101 CITATIONS IN ALPHABETICAL ORDER

Our list of 101 begins with the root paper:

Zadeh, L. A., "Fuzzy sets," ***Information and Control*****, 8(3), 1965, 338-353.**

There is not much we can say about this paper that has not already been said. Without it, there is no 101 list, and many of us would be herding cows, painting houses, riding motorcycles, drinking beer (ok, some of us would be doing that anyway) or playing guitars in seedy juke joints. So, on to the subsequent 100 papers and books supplied by the 20 pioneers.

[1] Atanassov, K. T., "Intuitionistic fuzzy sets," *FSS*, 20, 87-96, 1986.

> In this paper Atanassov introduced his ideas about intuitionistic fuzzy sets to the fuzzy set community, and the basic definitions.

[2] Baldwin, J. F., "A new approach to approximate reasoning using fuzzy logic," *FSS*, 2, 1979, 309-325.

> This is one of the first papers, that focused on the extension of fuzzy logic to approximate reasoning on the basis of logical considerations. In contrast to fuzzy control, Baldwin used human argumentation rather than the control of artificial systems (machines etc.). It is still computationally simple and efficient and eventually led to the development of the fuzzy computer language ***Fril***.

[3] Bellman, R. E. and L. A. Zadeh, “Decision-making in a fuzzy environment”, *Management Sciences*, 17, 1970, 141-154.

> Presumably the most influential paper in the entire fuzzy sets literature, this article provides a simple yet extremely powerful fuzzy setting for all kinds of decision problems. It has inspired research in fuzzy decision making, control, optimization, and in a multitude of problems in which a choice is to be made under fuzzy goals, conditions, intentions, etc.

[4] Bezdek, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981.

> One of the first textbooks to present classical pattern recognition problems (clustering and classifier design) in the framework of fuzzy sets and models. Special emphasis on algorithms that use alternating optimization as a means for approximating solutions of fuzzy objective function problems.

[5] Bezdek, J. C., "On the relationship between neural networks, pattern recognition, and intelligence, *IJAR*, 6(2), 1992, 85-107.

> Perhaps the first publication to define and use the term "Computational Intelligence," subsequently adopted by the *Neural Networks Council* (NNC). The NNC attached the term to its triumvirate of flagship conferences (WCCI), and eventually changed their name to the IEEE Computational Intelligence Society. For more information on the history of this term and its relationship to the Canadian journal *Computational Intelligence* published by Wiley, visit ieee-cis.sightworks.net/documents/History/Bezdek-eolss-CI-history.pdf.

[6] Bezdek, J. C. and Harris, J. D., "Fuzzy partitions and relations: an axiomatic basis for clustering," *FSS*, 1, 1978, 111-127.

> This paper derives a hierarchy of fuzzy similarity relation spaces (FSRs) whose minimal member is the set of crisp equivalence relations, and whose maximal member is the set of $\max-\Delta$ transitive FSRs. A transformation of fuzzy partitions based on sum-min matrix multiplication is shown to induce a pseudo metric on the data.

[7] Bezdek, J. C. and R. J. Hathaway, "Clustering with relational c-means partitions from pairwise distance data, *Math. Modelling*, 9(6), 1987, 435-439.

> This paper introduced the idea of relational duals for the hard and fuzzy c-means algorithms. It is the basis for the branch of soft clustering that includes possibilistic and non-Euclidean versions of relational c-means.

[8] Bonissone, P. "Soft computing: the convergence of emerging reasoning technologies", *Soft Computing*, 1(1), 1997, 6-18.

> One of the first studies of Hybrid Soft Computing, jointly using fuzzy logic (FL), neural networks (NN) and genetic algorithms (GA). The paper presents several cases studies of hybridization of two or more soft computing techniques, such as the use of FL to control GAs and NNs parameters, the application of GAs to evolve NNs topologies or weights, or to tune FL controllers, and the implementation of FL controllers as NNs tuned by back-propagation type algorithm. This paper has inspired many other subsequent works in hybrid soft computing.

[9] Bonissone, P. and K. Decker, "Selecting uncertainty calculi and granularity: An experiment in trading-off precision and complexity", in *Uncertainty in Artificial Intelligence*, L. Kanal, and J. Lemmer (Eds.), 217-247, North-Holland, 1986.

> This paper is the first study of term sets granularity and triangular norms distinguishability. In the paper it is noted that, when using term sets typical for knowledge elicitation, many t-norms collapse into a small number of similarity classes. As a result, five t-norms are enough to cover most situations.

[10] Bosc, P. and O. Pivert, "SQLf: a relational database language for fuzzy querying," *TFS*, 3(1), 1995, pp. 1-17.

> This paper describes how to extend well-known languages and algorithms for handling queries to relational databases, when queries involve preferences described in terms of fuzzy sets.

[11] Bouchon-Meunier, B., Rifqi, M. and S. Bothorel, "Towards general measures of comparison of objects," *FSS* 84 (2), 1996, pp. 143-153.

> This paper is an extensive study on indices of similarity between fuzzy sets, that bridges the gap between the fuzzy set literature and the mathematical psychology literature on similarity.

[12] Buckles, B. P. and Petry, F. E., "A fuzzy representation of data for relational databases," *FSS*, 7(3), 1982, 213-226.

> One of the earliest and most influential papers on the use of fuzzy sets and models in the context of relational databases.

[13] De Luca, A. and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory", *Inf. and Control*, 20(4), 301-312, 1972.

> One of the earliest papers to consider the concept of entropy, defined in the context of fuzzy information.

[14] Dubois, D. and H. Prade, "Operations on fuzzy numbers," *Int. J. Systems Science*, 9(6), 1978, pp. 613-626.

> An influential paper in the arithmetic of fuzzy intervals, studying the four operations, as well as the maximum and the minimum. While the basic definitions were proposed by Zadeh and had been studied by some scholars in Japan and the United States, this paper proposed a parametric representation (LR-fuzzy numbers) of fuzzy intervals and showed how to compute practical results with it. It also proved a general shape-invariance result for the addition of fuzzy numbers.

[15] Dubois, D. and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, 1980.

> This is the first extensive monograph describing the state of the art of the field after 15 years of fuzzy set research. It covers all aspects of the theory and its applications and contains a very extensive list of references on fuzzy sets at the time. Moreover it provides for the first time extensive accounts on topics such as the arithmetic of fuzzy intervals and fuzzy analysis, possiblity theory and its relation to the theory of evidence, fuzzy linear programming and fuzzy logic control.

[16] Dubois, D. and H. Prade, "*Possibility Theory – An Approach to Computerized Processing of Uncertainty*", New York, London, Plenum Press, 1988.

> Possibility theory, independently outlined by the economist G. L. S. Shackle, and reintroduced on another basis by L. A. Zadeh (Fuzzy sets as a basis for a theory of possibility, *FSS* 1(1), 3-28, 1978), is an approach to the processing of epistemic uncertainty. This book describes and explains possibility theory from the underlying mathematics to database applications in a very concise and understandable way (with the collaboration of H. Farreny, R. Martin-Clouaire, and C. Testemale)

[17] Dubois, D. and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *IJGS,* 17(2-3), 1990, pp. 191-209.

> This paper shows that fuzzy sets and rough sets address different issues and are complementary. It applies for the first time the machinery of rough sets to fuzzy sets, thus yielding upper and lower fuzzy approximations, and replaces the equivalence relation underlying rough sets by a fuzzy similarity relation in the sense of Zadeh.

[18] Dubois, D., Lang, J. and H. Prade, "Possibilistic logic," in: *Handbook of Logic in Artificial Intelligence and Logic Programming*, D. M. Gabbay, C. J. Hogger, J. A. Robinson, D. Nute, eds., Oxford University Press, 3, 1994, pp. 439-513.

> This paper defines an extension of classical logic to the case where propositions have various levels of certainty. It is based on the old principle that the validity of a reasoning chain is the validity of its weakest link. The model-theoretic semantics is in terms of fuzzy sets of models. This logic is inconsistency-tolerant. This work demonstrates a close connection between fuzzy sets and the literature on nonmonotonic reasoning and belief revision in artificial intelligence.

[19] Dunn, J. C., "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *Jo. Cyber.*, 3(3), 1974, 32-57.

> The first paper to derive a fuzzy version of the classical batch hard c-means (aka k-means) clustering model and alternating optimization algorithm, which was subsequently generalized as described in [4].

[20] Filev, D. "Fuzzy modeling of complex systems", *IJAR*, 5(3), 1991, 281-290.

> This paper introduces state space and polytopic Takagi-Sugeno type models as alternative to the conventional dynamic state space models of nonlinear systems.

[21] Fodor, J. C. and M. R. Roubens, *Fuzzy Preference Modelling and Multicriteria Decision Support*, Springer Theory and Decision Library, 14, 1994.

This monograph achieved a breakthrough in the study of fuzzy relations meant to model the idea of preference, a topic pioneered by Sergei Orlowski in the 1970's. It relied on the state of the art in fuzzy aggregation operations, especially t-norms and co-norms, and applied it to the decomposition of a preference relation into its strict part, its equivalence part and its incomparability part. It shows the difficulty of carrying preference modeling techniques over to the valued case in the max-min setting, indicating the need for algebraic structures like MV-algebras.

[22] Goguen, J. A., "L-fuzzy sets," *JMAA*, 18, 1967, 145-174.

This paper laid bare the mathematical nature of fuzzy sets as mappings from a set to a complete lattice. It can be considered as the seminal work that motivated much of the mathematical literature on fuzzy sets.

[23] Goguen, J. A., "The logic of inexact concepts," *Synthese*, 19, 1968/69, 325-373.

This paper develops a remarkably sophisticated foundation for this new logic. The results provide the framework for both the development of Zadeh's agenda of fuzzy logic in the broad sense as well for the parallel development on the agenda of fuzzy logic in the narrow sense.

[24] Grabisch, M. and Labreuche, Ch. "Bi-capacities, Part I: definition, Möbius transformation and interaction", *FSS*, 151, 211-236, 2005.

Bi-capacities arise as a natural generalization of capacities (or fuzzy measures) in a context of decision making where underlying scales are bipolar. They are able to capture a wide variety of decision behaviours, encompassing models such as Kahneman and Tversky' Cumulative Prospect Theory. The paper extends all familiar notions used for fuzzy measures in this more general framework, and introduces the interaction index for bi-capacities, generalizing the Shapley value in a cooperative game theoretic perspective.

[25] Gustafson, D. E. and Kessel, W. C. "Fuzzy clustering with a fuzzy covariance matrix," *Proc. IEEE CDC*, 1979, 761-766.

This is the first fuzzy clustering model with an objective function that attempts to match cluster shapes by adapting the individual norm associated with each cluster. As alternating optimization proceeds, the norm associated with each cluster adapts to fit the local structure of the cluster via the fuzzy covariance matrix.

[26] Hajek, P. *Metamathematics of Fuzzy Logic*, Kluwer, Dordrecht, 1998.

This book is the culmination of seminal contributions by Peter Hajek to fuzzy logic in the narrow sense. The book is the first comprehensive axiomatic presentation of important fuzzy logics, each based on a distinct t-norm and its residuum, with the rigorous proofs that these fuzzy logics are both sound and complete. Contrary to other contributors to fuzzy logic in the narrow sense, Hajek has always considered advances in fuzzy logic in the broad sense as an important source of inspiration for research in fuzzy logic in the narrow sense.

[27] Hall, L. O.; Ozyurt, I. B. and J. C. Bezdek, "Clustering with a genetically optimized approach," *TEC,* 3(2), 1999, 103-112.

This paper introduces a new way to optimize a fuzzy objective function for clustering. The evolutionary approach is shown to provide partitions with a better optimized objective function value than the classical alternating optimization scheme.

[28] Herrera, F., Herrera-Viedma, E. and J. L. Verdegay. "A model of consensus in group decision making under linguistic assessments, *FSS*, 78 (1), 1996, 73-87.

This paper proposes the use of linguistic preferences to represent individuals' opinions, and a definition of fuzzy majority of consensus, represented by means of a linguistic quantifier. Several linguistic consensus degrees and linguistic distances are defined to indicate how far a group of individuals is from the maximum consensus, and how far each individual is from current consensus labels over the preferences.

[29] Inuiguchi, M. and Ramık, J. "Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problems, *FSS*, 111 (1), 2000, 3-28.

This survey reviews the application of possibility theory to fuzzy optimization, augmenting flexible constraints in fuzzy linear programming with uncertainty about coefficients, represented by fuzzy numbers. Then degrees of possibility and necessity of satisfying constraints can be used in the spirit of chance-constrained programming.

[30] Jang, J. S. R. "ANFIS: Adaptive-network-based-fuzzy-inference-system," *TSMC*, 23, 1993, 665–685.

This paper describes a very useful algorthm which is a staple in the Matlab™ Fuzzy Toolbox. The author presents a Takagi-Sugeno (TS) fuzzy system in network form and combines it with a back-propagation like learning algorithm to provide automated tuning of membership functions and polynomial coefficients. This algorithm enabled a large number of useful applications during the 1990s.

[31] Kacprzyk, J. Group decision making with a fuzzy majority, *FSS,* 18, 1986, 105-118.

Introduction of a fuzzy majority – equated with fuzzy linguistic quantifiers and dealt with in terms of a calculus of linguistically quantified propositions. One of the primary references for fuzzy models of group decisions, social choice, and voting schemes.

[32] Kacprzyk, J., *Multistage Fuzzy Control: A Model-Based Approach to Control and Decision-Making*, Wiley & Sons, 1997.

The first comprehensive coverage of multistage optimal fuzzy control, viz., fuzzy dynamic programming, for deterministic, stochastic and fuzzy systems under control. Real world applications include socio-economic regional development, and power systems planning.

[33] Kacprzyk, J., Zadrozny S., Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools. *Information Sciences*, 173 (4), 2005, 281-304.

This paper puts together an approach to linguistic summaries of databases after Yager and Zadeh's notion of protoform, in connection with the handling of queries in fuzzy databases.

[34] Kandel, A. *Fuzzy Techniques in Pattern Recognition*, John Wiley & Sons, New York, 1982.

One of the first comprehensive and pioneering treatises of the subject of pattern recognition in the framework of fuzzy sets. The fundamentals of fuzzy sets are discussed in the framework of constructive ways to use this technology to formulate and solve certain pattern recognition problems.

[35] Karnik, N., J. M. Mendel and Q. Liang, "Type-2 fuzzy logic systems," *TFS*, 7, 1999, 643-658.

This is a foundational paper that established many of the basic concepts in the field of type-2 fuzzy logic systems.

[36] Kasabov N. and Qun Song, “DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and its application for time-series prediction,” *TFS*, 10(2), 2002, 1-37.

This paper introduces a new type of fuzzy inference system, DENFIS, for adaptive on-line and off-line learning, and shows how to apply it to dynamic time series prediction.

[37] Kasabov, N., *Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering*, MIT Press, 1996.

This book provides an understandable approach to knowledge-based systems for problem solving by combining different methods of AI, fuzzy systems, and neural networks.

[38] Kaufmann A. *Introduction to the Theory of Fuzzy Subsets,* Academic Press, 1975.

This book is the English translation of the first monograph ever written (in French) on fuzzy set theory. It contains elementary definitions of fuzzy sets and related topics, covering the first papers by Zadeh, with special emphasis on max-min-transitive fuzzy similarity relations. This book is tutorial and contains many examples and exercises.

[39] Keller, J., and Hunt, D., "Incorporating fuzzy membership functions into the perceptron algorithm," *TPAMI*, 7(6), 1985, 693-699.

This paper develops a fuzzy perceptron model and algorithm that (unlike the classical crisp perceptron) terminates on non-linearly separable data sets. The article includes a proof of convergence for iterative optimization of the fuzzy perceptron objective function.

[40] Klement, E. P., Mesiar, R. and E. Pap “*Triangular Norms”*, Springer, 2000

This book gathers many mathematical results concerning fuzzy set connectives in an organized ways, with a stress on solving functional equations.

[41] Klir, G. J. and B. Yuan, “*Fuzzy Sets and Fuzzy Logic: Theory and Applications”*, Prentice-Hall, 1995.

A very complete and comprehensive textbook that covers the basic elements of fuzzy models from a mathematical point of view.

[42] Kosko, B. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Prentice Hall, 1992.

This book is of historical significance due to its important role in the genesis of neurofuzzy systems.

[43] Krishnapuram, R. and J. M. Keller, "A possibilistic approach to clustering", *TFS*, 1(2), 1993, 98-110.

This paper generalized (hard and fuzzy) c-means clustering by eliminating the constraint that the sum of cluster memberships for any object must equal 1. It also introduced the idea of a possibilistic partition as one consisting of typicalities.

[44] Kruse, R. and K. D. Meyer, *Statistics with Vague Data*, Springer, 1987.

This monograph develops an approach to fuzzy random variables originally proposed by Huibert Kwakernaak in the late 1970’s. A fuzzy random variable is viewed as an ill-known random variable in contrast with the Madan Puri - Dan Ralescu approach.

[45] Lee, S. C. and E. T. Lee, "Fuzzy neural networks," *Math. Biosciences*, 23, 1975, 151-177.

This was the first paper to define the idea of a fuzzy neuron as a generalization of the McCulloch-Pitts neuron. Although cast in the more formal language of automata theory, it is the first paper about a fuzzy neural network.

[46] Lin, C. T. and George Lee, C. S. , "Neural-network-based fuzzy logic control and decision system," *TC*, 40(12), 1991, 1320-1336.

This paper introduces an innovative five-layer neural architecture for realizing a fuzzy rule based system for control and other decision making applications. It uses a hybrid learning scheme involving an unsupervised phase for defining the membership functions and a supervised phase for refining neuro-fuzzy system is proposed.

[47] Mamdani, E. H. and Assilian, S. "An experiment in linguistic synthesis with a fuzzy logic controller," *IJMMS*, 7, 1975, 1-13.

The starting point of fuzzy control whose continuation was a turning point for the acceptance of fuzzy logic in engineering.

[48] Marinos, P. N., "Fuzzy logic and its applications to switching systems," *TC*, 18(4), 1969, 343-348.

This is the first paper that presents a technique for analysis and synthesis of fuzzy logic functions with implementation in terms of logic gates. This paper led to the implementation of real fuzzy information processing hardware systems such as high-speed fuzzy logic controllers.

[49] Mendel, J. M., *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*, Prentice-Hall, 2001.

This textbook offers comprehensive coverage of both type-1 and type-2 fuzzy sets and rule-based systems for singleton and non-singleton fuzzifications.

[50] Mendel, J. M. and R. John, “Type-2 fuzzy sets made simple,” *TFS*, 10, 2002, 117-127.

This paper provides a representation theorem that shows a new way to represent a type-2 fuzzy set in terms of simpler embedded type-2 fuzzy sets.

[51] Mizumoto, M., "Fuzzy controls by product-sum-gravity method dealing with fuzzy rules of emphatic and suppressive types," *Int. Jo. of Uncertainty, Fuzziness and Knowledge-Based Systems,* 2(3), 1994, 305-319.

This paper shows that emphatic or suppressive effects on fuzzy inference results are observed under the product-sum-gravity method by using fuzzy control rules whose consequent part is characterized by a membership function whose grades are greater than 1, or a negative-valued membership function. The use of negative-valued membership functions is beneficial to the construction of fuzzy control rules.

[52] Mizumoto, M. and Tanaka, K., "Some properties of fuzzy sets of type 2," *Inf. and Control*, 31(4), 1976, 312-340.

This paper investigates the algebraic structure of Type 2 fuzzy sets under set operations defined by means of the extension principle on fuzzy numbers on the unit intervals serving as fuzzy membership grades.

[53] Murofushi. T and Sugeno, M. "An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure", *FSS*, 29, 1989, 201-227.

This paper first showed with concrete examples that (1) a non-additive measure (capacity in the sense of Choquet or fuzzy measure in the sense of Sugeno) represents interactions among subsets and (2) the Choquet integral is a reasonable integral with respect to such a non-additive measure.

[54] Negoita, C. V. and Ralescu, D. A., *Application of Fuzzy Sets to Systems Analysis,* Wiley, 1975.

This is the first book written on the basics of fuzzy sets, fuzzy theories (categories, topologies, etc.) and fuzzy logic, and also their possible applications to systems, automata, clustering, etc.

[55] Nguyen H. T. "A note on the extension principle for fuzzy sets," *JMAA*, 64, 1978, 369-380.

This pioneering paper describes the connection between the extension principle and the calculation of functions with set-valued arguments using alpha-cuts. It shows that fuzzy number calculations commute with cuts.

[56] Pal, N. R. and J. C. Bezdek, "Measuring fuzzy uncertainty." *TFS*, 2(2), 1994, 107-118.

This paper introduces two new classes, additive and multiplicative classes, of measures of fuzziness, which satisfy the five axioms of such measures. This paper also introduces the concept of weighted fuzziness to incorporate subjectivity in measures of fuzziness.

[57] Pedrycz, W., "Algorithms of fuzzy clustering with partial supervision," *Pattern Recognition Letters*, 3, 1985, 13 - 20.

This paper introduced the concept of partial supervision for fuzzy clustering and proposed algorithms that used it to do clustering in presence of partially labeled data.

[58] Pedrycz, W., *Fuzzy Control and Fuzzy Systems*, John Wiley. 1991

This research monograph is one of the first comprehensive and innovative publications that focuses on fuzzy control and fuzzy systems within a framework of fuzzy relational equations.

[59] Puri, M. L. and D. A. Ralescu, "Fuzzy random variables," *JMAA*, 114 (2), 1986, pp. 409-422.

This seminal paper proposed a mathematical extension of the theory of random sets to fuzzy random sets, 10 years after pioneering but largely ignored works by Robert Féron. In this approach, a fuzzy random variable is viewed as a mapping from a probability space to a space of membership functions, equipped with a suitable metric structure. Since then many scholars have followed this line to handle random linguistic variables.

[60] Rodriguez, R. O., Esteva, F., Garcia, P. and Godo, L., "On implicative closure operators in approximate reasoning," *IJAR*, 33, 2003, 159-184.

This paper clarifies the notions of graded implication and, through the imposition of reasonable constraints, characterization of the nature of implication operators.

[61] Ruspini, E. H. ,"A new approach to clustering," *Inform. Control*, 15(1), 1969, 22–32.

This is the first paper to define the notion of a fuzzy c-partition of data. As such, it is the root paper for the entire field of fuzzy clustering, which is now a very large part of the pattern recognition landscape.

[62] Ruspini, E. H., "On the semantics of fuzzy logic," *IJAR*, 5, 1991, 45-88.

This paper presents a formal characterization of the major concepts and constructs of fuzzy logic in terms of notions of distance, closeness, and similarity between pairs of possible worlds. The similarity logic developed in the paper allows a form of logical "'extrapolation'" between possible worlds. It is shown to have connections with possibility theory, in the setting of metric spaces.

[63] Ruspini, E. H., P. Bonissone, and W. Pedrycz, *Handbook of Fuzzy Computing*, Institute of Physics, 1998.

A handbook on fuzzy sets, systems, and applications that was state of the art in 1998. Co-edited by three fuzzy pioneers, it offers a coherent presentation and notation across multiple entries, which were written by a large number of other fuzzy pioneers.

[64] Saffiotti, A., Konolige, K., and Ruspini, E. H., "A multivalued logic approach to integrating planning and control," *Artificial Intelligence*, 76, 1981, 481-522.

This paper presents the first significant application of fuzzy logic methods to the planning and control of autonomous robots. This approach led to the SAPPHIRA architecture, which, until recently, was employed in many commercial autonomous mobile robots. The multilevel hierarchical, supervisor-controller, architecture introduced in this paper has been widely applied to other control systems.

[65] Sanchez, E. "Resolution of composite fuzzy relation equations," *Inf. and Control*, 30, 1976, 38-48.

This is the first, highly original and influential publication in the area of fuzzy relational equations. It is the root paper for a large ongoing research effort in relational theory.

[66] Seki, H. and Mizumoto, M., "On the equivalence conditions of fuzzy inference methods -part 1: Basic concept and definition," TFS, 19(6), 2011, 1097-1106.

This paper addresses equivalence conditions of a number of fuzzy inference methods such as the product-sum-gravity method, simplified fuzzy inference method, fuzzy singleton-type inference method, SIRMs inference method, and SIC inference method.

[67] Sugeno, M. *Theory of Fuzzy Integrals and Its Applications*, Ph.D. Thesis, Tokyo Institute of Technology, 1974.

Starting point of the fertile subject of fuzzy measures and the so-called Sugeno's integral. This paper introduced a family of measures (the lambda ones) that are either additive, or subadditive, or superadditive.

[68] Sugeno, M., and Yasukawa, T., "A fuzzy-logic-based approach to qualitative modeling. " *TFS*, 1(1), 1993, 7-31.

This paper proposes a two-step process, fuzzy modelling and its linguistic approximation, to generate qualitative models of systems based on input-output numerical data. Although the primary emphasis of this is on qualitative modelling, it also introduces another very important concept, the use of clustering to find fuzzy rules from numerical data, which drastically reduces the complexity of fuzzy rule generation.

[69] Tahani, H. and J. Keller, "Information fusion in computer vision using the fuzzy integral", *TSMC*, 20(3), 1990, 733-741.

This was the first journal paper (preceded by 2 conference papers) that framed the pattern recognition problem in terms of fuzzy integral fusion of information.

[70] Takagi, T. and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control", *TSMC*, 15, 1985, 116-132.

This paper established a link between conventional and fuzzy systems models and paved the way for the use of machine learning and control theory in fuzzy systems.

[71] Tanaka, H., Uejima, S. and Asai, K., "Linear regression analysis with fuzzy model," *TSMC*, 12(6), 1982, 903-907.

This paper was the first to propose a study of *fuzzy linear regression* (FLR), by adding fuzziness to regression analysis. It considered parameter estimation of FLR models under two factors: (i) the degree of fit; and (ii) the vagueness of the model. This paper inspired many subsequent works in fuzzy linear regression models.

[72] Tanaka, K and Wang, H., *Fuzzy Control Systems Design and Analysis: A Linear Matrix Inequality Approach*, John Wiley & Sons, 2004.

This book offers a systematic approach to the analysis and synthesis of stable fuzzy control systems based on Takagi-Sugeno type models.

[73] Trillas, E. and Riera, T. "Entropies in finite fuzzy sets", *Inf. Sciences,* 15(2), 1978, 159-168.

This paper first studied fuzzy entropies which are different from a Shannon-type (employed by De Luca and Termini) and considered relations between entropies and fuzzy integrals.

[74] Trillas, E. and Valverde, L., "On mode and implication in approximate reasoning," In *Approximate Reasoning and Expert Systems* (M. M, Gupta, A. Kandel, W, Bandler, and J. B. Kiszka, eds.), 1985, 157-166.

Trillas and Valverde's paper clarifies the nature of fuzzy implication - a central concept in fuzzy logic - while producing representation theorems that clearly define the proper structure of implication operators.

[75] Valverde, L. , "On the structure of F-indistinguishability operators," *FSS*, 17, 1985, 313-328.

Valverde's paper on the structure of fuzzy similarities brought clarity, through a principled approach, to the structure of fuzzy preorders and fuzzy similarity equations. Furthermore, this paper clarified the relationship between the notions of fuzzy preference and fuzzy similarity.

[76] Wang, L-X., and Mendel, J. M., "Generating fuzzy rules by learning from examples," *TSMC*, 22(6), 1992, 1414-1427.

This paper proposes a useful scheme for generating a fuzzy rule based system from numerical data for function-approximation type systems. It also proves that such a fuzzy rule based system has the universal approximation capability, which can approximate any nonlinear continuous function on a compact set to an arbitrary accuracy.

[77] Wang, L.-X. , "Fuzzy systems are universal approximators," *Proc. FUZZ-IEEE*, 1992.

This paper provided the first rigorous proof that a Mamdani fuzzy logic system is a universal approximator. cf. E. P Klement, "Are fuzzy systems universal approximators? *IJGS*, 28(2/3), 1999, 259-282.

[78] Wang, X., De Baets, B. and E. Kerre. "A comparative study of similarity measures," *FSS*, 73 (2), 1995, 259-268.

A systematic study of the notion of similarity between fuzzy sets and the properties of such similarity indices. This is used as a basis for defining a notion of approximate equality between fuzzy sets.

[79] Yager, R. R., "On a general class of fuzzy connectives," *FSS*, 4, 1980, 235-242.

One of the earliest papers to provide a generalization of the union and intersection operators used in fuzzy sets.

[80] Yager, R. R., "A procedure for ordering fuzzy subsets of the unit interval,", *Inf. Sci*., 24, 1981, 143-161.

This early work deals with the issue of comparing fuzzy sets of the unit interval, with an approach compatible with fuzzy arithmetics.

[81] Yager, R.R., "A new approach to the summarization of data", *Inf. Sci.*, 28, 1982, 69-86.

A breakthrough paper that introduces the concept of a linguistic data summary, which is equated to a linguistically quantified proposition with a fuzzy linguistic quantifier. As opposed to linguistic summarization of data (previously known for many years), this scheme accounts for imprecision in data and enables us to grasp the very essence of data in a human consistent way.

[82] Yager, R.R., "On Ordered Weighted Averaging aggregation operators," *TSMC*, 18, 1988, 183-190.

Perhaps the central paper that fueled many subsequent studies of aggregation functions in fuzzy logic. OWA operators were, and are, widely used in various applications.

[83] Yager, R. R., "Quantifier guided aggregation using OWA operators, *IJIS*, 11, 1996, 49-73.

In this paper Yager provides an approach for going from a linguistic specification of an aggregation imperative to its manifestation in terms of an OWA operator. It gives us an example of the concept of computing with words applied to aggregation.

[84] Yager, R. R. and Filev, D. P., "*Essentials of Fuzzy Modeling and Control",* John Wiley, 1994.

A textbook containing a systematic approach to fuzzy models and control, methods for developing and learning fuzzy models from data, and their applications.

[85] Yamakawa, T., "High-speed fuzzy controller hardware system : The mega-FIPS machine," *Inf. Sci*., 45, 1988, 113-128.

This article describes a high-speed fuzzy controller hardware system which facilitates approximate reasoning at 1,000,000 FIPS (fuzzy inferences per second). This was the first step in an approach to a fuzzy computer.

[86] Yamakawa, T., "A fuzzy inference engine in nonlinear analog mode and its application to a fuzzy logic control," *TNN*, 4(3), 1993, 496-522.

This is a tutorial on the utility of fuzzy systems that provides a broad scope overview of analog mode hardware.

[87] Yamakawa, T., "Silicon implementation of a fuzzy neuron," *TFS*, 4(4), 1996, 488-501.

This paper describes a fuzzy neuron chip which modifies an ordinary neuron model by fuzzy logic and facilitates high speed recognition (less than 0.5 microseconds) of handwritten characters.

[88] Zadeh, L. A., "Similarity relations and fuzzy orderings," *Inf. Sci*., 1971, 177-200.

The first paper that showed how to decompose a fuzzy similarity relation to discover cluster substructure in a partition tree on relational (usually dissimilarity) data. Also introduced the idea of transitive closures for fuzzy similarity relations.

[89] Zadeh, L. A., "Fuzzy logic and approximate reasoning," *Synthese* 30, 1975, 407-428.

This paper introduces two basic formalisms: fuzzy logic and approximate reasoning. Basically, fuzzy logic is a system of reasoning and computation in which the objects of reasoning and computation are classes with unsharp (fuzzy) boundaries. Fuzzy logic is much more than a logical system.

[90] Zadeh, L. A. "Outline of a new approach to the analysis of complex systems and decision processes", *TSMC*, 3(1), 1973, 28-44.

This paper introduces to the concepts of fuzzy systems, algorithms, models, and optimization from the perspective of conventional systems theory. It is the genesis of the fuzzy logic control literature.

[91] Zadeh, L. A. "The concept of a linguistic variable and its application to approximate aeasoning," Parts 1-3, *Inf. Sci*., p1: 8, 1975, 199-249; 1975, p2: 8, 301-357; 1976; p3: 9, 43-80.

This three part publication develops the definition, theory and applications of linguistic variables for use in approximate reasoning. It is a superb treatment of an integral component of all subsequent work in fuzzy logic, linguistic data processing, and computing with words.

[92] Zadeh, L. A., "A theory of approximate reasoning," in *Machine Intelligence,* 9, Hayes, J., Michie, D., and Mikulich, L. I., Eds., ed New York: Halstead Press, 1979, 149-194.

In this paper Zadeh very elegantly puts together many of his ideas on approximate reasoning in a wholistic framework. It provides the basis of much of Zadeh subsequent work on computing with words.

[93] Zadeh, L. A., "Fuzzy sets and information granularity," in Advances in Fuzzy Set Theory and Applications, eds. M. Gupta, R. Ragade and R. R. Yager, North Holland, 1979, 3-18.

This paper introduces the concept of granularity and relates it to information. Fuzzy granularity is a concept which is unique to fuzzy logic. The linguistic variable is a granular variable. A concept which is introduced in this paper is that of a fuzzy-set-value random variable, with fuzzy probabilities. This concept provides a basis for a generalization of the Dempster-Shafer Theory of Evidence.

[94] Zadeh, L. A., "Precisiation of meaning via translation into PRUF," In *Cognitive Constraints and Communication*, Vaina, L. and Hintica, J. (eds.), D. Reidel, Boston, 1984, 372-402.

This is the best paper that clearly and completely describes one of Lotfi Zadeh's greatest ideas - the one of precisiating the meaning of utterances in natural language by translating them into the meaning representation language PRUF. The language is based on a fuzzy-set interpretation of the theory of graded possibilities, whose expressive power is comparable to that of natural languages.

[95] Zadeh, L. A., "Fuzzy logic = computing with words," *TFS*, 2, 1996, 103-111.

Prof. Zadeh led the fuzzy community with innovative ideas that possessed deep insights. There were two phases: 1) propose fuzzy sets and their mathematical foundations and 2) propose of the idea "computing with words," which had significant value in expanding fuzzy logic from a scientific tool to the liberal arts. This was the first paper in that direction.

[96] Zadeh, L. A., "Generalized theory of uncertainty (GTU) - Principal concepts and ideas," *Comp. Stat. and Data Analysis*, 51, 2006, 15-46.

A basic premise in this paper is that there are many different kinds of uncertainty. The three principal kinds are possibilistic uncertainty, probabilistic uncertainty and bimodal uncertainty. GTU addresses the three principal kinds and others. GTU is a challenge to the Bayesian doctrine which posits that any kind of uncertainty can and should be dealt with through the use of probability theory. GTU has a unique capability--the capability to compute with probabilities, possibilities, events and relations which are described in natural language.

[97] Zadeh, L. A., "Towards a restriction-centered theory of truth and meaning (RCT), *Information Sciences,* 248, 2013, 1-14.

This paper is a radical departure from traditional approaches to representation of meaning and definition of truth. The meaning of a proposition is expressed as a restriction. A proposition is associated with two truth values: internal truth value and external truth value.

[98] Zimmermann, H.-J., "Fuzzy programming and linear programming with several objective functions". *FSS*, 1, 45-55, 1978.

This paper paved the way for many developments and applications in Operations Research. For example, classical linear programming requires crisp constraints that are often unrealistic. This paper showed how to soften the constraints, obtaining a more realistic model.

[99] Zimmermann, H.-J.. and Zysno, P., "Latent Connectives in Human Decision Making," *FSS*, 4, 1980, 37-51.

A paper published before t-norms and t-conorms were broadly used in fuzzy logic. The authors showed that fuzzy connectives cannot belong to universal classes, but should be contextually chosen. It also suggested the use of aggregation functions.

[100] Zimmermann, H.-J., "*Fuzzy Sets, Decision Making, and Expert Systems"*, Kluwer, 1987.

This book was one of the first texts that discusses how modeling with mathematics and empirical findings can be used to turn expert systems based on classical dichotomous logics into fuzzy expert systems.

IV. Discussion and Supplemental Reading

The 101 list was compiled using a very constrained method of sampling (i.e., only IEEE CIS pioneers were consulted). Consequently we feel justified in expanding the list a bit by adding some remarks and citations that might otherwise go unrecognized.

(i) Many important papers have been written in fields that are not directly germane to engineering applications. As you might expect, since our contributors are IEEE pioneers, this 101 list is heavily weighted towards the theory and applications in pattern recognition and control. However, there are very important papers in areas that might be called "pure mathematics, logic, philosophy, etc." such as topology, category theory, etc., that fall outside the natural interests of most members of a professional engineering society. Here are a few early citations, in chronological order, which were overlooked by our IEEE pioneers:

R. Lowen "A comparison of different compactness notions in fuzzy topological spaces," *JMAA*, 64, 1978, 446- 454.

S. Rodabaugh, "The Hausdorff separation axiom for fuzzy topological spaces," *Topology and its Applications*, 11, 1979, 225-233.

Pu Pao-Ming, Liu Ying-Ming, "Fuzzy topology. I. Neighborhood structure of a fuzzy point and Moore-Smith convergence," *JMAA*, 76, 1980, 571-599.

A. Di Nola, A. G. S. Ventre, "On some chains of fuzzy sets," *FSS*, 4, 1980, 185-191.

S. Gottwald "Fuzzy propositional logics", *FSS*, 3, 1980,181-192.

E. P. Klement "Construction of fuzzy σ-algebras using triangular norms," *JMAA*, 85, 1982, 543-565.

U. Höhle, "Fuzzy measures as extensions of stochastic measures," *JMAA*, 92, 1983, 372-380.

M. Togai and H. Watanabe, "A VLSI implementation of a fuzzy inference engine: Toward an expert system on a chip,", Inf. Sciences, 38(2), 1986, 147-163.

P. Diamond, P. Kloeden "Metric spaces of fuzzy sets", *FSS*, 35(2), 1990, 241-249.

(ii) There are some general papers and books which are not specifically related to fuzzy sets that nevertheless had an important impact on many people within our community. A very few of them are listed here:

G. Choquet, "Theory of capacities", *Annales de l'Institut Fourier*, 5, 1953/54, 131-295.

B. Schweizer and A. Sklar. "Associative functions and abstract semi-groups", *Pub. Math. Debrecen*, 10, 1963, 69-81.

R. Moore "*Interval Analysis*", 1966, Prentice-Hall, Englewood Cliffs N.J..

R. O. Duda and Hart, P. E. "*Pattern Classification and Scene Analysis*, 1973, John Wiley and Sons, NY.

G. Shafer. "*A Mathematical Theory of Evidence*", 1976, Princeton University Press.

(iii) There are also some works that do not appear in the 101 list because they were not necessarily foundational (at least, for the 20 contributing IEEE pioneers). We want to mention three of them here, with historical footnotes of a sort, that explain in part why we wanted to include them.

W. G. Wee and K. S. Fu. "A formulation of fuzzy automata and its application as a model of learning systems," *IEEE Trans. Syst. Science and Cyberns,* 5(3), 1969, 215-223.

> K. S. Fu was one of the really important "big guys" in the early history of fuzzy sets. The importance of his interest in the field at a time when it was quite embryonic and survival was a real issue cannot be overstated. He strongly encouraged the publication of the book [15]. He was also the first president of NAFIPS, the *North American Fuzzy Information Processing Society*, which in turn was the first professional society whose primary focus was fuzzy sets and models. As an example of his breadth of interest, this paper was a very early contribution to learning systems – now a hugely important field. Fu's student Bill Wee wrote the first PhD thesis on fuzzy pattern recognition, published just two years after Lotfi's 1965 paper.

A.Rosenfeld. "Fuzzy digital topology", *Information and Control*, 40, 1979, 76-87.

> Azriel Rosenfeld was a second "big guy" who helped keep the wolves from Lotfi's door in the early days. Rosenfeld, his students, and some of his colleagues produced a number of early papers on fuzzy graph theory, fuzzy geometry, and the use of fuzzy models in image processing. This paper is an early example.

R.L.P. Chang and T. Pavlidis. "Fuzzy decision tree algorithms," *TSMC*, 7(1), 1977, 28-35.

> Theodore Pavlidis was a third influential supporter who encouraged scientists and engineers to have an open mind about fuzzy sets. His stewardship of the *IEEE Transactions on Pattern Analysis and Machine Intelligence*, inherited from K. S. Fu, offered an important early repository for emerging research in various fuzzy disciplines. This paper of his about fuzzy decision trees was perhaps the first of its kind, but notice it appeared in another IEEE Transactions, *TSMC*, whose editor at that time was Andrew Sage, yet a fourth patron saint for early workers in fuzzy sets.

(iv) The above lists, mainly oriented towards papers by "IEEE pioneers", do not give much credit to the newer generation of fuzzy set researchers, active in the last twenty years. We could give yet another partially arbitrary list recognizing these newer papers and books, but refrain from doing it. Yet they fully belong to the posterity of Zadeh's 50-year-old paper. Many of them are named on the editorial boards of the numerous fuzzy sets and soft computing journals.

We conclude with this observation. Instead of the method of collection used here, we could have polled the past presidents of IFSA (the *International Fuzzy Systems Association*), all the editors of FSS, or only those researchers working in business, or for a government. Each poll would produce a somewhat different list. The intersection of our 101 list with any of these lists would in all likelihood not be empty. But, for example, a list of the 100 most cited references in fuzzy sets, would certainly not coincide with our 101 list either. But ... which citation engine? – each one would undoubtedly produce a slightly different set of rankings. Carrying this argument to its logical conclusion, there can obviously be an infinite number of lists, no two of which coincide. We can only hope that this list is of some value to readers and attendees at the 2015 FUZZ- IEEE.

That's all, Folks!