Chapter 14
Questions on Uncertain, Possible, and Probable

In plain language, the predicative words "possible" and "probable" are often used synonymously, but without argument for it, and what seems more suitable is to say of something probable that it is possible, but not reciprocally. In language, "possible" seems to be more largely applicable than "probable". For instance, if when throwing a single die it is probable to get five points, it is clear that it can be expected because this output is among "those that are possible"; obtaining eleven points is not possible, and consequently there is no sense in attributing to it the property of being probable. In addition, there are also life's ordinary situations linguistically qualified as "possible but not probable", even if they can actually have a very small probability. For instance, it is possible that in a few minutes my old friend John, 10 years older than I am, and from whom I have heard nothing in the last 10 years, can call me by phone, but it deserves to be qualified as something improbable or, at most and if John is still alive, with a very small probability provided it could be effectively computed.

The theoretical distinction between possible and probable can be seen, in principle, in the different axioms with which the mathematical theories on the measures of probability by Kolmogorov in 1933, and on the measures of possibility by Zadeh in 1978, formalized the measuring of probability and possibility, respectively. Notwithstanding, these theories suppose that the elements to which the predicative words probable and possible are applied, belong to actually strong types of lattices with a negation, something that, in plain language and as was said, is odd, risky, and even dangerous to always suppose it. Let's begin with an overview of these theories that, nevertheless, refer to the concept of probability and possibility but not, directly, to the use of the words "probable" and "possible" in plain language but, at most, in some particular and specialized part of it.

14.1.   Kolmogorov established his theory of probability on the following hypotheses:

1. A measure of probability, prob, assigns a number between 0 and 1, to "events" represented by subsets in a Boolean algebra on a set Ω included in the power-set 2^X of some universe X, prob: $\Omega \rightarrow [0, 1]$, and such that
2. $\text{prob}(\Omega) = 1$.

This presumes that all the laws valid in the Boolean algebra Ω are applicable to the "events", something supposed as actually happening; in particular, it is supposed the existence of perfect classifications, or partitions, of the nonempty subsets in Ω containing several elements.

3. The essential axiom for the mapping prob is its additive law,

$$\text{prob}(A \cup B) = \text{prob}(A) + \text{prob}(B),$$

provided the intersection of the "events" A and B were empty, $A \cap B = \emptyset$; that is, $\{A, B\}$ is a perfect classification or partition of $A \cup B$.

From this follows $\text{prob}(A') = 1 - \text{prob}(A)$, and $\text{prob}(\emptyset) = 0$. With it, is proven that prob is a measure for the graph $(\Omega, \subseteq)$ with the minimum $\emptyset$, and the maximum Ω, and thanks to the existence of relative complements in the Boolean algebras. That is, and as formerly proven, $A \subseteq B \Rightarrow \text{prob}(A) \leq \text{prob}(B)$.

Thus, probabilities can be seen in Boolean algebras as (additive) measures of the predicative word "probable" applied to the events, and by supposing that $\emptyset$ is the less probable subset, Ω the most probable of them, and that "A is less probable than B" is just done by $A \subseteq B$, identifying $<_{\text{probable}}$ with $\subseteq$; that (with finite subsets) "less elements" is equivalent to "less probable", with $\emptyset$ minimal, and Ω maximal.

All that comes from the fact that, genetically, probabilities come from problems in which the "represented events" can not only be precisely named and counted, but also perfectly classified into crisp parts. In addition, in Kolmogorov's interpretation of probability, it should be taken into account that events are not only considered random ones, in the sense of being obtainable by indefinitely repeating an experiment under exactly the same conditions at each repetition, but that frequencies of its appearances can be somehow computed.

Kolmogorov's theory arises from a crisp objectivistic interpretation of probability through random experiments; in itself, it is but an abstract mathematical theory of (normalized) additive measures, and, today is not only the most widely known interpretation of probability, but is also responsible for most of the applicative successes of mathematical statistics. The interpretation by Kolmogorov reflects the famous and wise statement, "Nothing is more practical than a good theory."

Notwithstanding, not all the successful measures in the applications are additive; for instance, the big family of Sugeno's λ-measures, m, verifying:

$$m(A \cup B) = m(A) + m(B) + \lambda.m(A).m(B), \quad \text{with } -1 < \lambda, \quad \text{if } A \cap B = \emptyset,$$

contains additive measures if it is $\lambda = 0$, but with $\lambda < 0$ has super additive and with $\lambda > 0$ has subadditive measures.

In "reality" there are relevant situations in which information, or evidence, is not so precise for allowing its splitting in separate pieces, and it should be remarked that, in language, the term "probable" is not only applied to precise, but also to imprecise, statements that cannot be represented by sets, a case in which the immediate applicability of the Kolmogorov theory is at least dubious; the existence of crisp partitions is so in some nebulous situations where what appears is not a clearly separable mixing of information. Sometimes the measure of a "totality" is greater than the sum of the measures of its parts, and sometimes it is less. There are also "events" whose repetition under the same exact conditions is not possible; even in some situations, former instances mean nothing for the next one.

14.2. Zadeh established his theory on the measures of possibility with fuzzy sets in the unique basic fuzzy algebra (BAF) that is a lattice, the De Morgan algebra, $([0, 1]^X; \min, \max, 1 - \mathrm{id})$; but, namely and for what concerns its basic hypotheses, it can be considered in an abstract De Morgan algebra $(M; \leq; 0, 1; \cdot, +;')$, where the order $\leq$ is that of the lattice, and supposing that all its elements can be qualified as "possible". These hypotheses are just the following.

A measure of possibility is a mapping $\pi: M \rightarrow [0, 1]$, such that:

- $\pi(0) = 0$,
- $\pi(1) = 1$, and
- $\pi(a + b) = \max(\pi(a), \pi(b))$, for any pair a, b in M, regardless of being $a \cdot b = 0$, or not.

Because $a \leq b$ is equivalent to $a + b = b$, it follows that $\pi(b) = \pi(a + b) = \max(\pi(a), \pi(b)) \geq \pi(a)$; thus, π can be seen as a measure of the predicative word "possible" when applied to the elements in a De Morgan algebra, and with qualitative meaning $(M, \leq)$, once the relation $\leq_{\text{possible}}$ is identified with the lattice's order $\leq$ of the De Morgan algebra.

Note that π is not additive but subadditive, because it verifies $\pi(a + b) = \max(\pi(a), \pi(b)) \leq \pi(a) + \pi(b)$; totalities measure less than their parts. Regarding $\pi(a')$, it not always holds its equality with $1 - \pi(a)$, that cannot be a law, because for any Boolean element a in M (i.e., verifying $a + a' = 1$), it follows that $\pi(a + a') = \pi(1) = 1 = \max(\pi(a), \pi(a')) \leq \pi(a) + \pi(a')$, and thus, $1 - \pi(a) \leq \pi(a')$.

An advantage of Zadeh's theory is that it models the applicability of "possible" to elements that do not need to be in a Boolean algebra, as is the case with membership functions of fuzzy sets, and for which in principle no crisp partition exists; neither the existence of crisp partitions, nor that the negation is a Boolean complement, is previously supposed. A serious disadvantage is, nevertheless and for plain language, that De Morgan algebras are still too strong algebraic structures with a lattice basis, in which some of their properties, such as they are the

conjunction's commutative law and the distributive laws, cannot always be presumed in language.

In both Boolean and De Morgan algebras, the weight of laws of a syntactic origin is actually too strong for language; this occurs, for example, in Boolean algebras with the law of "perfect repartition", $a = a \cdot b + a \cdot b'$. It is a law that neither holds in all ortholattices, nor in De Morgan algebras and that, as shown in the BAFs, forces that conjunction and disjunction are not dually linked. In addition, double negation, $(a')' = a$, and duality $(a + b)' = a' \cdot b'$, are laws not always verifiable in plain language. In De Morgan algebras, distributive laws are among those that cannot always be supposed in plain language.

Hence, both theories of Kolmogorov and Zadeh are only applicable to those parts of language in which their presumed laws can be accepted. Both presume that the relations $<_{\text{probable}}$ and $<_{\text{possible}}$ coincide with the partial order of the corresponding lattice.

Anyway, Zadeh's theory is not fully objectivistic as is Kolmogorov's theory; it neither supposes nor excludes that the statements whose possibility is to be measured should represent events obtainable in experiments repeatable under the same conditions; it admits events for which just some precise or imprecise information is known and can be represented by statements generating linguistic collectives or fuzzy sets. It does not force that statements should be endowed with an ortholattice algebraic structure, even if supposing it is a De Morgan one where some strong laws hold, such as the distributive ones. It proceeds through a different weakening of Boolean algebras than the quantum calculus does.

In addition, and as is well known, a membership function taking the value 1 at some point can be interpreted as a distribution of possibility conditioned by the previously available information on the use of its linguistic label but not, in general, as a probability distribution.

Were each fuzzy set's membership function μ equal to a probability p_μ (in a universe X being a Boolean algebra, or in an orthomodular lattice), because

$\mu \leq \lambda \Leftrightarrow \mu(x) \leq \lambda(x),$ for   all x in $X,$ it   is   also
$\mu(x') \leq \lambda(x') \Leftrightarrow 1 - p_\mu(x) \leq 1 - p_\lambda(x) \Leftrightarrow p_\lambda(x) \leq p_\mu(x),$ or $\lambda(x) \leq \mu(x) \Leftrightarrow \lambda \leq \mu,$

it would follow that $\mu = \lambda$; that is, the pointwise ordering between these fuzzy sets collapses in the identity. It follows a rare ordering of membership functions, under which two of them only can be coincidental or not comparable. With it, many useful applications of fuzzy sets would be lost. The pointwise ordering is not a natural form for ordering probability measures. Interpreting fuzzy sets in a Boolean algebra, or in an orthomodular lattice, as a measure of probability just leads to a very odd "theory" for both fuzzy sets and probabilities.

Notwithstanding, this does not mean that each particular numerical value of a fuzzy set cannot be obtained as the value of a probability. For some concretion, given the finite fuzzy set in $X = \{1, 2, 3, 4\}$,

$$\mu = 0.5/1 + 0.7/2 + 1/3 + 0/4,$$

there are many quadruplets (p_1, p_2, p_3, p_4) of probabilities, each able to give the corresponding value of μ; for instance:

$$\mu(1) = 0.5 = p_1(1); p_2(2) = 0.3; p_1(3) = 0.2; p_1(4) = 0,$$
$$p_2(1) = 0.2; \mu(2) = p_2(2) = 0.7; p_2(3) = 0; p_2(4) = 0.1,$$
$$p_3(1) = 0; p_3(2) = 0; \mu(3) = p_3(3) = 1; p_3(4) = 0,$$
$$p_4(1) = 0.5; p_4(2) = 0.2; p_4(3) = 0.3; \mu(4) = p_4(4) = 0.$$

This simple example shows that there is nothing against those cases in which each numerical value of the membership function comes from a specific random variable, that for some fuzzy sets it seems possible to design a series of random experiments from whose respective probabilities the values of its membership function can be obtained; something that, in general, is impossible by means of a single probability and corresponds to a statistical view based on managing random variables. Anyway, what is not clear enough is on which characteristics of the fuzzy set's context, such random experiments, or random variables, could be linked; it seems dependent on how the contextual information could be acquired.

In short, what cannot be excluded at all is the possibility of obtaining the values of a fuzzy set's membership function through a statistical methodology. But, in any case, there cannot be coincidence between a fuzzy set and a probability defined on the same ground. The theories of fuzzy sets and probabilities have different goals, although fuzzy sets and possibility measures have closer goals.

14.3.   As has been shown, Kolmogorov's theory can be imported (with some modifications) into algebraic structures weaker than Boolean algebras, as it is into orthomodular lattices in the so-called "quantum probability calculus," corresponding to a particular form of language in which statements are precise even if not all the information on them is known, and where, instead of crisp partitions $p = q + r$ with $q \cdot r = 0$, it is supposed with $q \leq r'$ that, only provided q and r were Boolean elements, would coincide with $q \cdot r = 0$; in orthomodular lattices, contradiction ($q \leq r'$) implies incompatibility ($q \cdot r = 0$), but not reciprocally. These are two concepts only coincidental in Boolean algebras and thanks to its basic perfect repartition law, but that in language (and even in most algebras of fuzzy sets) are actually independent of each other. It should be pointed out that the perfect repartition law comes directly from distributivity:

$$1 = p + p' = > q = q \cdot (p + p') = q \cdot p + q \cdot p'$$

and that distributivity is not valid in orthomodular lattices; perfect repartition is also not always valid in De Morgan algebras because in them $p + p' = 1$ is not a law.

The quantum case shows how the information available on what is stated affects the laws that can be supposed, but it should be noted that, in plain language, they

cannot be supposed to hold and that, when dealing with one of its parts, some previous checking of them is necessary. It can be said that fuzzy logic is the first approach breaking, in language, the usually supposed lattice structure of statements; something reflected, for instance, in the many applications where the conjunction is represented by the product instead of the minimum, and for which the usually presumed syntactic property $p \cdot p = p$ is lost.

In addition, it is also worth noting that in the so-called intuitionistic lattices (in which negation is not strong) a kind of relational, or conditional, probability can be defined, but that (as was proven) such lattices are included through a previous equivalence in a Boolean algebra of classes.

14.4. Once a short review of the mathematical theories of probability and possibility is done, let's overview the uses of possible, probable, and uncertain, in plain language.

As has been repeatedly said, in general, plain language is not submitted to all the laws of algebraic structures such as the algebras of Boole, De Morgan, and the like. Introducing an algebraic structure and formalizing it means constraining plain language to an artificial one. Plain language is almost all we count with to express what we experience with the senses, and what we elaborate intellectually; it has great and essential flexibility, and for "structuring language" it is necessary to consider only some of its specific parts, submit them to a regimentation, forget some particularities, and finally adopt a mathematical model that, like all models, at the end is but a more or less valid simplification of reality. In the particular case of a lattice's model, it is extremely rigid; for instance, it supposes the conjunction to be the greatest statement among all those from which the two statements submitted to conjunction follow forward. It also supposes some laws that are not always valid for statements such as the associative laws of conjunction and disjunction, or the existence of two statements considered as "neutral" and "absorbent" for the conjunction, and presuming that a clear and general concept of truth and falsehood is known in the corresponding piece of language. Syntax is important, but what really matters in language is semantics; a badly syntactically constructed linguistic statement is often comprehended well enough, but confusion of meaning is what actually produces a serious lack of comprehension.

What follows tries to shed some light on the words *uncertain*, *possible*, and *probable*, but without imposing too many laws and towards a not still existing total scientific domestication of the uses of these words in plain language, something for which there is yet a lack of practical knowledge.

14.5. What does it mean that something can be qualified as uncertain in plain language? Uncertain, as the opposite word of certain, implies (without a necessary equivalence) not-certain; hence, uncertain refers to some lack of certainty with respect to aspects surrounding what is stated but, instead, not-certain means a total lack of certainty. For instance, the statement, "It is uncertain that candidate C will win the election," just means that the statement "Candidate C will win the election" is, for what is contextually and currently known as certain, not necessary and

unsafe; that, if betting on it, any of those knowing what surrounds the election will risk a very low wager by believing that the bet can be easily lost. It is not a properly objectivistic view, but a subjective one, based on the experience of the player.

Anyway, for capturing a full-meaning of the predicative word U = uncertain in a universe X of statements p, q, r and so on, a quantity $(X, <_U, m_U)$ should be specified, and even an analysis of what certain means seems to be previously necessary for linking its qualitative meaning with the opposite word uncertain in the form $<_{\text{uncertain}} = <_U = <_{\text{certain}}^{-1}$.

Note that the coincidence of the relation $<_U$ (less uncertain than) with the order of a lattice (be it a Boolean, a De Morgan algebra, or whatever ordered structure) is but a supposition involving the hypothesis that X can be endowed with such an algebraic structure, something on which one must be, for what has been commented, extremely cautious and check, previously, if each of the laws defining such a structure can actually be supposed. To specify a quantity reflecting a full-meaning of U, and as happens with all words, some "experience" of the use of uncertain in plain language is necessary for establishing, at least, the empirical relation "less uncertain than", knowing which statements are maximal, which are minimal, and specifying a measure.

It can be supposed that in language the word uncertain inherits a "history relative to its use," coming from some experiences linked with contexts in which uncertain has been formerly used, and acquired either directly or by means of some "contagious" contact, oral or written, with others using it.

The relation $<_U$ will always depend on the context surrounding the use of U and, for instance, supposing it coincides with the order of a lattice (always verifying $p \leq q \Leftrightarrow p + q = q \Leftrightarrow p \cdot q = p$) could be very risky because, for instance, in language and due to time intervention, it can be $p \cdot q \neq q \cdot p$. Additionally, in language the exclusive disjunction is often managed $p \; \Delta \; q = (p + q) \cdot (p \cdot q)'$, instead of the inclusive disjunction $p + q$, but $p + q = q$ implies $p \Delta q = q \cdot (p \cdot q)'$ that, in a Boolean algebra means $p \Delta q = q \cdot p'$, and that because $q \cdot p' \leq p'$ is $p \Delta q$ contradictory with p and only coincides with q if it is $q \leq p'$, that is, if p and q are contradictory.

Inasmuch as no specific and satisfactory general theory of uncertainty comprising its measuring is currently well known, there is room for interpreting the word *uncertain* in the form each can be contextually able to do. For instance, those mastering probability theory or possibility theory tend to identify uncertain with probable or, respectively, with possible, even if obviously uncertain is more general than probable and possible; something probable or possible is uncertain, but the reverses are not clear enough. For instance, under which random experiment can a probability be computed for "Candidate C will win the election"? Of course, there is neither a way of designing experiments such as that of throwing a die, nor of considering a Boolean or a De Morgan algebra containing the presumed results of such an experiment, and less again ways of repeating it exactly under the same conditions at each repetition; for instance, a former election even with the same

candidates will not present the same surroundings of the next and, at least, some ideal similarities should be chosen.

It is debatable if Kolmogorov theory is directly applicable to this kind of "events", and, for instance, Zadeh introduced a calculus of probabilities for imprecise events represented by fuzzy sets but showing difficulties for defining conditional probability, such as also happens with quantum probability. In addition, and in the case of precise words, the identification of $<_U$ with the inclusion relation among subsets $\subseteq$ leads to identifying "less uncertain than" with "less elements than", perhaps a too risky identification. Possibility theory that involves subjective views could be more suitable even if its results can be subjected to a large lack of certainty.

14.6.   Regarding the use of $P = $ "probable" in language, for capturing its full-meaning the relation $<_P = <_{probable}$ should be captured specifying a measure m_P. What is obvious is that when the considered statements are imprecise, they cannot constitute a Boolean algebra, nor an orthomodular lattice, but that what is open is, in some cases, a De Morgan algebra, and even a different BAF. When the statements are precise, a Boolean algebra of sets representing them exists and hence what should be debated is if relation $<_P$ is, or is not, the inclusion of sets ($\subseteq$), and if there exists an additive measure m_P. What can be easily accepted is that $\subseteq$ is contained in $<_P$, but the reciprocal is in general dubious; thus, because $A \subseteq B$ implies $A <_P B$, it follows that $m_P(A) \leq m_P(B)$, provided m were a measure for $<_P$; $<_P$—measures are $\subseteq$—measures, but it could not be guaranteed that a $\subseteq$—measure is always $a <_P$—measure. Hence, it is not for sure than a probability could always, and in plain language, be a good enough measure of the word "probable", and less again that such measures should be additive.

For instance, to identify what "It is probable that candidate C will win the election" (shortened to "C is P") means it is first necessary to know the relationships "C is P is less probable than D is P", for all the candidates D in the election process; and once the graph (Candidates, $<_P$) and its maximal and minimal elements are known, more information is required on the context surrounding the election for specifying a measure. This does not mean, of course, that such a measure cannot be estimated by a statistical methodology.

There is a basic difference between $<_P$ and the relation of "comparative probability", introduced by T. Fine; such a difference lies in that comparative probability is a total or linear order, but $<_P$ cannot be always supposed to be so. Perhaps Fine's relation could be viewed as an extension of $<_P$; it deserves further study, but seems related to the linear relation of working meaning $<_m$ previously introduced for measures, and enlarging $<_P$.

14.7. Of course, similar comments can be made concerning the word $\prod = $ possible, as modeled in Zadeh's theory of possibility, and the use of the word $\prod$ in language by a full-meaning specified by a quantity $(X, <_{\prod}, m_{\prod})$ in which neither $<_{\prod}$ should necessarily coincide with the order of a De Morgan algebra, nor $m_{\prod}$

necessarily verifies the axioms of a possibility measure. Namely, that it does not always grow under the max operation.

Even accepting that "if probable, then possible, but not reciprocally", it does not imply that given a measure of probability prob and one of possibility poss, it should always be prob$(x) \leq$ poss(x), for all x in a universe where both measures prob and poss are defined; it can be easily checked in a finite universe. For this reason, Zadeh introduced the concept of consistent pairs (poss, prob): such a pair of measures is consistent if it is prob $\leq$ poss.

Such a concept is in agreement with the chain of inclusions, $<_P \subseteq <_\Pi \subseteq <_U$, concerning the qualitative meanings of the three words, and also with the typical expression "if it is probable, then it is possible", and in any case, "it is uncertain". In a finite Boolean algebra it is easy to prove there are probabilities that coincide with possibilities, but it is limited to degenerate measures, that is, those only taking the values 0 and 1; essentially, probabilities and possibilities are different measures. To specify a probability more information is needed than for specifying a possibility. $m_P \leq m_\Pi \leq m_U$ seems that it could be a coherence relation among the respective measures.

14.8. The ordinary use in language of the words *uncertain*, *possible*, and *probable*, is not yet well known as is that of the word *big* in a closed interval of the real line; it requires a kind of experimental work that, for instance, could help to shed some light into the debate on the different views of probability between the Kolmogorovians, frequentialists, or objectivists, and the Bayesians or subjectivists. If the first base their view in the previous application of probability to precise words denoting random events represented in Boolean algebras of crisp sets, the second even try to apply probability to imprecise words neither denoting random events nor representable in those algebras, nor in more general ortholattices; Bayesians seem to be closer to the use of "probable" in plain language and ordinary reasoning than are Kolmogorovians.

The essential point for objectivistic interpretation lies in decomposing events into disjoint events to allow accepting the additive law of probability. But, with imprecise events represented by membership functions of fuzzy sets, the situation is different because having a decomposition $\mu = \alpha + \beta$ with $\alpha \cdot \beta = \mu_0$ (the function constantly equal to zero, representing the empty set) implies working in a BAF that, if functionally expressible, requires solving the functional system of equations $a = G(b, c)$ and $F(b, c) = 0$.

For instance, in a standard algebra $([0, 1]^X, F, G, N)$, where G is a continuous t-conorm, F a continuous t-norm, and N a strong negation function, F should be a t-norm in the family of Lukasiewicz, $F = W_\Delta$ (with $\Delta: [0, 1] \to [0, 1]$ an order automorphism), because these continuous t-norms are the only ones with zero-divisors, and then $\Delta^{-1}(\max(0, \Delta(x) + \Delta(y) - 1)) = 0$ means $\Delta(y) \leq 1 - \Delta(x)$, or $y \leq N_\Delta(x)$, implying that the fuzzy sets α and β should be contradictory ($\beta \leq \alpha'$) in respect to the negation N_Δ.

That is, for counting with such kinds of "partitions", a very particular algebraic structure seems to be necessary; note that t-norms W_Δ are the only t-norms verifying

the principle of noncontradiction by conjunction ($\alpha \cdot \alpha' = \mu_0$) among the standard BAFs. These algebras are not lattices except if F = min and G = max, in which case they are De Morgan algebras. With these structures, one of the problems in plain language is the identification, inside a current problem, of a statement playing the role of "absorbent" for the conjunction, that is, being represented by the fuzzy set with membership function μ_0.

All the above needs to be searched in language through designing controlled processes of experimentation, and checking what is observed against some of the known mathematical models, as well as for getting hints towards establishing new models. For it, a search of the Web, as done in a recent paper by Sergio Guadarrama, Eloy Renedo, and myself, for studying how the linguistic conjunction "and" is actually used in language, could be a useful starting methodology. Without systematic observation in language, controlled experimentation, and adapting mathematical models to it, everything is but a kind of play with abstract ideas, perhaps of wishful thinking. It is difficult to imagine how a purely abstract formal kind of reasoning could arrive at, for instance, thermodynamics, and without not blind observation and controlled experiments. Our subject is empirical, and formalism is for helping its comprehension and for allowing computations.

Perfect partitioning is anchored in liking absolutes; a crisp set's partition suffices to name its parts by precise words specifying them. Perfect classification is an epitome for a "principle" of separation or isolation; that is, the belief that everything that is supposed to be composed can actually be separated in isolated parts whose union constitutes the composed totality, and that things grow by a clear superposition of pieces. It is an ideal that goes well with a lattice-type conception of language: a rigid conception that is artificial because, for instance, in language (and in life) time intervenes and situations are not always static; language is essentially dynamic. For instance, language also contains the opposites of words that, being several, sometimes are not comparable among them, and not coincidental with their negation; sometimes there is interplay with the given word, and not allowing perfect classifications. Zadeh's linguistic variables are a good example of it.

In language and reality there are many cloudy, gaseous like appearances, for which a principle of crisp separation is difficult to conceive, unless some crisp fixation, or representation of them, is accepted and that, in fact, does not appear in such a reality. Reality should be seen as it is, and represented in forms as close as possible to what it actually is; confusing name and thing is, jointly with believing that a name defines a real thing, just a dangerous "philosophical nominalism" based on believing that the universe of discourse can be perfectly classified in those x such that "x is P", and those such that "x is not P", whose intersection is empty, in which each set is isolated from the others. Paraphrasing Luigi Pirandello, a scientist should escape from words in search of where they can be applied; this could be metaphysics, but it is not science.

As commented before, what cannot be excluded is to modify measures by assigning, to what is to be measured, not a number in the real line but a mathematical object in a not linearly structured set. One reason for such a possible modification comes from the mentioned fact that often $<_P$ is not linear, but $<_m$ is

always so. For instance, and on one hand, it could be thought of as assigning to the measures values being complex numbers, or intervals in the real line, with which $<_m$ ceases to be linear and more possibilities for its coincidence with $<_P$ are open, although no safety on it can be stated. However, it often also happens that the available additional information on the behavior of P only allows us to recognize at each x in X an upper limit (b_x) and a lower limit (a_x) of the values m can take at point x, and then, defining $m(x) = [a_x, b_x]$, or $m(x) = a_x + ib_x$, cannot seem bizarre.

This comment opens the door to consider (when it can be suitable) what, in fuzzy set theory, are called type-two fuzzy sets, consisting in assigning fuzzy numbers in [0, 1] instead of crisp numbers in the unit interval; of course, crisp numbers and intervals are particular cases of fuzzy numbers and, in addition, type-two fuzzy sets are not only more general, but can immediately represent usual linguistic statements such as "the measure is high", by representing the linguistic label "high" as a fuzzy number in [0, 1], that is, by a membership function μ_{high} of the linguistic label "high" in such an interval, and as they are, for instance,

$$\mu_{high}(x) = x,$$

or

$$\mu_{high}(x) = 0 \quad \text{if} \quad 0 \leq x < 0.8,$$

and

$$\mu_{high}(x) = 1 \quad \text{if} \quad 0.8 \leq x < 1,$$

with the second equivalent to the interval [0.8, 1].

With this, several forms of interpreting and representing "high" are possible, and open a window not only for representing and measuring more statements in natural language, but also perhaps for enlarging the application of probabilities when both its arguments and values are themselves imprecise or uncertain. This is, for instance, the case of the typical linguistic statement, "It is with high probability that John is rich," that could be translated into fuzzy terms by

$$\text{prob}(\mu_{rich}(\text{John})) = \mu_{high},$$

once a theory of probability for imprecise events and whose values are fuzzy numbers can be established. Nevertheless, the laws under which such "prob" should be defined are not actually known. This is a very important topic towards computing with words.

14.9. This chapter is not, by its own nature and like the others in this book's Part II, a conclusive one. It only offers a reflection for exciting its potential readers' interest through some patterns relative to a new view of the meaning of words in plain language, and without conceiving this concept as one crisply definable by necessary

and sufficient conditions, but by quantities each specifiable through the contextual available information on the action of the word in the universe of discourse and, also, by adding reasonable hypotheses on it when necessary for its design.

In this view, meaning is not seen as a universal concept, isolated from both the context in which the word is used and the purpose of such use. The meaning of a word in a universe of discourse is not unique, but there are many possible meanings in each universe and, consequently, what cannot be thought is a single and universal meaning of the words *uncertain*, *possible*, and *probable*, like there is not a single probability for the events appearing in throwing a die, but several depending on the information available and acquired on the die by, for instance, through a nonde-structive analysis of it.

There is uncertainty even in specifying a measure, be it uncertain, or possible, or probable. Almost everything in plain language is endowed with uncertainty, and a point on which there is a coincidence between what is presented in this chapter and the Bayesian theory of probability is that everything depends, at least, on the previously available information on the subject currently under study. Measures depend on some a priori information on what surrounds the use of the corre-sponding word; indeed, each measure can be viewed as "conditioned" by such information, and, when it changes, an updating of what has been previously pre-sumed is necessary. A relevant difference between the Bayesian approach to probability, and the meaning concept in plain language, as it is approached here, is that the first mainly refers to events admitted as something that could actually happen, but meaning refers to the use of words in language that, often enough, not only refer to "physical" situations, but to virtual or informative ones, and that can count with a history of social uses. If defining a probability requires a strong algebraic structure, in plain language it does not seem to exist in a universal way; in each case, and as fuzzy logic shows, a particular algebraic structure should be searched for.

What is not done here is the more than debatable identification of uncertainty and risk; the risk taken when acting with uncertain knowledge is not considered as it properly belongs to the field of decision theory.

In sum, language is extremely complex; in the era of information it is, perhaps, the most complex system computer science is faced with, and that for compre-hending well what information is, needs to be scientifically domesticated inasmuch as information is basically conveyed by a language, be it plain or artificial, and employed for reasoning.

14.10. A comment trying to approach the quantity's model of meaning to the Bayesian interpretation of probability is still in order, even if it remains just as a naïve trial.

In such an interpretation, and once a priori nonnull probabilities $p(a)$ and $p(b)$ are established, the probability of a can be updated after knowing that of b, by means of the well-known Bayes' formula, $p(a/b) = p(a \cdot b)/p(b)$, in which the hidden con-ditional "If b, then a", is conjunctively represented by $a \cdot b$, and not by $b' + a$, as is always done in Boolean algebras but that does not allow $p(./b)$ to be a probability. It

shows that the interpretation of the formula does not properly concern the formal Boolean language of the model, but to something external to the model, that is, in the corresponding plain language. Note that it is $p(a/b) \leq p(a) \Leftrightarrow p(a \cdot b) \leq p$ (a). $p(b)$ and $p(a) < p(a/b) \Leftrightarrow p (a \cdot b) > p(a)$. $p(b)$. In any case, it seems that Bayesians could keep some linking with the presented conception of meaning and hence it deserves to be explored a little bit further.

Suppose a word P in X, with meaning given by a quantity $(X, <_P, m_P)$. Suppose that there is a word Q, with meaning $(X, <_Q, m_Q)$, such that it can be asserted "If x is Q, then x is P", for all x in X. Once this is known, is the a priori meaning of P affected? Provided the answer were affirmative, how would it be affected? Is there a new quantity $(X, <_P^*, m_P^*)$, giving an a posteriori meaning, conditioned by the information the conditional can furnish? How is such a new meaning defined?

Even without fully answering this question, let's turn around the conditionals by denoting p and q, respectively, the statements "x is P", and "x is Q". Thus, the first question is how to represent the conditional, or inference, $q < p$, that is, how to link all that with inference. What can be said about the character of an inference once $q < p$ is understood as a statement? Which property, or properties, should a degree of truth t enjoy for saying something, given the truth degree $t(q)$ of the antecedent, on the truth value $t(p)$ of the consequent p?

By presuming that t is such that $q < p \Rightarrow t(q) \leq t(p)$, the degree of true of the conclusion is greater than or equal to that of the antecedent; deduction propagates the truth by nondecreasing it. Note how counterintuitive it would be to suppose $q < p \Rightarrow t(p) \leq t(q)$, propagating the degree of true by retroceding instead of advancing it, something actually odd. What it yet lacks is studying the possibility of having formulae for the degrees $t(q < p)$, depending on $t(q)$ and $t(p)$, as they exist in the classical case in probabilized Boolean algebras by means of the Bayes' formula.

It is interesting, for instance, to bound the truth value of q starting from the modus ponens (MP) "inequality", $q \cdot (q < p) < p$, implying $t(q \cdot (q < p)) \leq t(p)$. Provided it were functions f and g such that,

$$t(q \cdot p) = f(t(q), t(p)), \quad \text{and} \quad t(q < p) = g(t(q), t(p)),$$

the inequality $f(t(q), g(t(q), t(p)) \leq t(p)$ would follow, and perhaps a bounding for t (p) only depending on $t(q)$ could be obtained. For instance, if as in the classical case, it were $f(a, b) = \min(a, b)$, and $g(a, b) = \max(1 - a, b)$, from

$$\min(t(q), \max(1 - t(q), t(p)) \leq t(p) \Leftrightarrow \max(\min(t(q), 1 - t(q)), \min(t(q), t(p))) \leq t(p),$$

the low bounding of $t(p)$: $\min(t(q), 1 - t(q)) \leq t(p)$ would follow.

Those expressions depend on the meaning, in the corresponding language's context, of the linguistic connectives *and, or, not,* whose specification is actually a contextual and partially open problem as formerly said.

In the limit situation in which all the statements are precise, and t can be taken as a Kolmogorov probability, $t = $ prob: Ω à $[0, 1]$, in a Boolean algebra Ω of subsets in the universe X, it appears an interesting case concerning the meaning of the

conditional. Provided it were prob(q) > 0, the Bayes' formula prob(p/q) = prob ($q \cdot p$)/prob(q) would represent the probability of $q < p$, of "p if q", a situation in which such a corresponding linguistic conditional cannot be interpreted in Ω as the material form $q' + p$, because it does not give a probability in Ω, as prob(p/q) does.

The conditional, once interpreted by "$p \cdot q$" actually facilitates a probability provided, as in a Boolean algebra, that the conjunction (·) were commutative. If prob were a probability in Ω, then prob(./q) would also be a probability in the set of the traces ($p \cdot q$) all its elements p show in q; that is, prob(./q) would also be a probability but in the restricted Boolean algebra constituted by the elements in $\Omega^* = \{p^* = p \cdot q; p$ in $\Omega\}$. In some sense, q represents a kind of diaphragm only allowing us to consider what "is inside q"; the antecedent constitutes the new universe, hiding what is out of it, and where, for instance, the negation of p is not p', but $p^\wedge = p' \cdot q$. Thus, it is

$$q^\wedge + p* = q' \cdot q + p \cdot q = 0 + p \cdot q = p \cdot q;$$

in Ω^*, inside the new and restricted universe q, but not outside; the Boolean and the conjunctive conditionals do coincide.

All this can suggest a view of measuring conditionals in forms similar to t ($q < p$) = $t(q \cdot p)/t(q)$, without t being a probability but just a particular measure of "true". Note, for instance, that in the Bayes' formula the conjunction ($p \cdot q$) is supposed commutative because the statements/events are precise/classical sets, but that, when the statements are imprecise/fuzzy sets, such commutative property cannot be generally presumed. The values $t(q)$ and $t(q \cdot p)$ can be called the a priori measures, and $t(q < p)$ =: $t(p/q)$, will be the a posteriori ones.

Provided the conjunction were commutative, it would be easy to find the truth value $t(p < q)$ in function of $t(q < p)$, because $t(q < p) \cdot t(p) = t(q \cdot p) = t(p \cdot q) = t$ ($p < q$) $\cdot t(q)$, or

$$t(p<q) = [t(p)/t(q)] \cdot t(q<p),$$

a formula expressing the truth degree of the inverted conditional, and showing that t ($p < q$) coincides with $t(q < p)$ if and only if $t(q) = t(p)$. Of course, it also shows that $t(p < q) \leq t(q < p) \Leftrightarrow t(p) \leq t(q)$. The case in which the conjunction is not commutative is an open one.

In some cases, the function $t(p/q)$ =: $t(q \cdot p)/t(q)$ is also a measure in the restricted universe q, inside q, and not only a simple degree of truth of $q < p$. In fact, it is:

1. $p_1 < p_2 = > q \cdot p_1 < q \cdot p_2 = > t(q \cdot p_1) \leq t(q \cdot p_2) = > t(p_1/q) \leq t(p_2/q)$;
2. If p is an a priori working antiprototype, that is, $t(p) = 0$, it is also an a posteriori working prototype, because $q \cdot p < p => t(q \cdot p) \leq t(p) = 0$ implies $t(q \cdot p) = 0$, and $t(p/q) = 0$. Nevertheless, a different problem arises when p is an a priori prototype, because from $t(p) = 1$ it is not immediate to arrive at, or to define, $t(p/q) = 1$, and adding some conditions seems to be necessary for

proving it. For instance, a sufficient condition is the validity, for the pair (q, p), of the law $t(p \cdot q) + t(p + q) = t(p) + t(q)$, because then from $t(p) = 1$ and "$p < p + q \Rightarrow t(p + q) = 1$", it follows that $t(p \cdot q) = t(q)$, and finally $t(p/q) = t(q)/t(q) = 1$. Note that the additive law follows in the particular case with $p \cdot q = 0$; hence such law implies the additive one is more general.

It should be remarked that a priori values are not always attainable from past historical data records; this is the case in some linguistic cases, such as the afore-mentioned "Candidate X will win the election" when, for instance, such a candidate participates in an election for the first time. In such cases, a priori values could be approximate through the experience, or subjective evidence, the observer can (directly or indirectly) have on the corresponding situation; furthermore the contextual knowledge (or evidence) can lead to attributing a value to the a priori ones. Only if there is "contextual ignorance", $t(q) = t(q')$, can it be accepted to take the value $t(q) = 0.5$, because such equality, and provided it were $t(q') = 1 - t(q)$, would imply $t(q) = 1/2$.

If the previous evidence is represented by e, the prior formula can be written in the form

$$t(p/q \cdot e) =: t_e(q < p) = t(p \cdot (q \cdot e))/t(q \cdot e),$$

and only in the ignorance case is it acceptable to take $t(q \cdot e) = 0.5$ that, because from $p \cdot (q \cdot e) < q \cdot e$ it follows that $t(p \cdot q \cdot e) \leq t(q \cdot e) = 0.5$, and the conclusion that follows, $t_e(q < p) \leq 1$, gives no actual information inasmuch as t_e is always less than or equal to one.

For instance, using additional information to specify either a measure m for a predicate P, or a degree t for the validity of a linguistic expression, is nothing more than a priori information (perhaps supplied by an expert, and not always coming from numerical data records; hence, and in a Bayes-style line of thought, it seems necessary for upgrading a priori measures. The use of additional a priori information, or evidence, seems to be a proper and unavoidable resource for the mathematical modeling of commonsense reasoning.

Nevertheless, the measuring of the extent to which a conditional can be a valid one in plain language is indeed an open problem and, when no numerical records are known, it should rely on some subjective information. Note that holding $q \cdot p < p$, and hence $t(q < p) \leq t(p)$, from the former expression follows the upper bound

$$t(q < p) \leq \min(1, t(p)/t(q))$$

for the measure of the conditional.

14.11. For ending this chapter, let's add a comment on conditional statements $p < q$, not from the point of view of deducing, but from that of guessing, and further than was advanced in the former Sects. 5.4 and 11.1. That is, by avoiding deducing from either modus ponens $(p \cdot (p < q)) < q$, or modus tollens (MT)

$(q' \cdot (p < q)) < p'$, and analyzing what can be conjectured or refuted from the sets of premises $\{p, p < q\}$, or $\{q', p < q\}$.

A conjecture and a refutation of the first set of premises (MP) are, respectively, elements c and r, such that $p \cdot (p < q) \leq / c'$, and $p \cdot (p < q) < r'$; concerning the second set of premises (MT), a conjecture is an element d such that $q' \cdot (p < q) \leq /d'$, and a refutation is an s such that $q' \cdot (p < q) < s'$. The problem consists in finding these kinds of four elements.

In principle, it seems easier to confront the refutations, and, concerning the conjectures, those in which it is, respectively, $c' < p \cdot (p < q)$, and $d' < q' \cdot (p < q)$. To "solve" these inequalities, it is necessary to count with a calculus in an algebraic framework, in which both the symbol < and an expression with connectives equivalent to $p < q$ could be identified.

For instance, in the framework of a Boolean algebra, from $q' \cdot (p' + q) \leq r'$, equivalent to $q' \cdot p' \leq r'$, or to $r \leq q + p$, it follows that what refutes a conditional are those r that are below the union of the antecedent and the consequent. Analogously, from $d' \leq q' \cdot (p' + q) \Leftrightarrow d' \leq q' \cdot p' \Leftrightarrow p + q \leq d$ follows that the elements that are greater than the union of antecedent and consequent are conjectures.

14.12. What about the counterfactual conditionals? This is a strongly semantic and scarcely syntactic subject in which background knowledge is, in addition to context and purpose, essential. Without imagining a nonexisting actual situation where the antecedent could be true, the conditional could lose sense.

For instance, and returning to the former example,

– $t(p < q) = \max\ (1 − t(p),\ t(q))$, provided $t(p) = 1$, it would follow that $t(p < q) = t(q)$: If the antecedent is true, the truth value of the conditional coincides with that of the consequent; both should be simultaneously true or false. But, were the antecedent false, $t(p) = 0$, it would follow that $t(p < q) = 1$: a false antecedent produces a true conditional.
– With the conjunctive representation the situation changes for what refers to false antecedents, because $t(p < q) = t(p \cdot q) = \min(t(p), t(q))$, and $t(p) = 1$ implies $t(p < q) = t(q)$, as in the former case, but $t(p) = 0$ implies $t(p < q) = 0$: the conditional only can be true provided its antecedent and consequent were true.

Anyway, with counterfactuals $p < q$ the sources for the truth or falsity of p and q are actually different: p comes from an imaginary source, but q from a real one. It seems that for the analysis of counterfactuals, the usual Boolean model is insufficient, even if the conjunctive one can seem to facilitate a more realistic representation. Of course, there are many more ways of interpreting $p < q$ in Boolean algebras, even without a single expression; for instance, identifying $p < q$ with

– $p' + q$, if p is contradictory with q, and
– $p \cdot q$, if they are not contradictory,

an interpretation satisfying the modus ponens inequality. Namely, in Boolean algebras, the equivalence $p \cdot (p \to q) \leq q \Leftrightarrow p \to q \leq p' + q$, allows us to construct expressions such as the former easily; for instance,

– q, if $p = 0$, and
– $p \cdot q$, otherwise.

In Boolean algebras there are many possible representations of the conditionals, and finding some of them appropriate to represent a counterfactual conditional is an open topic. For instance, checking if the last form for the conditional, or other similar, can represent some type of counterfactual, could be a way to start such a study, as it also can be by understanding the antecedent as a hypothesis for the consequent.

14.13. Another question with conditionals refers to when a conditional refutes "p and q"; that is, in Boolean algebras, $p \cdot q \leq (p \to q)' \Leftrightarrow p \to q \leq p' + q'$. In addition, $p < q$ is a type-one speculation of $p \cdot q$, provided $p \cdot q$ were not comparable with $p \to q$, but $(p \cdot q)' \leq p \to q \Leftrightarrow p' + q' \leq p \to q$. Still, a conditional is a hypothesis for $p \cdot q \Leftrightarrow p \to q \leq p \cdot q$. Obviously, for the wild type-two speculations there is no equational way for directly posing and trying to answer the question.

A more realistic and deep study of the conditional's representation problem could come from designing suitable experiments in language, establishing mathematical models and testing them against "reality" and concerning all types of conjectures, doing it, of course, in a not blind but systematic form, within background knowledge, and inside the context-dependent and purpose-driven praxis in plain language and ordinary reasoning.

Nevertheless, a complete study of the symbolic representation and measuring of conditionals still is, in almost all its aspects and for plain language, an open problem affecting, for example, the symbolic representation of children's stories in view of their automatic computer mechanizing, and it is manifested by the more than 40 operators that, in fuzzy logic, have been used to represent conditionals.

References

1. E. Trillas, Some uncertain reflections on uncertainty. Arch. Philos. Hist. Soft. Comput **1**, 1–16 (2013)
2. E. Castiñeira, S. Cubillo, E. Trillas, On possibility and probability in finite Boolean algebras. Soft. Comput **7**(2), 89–96 (2002)
3. L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst **1**, 3–28 (1978)
4. P. Roeper, H. Leblanc, *Probability Theory and Probability Semantics* (University of Toronto Press, Toronto, 1999)
5. B. De Finetti, Sul significato soggettivo della probabilità. Fundamenta Mathematicae. **XVII**, 298–329 (1931)

6. F.S. Roberts, A note on Fine's axioms for qualitative probability. Ann. Probab **1**(3), 484–487 (1973)
7. E. Trillas, Relational probabilities in intuitionistic lattices. Stochastica **12**(2-3), 85–102 (1988)
8. F. Esteva, On the structure of intuitionistic algebras with relational probabilities. Stochastica **12**(2-3), 103–111 (1988)
9. L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning —Part I. Inf. Sci **8**, 199–249 (1975)
10. E. Trillas, L. Eciolaza, *Fuzzy Logic* (Springer, Berlin, 2015)
11. S. Gudder, What is fuzzy probability theory? Found. Phys **30**(10), 1663–1678 (2000)
12. S. Guadarrama, E. Renedo, E. Trillas, *A First Inquiry on Semantic-Based Models of And*, eds. by R. Seising et al. Accuracy and Fuzziness (2015), pp. 331–350
13. L.A. Zadeh, Probability measures of fuzzy events. J. Math. Anal. Appl **23**, 421–427 (1968)
14. P. Weley, G. de Cooman, A behavioural model for linguistic uncertainty. Inf. Sci **134**(1–4), 1–37 (2001)