University of Texas at El Paso

ScholarWorks@UTEP

Departmental Technical Reports (CS)　　　　　　　　　　　　　　　　Computer Science

4-2005

Use of Maxitive (Possibility) Measures in Foundations of Physics and Description of Randomness: Case Study

A. M. Finkelstein

Olga Kosheleva
The University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Scott A. Starks
The University of Texas at El Paso, sstarks@utep.edu

Hung T. Nguyen

Follow this and additional works at: https://scholarworks.utep.edu/cs_techrep

Part of the Computer Engineering Commons

Recommended Citation

Use of Maxitive (Possibility) Measures in Foundations of Physics and Description of Randomness: Case Study

A. M. Finkelstein
Institute of Applied Astronomy
Russian Academy of Sciences
St. Petersburg, Russia

O. Kosheleva, V. Kreinovich, S. A. Starks
Pan-American Center for Earth
and Environmental Studies
U. Texas, El Paso, TX 79968, USA
vladik@cs.utep.edu

H. T. Nguyen
Dept. of Math. Sciences
New Mexico State University
Las Cruces, NM, 88003, USA
hunguyen@nmsu.edu

Abstract— According to the traditional probability theory, events with a positive but very small probability can occur (although very rarely). For example, from the purely mathematical viewpoint, it is possible that the thermal motion of all the molecules in a coffee cup goes in the same direction, so this cup will start lifting up.

In contrast, physicists believe that events with extremely small probability cannot occur. In this paper, we show that to get a consistent formalization of this belief, we need, in addition to the original probability measure, to also consider a maxitive (possibility) measure.

I. PHYSICISTS ASSUME THAT INITIAL CONDITIONS AND VALUES OF PARAMETERS ARE NOT ABNORMAL

To a mathematician, the main contents of a physical theory is the equations. The fact that the theory is formulated in terms of well-defined mathematical equations means that the actual field must satisfy these equations. However, this fact does *not* mean that *every* solution of these equations has a physical sense. Let us give three examples:

Example 1. At any temperature greater than absolute zero, particles are randomly moving. It is theoretically possible that all the particles start moving in one direction, and, as a result, a person starts lifting up into the air. The probability of this event is small (but positive), so, from the purely mathematical viewpoint, we can say that this event is possible but highly unprobable. However, the physicists say plainly that such an abnormal event is *impossible* (see, e.g., [5]).

Example 2. Another example from statistical physics: Suppose that we have a two-chamber camera. The left chamber is empty, the right one has gas in it. If we open the door between the chambers, then the gas would spread evenly between the two chambers. It is theoretically possible (under appropriately chosen initial conditions) that the gas that was initially evenly distributed would concentrate in one camera. However, physicists believe this abnormal event to be impossible. This is an example of a "micro-reversible" process: on the atomic level, all equations are invariant with respect to changing the order of time flow ($t \rightarrow -t$). So, if we have a process that goes from state A to state B, then, if while at B, we revert all the

velocities of all the atoms, we will get a process that goes from B to A.

However, in real life, many processes are clearly irreversible: an explosion can shatter a statue but it is hard to imagine an inverse process: an implosion that glues together shattered pieces into a statue. Boltzmann himself, the 19th century author of statistical physics, explicitly stated that such inverse processes "may be regarded as impossible, even though from the viewpoint of probability theory that outcome is only extremely improbable, not impossible." [1].

Example 3. If we toss a fair coin 100 times in a row, and get heads all the time, then a person who is knowledgeable in probability would say that it is possible – since the probability is still positive. On the other hand, a physicist (or any person who uses common sense reasoning) would say that the coin is not fair – because if it is was a fair coin, then this abnormal event would be impossible.

In all these cases, physicists (implicitly or explicitly) require that the actual values of the physical quantities must not only satisfy the equations but they must also satisfy the additional condition: that the initial conditions should *not* be *abnormal*.

Comment. In all these examples, a usual mathematician's response to physicists' calling some low-probability events "impossible", is just to say that the physicists use imprecise language.

It is indeed true that the physicists use imprecise language, and it is also true that in the vast majority of practical applications, a usual probabilistic interpretation of this language perfectly well describes the intended physicists' meaning. In other words, the probability language is perfectly OK for most physical applications.

However, there are some situations when the physicists' intuition seem to differ from the results of applying traditional probability techniques:

- From the probability theory viewpoint, there is no fundamental difference between such low-probability events as a person winning a lottery and the same person being

lifted up into the air by the Brownian motion. If a person plays the lottery again and again, then – provided that this person lives for millions of years – he will eventually win. Similarly, if a person stands still every morning, then – provided that this person lives long enough – this person will fly up into the air.

- On the other hand, from the physicist viewpoint, there is a drastic difference between these two low-probability events: yes, a person will win a lottery but no, a person will never lift up into the air no matter how many times this person stands still.

We have just mentioned that the traditional mathematical approach is to treat this difference of opinion as simply caused by the imprecision of the physicists' language. What we plan to show is that if we take this difference more seriously and develop a new formalism that more accurately captures the physicists' reasoning, then we may end up with results and directions that are, in our opinion, of potential interest to foundations of physics.

In other words, what we plan to show is that if we continue to use the traditional probability approach, it is perfectly OK but if we try to formalize the physicists' opinion more closely, we may sometimes get even better results.

Comment. It is known that the probabilistic approach is, in principle, quite capable of describing uncertainty in physical systems; in particular, the probabilistic approach is capable of describing rare events in physical systems. Some problems related to rare events require new probabilistic techniques, but overall, within a traditional probabilistic approach, there seems to be no need to introduce possibility measures.

It is also known that possibility measures can be useful for physical applications, e.g., to reflect considerable vagueness in the dynamical laws of complex physical systems, to describe the vague, fuzzy character of the experts (= physicists) knowledge about these systems.

What we plan to show, in this paper, is that there is one more area where possibility measures can be helpful in physical applications: in describing the physicists' *intuition* about rare events, an intuition that is somewhat different from its traditional probabilistic description.

II. A Seemingly Natural Formalization of This Idea

The above-mentioned property of being "not abnormal" ("typical") has a natural formalization: if a probability $p(E)$ of an event E is small enough, i.e., if $p(E) \leq p_0$ for some very small threshold p_0, then this event cannot happen.

In other words, there exists the "smallest possible probability" p_0 such that:

- if the computed probability p of some event is larger than p_0, then this event can occur, while
- if the computed probability p is $\leq p_0$, the event cannot occur.

For example, the probability that a fair coin falls heads 100 times in a row is 2^{-100}, so, if the threshold probability p_0

satisfies the inequality $p_0 \geq 2^{-100}$, then we will be able to conclude that such an event is impossible.

III. The Above Formalization of the Notion of "Typical" is Not Always Adequate

The problem with this approach is that *every* sequence of heads and tails has exactly the same probability. So, if we choose $p_0 \geq 2^{-100}$, we will thus exclude all possible sequences of 100 heads and tails as physically impossible. However, anyone can toss a coin 100 times, and this proves that some such sequences are physically possible.

The threshold should depend on the complexity of the property: for simple properties, it should be larger, but for complex properties (e.g., described by a specific sequence of 0s and 1s), it should be smaller.

Historical comment. This problem was first noticed by Kyburg under the name of *Lottery paradox* [10]: in a big (e.g., state-wide) lottery, the probability of winning the Grand Prize is so small that a reasonable person should not expect it. However, some people do win big prizes.

IV. Relation to Non-Monotonic Reasoning

Lottery paradox has been known for several decades, and many solutions have been proposed to resolve this paradox.

One possible solutions comes from the fact that in deriving the above paradox, we used classical logic, a logic that is *monotonic* in the following sense: once we made a logical conclusion, this conclusion remains valid no matter what new knowledge we acquire. In classical logic, if we increase the set of facts and rules, the set of conclusions can only increase.

Our objective, however, is to formalize expert reasoning (specifically, physicists' reasoning), and it is known that the expert reasoning is, in general, not monotonic. For example, if we know that birds normally fly, and we see a bird, then we normally conclude that this bird can fly. However, if it later turns out that this bird is an abnormal bird, e.g., a penguin, that we take back our original conclusion and conclude that this particular bird does not fly.

It is known that if we take the non-monotonic character of expert reasoning into consideration, then the lottery paradox stops being a paradox, it becomes simply one of the non-monotonic features of expert reasoning; see, e.g., Poole [11], [12] (see also [8]). Specifically, if we use formalisms like default logic that have been designed to capture commonsense reasoning, we can explain the above paradox.

From the pragmatic viewpoint, this approach is very satisfactory; however, from the foundational viewpoint, the existing description of commonsense non-monotonic reasoning is still being developed, better and better semantics of non-monotonic reasoning – in particular, reasoning about what is typical and what is normal – are appearing all the time.

What we plan to do in this paper is restrict ourselves only to the description of rare events. For this narrow specialized area of reasoning, we will provide a formalization of the notions of "typical" and "normal", and thus, in effect, we provide a *specific* non-monotonic logic – a logic is reasonably final (and

thus, does not depend on the fact that the *general* description of non-monotonic reasoning is still far from being final).

Comment. It is worth mentioning that there is an interesting alternative approach to the above solution of the lottery paradox: namely, we can, alternatively, conclude that our intuition is simply wrong and that events with very small (even 0) probability can actually happen.

This alternative approach was pioneered by such well-known specialists in philosophical foundations of probability theory as K. Popper and B. de Finetti (see, e.g., [3]). This approach is currently being successfully developed by G. Coletti, A. Gilio, R. Scozzafava, W. Spohn, and others (see, e.g., [2] and references therein). Within this alternative approach, there is a natural hierarchy of zero probability events (induced by the corresponding conditional probabilities), and this hierarchy also leads to a maxitive measure!

Since our objective is to formalize the physicists' intuition, not to reject it, we do not follow this alternative approach. However, the very fact that both approaches lead to the same formalism of maxitive measures makes us think that maybe there is a deep relation and similarity between these two approaches.

V. New Idea

"Abnormal" means something unusual, rarely happening: if something is rare enough, it is not typical ("abnormal"). Let us describe what, e.g., an abnormal height may mean. If a person's height is ≥ 6 ft, it is still normal (although it may be considered abnormal in some parts of the world). Now, if instead of 6 ft, we consider 6 ft 1 in, 6 ft 2 in, etc., then sooner or later we will end up with a height h_0 such that everyone who is taller than h_0 will be definitely called atypical, abnormal (to be more precise, a person of abnormal height). We may not be sure what exactly value h experts will use as a threshold for "abnormal" but we are sure that such a value exists.

While every person whose height is $> h_0$ is definitely atypical, a person whose height is below h_0 is not necessarily typical: he may be atypical because of some other properties.

For example, we may consider people atypical because of an unusual weight. Similarly, there exists a weight w_0 such that everyone whose weight exceeds w_0 will be called atypical.

Let us express the above idea is general terms. We have a *universal set*, i.e., the set U of all objects that we will consider. In the above example, U is the set of all people. Some of the elements of the set U are abnormal (in some sense), and some are not. Let us denote the set of all elements that are *typical* (*not abnormal*) by T.

On the set U, we have several decreasing sequences of sets $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$ with the property that $\bigcap_n A_n = \emptyset$.

In the height example, A_1 is the set of all people whose height is ≥ 6 ft, A_2 is the set of all people whose height is ≥ 6 ft 1 in, A_3 is the set of all people whose height is ≥ 6 ft 2 in, etc.

In the weight example, A_1 is the set of all people whose weight is ≥ 150 lb, A_2 is the set of all people whose weight is ≥ 160 lb, A_3 is the set of all people whose weight is ≥ 170 lb, etc.

We know that for each of these sequences, if we take a sufficiently large n, then all elements of A_n are abnormal (i.e., none of them belongs to the set T of not abnormal elements). In mathematical terms, this means that for some integer N, we have $A_N \cap T = \emptyset$.

Let us describe this idea in precise terms [8]. Let $\mathcal{L}$ be a theory. A set is $\mathcal{L}$-definable if we can explicitly *define* it in $\mathcal{L}$. (The set of all real numbers, the set of all solutions of a well-defined equation, every set that we can describe in mathematical terms is $\mathcal{L}$-definable.)

Our objective is to be able to make mathematical statements about $\mathcal{L}$-definable sets. Therefore, in addition to the theory $\mathcal{L}$, we must have a stronger theory $\mathcal{M}$ in which the class of all $\mathcal{L}$-definable sets is a set – and it is a countable set.

Definition 1. *Let U be a universal set. A non-empty set $T \subseteq U$ is called a* set of typical (not abnormal) elements *if for every $\mathcal{L}$-definable sequence of sets A_n for which $A_n \supseteq A_{n+1}$ for all n and $\bigcap_n A_n = \emptyset$, there exists an integer N for which $A_N \cap T = \emptyset$.*

Example. In the above coin example, $U = \{\mathrm{H}, \mathrm{T}\}^{\mathbb{N}}$, and A_n is the set of all the sequences that start with n heads and have at least one tail. The sequence $\{A_n\}$ is decreasing and $\mathcal{L}$-definable, and its intersection is empty. Therefore, for every set T of typical elements of U, there exists an integer N for which $A_N \cap T = \emptyset$. This means that if a sequence $s \in T$ is not abnormal and starts with N heads, it must consist of heads only. In physical terms, it means a random sequence (i.e., a sequence that contains both heads and tails) cannot start with N heads – which is exactly what we wanted to formalize.

VI. Main Result: Relation to Possibility Measures

To describe a set of typical elements, we ascribe, to each definable monotonic sequence $\{A_n\}$, the smallest integer $N(\{A_n\})$ for which $A_N \cap T = \emptyset$. This integer can be viewed as measure of complexity of the sequence: for simple sequences, it is smaller, for more complex sequences, it is larger.

In terms of this complexity, the above definition of typical elements can be reformulated as follows: an element $x \in U$ is typical if and only if for every definable decreasing sequence $\{A_n\}$ with an empty intersection, $x \notin A_N$, where $N = N(\{A_n\})$ is the complexity of this sequence.

It turns out that $N(\{A_n\})$ is a maxitive (possibility) measure in the following sense:

Proposition 1. $N(\{A_n \cup B_n\}) = \max(N(\{A_n\}), N(\{B_n\}))$.

Proof. Indeed, $N(\{A_n \cup B_n\})$ is the smallest value N for which $(A_N \cup B_N) \cap T = \emptyset$, i.e., for which no element of $A_N \cup B_N$ is typical. This is equivalent to saying that no element of A_N is typical and no element of B_N is typical. In other words,

$$(A_N \cup B_N) \cap T = \emptyset \leftrightarrow ((A_N \cap T = \emptyset) \,\&\, (B_N \cap T = \emptyset)).$$

Hence, $N(\{A_n \cup B_n\})$ is indeed the largest of the two numbers $N(\{A_n\})$ and $N(\{B_n\})$). Q.E.D.

Comment. If $A_n \subseteq B_n$ for all n, then Proposition 1 implies that $N(\{A_n\}) \leq N(\{B_n\})$, i.e., that this complexity measure is monotonic.

Since $A_n \cap B_n \subseteq A_n$ and $A_n \cap B_n \subseteq B_n$, we can thus conclude that $N(\{A_n \cap B_n\}) \leq \min(N(\{A_n\}), N(\{B_n\}))$.

Comment. One can easily see that a proof similar to our proof of Proposition 1 leads +to a more general result:

Proposition 2. *For every definable sequence of sequences* $\{A_n^\alpha\}$, $N(\{\bigcup_\alpha A_n^\alpha\}) = \max_\alpha(N(\{A_n^\alpha\}))$.

Comment. Another relation between the set of typical elements and possibility (maxitive) measures is described in [7].

VII. POSSIBLE PRACTICAL USE OF THIS IDEA: WHEN TO STOP AN ITERATIVE ALGORITHM

In numerical mathematics, we often know an iterative process whose results x_k are known to converge to the desired solution x, but we do not know when to stop in order to guarantee the given accuracy ε – i.e., to guarantee that the distance $d_X(x_n, x)$ between the numerical result x_n and the actual solution x does not exceed ε. In these cases, heuristic methods are used:

Usually, in iterative methods, if $x_k = x_{k+1}$, then x_k is the required solution. Therefore, if x_k and x_{k+1} are close, we can conclude that we are *close* to the solution. Hence, we stop when the consequent values x_k become close enough, i.e., when $d_X(x_k, x_{k+1}) \leq \delta$ for some $\delta > 0$.

This method is often used in physics, if, e.g., we have the expression of x as a sum of the infinite series (e.g., Taylor series in perturbation methods). Then, if, e.g., second order terms are negligibly small, we neglect quadratic *and* higher order terms, and use the linear expression as an approximation to the desired solution (see, e.g., [6]).

If we are solving the equation $f(x) = y$ with known y, then we stop when $f(x_k)$ becomes close enough to y – i.e., when $d_X(f(x), y) \leq \delta$ for some $\delta > 0$.

Both stopping criteria can be viewed as particular cases of the following general definition:

Definition 2. *Let X be a definable metric space, and let S be a definable set of convergent sequences of X.*

- *Let $\{x_k\} \in S$, k be an integer, and $\varepsilon > 0$ a real number. We say that x_k is $\varepsilon-$accurate if $d_X(x_k, \lim x_p) \leq \varepsilon$.*
- *Let $d \geq 1$ be an integer. By a* stopping criterion*, we mean a function $c : X^d \to R_0^+ = \{x \in R \,|\, x \geq 0\}$ that satisfies the following two properties:*
    - *If $\{x_k\} \in S$, then $c(x_k, \ldots, x_{k+d-1}) \to 0$.*
    - *If for some $\{x_n\} \in S$ and for some k, $c(x_k, \ldots, x_{k+d-1}) = 0$, then $x_k = \ldots = x_{k+d-1} = \lim x_p$.*

The two above-described criteria correspond to $c(x, x') = d_X(x, x')$ and $c(x) = d_X(f(x), y)$.

Proposition 3. *Let c be a stopping criterion. Then, for every ε, there exists a $\delta > 0$ such that if a sequence $\{x_n\}$ is not abnormal, and $c(x_k, \ldots, x_{k+d-1}) \leq \delta$, then x_k is $\varepsilon-$accurate.*

Proof. As A_n, we will take the set of all sequences for which for some k, $c(x_k, \ldots, x_{k+d-1}) \leq 2^{-n}$ and $d_X(x_k, x) > \varepsilon$. Clearly, $A_n \supseteq A_{n+1}$.

Let us show that the intersection $\cap A_n$ is empty. Indeed, suppose that the sequence $\{x_k\}$ belongs to this intersection. This means that for every n, there exists a $k(n)$ such that $c(x_{k(n)}, \ldots, x_{k(n)+d-1}) \leq 2^{-n}$ and $d_X(x_{k(n)}, x) > \varepsilon$. If some value k is equal to $k(n)$ for infinitely many n, this means that $c(x_k, \ldots, x_{k+d-1}) \leq 2^{-n}$ for all n and hence, that $c(x_k, \ldots, x_{k+d-1}) = 0$. From the definition of a stopping criterion, it then follows that $x_k = x$, so $d(x_k, x) = 0 \not> \varepsilon$. Hence, $k(n) \to \infty$, so (since $\{x_k\}$ is convergent), $d_X(x_{k(n)}, x) \to 0$ and $d_X(x_{k(n)}, x) \not> \varepsilon$. The contradiction shows that the intersection is empty.

So, there exists an N for which $A_N \cap T = \emptyset$. Hence, we can take $\delta = 2^{-N}$. Q.E.D.

Comment. So, if we restrict ourselves to not abnormal sequences only (i.e., sequence that stem from not abnormal, physical observations), then $c(x_k, \ldots, x_{k+d-1}) \leq \delta$ guarantees that we are ε-close to the desired solution.

In particular, $d_X(x_k, x_{k+1}) \leq \delta$ and $d(f(x_k), y) \leq \delta$ guarantee that $d_X(x_n, x) \leq \varepsilon$.

In case we are summing a numerical series $x_k = a_1 + \ldots + a_k$, we have $d(x_k, x_{k+1}) = |a_{k+1}|$, so, this stopping criterion means that means if the next term is negligible ($|a_{k+1}| \leq \delta$), then we are $\varepsilon-$close to the sum: $|x_k - x| \leq \varepsilon$.

VIII. WHEN RESTRICTED TO TYPICAL OBJECTS, MANY NUMERICAL PROBLEMS BECOME ALGORITHMICALLY DECIDABLE

Another possible application involves constructive real numbers, i.e., real numbers x for which we can – by computations (like with π) or by measurement – obtain, for every k, a 2^{-k}-approximation r_k (for which $|r_k - x| \leq 2^{-k}$). We can have several constructive real numbers $x, x', \ldots, x''$ corresponding to different measurable physical quantities. Based on the results of more and more accurate measurements of these numbers, we would like to check whether the actual values satisfy a desired property; e.g., whether for the vector $X \overset{\text{def}}{=} (x, x', \ldots, x'')$, we have $f(X) \geq 0$ for a given multi-variate function f. (Sometimes, we would like to check whether $f(X) > 0$, this is equivalent to checking whether $g(X) \geq 0$, for $g(X) \overset{\text{def}}{=} -f(X)$.) In general, it is algorithmically impossible to decide whether a given constructive number is non-negative or not; see, e.g., [9].

We can (easily) prove that there exists a k such that for every not abnormal vector X, either $f(X) \geq 0$ or $f(X) < -2^{-k}$.

In practice, the function $f(X)$ is continuous, hence there exists l such that if we know X with accuracy 2^{-l}, then we can determine $f(X)$ with accuracy $2^{-(k+1)}$ – and thus, tell whether $f(X) \geq 0$ or whether $f(X) < -2^{-k}$. So, if we

know that the actual vector is typical, then, after computing (or measuring) it with accuracy 2^{-l}, we will be able to tell whether $f(X) \geq 0$ or $f(X) < 0$. Thus, for *typical* values, the generally undecidable problem becomes algorithmically decidable. (Of course, the catch is that we do not know l – it depends on the selection of a typical set T.)

We can now use this result to find out, for a given spatial accuracy ε, where a given constructive function $f(X)$ attains it maximum on a given box $[\underline{x}_1, \overline{x}_1] \times \ldots \times [\underline{x}_n, \overline{x}_n]$. Indeed, we can divide the box into sub-boxes of size $\leq \varepsilon$, and algorithmically compute the maximum of f over each subbox [9]. When a function is typical, the difference between its maxima on two constructive subbox is also typical – so we can algorithmically decide whether the two maxima are equal or one is larger – and thus find out for which subbox the maximum is the largest; this is where the maximum is located – so have determined this location with the given spatial accuracy ε.

IX. ADDITIONAL IDEA: DEGREE OF TYPICALNESS

In the main text, we largely considered a set of typical elements T of a given universal set. It makes sense to also consider typical elements of different subsets of T: e.g., in addition to a typical height of a person, it may be reasonable to consider a typical height of a child or a typical height of a person living in Russia. For that, we must assume that we for some universal set U, we have a mapping that assigns, to every $\mathcal{L}$-definable set $A \subseteq U$, a non-empty set $T(A)$ that is a set of typical elements for A.

In the above text, we implicitly assumed that every object $a \in A$ can be classified as either a typical element of A or an abnormal element of A. According to common sense, however, being typical or abnormal may be a matter of degree. How can we extend our description to this "degree of typicalness"?

Typical means, e.g., that if we have no prior information about a real number, that it is reasonable to assume that this real number is not abnormal. Later on, we may learn that there are some features of this particular situation that make it atypical; however, if all we know is that the object is atypical, we should still be able to conclude that it a "typical" element of the class of abnormal objects, "typical exception" – crudely speaking, an abnormal object of degree 1. Alternatively, it may turn out to be an abnormal object is an exception even among exceptions – i.e., abnormal of degree at least 2, etc.

How can we describe this formally? We have shown that once a logico-mathematical theory $\mathcal{L}$ that formalizes the physical knowledge is fixed, in a larger (meta-)theory $\mathcal{M}$, we can talk about definable and typical objects. In this meta-theory $\mathcal{M}$, we can define what it means for a set $T(A) \subseteq A$ to be the set of all typical (not abnormal) elements of the set A.

As we have mentioned, to fully describe physicists' reasoning, we should assume that one such set $T(A)$ was selected for all (or at least for some) definable sets A. The model of the original theory $\mathcal{L}$ + this selection provides, in effect, a model of an *extended* theory $\mathcal{L}'$, in which, in addition to the original basic properties, relations, and basic functions, we also have a new basic property – that a is a typical element of A. In this extended theory $\mathcal{L}'$, the term "typical" is part of the theory, thus, e.g., the set $T(A)$ is explicitly definable in the new theory – while it was not definable in the original physical theory $\mathcal{L}$.

On top of this extended theory $\mathcal{L}'$, we can also build a new meta-theory $\mathcal{M}'$, and in this new meta-theory, we can talk about the typical elements of sets which are definable in $\mathcal{L}'$. In particular, in this new meta-theory $\mathcal{M}'$, we can talk about typical elements of the set $Ab(A) \stackrel{\text{def}}{=} A \setminus T(A)$ of all abnormal elements. The set $T(Ab(A))$ of all such typical elements consists of "typical exceptions", i.e., abnormal elements of degree 1. Remaining elements, i.e., elements of the set $Ab_2(A) \stackrel{\text{def}}{=} Ab(A) \setminus T(Ab(A))$, are elements which are exceptions even among exceptions, i.e., they are abnormal of degree 2.

Similarly, we can defined a yet another meta-theory and talk about elements of degree at least 3, etc. The general definition of the set $Ab_k(A)$ of all elements abnormal of degree $\geq k$ is $Ab_1(A) = A \setminus T(A)$ and $Ab_{k+1} = Ab_k(A) \setminus T(Ab_k(A))$. The *degree of abnormality* of an abnormal element $a \in A$ can then be defined as the largest integer k for which $a \in Ab_k(A)$.

X. BEYOND MAXITIVE MEASURES

We have mentioned that we also want to exclude events A of low probability. We have also mentioned that the threshold probability of an event – below which this event is impossible – should depend on the complexity of this event.

In probability theory, it is usually sufficient to consider events belonging to the Borel σ-algebra, i.e., the smallest σ-algebra generated by open and closed subsets of the original set U. It also makes sense to only consider probability measures defined on this σ-algebra; for every such event A and probability measure p, the probability $p(A)$ is well-defined.

Let $c(A) > 0$ denote this threshold "complexity" of event A. Then, for a probability measure p, if $p(A) < c(A)$, then A is impossible, i.e., $T(p) \cap A = \emptyset$, where $T(p)$ denotes the set of typical elements corresponding to p.

Definition 3. *Let p be a definable probability measure, and let c be a function that maps every definable set into a positive real number. We say that a set $T(p)$ is* consistent *with c and p if $T(p) \cap A = \emptyset$ for every definable set A for which $p(A) < c(A)$.*

What are the conditions on the "complexity measure" c under which a consistent set $T(p)$ exists for every p? It turns out that these conditions can be formulated in terms of the following property:

Definition 4. *We say that a sequence of sets $\{X_i\}$ is $\cup$-independent if for all i, $X_i \nsubseteq \bigcup_{j \neq i} X_j$.*

Proposition 4.

- *If for every definable probability measures p, there exists a non-empty set $T(p)$ that is consistent with c and p, then for all $\cup$-independent definable families $\{X_i\}$, we have $\sum c(X_i) \leq 1$.*

- *Let $0 < \varepsilon < 1$. If for every definable probability measures p, there exists a set $T(p)$ that is consistent with c and p and for which $p(T(p)) \geq 1 - \varepsilon$, then $\sum c(X_i) \leq \varepsilon$ for all $\cup$-independent definable families $\{X_i\}$ for which $\cup X_i \neq U$.*
- *Let $0 < \varepsilon < 1$. If $\sum c(X_i) \leq \varepsilon$ for all $\cup$-independent definable families $\{X_i\}$, then for every definable probability measure p, there exists a set $T(p)$ that is consistent with c and p and for which $p(T(p)) \geq 1 - \varepsilon$.*

Proof. Let us first consider the case when a non-empty consistent set $T(p)$ exists for every definable probability measure p. Let us assume that, vice versa, $\sum\limits_{i=1}^{\infty} c(X_i) > 1$. In this case, for some finite n, we have $s \stackrel{\text{def}}{=} \sum\limits_{i=1}^{n} c(X_i) = 1$. Since the family $\{X_i\}$ is $\cup$-independent, for every $i \leq n$, there exist $x_i \in X_i \setminus \bigcup\limits_{j \neq i} X_j$. So, all n elements x_i are different, and each X_i contains x_i but not any other x_j.

We can now define a probability measure p concentrated on x_i, with $p(\{x_i\}) = c(X_i)/s$. Here, $p(X_i) = p(\{x_i\})$. Since $s > 1$, we have $p(X_i) < c(X_i)$, hence each event $X_i \neq x_i$ is impossible. So, for this p, we have $x_i \notin T(p)$. On the other hand, for $A \stackrel{\text{def}}{=} U \setminus \{x_1, \ldots, x_n\})$, we have $p(A) = 0 < c(A)$, so elements $\neq x_i$ are also impossible – which contradicts to $T(p) \neq \emptyset$.

Let us now assume that for every p, there exists a consistent set $T(p)$ for which $p(T(p)) \geq 1 - \varepsilon$. Let us prove that $\sum c(X_i) \leq \varepsilon$. Indeed, if $\sum c(X_i) > \varepsilon$, then there exist a finite $\cup$-independent family $\{X_1, \ldots, X_n\}$ for which $s \stackrel{\text{def}}{=} \sum c(X_i) > \varepsilon$. Select x_i as in the first part of this proof. Since $\cup X_i \neq U$, there exists an element $x_0 \notin \cup X_i$. Pick $\delta > 0$ so small that we still have $s/(1 + \delta) > \varepsilon$, and take $p(\{x_i\}) = c(X_i)/(1 + \delta)$ and $p(\{x_0\}) = 1 - \sum p(\{x_i\})$. Here, $p(X_i) = p(\{x_i\}) = c(X_i)/(1 + \delta) < c(X_i)$ hence $T(p) \cap X_i = \emptyset$. So, $T(p) \subseteq X \setminus \cup X_i$. However, $p(\cup X_i) = \sum p(\{x_i\}) = s/(1+\delta)$, so $p(\cup X_i) > \varepsilon$ hence $p(T(p)) < 1 - \varepsilon$ – a contradiction with our assumption.

Finally, let $\sum c(X_i) \leq \varepsilon$ for all $\cup$-independent families $\{X_i\}$. Let us show that as $T(p)$, we can take whatever remains after we exclude all $\mathcal{L}$-definable sets A with $p(A) < c(A)$. To prove that $T(p) \neq \emptyset$, we will prove that for every n, after excluding n such sets, the remainder T_n has a measure $\geq 1 - \varepsilon$. Thus, the limit set $T(p)$ – after excluding (countably many) definable sets for which $p(A) < c(A)$ – will also have measure $\geq 1 - \varepsilon > 0$.

Let $A_1, \ldots, A_n$ be sets that we excluded because $p(A_i) < c(A_i)$. Overall, we excluded the union $\cup A_i$. If one of these set A_i is contained in the union of others, then we do not need to explicitly exclude it – it is excluded automatically when we exclude others. So, we can delete this set from the list A_i of explicitly excluded sets. By repeating this procedure, we will eventually end up with a $\cup$-independent subset A_i – with the same union $\cup A_i$ as before. For this subset, $\sum c(A_i) \leq \varepsilon$, so, since $p(A_i) < c(A_i)$, we have $\sum p(A_i) < \varepsilon$.

The probability measure $p(\cup A_i)$ of the excluded union $\cup A_i$ cannot exceed the sum $\sum p(A_i)$ and is, hence $\leq \varepsilon$. Thus, the remainder has a measure $\geq 1 - \varepsilon$. The proposition is proven.

REFERENCES

[1] L. Boltzmann, "Bemrkungen über einige Probleme der mechanischen Wärmtheorie", *Wiener Ber. II*, 1877, Vol. 75, pp. 62–100.
[2] G. Coletti and R. Scozzafava, *Probabilistic Logic in a Coherent Setting*, Kluwer Academic Publishers, Dordrecht, 2002.
[3] B. de Finetti, *Theory of Probability*, Wiley, New York, 1970.
[4] D. Dubois and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York, 1988.
[5] R. P. Feynman, *Statistical Mechanics*, W. A. Benjamin, 1972.
[6] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures*, Addison-Wesley, 1965.
[7] A. M. Finkelstein, O. Kosheleva, V. Kreinovich, S. A. Starks, and H. T. Nguyen, "To Properly Reflect Physicists' Reasoning about Randomness, We Also Need a Maxitive (Possibility) Measure", *Proc. FUZZ-IEEE'2005*, Reno, Nevada, May 22–25, 2005 (to appear).
[8] V. Kreinovich and A. M. Finkelstein, "Towards Applying Computational Complexity to Foundations of Physics", *Notes of Math. Seminars of St. Petersburg Dept. of Steklov Inst. of Math.*, 2004, Vol. 316, pp. 63–110.
[9] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1997.
[10] H. E. Kyburg, Jr., *Probability and the logic of rational belief*, Wesleyan Univ. Press, 1961.
[11] D. Poole, "The effect of knowledge on belief: conditioning, specificity and the lottery paradox in default reasoning", *Artificial Intelligence*, 1991, Vol. 49, pp. 281–307.
[12] D. Poole, "The effect of knowledge on belief: conditioning, specificity and the lottery paradox in default reasoning", In: R. J. Brachman, H. J. Levesque, and R. Reiter (Eds.), *Knowledge Representation*, MIT Press, Cambridge, Massachusetts, 1991.
[13] L. Zadeh, "Fuzzy sets as a basis for a theory of possibility", *Fuzzy Sets and Systems*, 1978, Vol. 1, pp. 3–28.

APPENDIX

What is the relation between $T(A)$ for different sets A? Some such properties are described in [8]; here is one more.

Proposition 5. *Let U be the set of all real numbers, and let T be a mapping that maps every $\mathcal{L}$-definable non-empty subset $A \subseteq U$ into a non-empty set $T(A)$ of typical elements w.r.t. A. Then there exist non-empty sets A, B, and C for which $A \subseteq B$, $T(B) \subseteq C$, and $T(A) \not\subseteq C$.*

Proof. Let $B = [0, 1]$. As a sequence B_n, let us take $B_n = (0, 1/n)$. This is a $\mathcal{L}$-definable monotonic sequence with the empty intersection, so, due to our Definition, there exists an integer N for which $B_N \cap T(B) = \emptyset$. Hence, for $A = A_N$ and $C = B \setminus B_N$, we have $A \subseteq B$, $T(B) \subseteq B \setminus B_N = C$, but $T(A) \subseteq A = B_N$ and therefore $T(A) \cap C = \emptyset$. Q.E.D.

Comment. A natural interpretation of a commonsense statements like "normally, A implies B" is that *typical* elements of A have the property B, i.e., that $T(A) \subseteq B$. It is well known that such commonsense implication is sometimes not transitive: e.g., penguins (A) are birds (B), birds normally fly (C), but penguins do not normally fly. Our proposition proves that the existence of a non-transitivity example is not accidental: it follows from the very definition of typicality.