Journal of Uncertain Systems
Vol.1, No.2, pp.84-108, 2007
Online at: www.jus.org.uk

To Properly Reflect Physicists' Reasoning about Randomness, We Also Need a Maxitive (Possibility) Measure and a General Notion of Boundedness

Andrei M. Finkelstein[1], Olga Kosheleva[2], Tanja Magoc[3],
Erik Madrid[3], Scott A. Starks[4], and Julio Urenda[3]

[1]*Institute of Applied Astronomy, Russian Academy of Sciences, St Petersburg, Russia*
[2]*Department of Teacher Education*
[3]*Department of Computer Science*
[4]*Department of Electrical and Computer Engineering*
University of Texas at El Paso, El Paso, TX 79968, USA
Emails: {olgak, tmagoc, semadrid, sstarks, jcurenda}@utep.edu

Received 15 November 2006; Accepted 20 December 2006

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. We also show that the resulting advanced and somewhat difficult-to-described definition can be actually viewed as a particular case of something very natural: the general notion of boundedness.

1 Introduction

To a physicist, clearly, real numbers 1.0 or 0.5 or e (2.781828...) are "not random", while numbers produced by real-life random physical processes are "random". This distinction between random and non-random objects has been known for several centuries, but it is very difficult to describe in precise mathematical terms. The formal definition was only given in the 20 century, via the work on von Mises, Kolmogorov, Martin-Löf and others. One known description is a description via Kolmogorov complexity. The original description is not in complete agreement with our intuition because according to this definition, a sequence of 1,000,000 0s followed by a 1 is random.

Leonid Levin used Kolmogorov complexity to provide a more intuitive definition. Crudely speaking, if we have a property that few objects satisfy, then it is not very probable that a random object has this property. So, if we have a sequence of such properties, rarer and rarer, then there is a threshold after which "random" ("typical") objects cannot satisfy one of them. For the properties "start with n zeros", we conclude that there is an integer N such that a truly random sequence cannot start with N zeros.

In the first part of 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. In the second part, we relate this to Levin's idea and formulate a new abstract notion of randomness.

We then show that the resulting advanced and difficult-to-described definition can be actually viewed as a particular case of something very natural: the general notion of boundedness. This notion – like many other general notions – is best formulated in terms of a very general language called category theory. In this language, we have objects and mapping (called "morphisms"). For example, in the category Set objects are sets, and mapping are arbitrary functions. In the category Ord, objects are ordered sets, and mapping are monotonic functions. We can define a subset S of an object A to be bounded if for every morphism f from A to real numbers R, the range $f(S)$ is bounded. It turns out that for many categories, this notion of boundedness is equivalent to a natural one: for Set, S is bounded if and only it is finite; for linearly ordered sets, S is bounded if and only if it is bounded in the normal sense; for continuous mappings, bounded means compact etc.

We show that if instead of the category of *all* sets and functions, we consider only *definable* sets and functions, then boundedness is exactly randomness (typicalness) in the sense of the above definition. Thus, the seemingly complex notion of randomness is, in fact, very mathematically natural.

Comment. Some of the results from this paper first appeared in [9, 10, 14]

2 Formalizing Physicists' Reasoning about Typicalness: Formulation of the Problem

2.1 Physicists Assume That Initial Conditions and Values of Parameters Are Not Abnormal

To a mathematician, the main contents of a physical theory is its 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., [11]).

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 ($t \to -t$). 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." [3].

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.

2.2 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 $p(E) \le p_0$ 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 $\le p_0$, 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 $p_0 \ge 2^{-100}$, then we will be able to conclude that such an event is impossible.

2.3 The Above Formalization of the Notion of "Typical" Is Not Always Adequate

In the previous section, we described a seemingly natural formalization of the notion "typical" ("not abnormal"): if a probability of an event is small enough, i.e., $\le p_0$ for some very small p_0, then this event cannot happen.

The problem with this approach is that *every* sequence of heads and tails has exactly the same probability. So, if we choose $p_0 \ge 2^{-100}$, 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.

Historical comment. This problem was first noticed by Kyburg under the name of *Lottery paradox* [18]: 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.

2.4 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 [20, 21] (see also [15]). 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., [6]). This approach is currently being successfully developed by G. Coletti, A. Gilio, R. Scozzafava, W. Spohn, and others (see, e.g., [5] 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.

3 Kolmogorov's Idea: Using Complexity

Crudely speaking, the main problem arises because we select the same threshold p_0 for all events. For example, if we toss a fair coin 100 times then a sequence consisting of all heads should not be possible, and it is a reasonable conclusion because the probability that tossing a fair coin will lead to this sequence is extremely small: 2^{-100}.

On the other hand, whatever specific sequence of heads and tails we get after tossing a coin, this sequence also has the same small probability 2^{-100}. In spite of this, it does not seem to be reasonable to dismiss such sequences.

Several researchers thought about this, one of them A. N. Kolmogorov, the father of the modern probability theory. Kolmogorov came up with the following idea: the probability threshold $t(E)$ below which an event E is dismissed as impossible must depend on the event's complexity. The event E_1 in which we have 100 heads is easy to describe and generate; so for this event, the threshold $t(E_1)$ is higher. If $t(E_1) > 2^{-100}$ then, within this Kolmogorov's approach, we conclude that the event E_1 is impossible. On the other hand, the event E_2 corresponding to the actual sequence of heads and tails is much more complicated; for this event E_2, the threshold $t(E_2)$ should be much lower. If $t(E_2) < 2^{-100}$, we conclude that the event E_2 is possible.

The general fact that out of 2^n equally probable sequences of n 0s and 1s some are "truly random" and some are not truly random was the motivation behind Kolmogorov and Martin-Löf's formalization of randomness (and behind the related notion of Kolmogorov complexity; the history of this discovery is described in detail in [19]).

This notion of Kolmogorov complexity was introduced independently by several people: Kolmogorov in Russia and Solomonoff and Chaitin in the US. Kolmogorov defined complexity $K(x)$ of a binary sequence x as the shortest length of a program which produces this sequence. Thus, a sequence consisting of all 0s or a sequence 010101... both have very small Kolmogorov complexity because these sequences can be generated by simple programs; on the other hand, for a sequence of results of tossing a coin, probably the shortest program is to write print(0101...) and thus reproduce the entire sequence. Thus, when $K(x)$ is approximately equal to the length $\mathrm{len}(x)$ of a sequence,

this sequence is random, otherwise it is not. (The best source for Kolmogorov complexity is a book [19].)

However, the existing Kolmogorov complexity theory does not yet lead to a formalism describing when low-probability events do not happen; we must therefore extend the original Kolmogorov's idea so that it would cover this case as well.

4 First Formalization of Typicalness and Its Relation to Maxitive (Possibilistic) Measures

4.1 A More Adequate Formalization of Typicalness and the Corresponding Result

Let us start with motivations. We have mentioned that we cannot consistently claim that an event E is possible if and only its probability $p(E)$ exceeds a certain threshold p_0; instead, we must take into consideration that "complexity" $c(E)$ of an event, and claim, e.g., that an event E is possible if and only if $p(E) > p_0 \cdot c(E)$, i.e., equivalently, $m(E) > p_0$, where we denoted $m(E) \stackrel{\text{def}}{=} p(E)/c(E)$.

Comment. To handle events with 0 probability, we must extend the ratio $m(E)$ to such events – otherwise, e.g., for the uniform distribution on the interval $[0,1]$, we would have $p(\{x\}) = 0 \le p_0 \cdot c(\{x\})$ hence no point x would be possible.

We would like to characterize the "ratio measures" $m(E)$ for which this definition is, in some reasonable sense, "consistent" for all possible thresholds p_0. In order to do that, let us first find out how to formalize the notion of "consistency".

Let X be the set of all possible outcomes. An *event* is then simply a subset E of the set X, and p is a probability measure on a σ-algebra of sets from X.

Let $T \subseteq X$ be the set of all outcomes that are actually possible. Then, an event E is possible if and only if there is a possible outcome that belongs to the set E, i.e., if and only if $E \cap T \neq \emptyset$.

Now, we are ready for the main definition:

Definition 1. *Let X be a set, and let $\mathcal{A} \subseteq 2^X$ be a σ-algebra of subsets of the set X. By a* ratio measure *m we mean a mapping from $\mathcal{A}$ to the set of non-negative real numbers (and, possibly a value $+\infty$) such that for every real number $p_0 > 0$, there exists a set $T(p_0)$ for which*

$$\forall E \in \mathcal{A}\,(m(E) > p_0 \leftrightarrow E \cap T(p_0) \neq \emptyset). \tag{1}$$

To describe our main result, we need to recall the definition of a maxitive (possibility) measure [8, 22, 24]:

Definition 2. *A mapping m from sets to real numbers (and possibly a value $+\infty$) is called a* maxitive (possibility) *measure if for every family of sets E_α for which $m(E_\alpha)$ and $m(\cup E_\alpha)$ are defined, we have*

$$m\left(\bigcup_\alpha E_\alpha\right) = \sup_\alpha m(E_\alpha).$$

Comment. Our definition is slightly more general than the definitions from [8, 22, 24]:

- in [8, 24], only values $m(E) \in [0,1]$ are allowed, while we allow arbitrary non-negative values;
- in [22], the maxitive property is only required for finite families E_α, while we allow arbitrary (in particular, infinite) families of sets.

Theorem 1. *For a given probability measure $p(E)$, a function $m(E)$ is a ratio measure if and only if it is a maxitive (possibility) measure.*

Comment about the result. Since $m(E) = p(E)/c(E)$ is a possibility measure, we thus have $c(E) = m(E)/p(E)$. In other words,

$$\text{complexity} = \frac{\text{possibility}}{\text{probability}}.$$

Comment about the proof. For the reader's convenience, all the proofs are placed in a special Appendix.

Comment about the result and about the proof. This result is in perfect accordance with a recent paper by D. Dubois, H. Fargier, and H. Prade [7] in which the authors proved that the only uncertainty theory coherent with the notion of accepted belief is possibility theory.

Moreover, even our proof is similar to the proofs from [7, 8]. In principle, we could drastically shorten our proof if, instead of presenting our proof step by step, we would instead simply explain what needs to change in these proofs. This would be sufficient for specialists in possibility theory.

However, one of our main objectives is to win more converts for the possibility theory. For the benefit of these potential converts who are not yet familiar with the possibility theory and its proofs, we decided to present the proof "from scratch".

4.2 Auxiliary Result Related to the Above Formalization of Typicalness

Our definition of complexity depends on the choice of the probability measure. In other words, complexity of an event depends on the problem that we are trying to solve. This makes sense because what we are looking for is complexity relevant to the problem.

However, a natural question is: is it possible to have a "universal" complexity measure, i.e., a complexity measure that will serve all possible probability measures $p(E)$? The answer is "no", even if, instead of all possible thresholds p_0, we just consider a single one. This result is true even for X equal to the standard interval $[0, 1]$.

Let us describe this result in precise terms.

Definition 3. *Let $X = [0, 1]$, and let $\mathcal{A} \subseteq 2^X$ be a σ-algebra of all Lebesgue-measurable sets. By a* universal complexity measure *c we mean a mapping from $\mathcal{A}$ to the interval $[0, 1]$ for which $0 < c([a, b]) < 1$ for every interval $[a, b]$, and for every probability measure p on $\mathcal{A}$, there exists a set $T[p]$ for which*

$$\forall E \in \mathcal{A}\,(p(E) > 0 \to (p(E) > c(E) \leftrightarrow E \cap T[p] \neq \emptyset)).$$

Theorem 2. *A universal complexity measure is impossible.*

Comment 1. Since we gave a rationale that $m(E)$ has the form $p(E)/c(E)$, and another rationale for having $m(E)$ maxitive, the impossibility result is not unexpected.

Comment 2. The term "complexity" has many meanings; in this paper, we only use it as a way to describe which rare events are possible and which are not. From this viewpoint, Theorem 2 is not so much a theorem complex complexity but rather a result about the impossibility of representing physicists' reasoning about rare events in purely probabilistic terms – a result similar, in spirit, to results from [7].

5 Second Formalization of Typicalness

5.1 Motivations and Definition

"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 $\geq$ 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 in 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 $\geq$ 6 ft, A_2 is the set of all people whose height is $\geq$ 6 ft 1 in, A_3 is the set of all people whose height is $\geq$ 6 ft 2 in, etc.

In the weight example, A_1 is the set of all people whose weight is $\geq$ 150 lb, A_2 is the set of all people whose weight is $\geq$ 160 lb, A_3 is the set of all people whose weight is $\geq$ 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 [15]. 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 4. *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.

5.2 Possible Practical Use of the Second Formalization of Typicalness: 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 $d_X(x_k, x_{k+1}) \le \delta$ 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., [12]).

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 $d_X(f(x), y) \le \delta$ for some $\delta > 0$.

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

Definition 5. *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 $d_X(x_k, \lim x_p) \le \varepsilon$.*
- *Let $d \ge 1$ be an integer. By a* stopping criterion*, we mean a function $c : X^d \to R_0^+ = \{x \in R \,|\, x \ge 0\}$ 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 1. *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 $c(x_k, \ldots, x_{k+d-1}) \le \delta$, then x_k is $\varepsilon-$accurate.*

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

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

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 ($|a_{k+1}| \le \delta$), then we are $\varepsilon-$close to the sum: $|x_k - x| \le \varepsilon$.

5.3 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 $|r_k - x| \le 2^{-k}$). 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 $X \stackrel{\rm def}{=} (x, x', \ldots, x'')$, 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 $g(X) \stackrel{\rm def}{=} -f(X)$.) In general, it is algorithmically impossible to decide whether a given constructive number is non-negative or not; see, e.g., [17].

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 [17]. 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.

5.4 Relation between the Second Definition of Typicalness and 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 2. $N(\{A_n \cup B_n\}) = \max(N(\{A_n\}), N(\{B_n\}))$.

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 2 leads to a more general result:

Proposition 3. *For every definable sequence of sequences* $\{A_n^\alpha\}$, $N(\{\bigcup_\alpha A_n^\alpha\}) = \max_\alpha(N(\{A_n^\alpha\}))$.

5.5 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 $a \in Ab_k(A)$.

5.6 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 6. *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 7. *We say that a sequence of sets $\{X_i\}$ is* $\cup$-independent *if for all i, $X_i \not\subseteq \bigcup_{j \neq i} X_j$.*

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 $p(T(p)) \geq 1 - \varepsilon$.*

5.7 Auxiliary Result on the Second Definition of Typicalness: Relation between $T(A)$ for Different A

What is the relation between $T(A)$ for different sets A? Some such properties are described in [15]; 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 $T(A) \not\subseteq C$.*

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.

6 Typicalness and Boundedness

6.1 Second Definition of Typicalness: Advantages and Limitations

As we have seen, the above second definition of typicalness seems to be a reasonably adequate formalization of the physicists' reasoning. Not only the definition itself reflects the physicists' intuition, but the corollaries of this definition seems to be in agreement with this intuition: iterative processes stop, "physically" solvable problems become formally algorithmically solvable, and paradoxes of commonsense reasoning turn out to be natural consequences of our general definition.

In short, from the viewpoint of adequate formalization of physical intuition, the above definition has definite advantages. However, this definition also has a serious limitation:

- on the one hand, our main objective is to formalize physicist reasoning about typicalness, i.e., describe this reasoning in precise *mathematical* form;

- on the other hand, from the purely mathematical viewpoint, the above definition is very complex and difficult to understand.

Since the above definition seems adequate, we would like to encourage its use. So, the enhance the use of this second definition of typicalness, it is desirable to reformulate it in clearer more mathematically natural terms.

This is exactly what we plan to do in this part of the paper. Specifically, we will show that our definition of typicalness can be naturally viewed as a particular case of a general notion of *boundedness*.

6.2 Towards A General Notion of Boundedness

We would like to describe the general notion of boundedness: what it means for a set to be bounded, what it means for a topological space to be bounded, etc. In general, in mathematics, we use different terms to handle different types of objects:

- when we deal with sets, we talk about sets and functions between sets;
- when we talk about topological spaces, we take about such spaces, open, closed, and compact sets within these spaces, and continuous mappings (functions) between such spaces.

To be able to give a general over-the-board definition of boundedness, we need to be able to have a general language in which we can talk about different types of mathematical objects. In mathematics, there is such a language: it is the language of *category theory*; see, e.g., [1, 2].

A category is a pair consisting of *objects* and *morphisms*. To explain this notion, let us give a few examples:

- We may have a category Set in which objects are arbitrary sets, and morphisms are arbitrary functions between these sets.
- We may have a category Top in which objects are topological spaces, and morphisms are continuous functions.
- We may have a category Ord of ordered sets in which objects are ordered sets and morphisms are monotonic mappings between ordered sets.
- We may have a category of linear spaces in which objects are linear spaces, and morphisms are linear mappings, etc.

In this general language, we can talk about the properties of objects and morphisms and this will cover all above examples.

Formally, a category can be defined as follows. To every morphism f, we associate the "from" object a and the "to" object b; we usually say that f is a morphism from a to b and denote it by $f : a \to b$. When $f : a \to b$ and $g : b \to c$, then there exists a *composition* $f \circ g : a \to c$, for which $(f \circ g) \circ h) = f \circ (g \circ h)$. Finally, for every object a, there is a *identity morphism* id_a such that $f \circ id_a = id_a \circ f$ for every f.

In all above examples, composition is a normal composition of functions and id_a is an identity mapping, i.e., a mapping for which $f(x) = x$ for all x.

How can we use category theory to define a general notion of boundedness? The notion of boundedness is very clearly defined for subsets of the real line $\mathbb{R}$: a set $S \subseteq \mathbb{R}$ is bounded if and only if there exist bounds $\underline{b}$ and $\overline{b}$ such that $S \subseteq [\underline{b}, \overline{b}]$, i.e., such that $\underline{b} \le s \le \overline{b}$ for all $s \in S$. It is therefore reasonable to propose the following general definition.

In this paper, we will be restricting ourselves to *concrete* categories, i.e., categories in which all objects are sets (maybe with additional structure), and all morphisms are functions between these sets. All above examples of categories are concrete categories.

Definition 8. *Let $\mathcal{C}$ be a concrete category that contains the set of real numbers $\mathbb{R}$.*

- *We say that an object A is bounded if for every morphism $f : A \to \mathbb{R}$, the image $f(A)$ is bounded.*
- *We say that a subset $S \subseteq A$ of an object A is bounded if for every morphism $f : A \to \mathbb{R}$, the image $f(S)$ is bounded.*

This definition leads to reasonable properties:

Proposition 6.

- *a subset of a bounded set is always bounded;*
- *a finite union of bounded set is always bounded;*
- *an image of a bounded set under a morphism is bounded.*

Let us show that for natural categories, this definition leads to reasonable notions.

6.3 Boundedness for Sets

Proposition 7. *In the category of sets:*

- *a set A is bounded if and only if it is finite;*
- *a subset $S \subseteq A$ is bounded if and only if S is a finite set.*

6.4 Boundedness for Ordered Sets

Let us now describe a similar result for ordered sets and monotonic mappings, i.e., mappings f for which $x \le y$ implies $f(x) \le f(y)$.

An order set is called *linearly* (or *totally*) ordered if for every a and b, we have $a \le b$ or $b \le a$.

An ordered set is called *separable* if there exists a countable subset $\{x_n\}$ such that is $a < b$, there there exists an integer i for which $a \le x_i \le b$; see, e.g., [4].

Proposition 8. *In the category of separable linearly (totally) ordered sets:*

- *a set A is bounded if and only if it has the largest element and the smallest element;*
- *a subset $S \subseteq A$ is bounded if and only if it is bounded in A, i.e., if and only if there exist elements $\underline{b}$ and $\overline{b}$ such that $\underline{b} \le s \le \overline{b}$ for all $s \in S$.*

An interesting particular case of a linearly ordered set is an *ordinal* number; see, e.g., [23]. Such "numbers" include $1, 2, \ldots, n, \ldots, \omega, \omega + 1, \ldots, \omega + \omega, \ldots$. Some ordinal numbers like ω and $\omega + \omega$ are *limit ordinals*, i.e., ordinals which cannot be represented as $\beta + 1$ for some β. Alternatively, limit ordinals can be defined as the ones in which there is no largest element. So, we arrive at the following corollary:

Corollary. *In the category of linearly ordered sets and monotonic morphisms, a countable ordinal α is bounded if and only if it is not a limit ordinal.*

Our result about linearly ordered sets can be extended to partial orders which are *directed*, i.e., for which for every a and b, there exists elements l and u such that $l \le a \le u$ and $l \le b \le u$. It is worth mentioning that *lattices*, i.e., ordered sets in which every two elements has the greatest lower bound and the smallest upper bound, are a particular case of directed ordered sets.

Proposition 9. *In the category of separable directed ordered sets:*

- *a set A is bounded if and only if it has the largest element and the smallest element;*
- *a subset $S \subseteq A$ is bounded if and only if it is bounded in A, i.e., if and only if there exist elements $\underline{b}$ and $\overline{b}$ such that $\underline{b} \le s \le \overline{b}$ for all $s \in S$.*

For linear orders and directed orders, boundedness means that the set S is contained in an interval – because if it is contained in several intervals, then we can embed their union into a single one. For a general partially ordered set, we may have to have several intervals bounding a given set.

For this case, we need to somewhat modify the notion of separability.

Definition 9. *We say that a (partially) ordered set A is* p-separable *if there exists a countable set $\{x_i\}$ such that for every $a \in A$ there exist i and j such that $x_i \le a \le x_j$.*

Proposition 10. *In the category of p-separable partially ordered sets:*

- *a set A is bounded if and only if there exist elements $M_1, \ldots, M_k$ and $m_1, \ldots, m_l$ such that every element $a \in A$ is smaller than one of the M_i and larger than one of the m_j;*
- *a subset $S \subseteq A$ is bounded if and only if it is bounded in A, i.e., if and only if there exist elements $M_1, \ldots, M_k$ and $m_1, \ldots, m_l$ such that every element $s \in S$ is smaller than one of the M_i and larger than one of the m_j.*

6.5 Boundedness for Topological Spaces

The set of real numbers is not only an ordered set, it is also a topological space, so it is natural to analyze what boundedness means for topological spaces.

Let us recall that a topological space A is called *separable* if there exists a sequence $\{x_n\}$ which is everywhere dense in A, i.e., for which the closure of the set $\{x_n\}$ coincides with A. A topological space is called *normal* if for every tow disjoint closed sets S_1 and S_2 that exist open disjoint sets $U_1 \supseteq S_1$ and $U_2 \supseteq S_2$. In particular, all metric spaces are normal; see, e.g., [13].

Proposition 11. *In the category of separable normal topological spaces and continuous mappings:*

- *a space A is bounded if and only if it is compact;*
- *a subset $S \subseteq A$ is bounded if and only if its closure is compact in A.*

6.6 Boundedness for Abelian Groups

The set of real numbers is also an Abelian group with respect to addition.

Proposition 12. *In the category of groups and homomorphisms, an Abelian group G is bounded in the above general sense if and only if every element of this group has a finite order.*

6.7 Typicalness as a Particular Case of Boundedness

Now that we showed that the above definition of boundedness is natural, we can prove the following (rather surprising) result: that our seemingly complex second definition of typicalness is actually a simple particular case of this general notion of boundedness.

For that, let us consider the category Def in which objects are definable sets, and morphims are definable functions.

Proposition 13. *In the category Def of definable sets and definable mappings, a subset $T \subseteq U$ of a set U is bounded if and only if T is a set of typical elements.*

Comment. Thus, the seemingly complex notion of randomness is, in fact, very mathematically natural.

7 Conclusion

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 showed 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.

We also showed that the resulting difficult-to-described definition can be actually viewed as a particular case of something very natural: the general notion of boundedness.

Acknowledgments

This work was supported by NASA grant NCC5-209, by USAF grant F49620-00-1-0365, by NSF grants EAR-0112968, EAR-0225670, and EIA-0321328, by Army Research Laboratories grant DATM-05-02-C-0046, and by the NIH grant 3T34GM008048-20S1.

The authors are thankful to Vladik Kreinovich and Hung T. Nguyen for their help.

References

[1] J. Adámek, H. Herrlich, and G. E. Strecker, *Abstract and Concrete Categories*, John Wiley & Sons, New York, 1990.

[2] A. Awodey, *Category Theory*, Clarendon Press, Oxford, 2006.

[3] L. Boltzmann, "Bemrkungen über einige Probleme der mechanischen Wärmtheorie", *Wiener Ber. II*, 1877, Vol. 75, pp. 62–100.

[4] D. S. Bridges and G. B. Mehta, *Representations of Preference Orderings*, Springer Verlag, Berlin-Heidelberg, 1995.

[5] G. Coletti and R. Scozzafava, *Probabilistic Logic in a Coherent Setting*, Kluwer Academic Publishers, Dordrecht, 2002.

[6] B. de Finetti, *Theory of Probability*, Wiley, New York, 1970.

[7] D. Dubois, H. Fargier, and H. Prade, "Ordinal and probabilistic representations of acceptance", *J. Artificial Intelligence Research*, 2004, Vol. 22, pp. 23–56.

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

[9] 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", *Proceedings of the 2005 IEEE International Conference on Fuzzy Systems FUZZ-IEEE'2005*, Reno, Nevada, May 22–25, 2005, pp. 1044–1049.

[10] A. M. Finkelstein, O. Kosheleva, V. Kreinovich, S. A. Starks, and H. T. Nguyen, "Use of Maxitive (Possibility) Measures in Foundations of Physics and Description of Randomness: Case Study", *Proceedings of the 24th International Conference of the North American Fuzzy Information Processing Society NAFIPS'2005*, Ann Arbor, Michigan, June 22–25, 2005, pp. 687–692.

[11] R. P. Feynman, *Statistical Mechanics*, W. A. Benjamin, 1972.

[12] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures*, Addison-Wesley, 1965.

[13] J. L. Kelley, *General Topology*, Springer-Verlag, Berlin-Heidelberg-New York, 1975.

[14] O. M. Kosheleva, "On a general notion of boundedness", *Notices of the American Mathematical Society*, 1979, Vol. 26, No. 1, p. A-2 (publ. No. 79T-A3).

[15] V. Kreinovich, "Toward Formalizing Non-Monotonic Reasoning in Physics: the Use of Kolmogorov Complexity and Algorithmic Information Theory to Formalize the Notions 'Typically' and 'Normally' ", In: L. Sheremetov and M. Alvarado (eds.), *Proceedings of the Workshops on Intelligent Computing WIC'04 associated with the Mexican International Conference on Artificial Intelligence MICAI'04*, Mexico City, Mexico, April 26–27, 2004, pp. 187–194.

[16] V. Kreinovich and A. M. Finkelstein, "Towards Applying Computational Complexity to Foundations of Physics", *Notes of Mathematical Seminars of St. Petersburg Department of Steklov Institute of Mathematics* (to appear).

[17] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1997.

[18] H. E. Kyburg, Jr., *Probability and the logic of rational belief*, Wesleyan Univ. Press, 1961.

[19] M. Li and P. M. B. Vitanyi, *An Introduction to Kolmogorov Complexity*, Springer, N.Y., 1997.

[20] 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.

[21] 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.

[22] A. Puhalskii, *Large Deviations and Idempotent Probability*, Chapman and Hall/CRC, Boca Raton, 2001.

[23] P. Suppes, *Axiomatic Set Theory*, Dover, New York, 1972.

[24] L. Zadeh, "Fuzzy sets as a basis for a theory of possibility", *Fuzzy Sets and Systems*, 1978, Vol. 1, pp. 3–28.

A Appendix: Proofs

A.1 Proof of Theorem 1

Let us first prove that every ratio measure $m(E)$ is indeed a maxitive measure. By definition of a maxitive measure, we need to prove that if E_α is a family of sets from the σ-algebra $\mathcal{A}$ for which the union $E = \cup E_\alpha$ also belongs to $\mathcal{A}$, we have $m(E) = \sup_\alpha m(E_\alpha)$.

Let us prove this inequality by reduction to a contradiction. Let us assume that $m(E) \neq \sup_\alpha m(E_\alpha)$. In this case, we have two options:

- $m(E) > \sup_\alpha m(E_\alpha)$ and
- $m(E) < \sup_\alpha m(E_\alpha)$.

Let us show that in both cases, we have a contradiction.

Indeed, by definition of a ratio measure, for every p_0, there exists a set $T(p_0)$ such that for every set S, we have $m(S) > p_0$ if and only if $S \cap T(p_0) > 0$.

If $m(E) < \sup_\alpha m(E_\alpha)$, let us select p_0 for which

$$m(E) < p_0 < \sup_\alpha m(E_\alpha).$$

Since $m(E) < p_0$, we conclude that the event E is not possible, i.e.,

$$E \cap T(p_0) = \emptyset. \tag{2}$$

On other hand, since $p_0 < \sup_\alpha m(E_\alpha)$, there exists a value α_0 for which $p_0 < m(E_{\alpha_0})$. For this α_0, by definition of a ratio measure, the event E_{α_0} is possible, so there exists an outcome x from E_{α_0} that also belongs to the set $T(p_0)$ of possible events. However, since $E = \cup E_\alpha$, we have $x \in E$, so $x \in E \cap T(p_0)$ – which contradicts our previous conclusion (2). This contradiction shows that the inequality $m(E) < \sup_\alpha m(E_\alpha)$ is impossible.

If $m(E) > \sup_\alpha m(E_\alpha)$, let us select p_0 for which

$$m(E) > p_0 > \sup_\alpha m(E_\alpha).$$

Since $m(E) > p_0$, we conclude that the event E is possible, i.e., there exist an outcome x that belongs both to E and to $T(p_0)$. Since E is the union of the set E_α, this event x belongs to one of the sets E_{α_0}. Thus, $E_{\alpha_0} \cap T(p_0) \neq \emptyset$, so by definition of a complexity measure, we should have

$$m(E_{\alpha_0}) > p_0 \tag{3}$$

for this α_0. However, from our assumption $m(E) > \sup_\alpha m(E_\alpha)$ and from the fact that $\sup_\alpha m(E_\alpha) \geq m(E_{\alpha_0})$, we conclude that $m(E_{\alpha_0}) < p_0$ – a contradiction with our previous conclusion (3). This contradiction shows that the inequality $m(E) > \sup_\alpha m(E_\alpha)$ is also impossible.

Thus, every ratio measure $m(E)$ is indeed a maxitive measure.

To complete the proof of Theorem 1, we must now prove that if $m(E)$ is a maxitive measure, then it is a ratio measure. To prove this, we will show that for every positive real number p_0, there exists a set $T(p_0)$ that satisfies the condition (1). We will show that as such a set, we can take a complement to the union of all sets $S \in \mathcal{A}$ for which $m(S) \leq p_0$, i.e.,

$$T(p_0) = - \cup \{S \in \mathcal{A} \,|\, m(S) \leq p_0\}. \tag{4}$$

We must prove that for every $E \in \mathcal{A}$, $E \cap T(p_0) \neq \emptyset$ if and only if $m(E) > p_0$. Actually, we will prove an equivalent statement: that for every $E \in \mathcal{A}$, $E \cap T(p_0) = \emptyset$ if and only if $m(E) \leq p_0$.

If $m(E) \leq p_0$, then E is completely contained in the union $\cup\{S \in \mathcal{A} \,|\, m(S) \leq p_0\}$, thus, E cannot have common points with the complement $T(p_0)$ to this union.

Vice versa, let us assume that for some event $E \in \mathcal{A}$, we have $E \cap T(p_0) = \emptyset$. This means that the set E is completely contained in the complement to $T(p_0)$, i.e.,

$$E \subseteq \cup\{S \in \mathcal{A} \,|\, m(S) \leq p_0\}.$$

Thus,

$$E = \cup\{S \cap E \,|\, S \in \mathcal{A} \,\&\, m(S) \leq p_0\}. \tag{5}$$

If the set S and E belongs to a σ-algebra, then their intersection and their difference also belong to the σ-algebra. From $S = (S \cap E) \cup (S - E)$ and the definition of a maxitive measure, we thus conclude that $m(S) = \max(m(S \cap E), m(S - E))$ hence $m(S \cap E) \le m(S)$. So, if $m(S) \le p_0$, we have $m(S \cap E) \le m(S) \le p_0$ hence $m(S \cap E) \le p_0$.

Applying the definition of a maxitive measure to the formula (5), we can now conclude that $m(E) = \sup m(S \cap E)$, where supremum is taken over all $S \in \mathcal{A}$ for which $m(S) \le p_0$. We have already shown that for all such S, we have $m(S \cap E) \le p_0$. Thus, $m(E)$ is the supremum of a set of numbers each of which is $\le p_0$. We can therefore conclude that $m(E) \le p_0$.

The theorem is proven.

A.2 Proof of Theorem 2

To prove this theorem, we will assume that a universal complexity measure $c(E)$ exists, and from this assumption, we will deduce a contradiction.

First, let us show that if $A \subset B$ are two sets from the σ-algebra $\mathcal{A}$ for which $B \ne X$, then $c(A) \ge c(B)$. Indeed, let us prove that if $c(A) < c(B)$, then we get a contradiction.

If $c(A) < c(B)$, then we can set up a probability measure p for which $p(A) = c(B) > 0$ and $p(B - A) = 0$. For this probability measure, $p(B) = p(A) + p(B - A) = c(B)$, hence $p(B) \le c(B)$ and $p(B) > 0$. By definition of a universal complexity measure, this means that the set B has no common points with $T[p]$. Since A is a subset of B, it also has no common points with $T[p]$; due to $p(A) > 0$, we should have $p(A) \le c(A)$. However, $p(A) = c(B) > c(A)$ – a contradiction shows that the case $c(A) < c(B)$ is impossible.

Let us now show if $A \subset B$ and $c(B) > 0$, then $c(A) = c(B)$. We already know that $c(A) \ge c(B)$. Thus, it is sufficient to show that if $c(A) > c(B)$, then we get a contradiction.

Indeed, since $c(B) > 0$ and $B - A \subseteq B$, we have

$$c(B - A) \ge c(B) > 0.$$

Let $c(A) > c(B)$, then we can set up a probability measure p for which $p(A) = c(A)$ and

$$0 < p(B - A) \le c(B - A).$$

For this probability measure, $p(A) \le c(A)$ and $p(A) > 0$, hence the set A cannot have any common points with $T[p]$. Similarly, since $0 < p(B - A) \le c(B - A)$, the set $B - A$ cannot have any common points with $T[p]$. Since neither the set A nor the set $B - A$ can have common points with $T[p]$, their union $B = A \cup (B - A)$ also cannot have any common points with $T[p]$. According to the definition of a universal complexity measure and the fact that $p(B) > 0$, this would mean that $p(B) \le c(B)$, but $p(B) > c(A) > c(B)$. The contradiction shows that the case $c(A) > c(B)$ is also impossible.

So, $A \subseteq B$ and $c(B) > 0$ imply that $c(A) = c(B)$.

Let $[a, b]$ be an arbitrary interval $\ne [0, 1]$. Then, by definition, $c([a, b]) > 0$, so, for every set $E \subseteq [a, b]$, we have $c(E) = c([a, b])$. Let us select an integer $n > 1/c([a, b])$ and divide the interval $[a, b]$ into n subintervals of equal size. For the uniform distribution on the interval $[a, b]$, the probability $p(E)$ of each subinterval E is equal to $1/n$. Since $n > 1/c([a, b])$, we thus conclude that $p(E) = 1/n \le c([a, b])$, i.e., $p(E) < c(E) = c([a, b])$. Thus, none of these n subintervals can contain elements from $T[p]$. On the other hand, $p([a, b]) = 1 > c([a, b])$ hence the union $[a, b]$ of these n subintervals does contain elements from $T[p]$ – a contradiction.

The theorem is proven.

A.3 Proof of Proposition 1

As A_n, we will take the set of all sequences for which for some k, $c(x_k, \ldots, x_{k+d-1}) \le 2^{-n}$ 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 $c(x_{k(n)}, \ldots, x_{k(n)+d-1}) \le 2^{-n}$ and $d_X(x_{k(n)}, x) > \varepsilon$. If some value k is equal to $k(n)$ for infinitely many n, this means that $c(x_k, \ldots, x_{k+d-1}) \le 2^{-n}$ 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.

A.4 Proof of Proposition 2

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.

A.5 Proof of Proposition 4

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, $\sum_{i=1}^{\infty} c(X_i) > 1$. In this case, for some finite n, we have $s \stackrel{\text{def}}{=} \sum_{i=1}^{n} c(X_i) = 1$. Since the family $\{X_i\}$ is $\cup$-independent, for every $i \le n$, there exist $x_i \in X_i \setminus \bigcup_{j \ne i} X_j$. 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 $X_i \ne x_i$ 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 $\ne x_i$ are also impossible – which contradicts to $T(p) \ne \emptyset$.

Let us now assume that for every p, there exists a consistent set $T(p)$ for which $p(T(p)) \ge 1 - \varepsilon$. Let us prove that $\sum c(X_i) \le \varepsilon$. 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 $\cup X_i \ne U$, 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 $\sum c(X_i) \le \varepsilon$ 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 $T(p) \ne \emptyset$, we will prove that for every n, after excluding n such sets, the remainder T_n has a measure $\ge 1 - \varepsilon$. Thus, the limit set $T(p)$ – after excluding (countably many) definable sets for which $p(A) < c(A)$ – will also have measure $\ge 1 - \varepsilon > 0$.

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.

A.6 Proof of Proposition 5

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.

A.7 Proof of Proposition 6

Let us first prove that a subset S' of a bounded set S is always bounded. Indeed, by definition, the fact that the set S is bounded means that $f(S)$ is bounded for all morphisms f. Due to the fact that S' is a subset of S, we have $f(S') \subseteq f(S)$. Since the set $f(S)$ is bounded, its subset $f(S')$ is bounded as well. Thus, the set $f(S')$ is bounded for every morphism f, i.e., the subset S' is indeed bounded.

Let us now prove that the finite union $S = S_1 \cup \ldots \cup S_n$ of bounded sets $S_1, \ldots, S_n$ is bounded. Indeed, since each set S_i are bounded, for each $f : A \to \mathbb{IR}$, its image $f(S_i)$ is a bounded subset of the real line. The image $f(S)$ is equal to the union $f(S_1) \cup \ldots \cup f(S_n)$ of bounded subset of $\mathbb{IR}$ and is, therefore, also bounded. So, the union S is bounded in the sense of our definition.

Finally, let us prove that the image $S' = f(S)$ of a bounded set $S \subseteq A$ under a morphism $f : A \to B$ is also bounded. For that, we must prove that for every morphism $g : B \to \mathbb{IR}$, the image $g(S')$ is bounded. Indeed, for every g, the composition $f \circ g : A \to \mathbb{IR}$ is a morphism from A to $\mathbb{IR}$. Since S is a bounded subset of A, we thus conclude that the image $(f \circ g)(S)$ is a bounded subset of the real line. However, $(f \circ g)(S) = g(f(S)) = g(S')$. Thus, the image $g(S')$ of the set S' under every morphism is indeed bounded. The proposition is proven.

A.8 Proof of Proposition 7

If a set A is finite, i.e., if $A = \{a_1, \ldots, a_n\}$ for some n, then $f(A) = \{f(a_1), \ldots, f(a_n)\}$, so $f(A) \in [\underline{b}, \overline{b}]$ for $\underline{b} = \min(f(a_1), \ldots, f(a_n))$ and $\overline{b} = \max(f(a_1), \ldots, f(a_n))$. Similarly, if S is a finite subset of A, then the image $f(S)$ is always finite.

To complete the proof, let us show that if the set A is infinite, then there exists a function $f : A \to \mathbb{IR}$ for which the image $f(A)$ is unbounded. Since the set A is infinite, it contains an infinite sequence $\{x_n\} \subseteq A$: we pick any element x_1; since A is infinite, it cannot be exhausted by this element, so there exists $x_2 \in A - \{x_1\}$. Similarly, x_1 and x_2 cannot exhaust A, so there exists x_2 such that $x_3 \neq x_1$ and $x_3 \neq x_2$, etc. We can now define $f(x)$ as follows: $f(x_n) = n$ and $f(x) = 0$ for $x \notin \{x_n\}$. For this function f, the image $f(A)$ consists of all natural numbers and is, thus, unbounded.

A similar function can be defined if we instead of an infinite set A, we consider an infinite subset $\subseteq A$. The proposition is proven.

A.9 Proof of Proposition 8

Bounded in the sense of order → bounded in the sense of the new general definition. Let us first prove that an ordered set A with the largest element M and the smallest element m is bounded in the sense of our definition. Indeed, the fact that m is the smallest element means that $m \leq a$

for every $a \in A$, and the fact that M is the largest element means that $a \le M$ for all $a \in A$. For every monotonic mapping $f : A \to \mathbb{R}$ and for every $a \in A$, we have $m \le a \le M$ hence (due to monotonicity) $f(m) \le f(a) \le f(M)$. Thus, the image $f(A)$ is bounded by the values $f(m)$ and $f(M)$.

Similarly, let us assume that for a set $S \subseteq A$ there exist elements $\underline{b}$ and $\overline{b}$ such that $\underline{b} \le s \le \overline{b}$ for all $s \in S$. Then, for every monotonic function $f : A \to \mathbb{R}$, we have $f(\underline{b}) \le f(s) \le f(\overline{b})$. Hence, the image $f(S)$ is bounded by the values $f(\underline{b})$ and $f(\overline{b})$.

Bounded in the sense of the new general definition → bounded in the sense of order. Vice versa, let us now assume that a linearly ordered set A is not bounded. Without losing generality, we can assume that the set A does not have the largest element, i.e., that $\neg\exists a \forall b (a \ge b)$. This means that $\forall a \exists b (a \not\ge b)$. Since the set A is linearly ordered, $a \not\ge b$ means $a < b$. Thus, for every element $a \in A$, there exists a larger element $b > a$.

Since the set A is separable, there exists a sequence $\{x_n\}$ such that for every a and b, if $a < b$, then $a \le x_i \le b$ for some i. Let us use this sequence to build an auxiliary sequence $\{y_n\}$:

- As y_1, we take an arbitrary element which is larger than x_1 (it exists since the set A has no largest element).
- Once $y_1, \ldots, y_k$ are defined:
  - we find the largest $\max(y_k, x_{k+1})$ of the elements y_k and x_{k+1} (it exists since the order on A is linear), and
  - as y_{k+1}, we take an element which is larger than $\max(y_k, x_{k+1})$ (it exists since the set A has no largest element).

By definition, $y_{k+1} > \max(y_k, x_{k+1}) \ge y_k$, hence $y_{k+1} > y_k$ – i.e., the new sequence $\{y_n\}$ is strictly increasing. Similarly, from $y_{k+1} > \max(y_k, x_{k+1}) \ge x_{k+1}$ we conclude that $y_{k+1} > x_{k+1}$; together with $y_1 > x_1$, we conclude that $y_k > x_k$ for all k.

We can now define the monotonic function $f : A \to \mathbb{R}$ as follows: for each $a \in A$, $f(a)$ is the largest n for which $a \ge y_n$ (and $f(a) = 0$ if no such n exists). First, let us show that such a largest n always exists. Indeed, if it did not exist for some $a_0 \in A$, then we would have $a_0 \ge y_k$ for all k. Since $y_k > x_k$, we would thus have $a_0 > x_k$ for all k. On the other hand, since the set A does not have the largest element, the element a_0 is not largest, so there exists a b such that $a_0 < b$. Due to separability, there exists i for which $a_0 \le x_i \le b$ — which contradicts to our previous conclusion that $a_0 > x_i$ for all i. This contradiction shows that the function f is everywhere defined.

It is also easy to check that the function f is monotonic: indeed, if $a \le b$ and $a \ge y_n$, then $b \ge y_n$. In particular, for $n = f(a)$, we have $b \ge y_{f(a)}$. Thus, the integer $f(b)$ – the largest n for which $b \ge y_n$ – is larger than or equal to $f(a)$.

Since y_n is a strictly increasing sequence, we conclude that $y_i \ge y_n$ for all $i \le n$ and $y_i < y_n$ for $n > i$. Thus, by definition of the function f, we have $f(y_i) = i$. Hence, the image $f(A)$ of the set A under f contains all natural numbers and is, thus, unbounded. The first part of the proposition is proven.

Let us now prove that if a set $S \subseteq A$ is not bounded in the order sense, then it is not not bounded in the sense of our general definition as well. Without losing generality, let us assume that the set S does not have an upper bound in A. This means, in particular, that the set A does not have the largest element – because otherwise, this largest element would be an upper bound for S. Thus, as in the previous proof, we can construct a sequence $\{y_n\}$ and a monotonic function $f : A \to \mathbb{R}$.

Let us prove that for this function, the image $f(S)$ is unbounded. Indeed, if it was bounded by some number C, we would have $f(s) \le C$ for all $s \in S$. By definition of the function $f(s)$ as the

largest number n for which $s \geq y_n$, the inequality $f(s) \leq C$ means that for any $n > C$, e.g., for $n_0 = \lceil C \rceil + 1$, we have $s < y_{n_0}$. Thus, this number y_{n_0} is an upper bound for the set S – contrary to our assumption that the set S has no upper bound.

The proposition is proven.

A.10 Proof of Proposition 9

Bounded in the sense of order → bounded in the sense of the new general definition. This part is proven exactly as in the linearly ordered case.

Bounded in the sense of the new general definition → bounded in the sense of order. Vice versa, let us now assume that a linearly ordered set A is not bounded. Without losing generality, we can assume that the set A does not have the largest element, i.e., that $\neg\exists a \forall b (a \geq b)$. This means that for every $a \in A$, there exists an element $b \in A$ such that $a \not\geq b$. Since A is directly ordered set, there exists an element c such that $c \geq a$ and $c \geq b$. We cannot have $c = a$, since then we would have $a \geq b$ – and we know that $a \not\geq b$. Thus, $c > a$. Thus, for every element $a \in A$, there exists a larger element $c > a$.

Since the set A is separable, there exists a sequence $\{x_n\}$ such that for every a and b, if $a < b$, then $a \leq x_i \leq b$ for some i. Let us use this sequence to build an auxiliary sequence $\{y_n\}$:

- As y_1, we take an arbitrary element which is larger than x_1 (we have proven the existence of such larger element for every $a \in A$).
- Once $y_1, \ldots, y_k$ are defined:
  - we use the directed order property to find an element z_{k+1} such that $z_{k+1} \geq y_k$ and $z_{k+1} \geq x_{k+1}$, and
  - as y_{k+1}, we take an element which is larger than z_{k+1} (the existence of such a larger element has been proven).

By definition, $y_{k+1} > z_{k+1} \geq y_k$, hence $y_{k+1} > y_k$ – i.e., the new sequence $\{y_n\}$ is strictly increasing. Similarly, from $y_{k+1} > z_{k+1} \geq x_{k+1}$ we conclude that $y_{k+1} > x_{k+1}$; together with $y_1 > x_1$, we conclude that $y_k > x_k$ for all k.

We can now define the monotonic function $f : A \to \mathbb{R}$ as follows: for each $a \in A$, $f(a)$ is the largest n for which $a \not\leq y_n$ (and $f(a) = 0$ if no such n exists). First, let us show that such a largest n always exists. Indeed, if it did not exist for some $a_0 \in A$, then we would have $a_0 \not\leq y_k$ for all k. Since $y_k > x_k$, we would thus have $a_0 \not< x_k$ for all k. On the other hand, for the element a_0, there exists a b such that $a_0 < b$. Due to separability, there exists i for which $a_0 \leq x_i \leq b$ — which contradicts to our previous conclusion that $a_0 \not\leq x_i$ for all i. This contradiction shows that the function f is everywhere defined.

It is also easy to check that the function f is monotonic: indeed, if $a \leq b$ and $a \not\leq y_n$, then $b \not\leq y_n$. In particular, for $n = f(a)$, we have $b \not\leq y_{f(a)}$. Thus, the integer $f(b)$ – the largest n for which $b \not\leq y_n$ – is larger than or equal to $f(a)$.

Since y_n is a strictly increasing sequence, we conclude that $y_i \geq y_n$ for all $i \leq n$ and $y_i < y_n$ for $n > i$. Thus, by definition of the function f, we have $f(y_i) = i$. Hence, the image $f(A)$ of the set A under f contains all natural numbers and is, thus, unbounded. The first part of the proposition is proven.

Let us now prove that if a set $S \subseteq A$ is not bounded in the order sense, then it is not not bounded in the sense of our general definition as well. Without losing generality, let us assume that the set S does not have an upper bound in A. This means, in particular, that the set A does not have the

largest element – because otherwise, this largest element would be an upper bound for S. Thus, as in the previous proof, we can construct a sequence $\{y_n\}$ and a monotonic function $f : A \to \mathbb{R}$.

Let us prove that for this function, the image $f(S)$ is unbounded. Indeed, if it was bounded by some number C, we would have $f(s) \le C$ for all $s \in S$. By definition of the function $f(s)$ as the largest number n for which $s \not\le y_n$, the inequality $f(s) \le C$ means that for any $n > C$, e.g., for $n_0 = \lceil C \rceil + 1$, we have $s \le y_{n_0}$. Thus, this number y_{n_0} is an upper bound for the set S – contrary to our assumption that the set S has no upper bound.

The proposition is proven.

A.11 Proof of Proposition 10

It is easy to prove that if there exist the desired bounds $M_1, \ldots, M_k$, $m_1, \ldots, m_l$, then the set A (correspondingly, its subset S) is bounded. Indeed, for every $a \in A$, we have $m_j \le a \le M_i$, so for a monotonic morphism, we have $f(m_j) \le f(a) \le f(M_i)$. Thus, $\min_j f(m_j) \le f(a) \le \max_i f(M_i)$. Hence, the image $f(A)$ is bounded by the values $\min_j f(m_j)$ and $\max_i f(M_i)$.

Similarly, for a subset S, the image $f(S)$ is bounded by the same two numbers.

Vice versa, let us assume that the set A is p-separable, with a countable set $\{x_n\}$, and that $S \subseteq A$ is bounded in the sense of our general definition. Let us then prove that it is bounded in the above ordered sense. For this, we need to prove that there exist upper bounds $M_1, \ldots, M_k$ such that every $s \in S$ is smaller than or equal to one of them, and that there exist lower bounds $m_1, \ldots, m_l$ such every $s \in S$ is larger than or equal to one of them. Without losing generality, let us prove the existence of upper bounds. Indeed, suppose that such bounds do not exist. This means that for every finite set of elements $M_1 \in A, \ldots, M_k \in A$, there exists an $s \in S$ such that $s \not\le M_1$, and $s \not\le M_2$, ..., and $s \not\le M_k$.

Let us now define a function $f : A \to \mathbb{R}$ as follows. For every $a \in A$, the value $f(a)$ is defined as the largest n for which $a \not\le x_1$, ..., and $a \not\le x_n$ (and $f(a) = 0$ if there is no such n). Let us first prove that the value $f(a)$ is defined for every a. Indeed, by definition of p-separability, we have $a \le x_i$ for some i, thus, $f(a)$ cannot exceed this i. It is also easy to prove that the function f is monotonic: indeed, if $a \not\le x_i$ and $a \le b$, then $b \not\le x_i$, so $f(a) \le f(b)$.

Due to the above-proven property of the set S, for every n, there exists $s_n \in S$ such that $s_n \not\le x_1$, ..., and $s_n \not\le x_n$. By definition of the function f, this means that $f(s_n) \ge n$. So, the image $f(S)$ of the set S contains arbitrarily large numbers – which contradict to our assumption that the image $f(S)$ is bounded. This contradiction proves that the desired bounds $M_1, \ldots, M_k$ do exist. The proposition is proven.

A.12 Proof of Proposition 11

Compact → bounded in the sense of the general definition. It is known that an image $f(A)$ of a compact set A is always compact, and that every compact subset of a real line is bounded. Thus, if A is compact or if the closure $\overline{S}$ is compact, then $f(A)$ (correspondingly, $f(\overline{S})$) is bounded – and hence, the image $f(S) \subseteq f(\overline{S})$ is bounded as well.

Bounded in the sense of the general definition → compact. For separable spaces, compactness is equivalent to the fact that every sequence has a convergent subsequence. Thus, the fact that the space A (or the closure $\overline{S}$) is not compact means that it contains a sequence $\{y_n\}$ that has no converging subsequence. Since it has no converging subsequence and the space is separable, this means that the set $Y \stackrel{\text{def}}{=} \{y_n\}$ is closed. On this set, we can define a function $f(y_n) = n$, and it will be continuous.

Since A is a normal topological space, we can apply Tietze extension theorem and conclude that this function f can be extended to a continuous function $f : A \to \mathbb{R}$.

If A is not compact, this means that we have a continuous function f for which the image $f(A)$ contains all natural numbers and is, thus, unbounded. So, the space A is not bounded in the sense of our general definition.

If the closure $\overline{S}$ is not compact, then the image $f(\overline{S})$ is unbounded. Let us show that in this case, the image $f(S)$ is unbounded as well. Indeed, every point $s \in \overline{S}$ is a limit of points from S: $s = \lim s_n$. Since the function f is continuous, we conclude that the value $f(s)$ is the limit of the corresponding values $f(s_n) \in f(S)$. Thus, the unbounded image $f(\overline{S})$ is contained in the closure of $f(S)$. Since $f(\overline{S})$ is unbounded, the image $f(S)$ is unbounded too. Thus, the set S is not bounded in the sense of our general definition. The proposition is proven.

A.13 Proof of Proposition 12

If an element a of an Abelian group G has a finite order, i.e., if $n \cdot a = 0$ for some natural number n, then for every homomorphism $f : G \to \mathbb{R}$, we have $f(n \cdot a) = n \cdot f(a)$ and $f(0) = 0$. Thus, $n \cdot f(a) = 0$, hence $f(a) = 0$. So, the image $f(G)$ consists of a single number 0 and is, hence, bounded.

Vice versa, let $a \in G$ be an element with an infinite order. Then, its multiples $\{n \cdot a\}$ form a subgroup G_0. Since the group G is Abelian, we can represent G as a Cartesian product $G_0 \times G_1$ for the factor-group $G_1 = G/G_0$. Thus, every element $g \in G$ can be represented as a pair $(n \cdot a, b)$ with $b \in G_1$. We can thus define a homomorphism $f : G \to \mathbb{R}$ as follows: $f(n \cdot a, b) = n$. The image of this homomorphism contains all integers and is, therefore, unbounded. The proposition is proven.

A.14 Proof of Proposition 13

Bounded → typical. Let us first show that if T is a set of typical elements in the sense of Definition 4, then T is bounded, i.e., $f(T)$ is bounded for every definable function $f : U \to \mathbb{R}$. Indeed, let us consider the sets $A_n = \{u : |f(u)| > n\}$. Since f is a definable map, this sequence A_n is definable – the above formula is its definition in terms of f.

Let us prove that this sequence satisfies both conditions from Definition 4.

- Since $|f(u)| > n + 1$ implies $|f(u)| > n$, we have $A_n \supseteq A_{n+1}$ for all n.
- By definition, $\bigcap_n A_n$ is the set of all the elements $u \in U$ for which $|f(u)| > n$ for all n – and no real number can be larger than all integers, so this intersection is empty.

Thus, we conclude that there exists an integer N for which $A_n \cap T = \emptyset$, i.e., for which $|f(u)| \le N$ for all $u \in T$. Thus, the image $f(T)$ is always bounded – i.e., the set T is bounded.

Typical → bounded. Vice versa, let us assume that a set $T \subseteq U$ is bounded. Let us prove that it is a set of typical elements in the sense of Definition 4. Indeed, let A_n be a definable sequence for which $A_n \supseteq A_{n+1}$ and $\cap A_n = \emptyset$. Without losing generality, we can add $A_0 = U$ to this sequence, both properties of this sequence will remain true. Then, we can define the following function $f : U \to \mathbb{R}$: for each $u \in U$, $f(u)$ is the largest integer n for which $u \in A_n$. This function is everywhere defined: indeed, if for some u, we had arbitrarily large integers n_k for which $u \in A_{n_k}$, then, due to monotonicity $A_n \supseteq A_{n+1}$, we would have $u \in A_n$ for all n – which would contradict to our assumption that $\bigcap_n A_n = \emptyset$.

Since A_n is a definable sequence, the above definition shows that this function f is definable. Since T is a bounded set, we therefore conclude that the set $f(T)$ is bounded, i.e., there exists an integer N for which $f(u) < N$ for all $u \in T$. By definition of the function f, this means that $u \notin A_N$ for all $u \in T$, i.e., that $A_N \cap T = \emptyset$. This is exactly what we need to prove that T is a set of typical elements.

The proposition is proven.