A. Ya. Temkin

PHYSICS WITHOUT SPACE AND TIME

IN MICRO- AND MACROWORLD:

1. ***HIERARCHY BASED MICROWORLD SCALES' CLASSIFICATION AND MICROWORLD PHYSICS***
2. ***ON THE REFERENCE SYSTEM CONCEPT IN GENERAL RELATIVITY AND BEYOND***
3. ***SET PARTITION BASED DEFINITION OF ELEMENTARY EVENTS AS THE PROBABILITY AND INFORMATION THEORIES COMMON BASIS***

Tel-Aviv University Press ***Tel-Aviv 2007***

PHYSICS WITHOUT SPACE AND TIME

A. Ya. Temkin

Department of Interdisciplinary Studies

Faculty of Engineering

Tel-Aviv University

Ramat-Aviv

Tel-Aviv 69978

Israel

E-mail: **temkin@eng.tau.ac.il**

Fax: +972-3-6410189

Tel: +972-3-6408176

Tel-Aviv University Press ***Tel-Aviv 2007***

Printed by Tel-Aviv University Press Tel-Aviv 2007

This book contains three articles. The first of them dedicated to the consideration of the microworld description without the supposition on the existence of space and time. An alternative approach is formulated and developed. The second one is dedicated to the consideration of the macroworld also without supposition on the space and time existence. The conditions that the space and time can be defined are found. In the third article the information theory is formulated in the framework of the set theory. It is used as the mathematical background of some considerations made in the first one.

ARTICLE 1

HIERARCHY BASED MICROWORLD SCALES' CLASSIFICATION AND MICROWORLD PHYSICS

A. Ya. Temkin

Department of Interdisciplinary Studies

Faculty of Engineering

Tel-Aviv University

Ramat-Aviv

Tel-Aviv 69978

Israel

E-mail: temkin@eng.tau.ac.il

Fax: +972-3-6410189

Tel: +972-3-6408176

ABSTRACT

The penetration to smaller and smaller scales of the physical World by high energy elementary particles' collisions demands the definition and study of scales as the general concept. In the present paper scales of the microworld are defined on the grounds of the hierarchy that can be set up within a well ordered finite or infinite countable set of well ordered sets. Indirect measurements in pure mathematical as well as physical meaning are considered as the mean to obtain information on the occurring within each scale. The general concept of the physical laws within a certain scale is defined in the framework of the set theory. The hypothesis is proposed that quarks are physical objects existing not within the same scale as other elementary particles do, but within the following ("smaller", exactly of lower hierarchy) scale. Maybe, in particular, this is the cause of difficulties of the free quark detection. Some other consequences of the abovementioned hypothesis are discussed.

Possible connection of well known problems of the quantum field theory (divergences, renormalized electron charge zero) with the microworld scaling is discussed.

Limitations of the possibility to obtain information on the going on within sets of low hierarchy (in particular, "small" scales of the microworld, in physics) and to transfer it toward the set of the highest hierarchy (in particular, to the macroscopic observer in physics) are found as a consequences of the multi-step character of measurements. The existence of the "smallest" scale (*i. e.*, scale of the lowest hierarchy) is considered.

Keywords: microworld scales, set theory, information value and hierarchy, well-ordered set hierarchy, sub-quantum physics, indirect measurements, quarks

PACS: 02.10.Ab, 03.65.Ta

1. INTRODUCTION

The contemporary high-energy physics that uses particles of higher and higher energies penetrates to smaller and smaller scales of the microworld. In view of this it would be desirable to define the general concept of THE MICROWORLD SCALE and to find how the information on the occurring within each scale can be forwarded to the MACROSCOPIC OBSERVER possessing the exclusive ability to obtain, to process and to interpret it. It is necessary also to elucidate whether the total number of microworld scales is PRINCIPALLY limited or not, *i. e.*, whether this penetration to the depth of the microworld by the high-energy physics cannot/can be continued up to its infinitesimally small scales.

Classical and quantum physics correspond to macroscopic and microscopic scales of the world. The second of them, quantum physics' scale, corresponds to space region of characteristic linear dimension $\stackrel{<}{\sim} 10^{-8}\,cm$, *i. e.*, of atomic dimension or less. Whether "less" means "up to infinitesimally small linear dimension", or there exists its lowest limit $l_{\min,1}$ such that from linear dimension $\stackrel{<}{\sim} l_{\min,1}$ begins the new scale with its own physical laws that stretches up to $l_{\min,2}$, and so on? It is, at least, not impossible. To consider this and other problems of the microworld scaling it is first of all to define the concept "scale". The difficulty is that we cannot be sure that the space-time continuum and, therefore, dimension exists always. By this reason it would be desirable to define the concept SCALE without use the notions of upper and lowest limits of its dimension.

For this purpose in the present work we use as the starting point not dimension limits, but the hierarchy that is set up among well ordered sets forming a well ordered set. The microworld scales are considered as a particular case of this set theoretical consideration. They defined and classified with respect to their hierarchy, which is established, in this case, on the grounds of properties and characters of physical phenomena, and of the information on them. This approach to the microworld scales' classification does not demand the use of the notion DIMENSION of each scale (for example, by assignment scale's upper and lowest limits) for the scale definition. Thus, it delivers us from the necessity to use space and time (or space-time) as the area for physical event addresses' representation throughout all scales, as it is being done in classical and quantum mechanics, *i. e.*, within the macroscopic and atomic scales.

NOTE:

1) In the present paper we, for brevity, call events not only events themselves (*e. g.*, collisions), but also objects (*e. g.*, elementary particles).
2) When we consider the classical and quantum scales, and only in these cases, we call addresses of physical events

their co-ordinates in space-time, spin, isotopic spin, parity *etc.*. The general representation of event addresses is introduced.

………………………………………………

In Sec. 2 the concept of INFORMATION TYPES is defined and used to introduce the concept of set types for sets able to treat the information. It is proposed to define type of such a set according that what functions of the information treatment it executes. However, other possibilities of type definition are considered preliminary (detailed consideration is in Secs. 3 and 4).

Sec. 3 is dedicated to the general consideration of the information and set hierarchy on the grounds of the information value [1,2] as well as on the grounds of the Russell's theory of types [3]. The concept of the information value is considered and developed.

In Sec. 4 the hierarchy among sets $S^{(l)}$ forming the well-ordered set S is considered.

In Sec. 5 the set theoretical approach to physical events and their addresses (defined as elements of two corresponding sets) representation is formulated and developed.

In Sec. 6 one defines and considers mapping with the feedback.

In Sec.7 one continues the consideration of the Sec. 4 on the setting up hierarchy inside a well-ordered set of well-ordered sets and defines the concept of microworld scales on these grounds.

In Sec. 8 indirect measurements within different scales are considered. It is indicated that possibly quarks exist not within the same scales as hadrons and leptons, but within a "smaller" scale, or exactly, within a scale of lower hierarchy.

In Sec. 9 the set theoretical concept of physical laws within a certain microworld scale is introduced and considered.

In Sec. 10 one considers consequences of the fact that a measurement within a microworld scale made by a macroscopic observer is a sequence of indirect measurements within previous larger scales. It is indicated that it leads to a fundamental limitation of our knowledge on the microworld.

In Sec. 11 the transition from quantum to the first sub-quantum scale is considered as an example of the proposed theory application.

2. INFORMATION TYPES AND SET TYPES (PRELIMINARY)

The purpose of the Sections 2 - 4 is to consider different ways to set up the hierarchy of sets able to treat the information: A) on the grounds of the theory of information by definition of the concepts of information types and the information hierarchy, and B) on the grounds of the Russell's theory of types [3]. It will be shown that in the case when the information and its value [1,2] are expressed in a language and built according logical rules the both approaches lead to the same result and that Russell's types can be expressed in terms of the value of the information obtained as result of the information treatment.

As the starting point we consider here a well-ordered finite or countable set S of well-ordered sets supposing that some of these sets are able to treat the information. Now introduce and set up the hierarchy among these sets *based on their properties with respect to the information treatment* and thereupon reorder set S with respect to hierarchy of sets forming it. The information treatment includes the following functions: 1) the receipt of information (from), 2) the sending of information (to), 3) the information processing, 4) the information interpretation, and 5) the information storage in memories. Define that the hierarchy of such a set is determined by which of these functions it executes. However, it is to take into account that this criterion could be not sufficient one to determine the hierarchy because there is the possibility that more than one set have the same type of the information treatment and, therefore, the same hierarchy according to this criterion that prevents to set up the order among them, when one reorders the set S with respect to hierarchies of sets forming it. It seems to be like the quantum state degeneration. This analogy suggests an idea to search for some supplementary criterions that may allow one to attribute to these sets different hierarchies (to break this "degeneration"). Then they can be ordered among themselves also with respect to their hierarchies. In the case when no such supplementary criterion exists, they can be ordered among themselves on the grounds of reasons other than hierarchy or arbitrarily.

We shall accept that the lowest hierarchy is attributed to sets that, in general, do not treat the information while the highest one is attributed to each set executed all five functions. If there is only one set of the highest hierarchy, then order (or reorder) set S with respect hierarchies of sets forming it so that the set possessing the highest hierarchy will be the first (we shall attribute to it number 0), while the hierarchy of other sets decreases when the number augments.

Let us reinterpret the described approach in terms of the INFORMATION VALUE [1,2]. The information obtained by the interpretation of the processed information has the largest value because it is able to induce the most serious changes to the understanding of the obtained information meaning and, on these grounds, to invent its new applications creating material changes. For example, if to speak on physics, such an interpretation may mean the replace of existing physical laws to the new ones, which leads to a serious, maybe drastic change of our understanding of the going on in the World (*cf.* the replace of the classical mechanics to the relativity and quantum mechanics) and creation new applications, *e. g.*, nuclear energy, quantum computing *etc*... Following this way we define the hierarchy in accordance with the order of values of the information. In Sec. 3 the other approach based on Russell's theory of types [3] will be represented.

The information on going on within a certain set of those forming the set S should be transmitted step-by-step to the set of the highest hierarchy to be processed and interpreted. The process of the information extraction (on occurring within a set) we shall call MEASUREMENT considering it as a general mathematical notion. In its physical applications we, for brevity, shall use this term also for observation. For example, in the microworld one uses measurements while in the macroworld (*e. g.*, in the astronomy and astrophysics) mainly observations are used.

3. TYPES, INFORMATION TREATMENT AND SETS HIERARCHY (general approach)

Consider now the setting up hierarchy within the information with more details. The information is characterized by its amount (see, for example, [4,5]) and value [1,2]. Let us consider the following multi-step process:

primary information I_1 *- information creation* I_2 *- information creation* I_3 *- information creation* I_4 *- information creation* I_5 *- etc.*

The information value can be determined as the amount of the information I_{n+1} created in all n these steps divided into amount of the primary information I_1. However, it cannot be the only characteristic of the information value because it does not take into account properties of the information content. If properties of the created information content are taken into account, then the value of the primary information should be represented by a set of such characteristics + the number defined above. Let us try to represent these characteristics in general form. Denote each of them ξ_ν, where ν labels a certain characteristic. Then one has $\Xi = (\forall \nu)\{\xi_\nu\}$. Denote those of the primary information $\Xi_1 = (\forall \nu_1)\{\xi_{1,\nu_1}\}$, of the secondary information $\Xi_1 = (\forall \nu_2)\{\xi_{2,\nu_2}\}$,..., of the n^{ts} step $\Xi_{n+1} = (\forall \nu_{n+1})\{\xi_{n+1,\nu_{n+1}}\}$, ... *etc*.. The index at ν is necessary because for information obtained at each step the set of properties could be different from that for obtained at other steps.

Let us introduce the norm of a property $\|\xi_{l,\nu_l}\|$, which is a number. How the norm is defined depends on each concrete case, so we do not consider this problem for the general case. The complete representation of the primary information value is $\prod_{\forall l}^{\otimes} \Xi_l$, where $\prod^{\otimes}$ denotes the Cartesian product of sets. Using the norm of the property one can introduce a quantity characterizing the primary information value. It is $J_V = \left(\left\|\prod_{\forall l}^{\otimes} \Xi_l\right\| + I_{n+1}\right) I_1^{-1}$ that will be called INFORMATION VALUE. However, it must be kept in mind that, really, it is only a partial characteristic of the information value.

If only $n' < n$ steps are realized, while principally n steps are possible, one can define the concept of the POTENTIAL INFORMATION VALUE that is determined for all n steps, no matter how many of them are realized. One can define also the concept of the CONSTRAINED INFORMATION VALUE when constrains prohibit the realization of a part of steps.

Note that the use of the norm $\left\|\xi_{l,v_l}\right\|$ is not the only way to compare information values of primary information in different cases. It is possible to refuse from the use numbers for this purpose and instead of it to assign to each ξ_v quality Q (which is not obligatory a number) such that between any two Q_v and $Q_{v'}$ the relation of order, *e. g.*, $Q_{v'} \prec Q_v$, exists. One can interpret this relation so that the quality Q_v is higher than quality $Q_{v'}$. Whether this approach can be used instead of the use of $J_V = \left(\left\|\prod_{\forall l}^{\otimes} \Xi_l\right\| + I_{n+1}\right) I_1^{-1}$ to express the information value? It is possible, if relation of order like $Q_{v'} \prec Q_v$ can be set up between any two $\prod_{\forall l}^{\otimes} \Xi_l$ (for both cases of the initial information). But it seems questionable because different ξ_v with different Q_v enter to this Cartesian product in complicated combinations.

DEFINE that the type of the information is determined by its value, potential value or constrained value, accordingly to the considered problem. DEFINE that the information hierarchy is set up according types of the information.

Consider the case when the information and its value are expressed in a language and built according the rules of logic. Then one can define the types of information using the Bertran Russell's theory of types [3] as the starting point. We read in the abovementioned article of Bertran Russell: "A *type* is defined as the range of significance of a propositional function, *i. e.*, as the collection of arguments for which the said function has value." "Thus whatever contains an apparent variable must be of different type from the possible values of that variable; we will say that it is of a *higher* type." In our case, for example, the processed information can be considered as the set of values of apparent variables that are, in their turn, the result of the information interpretation.

Consider it in more detail. Let $V = \left(\forall r\left[r \in \mathbb{N}; r \in \left[r_0, r_{\max} > r_0\right]\right]\right)\{V_r\}$ are apparent variables [3,6] of the processed, but not yet interpreted information. Here and in the following text $\mathbb{N}$ denotes the set of all natural numbers. The interpretation consists in 1) the setting up connections between V_r with different values of r, 2) the setting up rules how values of V_r can be calculated and 3) the setting up connections with variables characterizing external factors influencing the considered system. We shall call a THEORY the result of the interpretation. Note that thereupon this new theory may be, in its turn, used for the interpretation of the new information, which itself was obtained on the grounds of a preceding theory, which also was obtained on the grounds …and so on. $V = \left(\forall r\left[r \in \mathbb{N}; r \in \left[r_0, r_{\max} > r_0\right]\right]\right)\{V_r\}$ can be obtained now as values of apparent variables $U = \left(\forall s\left[s \in \mathbb{N}; s \in \left[s_0, s_{\max} > s_0\right]\right]\right)\{U_s\}$ of interpreted information, *i. e.*, from the theory. Therefore, according Russell the interpreted information (expressed in terms of these apparent variables) is of higher type in comparison with the processed, but not yet interpreted information. The same can be said on the not yet processed and processed information: the second is of higher type than the first one. In general, elements of a set or their configurations can be considered as values of apparent variables of the information about this set. The information has, therefore, higher type than the set itself.

Let us return to the consideration of the approach when the type of the information is defined also on the grounds of its value [1,2]. It is more general than the written above Russel's approach because it is not limited with the condition that the information and its value must be expressed in a language. However, in the case when the information and it value are expressed in a language, this definition seems to be equivalent to the one based on Russell's theory of types and, in particular, leads to the same result that the interpreted information is of the highest type. Note that different levels of this interpretation may exist so that the information obtained by these kinds of interpretation could have different types. With the purpose to avoid such an uncertainty at the consideration of sets treating the information one defines the type of a set treating the information as the highest of the types of the information obtained by this treatment.

Let us now consider a well-ordered set containing sub-sets able to receive (also by performing measurements), to send, to process, to interpret and to store (in memories) the information (*cf*. Sec. 2.). We shall call such a subset OBSERVER, iff it is able to execute ALL these functions including measurements. We do not suppose that all considered subsets are observers, in other words, not each of them executes all the abovementioned functions.

Our purpose is to set up the hierarchy among such sets based on the information hierarchy defined above.

We define that the hierarchy order of two sets treating the information corresponds to their types order. The generalization to any finite or infinite countable well-ordered set of sets is evident. WE DEFINE that the hierarchy of a set containing a subset able to treat the information would be equal to the hierarchy of this subset. If this set contains a set of such subsets treating the information having different types, we shall define that the hierarchy of the considered set is determined by the highest of these types. Thus, the type attributed by definition to subsets able to receive, process, send and store the information (they are not *observers*) would be lower than that attributed to the observer which is able also to interpret the information.

The information value probably does not affect the original information entropy, but it may create the negative or positive entropy production. For example, at the explosive crystallization of an amorphous body by laser light the information carried by this light initiates the transformation of a disordered amorphous body to the ordered crystal. The value of the original information corresponds to the big (by the absolute value) negative information entropy production. At the initiation of the explosion of an explosive by electric signal the value of the original information corresponds to the big positive entropy production because an ordered structure is turned into a disordered one. This means, the value of the original information corresponds to the absolute value of the entropy production. DEFINE the specific absolute value of the entropy production as its absolute value divided to the original information amount. It can be an important characteristic of the information action. This connection between the information value and such a thermodynamic quantity as the entropy production suggests the idea that one can formulate the problem of the hierarchy also in terms of the thermodynamics.

Not all processes initiated by the original information are obligatory occurred at one step (*cf.* written in the beginning of this section). The "first creature" may initiate new processes creating "the second creature" *etc*. If only the "first creature" is taken into account or the following steps be prohibited by any conditions, the value of the original information would be less than in the case when the "second and following creatures" will be realized and taken into account. Therefore, as it was mentioned above, the value of the information only is not enough to characterize the ability of the original information, and it is so we introduced above the concept of potential value of information based on taking in the account those effects that the considered information potentially is able to produce (maybe in some steps), but not yet produced. For example, the information obtained by the observer can possess big value and potential value because its interpretation (possibly, even the creation of new physical laws) may produce remarkable effects.

Now we can define the concepts "VALUE OF INFORMATION" and "POTENTIAL VALUE OF INFORMATION" in terms of Russell's theory of types [3,6]. We shall accept that the type of the information is determined as the type of its expression in terms of mathematical logic notions [3].

DEFINITION 1. The value of the considered information (primary information) is the highest Russell's type of the information created by the activated primary information in maximum number of executed steps.

DEFINITION 2. The potential value of the considered information (primary information) is the highest Russell's type of the information that could be created by the activated primary information in maximum number of principally existing steps.

The activated information is the information that produces new information, physical, chemical, biological, industrial, social and other effects. Example: the prominent letter of Albert Einstein to Franklin D. Roosevelt, President of the USA, where Einstein proposed to begin researches aimed to create the nuclear weapon. It contained information that could be called frozen or potential one up to the moment when the President read it and decided to begin these researches. Then it became to be the active information. If the President did not read this letter or rejected the Einstein's proposal, the value of the information contained in this letter would be equal to the zero and only its potential value should be enormous.

A subset able to execute the information treatment must contain a subset formed of elements that, in their turn, are sets containing more than one element. Then in the considered subset different distributions of elements (for example, with respect to numbers of elements including to each element of this subset) can exist and, therefore, the probability and information can be defined.

The ability of such a subset to receive, send, process, store and interpret the information depends on the set structure. If there are a number of such subsets, their relative hierarchy is defined as their relative ability of the information treatment. The rough classification can be as follows: the lowest hierarchy (=0) have those which cannot receive, cannot send and cannot process the information; the hierarchy =1 is attributed to subsets which are able to receive, to send, to store, but cannot process the information; the hierarchy =2 is attributed to subsets which are able to receive, to send, to store and to process the information; the hierarchy =3 is attributed to subsets which are able to receive, to send, to process, to store and to interpret the information.

Inside each type could be different sub-types with different hierarchy among them. For example, inside type (3) could be different levels of the information interpretation. The highest hierarchy among them is attributed to the subset which extracts from the received and processed information general, in particular, physical laws.

Note that the active information can create new information, but it can create phenomena of different nature, *e. g.*, physical, chemical, biological, geophysical, emotions of human beings and animals, thoughts of human being expressed or not in a language, logical or not *etc.*. It must be taken into account at the consideration of the information value. The mathematical logic, in general, and Russell's theory of types, in particular, can be applied to the information value consideration only when the processes can be expressed in a language (or languages) according logical rules at all stages. Note that it must not negate without a serious consideration the possibility of existing of the information which is not expressed in a language, but despite it is built according logical rules. Of course, if it exists, these logical rules must be a *generalization* of those of the existing logic, for example, those connecting certain sets, but not propositions *etc.*.

4. HIERARCHY AMONG SETS $S^{(l)}$

Let there is a well-ordered not empty final or countable set S of not intersected not empty well-ordered sets $S^{(l)}\left[l\in\mathbb{N}, l\in\left[0,l_{\max}\vee\infty\right],\left(l'\neq l\right)\Rightarrow S^{(l)}\cap S^{(l')}=\emptyset\vee\left(l'=l\right)\Rightarrow S^{(l)}\right]$, where $\{\mathbb{N}\}$ is the set of all natural numbers. The written above allows one to set up the hierarchy between all well-ordered sets forming the well-ordered set S.

One can order the set $S = \left\{ S^{(l)} \right\}$ with respect hierarchies of sets $S^{(l)}$. Let us set up the hierarchy within the set S so that the set $S^{(0)}$ possesses the highest hierarchy and the hierarchy of sets $S^{(l)}$ decreases with the increase of l: $l' > l \Rightarrow \tilde{h}\left(S^{(l')}\right) < \tilde{h}\left(S^{(l)}\right)$, where $\tilde{h}\left(S^{(l)}\right)$ denotes hierarchy of the set $S^{(l)}$. Let us postulate that only the set $S^{(0)}$ is allowed to receive, to send, to process, to interpret and to store the information on all other sets. This means, we consider here the case when the observer(s) exists at only one scale, namely that having the highest hierarchy. At applications to microworld physics (see below) this means that there is only macroscopic observer(s). The corresponding situation arising at the macroworld study merits a special consideration; the consideration of the microworld (see below) cannot be automatically transferred to the macroworld. The information on events is provided by a certain mathematical procedure that we shall call *measurement* or *observation*. The use of these two terms is dictated by some applications of this mathematical theory: for example, in the microworld usually one uses measurements while in the astronomy and astrophysics observations are usually used to provide and obtain information. However, for brevity, we shall use the term measurement in both cases, keeping in mind that it includes also observation. The set $S^{(0)}$ contains a subset $U_{S^{(0)}} \subset S^{(0)}$ possessing the following properties: it is able to receive (also by measurements), to process, to interpret, and to store the information. We shall call this subset $U_{S^{(0)}}$ OBSERVER. The abovementioned interpretation is done on the grounds of certain laws (mathematical, physical *etc.*) that should be expressed in a convenient mathematical form and included to the subset $U_{S^{(0)}}$. If a certain deviation $\delta S^{(l')}\left(l' \in \{l\}, l' > 0\right)$ has occurred with the set $S^{(l')}$ itself, the information on it must be forwarded step-by-step to the subset $U_{S^{(0)}}$ to be processed, interpreted and stored.

5.SETS OF EVENTS AND THEIR ADDRESSES

Let there are two not empty well-ordered sets (see, for example, [7]) $A = \{a\}$ of elements a that we shall call EVENTS and $H = \{h\}$ of elements h that we shall call the SET OF ADDRESSES of elements a.

Consider a not empty subset $Y \subseteq A$. Set up the homomorphism keeping the order between Y and $H_Y \subseteq H$, where H_Y be homomorphic map of $Y \subseteq A$. We shall call H_Y the address of the subset Y of events. In particular, if $Y = a$, then $H_Y = H_a$ will be the address of the single event a. We use the homomorphic, but not the isomorphic mapping, taking into account that more than one event may have the same address.

Consider the following case

$$\left(\exists H, \exists \tilde{H}, (\forall i)\left[\exists H^{(i)}\right] \middle| (\forall i,\ \forall n)\left[n \wedge i \in \{\mathbb{N}\}\right]; \forall i \in [1, n]\right)\left[H \supseteq \tilde{H} = \prod_{i=1}^{n}{}^{\otimes} H^{(i)}\right], \qquad (1)$$

where $H, \tilde{H}$ and all $H^{(i)}$ are well- ordered sets.

Let us consider the following particular cases:

1. If $\tilde{H} = H$, the address of a subset Y of events can be represented as

$$H_Y = \tilde{H}_Y = \prod_{i=1}^{n} {}^{\otimes} H_Y^{(i)} \ , \qquad (2)$$

2. while if $\tilde{H} \subset H$, the following three cases are possible:

$$\text{a)} H_Y \subset \tilde{H} \text{, b) } H_Y \subset N = H \setminus \tilde{H} \text{, c) } H_Y \bigcap \tilde{H} \neq \emptyset \wedge H_Y \bigcap N \neq \emptyset \qquad (3)$$

In the case (a) Eqn. (2) is valid. However, the cases (b) and (c) demand special considerations.

We see two options. OPTION I: An expansion of the set H so that the obtained new set $H^{(*)} \supset H$ would be represented in the form

$$\big((\forall i, \forall m)\big[m \wedge n \wedge i \in \{\mathbb{N}\}\big]\big)\big[i \in [1, m > n]\big]\left[\exists H^{(i)}, H^{(*)} = \prod_{i=1}^{m} {}^{\otimes} H^{(i)}\right] \ , \qquad (4)$$

therefore, addresses will be considered as subsets of $H^{(*)}$, *i. e.*, as $H_Y^{(*)} \subseteq H^{(*)}$. OPTION II: That in two of abovementioned cases the address of $Y \subseteq A$ cannot be represented by Eqn. (2) and should be remained in the form H_Y . The property of the ordering of the set H allows one to write:

$$\left(\exists H_1 \subset H, \nexists H_2 \subset H \vee \left((\exists H_2 \subset H) \left[\mu(H_2) = \emptyset \right] \right) \right) \left[H_1 \subset H, H_1 \prec H_Y, H_1 \prec H_2 \prec H_Y \right] \tag{5}$$

and

$$\left(\exists H_3 \subset H, \nexists H_4 \subset H \vee \left((\exists H_4 \subset H) \left[\mu(H_4) = \emptyset \right] \right) \right) \left[H_3 \subset H, H_3 \succ H_Y, H_3 \succ H_4 \succ H_Y \right], \tag{6}$$

where $\mu(H)$ denotes the measure of the set *H*.

6.MAPPING WITH FEEDBACK

We wrote above on a mapping (homomorphism) of set A subsets to subsets of the set H. For the application of this formalism to a real system, e. g., physical system, a certain real procedure is necessary 1) to establish the demanded correspondences and 2) to make the result, *i. e.*, the address, known. The latter problem will be considered in this Section.

Let us consider a simple example. The set of all apartments in a building is mapped to a subset of the set of all natural numbers so that each apartment has its number. But usually this number is put on the apartment door, in other words, the apartment is labeled by its number. Then the number of the apartment, *i. e.*, the result of the abovementioned mapping, becomes known.

On the analogy of this example let us consider now how the address H_Y can be found out, in other words, by means of what the address of the subset of events $Y \subseteq A$ will become known. The ordinary procedure of the direct mapping of $Y \subseteq A$ to H_Y does not undertakes this task. At the same time several applications of the mapping, for example, measuring in physics, include this task and so a corresponding mathematical procedure is needed to accomplish it. Indeed, a measurement of space - time co-ordinates of an event would be useless, if the observer cannot obtain its result, in other words, if there is not feedback between the sending a (light) signal for a measurement and obtained results of it. That is why we want to label event by its address. It allows the observer at one go to get to know the event and its address. Of course, label may be changed from measurement to measurement. We shall call such an event (an element of the set *A*) with the label a LABELED EVENT (a labeled element of the set A). Probably, this labeling would be important also at the study of the macro-World (the Universe and its regions), but this is not a subject of the present paper.

Now introduce the necessary feedback procedure. Let each element $a \in A$ and each element $h \in H$ are themselves sets of two or more elements (this is the necessary condition that the feedback is possible):

$$a = \left(\forall n \in \mathbb{N}, \forall j \in [1,n]\right)\begin{Bmatrix} \tilde{a} \\ b_{a,j} \end{Bmatrix}, \qquad (7)$$

$$h = \left(\forall m \in \mathbb{N}, \forall l \in [1,m]\right)\begin{Bmatrix} \breve{h} \\ q_{h,l} \end{Bmatrix}, \qquad (8)$$

where $\tilde{a}$ and $\breve{h}$ denote the event and the address themselves, while $b_{a,j}$ and $q_{h,l}$ are intermediate objects using for the following mapping procedure. In the beginning for the sake of simplicity we shall consider the case when $n=1$ and $m=1$, *i. e.*, when there is only one $b_{a,1}=b_a$ and only one $q_{h,1}=q_h$. Map now $a\in A$ to an element $h\in H$. Then we shall obtain the pair $b_a q_h$. Now one can map this pair to the corresponding element a of the set A (according the index a of b_a). This means, we shall return element b_a to their place, but together with corresponding label q_h. Thus, now an element a of the set A is labeled by its address q_h. The presence of b_a in the pair $b_a q_h$ establishes the isomorphism between $\{b_a q_h\}$ and $\{a\}$ in the situation when the mapping $\{a\}$ to $\{h\}$ is a homomorphism, and, therefore, the inverse mapping made directly as $\{h\}\rightarrow\{a\}$ would be not single-valued.

The considered situation is like the one arising in the quantum mechanics in the case of quantum state degeneration. Then a new factor breaking the symmetry, for example, magnetic field, can remove the degeneration splitting the degenerated energy level into a number of closed different levels. In our case this role plays $b_{a,j}$ and $q_{h,l}$ that turn the homomorphism into isomorphism (of course, not for $\tilde{a}$ and $\breve{h}$, but for a and h defined by Eqns. (7) and (8), correspondingly).

The procedure described above can be done with each element $a \in Y \subseteq A$, and we shall obtain

$$Y = \left\{ a = \left(\forall j \in [1,n]\right) \left\{ \begin{matrix} \tilde{a} \\ b_a q_h \end{matrix} \right\} \right\} \tag{9}$$

Thus, the described procedure establishes the necessary feedback labeling the subset Y by the addresses of its elements. We shall call this mathematical procedure MEASUREMENT, though it could not be without fail the measurement in the physical meaning.

7.SETS' HIERARCHY

Let there is a countable (finite or infinite) well ordered set $S = \left\{ S^{(q)} \right\}$ of finite or infinite (countable or continuum) well ordered in pairs non-intersected sets $S^{(0)}, S^{(1)}, S^{(2)}, \ldots, S^{(q)}, \ldots$,

$\left(\forall (q, q')\right)\left[S^{(q)} \cap S^{(q')} \right] = \emptyset$, where $q \in [0, q_{\max} \vee \infty]$, $q_{\max} \in \mathbb{N}, q_{\max} \geq 1$.

In the pure mathematical framework the order within the set $S = \left\{ S^{(q)} \right\}$ can be set up, for example, according the order of types of $\forall S^{(q)}$, or by another way. However, for the applications to physical problems the order in the set $S = \left\{ S^{(q)} \right\}$ is to be set up on the grounds of physical reasons, including inferences arising from experimental results. If this order be set up on the grounds of mathematical reasons only, possibly the obtained mathematical theory would not be fit for the considered physical problems treatment.

Let the set S is reordered so that the hierarchy among the sets $S^{(q)}$ corresponds to their order such that the highest hierarchy is attributed to the set $S^{(0)}$. Define 1) that the ability to initiate indirect measurements in all $S^{(q)}$ is attributed only to the set $S^{(0)}$, and 2) that the information obtained from all such measurements can be extracted only from $H^{(0)}$, but not from any $H^{(q>0)}$.

This consideration allows one to define the notion SCALE OF THE MICROWORLD as follows: let sets $S^{(q)}$ with different q are ordered with respect to their hierarchy, then we shall call the set $S^{(q)}$ scale number q (q=0,1,2,3,...) of the microworld when 1)namely the microworld is studied and 2)number q=0 corresponds to the macroscopic scale. The observer is always macroscopic and makes measurements within the scale $S^{(0)}$.

8. MEASUREMENTS

In physics the information on a physical object is obtained by measurements. We shall keep this term also for the case of mathematical objects. Consider this problem in detail.

Measurement in physics can be performed DIRECTLY by a macroscopic observer (human being or automaton) using measuring instruments. The task of the observer includes the interpretation of measurement results. Because the observer is always macroscopic, his ability to make DIRECT MEASUREMENTS is limited with the atomic and nuclear scale. It is questionable whether they can be used within smaller scales, if they exist. Maybe it is possible because the development of particle accelerators to the direction of higher-and-higher energies, but, as it can be concluded from our consideration, the effective use of such measurement equipments is possible only in combination of direct and indirect measurements.

Mandelstam [8] introduced the concept of INDIRECT MEASUREMENT (see also [9,10]) to quantum mechanics. We shall try to use indirect measurements to penetrate step-by-step into smaller and smaller scales of the microworld, precisely speaking, into scales of lower and lower hierarchy. The general theory of indirect measurements is developed in Sec. 3.4* of the book [9]. However, this theory supposes that all systems participating in the indirect measurement process, in exception of the macroscopic observer, are quantum ones. In the present work we shall consider the penetration to the sub-quantum scale and beyond. There is no reason *à priori* to suppose systems within such scales to be also quantum. Therefore, the indirect measurement theory of Braginsky & Khalili [9] cannot be applied to the cases that are the subject of the present work, and so we develop here another theory of indirect measurements.

The application of Mandelstam's idea for this purpose can be represented by the following example. The macroscopic observer measures co-ordinate of a particle Υ within the atomic scale. Thereupon this particle collides with an object belonging to the nuclear scale, and the macroscopic observer measures the change of Υ's co-ordinate. The comparison of these two measurements provides information on the nuclear scale object.

Consider this problem firstly as the pure mathematical one. Let results of two subsequent direct measurements are different so that their difference exceeds the statistical error of measurement. This means that a subset $Y \subseteq A$ of events has two "addresses" $H_{Y,1}$ and $\mathrm{H}_{Y,2}$ such that $H_{Y,1} \prec H_{Y,2}$, $H_{Y,1} \cap H_{Y,2} = \emptyset$. This fact may have different interpretations. Possibly, something was changed by itself in the system $Y \subseteq A$ between these two measurements, if they were not made simultaneously. If it is proved that such a possibility does not exist in this case, then one of the remained possibilities is that there is one more subset $Y' \subseteq A$ of events, the "address" of which we shall denote $H_{Y'}$, that influences the address of the subset Y so that $\exists F\left(H_{Y,1}, H_{Y,2}\right) = H_{Y'}$. This equality means that this pair of addresses $H_{Y,1}$ and $H_{Y,2}$ contains the information sufficient to define the address $H_{Y'}$ of a subset Y' of events. The function $F\left(H_{Y,1}, H_{Y,2}\right)$ means a homomorphism of the pair $H_{Y,1} \subset H$ and $H_{Y,2} \subset H$ to a subset $H_{Y'} \subset H$. However, it is possible that the abovementioned pair of addresses does not contain the sufficient information on the address $H_{Y'}$, which means that $\nexists F\left(H_{Y,1}, H_{Y,2}\right) = H_{Y'}$. Consider now the general case when the number of measurements is not limited with two. Define the function

$$F_m\left(H_{Y,1} \prec H_{Y,2} \prec \cdots \prec H_{Y,m}; (\forall i)\left[H_{Y,i+1} \bigcap_{i=1}^{m} H_{Y,i} = \emptyset\right]\right), \qquad (10)$$

where $m \geq 1$. Denote by α type of subset of events. Then $H_{Y'_\alpha}$ is defined as the address of a subset Y'_α of events

$$\Leftrightarrow \left\{\left(\exists\left\{H_{Y,i}\right\}^{(\alpha)}\right)\left[\exists\left(\lim_{m_\alpha \to \infty} F^{(\alpha)}_{m_\alpha} = H_{Y'_\alpha} \neq \emptyset\right) \wedge \nexists\left(\lim_{m_\alpha \to \infty} F^{(\alpha)}_{m_\alpha} = H_c \neq H_{Y'_\alpha}\right)\right]\right\}, \qquad (11)$$

where $\exists\left\{H_{Y,i}\right\}^{(\alpha)}$ means, "Exists set (sequence) $\{\ \}^{(\alpha)}$ ".

It is evident from Eqn. (11) and Y'_α definition that 1) $H_{Y'_\alpha}$ cannot be detected directly, but only indirectly by its influence upon the subset Y of events, and 2) Y'_α itself cannot be detected by this way, but only its address.

Now let $Y \overset{def}{=} Y^{(q)} \subseteq S^{(q)}$. Then $Y' \overset{def}{=} Y^{(q+1)} \subseteq S^{(q+1)}$. If in all considerations and formulas above one replaces $Y^{(q)}$ to $Y^{(q+1)}$, and $Y^{(q+2)}$ to $Y^{(q+3)}$, one obtains the address of a subset $Y^{(q+2)}$ of events. The following step will be evidently the replace $Y^{(q+1)}$ to $Y^{(q+2)}$ and $Y^{(q+2)}$ to $Y^{(q+3)}$. Thus, the address of a new subset of events $Y^{(q+3)}$ will be obtained. This procedure can be continued.

The purpose of this Section is to find a way to detect events within a scale $(q+1)$ by use of indirect measurements made within the scale q. It will be one step of the multi-step indirect measurement made by an observer who always is macroscopic, *i. e.*, being and making measurements within the scale $q = 0$. Such a measurement consists of subsequent applications of described one step indirect measurements (that we denote $q' \to q'+1$) beginning from that $0 \to 1$ up to the desired measurement $q \to q+1$. Now rewrite Eqns. (10) and (11) for a measurement $q \to q+1$:

$$F_m^{(q)}\left(H_{Y^{(q)},1} \prec H_{Y^{(q)},2} \prec \cdots \prec H_{Y^{(q)},m}; (\forall i)\left[H_{Y^{(q)},i+1} \bigcap_{i=1}^{m} H_{Y^{(q)},i} = \emptyset\right]\right) \qquad (12)$$

Then $H^{(q+1)}_{Y^{(q+1)},\alpha}$ is defined as the address of a subset $Y^{(q+1)}_{Y^{(q+1)},\alpha}$ of events within the scale $S^{(q+1)}$

$$\Leftrightarrow \{\left(\exists\left\{H^{(q|\alpha)}_{Y^{(q)},i}\right\}^{(q|\alpha)}\right)$$
$$\left[\exists\left(\lim_{m_{q|\alpha}\to\infty} F^{(q|\alpha)}_{m_{q|\alpha}} = H^{(q+1)}_{Y^{(q+1)},\alpha} \neq \emptyset\right) \wedge \nexists\left(\lim_{m_{q|\alpha}\to\infty} F^{(q|\alpha)}_{m_{q|\alpha}} = H_c \neq H^{(q+1)}_{Y^{(q+1)},\alpha}\right)\right]\} \qquad (13)$$

It is important to find how many and what namely independent data are included to each of these addresses. The answer to these questions establishes the physical grounds within each scale and by this indicates what is to be determined by experiments to study physical events within a certain scale.

Let the Eqn. (13) is not satisfied. It is possible that it is induced by events taking place within a set $S^{(q+2)}$, the set (scale) which's existence was not yet known. If the Eqn. (13) be satisfied by the substitution scale of number $(q+1)$ as the smallest hierarchy scale to the one of number $(q+2)$, then this hypothesis will be confirmed. In this case the measurement is to be done firstly within the scale $S^{(q+1)}$ from the one $S^{(q)}$, but it must measure the address of an event for which the Eqn. (13) is satisfied. Then, as the second step, it is to make measurement within the scale $S^{(q+2)}$ from the one $S^{(q+1)}$ of the address of the event which is responsible for the phenomenon mentioned above, *i. e.*, that the Eqn. (13) was not satisfied.

Let $q=0$, and the measurements are made by a macroscopic observer. According to the written above he is able to measure not only within the scale $S^{(1)}$, but also within the scale $S^{(2)}$ and, apparently, beyond because the procedure described above can be continued to $q>2$. This means, if we begin from the macroscopic observer, *i. e.*, from $S^{(0)}$, this procedure opens him the way to make measures within smaller and smaller scales. According to the written above this penetration to smaller and smaller scales is realized if at each step, in exception of the last one, Eqn. (13) is not satisfied and, therefore, the result exists only at the last step of these multi-step measurement.

It could happen that results of measurements of taking place within a scale number (q) can be interpreted on the grounds of the hypothesis that they are produced by the existence of some new objects that were not detected in this scale. However, it is possible that these hypothetic objects are within the scale $(q+1)$, but not (q). Possibly, they can be detected there by direct measurements using up to date techniques, *e. g.*, accelerators producing extremely high energy particles. Then their direct detection will confirm the abovementioned hypothesis. Perhaps, this is namely the situation with quarks (see, for example, [11,12,13,14,15,16]: QUARKS EXIST NOT WITHIN THE SAME SCALE WHERE OTHER ELEMENTARY PARTICLE EXIST, BUT WITHIN THE FOLLOWING ("SMALLER", EXACTLY, OF LOWER HIERARCHY) SCALE. Among while this interpretation leads to the following question: whether the space-time continuum exists within the very "small" quark scale? It is very probably that the answer is negative. If so, notions of transformations of (non-existing) co-ordinate systems and corresponding groups are nonsense within the quark scale.

9. PHYSICAL LAWS WITHIN EACH SCALE

Represent in the general form physical laws within a certain scale $S^{(q)}$. Let measurements have determined the address $H^{(q)}_{Y^{(q)}}$ of the subset $Y^{(q)} \subseteq A^{(q)}$. Introduce a set $\left\{\mu^{(q)}_{Y^{(q)}}\right\} \overset{def}{=} \tilde{M}^{(q)}_{Y^{(q)}} \overset{def}{=} \tilde{M}^{(q)}\left(Y^{(q)} \subseteq A^{(q)}\right)$ that attributes certain properties to the set $H^{(q)}_{Y^{(q)}}$, which means, in particular, the choice of the interpretation of the obtained measurements' results.

Call $\tilde{M}^{(q)}_{Y^{(q)}} \overset{def}{=} \tilde{M}^{(q)}\left(Y^{(q)} \subseteq A^{(q)}\right)$ ASSOCIATED SET of the subset $Y^{(q)} \subseteq A^{(q)}$ of events. The interpretation of address measurement results for different subsets $Y^{(q)} \subseteq A^{(q)}$, *i. e.*, different $\tilde{M}^{(q)}\left(Y^{(q)} \subseteq A^{(q)}\right)$, may be interdependent. It suggests an idea to introduce the set associated with the set of all subsets of the set of events:

$$\left\{\mu^{(*)(q)}\right\} \overset{def}{=} M^{(q)} = \left(\forall Y^{(q)} \subseteq A^{(q)}\right)\left[\left\{\tilde{M}^{(q)}\left(Y^{(q)} \subseteq A^{(q)}\right)\right\}\right] \tag{14}$$

The set $M^{(q)}$ we shall call THE COMPLETE PHYSICAL THEORY WITHIN THE SCALE $S^{(q)}$ **,** while $\tilde{M}^{(q)}\left(Y^{(q)} \subseteq A^{(q)}\right)$ we shall call A PARTIAL PHYSICAL THEORY WITHIN THE SCALE $S^{(q)}$ **.** The choice of sets $M^{(q)}$ and $\tilde{M}^{(q)}\left(Y^{(q)} \subseteq A^{(q)}\right)$ really means the introduction of models because there is a certain freedom of their choice, but not a "categorical imperative" what namely is to be chosen as the theory.

One of possible options is to choose $\tilde{M}^{(q)}\left(Y^{(q)} \subseteq A^{(q)}\right)$ and $M^{(q)}$ as sets of operators $\left\{\hat{\mu}^{(q)}_{Y^{(q)}}\right\} \overset{def}{=} \hat{\tilde{M}}^{(q)}\left(Y^{(q)} \subseteq A^{(q)}\right)$ and

$\left\{\hat{\mu}^{(*)}\right\} \overset{def}{=} \hat{M}^{(q)} = \left(\forall Y^{(q)} \subseteq A^{(q)}\right)\left[\left\{\tilde{M}^{(q)}\left(Y^{(q)} \subseteq A^{(q)}\right)\right\}\right]$ over the set $H^{(q)}$. How these operators determine the interpretation of measurements' results? If $H^{(q)}$ is a space, they can map (project) the set $H^{(q)}$ or its subsets to another space $R^{(q)}$, *e. g.*, a Hilbert space, and its subsets. It can be written as follows:

$$\hat{\hat{M}}^{(q)}\left(Y^{(q)} \subseteq A^{(q)}\right)\left(H^{(q)}_{Y^{(q)}} \subseteq H^{(q)}\right) = R^{(q)}_{Y^{(q)}} \subseteq R^{(q)} \qquad (15)$$

The measurement results are none other than the set of elements $\left(H^{(q)}_{Y^{(q)},mes} \subseteq H^{(q)}\right)$. These results can be interpreted only after the following operator $\hat{\hat{M}}^{(q)}_{Y^{(q)}}$ action:

$$\hat{\hat{M}}^{(q)}_{Y^{(q)}}\left(H^{(q)}_{Y^{(q)},mes} \subseteq H^{(q)}\right) = R^{(q)}_{Y^{(q)},mes}, \qquad (16)$$

i. e., not the obtained results themselves, but their projection to the space $R^{(q)}$ should be used for the interpretation.

Thus, the transition from q to $q+1$ means the search for convenient models (=physical theories) for interpretation of results of indirect measurements made on the scale $S^{(q)}$ to determine addresses of events on the scale $S^{(q+1)}$ under the condition that the physical laws on the scale $S^{(q)}$ are already known. These laws are necessary for the theory of the abovementioned indirect measurements (*cf.* [9,10]). Without them the interpretation of such indirect measurements would be impossible.

Let us use the analogy with the approach to classical scale – quantum scale transition. It must keep in mind that the choice of this way is of the hypothetical character, and that other, probably also hypothetical ways may exist to be used for the construction of physical theories within different scales. Following this way one could replace physical quantities, *i. e.*, the addresses that principally can be measured, defined within the scale number q to operators within the scale number $(q+1)$, which reminds the transition from classical to quantum scale. The obtained operators just exactly will form the theory $\hat{M}^{(q+1)}$.

The set $\{h^{(q+1)}\} = H^{(q+1)}$ may, in particular, be Hilbert space or its subset, upon which act operators of the theory $\hat{M}^{(q+1)}$. This reminds the approach of the quantum mechanics where quantities of the classical mechanics are replaced by operators acting upon probability amplitudes. Try to keep this way of $q=0$ to $q=1$ transition for all values of q. In quantum mechanics the wave function is the function of all the addresses of the classical mechanics, *e. g.*, co-ordinate, or linear momentum, and time. So we shall accept that $(\forall q)$ addresses $h^{(q+1)} \in H^{(q+1)}$ would be "wave functions" (probability amplitudes) of addresses $h^{(q+1)}$. If one accepts this way, there is no reason to keep the classical notion of the address used within the quantum and other scales. So we shall consider $(\forall q)[h^{(q)}]$ as the address itself. By this definition we break the direct connection between a measurement result and the address because, for example, in quantum mechanics the wave function is the amplitude of probability of addresses (if to stay on the probabilistic interpretation of quantum mechanics). This means, if to accept this definition, the address is not obligatory measurable, but it can serve for the interpretation of measurement results in the framework of a certain theory, as it is being done in quantum mechanics.

10. MORE ON MACROSCOPIC OBSERVER MEASUREMENT WITHIN A SCALE $S^{(q+1)}$

It was considered above (Sec. 5) how a scale $(q+1)$ can be detected by indirect measurements with feedback within the scale q and how it can be continued to scales $q+2$ and beyond. Now, taking into account the written in Sec. 9, we can give more concrete expression to this procedure. These measurements may detect that there is an event or a set of events violating the laws $\hat{M}^{(q)}$. Then, one of possible ways to interpret this fact would be the assumption that a smaller (of smaller hierarchy) scale $S^{(q+1)} < S^{(q)}$ exists with its own physical laws $\hat{M}^{(q+1)}$. The task now is to find $\hat{M}^{(q+1)}$. Remind that the multi-step indirect measurement is a sequence of indirect measurements (*cf*. the end of Sec. 5) realized, as it was describing above, steps from $S^{(0)}$ to $S^{(1)}$ **,** from $S^{(1)}$ to $S^{(2)}$, from $S^{(2)}$ to $S^{(3)}$,…, from $S^{(q)}$ to $S^{(q+1)}$. All these measurements are made in succession by a macroscopic observer who himself, by the definition, is within the scale $S^{(0)}$ .Such a multi-step measurement can be represented by the following scheme.

1. The macroscopic observer finds out that results of certain measurements cannot be interpreted on the grounds of physical laws $\hat{M}^{(0)}$ within the scale $S^{(0)}$ because Eqn. (13) is not satisfied in this case.

2. He finds out that no change of these physical laws can change this fact. Then he supposes that the scale $S^{(1)}$ of lower than $S^{(0)}$ hierarchy exists with its physical laws $\hat{M}^{(1)}$.
3. Then the macroscopic observer finds out that results of some of his indirect measurements made in the scale $S^{(0)}$ to study occurring within the scale $S^{(1)}$ cannot be interpreted on the grounds of physical laws $\hat{M}^{(1)}$ or any other physical laws within this scale, and supposes that the scale $S^{(2)}$ of lower than $S^{(1)}$ hierarchy exists with its physical laws $\hat{M}^{(2)}$.
4. *Etc.*

Note. It is important to remind that at each step the set $\hat{M}^{(q)}$ is not the only possible physical laws. Therefore, those measurements' results that cannot be interpreted on the grounds of the physical laws $\hat{M}^{(q)}$ may be interpreted on the grounds of the other physical laws $\hat{M}'^{(q)}$ within the same scale $S^{(q)}$ without hypothesis on the $(q+1)$-th scale existence. Only if it be found impossible, the existence of a new scale of lower hierarchy can be supposed and considered. Note that it demands to be very careful because results of some other kinds of measurements described well by $\hat{M}^{(q)}$ may be incompatible with $\hat{M}'^{(q)}$.

Each measurement of event addresses within the scale $(q+1)$ performed by a macroscopic observer is a multi-step sequence of measurements. This fact is a matter of principal. Indeed, this sequence of measurements with the feedback (see Sec. 3) at each step demands a certain time $\vartheta_{0,q+1}$ determined by the observer's clock which is apparently an increasing function of the number of steps. Note that it is necessary to use measurements with the feedback, so the time $\vartheta_{0,q+1}$ includes the time of measurement itself and the feedback time.

In view of this it could be expected that the maximum value $\max q \overset{def}{=} q_{\max}$ exists that limits our sequential penetration into smaller and smaller microworld scales. The reason of such a limitation existence is that there is the maximum permitted time, $\max \vartheta_{0,q+1} = \vartheta_{0,q+1,\max}$ of measurement still allowing the observer to attribute a certain time moment (with a *reasonable* error) to the information provided by the measurement within the scale $S^{(q+1)}$, while for $\vartheta_{0,q+2} > \vartheta_{0,q+1,\max}$ it becomes impossible. Even if we shall refuse from a dynamic description of the event behavior, *i. e.*, from its description as function of time, and will limit ourselves with only the connection between the initial and final states (the basic idea of the S-matrix method in the quantum collision theory), physically the time interval between the initial state creation and the appearance of the final one really cannot be ∞, but must be finite. This demands issues from the necessity to avoid processes other than the studied one to occur during this time interval (remind: observer's clock!), simply, not to blend different processes in our consideration. One more argument exits in favor of this limitation. The information that can be provided by a multi-step indirect measurement made within the scale $S^{(q+1)}$ would be expected to be of small amount and scant as to its content in comparison with that provided by measurements made within the scale $S^{(q)}$. Indeed, for example, measurements within the scale $S^{(1)}$ provide information really scant as to its content in comparison with those made within the scale $S^{(0)}$: within the classical scale $S^{(0)}$ one can measure simultaneously particle co-ordinate and corresponding linear momentum, while within the quantum scale $S^{(1)}$ it is impossible. It would be natural to suppose that this effect occurs at each $q' \to (q'+1)$. Then the considered effects $\forall q' \in [0, q+1]$ contribute to the resulting one for the information on events' addresses within the scale $S^{(q+1)}$ obtained by the macroscopic observer.

Note that this fundamental limitation of the possibility of knowing the microworld and its physical laws creates the following problem. Let within the scale $S^{(q_{\max}+1)}$ an event or a subset of events exist violating physical laws $\hat{M}^{(q_{\max}+1)}$ of this scale. In this case scales $S^{(q_{\max}+2)}$ and smaller do not exist. Therefore, this effect cannot be created by events within a neighbor smaller scale (*cf.* the written above). Then, what is its nature and origin? One way to eliminate this problem could be a kind of renormalization, which means a relevant change of physical laws, for example, the replacement of $\hat{M}^{(q_{\max}+1)}$ to another $\hat{M}'^{(q_{\max}+1)}$, but it is not always possible, as it was indicated above.

11. QUANTUM → SUB-QUANTUM SCALE TRANSITION

We have denoted classical scale $S^{(0)}$ and quantum scale $S^{(1)}$. Consider the penetration to the closest sub-quantum scale $S^{(2)}$ starting from $S^{(1)}$. The set $H^{(q)}$ at $q = 1$ is the Hilbert space of states of a quantum system $Y^{(1)} \subseteq A^{(1)}$. Then its address will be $H^{(1)}_{Y^{(1)}}$ that can be wave function or density matrix defined over the set $A^{(1)}$.

At the transition from classical scale $S^{(0)}$ to the quantum scale $S^{(1)}$ one obtains addresses in the form of wave functions defined over space-time continuum instead those in the form of space-time points. Now we transit from $S^{(1)}$ to $S^{(2)}$. Of course, different versions are possible in this case, but in the present work we shall limit ourselves with one of them and shall consider model that the address of a subset $Y^{(2)} \subseteq A^{(2)}$ at the scale $S^{(2)}$ is a function (in the general set theoretical meaning) defined over elements of Hilbert space that are addresses on the scale $S^{(1)}$. This means, we use the approach resembling to that used for the transition from classical to quantum.

Let there the set of functions $\Omega \overset{def}{=} \Omega(\{\psi\})$ of quantum Hilbert space elements. Define now the set of all elements Ω:

$$\Theta \overset{def}{=} \{\Omega\} \tag{17}$$

Because the consideration of this Section is only an example, it allows us, also as example, to choose Θ an abstract mathematical space. This choice allows us not to go away too far from the quantum mechanical formalism. What type of space? As in quantum mechanics fundamental experimental data and postulates issued from them would be necessary to answer to this question (see, for example, [17,18,19]). However, at present they are absent, and so we shall consider one hypothetical way.

If Θ is chosen as a Hilbert space, the subquantum scale will be on principle like one more step following the second quantization or, which is equivalent, Fock configuration representation or Fock functionals (see, for example, [20-23]).

Let us continue this process and transit from the scale $S^{(2)}$ to the scale $S^{(3)}$. By each step we suppose as before, that, the corresponding set is a space and that this space is the Hilbert one. Thus, we shall obtain the set

$$\Xi \overset{def}{=} \{\Theta\} \tag{18}$$

If to suppose that the set Ξ is a Hilbert space, one will obtain sub-quantum scale physics as something on principle like the 4th quantization. Note that we use the term Hilbert space only for short.

Really, these spaces should be supposed to be like that used in quantum mechanics, which is, generally speaking, not the Hilbert one, in particular, because of δ-functions.

12. NOTES ON MICROWORLD SCALES AND RELATIVISTIC QUANTUM FIELD THEORY

This Section is dedicated to a discussion on possible applications of microworld scales' theory developed above in this work to problems of relativistic quantum field theories. We consider here how the existence and properties of microworld scales influence the relativistic quantum theory and quantum field theory. In this respect of such a consideration an important problem is whether one is allowed to "guillotiner" without heavy consequences the well-ordered set of microworld scales or such a use of the "guillotine" could lead to incorrect results. So it would be desirable, first of all, to clarify whether this sequence is in fact infinite (as it seems) or finite. We shall return below to the connection of this problem with the quantum field theory. Notice here only that it is important because such a "guillotineering" the sequence of microworld scales or *natural* limitation of the microworld scales' number (if it is finite) leads with the necessity to a certain upper limit of particle energies that is to be taken into account at summations (integrations) with respect to all possible states.

Possibly, Dirac electron – positron vacuum in his relativistic electron theory [17] is a kind of approximate, possible to say, phenomenological representation of the influence of microworld scales of small hierarchy ("small" scales). A correct consideration of relativistic electron must include smaller hierarchy microworld scales and its interactions within them because the electron of very high energy penetrates into such scales (as it penetrates, for example, inside nuclei (see, for example, [24-27]). In such a consideration the negative total electron energy cannot appear because the energy balance must include energies of electron interaction within all microworld scales, but not only within the elementary particle scale (EPS), for which Dirac equation is written. The Dirac theory can be really considered as a kind of the renormalization when the global influence of smaller than of EPS hierarchy scales is approximately taken into account by the phenomenological model of the sea of electrons with negative total energy (electron – positron vacuum). The written above does not mean that positrons exist in such small hierarchy scales. Something occurring within one or some such scales influences the occurring within the EPS, and in Dirac theory namely this influence is approximately expressed in terms of the electron – positron pairs' appearance, fluctuations of the electron – positron vacuum *etc.*.

Possibly, divergences, the renormalized electron charge zero [28-30] a. o. difficulties of the quantum field theory (see, for example, [31-34] are originated from the influence scales "smaller" (or exactly, of hierarchy smaller) than the scale where the corresponding quantum theory of field is defined (EPS). Indeed, quantum theory of fields is a relativistic theory and so includes, as in the case of Dirac equation, very high energies of particles. Thus, summations must include (virtual) particles of such energies. Such particles are able to penetrate into scales of very small hierarchy, their behavior depends on their interaction within such scales and by this way these scales contribute to results of calculations obtained within the scale where the quantum field theory is used. Therefore, in a theory logically complete and closed into it ALL these scales must be taken into account. Probably, by this way the quantum field theory (principally very beautiful and promising one) could be "reanimated".

It is important for this purpose that, as it was established above, there is a scale of the minimum hierarchy (the "smallest" scale), and, by this reason, the maximum number $q_{\max}$ of scales exists. This number "cuts" naturally the sequence of scales which also limits the highest value of particle energy that is to be taken into account at the summation with respect to all virtual states. It could be expected that this limitation naturally prevents the divergences appearance.

The consideration by the quantum field theory is performed within a certain scale of the microworld (EPS). It seems we "work" all the time within this scale, all mathematical formalism is constructed just for this scale without taking into account even the existence of other scales. However, as we have just seen, the occurring within scales of smaller (than of EPS) hierarchy may influence the occurring within the EPS. How to describe this influence? It depends on the theory. If one considers models based on (elementary) particles, in some cases (but not in the general case!) one can consider transitions of such particles from the considered scale to those of smaller hierarchy and vice versa. In the framework of more general approach processes, interactions and objects that are within smaller hierarchy scales must be taken into account at consideration of their consequences detected within the scale of elementary particles.

From the proposed point of view the existing concept itself of the vacuum seems to be none other than an attempt to take into account the influence of going on within low-hierarchy scales upon considerations made within EPS by use of an approximate phenomenological model. However, the real vacuum (not this model!) is the set of all microworld scales.

The classification of elementary particles based on the group representation theory (see, for example, [35]) also can be affected by taking into consideration low-hierarchy microworld scales. Indeed, the symmetry can be violated by the occurring within such scales when very high energies of particles are taken into account. Roughly speaking, the interaction of particles within microworld scales of hierarchies lower than that of the EPS could be an addition to the Lagrangian breaking the symmetry existing within the EPS. This effect becomes essential when energies of particles that are to be taken into account at summations (with respect to intermediate states) are sufficiently high to let particles penetrate into such low-hierarchy ("small") scales. Strictly speaking, a correct theory must take into account ALL microworld scales. There is no matter whether, for example, the Lagrangian formalism can be used or not.

13. CONCLUSIONS

In the present paper the general approach is proposed to the successive penetration to well ordered sets of smaller and smaller hierarchy from the set having the highest hierarchy. This hierarchy is set up within a well ordered set of well ordered sets. Scales of the microworld are defined as a particular case of this system, and the penetration to smaller and smaller scales is replaced to the penetration to sets of lower and lower hierarchy. The substitution is necessary because in the considered case the word "small" has only intuitive meaning, if possible to say so. We simply got accustomed that atomic and nuclear scales have characteristic sizes $\sim 10^{-8}$cm and $\sim 10^{-13}$cm, correspondingly, and so we think in terms of sizes. But such an approach becomes unclear when we try to study smaller scales. The physical meaning of "small" will be lost as a consequence of the impossibility to define the concept "size" because it demands the existence of space-time, which is very questionable there. The definition of the notion "scale of the microworld" must be based on physical properties of occurring events. While events possess properties of a certain class, they all are within a certain scale. By this way we get rid of the use the size of a space-time region with this purpose. This approach allows one to study mathematically microworld scales' system as hierarchic well ordered set of well ordered sets filling these mathematical objects with the physical content.

In Secs. 2-4 we considered and developed different ways to establish hierarchy among different kinds of the information and among sets. An important way to do it is based on the consideration of the information value. An attempt is made to represent the information value mathematically, though this problem is extremely complicated because each case differs from the other ones. Three notions were defined: information value, potential information value and constrained information value. The hierarchy between different kinds of the information can be established on the grounds of the information value, or potential information value, or constrained information value, depending on the considered case.

In the case when the information is expressed in a language (natural or formal) the hierarchy can be established also on the grounds of Russell's theory of types. The formalisms developed in Secs. 2-4 serve as the mathematical ground of the microworld scales theory.

Our general mathematical approach to the problem of physical laws within different microworld scales is based on the use of two well ordered sets. Elements of one of them we have denoted *events*. We mean that in applications to the physics this term includes objects (*e. g.*, electron) and events in proper meaning occurring with them (*e. g.*, electron scattering). Elements of the second set we have denoted events' addresses.

Address of a subset of the set of events is a certain subset of the set of addresses put to correspondence to this subset of events by a homomorphic mapping with the feedback. The feedback is necessary to "label" the considered subset of events by the corresponding subset of the set of addresses. The mapping is realized by measurements made always by an observer within the highest hierarchy set. In physics it corresponds to the macroscopic observer. Indirect measurements are considered as the only type of them allowing such an observer to obtain the information on occurring within a set of low hierarchy. Limitations of our possibility to penetrate to low hierarchy sets (to small scales of the microworld, in physics) are found as issued from the multi-step character of measurements.

It was indicated that possibly quarks are not within the same scale that elementary particles, but within the following ("smaller", exactly and without inverted commas, lower hierarchy) scale where the existence of space-time continuum is questionable. If it really does not exist, co-ordinate transformations and their groups are nonsense within this scale, and, therefore, the theory of corresponding group representations is not fit for quarks study.

The (physical) theory is defined within each scale as a set that may be a set of operators (this option seems us to be the most realistic one, but for the present yet only seems). Then the main task is to find this set for each scale in consideration. In Sec. 9 one proposed a hypothetical way how to do it within different scales.

It must warn that the penetration deeper and deeper to small scales of the microworld is not a high way, but very complicated and sometimes even contradictory process. If at a certain step one finds event that cannot be understood in the framework of physical laws of the considered scale, the solution of this problem is in introduction a neighbor lower hierarchy scale with its own physical laws, iff it is found out that it cannot be obtained in the framework of the considered scale by the change its physical laws. So it is necessary to search for experiments and theoretical arguments to distinct between these options. Really, in the framework of each of these two options many "sub-options" exist (*e. g.*, different versions of physical laws), which may complicate essentially each step.

We want to call the attention to the fact that the existence of scales is not postulated, the proposed theory allows one find out what scales of the microworld do exist.

In the contemporary considerations scales of the hierarchy smaller than the EPS hierarchy are not considered explicitly and instead, to take nevertheless their effect into account one considers a certain model, for example, the Dirac background of electrons having the negative energy. The main difficulty of the taking into account effects produced by smaller hierarchy scales is that we do not know the going on and the physical laws within these scales. The renormalization (see, for example, [31,33,36-39]) is an attempt to exclude or to isolate this effect or, at least, its parts that could be considered as not important, being, in fact, a kind of phenomenological approach.

The similar factors, such as the taking into account interactions not only within EPS, but also within lower-hierarchy scales, could affect the classification of elementary particles on the grounds of their symmetry group representations [35].

Thus, relativistic quantum field theories (independently of particle energies that should be taken into account) limit the consideration with only one EPS and physical laws existing within this only scale, forgetting other scales with their physical laws. This is the main cause of well known difficulties of these theories.

REFERENCES

1. Manfred Eigen, Naturwissenschaften **58**, 465 (1971)
2. M. V. Volkenstein, Found. Phys. **7**, 97 (1977)
3. Bertrand Russell, American J. Math. **30**, 222 (1908)
4. C. E. Shannon, The Bell System Technical Journal **27**, 379, 623 (1948)
5. Leon Brillouin, *Science and Information Theory*. Academic Press Inc., Publishers, New York 1956
6. Alfred North Whithead & Bertrand Russell, *Principia Mathematica*, Second Edition, Cambridge at the University Press 1963
7. Tomas Jech, *Set Theory, Springer Monographs in Mathematics.* Springer-Verlag, Berlin 2003
8. L. I. Mandelstam, *Lektsii po Optike, Teorii Otnositel'nosti i Kvantovoi Mekhanike (*in Russian), Moskva, Nauka 1972 (*Lectures on Optics, Relativity Theory and Quantum Mechanics*, Moscow, Nauka 1972).
9. V. B. Braginsky & F. Ya. Khalili, *Quantum Measurement.* Cambridge University Press 1992

10. G. Auletta, *Foundation and Interpretation of Quantum Mechanics.* World Scientific, Singapore - New Jersey - London - Hong Kong 2000

11. J. J. J. Kokkedee, *The Quark Model*, W. A. Benjamin, Inc., New York – Amsterdam 1969

12. Y. Nambu, *QUARKS*, World Scientific Publishing Co Pte Ltd, Singapore 1985

13. V. N. Gribov, *Gauge Theories and Quark Confinement*, Moscow: PHASIS 2002

14. Atsushi Hosaka & Hiroshi Toki, *Quarks,Baryons and Chiral Symmetry*, World Scientific Publishing Co Pte Ltd, Singapore-New Jersey-London-Hong Kong 2001

15. Yuval Ne'eman, Nucl. Phys. 26, 222 (1961)

16. Murray Gell-Mann & Yuval Ne'eman, *The Eightfold Way*, W. A. Benjamin, Inc., New York – Amsterdam 1964

17. P. A. M. Dirac, *The Principles of Quantum Mechanics*, Oxford at Clarendon Press 1958

18. L. I. Shiff, *Quantum Mechanics*, McGrow-Hill Book Company, Inc., New York-Toronto-London 1955

19. L. D. Landau & E. M. Lifshitz, *Quantum Mechanics: non-relativistic theory*, Oxford, Pergamon, Butterworth-Heinemann 1977

20. F. A. Berezin, *The Method of the Second Quantization*, New York, Academic Press 1966
21. A. S. Davydov, *Quantum Mechanics*, Oxford, Pergamon Press 1976
22. V. Fock, *Zeitschrift für Physik* **75**, 622 (1932)
23. V. Fock, Uchonyie Zapiski Leningradskogo Universiteta № 17, 1937 (in Russian) (Scientific Proceedings of Leningrad University № 17, 1937); V. Fock, Sowiet Physics **6**, 425 (1934)
24. Robert Hofstadter, *Electron scattering and nuclear and nucleon structure: a collection of reprints with an introduction.* New York, Benjamin 1963
25. R. Hofstadter, B. Hahn, A. W. Knudsen, and J.A.McIntyre, Phys. Rev. **95**, 512 (1954)
26. Robert Hofstadter, Rev. Mod. Phys. **28**, 214 (1956)
27. Robert Herman and Robert Hofstadter, *High-energy electron scattering tables.* Stanford, CA, Stanford University Press 1960
28. I. Pomeranchuk, Doklady Akademii Nauk SSSR (Proceedings of the Academy of Sciences of the USSR) **103**,1005 (1955) (in Russian)
29. V. B. Berestetskiĭ, Sov. Phys. Usp. 19, 934 (1976)
30. L. D. Landau, A. A. Abrikosov, I. M. Khalatnikov, Doklady Akademii Nauk SSSR (Proceedings of the Academy of Sciences of the USSR) **95**, 497 (1954) (in Russian)

31. Nikolai N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized fields.* 3rd ed., New York, Interscience 1980

32. Lewis H. Ryder, *Quantum Field Theory*. 2nd ed., Cambridge: Cambridge University Press 1996

33. H. Umezawa, *Quantum Field Theory*. North-Holland Publishing Company, Amsterdam 1956

34. Steven Weinberg, *The Quantum Theory Of Fields*. Cambridge: Cambridge University Press, vol. I – III 1995 – 2000

35. Yu. B. Rumer & A. I. Fet, *Teoria Unitarnoi Simmetrii* (*Theory of Unitary Symmetry). "*Nauka", Moscow 1970 (in Russian)

36. K. Hepp, *Théory de la renormalization.* Springer-Verlag, Berlin-Heidelberg-New York 1969

37. Edward B. Manoukian, *Renormalization.* Academic Press, New York 1983

38. John C. Collins, *Renormalization.* Cambridge: Cambridge University Press 1984

39. Manfred Samhofer, *Renormalization.* Berlin, Springer 1999

ARTICLE 2

ON THE REFERENCE SYSTEM CONCEPT IN GENERAL RELATIVITY AND BEYOND

A. Ya. Temkin

Department of Interdisciplinary Studies

Faculty of Engineering

Tel-Aviv University

Ramat-Aviv

Tel-Aviv 69978

Israel

E-mail: **temkin@eng.tau.ac.il**

Fax: +972-3-6410189

Tel: +972-3-6408176

ABSTRACT

The concept of ***macroscopic reference body*** in the Universe is defined and considered. Thereupon the ***reference system*** is defined as a certain set of reference bodies. The procedure of the reference system concept definition naturally leads to the possibility of the existence of holes in the Universe. It is important that the both concepts are defined without making use of the space (-time) concept and so it opens a way to construct new, more general theories instead the ones traditionally based on certain kinds of space (-time) and co-ordinate transformations. After the concept of reference system had been defined, the conditions were elucidated determining when the concept of space (-time) can be defined. Only then is possible to construct theories such as the General Relativity, Unified Field theories *etc*.. The use of high-dimensional Riemannian spaces, as well as of Hilbert, metric and normalized spaces for complex systems having large number of internal states (biological systems, interiors of stars *etc*.) is discussed.

Key words: General Relativity, reference system concept, cosmology.

GENERAL VIEW

This work was born from author's reflections on the Landau & Lifshitz (L. Landau & E. Lifshitz, The Classical Field Theory. Addison-Wesley Press, Inc., Cambridge, Mass., 1951; First Russian Edition, Moscow-Leningrad, 1948) definition of the reference system concept in General Relativity. I read it firstly when I was a student of Physical Faculty of Moscow University at the end of 40^{th} and the beginning of 50^{th}. Then I was satisfied with this definition. However, despite it I felt all the time a kind of "internal anxiety", maybe more subconscious then conscious, that, as it had seemed to me, possibly something lack in this definition. But only some years ago I was ready to think seriously on this problem with the purpose this time to elucidate it. The natural logical train of thoughts resulted me in the conclusion that Landau & Lifshitz definition of the reference system concept in General Relativity is, in fact, an excellent starting point, but not yet the definition. The obtained result was unexpected for me because it exceeded far the limits of the abovementioned problem. It turned out that this is the starting point not only for the search for the definition of the reference system concept in General Relativity. The Landau & Lifshitz definition is also the starting point for the construction of new physical theories much more general than the General Relativity. Among them theories that do not use the concept of space or space-time. It is the starting point also for the elucidation where the concept of space (space-time) is from and under what conditions it can be introduced. Perhaps, the main result of this work has to do namely with the emergence of the space and its dimension. We, physicists have got accustomed to the existence of space. For longtime we knew only 3D-Euclide space and could not imagine that spaces having more than 3 dimensions may exist. But the time did its job. And, as result, the majority of physicists already reconciled themselves (it took two or three dozens of years after the appearance of the Special Relativity) with the 4D-space where the time is the 4th dimension, nobody is astonished at 5D-, 7D- and so on physical theories. Briefly, no matter what space is used in a certain physical theory, any space is welcome, but to call in question the fundamental fact of the space unconditional existence seems us unbearable. Even the thought itself that in certain Even the thought itself that in certain cases it is to construct physical theories without use of space (because it cannot be defined) seems preposterous. However, in this paper is showed that *the absence of space is the general case*, while a space exists only in certain special cases. In Sec. 2 is showed that the definition of the reference system concept is the necessary condition for the definition of the space concept while customarily the reference system concept is defined in already existing space. For it one chooses a mathematical space of a certain type (*e. g.*, Riemannian space) that should be relevant to the fundamental tenets made on the grounds of physical experiments. Thereupon one defines measurements of a set of quantities in this space made by observers to establish an isomorphism between mathematical quantities of the chosen space and physical data. It transforms the chosen mathematical space into a physical one. But whether physical reasons allow us to introduce any space (of any type) always? It can be elucidated only if to build a theory that does not presuppose the existence of a physical space, in other words, does not presuppose that the physical information can, in general, be expressed in terms of any space. Thereupon it would be interesting to search for conditions under which the space can be introduced. The reference system is usually defined as a set of reference bodies, as of its elements. So we use the *concept of the reference body, i. e.,* ***a physical body with an observer***, as the starting point of the search for the definition of the reference system concept. As we shall see in Sec. 2, the definition of the reference system concept will allow us to consider how and under what conditions the space emerges.

The absence of space creates, among other problems, the problem of the communication between observers. From the first work of Albert Einstein "Zur Elektrodynamik bewegter Körper" observers play the fundamental role in the physical theory sending light signals to each other. As it is emphasized in Sec. 2, the light, *i. e.*, electromagnetic waves, needs a space for its existence. If our starting point in the general case is that the concept of space is not yet defined, one cannot yet suppose the light existence. Therefore, if one wants to build a theory, it is to assume the existence of a type of signals fit for the information transfer between observers. If it does not exist, no theory can be built. It is very serious phenomenon. Perhaps, the existence of the dark matter means just that observers being on two different types of physical bodies have no means for the information exchange. In Sec. 2 we proceed from the assumption that the necessary signals exist though we cannot yet concretize their nature. One important conclusion is issued from this assumption. Let there is one set of observers able to communicate among themselves. Let there is the second set of observers also able to communicate among themselves, but unable to communicate with observers belonging to the first set. In this case two different physical theories should be constructed, separately for each of these two sets of observers. If observers belonging to the second set are unable to communicate even between themselves, they are useless, and no theory for this set of observers can be built. These two possibilities must be taken into account at the dark matter study.

In the framework of Sec. 2 the problem of the Lorentz transformations' applicability to biological systems and automata with a great number of states is discussed. As it was elucidated, their applicability and, in general, validity for such systems are, at least, very questionable. In particular, the ordinary approach to the Twins Paradox is called in question. One of reasons of these doubts is as follows. It is well known that internal processes in living organisms proceed not during only the universal physical time as mechanical processes do. They proceed during biological times such as circadian time, thinking time *etc*. (for the general mathematical definition of time of any type and its direction see also our book "***SOME IDEAS ON INFORMATION PROCESSING, THINKING AND GENETICS***" (Tel-Aviv University Press, 1999; available also on-line (free): http//www.eng.tau.ac.il/~temkin).

1. INTRODUCTION

The problem of the ***reference system*** (***RS***) concept definition in the General Relativity is very complicated and difficult. In [1, p.249] is written: *"… **for the exact determination of the position of bodies in space in the presence of a gravitational field, it is necessary, strictly speaking, to have a system made up of an infinite number of bodies, filling all of space. Such a system of bodies together with clocks linked to each body and recording time in an arbitrary way constitutes a system of reference in the general theory of relativity**"*. However, such a situation is never realized because physical bodies never fill in the whole space, but there are empty regions between them. Moreover, this definition of RS has meaning only, if the concept of space is defined ("filling all of ***space***"), while the concept of space, in its turn, can be defined only, if the concept of the reference system is defined. Thus, one obtains the vicious circle. In the present paper we shall reconsider the whole problem of RS definition starting from the consideration of a ***reference body (RB).*** The next step will be the elucidation how and under what conditions RS can be constructed of a set of RBs. At these first and second stages of our consideration the concept of space (-time) is not defined and, therefore, the General Relativity [2] and other theories based on this concept, cannot yet be constructed. Thereupon it will be found under what conditions and how the space and time in the Universe and its regions can be defined, in particular, the space-time of the General Relativity Theory, Unified Field Theories *etc.*.

In Secs. 2 - 4 we define the concept of RS in the framework of an essentially macroscopic theory valid and fit for the Universe and its regions. Extrapolations to small bodies and the microcosm become possible only when the concept of space - time is defined (Sec. 5). This and other limitations will be seen clearly from the following consideration.

2. REFERENCE BODIES

Let us consider firstly the situation when the Universe contains only *one* ***macroscopic*** physical body *A* and there is a macroscopic observer ***A*** on this body. Observer can be a human being or automaton. We accept that the observer has instruments necessary to make measurements. The task of the observer is to interpret their results. Without this interpretation measurements provide no information. In the following text we shall write for short "observer" instead "observer and measuring instruments". The size of *A* must be not less than the one of ***A*****.** Let within the limits of the observer ***A*** and his closest vicinity on the body *A* he can define the concepts of direction and local time (*i. e.*, valid within the same limits), and can measure them. The defined concepts are needed to the observer only to determine the directions and time moments of emitted and accepted signals. Note that we do not introduce the notion of ***light*** signals because the concept of light, *i. e.*, electromagnetic waves, can be defined only if the notions of space and time are already defined outside the closest vicinity of the observer. ***We shall suppose that signals of any nature exist possessing only the following properties:***

1. ***Signal can be emitted and can be detected (accepted).***
2. ***Signal possesses the direction within the considered vicinity of the observer.***

Note that for a certain generalization of the proposed theory, the body *A* in the following consideration may be meant as a certain part of a larger body $\mathcal{A}$. If the body $\mathcal{A}$ is essentially larger than **A**, one will consider a region of $\mathcal{A}$ including **A** as the body *A*. Choice of $A<\mathcal{A}$ or $A=\mathcal{A}$ depends on the task: whether one wants to obtain the information on the body $\mathcal{A}$ itself and on the outside World or on the outside World only.

The task of the observer is to send signals out from *A* to different directions, to accept incoming signals from observers on other physical bodies responding to his signals, and to extract the information carried by them. It must emphasize that, naturally, the notion of the signal velocity is not introduced.

DEFINITION. *The body A* ***together with the observer A we shall call REFERENCE BODY*** (***RB***) and denote it $(A\,|\,\mathbf{A})$.

Let us consider secondly the case when the Universe contains only *two* physical bodies *A* and *B* with observers **A** and **B**, correspondingly. These observers are supposed to have the necessary measuring equipment.

Consider now the measurement. At a certain time moment $t_{A,1}$ the observer ***A*** sends a signal to a certain local direction $\Omega_{A,1}$ determined on A within a certain error $\delta\Omega_{A,1}$. He repeats such sending at different time moments $t_{A,i}$ in different directions $\Omega_{A,i}$ labeling each signal by the time and direction of its emission, as well as by his identity code. If at time moment $t^{l}_{A,i}$ he receives a signal from ***B*** in response to his signal sent at the time $t_{A,i}$, he registers that there is the body B reached from A in time $\Delta_{AB}t \overset{def}{=} \left(t^{l}_{A,i} - t_{A,i}\right)/2$ by the signal sent in the local direction $\Omega_{A,1}$. It must emphasize that this formula is the <u>definition</u> of the time $\Delta_{AB}t$. By the similar way one can consider a certain set of such bodies $A, C, C', C'',\ldots,C^{(n)}$. Then he will receive such response signals from all other observers and can register the time necessary to reach each of them and the directions in which he sent the corresponding signals. In more general case, such a set $\{C^{(i)}\}$ of physical bodies can be considered then a body $B\in \{C^{(i)}\}$ can be reached from another body $A \in \{C^{(i)}\}$ not only directly, but also by different ways (passing through different subsets of the set $\{C^{(i)}\}$) by means of signal transmission from one observer to another. If each observer labels his signals by his identity code then ***A*** must know details about the way passed by the signal, to extract the information carried by the obtained signal.

Return now to the case of two bodies A and B. Let the observer ***A*** emitted two signals at different time moments $t_{A,1}$ and $t_{A,2} > t_{A,1}$ and in different directions $\Omega_{A,1}$ and $\Omega_{A,2}$ (each of $\Omega_{A,l}$ is of unit length). Suppose that the response signals of the observer ***B*** to these signals of ***A*** reach the observer ***A*** at time moments $t^{B}_{A,1}$ and $t^{B}_{A,2}$, correspondingly, entering in directions $\Omega^{B}_{A,1}$ and $\Omega^{B}_{A,2}$ (determined in ***A***'s close vicinity on *A*), correspondingly. One can determine

$$\Delta^{(1)}_{AB}t = \tfrac{1}{2}\left(t^{B}_{A,1} - t_{A,1}\right), \Delta^{(2)}_{AB}t = \tfrac{1}{2}\left(t^{B}_{A,2} - t_{A,2}\right) \tag{1}$$

and

$$\Delta^{(1)}_{AB}\Omega = \Omega^{B}_{A,1} - \Omega_{A,1} \text{ and } \Delta^{(2)}_{AB}\Omega = \Omega^{B}_{A,2} - \Omega_{A,2} \tag{2}$$

If at least one of the differences

$$\Delta^{(21)}_{AB}t \overset{def}{=} \Delta^{(2)}_{AB}t - \Delta^{(1)}_{AB}t \text{ and } \Delta^{(21)}_{AB}\Omega \overset{\text{def}}{=} \Delta^{(2)}_{AB}\Omega - \Delta^{(1)}_{AB}\Omega \tag{3}$$

is not equal to the zero, this fact is the **definition** of the concept of the **relative movement** of *A* and *B*:

DEFINITION: ***TWO RBs ARE IN THE STATE OF RELATIVE MOTION, IFF***

$$\Delta^{(21)}_{AB}t \neq 0 \text{ or } \Delta^{(21)}_{AB}\Omega \neq 0, \text{ or } \Delta^{(21)}_{AB}t \neq 0 \text{ and } \Delta^{(21)}_{AB}\Omega \neq 0. \tag{4}$$

Now try to find characteristics of the relative motion. Suppose that the repetition frequency of signals sending by ***A*** can be done very high, and consider the dependence of $\Delta^{(21)}_{AB}t$ on the time interval Δt_{em} between two subsequent emissions of light signals. One can characterize the relative motion by the dimensionless "velocity"

$$v_{AB} = \lim_{\Delta t_{em} \to 0} \frac{\Delta^{(21)}_{AB}t}{\Delta t_{em}} \quad . \tag{5}$$

It is clear that it is impossible to realize $\Delta t_{em} \to 0$ because the direction to what the next signal must be send to reach ***B*** is not known, and it must spend time >0 to send signals to many different directions to find the right one. Therefore, maximum that could be done is to limit with the smallest value of Δt_{em}, which means the approximate determination of the "velocity" as a ratio of two finite differences. It must emphasize that ***principally*** cannot be done more, *i. e.*, the "velocity" ***principally*** cannot be determined exactly.

It is evident that this "velocity" is not analogous to the vector of velocity in classical or relativistic mechanics, but only to its length. If to introduce the notion

$$\vec{\omega}_{AB} = \lim_{\Delta t_{em} \to 0} \frac{\Delta_{AB}^{(21)} \Omega}{\Delta t_{em}} \quad , \tag{6}$$

then the set $\{v_{AB}, \vec{\omega}_{AB}\}$ would be analogous to the velocity vector. It must remember that $\Omega \overset{def}{=} \vec{\Omega}$ is a unit length direction in the closest vicinity of the observer on the body *A*.

Operations (*e. g.*, processing of the information carried by ***A***'s signal) that ***B*** must perform to interpret the incoming signal and to send the response signal to ***A***, can take the time τ. Taking it into account, one obtains

$$\Delta_{AB,\tau}^{(1)} t = \tfrac{1}{2}\left(t_{A,1}^{B} - t_{A,1}\right) - \tau \text{ and } \Delta_{AB,\tau}^{(2)} t = \tfrac{1}{2}\left(t_{A,2}^{B} - t_{A,2}\right) - \tau$$
$$\text{instead } \Delta_{AB}^{(1)} t \text{ and } \Delta_{AB}^{(2)} t \tag{7}$$

written above. If τ is constant the values of $\Delta_{AB}^{(21)} t$ remains unchanged for all pairs of neighbor signals.

However, τ may be different for different pairs, *e. g.*, when the content and amount of the information carried by different signals are different and demand different time for its processing and response to it. It can be expected that if only something analogous to mechanics is considered, $\tau = const$ because in this case the transferred information is very poor. But when ***A*** forwards to ***B*** more sophisticated and diverse information, *e. g.*, on processes occurring on/inside *A,* or search for such an information about *B* (*A* and *B* may be two stars), τ as well as $\Delta_{AB}^{(21)}t$ _depend on the amount and, probably, on the type of the information_.

This is a very important point. In mechanics (classical and relativistic, including the general relativity) only very poor information on considered physical bodies (for example, on co-ordinates, linear and angular momenta) is transferred from one RS to another and the co-ordinate transformations physically accompanied by just such a poor information transfer. At the same time, despite this fact it is almost commonly accepted that they are universal and valid for all kinds of objects, no matter what is the information that should be transferred from one RS to another. For example, Lorentz transformations of the special relativity are applied to the "Paradox of Twins" despite the evident fact that the state of a human being is determined by a lot of information of different types. So, in this case τ could have not the same value for all pairs of signals, which makes Lorentz transformations not valid for this case. It must be also taken into account that each biological system possesses of some types of the biological time. Processes inside it occur with respect to these types of time. The time in kinetic equations for such a process is the relevant (to this process) biological time, but not the universal physical time. There is no reason to suppose that biological time is transformed according the Lorentz transformation law for the universal physical time. Therefore, the interpretation of "Paradox of Twins" on the grounds of Lorentz transformations is at least doubtful. A correct approach to the "Paradox of Twins" demands to know how information on their internal states is transformed at the change of the RS, no matter whether it is expressed by use of biological time or not. An important question is whether an observer is able to accept, decode and interpret all this information in reasonable time interval? It must be clarified about what occurs, if ***only a certain part*** of the information mentioned above is transferred to the new

RS. For example, what occurs in the case when the whole information on one of two twins is transferred to the new RS, while only a part of the one about the second twin is transferred? In this case, if one RS is the Earth and the second one is the cosmic ship with *two* twins, what will be after their return to the Earth? This case is another version of Twin Paradox. ***The written here refers also to all biological systems, automata with a great number of states, interiors of stars and, probably, many other types of existing systems.***

Consider this problem with more details. In the case of such a complex system a RB obtains two types of information:

1. Information on this object considered simply as a physical body (mechanics of the object). It can be referred to a certain moment of the local time, or exactly to its environment of the duration τ_1.
2. Information on its internal states and their changes. This information can be referred to a certain local time interval of the duration $\tau_2 >> \tau_1$.

The total information is an inseparable (in the general case) combination of these two types of information. An important task is to find a representation of this information fit for its decoding and interpretation by the observer.

3. TWO-BODY REFERENCE SYSTEM

We considered above the primary concept that is the one of RB $(A \mid \boldsymbol{A})$.

Not always is possible to introduce RSs. However, it must not forget that it is very important concept, on use of which are based Classic Mechanics and Special Relativity, where RS is a set of RBs together with co-ordinate system. A purpose of the present work is to introduce the concept of RS to the general relativity. RS is a set of RBs **organized by a special way** that will be considered below. We shall begin this consideration from the RS formed of two RBs $(A_1 \mid \boldsymbol{A}_1)$ and $(A_2 \mid \boldsymbol{A}_2)$. Let Q is a macroscopic trial physical body without observer. Thus, the body Q cannot send signals. We suppose, however, that it reflects signals sent by RBs. Note that if there are some (more than one) types of signals that can be sent by observers, it is possible that the body Q reflects not all of them. Then the use of different types of signals by observers gives different pictures of the Universe. This possibility is not considered in the present paper, but it merits to be considered seriously. In the following text we shall limit ourselves with only one type of signals. Then each RB is able to send signals of this chosen type and by use of them to detect the body Q and to find its distance and direction (determined locally in the close vicinity of the observer) at a certain moment of RB local time. It is possible only if the body Q reflects signals. If it absorbs all signals reached its surface, this body cannot be detected and no measuring can be done. It may create a serious problem in the theoretical description of the Universe because it would be impossible to distinguish between its empty and not empty but non-reflecting parts.

Under what conditions more information on the physical body Q can be obtained when these two RBs form a RS $\mathscr{M}_2 = (A_1 \mid \boldsymbol{A}_1) \cup (A_2 \mid \boldsymbol{A}_2)$ than in the case of two independent RBs? A necessary condition of it is as follows. This RS is able to provide such an information that cannot be provided by RBs $(A_1 \mid \boldsymbol{A}_1)$ and $(A_2 \mid \boldsymbol{A}_2)$ considered separately, if the set of data concerning the body Q obtained by $(A_2 \mid \boldsymbol{A}_2)$ cannot be calculated as function of the one obtained by $(A_1 \mid \boldsymbol{A}_1)$, and vice versa. If this necessary condition is fulfilled, RBs $(A_1 \mid \boldsymbol{A}_1)$ and $(A_2 \mid \boldsymbol{A}_2)$ form RS $\mathcal{M}_2 = (A_1 \mid \mathbb{A}_1) \cup (A_2 \mid \mathbb{A}_2)$, iff $\boldsymbol{A}_1$ obtains an information about the occurring with $(A_2 \mid \boldsymbol{A}_2)$, including an information on its local co-ordinate system and local time transformations, its position and movement relatively A_1, as they were defined above; and vice versa. One can demand "the complete information" instead this uncertain "an information". However, it would introduce unnecessary restrictions expressing the tendency to act in the spirit of the existing theories. Denote $\xi_k \quad (k = 1,2)$ the set of data obtained by the RB $(A_k \mid \boldsymbol{A}_k)$ and $\Xi^{(2)} = \{\xi_k\}$. Here $\Xi^{(2)}$ is the set of results of measurements performed upon the body Q relatively two considered RBs, but not in two-RB RS. It is to eliminate from $\Xi^{(2)}$ the number of data equal to the number of functional dependencies between them. Then the new set of data $\Xi^{(1*2)}$ will be co-ordinates and time of the body Q in the RS $\mathcal{M}_2 = (A_1 \mid \boldsymbol{A}_1) \cup (A_2 \mid \boldsymbol{A}_2)$.

4. REFERENCE SYSTEM FORMED BY A FINITE OR COUNTABLE SET OF REFERENCE BODIES

Let us consider a finite set $\mathcal{M}_N = \{(\forall (1 \le k \le N))(A_k \mid \boldsymbol{A}_k)\}$ of RBs $\{(A_k | \boldsymbol{A}_k)\}$, where k and N are integral numbers. This consideration can be generalized to $N \to \infty$, *i. e.*, to countable set of RBs. Let all $(A_k \mid \boldsymbol{A}_k) \in \mathcal{M}_N$ send light signals. Let some of them

$\left(A_{k'} \mid \boldsymbol{A}_{k'}\right) \in \mathcal{M}' \subseteq \mathcal{M}_N$ have received signals reflected from Q. Then results of measurements made by each $\mathbb{A}_{k'}$ will be represented by Eqns. (1), (2) and (3) with the substitutions of B by Q, A by $A_{k'}$ and "response signal" by "reflected signal". Then Eqns. (4), (5), (6) and (7) with the same substitutions will characterize the relative movement of the bodies $A_{k'}$ and Q.

Up to this place we considered a set of independent RBs. The first necessary condition that a system formed of these bodies would be RS is that it must be able to provide information, which cannot be provided by the set $\left\{\left(A_k \mid \boldsymbol{A}_k\right)\right\}$ of independent RBs. A necessary condition of it is, in its turn, that the set of data concerning the body Q obtained by any of these RBs cannot be calculated as function of the one obtained by the rest of them. This statement opens the way to understanding of that how to determine the subset that can be used as RS. Let the data obtained by the RB $\left(A_{k^{(1)}} \mid \boldsymbol{A}_{k^{(1)}}\right) \subset \mathcal{M}'$ can be calculated from those obtained by all other RBs. Then this RB should be eliminated, and the subset $\mathcal{M}' \setminus \left(A_{k^{(1)}} \mid \boldsymbol{A}_{k^{(1)}}\right)$ will be considered as a possible RS. However, if after this elimination one of its RBs' data can be obtained as functions of other RBs' data, the same procedure must be repeated. Note that these procedures are based on the demand that the *necessary* condition formulated above must be fulfilled.

Functional dependencies between data on Q obtained by different RBs are none other than transformation laws from one RB to another. This means, if *all* data on Q obtained by a certain RB $\left(A_{k'} \mid \boldsymbol{A}_{k'}\right)$ can be obtained also by transformation of those obtained by, at least one, other RB $\left(A_{k''} \mid \boldsymbol{A}_{k''}\right) \in \mathcal{M}_N = \left\{\left(\forall\left(1 \le k \le N\right)\right)\left(A_k \middle| \boldsymbol{A}_k\right)\right\}$, then

$\left(A_{k'} \mid \boldsymbol{A}_{k'}\right)$ is superfluous and so must be eliminated from the considered set $\mathcal{M}_N = \left\{\left(\forall\left(1 \le k \le N\right)\right)\left(A_k \mid \boldsymbol{A}_k\right)\right\}$ of RBs.

Let us formulate one more necessary condition that a set of RBs can be used to construct RS. A countable set of RBs $\mathcal{M}_N = \left\{\left(\forall\left(1 \le k \le N\right)\right)\left(A_k \mid \boldsymbol{A}_k\right)\right\}$ forms RS, if each $\left(A_{k'} \mid \boldsymbol{A}_{k'}\right)$ obtains an information on all other RBs $\left(A_k \mid \boldsymbol{A}_k\right) \in \mathcal{M}_N$ (where $k \ne k'$), including information on their local directions and times, their positions and movement relatively all other A_k, as they were defined above. One can demand "the complete information" instead this uncertain "information". However, it would introduce unnecessary restrictions (even more severe than in the case of RS formed of two RBs) expressing the tendency to act in the spirit of the existing theories. This means, all the RBs forming a certain RS are not independent, but united *by the information exchange*.

Let us now compare it with the RS definition in the Classic Mechanics and Special Relativity. In the Classic Mechanics the RS can be defined as a ***reference body*** (***RB***) together with a 3-D co-ordinate system. The axes of this co-ordinate system are supposed to be infinitely long and infinitesimally thin absolutely rigid rods. Important fact: ***there is only one RB*** at each RS, but not a set of RBs. One can, however, to consider these rods as the continuous sets of RBs (or the countable sets of RBs of finite lengths). We shall use this interpretation below to compare RSs in different theories. Note that in the Classic Mechanics the time is considered to be universal, and so the presence of a clock on the RB is only a small technical detail.

The Special Relativity continuous to use the concept of RS, but not in a dark, as it was done in the Classic Mechanics before the appearance of the Special Relativity. The concept of RS in the Special Relativity as well as in the Classic Mechanics was then analyzed and defined. In the Special Relativity there is also one RB on each RS, as in the Classic Mechanics, but this RB is defined as ***a physical body together with an observer***. The observer is introduced explicitly and his role is essential in this theory [1, 2]. The rigidity of co-ordinate axes is not like the one in the Classic Mechanics because lengths of intervals depend on the relative velocity of two RSs. The difference between classic mechanics and special relativity becomes transparent when it is formulated in terms of the difference between transformations from one RS to another. In such terms can be formulated also the difference between Special and General Relativity: Lorentz transformations in the first one versus all possible continuum transformations in the second one [1, 2].

Thus, in the Classical Mechanics RBs forming a RS are supposed to be connected by rigid rods. In the Special Relativity the rods remain, but their lengths are changed according Lorentz transformations. In our terms this fact means, in particular, that each RB in the RS has the complete information on all other RBs of this RS. *In our consideration, when, starting from sets of independent RBs, we move to the introducing of RS, one supposes only the existence of very flexible, almost ephemeral* ***informational*** *connection between RBs forming the RS.*

Let us consider the case when *only a certain part* of data on *Q* obtained by a certain RB $\left(A_{k'} \mid \boldsymbol{A}_{k'}\right)$ can be obtained also by transformation of those obtained by one or more other RBs $\left(A_{k''} \mid \mathbf{M}_{k''}\right) \in \mathbf{M}_N = \left\{\left(\forall\left(1 \le k \le N\right)\right)\left(A_k \left|\mathbf{M}_k\right.\right)\right\}$, where $k'' \neq k'$. In this case the measurements performing by $\left(A_{k'} \mid \boldsymbol{A}_{k'}\right)$ that provide these superfluous data are to be stopped. The other way is to rearrange the measurements performed by all

$\left(A_k \mid \boldsymbol{A}_k\right) \in \mathcal{M}_N = \left\{\left(\forall\left(1 \le k \le N\right)\right)\left(A_k \left| \boldsymbol{A}_k\right.\right)\right\}$ to avoid this redundancy. Only when this way is not successful, it is to eliminate superfluous RBs. A set $\mathcal{M}_N = \left\{\left(\forall\left(1 \le k \le N\right)\right)\left(A_k \mid \boldsymbol{A}_k\right)\right\}$ of N RBs can serve as RS, if it is arranged so that there will be no redundant measurements. If such an arrangement is impossible, one or more RBs must be eliminated to obtain RS.

Let the set $\mathcal{M}_N = \left\{\left(\forall\left(1 \le k \le N\right)\right)\left(A_k \mid \boldsymbol{A}_k\right)\right\}$ can serve as RS (see above). Denote the set of data on Q obtained by $\forall\left(\left(A_{k'} \mid \boldsymbol{A}_{k'}\right) \in \mathcal{M}'\right)$ as follows:

$\left\{\left(A_{k'} \mid \boldsymbol{A}_{k'} \left| \xi_{k'} \right\rangle\right)\right\} \overset{def}{=} \mathcal{M}' \left| \Xi' \right\rangle$, where the set $\Xi' = \left\{\xi_{k'}\right\}$, $\xi_{k'}$ means the set of the data obtained by a RB $\left(A_{k'} \mid \boldsymbol{A}_{k'}\right)$. Let $\hat{\mathbf{P}}$ be the projection operator projecting the set of data $\Xi' = \left\{\xi_{k'}\right\}$ into the set obtained after the elimination of all redundant data from $\Xi' = \left\{\xi_{k'}\right\}$. Using this projection operator $\hat{\mathbf{P}}$, one can define the notion of the body Q location as $\hat{\mathbf{P}}\Xi'$.

Now it must find the lowest limit of the number of RBs (forming a RS) and their optimal configuration necessary to determine the location (including location in time) of the trial body Q ***with a given precision***. The question is not on errors of measurements themselves, but on errors arising as a consequence of a certain choice of the set of RBs. This is an essentially new element in our reasoning. Up to this place we classified the information into the necessary and superfluous ones without taking into account what precision (and of what meaning) is talking about. However, *this classification depends on the demanded precision*: less information is *necessary* when the low precision is demanded than when the high one is demanded, *i. e.*, less information refers to the type "necessary" and more to the type "superfluous" in the case of low precision, and vice versa. This means, without the indication of the demanded precision all other conditions formulated above remain uncertain.

Try to define the corresponding notion of error. Consider a subset $\mathcal{M}'' \subseteq \mathcal{M}'$. **a)** Express the ***expected*** results of measurements that should be performed by RBs $\left(A_{k''} \mid \boldsymbol{A}_{k''} \mid \xi_{k''}\rangle\right) \prec \mathcal{M}' \backslash \mathcal{M}''$, as certain functions of data obtained by all RBs forming $\mathcal{M}''$. **b)** Use the data ***really obtained*** by these observers to determine these functions. **c)** Thereupon use these already known functions to calculate the expected results of new measurements that are to be performed by RBs $\left(A_{k''} \mid \boldsymbol{A}_{k''} \mid \xi_{k''}\rangle\right) \prec \mathcal{M}' \backslash \mathcal{M}''$. Differences between calculated and observed data form the set of errors. One or few of numbers can characterize this set of errors. In the following text we shall call the set of these numbers *the error arises at a certain choice of the subset of RBs*. Perhaps, the standard deviation could be such a measure, but maybe more complicated mathematical constructions could be more relevant. It must define the maximum permitted value of error, which means that the set of RBs $\left(A_{k''} \mid \boldsymbol{A}_{k''} \mid \xi_{k''}\rangle\right) \prec \mathcal{M}' \backslash \mathcal{M}''$ cannot be used as RS, if the error is over this maximum permitted value. It is to take into consideration also that there are different sets $\mathcal{M}_j$ with the same number *j* of RBs distinguishing between themselves by different constructions γ_j, and, therefore, $\mathcal{M}_j = \left\{\mathcal{M}_{\gamma_j, j}\right\}$. Here $\mathcal{M}_{\gamma_j, j}$ denotes a set of *j* RBs of the construction γ_j. Consider now all possible sequences of subsets of RBs $\left\{\mathbf{M}_j\right\} = \mathbf{M}_1 \subset \mathbf{M}_2 \subset \mathbf{M}_3 \subset \cdots \subset \mathbf{M}_{j_{\max}}$, where each $\mathbf{M}_j = \left\{\mathbf{M}_{\gamma_j, j}\right\}$ and the corresponding sequences of errors determined as it was written. Let among these sequences of errors is a Cauchy sequence. Then, if at $j \leq j_{\min}$ the error, mentioned above, is for all $\mathcal{M}_{\gamma_j, j}$ over its maximum permitted value, but is within the limits of the permitted error at $j > j_{\min}$, sets of RBs $\mathcal{M}_{j > j_{\min}}$ and only such sets can be used as RSs.

The question arises: whether *in all cases* at least one sequence $\{\mathcal{M}_j\} = \mathcal{M}_1 \subset \mathcal{M}_2 \subset \mathcal{M}_3 \subset \cdots \subset \mathcal{M}_{j_{\max}}$ exists such that the corresponding sequence of errors would be Cauchy sequence? This "in all cases" means, in particular, "in any region of the Universe". If in a certain region of the Universe no such sequence exists, to what consequences this fact can lead? Moreover, whether it is possible to define the concept of region of the Universe when the condition mentioned above is not satisfied? Indeed, this concept would be different for different $\mathcal{M}_j$.

The approach described above to the determination of errors is practically not realizable. We shall consider here another approach. Consider now a sequence $(\forall(1 \leq j < \infty))[\mathcal{M}_j]$. Let the measurements performed by the sets $\mathcal{M}_{j'}$ and $\mathcal{M}_{j'+1}$ of RBs provided data $\Xi_{j',\gamma_{j'}}$ and $\Xi_{j'+1,\gamma_{j'+1}}$, correspondingly. Define the error as a function or functional $\delta_{j',\gamma_{j'},j'+1,\gamma_{j'+1}}[\Xi_{j',\gamma_{j'}}, \Xi_{j'+1,\gamma_{j'+1}}]$ that characterizes the change of data obtained by measurements made over a body Q, when one replaces the system of j RBs by the system of j+1 RBs. Denote $\delta^{(\min)}_{j',\gamma^{(0)}_{j'},j'+1,\gamma^{(0)}_{j'+1}}$ the minimum of $\delta_{j',\gamma_{j'},j'+1,\gamma_{j'+1}}$ at a given j' obtained by variation of $\gamma_{j'}$ and $\gamma_{j'+1}$ and reached at their values marked by the upper index (0). Consider now such pairs at different increasing values of j'. If at $j' > j_{\min}$ the sequence $\delta^{(\min)}_{j',\gamma^{(0)}_{j'},j'+1,\gamma^{(0)}_{j'+1}}$ is a Cauchy sequence, the corresponding

$\mathcal{M}_{j',\gamma_{j'}^{(0)}}$ can be used as RSs. The choice of $j' > j_{\min}$ is determined by the desirable value of $\delta_{j',\gamma_{j'},j'+1,\gamma_{j'+1}}$.

Thus, a **RS is a set** $\mathcal{M}_{j',\gamma_{j'}^{(0)}}$ **of** $j' > j_{\min}$ **RBs 1) united by the convenient information exchange and 2) satisfying the condition that for all measurements made with the purpose to get information on macroscopic physical bodies a certain value** ***j'₀ >0*** **exists such that the sequence** $\delta^{(\min)}_{j',\gamma_{j'}^{(0)},j'+1,\gamma_{j'+1}^{(0)}}$ **is a Cauchy sequence for all** $j' > j_0$**.**

It must be taken into account that, possibly, for no set $\mathcal{M}_{j',\gamma_{j'}^{(0)}}$ this condition is satisfied for measurements made on certain macroscopic physical bodies in the Universe. This means, ***a set of such macroscopic physical bodies may exist that for them no set*** $\mathcal{M}_{j',\gamma_{j'}^{(0)}}$ ***can serve as reference system. This lack of universality of RSs is contrary to the commonly accepted in the Classic Mechanics, Special and General Relativity, where each RS is valid for all physical bodies in the Universe***. ***It may lead to far-reaching consequences, in particular, in the cosmology***: what, for example, would be observed effects, when no RS will be valid for the description of a certain set of physical bodies? Maybe in this case we shall see a kind of hole in the Universe? Another case is also possible. Let some RSs (1st group of RSs) are valid for a certain set of macroscopic physical bodies, while some other RSs (2nd group) are not valid for them. Then it is expected that the transformation from one of RS belonging to the 1st group to another belonging to the 2nd group would lead to the loss of information on the existence itself of these bodies. Thus, it must choose very careful classes of RSs so that within each such a class transformations between different RSs should be permitted.

5. CO-ORDINATE SYSTEMS

We have defined the concept of the RS. This definition does not contain the condition that a co-ordinate system must be attributed to each RS $\mathcal{M}_{j',\gamma_{j'}^{(0)}}$. Let us try to find a way to introduce co-ordinate systems to RSs, and to define such a class of RSs, for which it is principally possible. This class of RSs would be fit for Classic Mechanics, Special and General Relativity, where RS without fail includes a co-ordinate system [1 - 9]. Pay the attention that co-ordinate system is not figured in the definition [1] of the RS. This means, the definition [1] of the RS really refers to theories more general than the General Relativity because the mathematical formalism of the General Relativity is based on 4-D Riemannian geometry that demands the co-ordinate systems.

As it was found above, each RB can measure distance between itself and a macroscopic physical body Q , direction to this body in RB's local co-ordinate system, and the time by a local clock. If this RB belongs to a certain RS, the obtained data are forwarded to *some* other RBs that also belong to this RS. Pay attention: to ***some*** other, but *not, without fail, to all* other RBs forming the RS. It is a very important point. The case *without fail* corresponds only to the Classical Mechanics and Special Relativity. Indeed, in these theories the co-ordinate axes are considered as infinitesimally thin solid rods. A perturbation is transferred throughout each rod with the infinitely high velocity (Classical Mechanics) or with the velocity of light c (Special Relativity). Principally this transfer of a perturbation can be expressed in terms of the information transfer. In our consideration the more general case is considered when is possible also that only a part, maybe a small part of the RS is involved into this information transfer. In other words, only a certain part of RBs of the considered RS "knows" on the body Q and participates in the corresponding information processing, while in Classical Mechanics and Special Relativity it is known to the whole RS that participates as a whole in the information processing.

Denote $\mathcal{M}'_\delta \subseteq \mathcal{M}'$ the subset of these involved RBs that in certain cases may coincide with the whole RS. $\mathcal{M}'_\delta \subseteq \mathcal{M}'$ processes the data obtained by RBs forming it, and eliminates the redundant information. Then, according the written above, the location of Q determined by $\mathcal{M}'_\delta \subseteq \mathcal{M}'$ is $\hat{\mathrm{P}}_\delta \Xi'_\delta$, where $\hat{\mathrm{P}}_\delta$ is the projection operator (defined above) for the considered part of RS $\mathcal{M}'_\delta \subseteq \mathcal{M}'$ and Ξ'_δ are the data on Q obtained by all RBs forming $\mathcal{M}'_\delta \subseteq \mathcal{M}'$. Let us **suppose** that, at least, one way exists allowing us to represent all $\hat{\mathbf{P}}_\delta \Xi'_\delta$ as linear combinations $\hat{\mathbf{P}}_\delta \Xi'_\delta = \left\{ \sum_{k'=1}^{n} \alpha_{kk'} f_{k'} \right\}$, where f_k are n numbers (k=1,2,…,n) and $\alpha_{kk'}$ are coefficients of linear combination of these numbers. ***Really this means that all the information, which cannot be represented in this form, is neglected.*** It returns us to the consideration of a material point and sets of them, and limits the consideration with very simple objects. For example, biological systems and automata cannot be considered in the framework of any theory that accepted this limitation.

These n numbers can be interpreted geometrically as co-ordinates in a certain n-dimensional space. The choice of n and of the geometry is not considered in the present work. For example, it is done in the General Relativity and Unified Field theories on the grounds of physical and mathematical reasons. In the General Relativity n=4, and the geometry is chosen to be Rimannien, while in Unified Field theories one finds n=4 and n=5, and different types of geometry. There are theories using 7 and more dimensions also for the description of only mechanical properties of the system without taking into account its internal states. As it can be seen, ***our consideration shows where the choice of dimension of the used space is from***. It opens a ***natural way*** to construct theories of different dimensions and to take into account internal states of a system.

The concept of space is defined here simply as a mathematical formalism for the interpretation of results of measurements made by an RS consisting of a finite or countable set of RBs. The physical picture was that there is only a finite or countable set of physical bodies (a part of them are RBs) in the Universe and the question about what is between them has no meaning. This picture was not changed. However, now in this picture physical bodies, including RBs, are placed to the space - time continuum. Then the next step is as follows. One can "forget" how this concept of space was defined really, "overturn" the logic and, after this logical leap, consider the space as the primary fundamental concept. Now the concept of light can be introduced and measurements would be made by *light* signals. In the framework of this picture the limitation (see INTRODUCTION) can be removed that only large macroscopic bodies are considered. The last fact makes possible the application of the theory to such small objects as biological systems are.

As it was written in Sec. 2, the information on a complex object with a great number of internal states when it is considered simply as a structureless physical body, *i. e.*, its mechanics, and the one on its internal states form an inseparable mixture. In the case when a space R is defined and used for the representation of the information on this object, its subspace R_1, probably ***low dimensional***, serves to represent the information of the 1st type (mechanics). At the same time, if the considered object possesses of a large number of internal states (*e. g.*, biological systems, stars), the information of the 2nd type is so rich that can be approximately represented in a ***high dimensional*** Riemannian space (see, for example, [8]). Perhaps, better representation can be achieved in Hilbert space. However, even Hilbert space may be not enough fit for this purpose, and then it is to use normalized or metric spaces. Thus, the construction of theories with high or infinite number of space dimensions, using metric or normalized spaces is necessary to represent the information on object internal states. However, this is a very difficult task because it must find the transformation laws of the information on these internal states. Note that in theories using the abovementioned Hilbert, normalized or metric spaces the picture of the World will be more like the quantum physics picture than the Einstein one.

According to the written in Sec. 2, the information of the 2*nd* type cannot be referred to a certain moment of time, but to a certain local time interval of the duration τ_2. It leads to uncertainty of this information itself. Therefore, states of a measured system will be represented not by points of the space-time of chosen type and number of dimensions, but by regions such that projection of each of them to the axe of time will have the length $\tau_2 >> \tau_1$.

The projections of a set of such regions representing the information on the considered object into one of these subspaces R_1 and R_2 describe the mechanical properties of the considered object or its internal properties, correspondingly. It is important that, generally speaking, such projections into subspace R_1 also will be regions, but not points because states of the system are represented by regions in the space R, as it was explained above. It is a manifestation of the connection between mechanical and internal properties of the object, in other words, that they are inseparable. These properties can be approximately considered separately only if the connection between them is negligible.

6. CONCLUSIONS

In the present paper we tried to approach to the problem of the RS concept definition in the General Relativity and beyond it starting from the analysis of L. Landau and E. Lifshitz definition [1] cited in the INTRODUCTION. We began from the concept of macroscopic RB and postulated the following:

1. there are macroscopic physical bodies in the Universe;
2. on some of them are observers (such a physical body A with the observer $\boldsymbol{A}$ is called RB ($(A \mid \boldsymbol{A})$);
3. the observer $\boldsymbol{A}$ can measure the local time, *i. e.*, the time valid only within his close vicinity on the physical body A surface;
4. the observer $\boldsymbol{A}$ can measure local direction valid only within his close vicinity on the physical body A surface;
5. there exist signals of a certain nature and type such that each signal can be emitted and detected (accepted), and within the close vicinity of the observer signal has the direction;
6. each observer can send and receive signals.

Contrary to the made in other theories, such as Classic Mechanics, Special and General Relativity, the existence of space-time in the whole Universe and even in its parts is not postulated, and in the general case it does not exist. Even locally, in the close vicinity of the observer on the macroscopic physical body surface the existence of only local time and direction is postulated. It is important also that the existence of the light is not and cannot be postulated. This is in complete distinct from the accepted in the above mentioned theories, where namely light signals are used as the only mean of the information transfer between observers (because the velocity of light is the largest existing velocity). However, electromagnetic waves propagate in a space, while in our consideration space (- time) exists only in special particular cases.

The second step was the consideration of sets of RBs aimed to formulate the conditions, under which a set of RBs becomes RS. One of these conditions is the information connection between RBs forming RS. It was emphasized that this type of connection is much more flexible than the rigid connection between parts of RS in Classic Mechanics and Special Relativity. In our consideration, in distinct to the Classic Mechanics and Special Relativity, an information about a macroscopic physical body obtained by one RB of RS may be transferred to a few of other RBs of the same RS, but not, without fail, to all its RBs. From these few of RBs the information may be (or may not be) forwarded to a number of other RBs of the same RS.

Another important condition is that the RS must provide more information on a macroscopic physical body than all RBs forming this RS, if they are considered independent. On the other hand, the information that was provided by these independent RBs might contain a redundant information. It was considered how to eliminate this redundancy at the creation of RS from these RBs.

The definition of the concept of RS principally allows a possibility of the existence of such macroscopic physical bodies and their sets that can be detected and studied by ***no***

RS. If this possibility is realized it may lead, for example, to the existence of a specific kind of holes in the Universe.

The conditions were found when the space - time of different dimensions can be defined, in particular, the 4-D World of the General Relativity, 5-D and higher dimensions Worlds of different Unified Field Theories. These conditions show where the dimension of space - time is from.

A special attention was paid to the problem of the representation of the information on internal states of complex systems such as biological ones, automata with great number of states, interior of stars *etc*. and its transfer from one RB or RS to another. At this consideration was taken into account the time necessary to the observer for decoding and interpretation of such an information. It was concluded that in the case when the concept of space - time is defined, the representation of this type of information demands the use of high dimensional Riemannian space, or, better, Hilbert space, or maybe normalized or metric space. It was shown that the existence of a great number of internal states influences the mechanics of the considered object, *i. e.*, its behavior as a structureless physical body.

It would be interesting to consider another approach when the behavior of an object as a structureless physical body is represented inside a low dimensional brane or branes (see, for example, [10 - 15]) *imbedded into a high dimensional or Hilbert space serving for the representation of the information on internal states of the object.*

REFERENCES

1. L. Landau & E. Lifshitz, ***The Classical Theory of Field***. Addison-Wesley Press, Inc., Cambridge, Mass., 1951

2. P.G. Bergmann, ***Introduction to the Theory of Relativity.*** Prentice-Hall, New York, 1942

3. A. S. Eddington, ***Space, Time and Gravitation***. Cambridge University Press, 1920

4. A. Lichnerowitz, ***Théories Rélativistes de la Gravitation et de l'Électromagnétisme***. Paris, Masson, 1955

5. C. W. Kilmister, ***General Theory of Relativity***. Oxford, Pergamon, 1973

6. A. Papapetrou, ***Lectures on General Relativity***. Dordrecht-Holland, Reidel, 1974

7. V. Hlavaty & B. Hoffmann, ***Perspectives in Geometry and Relativity***. Bloomington, Indiana University Press, 1966

8. A. Z. Petrov, ***Einstein Spaces***. Oxford, Pergamon, 1969

9. V. Fock, ***The Theory of Space, Time and Gravitation.*** Oxford, Pergamon, 1964

10. D. Ida, JHEP 09 (2000) 014

11. B. Carter, Int. J. Theor. Phys. **40**, 2099 (2001)

12. D. Langlois, M. Rodrígez-Martínez, Phys. Rev. D**64**, 123507 (2001)

13. C. Germani, R. Maartens, Phys. Rev. D**64**, 124010 (2001)

14. R. A. Battye, B. Carter, A. Mennim, J. - P. Uzan, Phys. Rev. D**64**, 124007 (2001)

15. V. A. Rubakov, Physics - Uspekhi **44**, 871 (2001)

ARTICLE 3

Set partition based definition of elementary events as the probability and information theories common basis

A. Ya. Temkin

January 26, 2007

Department of Interdisciplinary Studies, Faculty of Engineering, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel, e-mail: temkin@eng.tau.ac.il; fax: 972-3 6410189; tel.: 972-3 6408176

Abstract

The set theoretical basis of the theory of information is constructed on the grounds of the one of the theory of probability by the *set of events partitioning* and its use for *elementary events of different orders* definition.

In the present paper are considered, in particular, *sets able to treat the information*, *i. e.*, to PROCESS, to GENERATE, and to INTERPRET it, to ACCEPT THE INFORMATION came from outside sources and to SEND THE INFORMATION to outside targets. It is the general approach that could be applied to the information treatment by computer, human brain, biological systems, in general, *etc.*.

In the present paper Bertrand Russell's theory of types is used for the *classification of the information* into different types that, in its turn, is used for *classification of sets* (treating it) into different types and the setting up (on this grounds) the hierarchy among sets. The same thing is done here also by another way, on the grounds of the consideration of that what of aboutmentioned functions of the information treatment are executed by different sets. In distinct from the first approach based on the Russell's theory of types, the second approach is not limited with the condition that the information must be expressed in a language.

The *information content* and *the information value* concepts are defined mathematically.

Keywords: sets, probability, elementary event, information, Russell's types, hierarchy (information), hierarchy (sets), scaling

MATHEMATICAL SUBJECT CLASSIFICATION: 03E04, 03E75, 03B15, 60A05, 94A15

1 INTRODUCTION

Systems able to treat, to generate, to obtain and to send the information are customarily macroscopic devices. However, it is evident that each such a system is a set of elements that, in their turns, may be macroscopic or microscopic. The most complicated are systems able to interpret the information. Usually is supposed that the interpretation is being done by a human being. For example, when the information contains physical laws, the purpose of its interpretation is to extract them. But if we consider systems treating the information (including its interpretation) as sets (in mathematical meaning of this notion), it suggests an idea to check what are the weakest conditions that a set is able to execute the abovementioned functions, in other words, what is the most "primitive" set able to do it.

An important purpose of the present work is to use properties of the information treatment by different well-ordered sets to classify them into different type and, thereupon to establish hierarchy relations between them. We use the way to do it based on the consideration of that what functions of the information treatment are executed by the considered set. One approach is based on the application of Bertrand Russell's theory of types [15] to the information treated by different sets, while the second one based on direct consideration of the information treatment by different sets. The second approach is fit for the treatment of the information expressed in a language (formal or natural) as well as of the one that cannot be expressed in any language (*cf.* [18] and [9]), while the theory of type is fit only for the information expressed in a language.

The information amount is defined by use of a priori and a posteriori probabilities [17]. For this purpose the concept of elementary event in the theory of probability is formulated so that a priori and a posteriori probabilities could be defined.

The probability theory is based on the definition of the space of elementary events (Ω, A, P), where $\Omega = \{\varepsilon\}$, ε is an elementary event, A is a σ - algebra and $P = P(A)$ is the probability measure such that $P(\Omega) = 1$ (see, for example, [4] , [13], [12], [11], [8], [10]). This definition does not demand that an elementary event ε would be without fail a *change* occurring within the considered set. In the present paper we shall define such types of elementary events that are elementary **changes** occurring within a given set of elements. They will be called *higher* (than zero) *order* elementary events, while ε will be called *zero-order* elementary event.

2 DEFINITION OF ELEMENTARY EVENT

Let there is a well-ordered set $\Phi = \{\varphi\} \neq \emptyset$ of elements φ(see, for example, [7],[14], [5], [6]). The set Φ can be finite or countable, or continuum. Let $\exists\,[\Theta \subseteq \Phi]\,(\Theta \neq \emptyset \wedge \mu\,(\Theta) \neq 0 \wedge \iota\,(\Theta) > 1)$, where $\iota\,(\Theta)$is the number of elements of the subset Θ, and $\mu\,(\Theta)$ is its measure. The subset Θ also can be finite or countable, or continuum.

The PARTITION is defined as follows [2]: A partition ϑ of a nonempty set Φ is a set of subsets of Φ, satisfying all of the following:

1. Each set in ϑ is nonempty.

2. For any x, if $x \in \Phi$, there is $C \subseteq \vartheta$ such that $x \in C$.

3.$(\forall\,(C \wedge D))\,[C \subseteq \vartheta \wedge D \subseteq \vartheta \wedge C \neq D] \Rightarrow C \cap D = \emptyset$.

The sets in ϑ are called *cells*.

Let

$$\Lambda = \{\forall\Theta, \left(\forall\Theta, \forall\left(\Theta \neq \Theta' \subseteq \Phi\right)\right)\,[\cup_{\forall\Theta\subseteq\Phi}\Theta]\,G\}, \tag{1}$$

where

$$G = [\Theta \neq \emptyset \wedge \mu\,(\Theta) \neq 0 \wedge \iota\,(\Theta) > 1] \wedge \left[\Theta \cap \Theta' = \emptyset \vee \mu\left(\Theta \cap \Theta'\right) = 0\right], \tag{2}$$

is a set of all *cells* $\Theta \subseteq \Phi$ *at a given partition of* Φ *into these subsets. Each cell is a* well-ordered set because it is a subset of a well-ordered set. We shall call such a partition ϑ also CONFIGURATION ϑ of the set Φ and shall write Λ_ϑ and $\Phi_\vartheta\,(\forall\vartheta)\,[\Phi_\vartheta \equiv \Phi]$correspondingly, to indicate that the both Λ and Φ are represented by a certain partition ϑ . Thus, we have the set $\{(\forall\vartheta)\,[\Phi_\vartheta]\}$.

Elements φ are supposed to be 1) identical and 2) such that each of them is a set containing one and only one element. However, in some applications of the proposed theory the second and, therefore, the first conditions could be violated. It is possible if there are some conditions issued not from the mathematical abstract set theory, but from outside of this field, *e. g.*, from the nature, for example, from physical interactions between elements φ leading to the creation of bound states, spin states *etc.*. One way to treat such cases without to introduce explicitly into consideration the "nature", *e. g.*, physics, is to define that elements φ are not all identical, but can be different. For example, the set Φ can be defined as $(\forall\,(s \wedge \Phi_s))\,[(s \wedge s') \in \mathbb{N}, (s \wedge s') \in [1, s_{max}] \subset \mathbb{N}, s' \neq s]\,(\Phi = \bigcup_{s=1}^{s_{max}} \Phi_s)$, where $\forall\,(s \wedge s')\,(\Phi_s \bigcap \Phi_{s'} = \emptyset)$and, therefore, each element should be $\varphi_s \in \Phi_s \subset \Phi$, which breaks the identity of all elements. Then at the partitioning of Φ each its subset (cell) can contain elements with different s: $\Theta = \bigcup_{s=1}^{s_{max}} \Theta_s\,(\alpha_s \mid \Theta)$, which means that the set of coefficients $\{\alpha_s\}$ is determined for each Θ separately. A delicate mathematical problem is connected with the fact that each element must have an attached label s indicating its origine.

Let $\exists L = \{(\forall\vartheta)\,\Lambda_\vartheta\}$ that can be finite or countable set, or continuum. In the latter case some times we can suppose that the set $L = \{(\forall\vartheta)\,\Lambda_\vartheta\}$ is a metric space, in which the distance $(\forall\,(\vartheta, \vartheta'))\,\rho\,(\Lambda_\vartheta, \Lambda_{\vartheta'}) \geq 0$ is defined so that $\rho\left(\Lambda_{\vartheta},\Lambda_{\vartheta'}\right)=0\Leftrightarrow\vartheta=\vartheta'$.

Then at each partition ϑ the set Φ_{ϑ} can be represented as the following sum

$$\Phi_{\vartheta}=\left(\forall\Theta,\forall\left(\Theta\neq\Theta'\right)\right)G\left(\cup_{\forall\Theta\in\Lambda_{\vartheta}}\Theta\right) \tag{3}$$

and one can rewrite

$$\Phi_{\vartheta}=\left(\forall\left(\Theta|\vartheta\right),\forall\left(\left(\Theta|\vartheta\right)\neq\left(\Theta'|\vartheta\right)\right)\right)G_{\vartheta}\cup_{\forall\left(\Theta|\vartheta\right)\in\Lambda_{\vartheta}}\left(\Theta|\vartheta\right), \tag{4}$$

where G_{ϑ} is obtained from G defined by Eqn. (2) by replacing each Θ and Θ' to $\Theta|\vartheta$ and $\Theta'|\vartheta$ correspondingly.$\{(\forall\vartheta)\,\Lambda_{\vartheta}\}$ can be a finite set, or a countable set, or the continuum, which in certain cases could be converted to a metric space with a given definition of the distance

$$\rho\left(\left\{\left(\forall\vartheta'\right)\Lambda_{\vartheta'}\right\},\left\{\left(\forall\vartheta''\right)\Lambda_{\vartheta"}\right\}\right)>0\vee\rho\left(\vartheta',\vartheta''\right)=0\Leftrightarrow\vartheta'=\vartheta'' \tag{5}$$

Define an *elementary event* as the transition from a configuration ϑ to another configuration ϑ', *i. e.*, $\vartheta\rightarrow\vartheta'$transition, which can be represented as follows $\left\langle\vartheta'\left|\mathbf{T}\right|\vartheta\right\rangle$, or by the equation

$$\mathbf{T}_{\vartheta'\vartheta}\mathit{\Phi}_{\vartheta}{=}^{def}\ \mathit{\Phi}_{\vartheta'} \tag{6}$$

where Φ_{ϑ} is defined by Eqn. (4), and $\mathbf{T}$ is an operator, matrix elements of which $\mathbf{T}_{\vartheta'\vartheta}=\left\langle\vartheta'\left|\mathbf{T}\right|\vartheta\right\rangle$ are elementary events.

Define the space of elementary events $((\Omega_1|\Phi), A_1, P_1)$, where

$$\Omega_1 = \left\{ \left(\forall\vartheta_1, \left(\forall\left(\vartheta_1' \neq \vartheta_1\right)\right)\right) \left\langle \vartheta_1' \left|\mathbf{T}\right| \vartheta_1 \right\rangle \right\}, \tag{7}$$

$$(\Omega_1|\Phi_{\vartheta_1}) =^{def} \left(\left\{ \forall\vartheta_1, \left(\forall\left(\vartheta_1' \neq \vartheta_1\right)\right) \left\langle \vartheta_1' \left|\mathbf{T}\right| \vartheta_1 \right\rangle |\Lambda_{\vartheta_1} \right\} \right), \tag{8}$$

$\vartheta_1 \in [1, 2, \ldots, n] \subset \mathbb{N}$, $\mathbb{N}$ denotes the set of all natural numbers, $A_1$is the corresponding σ-algebra, $P_1 (a_1 \subseteq A_1)$ is the corresponding probability measure and $P_1 (\Omega_1|\Phi) = 1$ (see below). The set of partitions can be an infinite countable one ($n \to \infty$) or the continuum.

3 INFORMATION TREATMENT BY SUBSETS OF THE SET Φ

The proposed approach allows one to define the Shannon information amount corresponding to an event. Consider a set Φ closed, but not isolated. This means, an exchange of the information between Φ and other sets exists, but not an exchange of elements of sets. Let

$$\exists\left(\forall\nu \in [1, \nu_{inf,max}] \subset \mathbb{N}\right) \left[\Theta_{inf,\nu} \subseteq \Phi\right] \tag{9}$$

that are able to detect and process the effect upon the set Φ produced by the information penetrated into the set Φ from another set $\Phi^\star$ from $(\forall\Phi^\star)\left[\Phi^\star \bigcap \Phi = \emptyset \vee \mu\left(\Phi^\star \bigcap \Phi\right) = 0\right]$ (the number of such sets $\Phi^\star$ could be >1). This effect is expressed in a change of the configuration of Φ , *i. e.*,of the distribution with respect to ϑ. We shall try to find a way that processes arising by spontaneous changes of configuration inside the set Φ could be represented in terms of the information.

Let each subset $(\forall \nu \in [1, \nu_{inf,max}] \subset \mathbb{N}) [\Theta_{inf,\nu} \subseteq \Phi]$ is able to generate information and to treat the information generated by it as well as the information coming in from different sets $\Phi^\star [\Phi^\star \cap \Phi = \emptyset \vee \mu (\Phi^\star \cap \Phi) = 0]$ or from other than $\Theta_{inf,\nu}$ subsets of Φ. The latter means that subsets

$$(\forall \nu \in [1, \nu_{inf,max}] \subset \mathbb{N}) [\Theta_{inf,\nu} \subseteq \Phi]$$

are able to *express in terms of the information* changes of configurations occurring within the set Φ. This ability should be based on certains properties of each set $(\forall \nu \in [1, \nu_{inf,max}] \subset \mathbb{N}) [\Theta_{inf,\nu} \subseteq \Phi]$. What could be these properties from? One option is that they are generated by some connections existing between elements of Φ or between its subsets. We shall call, for short, such connections *interactions*. Another option is a kind of the *information connections* between subsets of the set Φ. For $\Theta_{inf,v}$ one can write

$$(\Lambda_{inf,\nu} | \Theta_{inf,\nu}) = \left\{ \forall \phi, \left(\forall \phi, \forall \left(\phi, \phi' \right) \right) G_\phi \left[\cup_{\forall \phi \subseteq \Theta_{inf,\nu}} = \Theta_{inf,\nu} \right] \right\}, \qquad (10)$$

where Θ is replaced to ϕ and ϑ is replaced to Υ indicating configurations of the subset $\Theta_{inf,\nu}$ instead those of the set Φ. Then instead Eqn.(4) one obtains under all conditions accompanying Eqn. (3):

$$\Theta_{inf,\nu} = \left(\forall\left(\phi|\Upsilon\right), \forall\left(\left(\phi|\Upsilon\right) \neq \left(\phi'|\Upsilon\right)\right)\right) G_{\Upsilon} \cup_{\forall(\phi|\Upsilon)\in\Lambda_{\gamma}} \left(\phi|\Upsilon\right), \qquad (11)$$

where $\forall\phi \subseteq \Theta_{inf,\nu}$, $\Lambda_{\Upsilon} =^{def} (\Lambda_{inf,\nu}|\Upsilon)$. However, in distinct from the considered case of the set Φ representation by Eqn. (3), the subset $\Theta_{inf,\nu}$ possesses not only the representation (11), but also representations including intersections of the subset $\Theta_{inf,\nu}$ with subsets Θ appearing in Eqn. (3). The said can be expressed as follows:

$$\Theta_r =^{def} \Theta_{inf,\nu} \setminus \cup_{\forall u} \left(\Theta_{inf,\nu} \bigcap_{u} \Theta_u \subseteq \Lambda_{\vartheta}, u \in \mathbb{N}\right), \qquad (12)$$

$$\Theta_{inf,\nu} = \Theta_r \cup \cup_{\forall u} \left(\Theta_{inf,\nu} \bigcap_{u} \Theta_u \subseteq \Lambda_{\vartheta}, u \in \mathbb{N}\right), \qquad (13)$$

or

$$\Theta_{inf,\nu} = \left(\forall\phi, \forall\left(\phi, \phi'\right)\right) K\left(\phi, \phi'\right) \cup_{\forall\phi\subseteq\Lambda_{(inf,\nu|\Theta_r)}} \phi, \qquad (14)$$

were $K = K_1 \cup K_2$,

$$K_1\left(\phi, \phi'\right) = \left[\phi \neq \emptyset \wedge \mu\left(\phi\right) \neq 0 \wedge \iota\left(\phi\right) > 1, \phi \neq \phi', \phi \cap \phi' = \emptyset \vee \mu\left(\phi \cap \phi'\right) = 0\right], \qquad (15)$$

and

$$K_2 = \left(\Theta_{inf,\nu} \bigcap_u \Theta_u \subseteq \Lambda_\vartheta, u \in \mathbb{N} \right) \tag{16}$$

4 INFORMATION AMOUNT CARRIED BY AN EVENT

There are different definitions of the information amount in different versions of the information theory. We shall define the information amount using the *apriori* and *aposteriori* probabilities of the event, demanding the existence of time because they correspond to two different time moments $t_1 \prec t_2$. The definitions of *apriori* and *aposteriori* probabilities could be done by a more general way than the one based on the use of the time. For this purpose we shall employ here the approach introduced in our previous work [18] (Ch. 1), where instead of time one uses a well-ordered set $(H = \{h\})\,[H \cap \{\vartheta\} = \emptyset]$ of elements h possessing the power of continuum $\aleph$. Note that in a particular case it may be $H = t$, where t is the time. Following [18](Ch. 1), in the general case one establishes homomorphism between a subset $Y \subseteq \{\vartheta\}$ and a subset $H_Y \subseteq H$.

Consider now a particular case when $Y = \vartheta \in \{\vartheta\}$, *i. e.*, Y contains only one element. Then one can write $H_Y = H_\vartheta$. Let be $H_\vartheta \prec H_{\vartheta'}$. Then **DEFINE** *à priori* and *à posteriori* probabilities of an event $\langle \vartheta' \,|\mathbf{T}|\, \vartheta \rangle$ as follows:

$$\mathbf{P}_{a\,pr} =^{def} \mathbf{P}\left(\vartheta | H_\vartheta\right), \, \mathbf{P}_{a\,post} =^{def} \mathbf{P}\left(\vartheta' | H_{\vartheta'}\right) \tag{17}$$

This means, these are probabilities of the configurations ϑ and ϑ' labeled by H_ϑ and $H_{\vartheta'}$, correspondingly, exactly as by corresponding time moments before and after the transition $\vartheta \to \vartheta'$.

We have used above the probability $\mathbf{P}(\vartheta)$ of a configuration ϑ, which is not the probability of a configuration *change*. We shall define $\mathbf{P}(\vartheta)$ as the probability to find a certain configuration ϑ within the set $\{\vartheta\}$, in other words, putting $\Omega = \{\vartheta\}$in space of elementary events $((\Omega \mid \Phi), A, P)$ (the result we shall call space of zero-order elementary events). In a particular case when $\{\vartheta\}$ is a finite set of elements possessing equal probabilities, $\mathbf{P}(\vartheta) = \frac{1}{\mathcal{N}}$, where $\mathcal{N}$ is the total number of elements in $\{\vartheta\}$. In other cases the definition of $\mathbf{P}(\vartheta)$ must be relevant to the properties of the set $\{\vartheta\}$. In particular, if $\{\vartheta\}$ is the continuum, probably it is to define firstly the probability density and thereupon to use it to define the probability of a subset $Y \subseteq \{\vartheta\}$.

Now we can define the information amount $I(\langle\vartheta' |\mathbf{T}| \vartheta\rangle)$ corresponding to the transition $\vartheta \to \vartheta'$, *i. e.*, to the event $\langle\vartheta' |\mathbf{T}| \vartheta\rangle$, as follows:

$$I\left(\left\langle\vartheta' |\mathbf{T}| \vartheta\right\rangle\right) =^{def} I_{\vartheta'\vartheta} =^{def} K\, log_2 \frac{\mathbf{P}_{a\,post}}{\mathbf{P}_{a\,pr}} = K\, log_2 \frac{\mathbf{P}_{\vartheta'}}{\mathbf{P}_{\vartheta}}, \tag{18}$$

where $K = const.$

Let us consider now more general case of the transition $Y \to Y'$ $[(Y \wedge Y') \subset \{\vartheta\}]$, which corresponds to the event $\langle Y' |\mathbf{T}| Y\rangle =^{def} \mathbf{T}_{YY'}$. Then for this case one obtains formulas like (17) and (18) as follows:

$$\mathbf{P}_{a\,pr} =^{def} \mathbf{P}(Y \mid H_Y), \mathbf{P}_{a\,post} =^{def} \mathbf{P}\left(Y' \mid H_{Y'}\right), \tag{19}$$

$$I\left(Y'\left|\mathbf{T}\right|Y\right) =^{def} I_{Y'Y} =^{def} K\,log_2\frac{\mathbf{P}_{a\,post}}{\mathbf{P}_{a\,pr}} = K\,log_2\frac{\mathbf{P}_{Y'}}{\mathbf{P}_{Y}} \qquad (20)$$

Note that is not clear whether in the general case $\mathbf{T}_{Y'Y}$ can be expressed in terms of $\mathbf{T}_{\vartheta'\vartheta}$.

Consider now the case when one or more kinds of interaction (see above) exist between elements and subsets of the considered subset $\Theta_{inf,\nu}$. It may provoke changes of configurations. In particular, it may provoke such changes following the first detected transition $\vartheta \to \vartheta'$ or $Y \to Y'$. The existence of such an interaction generates laws determining how a change within the considered subset $\Theta_{inf,\nu}$ initiates the next ones. How it could be connected with configuration changes? While $\Theta_{inf,\nu}$ remains an abstract mathematical set, configurations are also of the abstract character and no one of them has advantages in comparison with the others. But the interaction changes profoundly this situation making different configurations unequal, which means that at least the probabilities $(\mathbf{P} \mid \vartheta) =^{def} \mathbf{P}_{\vartheta}$ (or $(\mathbf{P} \mid Y) =^{def} \mathbf{P}_{Y}$) will be different for different configurations (generally speaking, other differences also may exist).

Thus, our definition of the concept "ELEMENTARY EVENT" as representing elementary change allows one to express changes going on within a set in terms of the information amount and, therefore, possibly, under certain conditions, makes the set able to create and process the information (*cf.* [19]).

5 INFORMATION TREATMENT

If one wants to approach to the logical or another development of the information processing by a set, then, in view of the said in Section 5, it is $\forall\Theta_{inf,\nu}$ to 1) introduce or find laws according of which a change of the configuration causes the next ones, and 2) find or define how these changes are connected with the AMOUNT, CONTENT and VALUE of the information. In the following text we shall denote $\phi \subseteq \Theta_{inf,\nu}$ (indices inf, ν at ϕ will be omitted) and $\Upsilon_{inf,\nu}$, or simply Υ the configuration (*cf.* Sec. 2).

Let $\exists\left(H_\Upsilon \prec H_{\Upsilon'}\right)\left[\left(\Upsilon \mid H_\Upsilon\right), \left(\Upsilon' \mid H_{\Upsilon'}\right)\right]$, no matter why and how it was created. The task is now to find

$$\left(H_{\Upsilon'} \prec H_{\Upsilon''}, \nexists\left[H_{\Upsilon^\star}\right]\left(H_{\Upsilon'} \prec H_{\Upsilon^\star} \prec H_{\Upsilon''}\right)\right)\left[\left(\Upsilon'' \mid H_{\Upsilon''}\right)\right],$$

i.e., to find how this configuration change propagates within the set of configurations.

Let within the given subset $\Theta_{inf,\nu}$ the law exists

$$\left(\Upsilon'' \mid H_{\Upsilon''}\right) = \Psi\left(\left(\left(\Upsilon \mid H_\Upsilon\right), \left(\Upsilon' \mid H_{\Upsilon'}\right)\right) \mid H_{\Upsilon''}\right),\ H_{\Upsilon'} = H_\Upsilon, \qquad (21)$$

where Ψ is a given function of configuration perturbation created by the transition $\Upsilon \to \Upsilon'$. Let us consider some simple cases of Eqn. (21).

Let us consider the case when $\{\Upsilon\}$ is the continuum. Firstly we shall suppose that the set $\left\{\forall\Upsilon_{\Theta_{inf,\nu}}\right\} =^{def} \left\{\left(\exists\Theta_{inf,\nu}, \forall\Upsilon\right)\left[\Upsilon \in \Theta_{inf,\nu}\right]\right\}$ is a metric space (index inf, ν at Υ we omitted). Denote $\delta\Upsilon = \Upsilon' \setminus \Upsilon_0$ and call it configuration perturbation (Υ_0 denotes the Υ value in abscence of changes of the set $\{\Upsilon\}$). Taking into account that $H_{\Upsilon'} = H_{\Upsilon_0}$ (see Eqn. (21)), one can write $(\delta\Upsilon \mid H_{\Upsilon_0}) = (\delta\Upsilon \mid H_{\Upsilon'})$, which means that $\delta\Upsilon$ is mapped to $H_{\Upsilon_0} = H_{\Upsilon'} \subseteq H$. Therefore,

$$\left(\Upsilon' \mid H_{\Upsilon'=\Upsilon_0}\right) = (\Upsilon_0 \mid H_{\Upsilon_0}) \cup (\delta\Upsilon \mid H_{\Upsilon_0}) .$$

Then one can write

$$\left(\Upsilon'' \mid H_{\Upsilon''}\right) = \Psi\left((\Upsilon_0 \mid H_{\Upsilon_0}) \cup (\delta\Upsilon \mid H_{\Upsilon_0}) \mid H_{\Upsilon''}\right) \tag{22}$$

One can consider the set $\mathcal{R} = (\{\Upsilon\} \bigotimes H)$ and its subsets such as $(\Upsilon \bigotimes H_\Upsilon) \subseteq (\{\Upsilon\} \bigotimes H)$. Some of the equations written above can be rewritten in terms of the set $\mathcal{R} = (\{\Upsilon\} \bigotimes H)$.

Define now the following set $(\forall\Upsilon)\,[\Upsilon \subset \{\Upsilon\}]\,(\exists H_\Upsilon)\,[H_\Upsilon \subset H]\,[(\Upsilon \bigotimes H_\Upsilon) \subset \mathcal{R}]$, where H_Υ is the map of Υ to H.

Consider now a certain $\Upsilon' \subset \{\Upsilon\}$ such that Υ', which, in its turn, is a set contains more than one element. Represent Υ' as follows:

$\Upsilon' = \left(\forall\left(\Upsilon'_\gamma \wedge \Upsilon'_\delta\right)\right)\left[\Upsilon'_\gamma \subset \Upsilon', \Upsilon'_\delta \subset \Upsilon', \Upsilon'_\gamma \cap \Upsilon'_\delta = \emptyset\right] \Upsilon'_\gamma \cup \Upsilon'_\delta$.

Map (homomorphism) $\Upsilon' : \Upsilon' \to H_{\Upsilon'}$, where

$H_{\Upsilon'} = \left(\exists\left(H_{\Upsilon'_\gamma} \wedge H_{\Upsilon'_\delta}\right)\right) W_{\Upsilon'\gamma\Upsilon'_\delta} H_{\Upsilon'_\gamma} \cup H_{\Upsilon'_\delta}$,

where $W_{\Upsilon'_\gamma\Upsilon'_\delta} = \left[H_{\Upsilon'_\gamma} \subset H_{\Upsilon'}, H_{\Upsilon'_\delta} \subset H_{\Upsilon'}, H_{\Upsilon'_\gamma} \cap H_{\Upsilon'_\delta} = \emptyset\right]$.

Consider now the pair $\Upsilon'_\delta \wedge H_{\Upsilon'_\delta}$ and map (*isomorphism*) $\Upsilon'_\delta \wedge H_{\Upsilon'_\delta} : \Upsilon'_\delta \wedge H_{\Upsilon'_\delta} \to \Upsilon'$. Then one will obtain the combination $\left\langle \Upsilon' \mid \Upsilon'_\delta \wedge H_{\Upsilon'_\delta} \right\rangle$, which means that $H_{\Upsilon'_\delta}$ labels Υ' by its map to H. The presence of Υ'_δ in the pair $\Upsilon'_\delta \wedge H_{\Upsilon'_\delta}$ targets this inverse mapping namely to Υ'.

Now replace the set $(\Upsilon \bigotimes H_\Upsilon)$ to the set

$$\mathcal{L} = (\forall \Upsilon)\,[\Upsilon \subset \{\Upsilon\}]\,(\exists H_\Upsilon)\,[H_\Upsilon \subset H]\left[\left(\langle \Upsilon \mid \Upsilon_\delta \wedge H_{\Upsilon_\delta}\rangle \bigotimes H_\Upsilon\right) \subset \mathcal{R}\right] \tag{23}$$

The Eqn. (23) means that, unlike $(\Upsilon \bigotimes H_\Upsilon)$, for a given Υ *namely this* Υ corresponds to its map into H. Indeed, $\Upsilon : \Upsilon \rightarrow H$ is homomorphic and, therefore, the inverse operation is not single-valued. Homomorphism, but not isomorphism, was chosen because the isomorphic $\Upsilon : \Upsilon \rightarrow H$ could contradict to the fact that more than one configuration may have the same map, as, for example, in the case when $H = t$, where t is the time, different configurations can exist at the same time moment.

Let us consider now the case when

$$\left(\Upsilon'' \bigotimes H_{\Upsilon''}\right) = \left(\Upsilon_0 \bigotimes H_{\Upsilon_0}\right) \cup \alpha\left(\delta\Upsilon_{H_{\Upsilon_0}} \bigotimes (H_{\Upsilon''} \setminus H_{\Upsilon_0})\right), \tag{24}$$

where $\alpha \in \mathbb{R}$ is a coefficient specific for the considered problem. We shall consider the case $\alpha \ll 1$.

Transition $\Upsilon_0 \rightarrow \Upsilon''$ can be represented by operators $\mathbf{T}_{\Upsilon''\Upsilon_0}$, no matter whether it occurs in one (as it was considered above) or more steps. In the latter case it would be interesting to express the multi-step transition operator $\mathbf{T}_{\Upsilon''\Upsilon_0}$ in terms of single-step ones $\mathbf{T}_{\Upsilon^{(l)}\Upsilon^{(l-1)}}$, $l \in [0, l_{max}] \in \mathbb{N}$ such that $\left(\Upsilon^{(l_{max})} = \Upsilon''\right)$. As an example this approach can be applied to the transition represented by the Eqn. (24). We shall use for it operators

$$\mathbf{\Gamma}_{l,l-1} =^{def} \left\langle \Upsilon^{(l)} \bigotimes H_{\Upsilon^{(l)}} \left| \mathbf{\Gamma}_{l,l-1} \right| \Upsilon^{(l-1)} \bigotimes H_{\Upsilon^{(l-1)}} \right\rangle \tag{25}$$

and construct from them the operator

$$\mathbf{\Gamma}_{l_{max},1} =^{def} \sum_{l=1}^{l_{max}} \mathbf{\Gamma}_{l,l-1} \tag{26}$$

Generalize the Eqn. (24) for multi-step process. We can write

$$\Upsilon^{(l)} \bigotimes H_{\Upsilon^{(l)}} = \left(\Upsilon^{(l-1)} \bigotimes H_{\Upsilon^{(l-1)}} \right) \bigcup \alpha_{l,l-1} \left(\delta \Upsilon^{(l-1)}_{H_{\Upsilon^{(l-1)}}} \bigotimes \left(H_{\Upsilon^{(l)}} \setminus H_{\Upsilon^{(l-1)}} \right) \right) \tag{27}$$

Let $\mathcal{P}$ is the projection operator that projects $(\Upsilon \bigotimes H_\Upsilon)$ into the set $\{\Upsilon\}$. We denote $\mathcal{P}\left(\Upsilon \bigotimes H_\Upsilon\right) = \Upsilon_{H_\Upsilon}$ that makes clear the meaning of the corresponding member in Eqn. (27).

Following Eqns. (19) and (20), one can write

$$\mathbf{P}_{a\,pr} =^{def} \mathbf{P}\mathcal{P}\left(\Upsilon_0 \bigotimes H_{\Upsilon_0} \right), \tag{28}$$

$$\mathbf{P}_{a\,post} =^{def} \mathbf{P}\mathcal{P}\left(\Upsilon'' \bigotimes H_{\Upsilon''} \right), \tag{29}$$

$$I = K log_2 \frac{\mathbf{P}_{a\,post}}{\mathbf{P}_{a\,pr}} = K log_2 \frac{\mathbf{P}\mathcal{P}\left(\Upsilon'' \bigotimes H_{\Upsilon''}\right)}{\mathbf{P}\mathcal{P}\left(\Upsilon_0 \bigotimes H_{\Upsilon_0}\right)} \tag{30}$$

Using Eqn. (24), one obtains

$$I = Klog_2 \frac{\mathbf{P}\mathcal{P}\left[\left(\Upsilon_0 \bigotimes H_{\Upsilon_0}\right) \bigcup \alpha \left(\delta \Upsilon_{H_{\Upsilon_0}} \bigotimes \left(H_{\Upsilon''} \setminus H_{\Upsilon_0}\right)\right)\right]}{\mathbf{P}\mathcal{P}\left(\Upsilon_0 \bigotimes H_{\Upsilon_0}\right)} \tag{31}$$

The amount $I_{l,l-1}$ of the information obtained at the transition $l-1 \rightarrow l$ is expressed by the formula like Eqn. (31):

$$I_{l,l-1} = Klog_2 \frac{\mathbf{P}\mathcal{P}\left[\left(\Upsilon_{l-1} \bigotimes H_{\Upsilon_{l-1}}\right) \bigcup \alpha_{l,l-1} \left(\delta \Upsilon_{H_{\Upsilon_{l-1}}} \bigotimes \left(H_{\Upsilon_l} \setminus H_{\Upsilon_{l-1}}\right)\right)\right]}{\mathbf{P}\mathcal{P}\left(\Upsilon_{l-1} \bigotimes H_{\Upsilon_{l-1}}\right)} \tag{32}$$

Then the amount of the information obtained at the multi-step transition is as follows:

$$I_{0,l_{max}} = Klog_2 \prod_{l=1}^{l_{max}} \mathcal{I}_{l,l-1}, \tag{33}$$

where $\mathcal{I}_{l,l-1}$ is the expression following the sign of logarithm in Eqn. (32).

Let us introduce $\delta_{l_{max},0}H$ as follows

$$\delta_{l_{max},0}H =^{def} \left(H_{\Upsilon_{l_{max}}} \setminus H_{\Upsilon_0}\right) = \sum_{l=1}^{l_{max}} \left(H_{\Upsilon_l} \setminus H_{\Upsilon_{l-1}}\right) \tag{34}$$

Then one can define the "effective" representation of the information obtained at all these multi-step transitions

$$I_{0,l_{max},eff} = I\left(\left(\Upsilon_{l_{max}} \setminus \Upsilon_0\right) \bigotimes \left(H_{\Upsilon_{l_{max}}} \setminus H_{\Upsilon_0}\right)\right) \tag{35}$$

However, this representation does not mean that the multi-step formula (33) can be replaced to the as if one-step formula (35) because $\left(H_{\Upsilon_{l_{max}}} \setminus H_{\Upsilon_0}\right)$

is to be calculated according the formula (34) by the multi-step way.

6 NOTE ON LANGUAGE

A subset $\{\Upsilon\} \subseteq \left(\left\{\forall\Upsilon_{\Theta_{inf,\nu}}\right\}\right)$ could be finite or countable even if $\Theta_{inf,\nu}$ is the continuum. In this case $\{\Upsilon\}$ should be built of elements chosen from the set $\left\{\forall\Upsilon_{\Theta_{inf,\nu}}\right\}$. If we want to express the information by a language (formal or natural), it is to choose a finite set of elements $(\iota \in \mathbb{N}, \iota \in [1, \iota_{max}])$ $\{\Upsilon_\iota\}$ so that each environment of each element $\Upsilon_\iota$would contain an infinite set $\{\Upsilon\} \in \left\{\forall\Upsilon_{\Theta_{inf,\nu}}\right\}$, *i. e.*, all $\Upsilon_\iota$must be limit points of the set $\left\{\forall\Upsilon_{\Theta_{inf,\nu}}\right\}$. These elements Υ_ι will be the letters of the language alphabet, their combinations will be words, phrases *etc.*. The finite set of of all Υ_ι included to these combinations of letters we shall denote $Doc_\tau = \{\forall\Upsilon_{\iota,\tau}\}$ and call "document τ". However, all possible combinations of letters form an infinite countable set of limit points of the set $\left\{\forall\Upsilon_{\Theta_{inf,\nu}}\right\}$. All letters contained in all *permitted* combinations of them form a language. What combinations of letters, strings (words) and phrases are permitted is defined by the grammar. Thus, it is to define the *generative device* that generates a language. This device is the *generative grammar [16]*

$$\mathbf{G} = \left(\Upsilon_{\iota,N}, \Upsilon_{\iota,T}, \Upsilon_{\iota,N,0}, F\right), \tag{36}$$

where $\Upsilon_{\iota,N}$ and $\Upsilon_{\iota,T}$ are disjoint alphabets, $\Upsilon_{\iota,N,0} \in \Upsilon_{\iota,N}$, and $F = \{(P, Q)\}$ is a finite set of ordered pairs (P, Q) such that Q is a word over the alphabet $\Upsilon_\iota = \Upsilon_{\iota,N} \bigcup \Upsilon_{\iota,T}$ and P is a word over Υ_ι containing at least one letter of $\Upsilon_{\iota,N}$. The elements of $\Upsilon_{\iota,N}$ are called *nonterminals* and those of $\Upsilon_{\iota,T}$

terminals; $\Upsilon_{\iota,N,0}$ is called the *initial letter.* Elements (P, Q) of F are called *rewriting rules* or *productions* and are written $P \to Q$ ([16]). We limit ourselves with these short notes concerning formal languages (taken from [?]) and refer the reader to the book of Salomaa [16]. An interesting problem that we shall consider is what could happen with the information written in terms of the continuum set $\{\Upsilon\}$if it will be rewritten in terms of a language, *i. e.*, of a countable set. The condition that $\Upsilon_\iota$must be limit points of the set $\left\{\forall \Upsilon_{\Theta_{inf,\nu}}\right\}$ hints at the way to clarify it.

All possible documents $(\forall Doc_\tau)$ form an infinite countable set (of letters) that we shall denote $\Theta_{inf,\nu,\alpha} \subset \Theta_{inf,\nu}$ and its configurations φ_α, where α indicates the chosen subset $\Theta_{inf,\nu,\alpha}$ among all possible ones. The set $\{\forall \varphi_\alpha\}$ for each α evidently will be countable. We shall consider the information processing by use of a language in the countable subset $\Theta_{inf,\nu,\alpha}$ and the transfer a without-language information contained in $\Theta_{inf,\nu}$ to $\Theta_{inf,\nu,\alpha}$, where it will be expressed in a language.

Let $(\exists \{\Upsilon\}) \left[\{\Upsilon\} \subseteq \left\{\forall \Upsilon_{\Theta_{inf,\nu}}\right\}\right]$. Let C is a combination of alphabet letters. Map $\forall \{\Upsilon\} : \forall \{\Upsilon\} \to (\{\forall C\}) [\{\Upsilon_C\}]$. Let each set $\{\Upsilon\}$ possesses a set $\left\{\mathcal{F}_{\{\Upsilon\}}\right\}$ of properties, among them may be also the information. We demand that this mapping would satisfy the condition that the difference between the set of properties $\mathcal{M} = \mathcal{M}_{\{\Upsilon_C\}}$ obtained as result of the mapping and $\left\{\mathcal{F}_{\{\Upsilon\}}\right\}$ for all $\{\Upsilon\}$ should be minimum. Of course, at each concrete case the existence of constraints could be expected, *e. g.*, in the case of a language its grammar should provides constraints. Note, that the described procedure could be performed when C is any countable set.

The described pure mathematical procedure opens a way to the elucidation of the language origin. To achieve this goal it is necessary, in particular, to find out what a *real* process (for example, going on at the human thinking) plays role of the abovementioned mapping.

7 ON A POSSIBLE APPROACH TO THE INFORMATION CONTENT AND VALUE

Our approach used above to define the concept of the information amount was based on the consideration of subsets of the set of possible configurations of a given set. On the face of it this basis would be too poor to use it for the definition of such complicated concepts as the information content and value are. However, in the consideration made above we did not use the majority of properties of configuration set and its subsets. We expect that their use could allow us to perform the task to define concepts of the information content and value in the framework of the configurations' consideration. Let us do a step in this direction.

Let us call Υ the SUBJECT OF THE INFORMATION*,* or, by other words, *what* is the considered information *about*. But, really, what could be gotten to know about Υ, what information could be obtained on Υ, in general? This information is only how much is theΥ contribution $C(\Upsilon)$ to $\{\Upsilon\}$that can be expressed as follows:

$$C(\Upsilon) =^{def} \frac{\mu(\Upsilon)}{\mu(\{\Upsilon\})}, \tag{37}$$

where μ is the measure of the set in backets and $\{\Upsilon\}$ is the set of all configurations Υ. Thus, the information on the value of $C\left(\Upsilon\right)$ is the only *subject of the information on* Υ. The set $\{\Upsilon\}$, represents the one of different subjects of the information such that each of them corresponds to a certain Υ, while $\{C\left(\Upsilon\right)\}$is the corresponding set of information subjects' contributions. It can be called the *distribution function* of subjects' contributions and denoted

$$\mathbb{F} = \mathbb{F}\left(C_{\Upsilon}\right), \tag{38}$$

where $C_{\Upsilon} =^{def} C\left(\Upsilon\right)$. Now one can define the Shannon information amount corresponding to a change of this distribution function (we denoted $\mathbb{P}$ the probability):

$$\exists\left(H_1 \prec H_2\right)\mathbb{I}_{2,1} = Klog_2\frac{\mathbb{P}\left(\mathbb{F}_2\left(C_{\Upsilon}\right) \mid H_2\right)}{\mathbb{P}\left(\mathbb{F}_1\left(C_{\Upsilon}\right) \mid H_1\right)}, \tag{39}$$

no matter by what this change was produced, $K = const$. The Eqn. (39) is valid only when the set of all functions $\mathbb{F} = \mathbb{F}\left(C_{\Upsilon}\right)$ is finite or infinite countable one. In the Eqn. (39) probabilities appearing in the numerator and denominator are *aposteriory* and *apriory* ones correspondingly.

For continuum set of functions $\mathbb{F} = \mathbb{F}\left(C_{\Upsilon}\right)$ the Eqn. (39) can be replaced by the following one

$$\mathbb{I} = -K_c \int \delta\mathbb{F}\mathbb{P}\left(\mathbb{F}\right) log_2 \mathbb{P}\left(\mathbb{F}\right), \tag{40}$$

where $K_c = const$.

For example, varieties of possibilities could be issued from the taking into account compositions of different subsets of configurations and the composition of the given set as consisting of such subsets. Then the *content of an information* can be defined as *what kinds* of the abovementioned composition changes this information provokes, while the *information value* can be defined as the *information change* at such composition changes (as compared with the primary information).

Here we limit ourselves with these remarks because these two problems merit to become the subjects of special researches.

8 INFORMATION TYPES AND SET TYPES

The purpose of the Sections 8 - 10 is to consider different ways to set up the hierarchy of sets able to treat the information: A) on the grounds of the theory of information by definition of the concepts of information types and the information hierarchy, and B) on the grounds of the Russell's theory of types [15]. It will be shown that in the case when the information and its value [3],[20] are expressed in a language and built according logical rules the both approaches lead to the same result and that Russell's types can be expressed in terms of the value of the information obtained as result of the information treatment.

As the starting point we consider here a well-ordered finite or countable set S of well-ordered sets supposing that some of these sets are able to treat the information. Now introduce and set up the hierarchy among these sets based on their properties with respect to the information treatment and thereupon reorder set S with respect to hierarchy of sets forming it. The information treatment includes the following functions: 1) the receipt of information (from), 2) the sending of information (to), 3) the information processing, 4) the information interpretation, and 5) the information storage in memories. Define that the hierarchy of such a set is determined by which of these functions it executes. However, it is to take into account that this criterion could be not sufficient one to determine the hierarchy because there is the possibility that more than one set have the same type of the information treatment and, therefore, the same hierarchy according to this criterion that prevents to set up the order among them, when one reorders the set S with respect to hierarchies of sets forming it. It seems to be like the quantum state degeneration. This analogy suggests an idea to search for some supplementary criterions that may allow one to attribute to these sets different hierarchies (to break this "degeneration"). Then they can be ordered among themselves also with respect to their hierarchies. In the case when no such supplementary criterion exists, they can be ordered among themselves on the grounds of reasons other than hierarchy or arbitrarily.

We shall accept that the lowest hierarchy is attributed to sets that, in general, do not treat the information while the highest one is attributed to each set executed all five functions. If there is only one set of the highest hierarchy, then order (or reorder) set S with respect hierarchies of sets forming it so that the set possessing the highest hierarchy will be the first (we shall attribute to it number 0), while the hierarchy of other sets decreases when the number augments.

Let us reinterpret the described approach in terms of the INFORMATION VALUE [3],[20]. The information obtained by the interpretation of the processed information has the largest value because it is able to induce the most serious changes to the understanding of the obtained information meaning and, on these grounds, to invent its new applications creating material changes. For example, if to speak on physics, such an interpretation may mean the replace of existing physical laws to the new ones, which leads to a serious, maybe drastic change of our understanding of the going on in the World (*cf.* the replace of the classical mechanics to the relativity and quantum mechanics) and creation new applications, e. g., nuclear energy, quantum computing etc... Following this way we define the hierarchy in accordance with the order of values of the information. In Sec. 9 the other approach based on Russell's theory of types [15] will be represented.

The information on going on within a certain set of those forming the set should be transmitted step-by-step to the set of the highest hierarchy to be processed and interpreted. The process of the information extraction (on occurring within a set) we shall call MEASUREMENT considering it as a general mathematical notion. In its physical applications we, for brevity, shall use this term also for observation. For example, in the microworld one uses measurements while in the macroworld (e. g., in the astronomy and astrophysics) mainly observations are used.

9 TYPES, INFORMATION TREATMENT, AND SETS HIERARCHY

9.1 HIERARCHY WITHIN THE INFORMATION

Consider now the setting up hierarchy within the information with more details. The information is characterized by its amount (see, for example, [17], [4]) and value [3], [20]. Let us consider the following multi-step process:

primary information I_1 - information creation I_2 - information creation I_3 - information creation I_4- information creation I_5- etc.

The information value can be determined as the amount of the information I_{n+1} created in all n these steps divided into amount of the primary information I_1. However, it cannot be the only characteristic of the information value because it does not take into account properties of the information content. If properties of the created information content are taken into account, then the value of the primary information should be represented by a set of such characteristics + the number defined above. Let us try to represent these characteristics in general form. Denote each of them ξ_ν, where ν labels a certain characteristic. Then one has $\Xi = (\forall\nu)\{\xi_\nu\}$. Denote those of the primary information $\Xi_1 = (\forall\nu_1)\{\xi_{1,\nu_1}\}$, of the secondary information $\Xi_2 = (\forall\nu_2)\{\xi_{2,\nu_2}\}$,..., of the n^{th} step $\Xi_{n+1} = (\forall\nu_{n+1})\left\{\xi_{n+1,\nu_{n+1}}\right\}$, etc.. The index at ν is necessary because for information obtained at each step the set of properties could be different from that for obtained at other steps.

Let us introduce the norm of a property $\|\xi_{l,\nu_l}\|$, which is a number. How the norm is defined depends on each concrete case, so we do not consider this problem for the general case. The complete representation of the primary information value is $\prod_{\forall l}^{\otimes} \Xi_l$, where $\prod^{\otimes}$ denotes the Cartesian product of sets. Using the norm of the property one can introduce a quantity characterizing the primary information value. It is $J_V = \left(\| \prod_{\forall l}^{\otimes} \Xi_l \| + I_{n+1}\right) I_1^{-1}$ that will be called INFORMATION VALUE. However, it must be kept in mind that, really, it is only a partial characteristic of the information value.

If only $n' < n$ steps are realized, while principally n steps are possible, one can define the concept of the POTENTIAL INFORMATION VALUE that is determined for all n steps, no matter how many of them are realized. One can define also the concept of the CONSTRAINED INFORMATION VALUE when constrains prohibit the realization of a part of steps. Notice that it must distinguish between "potential information value" and "value of the potential information". An example of the potential information is the genetic information written by the genetic code on a DNA molecule. Its value is not the potential one. The potential information is the one contained in a text *etc.*. **Really, the potential information is not an information, in general, but CONSTRAINS limiting the future development and behavior of the considered system, *e. g.*, the development and behavior of biological system originated of a certain DNA double helix.**

Note that the use of the norm $\|\xi_{l,\nu_l}\|$ is not the only way to compare information values of primary information in different cases. It is possible to refuse from the use numbers for this purpose and instead of it to assign to each ξ_ν quality Q (which is not obligatory a number) such that between any two Q_ν and $Q_{\nu'}$ the relation of order, e. g., $Q_{\nu'} \prec Q_\nu$, exists. One can interpret this relation so that the quality Q_ν is higher than quality $Q_{\nu'}$. Whether this approach can be used instead of the use of $J_V = \left(\| \prod_{\forall l}^{\otimes} \Xi_l \| + I_{n+1}\right) I_1^{-1}$ to express the information value? It is possible, if relation of order like $Q_{\nu'} \prec Q_\nu$ can be set up between any two $\prod_{\forall l}^{\otimes} \Xi_l$ (for both cases of the initial information). But it seems questionable because different ξ_ν with different Q_ν enter to this Cartesian product in complicated combinations.

9.2 BERTRAND RUSSELL'S THEORY OF TYPES APPLICATIONS

DEFINE that the type of the information is determined by its value, potential value or constrained value, accordingly to the considered problem.

DEFINE that the information hierarchy is set up according types of the information.

Consider the case when the information and its value are expressed in a language and built according the rules of logic. Then one can define the types of information using the Bertran Russell's theory of types [15] as the starting point. We read in the abovementioned article of Bertran Russell: "A *type* is defined as the range of significance of a propositional function, i. e., as the collection of arguments for which the said function has value." "Thus whatever contains an apparent variable must be of different type from the possible values of that variable; we will say that it is of a *higher* type." In our case, for example, the processed information can be considered as the set of values of apparent variables that are, in their turn, the result of the information interpretation.

Consider it in more detail. Let $V = (\forall r \, [r \in \mathbb{N}; r \in [r_0, r_{max} > r_0]]) \{V_r\}$

are apparent variables [15], [16] of the processed, but not yet interpreted information. Here and in the following text $\mathbb{N}$ denotes the set of all natural numbers. The interpretation consists in 1) the setting up connections between V_r with different values of r, 2) the setting up rules how values of V_r can be calculated and 3) the setting up connections with variables characterizing external factors influencing the considered system. We shall call a THEORY the result of the interpretation. Note that thereupon this new theory may be, in its turn, used for the interpretation of the new information, which itself was obtained on the grounds of a preceding theory, which also was obtained on the grounds... and so on. $V = (\forall r\,[r \in \mathbb{N}; r \in [r_0, r_{max} > r_0]])\,\{V_r\}$ can be obtained now as values of apparent variables $U = (\forall s\,[s \in \mathbb{N}; s \in [s_0, s_{max} > s_0]])\,\{U_s\}$ of interpreted information, i. e., from the theory. Therefore, according Russell the interpreted information (expressed in terms of these apparent variables) is of higher type in comparison with the processed, but not yet interpreted information. The same can be said on the not yet processed and processed information: the second is of higher type than the first one. In general, elements of a set or their configurations can be considered as values of apparent variables of the information about this set. The information has, therefore, higher type than the set itself.

Let us return to the consideration of the approach when the type of the information is defined also on the grounds of its value [3], [20]. It is more general than the written above Russel's approach because it is not limited with the condition that the information and its value must be expressed in a language. However, in the case when the information and it value are expressed in a language, this definition seems to be equivalent to the one based on Russell's theory of types and, in particular, leads to the same result that the interpreted information is of the highest type. Note that different levels of this interpretation may exist so that the information obtained by these kinds of interpretation could have different types. With the purpose to avoid such an uncertainty at the consideration of sets treating the information one defines the type of a set treating the information as the highest of the types of the information obtained by this treatment.

Let us now consider a well-ordered set containing sub-sets able to receive (also by performing measurements), to send, to process, to interpret and to store (in memories) the information (cf. Sec. 7.). We shall call such a subset OBSERVER, iff it is able to execute ALL these functions including measurements. We do not suppose that all considered subsets are observers, in other words, not each of them executes all the abovementioned functions.

Our purpose is to set up the hierarchy among such sets based on the information hierarchy defined above.

We define that the hierarchy order of two sets treating the information corresponds to their types order. The generalization to any finite or infinite countable well-ordered set of sets is evident. WE DEFINE that *the hierarchy of a set containing a subset able to treat the information would be equal to the hierarchy of this subset.* If this set contains a set of such subsets treating the information having different types, we shall define that the hierarchy of the considered set is determined by the highest of these types. Thus, the type attributed by definition to subsets able to receive, process, send and store the information (they are not observers) would be lower than that attributed to the observer which is able also to interpret the information.

The information value probably does not affect the original information entropy, but it may create the negative or positive entropy production. For example, at the explosive crystallization of an amorphous body by laser light the information carried by this light initiates the transformation of a disordered amorphous body to the ordered crystal. The value of the original information corresponds to the big (by the absolute value) negative information entropy production. At the initiation of the explosion of an explosive by electric signal the value of the original information corresponds to the big positive entropy production because an ordered structure is turned into a disordered one. This means, the value of the original information corresponds to the absolute value of the entropy production. DEFINE the specific absolute value of the entropy production as its absolute value divided to the original information amount. It can be an important characteristic of the information action. This connection between the information value and such a thermodynamic quantity as the entropy production suggests the idea that one can formulate the problem of the hierarchy also in terms of the thermodynamics.

Not all processes initiated by the original information are obligatory occurred at one step (cf. written in the beginning of this section). The "first creature" may initiate new processes creating "the second creature" etc. If only the "first creature" is taken into account or the following steps be prohibited by any conditions, the value of the original information would be less than in the case when the "second and following creatures" will be realized and taken into account. Therefore, as it was mentioned above, the value of the information only is not enough to characterize the ability of the original information, and it is so we introduced above the concept of potential value of information based on taking in the account those effects that the considered information potentially is able to produce (maybe in some steps), but not yet produced. For example, the information obtained by the observer can possess big value and potential value because its interpretation (possibly, even the creation of new physical laws) may produce remarkable effects.

Now we can DEFINE the concepts "VALUE OF INFORMATION" and "POTENTIAL VALUE OF INFORMATION" in terms of Russell's theory of types [15], [16]. We shall accept that the type of the information is determined as the type of its expression in terms of mathematical logic notions [15].

DEFINITION 1. The value of the considered information (primary information) is the highest Russell's type of the information created by the activated primary information in maximum number of executed steps.

DEFINITION 2. The potential value of the considered information (primary information) is the highest Russell's type of the information that could be created by the activated primary information in maximum number of principally existing steps.

The activated information is the information that produces new information, physical, chemical, biological, industrial, social and other effects. Example: the prominent letter of Albert Einstein to Franklin D. Roosevelt, President of the USA, where Einstein proposed to begin researches aimed to create the nuclear weapon. It contained information that could be called frozen or potential one up to the moment when the President read it and decided to begin these researches. Then it became to be the active information. If the President did not read this letter or rejected the Einstein's proposal, the value of the information contained in this letter would be equal to the zero and only its potential value should be enormous.

A subset able to execute the information treatment must contain a subset formed of elements that, in their turn, are sets containing more than one element. Then in the considered subset different distributions of elements (for example, with respect to numbers of elements including to each element of this subset) can exist and, therefore, the probability and information can be defined.

The ability of such a subset to receive, send, process, store and interpret the information depends on the set structure. If there are a number of such subsets, their relative hierarchy is defined as their relative ability of the information treatment. The rough classification can be as follows: the lowest hierarchy (=0) have those which cannot receive, cannot send and cannot process the information; the hierarchy =1 is attributed to subsets which are able to receive, to send, to store, but cannot process the information; the hierarchy =2 is attributed to subsets which are able to receive, to send, to store and to process the information; the hierarchy =3 is attributed to subsets which are able to receive, to send, to process, to store and to interpret the information.

Inside each type could be different sub-types with different hierarchy among them. For example, inside type (3) could be different levels of the information interpretation. The highest hierarchy among them is attributed to the subset which extracts from the received and processed information general, in particular, physical laws.

Note that the active information can create new information, but it can create phenomena of different nature, e. g., physical, chemical, biological, geophysical, emotions of human beings and animals, thoughts of human being expressed or not in a language, logical or not etc.. It must be taken into account at the consideration of the information value. The mathematical logic, in general, and Russell's theory of types, in particular, can be applied to the information value consideration only when the processes can be expressed in a language (or languages) according logical rules at all stages. Note that it must not negate without a serious consideration the possibility of existing of the information which is not expressed in a language (*cf.*, [9]), but despite it is built according logical rules. Of course, if it exists, these logical rules must be a *generalization* of those of the existing logic, for example, those connecting certain sets, but not propositions etc..

10 HIERARCHY AND SCALING AMONG SETS $S^{(l)} \subseteq \Phi$ ON THE GROUNDS OF HIERARCHY WITHIN THE INFORMATION

Let us try now to construct the theory of the scaling using the considerations of previous Sections.

Let there is a well-ordered not empty final or countable set $S \subseteq \Phi$ of not intersected not empty well-ordered sets

$$S^{(l)} \left[l \in \mathbb{N}, l \in [0, l_{max} \vee \infty], \left(l' \neq l\right) \Rightarrow S^{(l)} \cap S^{(l')} = \emptyset \vee \left(l' = l\right) \Rightarrow S^{(l)} \right]$$

The written above allows one to set up the hierarchy between all well-ordered sets forming the well-ordered set S and to approach to the *scaling*.

One can order the set $S = \left\{S^{(l)}\right\}$ with respect to hierarchies of sets $S^{(l)}$. Let us set up the hierarchy within the set S so that the set $S^{(0)}$ possesses the highest hierarchy and the hierarchy of sets $S^{(l)}$ decreases with the increase of l: $l' > l \Rightarrow \tilde{h}\left(S^{(l')}\right) < \tilde{h}\left(S^{(l)}\right)$, where $\tilde{h}\left(S^{(l)}\right)$ denotes hierarchy of the set $S^{(l)}$. Let us postulate that only the set $S^{(0)}$contains subsets $\Theta_{inf,\nu}$ allowed to receive, to send, to process, to interpret and to store the information about all other sets. This means, we consider here the case when the observer(s) exists at only one scale, namely that having the highest hierarchy. The set $S^{(0)}$ contains a subset $\Theta_{S^{(0)},inf,\nu} \subset S^{(0)}$ possessing the following properties: it is able to receive (also by measurements), to process, to interpret, and to store the information. We shall call this subset $\Theta_{S^{(0)},inf,\nu}$ OBSERVER. The abovementioned interpretation is done on the grounds of certain laws (mathematical, physical etc.) that should be expressed in a convenient mathematical form and included to the subset $\Theta_{S^{(0)},inf,\nu}$. If a certain deviation $\delta S^{(l')}$ $(l' \in \{l\}, l' > 0)$ has occurred with the set $S^{(l')}$ itself, the information on it must be forwarded step-by-step to the subset $\Theta_{S^{(0)},inf,\nu}$ to be processed, interpreted and stored.

Let each $S^{(l)}$ possesses some properties α: $\left(l \in \mathbb{N}, \gamma^{(l)} \in \mathbb{N}\right)\left\{\alpha^{(l)}_{\gamma^{(l)}}\right\}$. These properties may be very various. In particular, they may be laws determining the perturbation propagation within the set $\{\Upsilon\}$ (*cf.* above). Choose $\forall S^{(l)}$ so that $\forall\left(S^{(l)} \wedge S^{(l')}, l \neq l'\right) \nexists \left(\alpha^{(l)}_{\gamma^{(l)}} \wedge \alpha^{(l')}_{\gamma^{(l')}}\right)\left[\alpha^{(l)}_{\gamma^{(l)}} = \alpha^{(l')}_{\gamma^{(l')}}\right]$. Under this condition one can define sets $S^{(l)}$ as SCALES of the set S.

Properties $\left(l \in \mathbb{N}, \gamma^{(l)} \in \mathbb{N}\right)\left\{\alpha^{(l)}_{\gamma^{(l)}}\right\}$ of the set $S^{(l)}$ are defined on the grounds of results of measurements made by observer, *i. e.*, a certain subset $\Theta_{S^{(0)},inf,\nu}$ of the set $S^{(0)}$. Any change of the set of measurements may lead to changes of properties of some $S^{(l)}$. As result the condition

$$\forall \left(S^{(l)} \wedge S^{(l')}, l \neq l'\right) \nexists \left(\alpha^{(l)}_{\gamma^{(l)}} \wedge \alpha^{(l')}_{\gamma^{(l')}}\right) \left[\alpha^{(l)}_{\gamma^{(l)}} = \alpha^{(l')}_{\gamma^{(l')}}\right]$$

may be violated for some pairs $S^{(l)}$and $S^{(l')}$. Then it must change the representation of S in terms of $S^{(l)}$ to define new $S^{(l)}$ as scales.

If within a scale $S^{(q)}$ a phenomenon is detected , which cannot be interpreted in the framework of $S^{(q)}$ properties $\left(q \in \mathbb{N}, \gamma^{(q)} \in \mathbb{N}\right) \left\{\alpha^{(q)}_{\gamma^{(q)}}\right\}$, then two possibilities exist. The first of them is to define for the same set $S^{(q)}$ the new set of properties $\left(q \in \mathbb{N}, \gamma'^{(q)} \in \mathbb{N}\right) \left\{\alpha^{(q)}_{\gamma'^{(q)}}\right\}$, in the framework of which the abovementioned phenomenon could be interpreted. This step will be successful, iff no phenomenon exist that cannot be interpreted in this framework, even, if it is such a phenomenon that did was interpreted by $\left(q \in \mathbb{N}, \gamma^{(q)} \in \mathbb{N}\right) \left\{\alpha^{(q)}_{\gamma^{(q)}}\right\}$. If this attempt be unsuccessful, then the second possibility should be used, namely that this phenomenon occurs within the scale $S^{(q+1)}$ possessing the set of properties $\left((q+1) \in \mathbb{N}, \gamma^{(q+1)} \in \mathbb{N}\right) \left\{\alpha^{(q+1)}_{\gamma^{(q+1)}}\right\}$ or, maybe,

$$\left(m \in [1, m*], (m \wedge m*) \in \mathbb{N}, q \in \mathbb{N}, \gamma^{(q+m)} \in \mathbb{N}\right) \left\{\alpha^{(q+m)}_{\gamma^{(q+m)}}\right\}.$$

Whether it is possible that $m^* \to \infty$ or $\exists q_{max} < \infty$ and, therefore, $m^* \leq q_{max} - q$? This problem was considered in detail in our work [19] in the connection with the microworld physics. The result was that $\exists q_{max} < \infty$. This consideration, at least in general outline, can be repeated on the general mathematical level without any connection with the physics.

11 CONCLUSIONS

In the present paper we considered the origin of the information directly from the set theoretical definition of probability that was extended for this purpose. This extention was based on the generalization of elementary event concept used in the theory of probability, and definition of elementary events of *different (higher) orders*. They represent *changes* of a given *set configuration*, *i. e.*, *its partition into subsets*. Then elementary events can be ordered which, in particular, allows one to define *apriori* and *aposteriori* probabilities necessary for the definition of the Shannon *information amount* [17] *carrying by an event*.

Our definition of elementary events on the grounds of set configurations (partitionings) allows one to consider the *information treatment by sets*. Naturally, at a certain stage such a consideration demands the taking into account certains properties attributed to the considered set that could be not in the framework of its properties as an abstract mathematical set, *e. g.*, physical properties, if this set is a physical system. In other words, at this stage it is necessary to join the pure theory of *abstract sets* and the nature. It was demonstrated on some simple examples.

Principally, a set treating the information may be extremelly complicated, *e. g.*, human brain. It is interesting to find the minimum degree of the set complexity when it is still able to execute a certain selected task of the information treatment. These tasks are as follows: the information acceptance (import), the information processing, the information interpretation, the information generation, the information export, the use of the information for some actions. The most complicated and complex of them is the *information interpretation*. This is usually the task of an *observer*. In physics it is customary to suppose that observer is a human being. Human being is a set of elements (nuclei, electrons, atoms, molecules *etc.*). It suggests an idea to consider sets able to interpret the information as observer, no matter whether such a set is a human being or not. It would be natural to attribute to set-observer the highest hierarchy among sets treating the information, if to define the information interpretation as the highest function of the information treatment. In the present paper the hierarchy among sets treating the information was established according the level of the information treating by each of them.

In Sec. 6 the mathematical way is outlined to find the general grounds of the language origine.

In Sec. 7 the concepts of the information content and value are defined in the framework of the set theory. Subject of the information is defined as a set, namely, a configuration defined above, and, thereupon, the information about this subject is reduced to the one what is its contribution to the set of all possible subjects (in the considered case). The distribution function of these contributions is introduced and used to define the Shannon information amount corresponding to a change of this distribution. The content of the information is defined as what kinds of composition (of configurations) changes the information initiates.

In Secs. 8 and 9 the hierarchy establishing between sets treating the information is considered on the grounds types of the information treating, as well as on the grounds of information types defined by use of the Bertrand Russell's theory of types [15]. The concepts of the *information value* and *potential information value* were redifined on this grounds. In Sec. 10 the results obtained in Secs. 8 and 9 are applied to the problem of scaling within a set of well-ordered sets able to treat the information.

References

[1] Leon Brilloin, Science and Information Theory, Academic Press Inc., Publishers, New York 1956

[2] Robert L. Causey, Logic, Sets, and Recursion, Second Edition, Jones and Bartlet Publishers, Sudbury, Massachusetts, 2006

[3] Manfred Eigen, Selforganization of Matter and the Evolution of Biological Macromolecules, Naturwissenschaften **58**(1971), 465

[4] William Feller, An Introduction to Probability Theory and its Applications, vol. 2, Ch. IV: Probability Measures and Spaces, *Willey, New York,* 1971

[5] Seymour Haiden, & John F. Kennison, Zermelo-Fraenkel Set Theory, *Charles E. Merrill Publishing Company, Columbus, OHIO, A Bell & Howell Company,* 1968

[6] J. L. Hickman, General-Well-Ordered Sets, J. Austral. Math. Soc. A**19**(1975)**,** 7-20

[7] Tomas J. Jech, Set Theory, The 3d millenium edition, *Springer, Berlin-New York*, 2003

[8] A. N. Kolmogorov, Foundations of the Theory of Probability, *Chelsea Publishing Company, New York,* 1956

[9] Dominique Laplane, Penser, c'est-à-dire? Enquête neurophilosophique, ARMAND COLIN, Paris, 2005

[10] Michel Loève, Probability Theory, 3d ed., *D. Van Nostrand Company, Inc., Princeton-New Jersey-Toronto-London-Melbourn,* 1963

[11] M. Métivier, Notions fondamentales de la théory des probabilité. Mâitrises de Mathématiques, *Dunod, Paris,* 1968

[12] Dietrich Morgenstern, Einführung in die Wahrscheinlichkeitsrechnung und mathematishe Statistik, *Springer-Verlag*, 1968

[13] Jacques Neveu, Bases mathématiques du calcul des probabilités, *Masson, Paris*, 1964

[14] Patric Suppes, Axiomatic Set Theory, *Dover Publications, Inc., New York,* 1972

[15] Bertrand Russell, Mathematical Logic as Based on the Theory of Types, Am. J. Math. **30**(1908), 222-262

[16] Arto Salomaa, Formal Languages, Academic Press, New York,1973

[17] C. E. Shannon, The Bell System Technical Journal **27**(1948) 379-423, 623-656

[18] A. Ya. Temkin, Some Ideas on Information Processing, Thinking and Genetics, *Tel-Aviv University Press,* 1999; available also on-line: http://www.eng.tau.ac.il/~temkin/

[19] A. Ya. Temkin, Hierarchy Based Microworld Scales' Classification and Microworld Physics, http://arxive.org/abs/math-ph/0507016 , 2005

[20] M. V. Volkenstein, The Amount and Value of Information in Biology, Found. Phys. **7**(1977)**,** 97-109

[21] Alfred North Whithead & Bertrand Russell, Principia Mathematica, Second Edition, Cambridge at the University Press, 1963