Consciousness via Phase Transition of Learning: Qualia as a Property of Concepts

Boris Rusakov

August 2025

Abstract

Building on recent advances in understanding of consciousness as constituted by a set of concepts, this work proposes a formal description of this set as a *space of consciousness*. Geometrically, this is an N-dimensional vector space, where each of the N concepts corresponds to a basis vector. Algebraically, it is a space of N continuous functions, each corresponding to a concept. The concepts and the space of consciousness they constitute emerge via a phase transition of learning (or training) that is analogous to the discrete-to-continuum transition (or clustering transition) in theoretical physics. A concept is a dual elementary entity that combines meaning and quale as its intrinsic properties. Qualia are therefore fundamental intrinsic properties of the space of consciousness and serve as mediators between the concepts and their physical substrates (neurons or neural networks). Since concepts emerge via learning, both their meanings and qualia can be modified through retraining. The role of correlators of concepts, or associations, is discussed. They are the reason for the non-Abelian character of information and therefore for the emergence and irreversibility of phenomenological time.

1 Introduction

In previous publications [1, 2, 3], the functionality of consciousness was described in terms of its fundamental constituents, namely concepts [1]. The important role of concepts in consciousness has long been recognized. For example, Lewis [4] noted this within the context of the Theory of Knowledge, and long before that, Kant [5] referred to them as “minimal categories” in a similar philosophical context. The novelty of [1, 2] lies in the argument that consciousness consists entirely of concepts and operates solely with them. They are not merely scientific or educational entities as they were regarded. Every ordinary word, every phrase, every primitive notion, even an interjection, denotes a concept.

A significant part of the book [2] was devoted to discussing how concepts (and thus consciousness) might have emerged and what their physical basis is. It was suggested that the emergence of a concept must result from some phase transition.

This idea was recently explored by A. Gorsky [6], who proposed that this phase transition is the measurement-induced phase transition (MIPT) extensively studied in the context of quantum physics over the last decade (see Gorsky’s paper for references). This hypothesis seems to be quite plausible and fits the picture of the brain and its space of consciousness that we present here. According to the MIPT picture, the physical brain is a "probe", while the consciousness it produces is its "target". In this picture, the act or process of measurement constitutes probing the target by neural networks. This is a continuous activity of the brain. When the intensity of measurements (in frequency and amplitude) reaches a threshold, MIPT occurs and produces a new concept, corresponding to an instant increase in the dimension of the target space (the *space of consciousness*, see below).

The idea of explaining consciousness in terms of a phase transition has been discussed previously. See, for instance, [7], where consciousness was assumed to be a specific state of matter, or [8], which related it to self-organized criticality.

The MIPT picture resonates with the idea of [9] that consciousness may emerge as a result of the collapse of the wave function and does not exclude the scenario suggested in [10].

In addition, one should mention the numerous theories of consciousness that currently exist (see, for example, [11, 12, 13] and the references therein). A detailed evaluation of their relevance to the framework presented here lies beyond the scope of this work and is better left to the proponents of those theories.

2 Phase Transition

2.1 Animal-human phase transition

Although in [6] this transition is referred to as the "newborn-adult" phase transition, we will call it the "animal-human" phase transition. Technically, they are the same for one simple reason: the outcome of this transition is the emergence of a *concept* [1], a dual entity that is abstract in the target space and physical in the nervous system[1].

What do we know about both phases of this transition?

The animal phase, or disorder phase, is characterized by the absence of concepts. Instead, there are unrelated (discrete, isolated) facts or relational pairs: $x_1 \to a_1$, $x_2 \to a_2$, $\ldots$, $x_n \to a_n$. In the following, we will call them *proto-concepts*. In the human phase, these facts become values of a function $f(x)$ (a concept) that is defined on a continuous interval x, so that $f(x_1) = a_1$, $f(x_2) = a_2$, $\ldots$, $f(x_n) = a_n$. It is therefore the ordered phase, in the sense that the isolated facts are now unified by a function $f(x)$ that plays the role of a new dimension.

The emergence of concepts is an evolutionary necessity [2]. Although neither Nature nor Evolution (understood as survival of the fittest) has any inherent purpose, here the term 'purpose' is used for brevity to describe the results of their workings. Thus, the purpose (task) of brain activity is to support the survival and reproduction of its owner (or its owner's DNA [14]) by constantly monitoring the environment for danger, food, or mating partners.

Consequently, the task is to interpret the incoming signals and provide adequate responses. In the long term, evolution eliminates species whose interpretations and responses are inadequate. Since incoming signals can never provide a complete and defini-

[1]The true meaning of *duality* will be explained below.

tive picture, the brain can accomplish these tasks only by *guessing*.

Guessing is abstract thinking by definition, since it means creating details that were absent in the original incoming signal. The ability to guess is a crucial prerequisite for the emergence of concepts (and qualia, as will be discussed later). This ability becomes inheritable, as evolution eliminates those whose guessing leads to inadequate responses.

The hypothesis that the animal-human phase transition occurs spontaneously must be rejected, since it requires a substantial influx of energy that cannot arise without external input, despite the long-term benefits noted above.

There must be some local (short-term, immediate) reasons that can push the brain to undergo such a transition.

The first reason that comes to mind is the necessity to memorize more facts than memory can handle due to limitations of its capacity. It is much more economical to have a concept that relates seemingly unrelated facts than to memorize them as isolated. It is well established, both experientially and experimentally, that a person challenged to memorize a sequence of numbers can recall a longer sequence if able to relate the numbers and discern a pattern. There are several non-mutually exclusive models [15, 16] that successfully explain the "magic 7" limit of working memory capacity. It should be emphasized that this is a short-term, immediate, individual reason, unlike the one described before (which is beneficial for the entire species and in the long term).

Another possible reason is the influence of a strong emotion, according to the scenario described in [2][2]. In this scenario, the emotion serves as a trigger for a spike of energy influx that causes the phase transition, while the physical system is in a near-critical state. This scenario is analogous to the case of an infant, who is constantly challenged by parents and the surrounding environment, which together exert an 'artificial' and accelerated form of evolutionary pressure.

In both scenarios, the pre-human brain must be in a near-critical state, which appears to result from its constant necessity to guess.

Regardless of concrete motivations, the technical reason as suggested in [6] is intensive 'probing' ('measuring', 'learning'). The two scenarios described above are just a few among many possible reasons that push the brain to intensify the 'probing' and thus undergo MIPT.

2.2 Proto-concepts

Proto-concepts are phenomenological facts that an animal or an infant observes and memorizes. They are predecessors of concepts. Unlike concepts, they are **not abstract** entities. They are learned reactions to observed phenomena. For example, a dog recognizes its owner but does not "know" that other dogs also have owners. An infant recognizes their mother but does not yet realize that all people have mothers. This is because neither a dog nor an infant can create (or acquire) the concept of an owner or a mother. The crucial difference between a dog and an infant is that a dog will never acquire such concepts, whereas an infant can and typically will at some point (though not inevitably). When a child acquires their first concept, they create their *space of concepts*, or *space of consciousness*, and thus become a human. Some claim that animals and infants also

[2]One should note, however, that for this scenario to be realized, some basic concepts, at least 'fear' and 'suffering', must have already existed.

possess consciousness. While the terminology may be debated, it is important to note that their space of consciousness is zero-dimensional, as it consists only of discrete objects, that is, proto-concepts.

The term "proto-concept" has been used in several different contexts in cognitive science, philosophy, and developmental psychology. Apparently, it does not have a single recognized author. It was used, for example, by Gauker [17] and Prinz [18]. There are many synonyms for this word as well. For example, Peter Gärdenfors [19] uses the term "conceptual spaces". Other authors have also used the term or its synonyms in various contexts.

The proto-concepts are prerequisites for concepts. A concept *may* emerge when a group of proto-concepts (or even a single one) is recognized as sharing a common denominator, which thereby becomes the concept. For example, when a child realizes that not just he, but everyone he knows or does not know, has a mother and that even mothers have their mothers, then the notion 'mother' *may* become a concept, an abstraction, a common feature shared by all of these women. The note '*may*' is important, since a concept may or may not emerge. Having proto-concepts does not guarantee the emergence of concepts. For that to happen, there must be a phase transition.

On the other hand, a concept can emerge from a single proto-concept or without proto-concepts at all (for example, most scientific concepts do). One such scenario was proposed in [2]. The crucial point here is the emergence of a concept and thus of a space of consciousness, no matter how primitive it was initially. At the very beginning it had only one dimension, since each concept is a basis vector, and therefore the dimension of this space equals the number of concepts it contains.

It is important to emphasize that a proto-concept is not a component of a concept. A concept is neither built from nor does it consist of proto-concepts. A concept is a completely new *abstract* and *elementary* entity, a function, even though proto-concepts are its prerequisites, its animal predecessors.

2.3 Emergence of space of consciousness

Although [6] appears to correctly describe the transition in the physical brain (as MIPT), it does not adequately address its counterpart in the target space, which is the focus here.

Let us recall the earlier example of the concept of mother. The transition in the target space can be illustrated by the following two graphs:

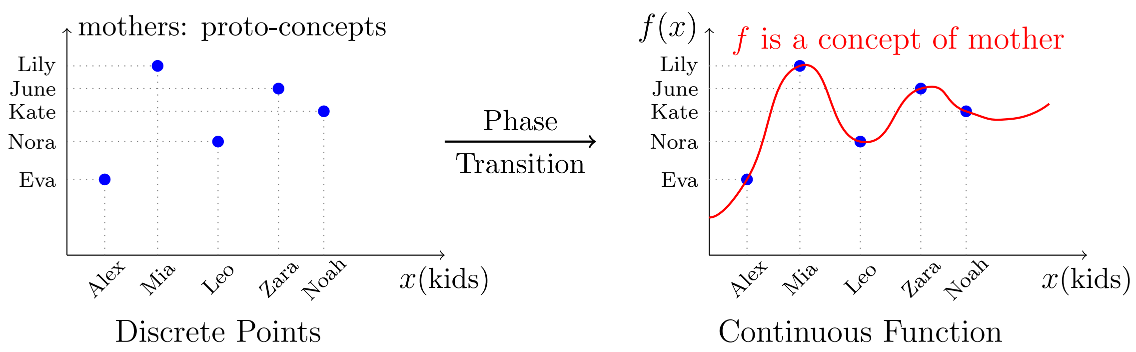

Fig. 1

We might call this transition the *phase transition of learning* (PTL). However, in order to align the name with MIPT, which is its counterpart in the physical brain, we will call it the **learning-induced phase transition** (LIPT).

On the left, we have the situation before the transition (animal phase, infant phase, or disordered phase). There are only proto-concepts, i.e. discrete observed facts. There is nothing beyond these discrete points and their corresponding values. One can imagine that this is how an infant who has not yet acquired the concept of mother sees it: Alex has Eva, Mia has Lily, etc.

On the right, there is the situation after the transition (human phase or ordered phase). Here, we have a continuous function f representing the concept of mother. Having this concept means that the child now realizes that every person, including those they do not know or will never know, has a mother. Formally, it means that there are infinitely many points, both between and beyond previously observed facts. Thus, mother becomes an abstraction, a function, i.e. a concept.

The emergence of the very first concept signifies the emergence of the **space of consciousness** to which it belongs (algebraically) and serves as a basis vector (geometrically). With each new concept, the dimensionality of this space increases by one.

From the mathematical point of view, the space X corresponding to the situation on the left is a finite set of discrete points and has dimension zero.

Algebraically, the **Space of Consciousness** is the space of N continuous functions (concepts), like the one pictured on the right side of Fig.1, each defined on its own continuous interval x of *all possible* observables. If we define N functions on this set, such that $f : X \to \mathbb{R}$, the space of all such functions is $\mathbb{R}^N$, which is a finite-dimensional vector space with $\dim(\mathbb{R}^N) = N$.

This is to some extent equivalent to quantum field theory, in which the configuration space is: $\mathcal{C} = \left\{ f : \mathbb{R}^d \to \mathbb{R} \right\}$, i.e., all functions $f(x)$, $x \in \mathbb{R}^d$. This space is infinite-dimensional because there are infinitely many degrees of freedom.

By contrast, a discrete model (e.g., the Ising model) has a finite configuration space:

$$\mathcal{C}_{\text{discrete}} = \left\{ \sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N) \,\middle|\, \sigma_i \in \{-1, +1\} \right\},$$

which has cardinality: $|\mathcal{C}_{\text{discrete}}| = 2^N$. This is a finite-dimensional space.

2.3.1 Emergence of Continuum Limit

The described scenario resembles the lattice-to-continuum transition. The lattice discretization introduces a finite lattice spacing a, and the fields are defined at the K lattice sites:

$$f_i = f(x_i), \quad i = 1, \ldots, K.$$

The configuration space is then:

$$\mathcal{C}_a = \mathbb{R}^K.$$

As $a \to 0$ and $K \to \infty$, the space becomes:

$$\lim_{a \to 0} \mathcal{C}_a = \mathcal{C} = \{ f(x) \},$$

which is infinite-dimensional.

This passage to the continuum can be regarded as a phase transition (or, more formally, the approach to a critical point):

$$\mathcal{C}_{\text{discrete}} \quad \xrightarrow[\text{continuum limit}]{a \to 0} \quad \mathcal{C}_{\text{continuous}}.$$

2.3.2 Analogy: Ising Model to ϕ^4 Theory

In the Ising model, we have spins:

$$\sigma_i \in \{-1, +1\}.$$

Near criticality, coarse-graining yields a continuous field $\phi(x)$. The effective action becomes the Landau-Ginzburg functional:

$$S[\phi] = \int d^d x \left[\frac{1}{2} \left(\nabla\phi(x)\right)^2 + \frac{r}{2}\,\phi(x)^2 + \frac{u}{4!}\,\phi(x)^4 \right].$$

The partition function is:

$$Z = \int \mathcal{D}\phi \, \exp\bigl(-S[\phi]\bigr).$$

Here, the integral is taken over all configurations in infinite-dimensional space $\mathcal{C}$.

2.3.3 Lattice-to-Continuum Phase Transition

The emergence of a continuous field can be regarded as a phase transition because:

- The degrees of freedom reorganize from finite discrete variables σ_i to continuous fields $\phi(x)$.
- The topology and dimension of the configuration space jump:

$$\dim \mathcal{C}_{\text{discrete}} = 0,$$

$$\dim \mathcal{C}_{\text{continuous}} = \infty.$$

Despite the similarity, it is important to emphasize that this transition differs from the LIPT that does not necessarily occur simply by increasing the number of discrete points (proto-concepts). For LIPT to occur, there must be an additional factor causing it, such as strong motivation, equivalent to the sufficiently high (over-the-threshold) intensity of learning. As discussed in [6], its counterpart in the brain, MIPT, occurs due to intensive measurement (probing).

Another difference is that the dimension of the space of consciousness is not ∞; it is N, the number of concepts.

At a certain critical intensity (in frequency and amplitude) of learning, the learning-induced phase transition (LIPT) occurs and produces a concept as a result.

3 Qualia

3.1 The Problem

Although we have now understood not only the role of concepts [1, 2] but also how they emerge, the qualia, which have puzzled researchers for so long (the issue is commonly attributed to Chalmers [20]), have remained a mystery. The questions are what they are, how they emerge and how they fit into the overall picture of consciousness.

Some philosophers and cognitive scientists argue that qualia are not primitive, independent entities but rather conceptual constructs (or even illusions). This view is sometimes called illusionism or representationalism [21, 22, 23, 24, 25]. Their argument is that qualia themselves are cognitive illusions. Experience consists only of representational states. If you fully specify the representational content, nothing is left out. The core claims of these views are as follows: 1) The intuition that qualia are "basic irreducible mental atoms" is itself a conceptual or representational construction, not a metaphysical fact. 2) There is no ontological qualia beyond information processing and representation. 3) Treating qualia as basic independent entities confuses feeling with the way it is structured.

In contrast, Chalmers and others [20, 26, 27] argue that qualia are a fundamental intrinsic property of consciousness.

Remarkably, both perspectives can be reconciled.

As we discussed earlier, there is no place for qualia as a separate, independent entity in the space of consciousness. Since the space of consciousness contains nothing but concepts, the qualia can belong to this space only through the concepts. Therefore, they must be intrinsically related. At least, qualia cannot be independent. If there are no concepts, there are no qualia. This matches our experience that we do not feel anything when we are unconscious.

3.2 Proto-qualia

However, animals and human infants exhibit reactions similar to ours, even though they have no concepts. Do they have feelings or senses? We know they have proto-concepts. Could it be that, just as our qualia are tied to concepts, animals and human infants possess something like *proto-qualia* tied to proto-concepts? If so, what are these and, most importantly, what is the relation?

Just as proto-concepts are isolated memorized facts, the *proto-qualia* are purely mechanical reactions to stimuli. From the outside (from a human point of view) they *look exactly* like qualia, but they are not, since they are not accompanied by subjective experience. They are programmed by evolution. In pure form, they are most clearly manifested in infants and animals. For example, when we see a dog 'joyfully' jumping at its owner, we may assume it has qualia, although in fact we are witnessing mere proto-qualia, a purely reactive behavior. 'Joyfulness' is our projection based on human experience and is misleading because the dog has no qualia, as will be explained in the next subsection.

Unlike qualia, the proto-qualia are directly observable, both visually and in neuroimaging as reactions of the brain. The key issue is when and how such reactions (proto-qualia) become accompanied by subjective experience (qualia).

3.3 What the qualia are and how they emerge

Although qualia are not independent entities and must be linked to concepts, the persistent issue remains: why and how we feel, and more importantly, what **feeling** is.

One might argue that we feel only because we believe so. If this is correct, then the question is: could the feeling be eliminated by abandoning the belief? Theoretically, the answer is yes. Practically, it is difficult but achievable.

There are numerous instances in which a person experiencing pain "forgets" about it (i.e. stops experiencing it) when their attention is diverted. There are techniques for pain relief in healing that induce another strong sensation in order to "distract" a person from experiencing pain.

The art of implementing these techniques is perfected by yogis, who train to control sensation, and by elite commandos, who learn to ignore pain.

Hence, the proposition that feeling can be eliminated by abandoning belief is not nonsensical.

The answer to the above questions becomes easier to achieve if one recognizes that each concept, in addition to its meaning (which is naturally implied, since the meaning and the concept are viewed as almost the same, and sometimes even used interchangeably as synonyms), also carries the *quale* (the feeling) associated with it.

Therefore, the answer we propose is:

A quale is an intrinsic property of a concept that connects it to its physical carrier (neural network).

In other words, a quale is a communicator of the meaning of a concept to its neural network. Meaning and quale are just two sides, or two intrinsic properties, of a concept.

Therefore: **The qualia are the fundamental intrinsic properties of the space of consciousness.**

Functionally, they serve as a tool or a mechanism by which the brain *interprets* its own state or the processes occurring within it.

The illustration of these relations can be represented by the following figure:

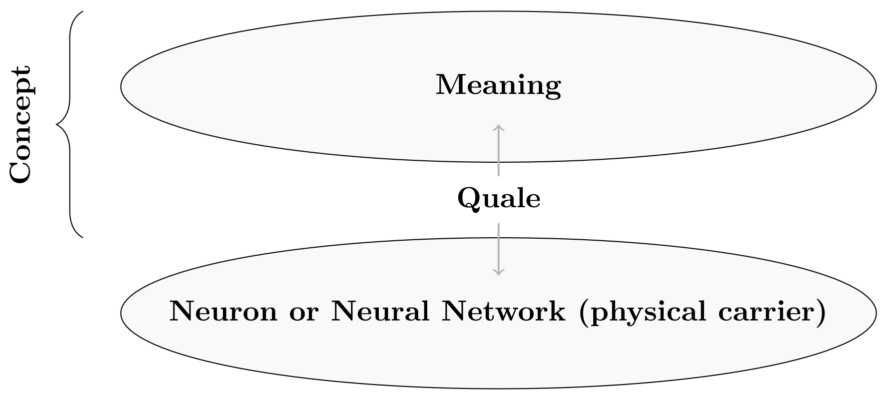

Fig. 2

We believe that this realization solves the problem and reconciles the two philosophical camps mentioned above.

For the whole brain and for its target space, the space of consciousness, the picture based on the previous can be illustrated as the following diagram:

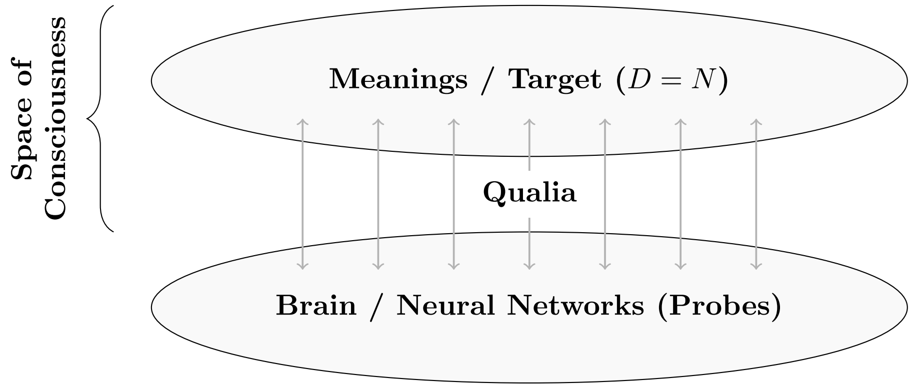

Fig. 3

Hence, consciousness does not exist without qualia. The loss of qualia entails the loss of consciousness, and vice versa.

The dimension of the space of consciousness is $D = N$, where N is the number of concepts.

Unlike qualia, the proto-qualia do not provide this communication. For an animal (or an infant), the corresponding diagram transforms into the following (notice the absence of feedback arrows):

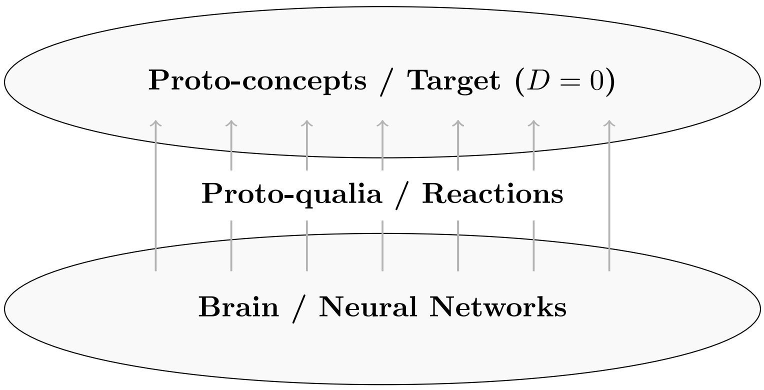

Fig. 4

As outlined before, in this case, the dimension of the target is zero because it consists of discrete points only.

Thus, qualia must be regarded as integral properties of concepts and, consequently, fundamental intrinsic properties of the space of consciousness.

As an example, consider the "feeling" of redness. There is no such property in the physical world. Then how and why do we experience it so distinctly? The answer is: 1) because we learned it, training our consciousness to perceive it in this way, and 2) because the brain is capable of doing so. It has direct access, by means of qualia, to the space of consciousness, where it can see (feel, sense, experience) all of its dimensions with their strange qualities.

It is necessary to emphasize that the ability to feel redness (or any other color) must not be confused with the ability to differentiate colors (more correctly, frequencies of the spectrum of light). Many animals possess this ability.

However, only humans possess the ability to see each frequency as a *color*. Moreover, each person experiences it uniquely and cannot experience it within the space of consciousness of another person.

Since the proto-qualia are directly observable by neuroimaging, it would be very instructive to compare them for different colors, both for animals and humans. My prediction is that the human proto-qualia for colors are not very different from the animal ones. However, qualia are not directly observable and are possessed only by humans. The indirect way to observe a quale is to activate its concept, either verbally or visually, and identify it via neuroimaging.

This is not the first time we have accepted notions that we cannot imagine, comprehend, or grasp. Eventually, we do it quite often, sometimes without even noticing. The consciousness is the cleverest illusionist in the world. Recall the so-called "optical illusions" where we "see" things that are not there and, conversely, fail to see what is in front of our noses. In addition, we still do not understand *what* many physical entities, such as electric charge, mass, or spin *are*. We simply know *what they do*, and we accept that.

Similarly, we need to accept qualia as the ***fundamental property*** of the space of consciousness. To add the similarity, we know very well *what they do*. When we are conscious, we see, smell, hear, and feel in various ways. Certainly, the feeling of numbers or other "abstract"[3] concepts is much subtler than the feeling of pain or suffering. However, they do exist, even if they are not normally noticed. Some people say that they like *blue* and dislike *orange*, while others *feel* the opposite[4]. One theoretical physicist said that he *feels* fermions differently from bosons: fermions are rough, rude, and asocial, while bosons are the opposite[5]. Later in this text, in the subsection on associations, we will explain that this analogy is perfectly reasonable, as making associations not only helps to create a concept, but also to memorize it. Which particular associations we make and how we feel a particular concept depends on how we have trained ourselves. What is remarkable is that we can be retrained.

The picture would be incomplete without revisiting the phase transition described above (LIPT), this time with respect to qualia.

Since qualia are integral properties of concepts, they emerge simultaneously as a result of the same phase transition[6].

Just as proto-concepts do not necessarily lead to concepts, the proto-qualia are not necessarily accompanied by qualia. The necessary condition for their emergence is LIPT.

3.4 Qualia and memory

Since qualia provide a direct link between concepts and the brain, they play a crucial role in how concepts are memorized and recognized by the brain.

[3] All concepts are abstract.

[4] "Like" and "dislike" signify qualia, although this explanation is unnecessary as *everything* is accompanied by qualia (as everything is a concept, and quale is its property).

[5] For those unfamiliar with these terms: fermions cannot co-exist with each other, while bosons can.

[6] This conclusion contradicts [6] where the qualia were considered separate entities resulting from another phase transition.

We can now revisit the problem discussed in [2] regarding how we remember concepts and the words or phrases that denote them. The example goes as follows: you realize that you have forgotten a word or a name and try to retrieve it. You scroll through suggestions that appear in your mind, dismissing the wrong ones until the correct one comes back to you. You unmistakably recognize it, but how? If you forgot it, how can you instantly recognize it when memory throws it at you? And how do you know all the others are wrong? I suggested that it is because we somehow remember *the feeling* that the correct one evokes. Now we understand exactly how: each concept (and therefore the word that denotes it) has a quale associated with it. The quale serves as a handle, an indicator, which we remember even if we forget the word. Thus, when the correct one appears, it feels and sounds right, unlike all others that could be close but do not *feel* quite right.

Another good example of the role of qualia in memorization is rhyme. It is well known that we memorize rhymed verses much easier than prose. The reason, as we now understand, is qualia. The rhymes (which we feel) significantly reduce the choice of candidates for retrieval from memory, and thus reduce the time and effort required for it.

Hence, one can view a quale as a unique identifier linking a concept to its neuron or neural network.

It is well-known that memorization requires multiple repetitions. This aligns with the conclusion that creating a concept requires an increase in the intensity of learning, since the best way to memorize something is to have a concept of it.

In addition, memorization requires making associations. This will be explained in the subsection on associations.

3.5 Qualia and evolution

We know that concepts and consciousness provide *Homo Falsus* [2] with a significant evolutionary advantage over animals. This is extensively addressed in many publications, including [1, 2]. But what is special about qualia, besides the fact that, as we now understand, the concept (and therefore consciousness) simply does not exist without qualia? Does the ability to feel provide an advantage over having the reactions simply "hardwired" without any subjective experience?

The answer is that having qualia prompts an immediate correction in response at the individual's level, which is a significant advantage for quickly adapting to changes, compared to animals. Qualia provide a back channel of communication with the brain, indicating whether the response to a stimulus is adequate or not. In contrast, the animals cannot change their behavior immediately. They are "programmed" by evolution and can only adapt over generations[7]. In contrast, humans have adapted quickly and survived in any environment that they encountered.

3.6 Thought experiment: how do animals see?

Answering this question may help us better understand how *humans* see the space of consciousness. To do this experiment, imagine what you would see if you lacked concepts and qualia, as is the case for animals. Just look at some random object, for example, a

[7] Historically, this was often insufficient for the survival of a species, especially in the cases of large animals with long reproductive cycles. They simply did not have enough time to adapt.

green tree, and imagine what you would see without any of the concepts, such as tree, leaves, branches, green, beauty, environment, and so on, and without any knowledge or feelings (so that you cannot enjoy its beauty or appreciate the benefits it provides). It is not difficult to realize that, in fact, you would see almost nothing. As if it does not exist. Try this exercise with other objects, and you will find that the only things an animal sees or is "looking for" (and reacting to) are those that are necessary for survival: food, danger, mating partners, and very few others. All the rest are almost irrelevant, as if they do not exist. They are not completely invisible, of course, but they mean very little. A tree could become an obstacle in pursuit or escape, so it is something to avoid, or vice versa, for some other animals, it is a shelter and retreat.

Therefore, the animal component of vision is negligible compared to what we see in the space of consciousness. Animal vision merely triggers consciousness to generate the "complete picture".

This is yet another argument supporting the assertion, made in the book [2], that consciousness and vision are one and the same.

4 Space of Consciousness and Information

4.1 Correlators of concepts – associations

The correlators of concepts and their role were discussed in [3]. The corresponding mathematical framework is briefly outlined in the next subsection. In addition, we now understand how these correlators are formed. The concepts correlate with each other through qualia that connect them to their respective neurons or neural networks, which in turn connect them via synapses, collectively forming the *connectome* [28]. Obviously, the correlation is the strongest when the same neuron participates in both networks carrying two concepts.

Here is a diagram showing how the correlators of concepts are formed:

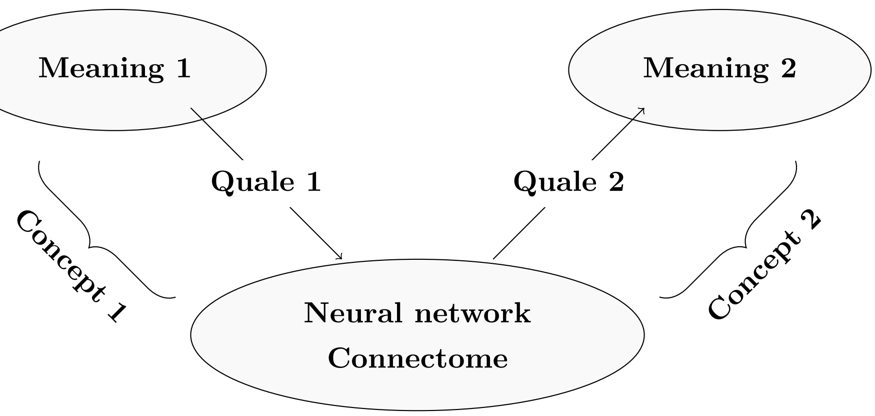

Fig. 5

Just like the entire set of neural connections in the brain forms the *connectome* [28], the complete set of concepts with their correlators forms an abstract structure that we may call *conceptome*.

The correlators of concepts are commonly referred to as *associations*. Combined together they form *the matrix of information.* As shown in [3] and in the next subsection, they are the reason why information is a non-Abelian (non-commutative) entity. The off-diagonal elements of this matrix are correlators of concepts, i.e. associations, and the diagonal elements are self-correlators. This also explains why time is unidirectional: matrices of information at different moments do not commute (see below).

4.2 Formal description

In more formal terms, the correlators of concepts are the components of the metric tensor in the N-dimensional space of consciousness.

Let N be the number of concepts. Then the information Ψ is an $N \times N$ matrix:

$$\Psi = \begin{pmatrix} \psi_{11} & \dots & \psi_{1N} \\ \vdots & \ddots & \vdots \\ \psi_{N1} & \dots & \psi_{NN} \end{pmatrix} \tag{1}$$

where each diagonal element $\psi_{\mu\mu}$ is a reaction of a concept μ on itself, and thus it is a proto-quale (animal component of consciousness), while the off-diagonal element $\psi_{\mu\nu}$ is an association, a correlator of concepts μ and ν.

One can assume, without losing generality, that all $\psi_{\mu\nu}$ are normalized by their maximum values and thus fall in the range from 0 to 1.

Empirically, the matrix $\Psi = \{\psi_{\mu\nu}\}$ exhibits a block-diagonal structure corresponding to conceptual clusters. Within a given cluster, $\psi_{\mu\nu} \approx 1$, indicating a strong intra-cluster association, whereas for concepts belonging to different clusters $\psi_{\mu\nu} \approx 0$.

All $\psi_{\mu\nu}$ are measurable observables. As suggested in [1] and [3], they can be detected and measured by neuroimaging within any given information Ψ. Having a catalog of all concepts, one can cross-reference their correlators, and thus measure and compute all elements of the matrix of information Ψ.

4.2.1 Time and operator of consciousness

Consider two different matrices of information Ψ and denote them by a parameter t as $\Psi(t_1)$ and $\Psi(t_2)$. Generally, they do not commute:

$$\Psi(t_1) \times \Psi(t_2) \neq \Psi(t_2) \times \Psi(t_1) \tag{2}$$

unless they are diagonal (Abelian: all off-diagonal elements, i.e. associations, are zeros).

Due to the qualia, we feel all the elements of matrices Ψ. Moreover, since the matrices do not commute, the parameter t is also felt, though indirectly, as a marker for each Ψ. Due to the non-commutativity, it feels unidirectional, which gives rise to the emergence of a concept of time[8].

We *feel* time because of associations[9]. We *associate* events: we remember *what* happened and *when*, or what was *before* or *after* what, unless memory fails us.

[8]Here, we certainly imply the *phenomenological time*, as we are far from explaining the emergence of *physical time* that must be a fundamental property of another space.

[9]Off-diagonal elements of Ψ.

Now, let us introduce an operator S that converts one matrix to another: $S \times \Psi(t_1) = \Psi(t_2)$. Following [3], we call S the *operator of consciousness.* Its action is:

$$S \times \Psi(t) = \Psi(t + \Delta t) \tag{3}$$

which results in

$$S = \Psi(t + \Delta t) \times \Psi^{-1}(t) \tag{4}$$

Equation (3) describes a single act of interaction. The operator S acts on the information matrix Ψ and transforms it into a new information matrix. This is exactly what we attribute to the work of consciousness. Therefore, we can truly treat the operator S as an operator of consciousness.

The formula (4) defines the operator of consciousness through the transformation of information. It also shows that the operator S is a function of time: $S = S(t)$.

One can try to obtain a dynamic equation that relates the operator of consciousness and information.

Let us recall that

$$\Psi(t + \Delta t) = \Psi(t) + \Delta t \frac{d}{dt}\Psi(t) + \frac{1}{2}(\Delta t)^2 \frac{d^2}{dt^2}\Psi(t) + \dots \tag{5}$$

In the adiabatic approximation (5) takes the Schrödinger-like form:

$$(S(t) - I) \times \Psi(t) = \Delta t \frac{d}{dt}\Psi(t) \tag{6}$$

where I is the identity matrix.

Solving (6) with respect to Ψ we arrive at the following:

$$\Psi(t) = \exp \frac{1}{\Delta t} \int dt (S(t) - I) \tag{7}$$

which thus relates the matrix of information Ψ to the operator of consciousness S.

It is interesting to note that due to the subtraction of diagonal elements $(-I)$, information is defined by associations only.

On the other hand, the matrix Ψ is the metric tensor of the N-dimensional space of consciousness. Therefore, it defines its curvature, its Ricci tensor, Ricci scalar, and the Einstein field equation on this space.

4.2.2 Partition Function of Space of Consciousness

The theoretical physics description of this space can be based on the fact that it is a space of N continuous functions f mentioned in Section 2. It is defined by the partition function:

$$Z = \int Df \ \exp(-S[f]) \tag{8}$$

where N is a number of concepts, S is an action, and the integral is taken over all functions f. The action S can be defined from the dynamic equation, such as (7), and not surprisingly coincides with the time integral of the operator of consciousness:

$$S = \int dt \ S(t) \tag{9}$$

Equation (8) can be replaced by the integral over all matrices of information:

$$Z = \int D\Psi \ \exp(-S[\Psi]) \tag{10}$$

where action $S = S[\Psi]$ is determined by the operator of consciousness as:

$$S[\Psi] = \int dt \ S[\Psi(t)] \tag{11}$$

4.3 Change of concept, reassignment, and change of association

Another issue addressed in the book [2] has become clearer now: whether a concept (and therefore not just its meaning but also its quale) can be changed. The intuitive answer is yes, but now we have also understood how.

Since the concept, its meaning, and its quale result from LIPT, they can be changed, removed, or reassigned by the same, i.e. by extensive learning, or re-training.

This concerns not only the concepts but also their correlators, and thus the matrix of information. Changing it requires considerable individual effort. In the book [2] it was assumed and demonstrated by examples[10].

We now understand why and how any concept can be changed, its meaning or quale reassigned, or removed altogether. This certainly includes the concept of pain and, remarkably, its quale and, even more importantly, its correlation with the concept of suffering. However, one has to emphasize that this is only in theory. The technique for achieving such a result is not simple. Studying yoga techniques might be a good starting point.

4.4 Matrix of information in sleep

As is well known, our senses and feelings in sleep are significantly suppressed and many are absent altogether. In terms of the present paper, this corresponds to absent or negligible qualia. Consequently, the associations become negligible or disappear altogether. Therefore, the matrix of information is nearly diagonal and thus commutative (Abelian). Consequently, time is absent or significantly distorted.

I do not mean to suggest that in sleep we are almost animals, but we resemble early humans in some respects. It should be noted that while most modern, civilizational senses and concepts are absent in sleep, the most ancient ones remain clearly present, often in an amplified or even exaggerated form. Examples include irrational fear or causeless joy. Thus, in sleep, we are still humans, although very different from our awake (and civilized) selves.

Moreover, since there are several distinct stages of sleep, I suspect that there are several different matrices of information, each corresponding to its own stage.

An important remark: Unlike the phase transition discussed here, the phase transition of awakening discussed in [29] (see also [6]) involves the *activation and retrieval* of concepts, not their *emergence* as in LIPT. Therefore, these two different transitions should not be confused with each other.

[10] Such as concepts of political views – they obviously can be changed, and we observe this change around us on a regular basis.

4.5 Concepts and education

To teach a student a new concept, an educator can only communicate its meaning. Unless the student acquires their own concept with its own quale, the concept has not truly been learned. Learning a concept means owning it, that is, feeling it. Otherwise, it would be mere memorization of facts (proto-concepts discussed above) that can be replaced by Google search and easily forgotten. Scientists know this very well: to really learn (understand) a new concept, one needs to *feel* it and *associate* it with other concepts.

Owning a concept requires substantial self-training to ensure that the phase transition discussed above actually occurs.

5 Conclusion

Thus, we have described the space of consciousness as an N-dimensional space where each of the N concepts is a basis vector. This space is a mental (or abstract) space produced by the brain, which is its physical carrier.

We have also addressed the problem of qualia. The concepts and the brain communicate via qualia. Each concept corresponds to one and only one quale, which in turn connects it to its neural network. However, that does not preclude a neural network from participating in multiple other networks and thus serving as a physical carrier of other concepts.

In addition, since neural networks are connected (forming the connectome), the concepts correlate and give rise to associations (and the conceptome).

Therefore, we have addressed the long-standing question of where we see red, smell lavender, hear Mozart, and feel pain. The answer is that these experiences occur in the person's own space of consciousness, to which the person's brain has direct access via qualia that are the intrinsic properties of this space.

Just as electric charge is the fundamental intrinsic property of the electron (and thus of the electromagnetic field) and mass is the intrinsic property of many elementary particles, qualia are the fundamental properties of concepts (and thus of the space of consciousness). The difference is that we can directly experience qualia: see, smell, etc.

Consequently, we agree with both camps in the debate on consciousness: with Dennett and others that qualia do not exist without concepts (cognition) and with Chalmers that qualia are fundamental intrinsic properties of consciousness. The antagonism is resolved by the recognition that quale is simply one side of a concept (where the other side is its meaning). They do not exist without each other and together are attributes of a concept, a fundamental elementary constituent of the space of consciousness.

Although the picture presented here removes some fog surrounding consciousness and qualia, there remain many important problems to be solved.

One example is how this space relates to the "real world", and what the latter actually means. This becomes especially demanding as we now realize that the space of consciousness emerges via the same mechanism (MIPT/LIPT) as quantum physics phenomena such as quantum entanglement.

The most intriguing question is certainly the one we cannot answer: what else consciousness is capable of, what other concepts and qualia it will produce? Whether it will eventually produce a concept or concepts that reveal to us the mystery of the "real world"? We will never know until it happens. What we could try to do is to facilitate the phase transitions. If we can do this, then perhaps we will jump to the next level of intelligence and would not even need an AI as an assistant.

On the practical side, we can now answer a question that has long concerned AI experts: whether qualia are necessary for consciousness. The answer is that they are. As we have just concluded, the concepts that constitute consciousness do not exist without qualia. Moreover, both emerge as a result of the same phase transition. Whether qualia can be programmed (hardwired) in AI is a separate question. However, one needs to understand that the phase transition does not occur spontaneously. Even having all the necessary prerequisites for it, such as proto-qualia and proto-concepts (both are relatively easy to program in AI), does not guarantee that the transition will occur. There must be external motivational factors, equivalent to natural dangers driving the animal-human transition or family pressure facilitating the newborn-adult transition.

As we now understand, neuroimaging deals with proto-qualia. In order to narrow the study to qualia, one has to invoke concepts. This can be done verbally or visually by showing the corresponding images to a person[11].

The theoretical picture presented here invites empirical validation through investigation of the neural correlators of concepts (associations) and their dynamic transformations over time. Neuroimaging techniques could be employed to measure functional connectivity patterns corresponding to correlators of concepts, revealing the structure of the conceptome. Time-resolved studies during learning or sleep could clarify the temporal evolution. This might help to achieve a comprehensive description of the information matrix Ψ and the operator of consciousness S. Additionally, interventions such as neuro-feedback or brain stimulation could probe the causal relationship between neural activity and qualia, while computational modeling in artificial intelligence may help simulate and better understand these processes.

Declaration of generative AI and AI-assisted technologies in the writing process. During the preparation of this work, the author used ChatGPT (OpenAI GPT-5) to correct language errors. After using this tool/service, the author reviewed and edited the content as needed and assumes full responsibility for the content of the publication.

References

[1] Boris Rusakov. Concepts as elementary constituents of human consciousness. *arXiv: 2208.09290*, Jul-Oct 2022.

[2] Boris Rusakov. *Homo Falsus – Lying Man: Physics, Consciousness, Language, Information, and I.* Xpertnet Inc, ISBN 979-8988291008, May 2023.

[3] Boris Rusakov. Mathematical philosophy of consciousness. *Journal of Consciousness Exploration and Research*, 15(2):179–189, 2024.

[4] Clarence Irving Lewis. *Mind and the world-order: Outline of a theory of knowledge.* Courier Corporation, 1956.

[11] Obviously, this implies that the subject does indeed possess the corresponding concept.

[5] Immanuel Kant. *Critique of Pure Reason*. Hartknoch, 1781.

[6] Alexander Gorsky. Consciousness via MIPT? *arXiv preprint arXiv:2506.04875*, 2025.

[7] Max Tegmark. Consciousness as a state of matter. *Chaos, Solitons & Fractals*, 76:238–270, 2015.

[8] P. Bak. *How Nature Works: The Science of Self-Organised Criticality*. Copernicus Press, New York, 1996.

[9] R. Penrose. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford University Press, 1989.

[10] S.Hammeroff and R.Penrose. Orchestrated reduction of quantum coherence in brain microtubules: a model for consciousness. *Toward a Science of Consciousness – The First Tucson Discussions and Debates.*, pages 507–540, Cambridge, MA: MIT Press., 1996.

[11] Giulio Tononi and Christof Koch. Consciousness: here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668):20140167, 2015.

[12] Giulio Tononi. An information integration theory of consciousness. *BMC neuroscience*, 5(1):1–22, 2004.

[13] Francis Crick and Christof Koch. A framework for consciousness. *Nature neuroscience*, 6(2):119–126, 2003.

[14] Richard Dawkins. *The Selfish Gene*. Oxford Press. ISBN 0-19-857519-X, 1976.

[15] Olga Valba and Alexander Gorsky. K-clique percolation in free association networks and the possible mechanism behind the 7 ± 2 law. *Scientific Reports*, 12:5540, 2022.

[16] Christian Bick and Mikhail Rabinovich. Dynamical origin of the effective storage capacity in the brain's working memory. *Phys. Rev. Letters*, 103(21):218101(1–4), 2009.

[17] C Gauker. Words without meaning. *MIT Press*, 2003.

[18] J Prinz. Furnishing the mind: Concepts and their perceptual basis. *MIT Press*, 2002.

[19] Peter Gärdenfors. Conceptual spaces: The geometry of thought. *MIT Press*, 2000.

[20] D. Chalmers. Facing up to the problem of consciousness. *Journal of Consciousness Studies 2: 200-19*, 1995.

[21] Daniel Dennett. *"Quining Qualia". In A. Marcel E. Bisiach (eds.), Consciousness in Contemporary Science*. Oxford University Press, 1988.

[22] Daniel Dennett. *Consciousness Explained*. Little, Brown, 1991.

[23] K. Frankish. Illusionism as a theory of consciousness. *Journal of Consciousness Studies, 23(11–12), 11–39*, 2016.

[24] M. Tye. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press, 1995.

[25] T. Metzinger. *Being No One: The Self-Model Theory of Subjectivity*. MIT Press, 2003.

[26] T. Nagel. What is it like to be a bat? *The Philosophical Review*, 83, No. 4:435–450, 1974.

[27] F. Jackson. Epiphenomenal qualia. *The Philosophical Quarterly*, 32, No. 127:127–136, 1982.

[28] S.Seung. *Connectome: How the Brain's Wiring Makes Us Who We Are*. MIT press, 2013.

[29] Alexander Gorsky. Page time and the order parameter for a consciousness state. *arXiv preprint arXiv:2212.10602*, 2022.