Qualia and the Brain

-synopsis- with 7 Figures and 2 tables

Ken Mogi

Sony Computer Science Laboratory Inc. Takanawa Muse Bldg. 3-14-13, Higashigotanda Shinagawa-ku, Tokyo, 141-0022 Japan Tel +81-3-5448-4380 Fax +81-3-5448-4273 e-mail kenmogi@csl.sony.co.jp http://www.csl.sony.co.jp/person/kenmog.html (Office) http://www.qualia-manifesto.com (The Qualia Movement)

The concept of qualia describes the unique properties that accompany our senses. It is an essential concept when we try to understand the principle that bridges the neural firings in our brain and our perception. The idea of qualia is also of crucial importance when we try to study the functions of the brain from an objective point of view. Qualia must be part of the mathematical formulation of information we use to understand the function of the brain.

1. Introduction

We perceive the world through our senses. If you go out in the garden on a summer morning and look up into the sky, the blackish blue color comes into your mind. When you see morning dew on the leaves, you become aware of their round shape and the sheen on their surface. Putting your finger on the dew, you feel a sensation of coolness on the skin. Sniff the leaves, and you smell the distinctive scent of the plant. Then you realize that all the while, you have been bombarded with the auditory sensation of the sparrow singing on the roof...The characteristic qualities that accompany these sensations have been called "qualia". It is a patent fact that we perceive the world through a collection of qualia. Any attempt to understand the relation between the mind and the brain would be impossible without incorporating qualia in its framework.

On the other hand, we find the brain to be a tremendously complex and well-made molecular machine. The action potential consists of the inward flow of Na^+ ion through the membrane. At the synapses, neurotransmitters such as glutamate and GABA are released into the synaptic cleft. The substances produced near the are transported to the synapses on the track of nucleus microtubules that extend through the axon. The gene regulation nucleus plays an essential role in in the functional the specification of the neurons, and is believed to play a role in the formation of long term memory. An orchestrated system of biomolecules of unimaginable complexity underlies the neural activity in the brain.

The various sensations that occur in our mind are accompanied by various qualia. In contrast, the brain is a complex molecular machine. How are these two worlds related to each other? What is the neural correlate of qualia? This is the central question of the so-called mind-body problem.

The question of qualia is not a solely philosophical one. The origin of qualia, ultimately, should be regarded as part of the natural law, just as the laws of physics or chemistry. Even from the functionalist's point of view, qualia would prove essential in understanding the principle of information processing in the brain, as I argue below. For example, one cannot hope to construct a computer that thinks like a human unless we implement qualia. Qualia should be an integral part of the

2

concept of information when we try to describe the function of the brain.

Despite its crucial importance in understanding the brain, the enigma of qualia is a deep one. How profoundly difficult the problem of qualia is testified by the words of Francis Crick, in his book *The Astonishing Hypothesis*.

The reader might accept all this but could well complain that I have talked all around the topic of consciousness, with more speculation than hard facts, and have avoided what, in the long run, is the most puzzling problem of all. I have said almost nothing about qualia--the redness of red--except to brush it to one side and hope for the best.

The quest for a solution of the enigma of qualia, the most difficult and important scientific puzzle that confronts humanity, is just beginning.

2. The neuron doctrine in perception

One of the central questions that concerns us is to understand how our perception is related to the neural firings in the brain. In 1972, Horace Barlow of Cambridge University applied the neuron doctrine to the problem of perception. Barlow suggested that the characteristics of our perception is specified by the nature of neural firings in our brain only. That is to say, no matter what mechanisms are involved, the properties of our mind is to be determined by the neural firing state only. This seems to be a reasonably sound starting point, judging from the experimental clues available today.

"Response selectivity" is a concept of central importance in neurophysiology. For example, in the primary visual cortex (V1), we find neurons that selectively responds to a bar with a certain orientation. In areas MT, V4, IT, we find neurons that respond to motion, color (in the context of "color constancy"), and form. As we go to the higher visual area, we find neurons with more complex response selectivities, and larger receptive fields. The objective of the single unit recording study can be defined as the determination of the response selectivity of a particular neuron.

One idea that emerges is to assume that when a neuron with a response selectivity to a particular visual feature fires, the perception of that feature occurs. For example, when a neuron selectively responsive to a bar slanted by 45 degrees to the right fires, the perception of the slanted bar would be incurred. When a neuron selectively responsive to a "face" fires in area IT, the perception of a "face" is invoked, and so on.

Unfortunately, there is a fundamental flaw in this line of argument, which becomes apparent when one tries to answer the following question. "When a neuron selectively responsive to a feature A fires, how does the brain (or the subject) know that it is selectively responsive to feature A?" The visual feature space is vast and complex. The fact that a neuron fires vigorously to a particular feature A does not necessarily mean that the neuron has a response selectivity to feature A only. In fact, in order to confirm the response selectivity of a neuron, every possible visual feature should, in principle, be presented to the neuron. Of course this is impossible. Moreover, our perception is constructed based on the neural firings at a particular psychological moment. (This is what the neuron doctrine dictates). It is impossible, just based on the firing of the neurons at a particular time, to establish the response selectivity of a a particular neuron. The reason for this being that response selectivity is based on the idea of an "ensemble", a set of stimulus-response relations which is at the heart of an statistical approach.

In general, as we go higher up the visual system, the more difficult it becomes to define the response selectivity of a particular neuron in an operational manner. This contradicts the idea that the neurons in the higher visual areas play a crucial role in our perception through their response selectivity. Even in the case where our perception is evoked by the successive neural firings from the lower visual areas to the higher visual areas, the significance of the neurons in higher visual area becomes obscure, if indeed the response selectivity plays an essential role in perception.

From these considerations, I conclude that response selectivity cannot be the foundation for the relation between neural firing and perception. Some other bridging principle(s) should come into the picture.

3. Mach's principle in Perception

Ernst Mach (1838-1916) was a physicist, philosopher, psychologist who had a great influence on Albert Einstein in his development of the theory of relativity. "Mach's principle" states that the mass of a particle is determined by its relation to all the other particles in the universe. If there was only one particle in the universe, it is meaningless to question how large its mass is. In general, the properties of an individual is determined by its relation to other individuals in the system.

A similar line of thought is relevant when we consider how our perception is formed through the neural firings in our brain. Namely, a neural firing plays a particular role in our perception, not because it is selectively responsive to a visual feature, but because the neural firing is related to other neural firings in the brain in such a way that the particular role in perception is endowed on the neural firing in question. We shall call this idea "Mach's principle in perception".

For example, suppose a neuron in area IT fired, and a perception of "rose" is invoked in our mind. In this case, perception of the "rose" is invoked not because the neuron selectively responds to the presentation of a rose, but because the neural firing in question is endowed with the property of "rose" through its relation to other neural firings in the brain. Specifically, the cluster of neural firings connected by interaction through action potentials that is initiated in area V1 and leads up to the neural firing in area IT codes the perception of "rose". It is meaningless to consider a single neural firing in isolation and assess its significance in perception, even if the response selectivity of that neuron could be established unequivocally.

In this picture, a percept (element of perception) is coded not by a single neural firing, but by a cluster of interactionconnected neural firings. This is the bridging principle that is consistent with the neuron doctrine in perception, and Mach's principle in perception. A percept is coded non-locally. Neural firings in spatially distant areas of the brain are integrated into a cluster through their mutual interactions, and form a percept. The distinction between the excitatory and inhibitory connections now becomes important. Specifically, it appears that only excitatory connections are included *explicitly* in the cluster of neural firings that forms a percept. Inhibitory connections influence the formation of percepts indirectly. For example, in the color constancy mechanism, inhibitory inputs from surrounds will lead to a non-formation of a percept of a color in the center. Inhibitory connections have significance in that they can "veto" the formation of a percept. However, inhibitory connections are not included explicitly in the cluster of neural firings that forms a percept. In order to see the intuitive meaning of this arrangement, consider a white bar in a black surround. In order that the white bar is a bar, it is necessary that the area surrounding the bar is black, rather than white. If the surround was white, then the bar would not be a bar! So the surround contributes to the formation of a white bar by *not* being white. However, the black surround does not constitute an *explicit* part of the percept "white bar".

The above picture gives some hints to the solution of some long-standing problems in visual perception. One of them is the so-called "binding problem". The various visual features, such as form, motion, color, and texture are represented in spatially separate cortical areas. Yet we are able to perceive these features as integrated into one visual object. For example, even though the features "rose", "moving rightward", "red", "velvety" are coded in areas IT, MT, V4, and V2 respectively, we have the integrated perception of "red velvety rose moving rightward". Several solutions, including the ideas "convergence zone" and "synchronous firing" are being proposed. However, none of them are immune from some fundamental difficulties. In order to arrive at a rudimentary solution of the binding problem, we need to go back to the very foundations of the neural correlates of perception. Namely, we need to adopt the idea that a percept is to be represented as a cluster of interaction-connected neural firings, which are spatially distributed among the various cortical areas.

Now suppose the percept "red" is invoked by a cluster of neural firings that extends from area V1 up to area V4. Under such circumstances, the qualia of "redness of red", whatever we may name it, should be specified by the spatio-temporal pattern of the neural firings which belong to the cluster, and by nothing else. This is a direct consequence of our assumption that the nature of our perception is determined by the characteristics of the neural firings that sustain it (the neuron doctrine in perception). Namely, the picture emerges that the correspondence between a particular pattern of cluster of neural firings, and the qualia invoked by it, is given in an *a priori* manner. Namely, we assume that the correspondence between the neural firing patterns and qualia is one to one, and there is no arbitrariness involved in it. Of course, one could argue that there is no guarantee that the correspondence is one to one. But then there is also no plausible reason to assume otherwise.

This above idea is of a fundamental importance. A natural law in the conventional sense is complete if it succeeds in predicting the spatio-temporal pattern of neural firings in the brain with a full precision. It is yet another type of natural law, which determines what kind of qualia is invoked in our perception once the spatio-temporal pattern of neural firings in the brain is given. As the correspondence between the firing pattern and qualia is one to one, it should be regarded as a kind of natural law, although of a rather different character from the conventional natural laws.

4. Interaction Simultaneity

Our perception is organized under a certain spatio-temporal order. For example, the space-time structure in which our visual perception occurs is different from that in which our auditory perception occurs. In audition, we don't have the massive parallel processing organized in an spatial order as in the case of vision. The "modality" of senses are marked by differences in the spatiotemporal structures that characterize them. Such differences between the modalities are usually taken for granted, and neural mechanisms of perception is discussed on top of them. That may be effective as a phenomenological and intermediate method of description. However, ultimately, the difference between the sensory modalities should be derived, from the first principles, based on the nature of the spatio-temporal patterns of the neural firings that sustain vision, audition, taste, smell, and touch.

How is our psychological time constructed? Albert Einstein, in his first paper of relativity theory published in 1905, stated

One thing should be remarked here. Such a mathematical description is physically meaningless unless the way we construct time is made clear. All our judgments about time is one about events that occur *simultaneously*.

What kind of mechanism determines the nature of simultaneity in our mind, namely the nature of psychological "now"?

Let us start from the neuron doctrine in perception. Namely, we assume that knowledge about the firing neurons is necessary and sufficient to determine the content of perception. The concept of "interaction simultaneity" dictates how to determine the nature of psychological time in a way consistent with the neuron doctrine. Under the principle of interaction simultaneity, when a neural firing and another neural firing are connected by interaction (*i.e.*, action potential propagation and subsequent synaptic interaction), these are considered to be simultaneous events. It takes a finite length of time (say 5 milliseconds) for the effect of a neural firing to propagate to a postsynaptic neuron. However, under the principle of interaction simultaneity, these firing events should be regarded as "simultaneous". We call the time parameter thus constructed "proper time" and write it as .

When the presynaptic neuron fires at time t, and the postsynaptic neuron fires at time t+t (as a result in part of the EPSP caused by the presynaptic firing), we assign the same proper time to the pre- and postsynaptic events.

Interaction simultaneity is derived from a more fundamental principle, that of "causality". Here, "causality" is taken to mean that given the state of the system at proper time , we are able to derive the state of the system at a slightly later time + . We need to use the proper time in order to describe the dynamical evolution of the neural network in a causal way.

Interaction simultaneity is based on the idea that if we are to derive the properties of our perception from neural firings, we should not adopt the position where we "observe" the neural firings from "outside the brain". If we observe the brain from outside, we can describe the dynamical evolution of the neural network with any desired temporal accuracy. We may, for example, describe the release and diffusion of neurotransmitters at synapses with submillisecond temporal resolution. However, under the neuron doctrine in perception, only the neural firing enter explicitly in our perception. Therefore, properties of our perception should be obtained without resorting to the idea of an outside observer. Thus, it becomes necessary to adopt the principle of interaction simultaneity.

We can obtain some interesting conclusions about the nature of psychological time. Firstly, the psychological "present" has a finite duration, when measured by the physical time t. The duration corresponds to the transmission delay present when the cluster of interaction-connected neural firings is formed. This could be of the order of ~ 50 ms. In other words, there is a "unit" of the psychological time, with a duration of ~50ms. Despite the existence of such a finite duration of the "moment", the flow of psychological time is shown to be smooth. Specifically, the displacement between the adjacent "moments" can be made arbitrarily small. This in turn mean that there is an "overlap" between adjacent moments of psychological time. A particular neural firing is shared by the neighboring moments of psychological time.

Libet reported that in order that we become conscious of a percept, neural firing of at least 500 milliseconds is required. Such a property of conscious perception may ultimately be linked to the idea outlined above.

5. Causality and Twistor

Interaction simultaneity is concerned with the construction of the psychological time. This in turn is part of a more general problem, namely, how our perceptual space-time structure emerges from the neural firings in our brain. "Causality" is expected to be a leading principle here. Namely, our perceptual space-time structure is constructed in such a way that within that framework, it becomes possible to describe the dynamical evolution of the neural network in a causal way.

Penrose's "twistor" is a hint for the mathematical structure to be developed from such an approach. Penrose constructs a "twistor space" separately from the physical space-time. The trajectory of light is represented as a straight line in physical space. In twistor space, it is mapped to a point. The trajectory of light represents the world-line along which causal interaction propagates. In twistor space, the set of points in physical spacetime that are connected by the interaction (which is represented by the trajectory of light) is mapped into a point. The twistor space is in a sense a more fundamental framework for natural law than physical space-time itself. Penrose writes

We should think of twistor space as the space in terms of which we should describe physics.

We cannot apply the twistor formalism directly to neural network. However, the basic idea of the twistor approach, namely to regard the causal relation between the individuals as more fundamental than the individuals themselves, corresponds to our arrangement of regarding the cluster of interactionconnected neural firings as the percept, that is, the elementary unit in perception. The nature of our perception is ultimately determined by the dynamics of the neural network. Accordingly, the cluster of interaction-connected neural firings functions not only as an element of perception, but also as an element in the dynamics of neural network. In order to describe such a dynamics, we would need a mathematical structure similar to that of twistor space. The twistor formalism is consistent with the principle of interaction simultaneity outlined in the previous section, and is indeed a natural mathematical embodiment of the principle.

An intriguing possibility is that a twistor-like space can be constructed to describe the dynamics of a neural network, and the space thus constructed corresponds to our perceptual space-time. If such a picture is found to be the case, our mind would inhabit the twistor-like space that describes the dynamics of neural networks. Of course, at present this is merely a conjecture.

6. New Concept of Information.

Shannon opened up the new field of information theory in his seminal 1948 paper. In this paper, he states as follows.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of the design.

Information theory that was developed following Shannon's initiative is mathematically a subset of probability theory. The relevance of probability theory in the elucidation of the mechanisms of perception is a limited one. We have seen that response selectivity, which is based on the concept of "ensemble", cannot be the foundation for the neural correlates of perception. By the same token, information theory, or probability theory in general, cannot explain the neural origin of perception *per se*. Shannon's approach is relevant when we consider such problems as the efficient coding of message sent through a channel where a certain level of noise is present. Such a question is precisely where probability theory finds its powerful application. In the brain, information theoretic approach in the above sense would be relevant to the analysis of peripheral

13

receptors, such as the retina. However, information theory, or probability theory in general, has limited relevance, if at all, to the more central information processing, where our perception and thinking take place.

If any definition of information is to be relevant to the function of the brain, it should reflect the dynamics of the neural network. It is not sufficient to treat the spike trains as an abstract collection of bits. We should start from the effect that a particular firing pattern has on the postsynaptic neurons. We should ask what kind of postsynaptic firing pattern is generated as a result. In other words, the concept of information should be "embedded" in the dynamics of neural networks.

I conjecture that qualia should be included in the mathematical formula of information as is relevant to the functions of the brain. This conjecture is justified in view of the important role played by qualia in the cerebral information processing.

For example, we can perceive "a red velvety rose" as an object located in a position in the visual field. How the brain does it is the so-called "binding problem". As far as we are concerned, this integration is possible, because each visual feature "red", "velvety", "rose" has distinct qualia, and there is no possibility of mixing them up. Suppose these features were represented by abstract numbers, such as "3", "7", "-1". Then it would be quite difficult, if not impossible, to "integrate" them in the unity of a single object!

The neural mechanism underlying the generation of the meaning of words is expected to be similar to the mechanism underlying the generation of percept accompanied by qualia. Namely, the meaning of words are determined by the relation between the neural firings in Wernicke's area, which is supposed to be responsible for semantic processing.

Under usual circumstances, we can make a distinction between our perception of an external object and the mental imaging of an object. Also, it is easy to separate the percept corresponding to an object that is "here and now" from the object encountered in the past. memory of an Such differentiations of percepts may seem self-evident. However, it is useful to remember that all these percepts are invoked by neural firings, and all neural firings are basically the same physical phenomenon, no matter what aspects of our perception they might be coding for. Significantly, the differentiation of percepts stated above appears as the difference in qualia associated with respective percepts. A percept corresponding to an object "here and now" is accompanied by a very vivid, conspicuous qualia. In contrast, when we remember an object that we encountered in the past, the accompanying qualia is of a more abstract, subdued nature. We tend to take such distinctions for granted. At the end of the day, these percept are invoked by neural firings in the brain. How are the remarkably different qualia generated from the collection of neural firings in the brain? How is the structuring of our perception, such as "inside the self" vs "outside the self", "present" vs "past" made possible? This is a question central to the remarkable information processing capability that is exhibited by our brain.

7. The Future of Qualia Research

Qualia are unique properties of our perception. It is difficult to convey to others how we feel when we munch a peach. However, beyond any reasonable doubt, qualia accompanying our senses are caused by the neural firings in the brain. Our chance of studying qualia in an objective manner lies in the relation between qualia and the neural firings. Qualia is a crucial concept when we try to understand the nature of information processing in our brain. If we are to understand the nature of processes occurring in the brain, we must construct a new concept of information, something that is different from Shannon's statistical construction of information. The concept of information to come should be embedded in the dynamics of neural network, reflect the interaction between the neurons, and include qualia as a natural ingredient of its mathematical structure.

Of course, no matter how the objective understanding of the principle of information processing in the brain is advanced, the mystery of our perception, and the qualia accompanying it, is likely to remain. The fundamental enigma, namely, the origin of self-consciousness and the uniqueness of personality will probably never be dissolved. However, we may say without much venturing that the time has come, when qualia are to be liberated from the studies of philosophers, and to become a subject of an empirical science. The difficulties that we face in the binding problem, and the neural mechanism of the semantics of language, cannot be dissolved without tackling the neurophysiological mechanism underlying qualia.

Qualia are central to our research program to study the relation between mind and brain as a field of empirical science. The experimental and theoretical study of qualia is too important to be ignored. As I close this article, it is fitting to refer to the words of Michael Faraday, whose pioneering work opened a whole new world of electromagnetism.

Nothing is too wonderful to be true, if it be consistent with the laws of nature.

Acknowledgements

I thank Horace Barlow for critical reading of the manuscript.

References

Barlow. H. B. (1972) Single units, sensation: a neuron doctrine for perceptual psychology? *Perception* **1**, 371-394.

Chalmers, D. The Puzzle of Conscious Experience. Scientific American, December, (1995)

Crick, F. (1994). in *The Astonishing Hypothesis*. (Simon and Schuster)

Einstein, A. (1905) Zur Elektodynamik bewegter Körper. Ann. der Phys. 17, 891-921.

Libet, B. (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav. Brain Sci.* **8**, 529-566

Penrose, R. (1994) in *Shadows of the Mind*. Oxford University Press, and (1989) in *Emperor's New Mind*. Oxford University Press.

Shannon, C.E. A mathematical theory of communication. *Bell System Technical Journal* **27**, 379-423, 623-656 (1948)



Neural Firing in the Brain

Perception

Figure 1

The neuron doctrine in perception (Barlow 1972) states that our perception is determined by, and only by, the pattern of neural firings in the brain. From this view point, the core of the mind-brain problem lies in finding the correspondence principle between the neural firings in the brain and our perception.



Visual Field

Figure 2

The various visual features are coded in spatially separate areas of the cortex. For example, color is coded in area V4, texture in area V2, and motion in area MT, and the form in area IT. We have the integrated perception such as "a red flower with fluffy texture in the middle is moving to the right". In order to achieve this, the various visual features represented in the cortex should be somehow bound. This is the so-called "binding problem".



(a) Convergence Zone

(b) Synchonous firing

Figure 3

Two schemes are at present proposed as the solution to the binding problem. In "convergence zone" theory, the visual features represented in multiple areas of the cortex are "integrated" in the "convergence zone" somewhere in the cortex. One obstacle stands in the way of implementing this idea. The problem of combinatorial explosion, namely, the possibility of combination of visual features leading to the shortage of coding capacity in the brain (sometimes called the yellow Volkswagen problem) should be accounted for. The other idea, that the synchronous firings of neurons in the cortex code for the visual features bound into one object, is not without its own difficulties. Among the difficulties are the ambiguity associated with the definition of synchronicity in a neural network where the time required for transmission of action potentials is finite. Another difficulty is how the visual features could be bound when there are multitude of objects within a small visual area, as in the case of marbles with various colors scattered over the visual field.



Figure 4

The apparent distinction between the modalities of senses, such as vision, audition, taste, touch, and smell is one of the most remarkable properties of our perception. The unique nature of qualia associated to each modality, and the spatio-temporal structures that accompany them, should ultimately be explained by the nature of the neural firings in the respective cortical areas that sustain each modality.

The Neuron Doctrine of Perception	Our perception is directly invoked by the neural firings in the brain. A non-firing neuron is as good as non- existent as far as perception is concerned. The characteristics of our perception should be explained by the nature of neural firings only.
Mach's Principle in Perception	In perception, the significance of a firing neuron is determined by its relation to other firing neurons at that psychological moment. (Response selectivity cannot be the ultimate bridging principle between the neural firings and perception.)
Principle of Interaction Simultaneity	The neural firings connected by interaction are simultaneous. There is no passage of proper time along the world line of interaction.
A priori correspondence of qualia	Interaction-connected cluster of neural firings constitute a percept. To a particular pattern of neural firing, a particular qualia corresponds. In other words, there is a one to one correspondence between neural firing pattern and qualia in an <i>a priori</i> manner.

Table 1

Basic assumptions



Figure 5

Under the principle of interaction simultaneity, the neural firings connected by interaction via the synapse are considered to be simultaneous. The time parameter thus constructed is called the proper time. In reference to the firing event of a particular neuron A, the presynaptic firings that produced EPSP in neuron A, and the postsynaptic firings that are influenced by the firing of neuron A, are considered to be simultaneous. The principle of interaction simultaneity is derived from the assumption of causality.



Fig.6

The twistor space is constructed on the basis of the causal relation between points in the physical space-time, and is more fundamental to the natural law than the physical space-time itself. In a similar manner, the perceptual space-time is constructed from the interaction of neural firings based on the causal relations.



Figure 7

Qualia plays an essential role in the structuring of our perception. For example, vivid qualia accompany our perception when we are looking at an object which is outside the "self" at present. In contrast, less vivid qualia accompany our perception of something within the "self" (i.e., mental imagery) or in the past (i.e., memory). Qualia thus reflect the structuring of information processing in our brain within the context of topological distinction between self and non-self, and the temporal distinction of present and past.

	Definition of Information	Foundation of Perception
Conventional Concept of Information (Statistical Picture)	Shannon type	Response Selectivity
New Concept of Information (Interaction Picture)	Unknown (Twistor-like mathematical framework?)	Interaction-connected neural firings (qualia)

Table 2

Conventional and New concept of Information