Title: How Consciousness Emerges from Ions

Authors: Peilei Liu[1]*, Ting Wang[1]

Affiliations:

[1]College of Computer, National University of Defense Technology, 410073 Changsha, Hunan, China.

*Correspondence to: plliu@nudt.edu.cn

Abstract: As Francis Crick said, neuroscience is a data rich but theory poor field, and it is missing a broad framework as in physics. We wish to put forward such a unified framework based on existing evidences. Unexpectedly, it is a very simple statistical model. Specifically, we find that neural mechanisms in the spatial and temporal dimensionalities follow similar statistical laws. And they are usually called neural coding and memory respectively. Moreover, memory can be divided into two types: long-term and short-term (or instantaneous). The instantaneous memory is the foundation of consciousness according to Crick. Then we indicate the physical and biological mechanisms behind these statistical laws. In general, they actually reflect random processes of particles such as ions. Detailed model and supporting evidences can be found in our previous work. And this simple model is really powerful in explaining most psychological phenomenon and advanced intelligence such as language.

Main Text: There are four greatest mysteries in the nature: universe, material, life, and consciousness. The former three correspond to relativity theory, quantum theory, and the double helix structure of DNA respectively. Until now however, the consciousness is still far from completely understanding (*1-3*). It is viewed as one of the most complex problem. In fact, many physicists ignore modern neurobiology when building models and some of them even hold the dualistic view (*3*). However, we believe that all biological organisms and neural mechanism must obey physical laws. On the other hand, most neurobiologists revel in producing experimental data instead of creating a systematic theory. Therefore there is still a gap between these two fields.

Consciousness has various meanings in different disciplines. This paper mainly discusses the neural mechanism correlative with consciousness, namely what happens when you think of an object. This can be viewed as a neural coding problem in essence. According to quantum physicist Erwin Schrödinger, neural coding should follow statistical laws as in quantum physics, because it contains too few particles for expressing “precise” physical laws (*1*). Computer scientist John Von Neumann also conjectured that neural coding should be statistical, in contrast to the precise coding in computer (*4*). However, specific statistical laws or principles are still unclear. Based on existing evidences, we have fortunately found these statistical laws and their neural implementations. They are unexpectedly simple but really powerful meanwhile. All c_i in this paper are constants, and they have different meanings in different paragraphs.

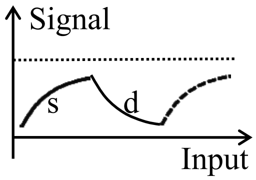

Fig. 1. Dorsal-fin curve. Rising phase s and falling phase d represent exponential summarization and exponential decay respectively. And they are approximately axisymmetric. Both of them have statistical implications. Specifically, if objects and events exist in continuous local space-time, their probability estimates will follow the dorsal-fin curve. On the other hand, this curve can be easily implemented in physics. Specifically, it is actually the reflection of random processes.

As shown in Fig. 1, the dorsal-fin-like curve widely exists in neurons and synapses (*5-7*). Examples include spike, EPSP (excitory postsynaptic potential), fatigue, adaption, forgetting, lateral inhibition, and so on. The dorsal-fin curve is composed of rising phase $F(x) = c_1(1-e^{-c_2 x})$ and falling phase $F(x) = c_3 e^{-c_4 x}$. In this paper, they are called exponential summarization and exponential decay respectively. Both of them have implications in statistics and can be implemented easily in physics and biology. Suppose that an object has many attributes, and only x of them is visible (see Fig. 2). On this condition, what's the probability of this object occurring? It can be inferred that the probability P(O)=P(O_1+O_2+…+ O_x)=1-P(¬O_1¬O_2…¬O_x)=$1-\prod_{i=1}^{x} q_i$ = $1-e^{-c_5\sigma}$, Where P(O_i)=P(O|A_i), O and A_i means object and attributes respectively, $\sigma = \sum_{i=1}^{x} |\log(q_i)|$, $q_i = P(\neg O_i)$. Therefore the ring phase of dorsal-fin curve can represent the probability estimate of an object occurring based on current visible features.

This can be easily implemented in physics and biology (*5, 8*). Specifically, exponential summarization can be transformed as $\frac{dy}{dx} = \mathrm{c}_1 - c_2 y$. Namely it is actually determined by two synchronous processes: linearly increase and exponential decay. And exponential decay could be the reflection of a simple random process. For example, if every atom decays randomly, the changes of radioactive materials follow exponential decay curve. The molecule mechanism of neural coding could follow similar mechanism (*5, 8*). Specifically, a neuron is like a leaky pool with continuous injection of multiple water pipes. And the leaky rate is proportional to the water amount itself. Then the relation between water amount and the number of water pipes follows the function of exponential summarization. Namely $f = c_1(1-e^{-c_3\sigma})$, where $\sigma = \sum_i f_i$, f and f_i are the spike frequencies of output and inputs respectively. In this case, a neuron's axon and dendrites can represent the occurring probabilities of an object and its attributes respectively (*8*).

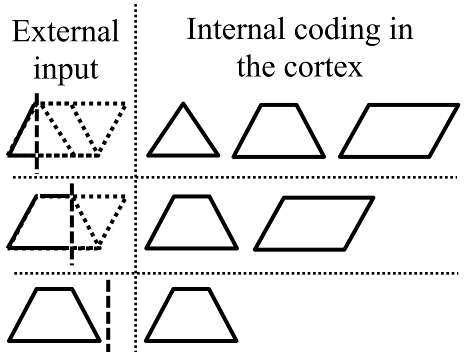

Fig. 2. Neural coding. Shapes on the left mean external inputs, and the dotted lines are invisible. Shapes on the right mean coding neurons. Based on limited information, many neurons will be fired. They compete with each other and only winners can strengthen their dendritic connections towards the input. As in the bottom row, if all information is available, there will be only one winner finally. And this winner neuron will fit the input better and better with the synaptic growth. This is the “grandmother cells” coding in essence.

As shown in Fig. 2, for a specific input, many neurons will have their own probability estimates. Therefore there are competitions between them. Suppose x kids saw a shape, and each of them gives a different answer such as trapezoid, parallelogram, triangle (see Fig. 2). In this case, who should you believe without having seen this shape? The best choice is the one with the highest IQ (intelligence quotient) or statistical confidence. However, how is the confidence or reliability of this choice influenced by the number x? In fact, it can be inferred that $q_i' = q_i e^{-c_1\sigma}$, where $\sigma = \sum_{j=1, j\neq i}^{x} |\log(q_i)|$, q_i are their initial confidences. Namely q_i' is actually negative correlative with x, and specially $q_i'=0$ when $x\to\infty$. Generally speaking, competitions between matched answers will result in dropping of every answer’s reliability. And this dropping follows exponential decay namely the falling phase of the dorsal-fin curve (see Fig. 1). This is like the famous watch law: it is hard to know the accurate time with many inconsistent clocks. Moreover, this conclusion can be extended to more general cases. For example, it could be allowed that partial kids give the same answer. In this case, the confidence of an answer is the exponential summarization of these kids’ confidences $q_i = (1 - e^{-c_1\alpha})$, where $\alpha = \sum_k |\log(q_{ik})|$. In most cases however, competitors or answers aren’t matched. On this condition, the strongest answer will enlarge the gap and result in “winner-take-all” finally. Specifically, the confidences of kids who giving the right answer will be increased, and this will bring them competitive advantage reversely in next time. Due to this “rich-get-richer” mechanism, kids will gradually become specialists good at detecting specific shapes with this game continuing. In other words, ultimately there will be a “one-to-one” or “mult-to-one” mapping from shapes to kids. This in essence is the “grandmother cell” coding (*9*).

Similarly, this process can be easily implemented by neurons and synapses (*8, 10*). Let a neuron represent an answer and its firing frequency represent confidence of this answer. Then corresponding kids can represent dendrites of this neuron. Then the competition can be implemented through lateral inhibition. Specifically, the inhibition from all revivals of neuron i follows exponential summarization $h_i = c_1(1-e^{-c_2\sigma})$ where $\sigma = \sum_{j\neq i} d_j$. And then the neuron's actual potential should be $p_i' = p_i - h_i = c_1 e^{-c_2\sigma} + c_3$, where $p_i = c_1 + c_3$. When $c_3 = 0$, $p_i' = c_1 e^{-c_2\sigma}$. Therefore this lateral inhibition actually implements the competitions mentioned above and has similar statistical meanings. In facts, this lateral inhibition is implemented through inhibitory lateral connections in retinas (*10*) but through retrograde messengers in the cortex (*8*). Moreover, synapses will be strengthened by stimulus according to Hebb conjecture (*11*). As mentioned above, such synaptic growth of "rich-get-richer" and lateral inhibition together will inevitably lead to "grandmother cells" instead of population coding (*9*). In other words, every object corresponds to a coding neuron in the cortex. And every neuron is a special detector of some similar inputs. In population coding however, you should give a new answer integrating all existing answers instead of picking one from them. This needs the "God's hand" or "binding mechanism" (*2*). Since neurons and synapses are the only computing units in the cortex, they should organize themselves other than seek help of outside force. Therefore in our opinion the brain waves should be byproducts rather than the binding mechanism. Certainly, the "grandmother cell" coding isn't precise. In fact, the precision is determined by the total number of neurons. For better survival however, our brain has to make a compromise between precision and the cost of material and energy.

The mechanisms mentioned above are in the spatial dimensionalities, and it is very similar in the temporal dimensionality. Suppose an event has been lasting for time x until now (see Fig. 3). On this condition, the probability of this event occurring should be $p = c_1(1-e^{-c_2x})$. Generally speaking, signals appearing occasionally could be noise. On the other hand, if this event has been interrupted for time x (see Fig. 3), the probability of its reoccurring should be $p = p_0 e^{-c_3x}$. After all, old clocks couldn't be as reliable as new ones. These two curves together will compose a dorsal-fin curve (see Fig. 1). Therefore the dorsal-fin curve can represent the probability estimate of an event occurring based on history information. And the premise is that events are continuous and local other than disperse in the timeline (see Fig. 3). Similarly, objects should exist in continuous local space as well. In fact, lateral inhibition and "grandmother cells" also reflect spatial localization in some degree. In conclusion, dorsal-fin curve actually reflects space-time localization. And it is usually called encoding or features binding in space (*2*), while in timeline it is usually called memory.

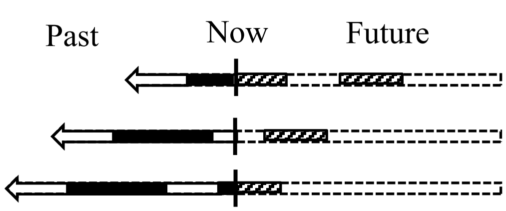

Fig. 3. Memory. Arrow lines represent timelines. Similar to the left side of Fig. 2, bars mean events and corresponding external inputs, and diagonal stripes are invisible at present. Every event corresponds to an external memory namely synaptic strengths or EPSPs. These two are long-term memory and short-term memory respectively. In essence, memory is the probability estimate of an event occurring based on input history. And the confidence of this estimate drops with time when without inputs, which actually corresponds to forgetting. Therefore memory and

forgetting correspond to the rising and falling phases of the dorsal-fin curve respectively. This reflects the temporal localization. Namely bars here are continuous and local in the timeline.

Similarly, this temporal mechanism can be easily implemented by synapses (*6-8*). Specifically, a synapse is like a leaky pool with only one water pipe. And the leaky rate is proportional to the water amount. Then the relation between water amount and time should follow function $p = c_1(1-e^{-c_2 t})$. Therefore, the accumulation of postsynaptic signals can represent the probability estimate of a feature occurring based on spikes history. And the synaptic strength can represent the statistical confidence of a feature based on longer stimulus history. These two are short-term (or instantaneous) and long-term memories respectively. On the other hand, when without water injection, the water amount will decay exponentially $p = c_3 e^{-c_4 t}$. And this is consistent with the probability dropping when inputs are interrupted.

According to Francis Crick, a kind of instantaneous memory such as EPSP should be the foundation of consciousness (*2*). According to our model, this instantaneous memory actually emerges from space-time concentration of ions. Specifically, ion's randomly passing membrane channels will result in the dorsal-fin curve and consciousness. Since every object corresponds to a single coding neuron, conscious of an object means this coding neuron's spikes. It might be hard to believe that the neural mechanism of consciousness is so simple. However, this fit experimental data well, and it can explain most psychological phenomenon and advanced intelligence such as language and reasoning (*8, 10, 12, 13*). Moreover, even a baby with billions of neurons is so simple, how complex could a single neuron's behavior be? As the old said, you can't recognize the mountain because you are in it.

In essence, the complex circuits in the cortex are caused by the complex postnatal experiences (*2*). In other words, the neural network is self-adaptive to external inputs like an ecosystem, in which neurons and synapses collaborate and compete like animals (*14*). At present, a focus of neuroscience is drawing the detailed circuits of cortex. In our opinion however, cortex is actually formed through self-organization of similar neurons (*8, 12*). Therefore it is more important to find out the interactive rules between neurons. Synaptic strength is a kind of long-term memory, which means its decay rate must be changeable (*8*). Otherwise it has no essential difference from short-term memory, no matter how slow its decay rate is. According to our model (*8*), forgetting is actually retrieval failure due to many factors such as synapse decay, lateral inhibition, lack of clues, and so on. Synapse decay composes the falling phase or dorsal-fin curve, and therefore it is an essential part of memory from the statistical viewpoint. Moreover, it is also a kind of resource recycling.

Consciousness and memory are somewhat like computer's memory and disk respectively. Differently from computer however, both of them are distributed. Moreover, contrast to the precise hard-coding in computer, neural coding are actually statistical and soft. However, neither of them is superior to the other. They just have different architectures and different abilities, and this is like comparing our legs with cars.

References and Notes:

1. E. Schrödinger, *What Is Life?* (Cambridge University Press, Cambridge, 1944).

2. F. Crick, *The Astonishing Hypothesis: The Scientific Search for the Soul* (Charles Scribner's Sons, New York, 1994).
3. C. Koch, K. Hepp, Quantum mechanics in the brain. *Nature* **440**, 611-612 (2006).
4. J. V. Neumann, *The Computer and the Brain* (Yale University Press, New Haven, CT, 2000).
5. A. L. Hodgkin, A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500-544 (1952).
6. E. L. Bienenstock, L. N. Cooper, P. W. Munro, Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* **2**, 28-32 (1982).
7. W. Gerstner, R. Kempter, J. L. Hemmen, H. Wagner, A neuronal learning rule for sub-millisecond temporal coding. *Nature* **383**, 76-78 (1996).
8. P. Liu, T. Wang, A quantitative neural coding model of sensory memory, in press (available at http://arxiv.org/abs/1406.6453).
9. C. G. Gross, Genealogy of the "grandmother cell". *Neuroscientist* **8**, 512-518 (2002).M.
10. P. Liu, T. Wang, A unified quantitative model of vision and audition, in press (available at http://arxiv.org/abs/1406.5807).
11. D. O. Hebb, *The Organization of Behavior: A Neuropsychological Approach* (Jhon & Sons, New York, 1949).
12. P. Liu, T. Wang, Motor learning mechanism on the neuron scale, in press (available at http://arxiv.org/abs/1407. 7027).
13. P. Liu, T. Wang, Neural mechanism of language, in press (available at http://arxiv.org/abs/1408. 5403).
14. G. Edelman, *The Theory of Neuronal Group Selection* (Basic Books, New York 1987).

Acknowledgments: This work was supported by the National Natural Science Foundation of China (Grant No. 61170156 and 60933005).