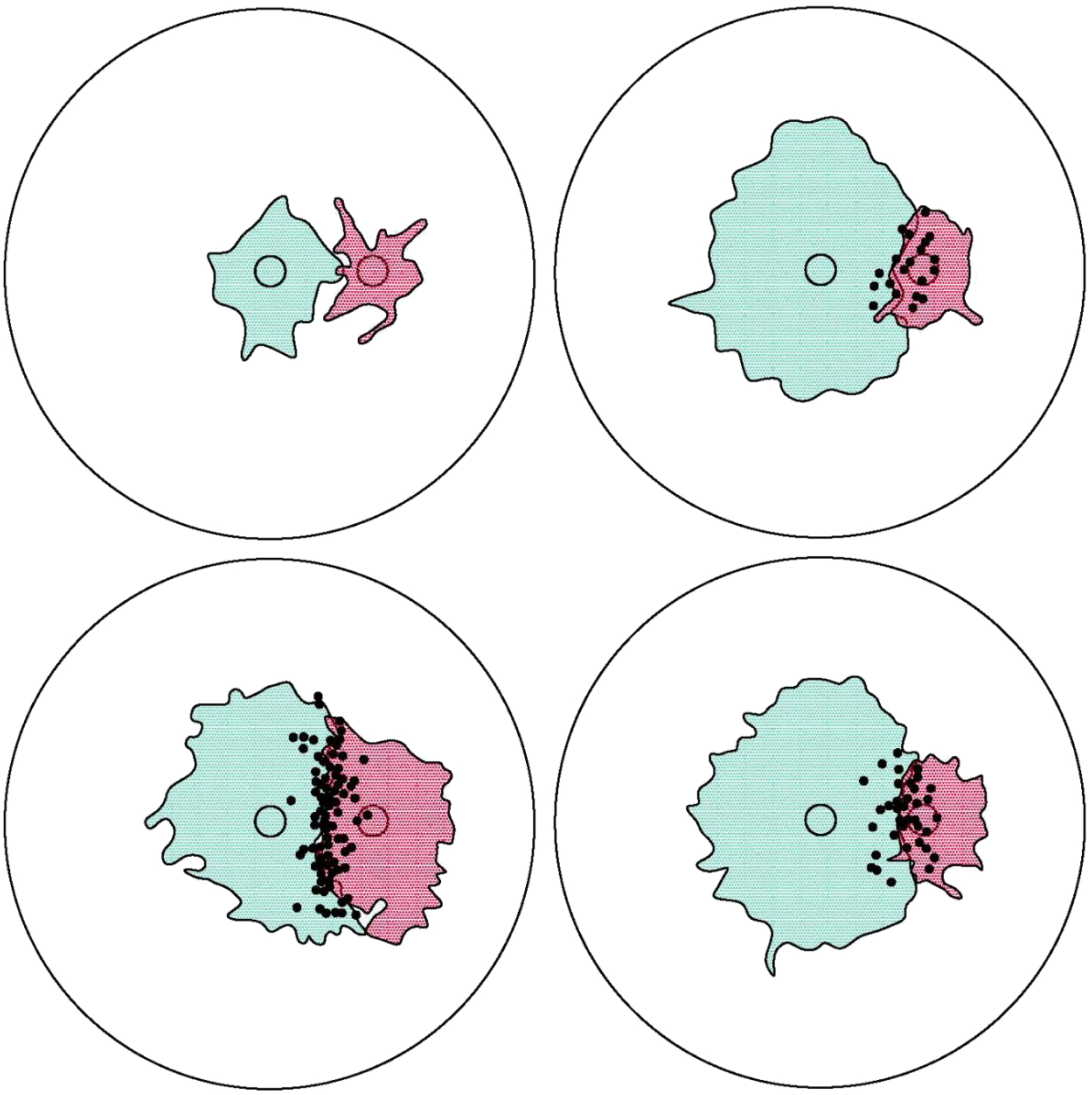
Consciousness as representation formation from a neural Darwinian perspective[*]

Anna Kocsis, mag.phil.
Institute of Philosophy Zagreb, Croatia
Vjeran Kerić, mag.phil.
Department of Psychology and Cognitive Science, New Bulgarian University, Sofia

The theory of neural group selection was proposed by Edelman (1978) and prompted by two sets of observation. The fact that individual nervous systems show a high level of structural and functional variability and the fact that in order to survive in its eco-niche an organism must either inherit or create criteria that enable it to partition the world into perceptual categories according to its adaptive needs (Edelman, 1993). The ability to categorize or the resulting categories themselves arise from processes of selection upon variation on the structural level. The selection processes suggested by Edelman take place on several levels; at the level of the genes, at the embriogenesis level and the level of dynamic changes of synaptic efficacy that is occurring in ones lifetime and are unique for each individual. Edelman was known for a leitmotive that he repeated at his talks and in his publications: „the mind is embodied and the body is embedded“ which means that for a satisfying characterisation of the properties of the mind we need to consider bodily functions and rhythms (such as circadian rhythms) but also the relevant features of the environment which the organism can detect as statistical regularities.

Computational models developed by Izikevich, Edelman and Gally (Izhikevich, Gally & Edelman 2004; Izhikevich 2006; Izhikevich & Edelman 2008) demonstrated that neuronal groups can spontaneously form, as predicted by Edelman, and exhibit cooperative and competitive behaviour. The necessary conditions for spontaneous formation of such groups are spike-time dependent plasticity (STDP) and axonal conduction delays. Both of these make the model biologically plausible. STDP is a spike-based formulation of a temporally asymmetric hebbian learning rule which is often characterized by the phrase „neurons that fire together wire together“. Repeated presynaptic arrivals result in long term potentiation (LTP) while repeated postsynaptic arrivals result in long term depression (LDP). The STDP profiles are different between synapse types and locations and depend on the presence or absence of neural modulators, especially dopamine that change the temporal profile of STDP. In the presence of dopamine, STDP turns from unsupervised learning into a reward-based learning paradigm. Depending on the pattern of conduction delays, STDP depresses some synapses and potentiates others. As a result of this plasticity, neurons self organize into robust and stable groups.

[*] This work has been supported by the Croatian Science Foundation under the project number 5343.



Schematically redrawn after Izhikevich, Gally & Edelman 2004.

These illustrations show the change of projective fields due to continuous coherent stimulation of neurons inside the black circles; first circle: starting groups, second circle: stimulation to the left circle, third circle: stimulation to both circles syncrhonously, forth circle: stimulation to both circles asynchronously. Black dots denote neurons shared by the projective fields. In the third circle neuronal groups cooperate while in the final illustration they compete where the left group is „stealing“ neurons from the right group due to the fact that these two groups have different spatio-temporal patterns of spikes so that the activation of one group can distort the firing pattern of the other group.

Neuronal groups can be seen as representing memories and experience where the content is robust but at the same time is being constantly changed and updated in accordance with incoming information or stimuli. Each neuron can be a part of many

different neuronal groups, resulting in a huge amount of possible neuronal groups. Which of them will be realized depends upon the fit between axon time delays, the STDP profiles and the characteristics of the input.

The theory of neuronal group selection has some interesting consequences for the problem of consciousness. Maybe the two most prominent features of consciousness are its unitary nature and phenomenal character. This means that an account of consciousness has to provide a mechanism which explains both the binding problem and the range of discriminations made by an individual. The relationship between binding and discrimination results in perceptual categorization. Perceptual categorization, in simple terms, is the way inputs are grouped into more robust groups, dependent on their statistical nature. Edelman's sketch for a solution to the binding problem consists of two elements: reentrant signaling and a value system. Reentrant signaling is a system of parallel, bidirectional pathways that enable communication between mental maps. This activity leads to the experience of spatiotemporal consistency in our interaction with the environment by increased synchronisation across distributed brain areas. The neuronal group that initiates this syncrhonous firing pattern „pushes out“ other competing groups firing at the same time (maybe as a reaction to the same stimulus). Which stimulus and by which groups will enter consciousness is to an extent determined by the activity of the underlying value systems. These systems affect the selectional process by altering the STDP profiles. The value system includes various thalamic nuclei and neuromodulator systems.

According to TNGS, the differences between the mental states of humans and other animals are a matter of degree and not kind (due to complexity). Consciousness was not selected for, its emergence depended on the selection of natural systems that give rise to it. The „goal” of the system is to represent the environment in a useful way. Therefore, consciousness and the mechanisms that underlie representational systems are tightly coupled while ”primitive” subcortical regions have an important modulatory role on conscious processing. In what follows, I will present some recent empirical findings that can elucidate to what extent have some of these tenets been related to consciousness thus far. Especially the claim of the close relationship between representational processes or processes updated by experience and consciousness within the context of value-driven categorization.

In their recently published study, Schurger et al. (2015) hypothesised that conscious perception involves temporary stabilization of distributed cortical networks in which case stability could be a measurement of conscious perception. Stability has previously been recognized by Edelman and other authors as an important feature of conscious processing. Schurger and colleagues asked participants to discriminate between masked houses and faces and tested the relationship between stability and conscious processing. They found that the amplitude of the P3 component was sensitive to whether the participant detected

the stimulans and to the category of the stimulus, while stability within trial was highly sensitive to the seen/unseen difference (less variability when the stimulans was seen): in this way one could claim that stability is a better indicator of consciousness than the P3 amplitude.

It has been previously shown that the P3 amplitude is closely related to the motivational significance and arousing nature of the stimulus (subjective probability of the stimulus, task relevance of the stimulus, its salience, intensity, novelty etc.) (Nieuwenhuis 2011; Polich 2007). The P3 component can more precisely be divided into two subcomponents: frontal P3a and temporal-parietal P3b. The P3a component is modulated by dopamine while the P3b component is modulated by norepinephrin (thus the P3 amplitude sensitivity to the stimulus category). What is the role of P3 is a difficult question since P3 is observed in any task that requires stimulus discrimination but the roles of frontal dopamine and temporal parietal noreprinephrin implicate P3 in learning, context updating and more generally with functions that maintain systems that represent the external world. Norepinephrine signals influence learning in a bottom-up way while dopaminergic modulation in frontal areas is associated with cognitive control attention and top-down regulation.

The second line of evidence comes from brain oscillations. Gamma-oscillations that occur in all cortical areas and most subcortical structures have received much attention and are considered to be a hallmark of conscious processing. However, recent studies have found that ongoing alpha band oscillations just prior to the stimulus onset can predict whether the stimulus will be consciously accessed. When alpha band power is lower, the stimulus is reported as perceived. The function of alpha oscillations is believed to be the regulation of cortical excitability by pulses of inhibition (Jensen et al. 2012). Changes in alpha oscillations are correlated with stimulans saliency (Jensen et al. 2012) and also the semantic content of the information that is retrieved from long-term memory (Klimesch 2012). The more semantically integrated the retrieved information is, the larger the alpha modulated release from inhibition is (event-related desynchronisation, ERD). Alpha oscillations are known to be generated in the thalamus (most likely pulvinar as lesions to the pulvinar are known to reduce cortical excitability and it is likely that pulvinar modulates neuronal synchrony between cortical areas and is also sensitive to stimulus salience) the exact mechanism is still unknown.

To sum up: the amplitude of P3 is closely connected to representation formation and updating processes. This points to a very close relationship between conscious processing and memory-related processing: as in conscious processing could be one type of integration of experiences and events by the system. These processes are value-driven in a sense that they are modulated by more primitive parts of the central nervous system. A very similar argument can be made in the case of alpha oscillations are generated in the thalamus and

are sensitive to salience and semantic content. These findings seem to be in line with Edelman's claims about the nature of consciousness.

Edelman, G. M. (1978). *The mindful brain: Cortical organization and the group-selective theory of higher brain function*. MIT Press.

Edelman, G. M. (1993). Neural darwinism: Selection and reentrant signaling in higher brain function. *Neuron* 115−125.

Izhikevich, E. M. (2006). Polychronization: Computation with spikes. *Neural Computation* 18:245–282.

Izhikevich, E. M., Gally, J. A. & Edelman, G. M. (2004). Spike-timing dynamics of neuronal groups. *Cerebral Cortex* 14: 933–944.

Izhikevich, E. M. & Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *PNAS* 105: 3593−3598.

Jensen, O., Bonnefond, M., & VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Science* 16:200–206.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Science* 16(12): 606–617.

Nieuwenhuis, S. (2011). Learning, the P3, and the locus coeruleus-norepinephrine system. In: R. Mars, J. Sallet, M. Rushworth & N. Yeung (Eds.) *Neural basis of motivational and cognitive control*. Oxford University Press, p. 209−222.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology* 118 (10): 2128−2148.

Schurger, A., Sarigiannidis, I., Naccache, L., Sitt, J. D. & Dehaene, S. (2015). „Cortical activity is more stable when sensory stimuli are consciously perceived“. doi:10.1073/pnas.1418730112.