

# A neural coarse graining theory of consciousness

Acer Y.C. Chang<sup>\*1</sup>, Martin Biehl<sup>†1</sup>, and Ryota Kanai<sup>‡1</sup>

<sup>1</sup>ARAYA, Inc., Tokyo, Japan

May 14, 2018

## Abstract

Neural systems process information through different levels of organisation in a hierarchical manner. Information at lower levels is finer-grained and can be coarse-grained for higher level computation. However, one is aware of information processed only at specific levels. Theorists have addressed this issue. For example, the intermediate level theory of consciousness suggests that the intermediate level seems to be privileged with respect to consciousness. It is true that we do not experience information processed by individual neurons which is always highly noisy. Besides, we have no conscious experience from interpersonal activities albeit massive interactions among individuals. Instead, neurophysiological evidence has been showing that conscious experience tends to covary with information encoded in coarse-grained neural states such as neural population codes. We argue that the neural states within the scope of the information closure determine the contents of consciousness and brain processes outside apart from the representations of that level remain unconscious. This argument suggests a distinction between conscious and unconscious processing and provides a generic computational framework. Finally, using the deep learning network, we can measure information closure in deep hidden layers. Our preliminary results show that information closure representation emerged after learning. We further decoded the information from the representation and compared it with human conscious perception.

## keywords:

Keywords: theory of Consciousness, information closure, neural coarse-graining, level of analysis

## 1 Introduction

## 2 neural coarse-graining

### 2.1 coarse-graining in neural system

### 2.2 Single neurons are noisy

*\* "neurons are noisy: the same pattern of activity never occurs twice, even when the same stimulus is presented."*

\* This fact leads to two important points. 1. coarse-graining is necessary to resist noise 2. The information processing related to consciousness should not be situated at this level. Otherwise our conscious percept should be very noisy and unstable.

### 2.3 firing rate coding

\* temporal coarse-graining

---

<sup>\*</sup>acercyc@araya.org

<sup>†</sup>martin@araya.org

<sup>‡</sup>Ryota Kanai

## 2.4 Population coding

\* Population coding is the idea that nervous systems signal and compute parameters using large populations of imprecise neurons, rather than one or a few precise neurons.

\* *"As single neurons are not very informative, to obtain accurate information about sensory or motor variables some sort of population averaging must be performed."*

\* *"Neural population code: the set of response features of a population of neurons that carry all information about the considered stimuli. These features consist of spatio-temporal sequences of action potentials distributed across neurons and/or time."*

\* The diverse response selectivity of sensory neurons

\* *"How a neural population represents information is partly determined by the diverse selectivity of individual neurons Shamir (2014)"*

\* *"Population code (also ensemble code) denotes a code by which neural information is encoded in the spatiotemporal activity patterns of many neurons." Binder et al. (2009)*

## 2.5 sensory hierarchy

\* Sensors can only receive partial information \* Through coarse-graining, higher level cortical areas can integrate those partial information to infer hidden causes.

## 2.6 The advantage of NCG

### 2.6.1 Resist to noise

\* *"The population will therefore be relatively insensitive to the loss of cells or noise in a small number of cells" Eurich & Wilke (2000)*

### 2.6.2 Operate on deterministic and abstract level

### 2.6.3 Predictive power

### 2.6.4 Reduce energy cost

## 2.7 neural coarse-graining in deep learning

# 3 Non-trivial Information Closure

*"Our theoretical interest concerns the type of system that is a unity for and by itself and not only for an external observer distinguishing some entity from the rest of the world. This requires a system that can be described as a whole without reference to its environment. In systems theory, this property is usually referred to as closure." (BERTSCHINGER et al., 2006)*

*"These concepts of closure play an important role in the architecture of systems theory, because they are used to 1. define the system (in distinction to its environment) and to 2. explain the autonomy of the system."*

Informational closure: The higher process is informationally closed, i.e., there is no information flow from the lower to the higher level. Knowledge of the microstate will not improve predictions of the macrostate (PFANTE et al., 2014, p. 4).

*"A system that is independent from its environment trivially achieves informational closure." (BERTSCHINGER et al., 2006)*

## 3.1 Mathematical definition of information closure

*"The notion of informational closure refers to a situation where the information flow between the environment and the system tends to zero." BERTSCHINGER et al. (2006)*

suggests that information closure can be defined as the degree of information flow between environment and the system.

$$\begin{aligned} J_n(E \rightarrow S) &= MI(S_{n+1}; E_n | S_n) \\ &= H(S_{n+1} | S_n) - H(S_{n+1} | S_n, E_n) \\ &= H(E_n | S_n) - H(E_n | S_n, S_{n+1}) \\ &= H(E_n | S_n) - H(E_n | S_n, S_{n+1}) \end{aligned} \tag{1}$$

### 3.2 Non-trivial information closure

Information closure could be trivial. That is, when a system is fully independent from the environment, no information can flow into the system from the environment and also leak to the environment from the system. Equa. xxxx shows the mathematical description of trivial information closure. When mutual information between the environment and system future state close to 0, and the system transition is independent from the environment, the system can reach informational closure. However, such systems do not have any functional meaning and evolutionary advantages.

Therefore, it's important to ensure the informational closure that the system reach is non-trivial.

*[[ cite 2006 and nic's paper ]]*

To achieve non-trivial information closure (BERTSCHINGER *et al.*, 2006) (Guttenberg *et al.*, 2016)

### 3.3 how to achieve non-trivial information closure

\* Through coarse-graining \* Accurately model deterministic environment " *This demonstrates that a system exhibiting certain internal regularities as measured by  $A^* = MI(S_{n+1}; S_n)$  can achieve informational closure either by gaining information about the environment or by increased autonomy, i.e. by becoming unpredictable or uncontrollable from the (13) environment. Therefore, information about the environment, i.e. modeling, and autonomy can be considered as complementary strategies for achieving informational closure.* " \* by Action

### 3.4 information closure and Level Identification

## 4 Our claim of consciousness

\* We claim that the coarse-grained state in the information closure determines contents of consciousness. \* It's important that not the neural states but the coarse-grained state determine contents of consciousness.

### 4.1 Measure of conscious level

### 4.2 Information and system theory

*[[ Closure can define a system, consciousness is information ]]*  
*[[ Could be linked to system theory, but may need help ]]*

*" This self-referential distinction from its environment therefore gives rise to the specific autonomy of such a system. Consequently, in systems theory, closure properties and autonomy are considered to be closely related concepts which are both at the heart of defining the system itself. "*

## 5 Emperical approaches of measuring neural information closure

## 6 Biological Evidence of information closure in neural system

\* Recent biological evidence suggests that low-level neurons encodes predictive information about external and

Sederberg *et al.* (2018) *"in recordings of populations of RGCs of salamander retina driven by a simple stimulus with partially predictable dynamics, joint activity patterns transmit information that is predictive of future stimuli "*

*"We examined whether the temporal correlations within populations of RGCs can be used to guide the search for readouts of predictive information."*

*" We showed that groups of cells with high external, stimulus-predictive information also had high internal predictive information on two classes of stimuli, "*

need to say this is exactly the non-trivial information closure, and refer to equation XXX

" Very simple learning rules could find near-optimal readouts of predictive information without any external instructive signal. This suggests that bottom-up prediction may play an important role in sensory processing "

"Internal predictive information can guide stimulus prediction without explicit reference to stimulus parameters."

" Wordword internal predictive information is correlated with wordstimulus predictive information across sets of four cells."

\* They also showed generalization of internal predictability across stimulus. This suggests .... what?

" The generalization of internal predictability across stimulus classes partially reflects the ability of the retina to adapt to the statistics of stimuli "

Palmer et al. (2015) " Groups of cells in the retina carry information about the future state of their own activity, and we show that this information can be compressed further and encoded by downstream predictor neurons that exhibit feature selectivity that would support predictive computations."

"It seems natural to phrase the problem of prediction in relation to the visual stimulus, as in Fig. 3, but the brain has no direct access to visual stimuli except that provided by the retina. Could the brain learn, in an unsupervised fashion, to predict the future of retinal outputs? More precisely, if we observe that a population of retinal ganglion cells generates the word  $w_t$  at time  $t$ , what can we say about the word that will be generated at time  $t + \tau$  in the future? The answer to this question is contained in the conditional distribution of one word on the other,  $P(w_{t+\tau} | w_t)$ "

" larger groups of cells carry predictive information for hundreds of milliseconds, "

<http://tinyurl.com/y9zmhl1n>

## 7 Relation to empirical findings in neuroscience and consciousness studies

### 7.1 reflexive behavioural is unconscious

### 7.2 Explain normal and abnormal conscious/non-conscious experience

### 7.3 Deterministic vs probabilistic

Dehaene et al. (2017) Vul et al. (2008); Moreno-Bote et al. (2011); Asplund et al. (2014); Vul et al. (2009)

## 8 Comparison with other theories

### 8.1 Intermediate Level Theory

Jackendoff (1987) Prinz (2007)

" In 1987, Ray Jackendoff published *Consciousness and the Computational Mind*. In it, he posed an important Where question: Where, in the flow of information, does consciousness arise? Most cognitive scientists agree that the mind is, in some sense, a computer. It is a device that processes information by transforming representations in accordance with rules. Computational devices decompose into various interconnected subsystems, each of which performs some aspect of a complex task. Given such a decomposition analysis, we can ask: in which subsystems does consciousness arise? If we depict the mind as a vast flow chart, and highlight the boxes whose rules and representations are conscious, which boxes should we mark? Jackendoff's answer is simple and elegant. He noticed that many of our mental capacities, including our senses and our language systems, are organized hierarchically. In each of these hierarchies, it makes sense to talk about low-, intermediate-, and high- level processing systems. We break down tasks into stages. Appealing to prevailing models of these stages, Jackendoff observed that the intermediate level seems to be privileged with respect to consciousness. Consciousness seems to arise in intermediate- level processing systems and not elsewhere." - Prinz (2007)

\* 2.5D

" Under the computational interpretation, the why question means: In what sense are interme-

about why it is this level

mediate level representations computationally important or distinctive? This is a tractable question, and it may lead to insights into the function of consciousness. Representations at the intermediate level are ideally suited for real-time deliberative behavioral responses. Low-level representations fail to bind features into coherent wholes. If we are going to interact with our environment, the low level is not a good guide. The high level is a good deal better for action, but also suboptimal. The high level tells us the category of the stimuli we perceive, but from an allocentric point of view. It abstracts away from stimulus features that are crucial for action. If we encounter a predator, the high-level visual representation does not tell us if it is facing toward us or away from us. Without that knowledge, we cannot decide what course of action to take. So, I propose that the intermediate level plays a distinctive role in information processing. It delivers representations that are coherent but viewpoint specific. These representations are useful for determining what to do here and now.<sup>arg1</sup>" Prinz (2007) "I conclude that the intermediate level plays a distinctive and important computational role. Some researchers find this puzzling. In wondering why the intermediate level is privileged, they note the arbitrariness of the fact that perceptual hierarchies have three levels. Could they not have two? Or a hundred? Do the 50 or so areas that contribute to visual processing really divide neatly into three levels? I think these puzzles depend on a particular understanding of what the word intermediate amounts to. One might get caught up in the fact that the intermediate level is the second stage in a sequence. Call this the serial understanding of the intermediate level. Alternatively, one might think of the intermediate level semantically. On this reading, the intermediate level is one that is not abstract and categorical, but not piecemeal or disunified. These notions are all in need of some refinement, but, as a first approximation, the idea is that intermediate-level representations are neither too specific nor too general." Prinz (2007)

### 8.1.1 The difference between ILT and our theory

Prinz suggested that "as. Cells in V1 are not promising candidates, because they do not reliably respond in ways that are consistent with features that we experience consciously. " Prinz (2007)

\* Our prediction is, if information that is carried by v1 neurons are necessary to complete information closure, the states of V1 neurons should also co-vary with conscious contents.

## 8.2 Global Workspace

### 8.2.1 Ignition in GWT

[[ Explain ignite ]]

## 8.3 Sensorimotor contingency

## 8.4 Predictive Brain

\* To achieve non-trivial information closure naturally maximal predictive power via coarse-graining.

I need math here!!

## 8.5 IIT

[[ Hoel's theory ]]

## 8.6 Internal simulation and self-modeling

## 9 Counter-intuitive prediction

## 10 Evolution of conscious mind

Dennett (2008)

## 11 Conclusion

\* NCC: To find NCC, it's not about where and when, it's about scale and level of description

## Funding

## Acknowledgements

## Supplemental Data

## References

- Asplund, Christopher L, Fougny, Daryl, Zughni, Samir, Martin, Justin W, & Marois, René. 2014. The attentional blink reveals the probabilistic nature of discrete conscious perception. *Psychological science*, **25**(3), 824–831.
- BERTSCHINGER, NILS, OLBRICH, ECKEHARD, AY, NIHAT, & JOST, JÜRGEN. 2006. Information and closure in systems theory. *the Proceedings of the 7Th German Workshop of Artificial Life*, 26–28.
- Binder, Marc D, Hirokawa, Nobutaka, & Windhorst, Uwe. 2009. Encyclopedia of neuroscience.
- Dehaene, Stanislas, Lau, Hakwan, & Kouider, Sid. 2017. What is consciousness, and could machines have it? *Science*, **358**(6362), 486–492.
- Dennett, Danile C. 2008. *Kinds of minds: Toward an understanding of consciousness*. Basic Books.
- Eurich, Christian W, & Wilke, Stefan D. 2000. Multidimensional encoding strategy of spiking neurons. *Neural Computation*, **12**(7), 1519–1529.
- Guttenberg, Nicholas, Biehl, Martin, & Kanai, Ryota. 2016. Neural Coarse-Graining: Extracting slowly-varying latent degrees of freedom with neural networks. *arXiv preprint arXiv:1609.00116*.
- Jackendoff, Ray. 1987. *Consciousness and the computational mind*. The MIT Press.
- Moreno-Bote, Rubén, Knill, David C, & Pouget, Alexandre. 2011. Bayesian sampling in visual perception. *Proceedings of the National Academy of Sciences*, **108**(30), 12491–12496.
- Palmer, Stephanie E., Marre, Olivier, Berry, Michael J., & Bialek, William. 2015. Predictive information in a sensory population. *Proceedings of the National Academy of Sciences*, **112**(22), 6908–6913.
- PFANTE, OLIVER, BERTSCHINGER, NILS, OLBRICH, ECKEHARD, AY, NIHAT, & JOST, JÜRGEN. 2014. COMPARISON BETWEEN DIFFERENT METHODS OF LEVEL IDENTIFICATION. *Advances in Complex Systems*, **17**(02), 1450007.
- Prinz, Jesse. 2007. The intermediate level theory of consciousness. *The Blackwell companion to consciousness*, 257–271.
- Sederberg, Audrey J, MacLean, Jason N, & Palmer, Stephanie E. 2018. Learning to make external sensory stimulus predictions using internal correlations in populations of neurons. *Proceedings of the National Academy of Sciences*, 201710779.
- Shamir, Maoz. 2014. Emerging principles of population coding: In search for the neural code. *Current Opinion in Neurobiology*, **25**, 140–148.
- Vul, Edward, Nieuwenstein, Mark, & Kanwisher, Nancy. 2008. Temporal selection is suppressed, delayed, and diffused during the attentional blink. *Psychological Science*, **19**(1), 55–61.
- Vul, Edward, Hanus, Deborah, & Kanwisher, Nancy. 2009. Attention as inference: selection is probabilistic; responses are all-or-none samples. *Journal of Experimental Psychology: General*, **138**(4), 546.