

# 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 31, 2018

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Non-trivial Information Closure</b>	<b>4</b>
2.1	Mathematical definition of information closure . . . . .	4
2.2	Non-trivial information closure . . . . .	4
2.3	how to achieve non-trivial information closure . . . . .	5
2.4	information closure and Level Identification . . . . .	5
2.5	Biological Creature may try to achieve closure even the environment is not . . . . .	5
<b>3</b>	<b>neural coarse-graining</b>	<b>5</b>
3.1	coarse-graining in neural system . . . . .	5
3.2	Single neurons are noisy . . . . .	5
3.3	firing rate coding . . . . .	5
3.4	Population coding . . . . .	5
3.5	sensory hierarchy . . . . .	6
3.6	The advantage of NCG . . . . .	6
3.6.1	Resist to noise . . . . .	6
3.6.2	Operate on deterministic and abstract level . . . . .	6
3.6.3	Predictive power . . . . .	6
3.6.4	Reduce energy cost . . . . .	6
3.7	neural coarse-graining in deep learning . . . . .	6
<b>4</b>	<b>A neural coarse graining theory of consciousness</b>	<b>6</b>
4.1	Measure of conscious level . . . . .	6
4.2	Conscious contents . . . . .	6
4.3	Sensory hierarchy and neural coarse-graining . . . . .	6
4.4	Information and system theory . . . . .	7
4.5	Emperical approaches of measuring neural information closure . . . . .	8
<b>5</b>	<b>Biological Evidence of information closure in neural system</b>	<b>8</b>
<b>6</b>	<b>Relation to empirical findings in neuroscience and consciousness studies</b>	<b>10</b>
6.1	reflexive behavioural is unconscious . . . . .	10
6.2	Explain normal and abnormal conscious/non-conscious experience . . . . .	10
6.3	Deterministic vs probabilistic . . . . .	10

@Martin:  
todo

This is super important. I  
need to deal with this very  
carefully.

---

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

<sup>†</sup>martin@araya.org

<sup>‡</sup>kanair@araya.org

<b>7</b>	<b>Comparison with other theories</b>	<b>10</b>
7.1	Intermediate Level Theory	10
7.1.1	The difference between ILT and our theory	11
7.2	first-order theories	11
7.3	Global Workspace Theory (GWT)	12
7.3.1	Global availability and Broadcasting	12
7.3.2	Code translation	12
7.3.3	Stability	12
7.3.4	broadcasting	12
7.3.5	Ignition	12
7.4	Higher Order Thought theory of consciousness	13
7.4.1	Metacognition	14
7.5	Sensorimotor contingency	14
7.6	Predictive Brain	14
7.6.1	statistical learning/prediction is not conscious	14
7.7	IIT	15
7.7.1	Information Integration	15
7.7.2	IIT	15
7.7.3	Hoel's theory	15
7.8	Internal simulation and self-modeling	15
<b>8</b>	<b>Counter-intuitive prediction</b>	<b>15</b>
8.1	information closure	15
8.2	Prediction on split-brain (vs IIT)	15
<b>9</b>	<b>Future work</b>	<b>15</b>
9.1	Brain area	15
9.2	How to coarse-grain in the neural system	15
<b>10</b>	<b>Evolution of conscious mind (not sure but if I have time)</b>	<b>16</b>
<b>11</b>	<b>Conclusion</b>	<b>16</b>

### 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

- Imagining you are a neuron in (a human) Bill's brain. Your "responsibility" is to collect neurotransmitters through dendrites from your other neurons, accumulate membrane potential, and send a signal to other neurons through an action potential.

- You have no idea that you are at a neural circuit in Bill's supplementary motor area and is now involved in a motor control processing while Bill reaches his hand to grab a cup.
- You are totally ignorant of the intension, the goal, and the motor plan that Bill is having now even though you are part of the physiological substrate of those "high-level" information.
- The same story also happens to Bill. In his conscious mind, he is not aware of what you do and even your existence if Bill never studies neuroscience.
- It seems that information you are processing and information shows in Bill's conscious percept are divorced from each other.
- It seems true that we cannot consciously access information processed at all scale in the neural system. Only information at a certain level of coarse-graining is able to become parts of conscious contents.
- When we focus on microscopic level of the neural system, information processing in individual neurons is highly noisy and state transitions of neurons are stochastic (Goldwyn & Shea-Brown, 2011; White *et al.*, 2000).  
The ion channels' state transitions are stochastic  
Neurons fire all the time. However, those massive action potentials do not appear in our conscious contents.
- Don't even mention ion channel dynamic.
- We also lack of experience of whole brain operation. Some parts of our brains process in an unconscious way.
- We don't have experience of inter-individual super-consciousness
- In this article, we argue that consciousness **correlates** information processing at a certain coarse-graining level which form information closure. More specifically, the conscious level is determined by the level of information closure and conscious contents is determined by the states of information closure.
- We first introduce non-trivial information closure and its importance for human scale agents.
- We then presume that through neural coarse-graining the human brain achieves a non-trivial information closure **realm** at a specific macroscopic level.
- We then argue that consciousness levels and contents are determined the state of consciousness for both levels and contents.
- We then discuss how this theory can concisely explain empirical results from previous studies and integrate several current major theories of consciousness.
- We finally provide theory predictions testable approaches.
- This framework provides a concise and quantifiable explanation for all current major theories of consciousness.
- This framework provides a more fundamental understanding about why some neural process are conscious but others are not.

Is this proper?

Need to unify this term

- =====
- consciousness is information
  - conscious contents
    - **[[ lower level ]]**
    - Information processing of individual neurons are noisy.
    - We don't have conscious content changes when individual neurons fire action potentials.
    - Don't even mention ion channel dynamic.
    - **[[ Higher-level dynamic ]]**
  - For human, physics at our body scale are nearly follow classic physics, which is deterministic.
  - **[[ Human tendency for active prediction ]]**
    - Stahl & Feigenson (2015) " *Given the overwhelming quantity of information available from the environment, how do young learners know what to learn about and what to ignore? We found that 11-month-old infants (N = 110) used violations of prior expectations as special opportunities for learning. The infants were shown events that violated expectations about object behavior or events that were nearly identical but did not violate*

*expectations. The sight of an object that violated expectations enhanced learning and promoted information-seeking behaviors; specifically, infants learned more effectively about objects that committed violations, explored those objects more, and engaged in hypothesis-testing behaviors that reflected the particular kind of violation seen. Thus, early in life, expectancy violations offer a wedge into the problem of what to learn."*

- Intuitive physics
- However, sensory signal are always noisy.
- We propose that consciousness is not related to particular neural architectures or location of the neural system.

Ref: Intuitive physics

## 2 Non-trivial Information Closure

@Martin:  
todo

- *" The notion of closure plays a prominent role in systems theory where it is used to identify or define the system in distinction from its environment"*  
[Bertschinger et al. \(2006\)](#)
- Formally, then, a system is informationally closed if (almost) no information flows into it from the environment.
- *" 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. \(2006\)](#)
- Mathematically, information closure can be define and measured by as the degree of information flowing from the environment to a system.

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

*" 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."*  
[\(Bertschinger et al., 2006\)](#)

### 2.1 Mathematical definition of information closure

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

### 2.2 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.  
 To achieve non-trivial information closure (Bertschinger *et al.*, 2006)  
 (Guttenberg *et al.*, 2016)

cite 2006 and nic's paper

## 2.3 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

## 2.4 information closure and Level Identification

## 2.5 Biological Creature may try to achieve closure even the environment is not

# 3 neural coarse-graining

- Definition of coarse-graining
- *"Coarse-graining: a method of reducing the complexity of a system by treating groups of atoms/molecules as single quasi-particles CG: coarse-grained"*
- *"a number of different microstates, First, typically a number of different microstates, will realize the same value of the macro-variables of upper level causal theories. "*
- One consequence of coarse-graining is that it makes it permissible to ignore certain causal factors that would be relevant at more fine-grained level of description. Price & Corry (2007)

## 3.1 coarse-graining in neural system

## 3.2 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
  - coarse-graining is necessary to resist noise
  - The information processing related to consciousness should not be situated at this level. Otherwise our conscious percept should be very noisy and unstable.

## 3.3 firing rate coding

- temporal coarse-graining

## 3.4 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)

### 3.5 sensory hierarchy

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

### 3.6 The advantage of NCG

#### 3.6.1 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)

#### 3.6.2 Operate on deterministic and abstract level

#### 3.6.3 Predictive power

#### 3.6.4 Reduce energy cost

### 3.7 neural coarse-graining in deep learning

@Martin:  
todo

## 4 A neural coarse graining theory of consciousness

### [[ Need to describe my intuition more concretely ]]

- 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.
- Consciousness is information
  - Information is level-dependent.
  - Information is different at microscopic and macroscopic levels
  - Consciousness must be level-dependent
- *" This illustrates how coarse-graining allows the formulation of incomplete generalizations that are relatively invariant, although at the cost of predictive precision regarding fine-grained details."*  
Price & Corry (2007)
  - conscious perception are very stable and time invariant.
  - Conscious perception is informative but we loss all the precise information about the states at microscopic level.

### 4.1 Measure of conscious level

### 4.2 Conscious contents

### 4.3 Sensory hierarchy and neural coarse-graining

- It's important to note that the dimension coarse-graining direction is not necessary to be align with anatomical sensory hierarchy in the neural system.

This is super important. I need to deal with this very carefully.

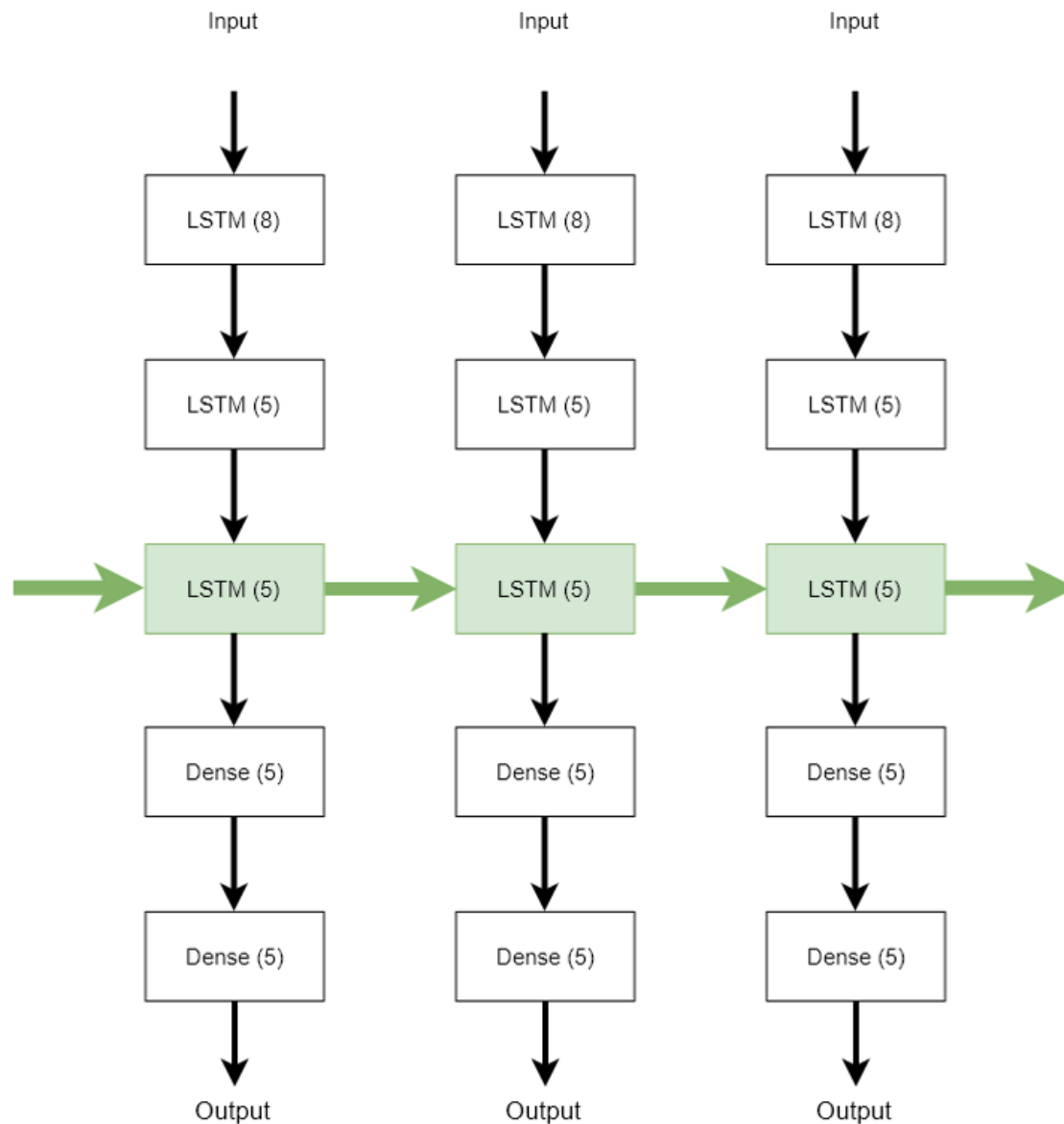


Figure 1: information closure through neural coarse-graining

- For example, pooling layers in CNN is an implementation of coarse-graining function. It's very plausible visual processing hierarchy in the neural system realises pooling layers.
- Of course, from evolutionary perspective, neural system should develop a better hardware to support neural coarse-graining
- It is possible that the neural system represents higher-level information processing by low level physical substrate. For example, studies have found representation for summary statistics of signal and neural population. From this point of view, sensory hierarchy may extend along with levels of coarse-graining.

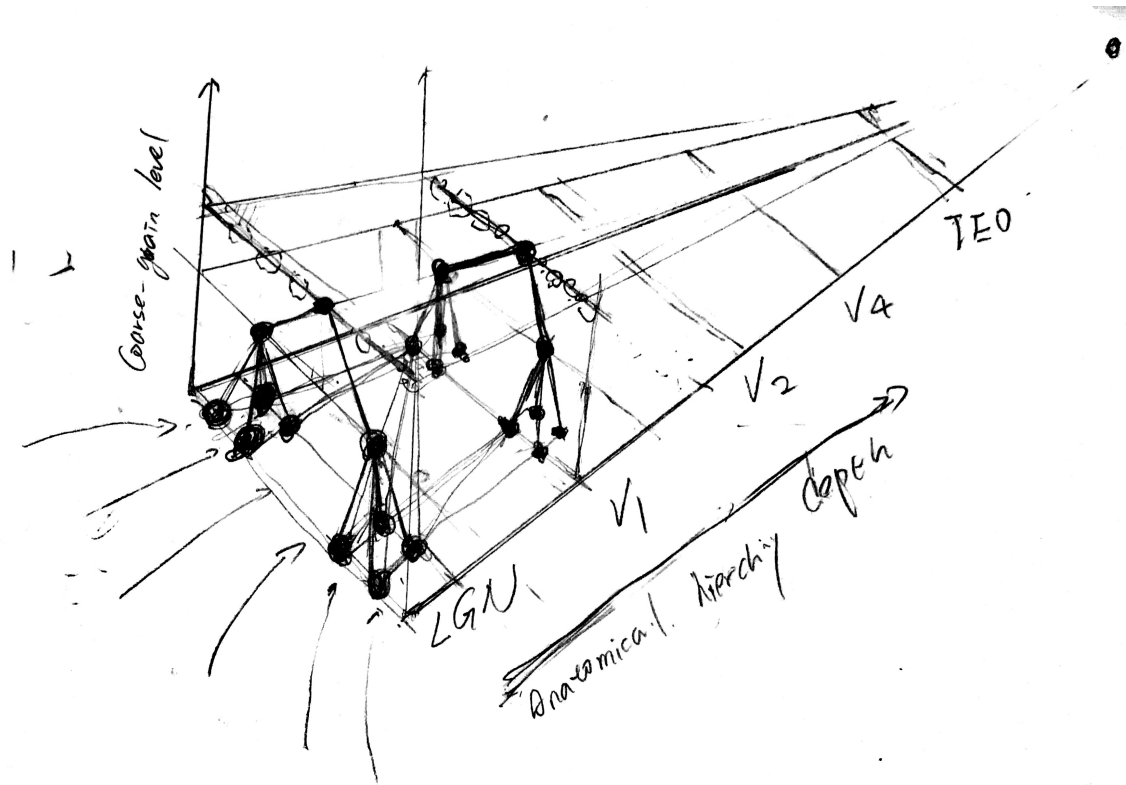
Ref: neural representation for summary statistics

#### 4.4 Information and system theory

[[ A 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."*



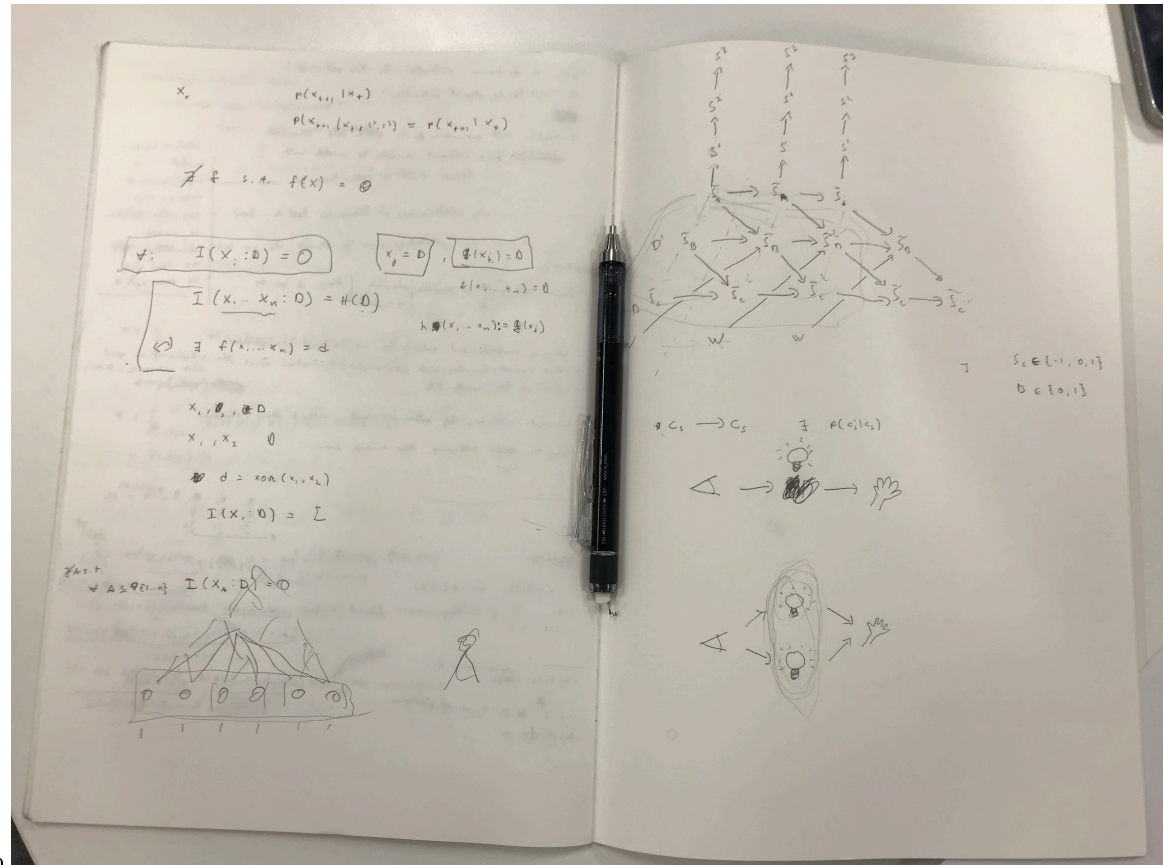
Coarse-grain and sensory anatomical hierarchies

#### 4.5 Empirical approaches of measuring neural information closure

### 5 Biological Evidence of information closure in neural system

- Recent biological evidence suggests that low-level neurons encodes predictive information about external and
- **[[ need to say this is exactly the non-trivial information closure, and refer to equation XXX ]]**
- [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, "*
- *" 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. "*





(1).jpg (1).bb

- " 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  $w$  at time  $t$ , what can we say about the word that will be generated at time  $t + t$  in the future? The answer to this question is contained in the conditional distribution of one word on the other,  $P(w_{t+t} | w_t)$  "
- <http://tinyurl.com/y9zml1n>
- " larger groups of cells carry predictive information for hundreds of milliseconds, "

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

### 6.1 reflexive behavioural is unconscious

- $x \rightarrow s \rightarrow a$ 
  - the system state is determined by environmental state, then it is not conscious.
- Policy is not conscious
- Patient B.F.

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

### 6.3 Deterministic vs probabilistic

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

## 7 Comparison with other theories

### 7.1 Intermediate Level Theory

- *" The intermediate level theory originally proposed by Ray Jackendoff and further defended and specified by Jesse Prinz proposes that within the hierarchy of representations that are used to describe the cognitive system, conscious experience occurs only for specific levels of representation. The theory is rooted in Jackendoff's analysis of different cognitive systems such as vision, language, and music and the subsequent observation that consciousness does not arise anywhere within these systems. According to Jackendoff, consciousness is not associated with low-level, nor with high-level representations, but rather with those implying intermediate levels of processing. For instance, in the domain of object recognition, it is assumed that the visual system comprises a low level with local computations of visual features, an intermediate level reflecting binding and object recognition, and a higher level computing viewpoint invariance and representing abstract categories. According to Jackendoff and Prinz, conscious experience is not comprised of a disunified picture of visual features, nor is it represented by viewinvariant categories. Rather it is composed of bound and specific instances of objects that are assumed to be computed at the intermediate level of representation. In an analogous manner, speech perception can be decomposed into three levels: an acoustic representation of speech sounds at the lower level, a high level involving abstract lexical and syntactic categories, and in between a word recognition level relying on phonological representations. This theory explains why the conscious experience associated with speech perception mostly involves phonological representations, rather than other levels of representations. In Jackendoff and Prinz theory, the privileged role of the intermediate level of processing is based on the need for real-time computational efficiency. Indeed, this level of representation is assumed to be the most relevant one regarding ecological and functional needs. Another important aspect of this theory concerns the central role of attention during conscious experience. Here, attention is defined as a selection process that acts as a gate to working memory mechanisms. It performs the function of selecting the relevant information that is amplified afterward and then becomes conscious. Indeed, Prinz acknowledges that activation of an intermediatelevel representation on its own cannot be a sufficient condition for consciousness, given that those representations can be activated during subliminal perception. However, this theory makes the crucial postulate that the amplification of intermediatelevel representations by attention is a necessary and sufficient condition for consciousness. In sum, for each domain of processing, the content of consciousness at a particular moment is supported by a representational structure of intermediate level for that domain, which is selected to enter short-term memory, and enriched by attentional processing."*  
Banks (2009)
- *" 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 compositional 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)

- Jackendoff (1987)
- Prinz (2007)
- 2.5D
- about why it is this level
  - *" Under the computational interpretation, the why question means: In what sense are intermediate 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."*

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)

### 7.1.1 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.

## 7.2 first-order theories

Lau & Rosenthal (2011)

- \* *" The first-order view maintains that conscious awareness is determined by early sensory activity alone, independently of higher-order representations. Thus the crucial difference between the first-order and higher-order views is that the latter but not*

*the former predict that conscious awareness is determined at least in part by pre-frontal and parietal activity.*

*Therefore, in the group of rst-order theorists we include here not only those commonly labeled as such in the philosophical literature [11] but also theorists, such as Ned Block [12,13], who hold that the phenomenal qualities of awareness depend on the biological substrate rather than merely the content of the rst-order representations, as well as scientists who hold that visual awareness depends on activity in content-specic regions of the extrastriate cortex [24,25] or on feedback from these regions to the primary visual cortex [26].*

*Most rst-order theories explain the difference between awareness and unawareness by positing that the latter is associated with weak information [29] or with representations in alternative (e.g. subcortical) sensory pathways. Thus this view suggests that conscious and unconscious processing will have different functional consequences. Whereas some rst-order theorists hold that some higher cognitive functions can be performed without conscious awareness [30,32], most rst-order theorists take a difference in perceptual task performance (e.g. hits vs misses in a detection task) as evidence for a difference in conscious awareness [12,13,25,26,33]. In other words, they typically associate task performance with awareness."*

## 7.3 Global Workspace Theory (GWT)

### 7.3.1 Global availability and Broadcasting

### 7.3.2 Code translation

- neural coarse-graining can naturally solve this problem.
- coarse-graining between levels can be thought as code translation

is this true?

### 7.3.3 Stability

- the presence of contexts as stable coalitions shaping access to the workspace.

### 7.3.4 broadcasting

### 7.3.5 Ignition

#### [[ Explain ignite ]]

- GWT is one of the current popular consciousness theories.
- *"AI system and corresponds to our first functional definition of consciousness: global availability"*  
Dehaene *et al.* (2017b)
- C1: Global availability of relevant information The need for integration and coordination  
Dehaene *et al.* (2017b)
- Evidence for integration and broadcasting
- Stability as a feature of consciousness *" meta-stability seems to be necessary for the nervous system to integrate information from a variety of modules and then broadcast it back to them, achieving flexible cross-module routing."*
- *"integration of computational resources in a large-scale coordination and for the exchange of information among processors"*
- *"Grounded on the distinction between conscious and unconscious processes, Bernard Baars global workspace theory is one of the most influential cognitive theories of consciousness. This theory relies on the metaphor of a theater. In this theater, unconscious specialized processors (equivalent to modules) are assumed to be the actors and the audience. While the audience represents the set of passive processors, actors represents active processors playing on the stage of the theater (i.e., the workspace). These actors are engaged in a competition for being seen by the audience: by broadcasting their information they actually compete for more broadcasting. Active processors with the highest coherent activity can form local coalitions that strengthen*

them in this competition process. The strongest coalition in this competition dominates the workspace, in a winner-take-all fashion, and corresponds to the content of consciousness. The workspace is equated by Baars to working-memory, in which only the most active content becomes conscious. Additionally, the dominant coalition benefits from global broadcasting, which allows it to recruit new processors from the audience in order to form a global coalition. Here, consciousness allows for the integration of computational resources in a large-scale coordination and for the exchange of information among processors that would otherwise remain separated. In this theory, each processor can operate in the conscious mode if it benefits from global broadcasting through the workspace, or it can operate in the unconscious mode when disconnected from the workspace. An important feature of the global workspace theory is the presence of contexts as stable coalitions shaping access to the workspace. Contexts are constituted of unconscious processors reflecting, in a hierarchical manner, our expectations, our beliefs, our goals, and ultimately our self. In particular, attention is implemented as a goal context in this theory. It is described as a mechanism that controls access to the workspace, acting as a filter and biasing the competition process toward a particular set of actors. At any given moment, the dominant coalition is under the spotlight of attention, and its informational content becomes the content of conscious experience. A crucial aspect of Baars theory is that it avoids the problem of the homunculus by reducing it to an audience of multiple unconscious processors. Here, there is no need for a hypothetical single conscious observer in the system, and thus there is no issue of infinite regression with a homunculus inside another homunculus. Instead, consciousness is considered to reflect the global broadcasting of information to an audience of unconscious processors. As the audience is unconscious, unsupervised, and receptive rather than attending to the information, it does not constitute an internal homunculus."

## 7.4 Higher Order Thought theory of consciousness

[[ People think a high order representation receives information from the first order representation. This may be a misunderstanding. This may be the result of neural coarse-graining. Because contents in information closure look like watching fine-grained information, people may have an wrong impression about it ]]

- 
- Information closure directly is critical for other systems to construct forward models.
  - Therefore, most of the high level cognitions can utilise the state of information closure.
  - For example, mental planning needs a precise forward model to infer the environmental dynamics.
  - [[ This is way HOT always involve consciousness ]]

- Rosenthal (2005)

" Higher-order theories of consciousness argue that conscious awareness crucially depends on higher-order mental representations that represent oneself as being in particular mental states."

Lau & Rosenthal (2011)

" According to this view, humans not only have first-order non-conceptual and/or analog perceptions of states of their environments and bodies, they also have second-order non-conceptual and/or analog perceptions of their first-order states of perception. And the most popular version of higher-order perception theory holds, in addition, that humans (and perhaps other animals) not only have sense-organs that scan the environment/body to produce fine-grained representations that can then serve to ground thoughts and action-planning, but they also have inner senses, charged with scanning the outputs of the first-order senses (i.e. perceptual experiences) to produce equally fine-grained, but higher-order, representations of those outputs (i.e. to produce higher-order experiences). A version of this view was first proposed by the British Empiricist philosopher John Locke (1690). In our own time it has been defended especially by Armstrong (1968, 1984) and by Lycan (1996)."

<https://plato.stanford.edu/entries/consciousness-higher/#SelRepHigOrdThe>

" neutral about whether conscious awareness adds significant utility or immediate impact on behavior and task performance [1,27]. This is because the view assumes that task performance in most perceptual and cognitive tasks depends mainly on rst-order rather than higher-order representations. Because conscious awareness can differ even if all rst-order representations remain completely un-

changed, such awareness itself might serve little function [1,27]."  
Lau & Rosenthal (2011)

My theory would have similar prediction. Because conscious contents are coarse-grained results from low-level information. It is possible that more than two low-level neural states all coarse-grained to a high-level state.

#### 7.4.1 Metacognition

" Metacognition: cognition that is about another cognitive process as opposed to about objects in the world. In this article we use it mainly to refer to sensory metacognition: a cognitive process that concerns the quality or efficacy of a perceptual process. The capacity of sensory metacognition is sometimes empirically assessed by the correspondence between subjective report and task performance: how closely are they associated with each other on a trial-by-trial basis"  
Lau & Rosenthal (2011)

### 7.5 Sensorimotor contingency

### 7.6 Predictive Brain

\* Because of this XXXX, the representation has the maximal non-trivial information closure (ntic) also has the maximal predictive power of the external environment.

I need math here!!

\* Therefore, it's reasonable to assume that if evolution wants to build predictive mechanisms, the mechanisms should be settled on the information at the coarse-grained (cged) level.

\* More specifically, a forward model, which received state and action as input and compute transition and then output next states, should be built on the representation that forms the information closure.

\* Therefore, the state of predictive model can be conscious while it is linked to information closure representation.

\* Note that, a system showing some predictive power is not sufficient to be conscious.

\* For example, the proportionalintegralderivative controller (PID) controller shows predictive behaviours due to its derivative which computes and predict error value in the future. Derivative action predicts system behaviour and thus improves settling time and stability of the system.

\* However, the state of the whole PID system still be determined by the inputs, i.e. the states of the controlled process, thus cannot complete information closure.

. It can compute position, velocity, and acceleration of objects, and output control signals (action). However, the state of the PID controller is not information-closed. The state still fully determined by the input signals. Therefore, it fails to reach information closure (ic). Hence, according to our theory, the PID system is not conscious.

\* Similarly, the neural system also has many circuits that show predictive power but do not have ic.

#### 7.6.1 statistical learning/prediction is not conscious

- considers the brain as a predictive machine
- According to predictive processing account, conscious contents are the inference results (posterior from Bayesian inference framework) of predictive mechanism.
- However, we can see that not all predictive mechanisms in the neural system can be consciously aware of.
- So what's the critical difference between them in terms of consciousness?
- We argue that when predictive mechanisms utilise the information in the information closure, the predictive result is conscious.
- Because the forward model built on the information closure representation can be seen as part of the closure.
- we can consider two scenarios
  - Predictive information is encoded in neuron populations but there is high entropy of conditional probability which means it's not closure. Therefore, the conscious level could be very low even though it still has some degree of predictability.
  - Another scenario is a forward model but its state fully depends on external sensory input. This system is not information closed and therefore is not conscious.

IS THIS TRUE? NEED DISCUSSION

- Based on this point of view, predictive power for a system does not guarantee consciousness.
- For example, in active inference , action is guided by prediction error so the state of the system is non-closed.

Ref: active inference

## 7.7 IIT

Help from Jun?

### 7.7.1 Information Integration

- **[[ Information integration without awareness ]]** Mudrik *et al.* (n.d.)
- *"consciousness is necessary for high-level but not low-level semantic integration"* Mudrik *et al.* (n.d.) I would argue an inverse relationship: "High-level information processing is necessary for consciousness." This is because higher-level information processing is more likely to create an information closure representation.
- Similarly, it is not *"consciousness is necessary for multisensory integration"* . It should be multisensory integration is more likely to create an information closure representation. Because in most situations, multisensory signals are generated from hidden causes which may have deterministic dynamic and relation with each other. Therefore, to infer the hidden causes and represent the deterministic relations can form information closure representation which is conscious.

### 7.7.2 IIT

### 7.7.3 Hoel's theory

Help from Marting?

## 7.8 Internal simulation and self-modeling

## 8 Counter-intuitive prediction

### 8.1 information closure

- We predict that through coarse-graining, conscious agents create

### 8.2 Prediction on split-brain (vs IIT)

\* IIT suggests that cutting, two subparts may both still have MIP, therefore, have two conscious mind.

\* However, our theory suggests that if the cut largely destroy the ic, then both part may not be conscious.

## 9 Future work

- Lack of direct evidence because no one focus on information closure

### 9.1 Brain area

### 9.2 How to coarse-grain in the neural system

- *" not all causal relationships (or relationships of nomological dependency) among micro-events aggregate up to causal relationships among coarse-grained macro-events that are constituted by those micro-events. Instead, whether one gets causation at the macroscopic level will depend (among other things) on the particular coarse-graining that is chosen."* Price & Corry (2007)
- *" It may seem surprising, even counterintuitive, that causal and statistical dependence relationships involving fine grained microscopic variables do not automatically show up in causal and dependence relationships among the macroscopic variables that are realized by the fine grained"*

*variables."*

Price & Corry (2007)

- Jonas & Kording (2017) well demonstrated that it is difficult to understand higher-level information processing by the current neuroscience methods.

## 10 Evolution of conscious mind (not sure but if I have time)

Dennett (2008)

## 11 Conclusion

- Why is this theory better than other theories of consciousness
- 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, Fougner, 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.
- Banks, William P. 2009. *Encyclopedia of consciousness*. Vol. 1. Academic Press.
- Bertschinger, Nils, Olbrich, Eckehard, Ay, Nihat, & Jost, Jürgen. 2006. Information and closure in systems theory. *Pages 9–21 of: Explorations in the Complexity of Possible Life. Proceedings of the 7th German Workshop of Artificial Life*.
- Binder, Marc D, Hirokawa, Nobutaka, & Windhorst, Uwe. 2009. *Encyclopedia of neuroscience*.
- Dehaene, Stanislas, Lau, Hakwan, & Kouider, Sid. 2017a. What is consciousness, and could machines have it? *Science*, **358**(6362), 486–492.
- Dehaene, Stanislas, Lau, Hakwan, & Kouider, Sid. 2017b. What is consciousness, and could machines have it?
- 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.
- Goldwyn, Joshua H., & Shea-Brown, Eric. 2011. The what and where of adding channel noise to the Hodgkin-Huxley equations. *PLoS Computational Biology*, **7**(11).
- 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.
- Jonas, Eric, & Kording, Konrad Paul. 2017. Could a neuroscientist understand a microprocessor? *PLoS computational biology*, **13**(1), e1005268.
- Lau, Hakwan, & Rosenthal, David. 2011. Empirical support for higher-order theories of conscious awareness. *Trends in cognitive sciences*, **15**(8), 365–373.
- 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.



- Mudrik, Liad, Faivre, Nathan, & Koch, Christof. Information integration without awareness. *Trends in Cognitive Sciences*.
- 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.
- Price, Huw, & Corry, Richard. 2007. *Causation, physics, and the constitution of reality: Russell’s republic revisited*. Oxford University Press.
- Prinz, Jesse. 2007. The intermediate level theory of consciousness. *The Blackwell companion to consciousness*, 257–271.
- Rosenthal, David M. 2005. *Consciousness and mind*. Oxford University Press.
- 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.
- Stahl, Aimee E., & Feigenson, Lisa. 2015. Observing the unexpected enhances infants’ learning and exploration. *Science*, **348**(6230), 91–94.
- 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.
- White, John A., Rubinstein, Jay T., & Kay, Alan R. 2000. Channel noise in neurons. *Trends in Neurosciences*, **23**(3), 131–137.