A neural coarse graining theory of consciousness/ Non-trivial Informational closure through neural coarse-graining as the Fundamental Informational Correlates of Consciousness

Acer Y.C. Chang[*1], Martin Biehl[†1], Yen Yu[‡1], and Ryota Kanai[§1]

[1]ARAYA, Inc., Tokyo, Japan

May 1, 2019

Contents

[*]acercyc@araya.org
[†]martin@araya.org
[‡]martin@araya.org
[§]kanair@araya.org

[Need Help]: Help from Martin?

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 co-vary with information encoded in coarse-grained neural states such as neural population codes. We argue that the neural states within the scope of the informational 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 informational closure in deep hidden layers. Our preliminary results show that informational 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, informational closure, neural coarse-graining, level of analysis

Not decided terminology

- information * in the NTIC process
- Full and short name of this theory
- core "?" (entity) for conscious content
- our theory? framework? proposal?
- system? process?
- variable? node? element? member?

1 Introduction

Imagine you are a neuron in Alice's brain. Your daily work is to collect neurotransmitters through dendrites from other neurons, accumulate membrane potential, and finally send signals to other neurons through action potentials along axons. However, you have no idea that you are one of neurons in Alice's supplementary motor area and involved in many motor control processes for Alice's action, for example, grabbing a cup. You are ignorant of intentions, goals, and motor plans that Alice has at every moment even though you are part of the physiological substrate responsible for all those actions. A similar story also applies to Alice's conscious mind. To grab a cup, for example, Alice is conscious of her intention and visuosensory experience of this action. However, her conscious experience does not show the dynamic of your membrane potential and about action potentials you send to other neurons every second. That is, information in Alice's conscious experience does not contain information involved in your information processing.

[[maybe need a transition sentence here]]

It seems to be true that we are unable to consciously access information processed at every scale in the neural system. Only information at certain levels of coarse-graining can become conscious contents. At microscopic levels of neural systems, information processing in individual neurons is highly noisy and state transitions of neurons are extremely stochastic (Goldwyn & Shea-Brown, 2011; White *et al.*, 2000). However, instead of stochastic action potentials and fluctuation of random noise at microscopic levels, what we are aware of in our conscious mind shows astonishing stability and robustness against the ubiquitous noise in the neural system (Mathis & Mozer, 1995). Likewise, conscious experience does not associate with states of the whole brain level because we have known that some parts of the neural system contribute very little to conscious experience (the cerebellum for example (Lemon & Edgley, 2010)). Additionally, we do not experience "super-consciousness" when we have massive inter-individual interactions with other human-beings. Even though evidence has shown several types of collective information processing at human social groups levels exist (Malone & Bernstein, 2015), information processed at this level seems to depart from our individual conscious contents.

In this article, we propose a new theory of consciousness, calledThe Neural Coarse-graining Theory of Consciousness (NCGT). This proposal states that consciousness does not rest on particular neural architectures or locations in the neural systems. Instead, we argue that consciousness correlates with information processing at a certain coarse-grained level where the neural systems create a high degree of non-trivial informational closure (NTIC). We further postulate that the *level* of consciousness corresponds to the degree of NTIC and the *contents* of consciousness corresponds to the state of the NTIC process (for the level versus content of consciousness see Laureys (2005); Overgaard & Overgaard (2010)).

Our theory takes a strong assumption that consciousness is information-based, information correlates of consciousness (ICC)

Complete this argument

Our theory rests on two core concepts: non-trivial informational closure and neural coarse-graining. In this article, we first introduce non-trivial informational closure and its importance to information processing for human scale agents. We next argue that through coarse-graining the human neural system can achieve non-trivial informational closure at a specific coarse-grained level. We then propose that the state of the informational closure correlates with the state of consciousness for both levels and contents which can be explained, quantified, and predicted computationally by this theory. We also address how this theory can concisely explain empirical results from previous studies and reconcile several current major theories of consciousness. Finally, we provide predictions and testable approaches for this theory.

[[Do I need a figure?, Something like this one?]]

To Be Modified ...

Levels of Representational Organization 267

Figure 10.1
Brain function is characterized by several aggregate levels of organization, rather loosely defined at different scales of measurement, beginning at the molecular level of ion channels and other proteins up to the highest level of cognition and behavior mediated by connected, macroscopic brain systems. Adapted from Shepherd (1990). Note the circularity in a person looking at his own ion channel activity displayed on an oscilloscope.

Pennartz (2017)

Figure 1: xxxx

2 Non-trivial informational closure

> -
>
> I am still not sure we should go for NTIC without describe my intuition about consciousness and closure? Or we should start with my intuition. Or we should postpone to explain the intuition in Chapter 4?

The notion of non-trivial informational closure (NTIC) is introduced by Bertschinger *et al.* (2006). The concept of closure is closely related to system identification in systems theory. One can define a system from its environment by computing the closureness of the system (Maturana & Varela, 1991; Rosen, 1991; Pattee, 2012; Luhmann, 1995). The closureness can be further quantified by information theory.

-

Writing Materials

- Consciousness as an intrinsic information. Therefore, information should be self-define. This means the boundary of conscious content should be related to informational closure.

- [[should we describe the intuition of boundary problem here?]]

- ===

- *" 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)
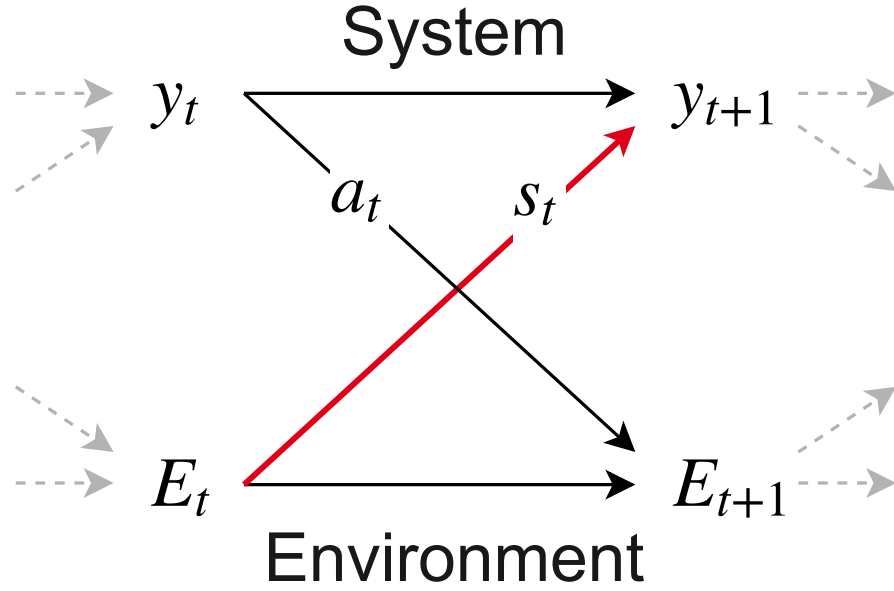
Figure 2: Figure adapted from Bertschinger *et al.* (2006)

- *" 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."*
(Bertschinger *et al.*, 2006)

- *" 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)

Writing Materials -

Information flow J_t from the environment E to a system Y at time t can be defined as the conditional mutual information I between the system future state Y_{t+1} and the current environment current state E_t given the current system state Y_t

$$
\begin{aligned}
J_t(E \to Y) &:= I(Y_{t+1}; E_t | Y_t) \\
&= I(Y_{t+1}; E_t) - (I(Y_{t+1}; Y_t) - I(Y_{t+1}; Y_t | E_t))
\end{aligned}
\tag{1}
$$

Bertschinger *et al.* (2006) defines that a system is informational closured when information flow from the environment to the system is close to zero.

$$
J_t(E \to Y) = 0
\tag{2}
$$

Achieving information closure (minimising J_t) could be trivial, which means the environment and the system are entirely independent of each other.

$$
I(Y_{t+1}; E_t) = 0 \quad \Rightarrow \quad J_t(E \to Y) = 0
\tag{3}
$$

6

However, informational closure can be achieved non-trivially. In the non-trivial case, the system trying to encode information about the environmental dynamics can also be informational closured. This means

$$I(Y_{t+1}; E_t) > 0 \tag{4}$$

This also implies

$$I(Y_{t+1}; Y_t) - I(Y_{t+1}; Y_t|E_t) > 0 \tag{5}$$

And, non-trivial informational closure can be defined as

$$
\begin{aligned}
NTIC \quad &:= I(Y_{t+1}; Y_t) - I(Y_{t+1}; Y_t|E_t) \\
&= I(Y_{t+1}; E_t) - I(Y_{t+1}; E_t|Y_t)
\end{aligned}
\tag{6}
$$

And, Maximising NTIC is equivalent to

$$
\begin{aligned}
\text{maximise} \quad &I(Y_{t+1}; Y_t) \qquad \text{and} \\
\text{minimise} \quad &I(Y_{t+1}; Y_t|E_t)
\end{aligned}
\tag{7}
$$

This implies the system contains in itself the information of its own future and the self-predictive information is gained from the information about the environment. Therefore, to achieve NTIC, the system can internalizes and synchronises the dynamics of the environment, i.e. modelling the environment. Furthermore, having high degrees of NTIC entails having high predictive power about the environment. This gives biological agents a great functional and evolutional advantage to achieve NTIC in their information processing systems. Practically, Guttenberg *et al.* (2016) proposed that NTIC can be achieved by maximising the predictive power about the environment and the self-predictive information of the system concurrently.

Though the significant advantage for biological systems to achieve NTIC, to our knowledge, studies directly examining NTIC in biological systems are limited. However, recent studies have showed relevant properties in the visual system of salamanders. By examining the salamander retina, in two studies (Palmer *et al.*, 2015; Sederberg *et al.*, 2018) the authors found that the a large groups of neural populations of retinal ganglion cells encoded predictive information about external stimuli also had high self-predictive information about their own future states (see Figure 3), therefore, in line with the characteristic of NTIC. Therefore, the self-predictive information of the neural populations can guides prediction about external stimuli without reference to stimulus parameters explicitly. Moreover, the internal predictability can be generalised to different stimulus classes suggesting the tendency to internalise the information of environmental dynamics in the neural populations. [[Add Some ending sentences]]
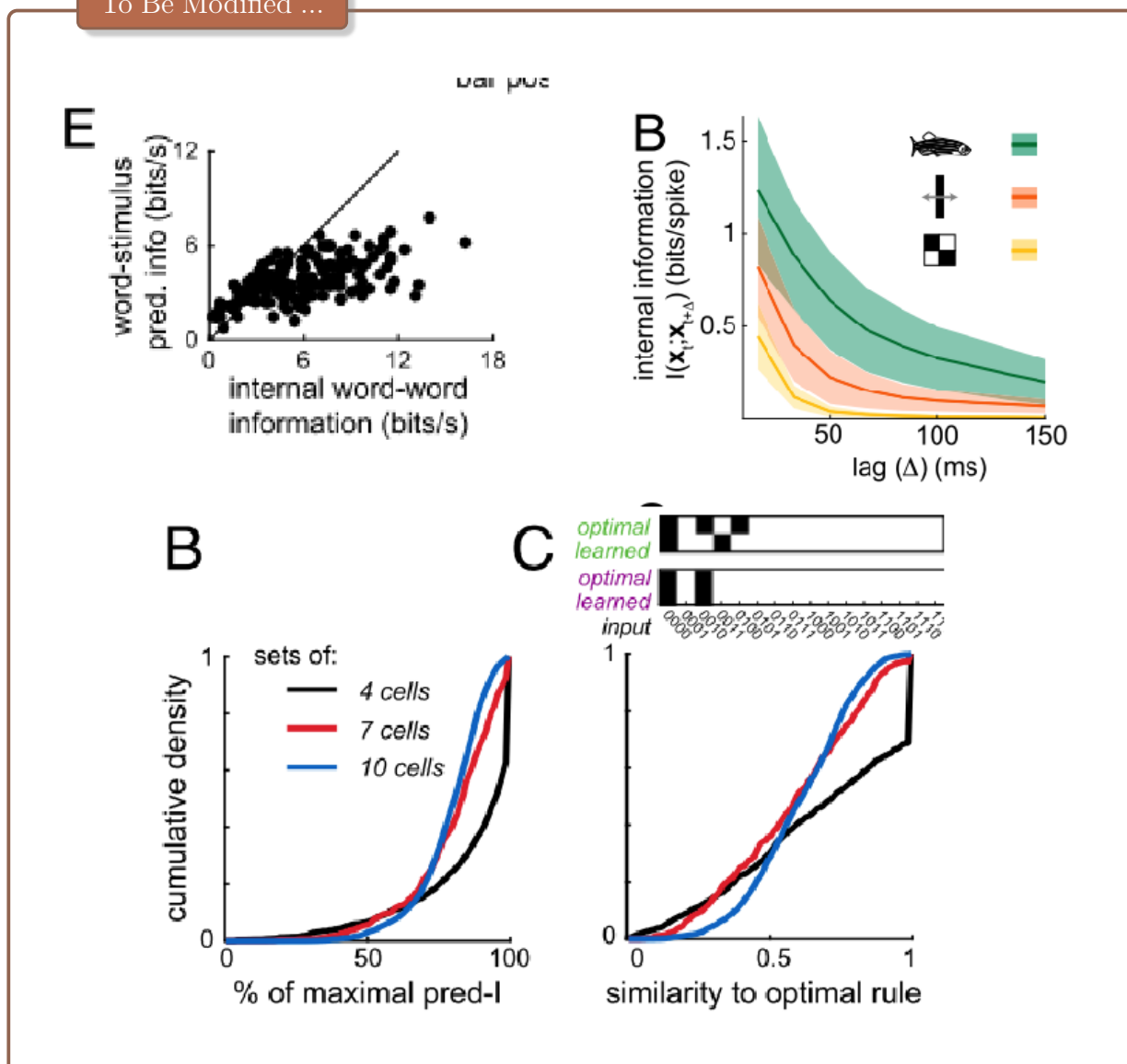
Figure 3: Figure adapted from Sederberg et al. (2018)

2.1 Informational closure and level identification

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

3 Neural coarse-graining

Nevertheless, to achieve information closure (so as NTIC) in highly stochastic systems (e.g., the neural system) is challenging. The current states of the system have low predictability of its future states due to the stochasticity. Information processing in the neural system is suffered from multiple noise sources throughout, including sensory, cellular, electrical, and synaptic noises. Previous studies have shown that states of neurons have large trial-to-trial variability at the cellular level. That is, the same pattern of activity never occurs twice even when the same stimulus is presented. Neurons are inevitably subject to thermal fluctuations and other physical noises (Faisal et al., 2008). Furthermore, environmental information is always only partial observable to individual sensors and neurons which only receive little amount of information from the environment. Therefore, self and environmental predictive information is difficult to be represented at the cellular level. It is implausible to construct NTIC processes at microscopic levels.

However, we argue that, by coarse-graining microscopic neural states, the neural system is able

to achieve NTIC at certain coarse-grained macroscopic levels. To encode information in coarse-grained variables is a way to encode stable and robust information in neural system and overcome the difficulties to achieve NTIC at microscopic levels. Coarse-graining in this paper is referred to many-to-one functions that map microscopic states to macroscopic states.. In other words, a number of different micro-states realise the same value of the macro-variable (Price & Corry, 2007). This is a common strategy that we can see in large number of previous neurophysiological studies. For example, in the temporal domain, neural spikes across time can be coarse-grained as spiking rate variables, known as rate coding (Adrian, 1926; Gerstner & Kistler, 2002; Maass & Bishop, 2001; Panzeri *et al.*, 2015; Stein *et al.*, 2005). Neurons encode stimulus intensity by increasing or decreasing firing rate. (Kandel *et al.*, 2000). In this coding scheme, individual neural spikes in a transient time span is not meaningful and stochastic (Stein *et al.*, 2005). In contrast, rate codes are relative stable and are highly robust against noise causing variation of inter-spike intervals.

Another example of coding strategy using coarse-graining is population coding. In this coding scheme, information is represented by joint spatiotemporal activities of several neurons. Therefore, the information is encoded in coarse-grained variables whose values are jointly determined by neural activation patterns within a neural population (Kristan Jr & Shaw, 1997; Pouget *et al.*, 2000; Binder *et al.*, 2009; Quian Quiroga & Panzeri, 2009). In such case, activities of every single neuron is not informative and stochastic. Instead, coarse-grained variables (such as population averaging and population vector) are able to represent information stably. Encoding information at coarse-grained levels allow the neural system ignores certain causal interactions relevant to fine-grained level of description (Woodward, 2007).

It has been empirically implemented that NTIC can be achieved by coarse-graining snesory inputs in artificial neural networks. (Guttenberg *et al.*, 2016) proposed a "neural coarse-graining (NCG)" architecture (fig.4) which aims to captures the emergent large-scale dynamics in the environment. The NCG network learns to coarse-grain input signals and extract latent parameters which are both predictive and predictable in the input signal and therefore achieves NTIC by building a higher-level representation of the underlying dynamics in the deep layers. More importantly, using this architecture, agents can process information and act at coarse-grained macroscopic levels.

It is important to note that NTIC should not be a monotonic function of scale of coarse-graining. Overly coarse-graining system states can lead to low degrees of NTIC. For example, in an extreme scenario, when all microscopic states map to a single macroscopic variable, the macroscopic level does not have enough information capacity and thus losses the self and environmental predictive power. This notion is also supported by empirical evidence. Sederberg *et al.* (2018) showed that the efficiency of reading out self-predictive information of neural population is higher when cell number in populations is 4 than 7 and 10-cell sets. This implies an optimal coarse-graining scale for NTIC processes.

At the physical scale which human being live in, the environmental dynamics is nearly deterministic following the laws of classical physics. It gives agents living at this physical scale great advantages if the agents can build NTIC process internally. Therefore, to coarse-grain microscopic states is necessary

---

Writing Materials

- *" Coarse-graining: a method of reducing the complexity of a system by treating groups of atoms/molecules as single quasi-particles CG: coarse-grained"*

- neural system use population coding and temporal coding to fight against noise.

- Therefore, even though the low level information processing is noisy, it is possible that the neural system achieve NTIC at a coarse-grained level.

- This implies that the information about the environmental dynamic should be represented at a certain coarse-grained level rather then other levels.

- The environment that human scale agents interact with is approximately govern by Newton's law and therefore the state transitions are nearly deterministic.

- [[Need something here]]

- Hence, to coarse-grain states of microscopic levels is necessary to represent and model the human-scale environment state and dynamics.

- Coarse-graining is necessary to resist noise

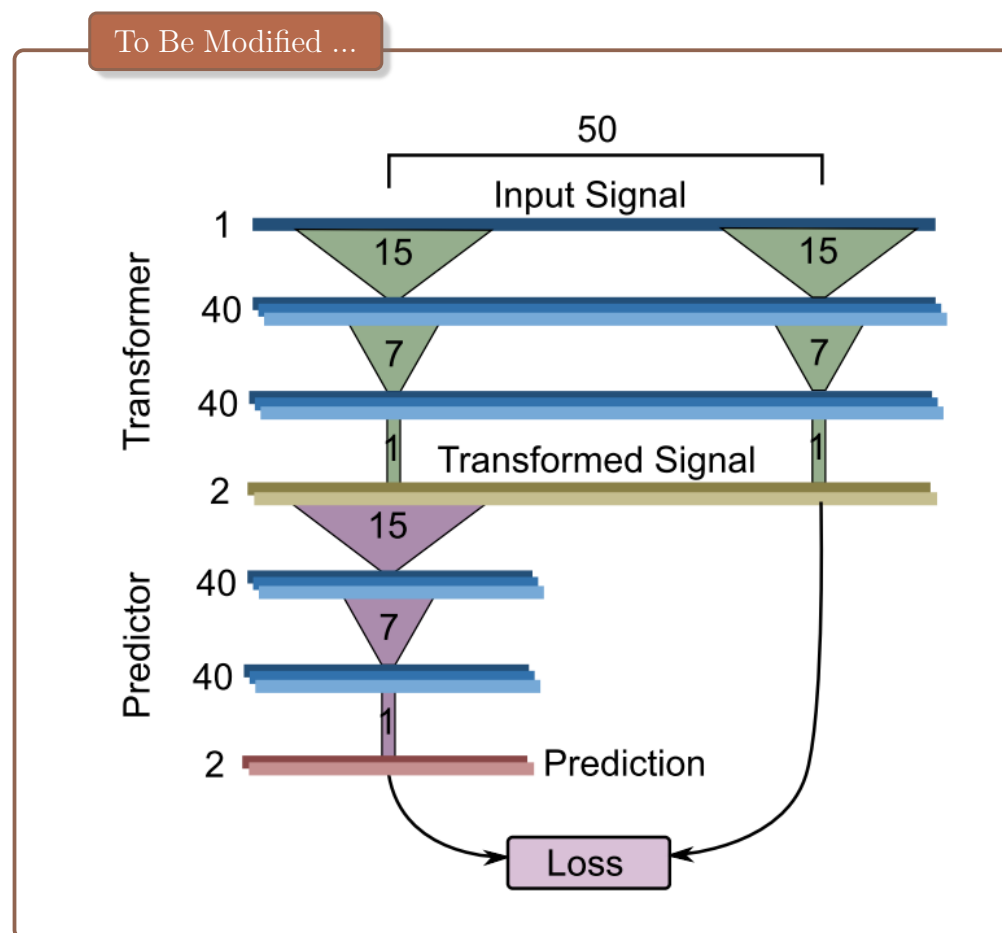- *" The population will therefore be relatively insensitive to the loss of cells or noise in a small*

Figure 4: NCG [[I need to simplify this figure]](Guttenberg *et al.*, 2016)

number of cells"
Eurich & Wilke (2000)

-

Writing Materials

- Studies have shown that coarse-graining is a common way that the neural system cope with inevitable noise at microscopic levels.

- *"Because the sequence of action potentials generated by a given stimulus varies from trial to trial, neuronal responses are typically treated statistically or probabilistically"*

- Note that, information processing related to consciousness should not be situated at this level. Otherwise our conscious percept should be very noisy and unstable.

-

Writing Materials

- Population coding
  - review ref
    * Stanley (2013)
    * Quian Quiroga & Panzeri (2009)
  - *" Averaging is used in many neural systems in which information is encoded as patterns of activity across a population of neurons that all subserve a similar function (for example, see REFs 140,141)"*
    140. Georgopoulos, A., Schwartz, A. & Kettner, R. Neuronal population coding of movement direction. Science 233, 1416–1419 (1986). 141. Lee, C., Rohrer, W. H. & Sparks, D. L. Population coding of saccadic eye movements by neurons in the superior colliculus. Nature 332, 357–360 (1988).
  - *" Electrophysiological studies in the monkey have shown that behaviourally relevant signals are averaged not only across neuronal populations but also over time in the formation of a behavioural decision 156 ."*
    156. Gold, J. I. & Shadlen, M. N. Representation of a perceptual decision in developing oculomotor commands. Nature 404, 390–394 (2000).
  - *"A distributed representation of information of this type is more robust to the effects of noise."*

  - *" 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)"*

- sensory hierarchy
  - sensory hierarchy also shows characteristic of coarse-graining
  - Due to partial information that sensors can receive
  - To form complex representation like shape, objects, coarse-graining is necessary.
  - Through coarse-graining, <mark>higher level</mark> cortical areas can integrate those partial information to infer hidden causes.

> might be confusing, need to think about the terminology here

-

Writing Materials

- Achieving NTIC is the key objective to to extract out latent control parameters in the environment.

- Through two objectives during training, the model can achieve NTIC.

- this can be achieved by maximising the self-predictive power of deep layers and its entropy through all training data.

- With the objectives, NCG network learns to coarse-grain input signals and extract latent parameters which are both predictive and predictable in the input signal and therefore achieves NTIC by building a higher-level representation of the underlying dynamics in the deep layers.

- Agents can process information and act at abstract levels

Writing Materials -

-

Writing Materials

- At the physical scale which human being live, the environmental dynamics is nearly deterministic following the laws of classical physics. This means that it is possible for agents living at this physical scale to build NTIC process internally.

- When the information in individual sensory channel is noisy and partial, coarse-graining becomes necessary for achieve NTIC.

Writing Materials -

4 A neural coarse graining theory of consciousness

iteration 5

In this section, we propose a new theoretical framework of consciousness {The Neural Coarse-graining Theory of Consciousness} . This framework proposes that NTIC is the fundamental properties for information processing to be conscious. Additionally, coarse-graining is necessary to achieve NTIC in various of different systems.

We need a good name

We starts our argument from the first presumptive fact that every human conscious precept is highly informative. Our conscious experience always consists of high multidimensional information and usually involves information from multiple modalities. The notion of information referred to in this article is considered as resolved uncertainty. Being a certain conscious state rules out other enormous possible conscious experience in the state space of conscious experience. Therefore, every conscious percept resolves a huge amount of uncertainty and provides considerable information.

This notion is in agreement with one of the axiom, *information*, in Integrated Information Theory (IIT 3.0) which claims that *"...an experience of pure darkness is what it is by differing, in its particular way, from an immense number of other possible experiences..."* (Oizumi *et al.*, 2014, p. 2).

Based on this presumptive fact, we postulate that the amount of information possessed of a conscious percept should be equivalent to information involved in a particular event or entity in the physical reality. Strictly speaking, we recognise information is a common language between conscious experience and the physical reality. We further take a radical standpoint and postulate that consciousness is fundamentally associated with information, i.e. information correlates of consciousness (ICC), Chalmers (1996); Tononi (2004); Gamez (2011, 2016)) and discard the notion of neural correlates of consciousness(NCC).

Next, we next consider the second presumptive fact that the information in consciousness is bound. Even though information involved in every conscious percept is extremely rich, it is still limited and we do not consciously feel everything in this universe. This leads to the boundary problems of consciousness Goff (2006). There should exist information boundaries for every conscious entity. Presuming consciousness is information-based phenomenon, we can identify the boundaries with measures from information theory. Particularly, in this case, boundaries can be defined by information closure. Therefore, information closure should be observed in conscious entities in the physical reality.

[[REWORDING | informational closure also suggests that there is no information flow from the lower to the higher level. Knowledge of the microscopic levels will not improve predictions of the

future states of the macroscopic levels (PFANTE *et al.*, 2014)]]. Information closure also implies that the system is ignorant of information outside of the system. Therefore, if a coarse-grained level achieves information closure, it suggests the NTIC process at the coarse-grained level is ignorant of information processing happening at microscopic levels and also higher coarse-grained levels. This virtually makes the informational closed macroscopic levels as a complete reality.

Finally, our conscious experience always reflects some degree of information about environment and bodily states, i.e. non-trivial information. To account for the information-based considerations, we propose that NTIC is a fundamental informational property associated with consciousness.

Writing Materials

For example, considering two physical systems (Alice and Bob) and there is no interaction (or information passing) in between. Assuming that Alice is conscious, the conscious experience that Alice has should not involve any information about Bob.

Similarly, we can apply this scenario to brain processing between two regions. We can consider a thought experiment. Imagining that in the future, one poor subject attend this experiment. We remove the subject's visual cortex from other brain areas. However, we keep all the connection between the two parts of the brain intact by replacing dendrites and axons between the two parts with micro electronic cables. Because these cables are very long, we can move the visual cortex to the surface of Mars. Now the experimenters give visual cortex a input signal. If the signal travels with the speed of light, it still takes three seconds to reach the other part on the earth. We can ask: Between the experimenters send the input signal and the information reach the part on the earth, what visual conscious content does the subject experience in the three seconds? We presume that in this three seconds subject's conscious experience should not contain any information about the visual input. Otherwise, the subject may be able to use the information to perform subsequent actions. This violates physics. This may look like a trivial questions.

However, in many systems, it is difficult to achieve NTIC at microscopic levels due to stochasticity. Through coarse-graining, NTIC processes become possible at macroscopic levels. Therefore, coarse-graining is necessary for many biological systems to implement NTIC processes, especially for the neural system (Fig.). In the following, we explained this framework and provide mathematical definitions and implications of ICC.

Writing Materials -

Writing Materials

- As we mentioned in Chapter 3, to form non-trivial informational closure has a huge evolutionary advantage for the creatures interacting with the environment at our scale.
- The physical environment dynamic at the human scale are nearly deterministic.
- To maximise fitness for individual human-being, neural system should try hard to infer and model the deterministic physical rules.
- This drives the neural system to create a non-trivial informational closure.
- Therefore, NTIC emerges from evolutionary processes is functionally plausible.
- Importantly, at the microscopic level, it is impossible to achieve NTIC as we mentioned in chapter 2.
- However, at a macroscopic scale, information processing can achieve non-trivial informational closure (Fig.)
- Therefore, the NTIC correlating conscious experience must settled at coarse-grained macroscopic levels.
- Through coarse-graining, we can define a new set of variables.
- information is scale-dependent
- If the new set of variables are informational closure, then this creates a new *reality* for all the variables inside.
- *Reality* here means that the all the future states of the system are driven by the current information (include variables and the relationships) within the boundary.
-
- Relationship between coarse-graining, scale and interaction with environment

- *" 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.
- To summaries, in our framework, consciousness is
  - Intrinsic
  - Informative
  - scale-dependent
- We claim that the state of a coarse-grained scale which realises non-trivial informational closured correlates the state of consciousness.
- We further postulate that level of consciousness corresponds to informational closureness of processes.
- The contents of consciousness correspond to the states of the non-trivial informational closure.
- In the following part, we....   `todo`
- This is the same as "if a tree falls in a forest" thought experiment.
- Therefore, to understand this system, you don't need more information from external environment.
- It also means everything in this system is self-defined.
-
  `I need to somehow acknowledge Pennartz (2017)`
  - *" In conclusion, for a well-behaved representational system, it does not matter whether a neural activation pattern represents an external world feature truthfully or not, as long as the system represents the feature consistently over time and in sensory space."*

  - pennartz levels are defined by functional elements call "functional ensembles". Pennartz (2017) *"This delineation is different from a scale-based distinction of levels, because different levels are distinguished here based on function , i.e., what each level can accomplish in terms of computation and representation, culminating in perception at the highest level."*

- *" Multilevel views on consciousness and cognition have been expressed before by many other theorists, such as Oppenheim and Putnam (1958), Attneave (1961), Hofstadter (1979, 1985), Churchland and Sejnowski (1992), Wimsatt (1994), and Lycan (1996)."*
  Pennartz (2017)
  - Oppenheim , P. , & Putnam , H. (1958). Unity of science as a working hypothesis . In H. Feigl , G. Maxwell , & M. Scriven (Eds.), Concepts, theories, and the mind – body problem (pp. 3 – 36). Minneapolis : University of Minnesota Press .
  - Attneave , F. (1961). In defense of homunculi . In W. A. Rosenblith (Ed.), Sensory communication (pp. 777 – 782). Cambridge, MA : MIT Press .
  - Hofstadter , D. R. (1979). Godel, Escher, Bach, an eternal golden braid . New York : Basic Books .
  - Churchland , P. S. , & Sejnowski , T. J. (1992). The computational brain . Cambridge, MA : MIT Press .
  - Wimsatt , W. C. (1994). The ontology of complex systems: levels of organization, perspectives, and causal thickets. Canadian Journal of Philosophy , 20 , 207 – 274 .
  - Lycan , W. G. (1996). Consciousness and experience . Cambridge, MA : MIT Press
- It's important that not the neural states but the coarse-grained state determine contents of consciousness.

4.1 Measure of consciousness/information correlate of consciousness

Based on our core assumption that NTIC is the core informational entity correlating conscious experience, we propose conscious levels and contents can be derived by computes the properties of a NTIC process.

We first hypothesise that, at given time t, levels of consciousness C_t^{Level} corresponds to the degree of $NTIC_t$ at a coarse-grained level.

$$C_t^{Level} = NTIC_t \tag{8}$$

We second hypothesise that, the content of consciousness corresponds to the state of the NTIC process.

$$C_t^{Content} = Y_t \tag{9}$$

We discuss the implications of the two hypotheses in the following.

4.2 Level of Consciousness correlates degrees of NTIC of a process

Our theory indicates that conscious levels are determined by two key elements based on the definition of NTIC (Equation 7). First, to achieve high level of NTIC, one needs to maximise the channel capacity between the current internal state and the future internal state $I(Y_{t+1}; Y_t)$. This indicates that the richer self-predictive information the NTIC process contained the higher level of consciousness the agent has. Considering two hypothetical systems composed of binary nodes, one has only three nodes and the other one has 1 million notes. Even thought the first system can reach the maximal the channel capacity, the maximal level of NTIC is only 3 bits. Comparing the first system, even though the second system does not have full self-predictability about its future state, the level of NTIC of this system can still far higher than the first system. This outcome naturally explains why people intuitively associate levels of consciousness with complexity of systems. [[I want to say something about animal consciousness here]]. `Need to end this part and may need a reference`

Second, one can also minimise the conditional mutual information $I(Y_{t+1}; Y_t | E_t)$. This suggests that conscious level positively correlates the level of information about the environment dynamic that the NTIC process encoded. Therefor, this implies that an agent who forms an NTIC process to lives and adapt to a more complex environment has a higher conscious level than an agent who interacts an environment with simpler dynamics. This also implies that agents who can explore, adapt model, and react to diverse environments should have higher level of consciousness.

Note that the non-monotonic relationship between levels of NTIC and scale of coarse-graining suggests that there exists a scale of coarse-graining with the maximal level of NTIC. The self-predictive information is the key to achieve high conscious level. If a system dynamics is highly scholastic and unpredictable, the level of consciousness should be low. As a result, information processed at neuronal scales which is heavily noisy and highly scholastic cannot form a high level of consciousness. Similarly, after coarse-graining, macro-variables lose predictive fine-grain information, the predictability become low and also the channel capacity between the current state and future state becomes very small, NTIC would be low as well. Consequently, human consciousness should be maximised at a certain coarse-grained level in the neural system.

Note that, the environment of a NTIC process in the nerual system includes not only the `Werite this`

Finally, which coarse-grained level can achieve high NTIC? agent-scale operation At the physical scale which human being live in, the environmental dynamics is nearly deterministic following the laws of classical physics. It gives agents living at this physical scale great advantages if the agents can build NTIC process internally. Therefore, to coarse-grain microscopic states is necessary
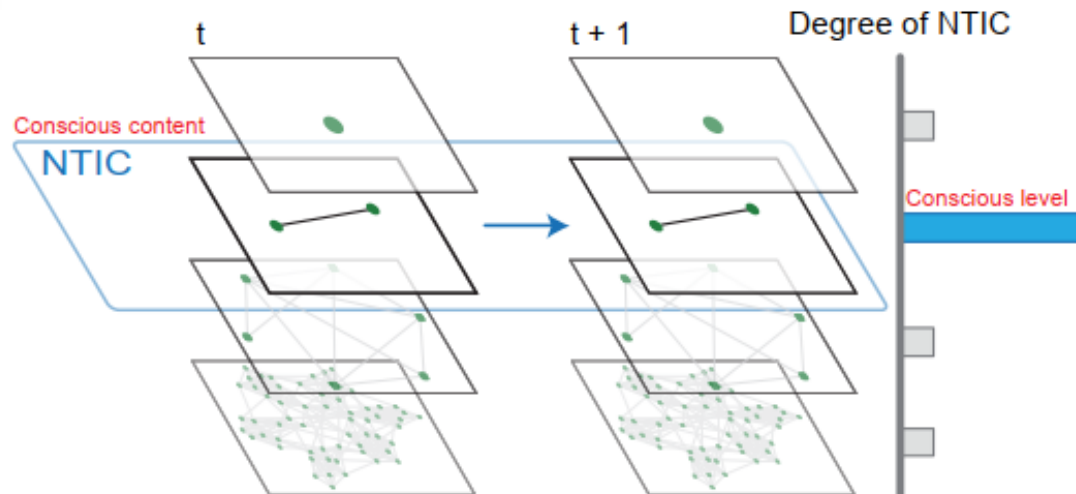
- `modify this figure`

Figure 5: Level of coarse-graining and level of consciousness. The degree of NTIC is not a monotonic function of coarse-graining levels, suggesting that high conscious level may exist at a certain coarse-grained level

4.3 Conscious Contents Corresponding to Coarse-grained States

As we mentioned above, human conscious contents always consist of rich, high dimensional, and integrated information. The state space of the correspondent physical system must have the same capacity of containing the same amount of information. We claim that conscious contents correspond to the state of the NTIC process.

This claim implies that the complexity of conscious contents correlates the amount of information encompassed in the NTIC process. Therefore, a crucial predict from our theory is that the larger state space of the NTIC process has the richer conscious experience the process can have (e.g., a NTIC process involving a large amount of coarse-grained variables). Crucially, even experience of full darkness contains rich information because it rules out other enormous potential alternative states of conscious contents. Equivalently, in the physical system, to know the state of the NTIC process when there is no sensory input (darkness) also requires the same amount of information from the system.

Follow the example in the last section, a small creature lives in a simple environment with very limited environment dynamics. It may forms a simple NTIC process. However, in this situation, the self-predictive information is very low and the state space of the NTIC process is small. As the result, the richness of the conscious experience this agent has will be very limited.

Importantly, because the closure is at a coarse-grained scale one state of the informational closure can be mapped to several states at microscopic scale, our theory supports multiple realisation thesis which suggests same conscious experiences can be mapped to different physical implementations.

Fig: maybe I need a figure here

Do I have this example?
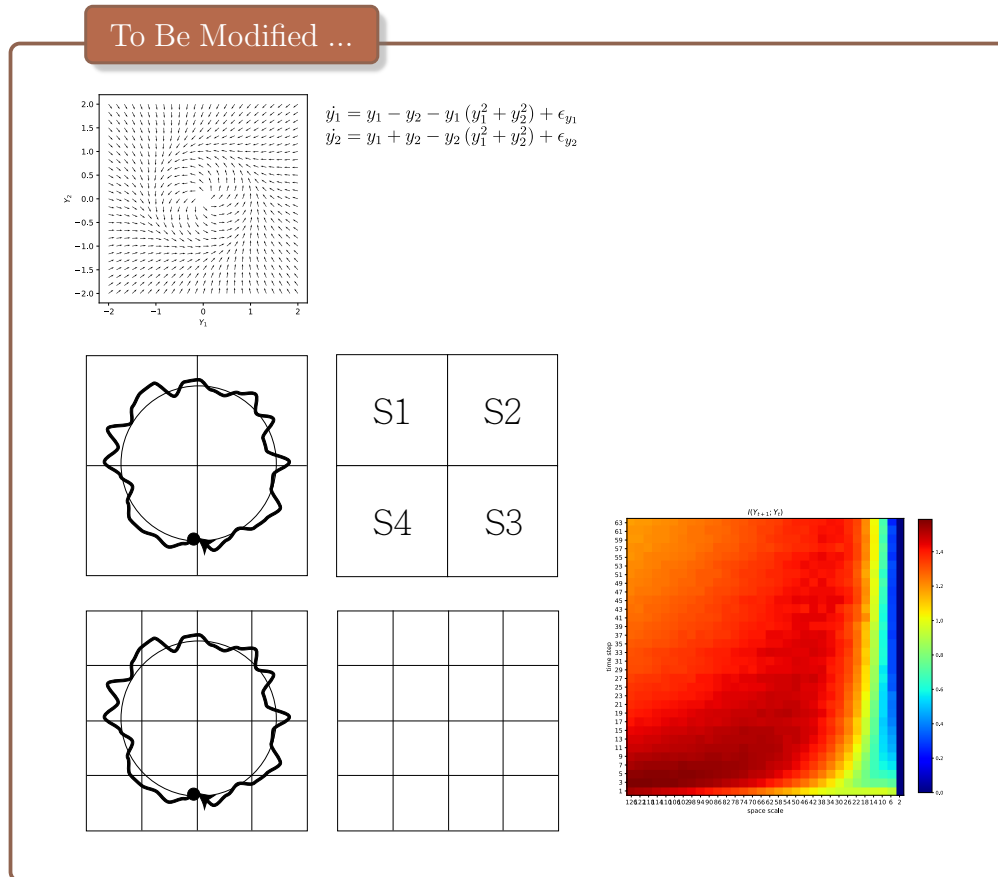
Ref: multiple realisation

16

Figure 6: Example of coarse-graining by a limit cycle process

$$\dot{y_1} = y_1 - y_2 - y_1 \left(y_1^2 + y_2^2 \right) + \epsilon_{y_1}$$
$$\dot{y_2} = y_1 + y_2 - y_2 \left(y_1^2 + y_2^2 \right) + \epsilon_{y_2} \tag{10}$$

The noise sources, ϵ_{y_1} and ϵ_{y_2}, are i.i.d with a normal distribution with mean $\mu = 0$ a and varance $\sigma^2 = 1$

4.4 Reconcile level and content of consciousness

Recently, the segregation of conscious level and conscious contents raise a debate (Bayne *et al.*, 2016; Fazekas & Overgaard, 2016). Our framework reconciles conscious level and conscious content. When the information encapsulate in the NTIC process is large and reaches a high NTIC, based on our definition, the system has high level of consciousness. Simultaneously, this provides a rich state space which is equivalent to encoding rich predictive environmental dynamic in the system. From this perspective, levels and contents of consciousness are just two different properties of NTIC processes. An important prediction from our theory is that in which the conscious level is determined by the size of the state space of the NTIC process rather the transient state of a system. For example, when we experience monotonous environment (a dark room), this does not degrade the conscious level.

Our framework also explains why in normal physiological states level and content are positively correlated (Laureys, 2005). To achieve high NTIC, a system also needs to have high dimensionality and a large state space which guarantees rich and high dimensional information in conscious contents. Therefore, this framework well integrates the two key but debated concepts in consciousness research.

Fig

Writing Materials

- This implies that, based on the Equation 7, the self-predictive information about the future states that the current state holds and the information about the environment encoded in the process determine the level of consciousness.

- This also suggests that the richness of the environment being modelled by the NTIC process has a direct contribution to the level of consciousness.

- To have high level non-trivial informational closure, the states of the coarse-grained scale need to

- First, maximise the mutual information between environment and the representation. This term implies that the agent need to maximise predictive power of environmental state. This also suggest that the environmental complexity is a crucial factor to the level of consciousness. The more rich and complex environment the agent try to predict the higher level of consciousness the agent has.

- Consciousness and complexity Tononi & Edelman (1998)

- Second, to minimise the mutual information between environment and the representation conditional on the past state of the representation. This term suggests that the representation is self-predicable and it mimics the environmental transition.

- Therefore, this theory predicts that level of consciousness determined by the environment that the agent adapts to. To form a high non-trivial informational closure, the agent will need to live in a complex environment and model the environmental transition precisely.

Writing Materials -

iteration 2

4.5   sensory hierarchy and neural coarse-graining

It's important to note that the coarse-graining direction is not necessary to be align with anatomical sensory hierarchy in the neural system. Our theory is different from consciousness theories which focus on the anatomical hierarchy, for example the intermediate Level Theory (Prinz, 2007; Jackendoff, 1987, see also 6.1.2). Due to the pervasive noise in the microscopic levels in the neural system reliable information processes need to be built upon the coarse-grained levels. Nevertheless, to exercise causal power in actuators and environments, to read out information from coarse-grained levels is necessary. . Therefore, the neural system still implement readout decoders for the subsequent signal output. In figure 4.5, the blue arrows indicate the readout mechanisms implemented in the neural system. We argue that previous data suggesting the neural correlates of consciousness between microscopic neural activities and conscious contents may be misled.

> Just saying anatomical is not enough. Need better expression

> Fig: Maybe need a figure discribing brain to hand

$\sim$ indicate biological implimentation
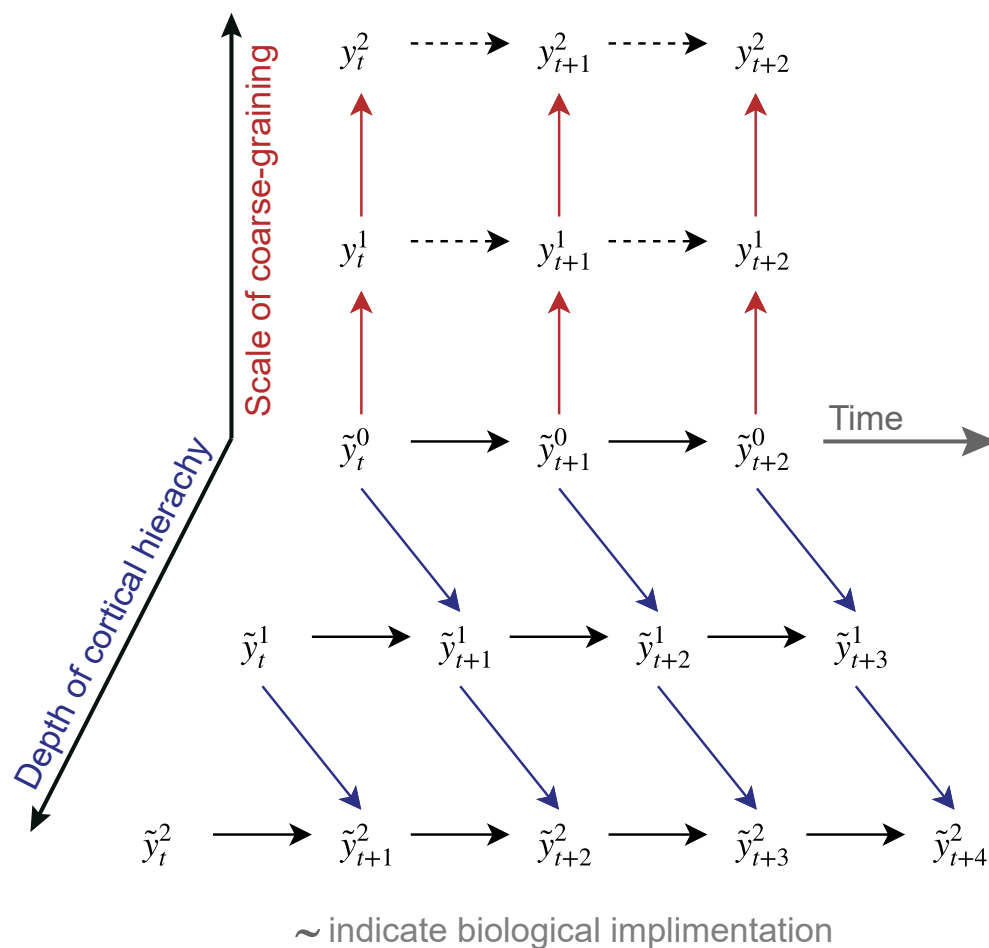
Writing Materials

- Of course, from evolutionary perspective, neural system should develop a better hardware to support neural coarse-graining

- For example, pooling layers in CNN is an implementation of coarse-graining function. It' is very plausible visual processing hierarchy in the neural system realises pooling layers.

- It is possible that the neural system represents higher-level information processing by low level physical subtract. For example, studies have found representation for summery statistics of signal and neural. population.
  From this point of view, sensory hierarchy may extend along with levels of coarse-graining.

Writing Materials -

Ref: neural representation for summery statistics

iteration 2

4.6   Dynamical Boundary

Our theory makes strong predicts that the physical boundary of a conscious system is dynamical. The boundary is determined by elements forming high NTIC. This implies that when the structure of interaction between elements changes the physical boundaries support consciousness should also change. If some coarse-grained variables encode environmental information, they may become necessary components to form high NTIC. In such case, these coarse-grained variables are "recruited" inside the boundary of conscious processes. Note that, due to synergetic information, even recruiting some new variables may largely increase the level of NTIC. This may explain why consciousness is

tightly associated with binding and integration of information in human perception and cognition. Another crucial implication is that the same neural substrates may not be always involved in conscious processing rather it depends on whether it creates high NTIC with other members in the neural system.

5 Conscious versus Unconscious Processing

iteration 3

> -
> The goal of this part is to combine empirical study and phenomenology

> -
> I need to expend this section much more and more references
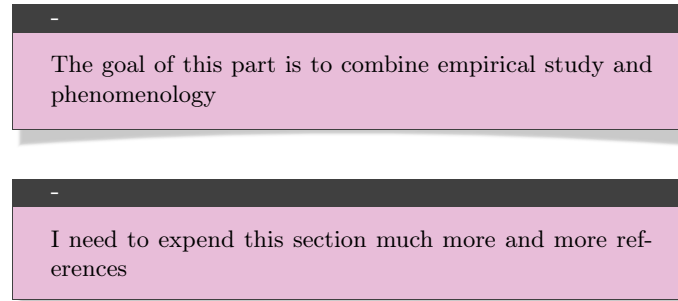
In this section, we suggest that our theory has a strong exploratory and predictive power but with minimal assumptions. Our theory indicates that, to make information conscious, a sufficient and a necessary conditions need to be satisfied.

1. Sufficient condition: Information is encoded in NTIC process.

2. Necessary condition: The system is capable of constructing NTIC processes at coarse-grained levels.

Take the neural system as an example, our theory can well explain previous empirical findings in consciousness science and further make strong and more precise predictions for future studies. In the following, we discussed the implication of the two necessary conditions with respect to the empirical data in the literature of consciousness.

5.1 Unconscious Processing

- As described above, if information processing fails to meet either one of the two necessary conditions, information remains unconscious.

- First, considering processes which fail to meet the first condition. That is, a information process is not NTIC. We can see many processes in the neural network system fall in this category.

- In this category, processes cannot form NTIC. The internal states of the processes are partially or fully determined by environment states, i.e. information is not closed. Reflective behaviours are one of the main subcategory.
  Reflective behaviours usually are considered as automatic and non-conscious behavioural controls. (Casali *et al.*, 2013).

  – In modern reinforcement learning framework, this usually is referred to policy-based agents. The agents constructs a policy function that maps sensory inputs to the action space (Sutton & Barto, 2018).

- Roughly speaking, a policy is a mapping from perceived states of the environment to actions to be taken when in those states. It corresponds to what in psychology would be called a set of stimulus–response rules or associations

- The same principle can be applied to blindsight and procedure memory which are often considered as the classical cases of unconscious processing

- This classic patient study of D.F. who has apperceptive agnosia (Whitwell *et al.*, 2014) also falls in this category. With online sensory inputs, the neural system can create functional processes without forming NTIC.
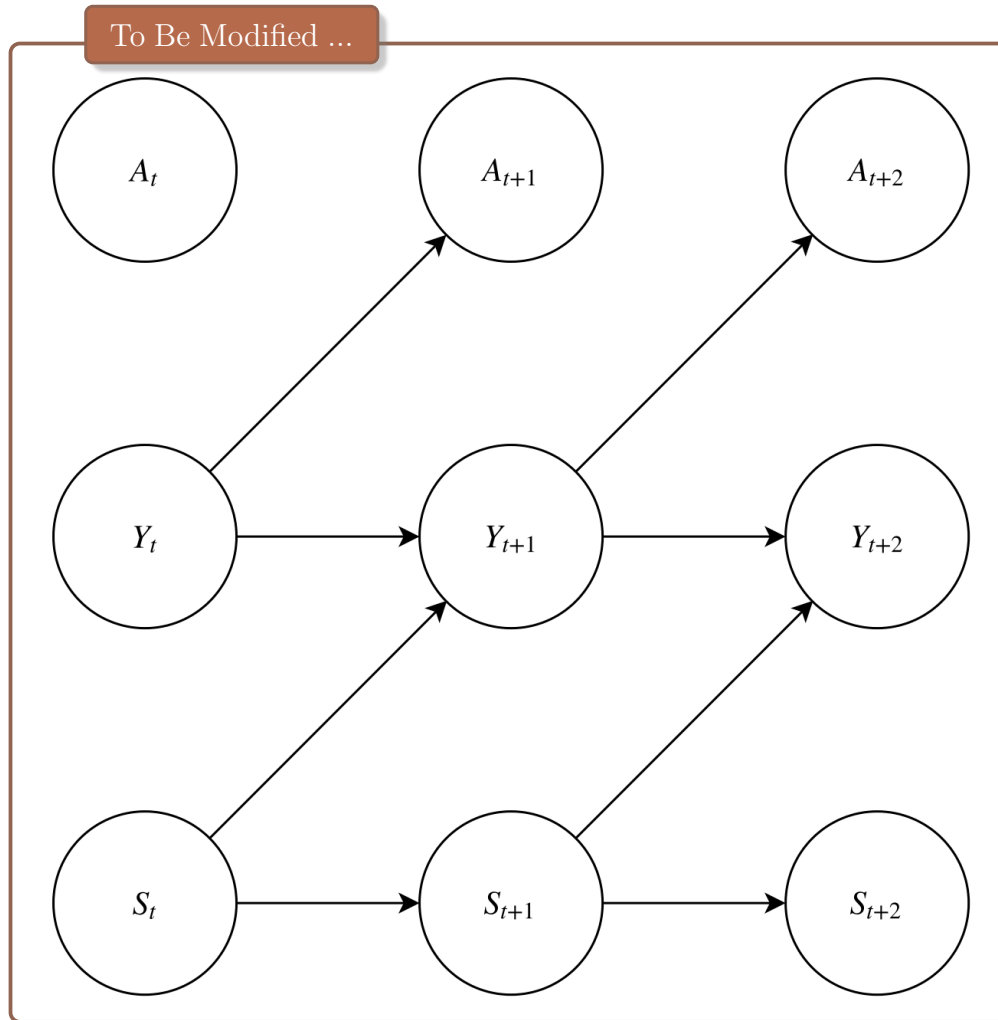
Figure 7: Policy network

- Note that, when we prolong the temporal interval between stimulus and action, usually consciousness is necessary for learning associations. Ref: time
- Another case is that, processes are information closed but it is trivial.
- ===
- Even though the system can form NTIC processes in some certain coarse-grained levels, information processing outside these levels remains unconscious.
- microscopic neural activities fall in this category.
- When processes at macroscopic levels becomes informational closed, information from other levels is redundant.
- Therefore, information from other levels does not influence the dynamic of this level.

——- === -

— procedure memory Reflective reaction, blindsight,

$$\pi(A|S)$$
$$J_t(E \to Y) := I(Y_{t+1}; E_t|Y_t) > 0 \tag{11}$$

[Need Help]: Math here

When the first condition is not satisfied,

In this case, the future state of this process is partially or fully determined by environment states.

We can see

- Conscious
- Boundary
    - binocular rivalry/CFS

21

- Libet's experiment
- Priming
- Bistable perception
- Perceptual overflow
- Unconscious determinants of free decisions in the human brain
- Threshold of unconscious to consciousness
- Probabilistic information is always unconscious

- Unconscious
  - Reflection
    * Reflexive systems is not informational closured.
    * Base d on our theory,
    * x -> s -> a
      · the system state is determined by environmental state, then it is not conscious.
    * Therefore, in general, a simple policy network is not conscious.
  - Blindsight
- Passive neural activity

5.2 reflexive behaviours is unconscious (or environment-state dependent processes)

5.2.1 Reflection

- Reflexive systems is not informational closured.
- Base d on our theory,
- $P(A|S)$ action policy is conditional on the sensory input data.
- x -> s -> a
  - the system state is determined by environmental state, then it is not conscious.
- Therefore, in general, a simple policy network is not conscious.

5.2.2 Blindsight

- This also blindsight.
- Blindsight patients and still make above chance judgement without conscious awareness of visual targets.
- The stimulus-reaction policy network still intact, but it cannot form informational closure and therefore, patients has no conscious awareness of visual targets
- Patient B.F. reactions are stimulus driven.

5.2.3 procedure memory

5.3 Conscious Processing

So far, previous consciousness research has identified a set of perceptual and cognitive functions which not exclusively but frequently associate with consciousness.
- response inhibition
- goal
- planning
- conflict resolution
- error detection
- task-switching
- working memory

- cognitive control
- overriding automatic responses
- self-monitoring
- global availability
- integration
- stabilisation
- information for off-line processing
- flexible routing of information
- integrates behaviour across time
- tasks for long duration
- flexibility and the strategic use of that information for complex operations and decision-making
- flexible and adjust our behavior to different contexts and goals
- complex visuomotor abilities
- conscious thought, reflection, reasoning, and temporally extended sense of self.
- multiple alternative possibilities
- simulate alternative realities
- imagination
- dreaming
- anticipation
- intending
- replaying
- interpreting
- reflecting
- reasoning
- problem solving
- deciding

To achieve high NTIC has evolutionary advantage. When a process achieves NTIC, the current state of the process contains the information about its future state (Eq. (5)). Therefore, the neural system can build forward models to predict the future state of the environment using the information encoded in the NTIC process. Cognitive functions recruiting forward models built upon it are conscious. (E.g., planning and imagination)

It is important to note that we do suggest any functionalism of consciousness. Rather, we believe that information is fundamentally linked to consciousness. The reason why consciousness is accompanied with some specific cognitive function is because NTIC is a powerful property that brings creatures a lot of advantage in the evolution Seth (2009).

Writing Materials

- importantly, evidence suggests that those functions in some conditions do not require the involvement of consciousness. This evidences an important fact that consciousness is not necessay to these functions.
    - Predictive information about the environment.
    - Stability
    - simulation environmental dynamics
- Replaying past events
- *"However, recent evidence also points out interesting dissociations between conscious and unconscious information processing when it comes to the duration, flexibility and the strategic use of that information for complex operations and decision-making"* van Gaal *et al.* (2012)

Writing Materials -

5.4 Unconscious Processing

When processes fail to satisfy either of the two conditions, they remain unconscious. First, when processes do not meet the first condition, i.e. the information processed not at the certain coarse-grained level, the information will not be included in conscious contents. For example, we are never conscious of neural spike activities at microscopic levels (e.g. single cell activities) because information passing among individual neurons are not at the certain coarse-grained level. Libet's free will experiment can be also explained by our theory. The random neural fluctuation under Libet's free will experiment happens at the microscopic scale (Schurger *et al.*, 2012).

Second, when processes do not meet the second condition, i.e. processes which do not achieve high NTIC will remain unconscious. This can explain why many "policy-based" processes are not conscious. Policy-based here features those processes taking sensory information as input and directly outputting actions. Therefore, the future states of policy-based processes largely depend on the environmental states and failed to achieve informational closure. Neural processes are included in this category, e.g. reflexive neural circuit, blindsight, procedure memory.

To summarise, our theory can well explain the various findings with parsimonious assumptions.

- *"the nonsemantic processing of primes is enhanced and priming effects are boosted when the experimental context allows the formation of automatic stimulus–response mappings."*
  Van den Bussche *et al.* (2009)

6 Comparison with other theories

> -
>
> Main points should be same/diff, our advantages, and predictions

In this section, we will discuss the relationship between our theory and other popular modern theories of consciousness.

iteration 2

6.1 Theories about multilevel views on consciousness and cognition

Previously, multilevel views on consciousness and cognition have been proposed. We shortly describe two theories which is related to our current proposal.

6.1.1 level of Organisation Wimsatt (1994)

As mentioned previously, the level of coarse-graining is different from the anatomical level. Pennartz focus more on multimodal representation as the most important factor (See the discussion in Pennartz (2015)). Our proposal is inline with Pennartz's Neurorepresentational theory (Neurorepresentationalism) Pennartz (2018, 2015). However, it's crucial to note that our theory claims only the level of coarse-graining achieving NTIC matters to consciousness rather than the anatomical level of sensory hierarchy.

> -
>
> Should I also deal with Marr's level of abstraction?

6.1.2 Intermediate Level Theory

One important theory linking consciousness to multilevel views is the Intermediate Level Theory (ILT). ILT proposes that conscious experience only correlates specific levels of representations in the sensory processing the hierarchy. This theory claims that consciousness only associates with the representation at intermediate levels rather than the ones at lower-level or higher-level discribed as 2.5D. However, ILT focuses on the levels in terms of anatomical direction. As mentioned previously, a critical difference between ILT and our theory is that our theory doesn't assume the direction

of coarse-graining should align with the direction of anatomical processing levels along cortical hierarchies (see fig 4.5).

Different from ILT which suggests that the neural substrates responsible for the 2.5D computation is conscious, we predict that if information that is carried by v1 neurons are necessary to complete informational closure, the states of V1 neurons should also co-vary with conscious contents.

6.1.3 Pennartz's neurorepresentational theory (Neurorepresentationalism)

Our theory shares some similar concepts with Pennartz's Neurorepresentationalism. However, Neurorepresentationalism focuses on multimodal representation Pennartz (2018, 2015). Another focus of neurorepresentationalism is on "making sense" between representations. Different from neurorepresentationalism, conscious contents in our framework are grounded by the informational relations between variables in NTIC because NTIC provides a closed relationship between states in a system.

Similar to Neurorepresentationalism, our framwork is more close to Marr ' s original distinction between the levels of (neural) implementation, algorithms, and computational theory [122], but focuses on how conscious representations correspond to computational operations at levels with lower complexity and representational capacity.Pennartz (2018, 2015)

Neurorepresentationalism is also a functionalist perspective. However, e don't assume any functional properties as the criteria of consciousness. Instead, we propose information as a more fundamental principle correlates consciousness.

Need something here

Writing Materials

- 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. "*
  ?

- *" 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)

25

- *" In 1987, Ray Jackendoff published Consciousness and the Computational Mind. In it, he posed an important Where question: Where, in the fl ow 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 decompositional analysis, we can ask: in which subsystems does consciousness arise? If we depict the mind as a vast fl ow 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? Th is 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. Th e high level is a good deal better for action, but also suboptimal. Th e 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 specifi c. Th ese representations are useful for determining what to do here and now.arg1"*
    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 disunifi ed. Th ese notions are all in need of some refi nement, but, as a fi rst approximation, the idea is that intermediate-level representations are neither too specifi c nor too general."*
    Prinz (2007)

- However, as we mentioned in Section 4.5, *"The notion of ' levels ' deployed here is functional and very different from that of successive anatomical processing stages along cortical hierarchies [75] or from anatomic scales in the brain. The multi-level concept in [2] is more akin to Marr ' s original distinction between the levels of (neural) implementation, algorithms, and computational theory [122], but focuses on how conscious representations correspond to computational operations at levels with lower complexity and representational capacity."*
  Pennartz (2018, 2015)

iteration 2

6.2 First-order theories

The first-order theories suggest that consciousness is determined by early sensory information or neural activities. The first-orde view maintains that conscious awareness is determined by early sensory activity alone, independently of higher-order representations. *"According to first-order views, phenomenal consciousness consists in analog or fine-grained contents that are available to the first-order processes that guide thought and action. So a phenomenally-conscious percept of red, for example, consists in a state with the analog content red which is tokened in such a way as to feed into thoughts about red, or into actions that are in one way or another guided by redness."*

We predict that, if early sensory information or neural activities after coarse-graining take parts in the NTIC process, early sensory information or neural activities also contribute consciousness. However, due to coarse-graining, changes of early sensory information or neural activities do not necessarily accompany with changes of conscious contents.

Writing Materials

- *" 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 prefrontal and parietal activity. Therefore, in the group of first-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 first-order representations, as well as scientists who hold that visual awareness depends on activity in content-specific regions of the extrastriate cortex [24,25] or on feedback from these regions to the primary visual cortex [26]. Most first-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 first-order theorists hold that some higher cognitive functions can be performed without conscious awareness [30–32], most first-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.."*
  Lau & Rosenthal (2011)

- Among those are first-order theories, on which a mental state is conscious if being in that state results in one's being conscious of something;

- According to first-order views, phenomenal consciousness consists in analog or fine-grained contents that are available to the first-order processes that guide thought and action. So a phenomenally-conscious percept of red, for example, consists in a state with the analog content red which is tokened in such a way as to feed into thoughts about red, or into actions that are in one way or another guided by redness.

Writing Materials -

---

iteration 1

6.3 Higher Order Thought theory of consciousness

In contrast to first-order theories, higher-order thought theories suggest that consciousness depends on higher-order representations representing agents as being in particular internal states. Meta-cognition is also associated with many conscious cognitive function. However, we propose a new perspective and indicate that HOT may be a misunderstanding . This may be the result of neural coarse-graining. Because contents in informational closure look like watching fine-grained information, people may have an wrong impression about it. Coarse-grained information can be summary statistics which may be misunderstood as higher-order representation. It is true that the neural

27

system does create meta-representation to represent xxx. The higher-order representation may resemble coarse-grained variables. This may cause misidentify the entity which correlates conscious experience.

Moreover, the predictive information in NTIC is critical for other systems to build forward models for a wild range of tasks. Therefore, most of the high level cognition can utilise the state of informational closure. For example, mental planning needs a precise forward model to infer the environmental dynamics.

-
Writing Materials

- In contrast to first-order theories, higher-order thought theories suggest that consciousness depends on higher-order representations representing agents as being in particular internal states.

- Coarse-grained information can be summary statistics which may be misunderstood as higher-order representation.

- It is true that the neural system does create meta-representation to represent xxx.

- The higher-order representation may resemble coarse-grained variables.

- This may cause misidentify the entity which correlates conscious experience.

- —————————————————————————-

- Information closure directly is critical for other systems to construct forward models.

- Therefore, most of the high level cognition can utilise the state of informational 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 first-order rather than higher-order representations. Because conscious awareness can differ even if all first-order representations remain completely unchanged, such awareness itself might serve little function [1,27]."*
  Lau & Rosenthal (2011)

- *" 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)

Writing Materials -

6.4 Global workspace theory

Global workspace theory (GWT)Baars (1988, 1997, 2002) or Global Neuronal Workspace theory (GNWT)Dehaene *et al.* (1998); Dehaene & Naccache (2001); Dehaene & Changeux (2011) states that the neural system consists of many separate specialised modules and a central global workspace (GW) which integrates and passes information amount the modules. Only the information that is accessed by the global workspace can be conscious aware of. The modules competes with each other to gain the access to the global workspace and the information from the winner can trigger an all-or-none "ignition" in the global workspace and be globally broadcasted to other modules. Conscious contents then associates information which gains access to the internal global workspace Dehaene *et al.* (2017a). Our framework can well accommodate and interpret these claims by GWT.

Stability: Conscious perception necessitates representational stability. Our theory rests upon self-predictability which formally encompasses stability.

Broadcasting: When new sensory information from a modality is involved in NTIC, the information can be preserved in the following dynamic of NTIC. The neural system can extract the information for specific task goals. This is equivalent to the function of broadcasting.

Non-linear ignition: Because mapping microscopic to macroscopic states is many-to-one, it commonly shows a nonlinear relationship between microstates and macrostates. In our theory, new information presents in conscious perception corresponds to a transition in the state space of the NTIC process. This can cause a significant change of the brain states and resembles non-linear ignition of brain activities. Conscious access as an accumulation of evidence leading to an all-or-none ignition What, if anything, remains unique to conscious processing? Although many cognitive operations can be partially launched non-consciously, these operations rarely if ever run to completion in the absence of consciousness. A subliminal stimulus may induce above-chance performance, behavioural priming, and a small amount of brain activity in narrowly defined brain networks, but these measures often increase dramatically as soon as the subject reports seeing the stimulus, especially in high-level areas [46 60,61]. Accumulation of evidence has been demonstrated with non-conscious stimuli [59], but only conscious stimuli cross the threshold beyond which an overt strategy can be fiexibly deployed.

A difference between our theory and GWT is that our theory does not suggest a neuronal core anatomical area for consciousness. . Coarse-grained information in NTIC is more crucial for consciousness rather than neural substrates supporting the global workspace.

Cite similar opinion from other theories

Link to Section 4

Writing Materials

- Global workspace theory (GWT)Baars (1988, 1997, 2002) or Global Neuronal Workspace theory (GNWT)Dehaene *et al.* (1998); Dehaene & Naccache (2001); Dehaene & Changeux (2011) states that the neural system consists of many separate specialised modules and a central global workspace (GW) which integrates and passes information amount the modules. Only the information that is accessed by the global workspace can be conscious aware of. The modules competes with each other to gain the access to the global workspace and the information from the winner can trigger an all-or-none "ignition" in the global workspace and be globally broadcasted to other modules. Conscious contents then associates information which gains access to the internal global workspace Dehaene *et al.* (2017a).

- Submodule access and NTIC

- Our theory can well accommodate several claims by GWT.

- Conscious perception necessitates representational stability: Our theory rests upon self-predictability which formally encompasses stability.

- Broadcasting: When new sensory information from a modality is involved in NTIC, the information can be preserved in the following dynamic of NTIC. The neural system can extract the information for specific task goals. This is equivalent to the function of broadcasting.

- Non-linear ignition: Because mapping microscopic to macroscopic states is many-to-one, it commonly shows a nonlinear relationship between microstates and macrostates. In our theory, new information presents in conscious perception corresponds to a transition in the state space

of the NTIC process. This can cause a significant change of the brain states and resembles non-linear ignition of brain activities.

- our theory re-explains the critical xxx from GWT.

- The critical difference is that our theory does not suggest a neuronal core anatomical area for consciousness. We suggest coarse-grained information form NTIC is crucial for consciousness.

- In sum, our theory is compatible with previous physiological evidence supporting GWT.

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

- e 'all-or-none' behavior, whereby the incoming evidence either quickly dies out (corresponding to subliminal processing) or is accumulated and amplified non-linearly into a full-blown state of high-level activity. This global 'ignition' has been proposed as a marker of conscious perception [70]. Indeed, empirically, when stimulus strength is varied, the early stages of non-conscious processing typically show a linear variation in activation, whereas conscious access is often characterized by a late non-linear amplification of activation which invades a distributed set of parietal, prefrontal and cingulate areas

6.4.1 Ignition

- Conscious access as an accumulation of evidence leading to an all-or-none ignition What, if anything, remains unique to conscious processing? Although many cognitive operations can be partially launched non-consciously, these operations rarely if ever run to completion in the absence of consciousness. A subliminal stimulus may induce above-chance performance, behavioural priming, and a small amount of brain activity in narrowly defined brain networks, but these measures often increase dramatically as soon as the subject reports seeing the stimulus, especially in high-level areas [46 60,61]. Accumulation of evidence has been demonstrated with non-conscious stimuli [59], but only conscious stimuli cross the threshold beyond which an overt strategy can be flexibly deployed.

- 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.* (2017a)

- C1: Global availability of relevant information The need for integration and coordination
  Dehaene *et al.* (2017a)

- 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*

Writing Materials -

6.5 Sensorimotor contingency

Sensorimotor contingency (SMC) theory of consciousness is proposed to account of conscious perception. It suggests that different types of SMCs give rise to different characteristics of conscious perception O'Regan & Noë (2001). The theory radically rejects the view that conscious contents is associated with information encoded in internal representations in the neural system. Rather, the quality of conscious perception depends on agents' mastery of SMCs. It emphasises the interactions between a system (the neural system) and its environment are crucial to form conscious perception

- Sensorimotor contingency (SMC) theory of consciousness is proposed to account of conscious perception. It suggests that different types of SMCs give rise to different characteristics of conscious perception O'Regan & Noë (2001).

- The theory radically rejects the view that conscious contents is associated with information encoded in internal representations in the neural system.

- Rather, the quality of conscious perception depends on agents' mastery of SMCs.

- It emphasizes the importance of the interactions between a system (the neural system) and its.

- Counter-intuitively, because consciousness inhabits in interactions between agents and environments, this theory suggests that consciousness is not driven by brain activities. Instead, consciousness should be shaped by the way of interaction, i.e. SMCs, between agents and environments.

- Similar to the SMC theory, our theory predicts that the interaction between the neural system and the environment shapes the quality of conscious perception.

- This is because to achieve NTIC the neural system should maximise the conditional mutual information between the future state of the NTIC process and the environment given the agent's current action (see equation XXX). I NEED A EQUATION!

- This suggests the to maximise NTIC the NTIC process needs to involve the information about the environmental transition and dynamic after agent preforms an action.

- Therefore, as a consequence of maximising NTIC, the representation forming NTIC implicitly encodes information about SMCs.

- In the long-term, once NTIC neural dynamic interoperates the environmental state transition given actions via maximising NTIC, the state space of XXX then is shaped by the new learned SMC

- Accordingly, the way how the agent interacts with the environment influences the structure of the NTIC representation.

-

- [[Do I need a figure here?]]

- However, in contrast to the SMC theory, our theory suggests the state of NTIC process correlate conscious contents. Therefore, if the NTIC process is formed inside the neural system, we should have correspondent conscious contents in our conscious experience without interacting with the environment.

- [[Not sure here]] This can easily explain that the conscious experience of dream is just induced by self-evolved state changes of the NTIC process.

- Further more the SMC and our theories make complete different prediction about the relationship between SMC and consciousness.

- Contrary to the SMC theory which suggests that consciousness resides on SMCs, our theory predicts that pure S-R contingency between the neural system and the environment can not achieve informational closure within the neural system and therefore is not conscious (please see section 0.2)

<div style="float:right; border:1px solid; background:orange;">Need to involve some sentence from</div>

Writing Materials

- O'Regan and Noë

- *"According to Noë, perceptual presence is explained by a " sensorimotor theory " on which perception depends on a practical mastery of sensorimotor dependencies or " sensorimotor contingencies " (SMCs) (O ' Regan & Noë, 2001). The theory inherits from Gibsonian notions of " affordances " (Gibson, 1979) and from enactive cognitive science (Thompson & Varela, 2001) which stress the importance of brain-body-world interactions in cognitive processes. On sensorimotor theory, the perception of a tomato as a (perceptually present, real, subjectively veridical) tomato is given by practical mastery of the SMCs governing how the sensory responses elicited by the tomato will behave in a variety of situations. A strong point of this theory is that it suggests why there are differences in qualitative character between modalities, the reason being that different modalities instantiate different SMCs (O ' Regan & Noë, 2001).*
"

- In their seminal article O'Regan and Noë (2001) introduce sensorimotor contingencies (SMCs) as the basis of a sensorimotor account of conscious perception.

- In short, different types of SMCs give rise to different qualities of sensory experience

- Second, vision, audition, touch, etc., correspond to domains of knowledge of the respective SMCs that are exercised by an agent as part of its habitual behavior.This "mastery of SMCs" is what lets an agent perceive its environment and adapt its behavior accordingly

- And third, sensory awareness arises when SMCs are integrated with planning, reasoning, and generation of behavior

- The theory is a radical departure from the classical view that the brain constructs an internal representation of the outside world on which higher cognitive processes such as memorizing, reasoning, and planning operat

- perception depends on a practical mastery of sensorimotor dependencies or "sensorimotor contingencies" (SMCs)

- s the importance of brain-body-world interactions in cognitive processes. On sensorimotor theory, the perception of a tomato as a (perceptually present, real, subjectively veridical) tomato is given by practical mastery of the SMCs governing how the sensory responses elicited by the tomato will behave in a variety of situations.

- ways of acting or as something we do, rather than an internal phenomenal experience or an internal representation of the world (O'Regan & Noë, 2001)

- *"I advance this truly astonishing hypothesis: To understand consciousness in humans and animals, we must look not inward, into the recesses of our insides; rather, we need to look to the ways in which each of us, as a whole animal, carries on the processes of living in and with and in response to the world around us . . . You are not your brain . . . Meaningful thought arises only for the whole animal dynamically engaged with its environment . . . And indeed the same is true for the quality of our conscious episodes . . . The taste of licorice is not something that happens in our brains."*
(Noë, 2009, pp. 7–8)

- Because consciousness resides in our behavioural interactions with the world rather than in our brain, the theory postulates that consciousness does not derive from brain activity at all. Consequently, there is no need to explain how brain activity causes or constitutes consciousness, because it does not.

- O ' Regan, J.K. and Noë, A. (2001) A sensorimotor account of vision and visual consciousness. Behav. Brain Sci. 24, 939 – 973

- SMC theory argues that consciousness arises from the execution of SMCs, resulting in brain – environmental interactions.

- The second variant of SMC theory postulates that ' it is not movement as such, but the consequences of potential movement that play a role ' (in perceptual experience) ([5], p. 1015; see also [9,10,19,20]). This ' non-acute ' or dispositional variant marks a signi fi cant shift in position because it de-emphasizes actual movement. Pennartz (2017)

- It stipulates that for us to be conscious of something in the environment (or of our way of interacting with the environment)

Writing Materials -

6.6 Predictive Processing

- The predictive processing is a powerful framework (PP) which integrates several influential ideas in the history of neuroscience including "perception as unconscious inference" from Helmholtz Helmholtz (1866), "analysis by synthesis", "predictive coding", "Bayesian brain hypothesis"

- is a powerful framework

- a recent emerged theoretical framework of perception and brain function, which suggests that the brain continuously generates and updates predictions about incoming sensory signals, represents a general principle of neural functioning.

- According to predictive processing (PP) account of perception, the neural system constantly performs unconscious statistical inference about hidden causes in the external environment. The perceptual contents are the "best guess" about the environment Clark (2013) Hohwy (2013)

- Bperceptual content is the result of probabilistic inference on the hidden causes of sensory signals.

- Because it is well integrated with Bayesian brain hypothesis, the PP framework has been used to interpret conscious perception in many perceptual domain. Hohwy (2013) Seth (2014)

- However, to link PP to consciousness, there are still two major gaps.

   - First, previous studies have demonstrated many brain processes following the principle of Bayesian inference and Bayesian integration. But not all these processes are conscious. This indicates that PP is not a sufficient mechanism for conscious perception. Some other factors may be involved to determine which PPs can be conscious or not.

   - Second, it has been emphasized that only the results of the inference or prediction can be consciously accessed. The process of inference in completely unconscious. So WHY?

- Even though PP has been acquiring attention in different discipline in neuroscience, So far, to our knowledge, no detailed explanation has been given to account for the two questions.

- We first hypothesize that predictive processing stays unconscious is because it is not able to achieve NTIC.

  - [[Example from PID controller?]]

- Second, according to our theory, it is obvious that the computation of predictive processing is unconscious and only the result is conscious. Because the computational details is not at the coarse-grained level which achieves NTIC.

  - So far, many neuronal models and coding schema have been proposed to explain how to implement Bayesian inference and prediction at the neuronal level.

  - However?

- We hypothesize that the distinction between conscious and unconscious predictive mechanisms is whether they build up on the information from the NTIC process.

Writing Materials

- Lamme (2015)

  - To my knowledge, this is the first time this has been so explicitly laid out—writers on predictive coding thus far have always stayed a little vague on where exactly consciousness sits in the Bayesian framework.

  - Yet the two points seem contradictory. In the Bayesian predictive coding framework, consciousness is the result of the unconscious inferential processes.

  - However, these authors have so far been very reticent about how exactly consciousness fits into the PC framework. For now, there is an emerging idea that understands consciousness as the final result of unconsciousness processing. According to Hohwy (2013), perceptual consciousness fits in the PC theory as the 'upshot' or 'conclusion' of unconscious perceptual inferences. Melloni (2014) also expects that inferential processes are conducted behind the curtain of consciousness and what we experience are the 'outcomes' or 'results' of an unconsciou inferential process. Lamme (2015) agrees with the idea that consciousness is based on the relation or more precisely conjunction of current inputs and priors, which together produce posterior beliefs. Karl Friston agrees to a certain extent with this interpretation of consciousness. According to Hobson and Friston (2016), consciousness is not a phenomenon but a process, a process of optimizing beliefs through inference (consciousness, dreams, and inference; the Cartesian theater revisited; a response to our theater critics).

- Link PP and c

  - Hobson & Friston (2016)

  - Melloni (2015)

- Seth (2014)

- When uncertainty is high, we should expect an increase in the ability of ever higher-level prior beliefs to determine conscious experience. I

- determine conscious experience.

- PP combines and builds upon previous ideas about the role of "unconscious inference" in perception (Helmholtz, 1925; Barlow, 1961; Gregory, 1970), the process of "analysis by synthesis" in psychology (Neisser, 1967), the "predictive coding" approach in neuroscience (Srinivasan et al., 1982; Rao and Ballard, 1999; Huang and Rao, 2011) and "generative models" and related probabilistic computational principles (MacKay, 1956; Mumford, 1992; Dayan et al., 1995; Hinton, 2007a,b; Tenenbaum et al., 2011).

- *"The brain is constantly confronted with a wealth of sensory information that must be processed efficiently to facilitate appropriate reactions. One way of optimizing this processing effort is to predict incoming sensory information based on previous experience so that expected information is processed efficiently and resources can be allocated to novel or surprising information. Theoretical and computational studies led to the formulation of the predictive coding framework (Friston 2005, Hawkins and Blakeslee 2004, Mumford 1992, Rao and Ballard 1999). Predictive coding states that the brain continually generates models of the world based on context and information from memory to predict sensory input. In terms of brain processing, a predictive model is created in higher cortical areas and communicated through feedback connections to*

lower sensory areas. In contrast, feedforward connections process and project an error signal, i.e. the mismatch between the predicted information and the actual sensory input (Rao & Ballard, 1999). The predictive model is constantly updated according to this error signal. "

- *"Predictive inference and perception The concept of PC overturns classical notions of perception as a largely 'bottom-up' process of evidence accumulation or feature detection, proposing instead that perceptual content is specified by top-down predictive signals that emerge from hierarchically organized generative models of the causes of sensory signals. According to PC, the brain is continuously attempting to minimize the discrepancy or 'prediction error' between its inputs and its emerging models of the causes of these inputs via neural computations approximating Bayesian inference (Figure 1 and Box 1). Importantly, prediction errors can be minimized either by updating generative models (perceptual inference and learning; changing the model to fit the world) or by performing actions to bring about sensory states in line with predictions (active inference; changing the world to fit the model)."*

- *"In terms of brain function, Hermann von Helmholtz's 19th Century notion of perception as unconscious inference has found new momentum in modern formulations of 'predictive processing', the Bayesian brain, and the 'free energy principle' (Clark, 2013; Friston, 2010; Hohwy, 2013). On these formulations, which can all be subsumed under the general term 'prediction error minimization' (PEM, (Hohwy, 2013)) perceptual content is the result of probabilistic inference on the hidden causes of sensory signals. "*

- *"Generative models encoded by brain structure and dynamics make predictions about sensory inputs, based on prior expectations. Action (e.g., body movements) and perception conspire to reduce sensory prediction errors, giving rise to perceptual content and"*

- *"More informally, what we perceive is the brain's 'best guess' of what's out there (or 'in here', for perceptions of body and self)."*

- *""predictive processing " (PP) 1 or the " Bayesian brain " this holds that perceptual content is determined by hierarchically organized generative (i.e., predictive) models (HGMs) of the external causes of sensory signals, induced by a process approximating Bayesian inference (Friston, 2009)."*

- Not all predictive mechanism are conscious.
- We hypothesize that the distinction between conscious and unconscious predictive mechanisms is whether they build up on the information from the NTIC process.
- we can also explain why only the result (posterior) of predictive processing is conscious.
- Because of this XXXX, the representation has the maximal non-trivial informational closure (ntic) also has the maximal predictive power of the external environment.

  > I need math here!!

- Therefore, it's reasonably 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 build on the representation that forms the informational closure.
- Therefore, the state of predictive model can be conscious while it is linked to informational closure representation.
- Note that, a system showing some predictive power is not sufficient to be conscious.
- For example, the proportional–integral–derivative 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 informational 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 informational 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.

6.6.1 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 informational closure, the predictive result is conscious.
- Because the forward model build on the informational closure representation can be seen as part of the closure.

IS THIS TRUE? NEED DISCUSSION

- 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 the 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 informational closured and therefore is not conscious.
    - 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-closured.

Ref: active inference

iteration 1

6.7 Integrated information theory

- One common principle between our theory and IIT is that instead of NCC, both theories suggest information is the fundamental entity which correlates consciousness.

Ref: cite David Chalmers

- Our theory is different from IIT which proposes that consciousness is information integration in a system.
- For level of consciousness, IIT predicts a system with higher level of integration produces a higher level of consciousness. In our theory, the level of consciousness corresponds to the level of NTIC, which suggests that level of consciousness is determined by self-predictive information and the information about the environment encoded in the process.
- In terms of conscious content, IIT suggests that the causal structure of a system determines the quality of conscious contents. In the current version of our theory, how the state of NTIC process maps to conscious content has not been sophisticatedly specified. More work is needed to understand the relationship between the state of NTIC process and conscious contents.
- ————————————————————————————
- [[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.) we 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 informational closure representation.

- Similarly, it is not *"consciousness is necessary for multisensory integration"* . It should be multisensory integration is more likely to create an informational 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 informational closure representation which is conscious.

6.7.1 Hoel's theory

This is very important. Need to be careful here.

- Our theory is closely related to causal emergence described by Hoel et al. (2013) which was examined by a following work using integrated information theory (Hoel et al. 2016). More research on the precise mathematical relationship between causal emergence and our theory is needed

- Hoel *et al.* (2016)

- causal emergence

- Hoel *et al.* (2013) examined causal power at different coarse-grained scale and show that causal power can be stronger at macro rather than micro levels.

- To answer the question about what is the best scale to compute Phi, Hoel *et al.* (2016) examined causal power at different coarse-grained scale.

- Further instigation is needed to elucidate the relationship between our theory and causal emergence

6.8 Internal simulation and self-modelling theory

7 Limitation

7.1 Finding potential coarse-graining functions

To empirically verify the current theory, it is important to know what coarse-graining functions which maps microscopic variables to macroscopic variables in the neural system. So far, in this article, we have not speculated potential candidates of the coarse-graining function (or functions) that creates NTIC in the neural system.

Gamez (2016) has systematically describe a similar issue in terms of finding data correlates of consciousness amount different level of abstraction.

- Our theory does not intent to solve the Hard problems.

- Inverted spectrum Inverted Qualia (Shoemaker, 1982; Block, 1990; Locke & Nidditch, 1979)

- Considering that a system has only two possible physical states X and Y mapping to three different qualia (we can label them as "*red*" and "*blue*" respectively). A system cannot rest on two states at any moment, therefore, knowing the system state at X excludes the the other possible state Y and receives 1 bit information. Similarly, if conscious percept is "*red*", it rules out the "*blue*". Therefore, information provides by the conscious percept is also 1 bit.

- However, our theory cannot tell why X and Y map to *red* and *blue* ,respectively, rather than the other way around.

-

- Our theory does not explain the mapping

8 Future works

- Lack of direct evidence because no one focus on informational closure.

- Physiological location of NTIC

- Practical measure of NTIC.

8.1 Finding coarse-grain function

- It's not clear what the coarse-graining function is.

- *" 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.

[margin note: this should be explained by environmental deterministic]

9 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

Adrian, Edgar D. 1926. The impulses produced by sensory nerve endings. *The Journal of physiology*, **61**(1), 49–72.

Asplund, Christopher L, Fougnie, 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.

Baars, Bernard J. 1988. *A cognitive theory of consciousness*. Cambridge University Press.

Baars, Bernard J. 1997. In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, **4**(4), 292–309.

Baars, Bernard J. 2002. The conscious access hypothesis: origins and recent evidence. *Trends in cognitive sciences*, **6**(1), 47–52.

Banks, William P. 2009. *Encyclopedia of consciousness*. Vol. 1. Academic Press.

Bayne, Tim, Hohwy, Jakob, & Owen, Adrian M. 2016. Are there levels of consciousness? *Trends in cognitive sciences*, **20**(6), 405–413.

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.

Block, Ned. 1990. Inverted Earth. *Philosophical Perspectives*, **4**(n/a), 53–79.

Casali, Adenauer G, Gosseries, Olivia, Rosanova, Mario, Boly, Mélanie, Sarasso, Simone, Casali, Karina R, Casarotto, Silvia, Bruno, Marie-Aurélie, Laureys, Steven, Tononi, Giulio, *et al.* 2013. A theoretically based index of consciousness independent of sensory processing and behavior. *Science translational medicine*, **5**(198), 198ra105–198ra105.

Chalmers, David J. 1996. *The conscious mind: In search of a fundamental theory*. Oxford university press.

Clark, Andy. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, **36**(3), 181–204.

Dehaene, Stanislas, & Changeux, Jean-Pierre. 2011. Experimental and theoretical approaches to conscious processing. *Neuron*, **70**(2), 200–227.

Dehaene, Stanislas, & Naccache, Lionel. 2001. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, **79**(1-2), 1–37.

Dehaene, Stanislas, Kerszberg, Michel, & Changeux, Jean-Pierre. 1998. A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the national Academy of Sciences*, **95**(24), 14529–14534.

Dehaene, Stanislas, Lau, Hakwan, & Kouider, Sid. 2017a. What is consciousness, and could machines have it?

Dehaene, Stanislas, Lau, Hakwan, & Kouider, Sid. 2017b. What is consciousness, and could machines have it? *Science*, **358**(6362), 486–492.

Eurich, Christian W, & Wilke, Stefan D. 2000. Multidimensional encoding strategy of spiking neurons. *Neural Computation*, **12**(7), 1519–1529.

Faisal, A Aldo, Selen, Luc PJ, & Wolpert, Daniel M. 2008. Noise in the nervous system. *Nature reviews neuroscience*, **9**(4), 292.

Fazekas, Peter, & Overgaard, Morten. 2016. Multidimensional Models of Degrees and Levels of Consciousness. *Trends in Cognitive Sciences*, **20**(10), 715–716.

Gamez, David. 2011. Information and consciousness. *Etica & Politica/Ethics & Politics*, **13**(2), 215–234.

Gamez, David. 2016. Are Information or Data Patterns Correlated with Consciousness? *Topoi*, **35**(1), 225–239.

Gerstner, Wulfram, & Kistler, Werner M. 2002. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press.

Goff, Phillip. 2006. Experiences don't sum. *Journal of Consciousness Studies*, **13**(10-11), 53–61.

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*.

Helmholtz, H von. 1866. Concerning the perceptions in general. *Treatise on physiological optics,*.

Hobson, J Allan, & Friston, Karl J. 2016. A response to our theatre critics. *Journal of Consciousness Studies*, **23**(3-4), 245–254.

Hoel, Erik P, Albantakis, Larissa, & Tononi, Giulio. 2013. Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences*, **110**(49), 19790–19795.

Hoel, Erik P, Albantakis, Larissa, Marshall, William, & Tononi, Giulio. 2016. Can the macro beat the micro? Integrated information across spatiotemporal scales. *Neuroscience of Consciousness*, **2016**(1), niw012.

Hohwy, Jakob. 2013. *The Predictive Mind*. Oxford University Press. 187 cites:.

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.

Kandel, Eric R, Schwartz, James H, Jessell, Thomas M, of Biochemistry, Department, Jessell, Molecular Biophysics Thomas, Siegelbaum, Steven, & Hudspeth, AJ. 2000. *Principles of neural science.* Vol. 4. McGraw-hill New York.

Kristan Jr, William B, & Shaw, Brian K. 1997. Population coding and behavioral choice. *Current opinion in neurobiology*, **7**(6), 826–831.

Lamme, Victor. 2015. *Predictive coding is unconscious, so that consciousness happens now.* Open MIND. Frankfurt am Main: MIND Group.

Lau, Hakwan, & Rosenthal, David. 2011. Empirical support for higher-order theories of conscious awareness. *Trends in cognitive sciences*, **15**(8), 365–373.

Laureys, Steven. 2005. The neural correlate of (un) awareness: lessons from the vegetative state. *Trends in cognitive sciences*, **9**(12), 556–559.

Lemon, RN, & Edgley, SA. 2010. Life without a cerebellum. *Brain*, **133**(3), 652–654.

Locke, John, & Nidditch, Peter H. 1979. *The Clarendon Edition of the Works of John Locke: An Essay Concerning Human Understanding.* Clarendon Press.

Luhmann, Niklas. 1995. Probleme mit operativer Schließung. *Soziologische Aufklärung*, **6**, 12–24.

Maass, Wolfgang, & Bishop, Christopher M. 2001. *Pulsed neural networks.* MIT press.

Malone, Thomas W, & Bernstein, Michael S. 2015. *Handbook of collective intelligence.* MIT Press.

Mathis, Donald W, & Mozer, Michael C. 1995. On the computational utility of consciousness. *Pages 11–18 of: Advances in neural information processing systems.*

Maturana, Humberto R, & Varela, Francisco J. 1991. *Autopoiesis and cognition: The realization of the living.* Vol. 42. Springer Science & Business Media.

Melloni, Lucia. 2015. Consciousness as Inference in Time. *Open MIND*, **22**.

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.*

Oizumi, Masafumi, Albantakis, Larissa, & Tononi, Giulio. 2014. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS computational biology*, **10**(5), e1003588.

O'Regan, J Kevin, & Noë, Alva. 2001. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, **24**(5), 939–973.

Overgaard, Morten, & Overgaard, Rikke. 2010. Neural correlates of contents and levels of consciousness. *Frontiers in psychology*, **1**, 164.

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.

Panzeri, Stefano, Macke, Jakob H, Gross, Joachim, & Kayser, Christoph. 2015. Neural population coding: combining insights from microscopic and mass signals. *Trends in cognitive sciences*, **19**(3), 162–172.

Pattee, Howard Hunt. 2012. Evolving self-reference: matter, symbols, and semantic closure. *Pages 211–226 of: Laws, language and life.* Springer.

Pennartz, Cyriel MA. 2015. *The brain's representational power: on consciousness and the integration of modalities.* MIT Press.

Pennartz, Cyriel MA. 2017. Consciousness, Representation, Action: The Importance of Being Goal-Directed. *Trends in cognitive sciences.*

Pennartz, Cyriel MA. 2018. Consciousness, representation, action: the importance of being goal-directed. *Trends in cognitive sciences*, **22**(2), 137–153.

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.

Pouget, Alexandre, Dayan, Peter, & Zemel, Richard. 2000. Information processing with population codes. *Nature Reviews Neuroscience*, **1**(2), 125.

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.

Quian Quiroga, Rodrigo, & Panzeri, Stefano. 2009. Extracting information from neuronal populations: Information theory and decoding approaches. *Nature Reviews Neuroscience*, **10**(3), 173–185.

Rosen, Robert. 1991. *Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life.* Columbia University Press.

Rosenthal, David M. 2005. *Consciousness and mind.* Oxford University Press.

Schurger, Aaron, Sitt, Jacobo D., & Dehaene, Stanislas. 2012. An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences*, **109**(42), E2904–E2913.

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.

Seth, Anil K. 2009. Function of consciousness. *Pages 279–293 of:* Banks, William P. (ed), *Encyclopedia of consciousness.* Elsevier/Academic Press.

Seth, Anil K. 2014. A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, **5**(2), 97–118. 51 cites:.

Shamir, Maoz. 2014. Emerging principles of population coding: In search for the neural code. *Current Opinion in Neurobiology*, **25**, 140–148.

Shoemaker, Sydney. 1982. The Inverted Spectrum. *Journal of Philosophy*, **79**(July), 357–381.

Stanley, Garrett B. 2013. Reading and writing the neural code. *Nature Neuroscience*, **16**(3), 259–263.

Stein, Richard B, Gossen, E Roderich, & Jones, Kelvin E. 2005. Neuronal variability: noise or part of the signal? *Nature Reviews Neuroscience*, **6**(5), 389.

Sutton, Richard S, & Barto, Andrew G. 2018. *Reinforcement learning: An introduction.* MIT press.

Tononi, Giulio. 2004. An information integration theory of consciousness. *BMC neuroscience*, **5**(1), 42.

Tononi, Giulio, & Edelman, Gerald M. 1998. Consciousness and complexity. *Science*, **282**(5395), 1846–1851.

Van den Bussche, Eva, Van den Noortgate, Wim, & Reynvoet, Bert. 2009. Mechanisms of Masked Priming: A Meta-Analysis. *Psychological Bulletin*, **135**(3), 452–477.

van Gaal, Simon, De Lange, Floris P, & Cohen, Michael X. 2012. The role of consciousness in cognitive control and decision making. *Frontiers in human neuroscience*, **6**, 121.

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.

Whitwell, Robert L., Milner, A. David, & Goodale, Melvyn A. 2014. The Two Visual Systems Hypothesis: New Challenges and Insights from Visual form Agnosic Patient DF. *Frontiers in Neurology*, **5**, 255.

Wimsatt, William C. 1994. The ontology of complex systems: levels of organization, perspectives, and causal thickets. *Canadian Journal of Philosophy*, **24**(sup1), 207–274.

Woodward, James. 2007. Causation with a Human Face. *In:* Price, Huw, & Corry, Richard (eds), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford University Press.

Backup

Biological evidence of informational closure in neural system

- So far, not so much research on non-trivial informational closure in biological neural systems.
- However, [[REWORDING | XXX]]
- Recent biological evidence suggests that low-level neurons encodes predictive information about external and
- **[[need to say this is exactly the non-trivial informational 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."*
- *" Word–word internal predictive information is correlated with word–stimulus 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 + delta t in the future? The answer to this question is contained in the conditional distribution of one word on the other, P ð w t +delta t j w t Þ "*
http://tinyurl.com/y9zmhlln

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

Relation to empirical findings in neuroscience and studies of consciousness

0.1 Unconscious Processing

- According the theory, there are two factor that renders neural processes unconscious.
- First, processes are reflexive. This causes non-closured
- Second, the abstraction level of the information is below the non-trivial informational closure scale.

0.2 reflexive behaviours is unconscious (or environment-state dependent processes)

0.2.1 Reflection

- Reflexive systems is not informational closured.
- Base d on our theory,
- x -> s -> a
  - the system state is determined by environmental state, then it is not conscious.
- Therefore, in general, a simple policy network is not conscious.

0.2.2 Blindsight

- This also blindsight.
- Blindsight patients and still make above chance judgement without conscious awareness of visual targets.
- The stimulus-reaction policy network still intact, but it cannot form informational closure and therefore, patients has no conscious awareness of visual targets
- Patient B.F. reactions are stimulus driven.

0.2.3 procedure memory

0.2.4 cerebellum

- cerebellum feed-forward connections, maybe just a policy network.

Ref: cerebellum

0.3 Under the level of informational closure

0.3.1 statistical inference

- The computation of Bayesian brain is unconscious.

0.3.2 subliminal perception

- Above chance performance can be performed through non-closure pathway (reflexive behaviours)
- Due to coarse-graining, mapping between changes of microstate

0.3.3 Libet Experiments

0.4 Deterministic vs probabilistic

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

About Phenomenology (conscious processing)

0.5 Perceptual overflow

0.6 Volition and freewill

- Volition is a result of coarse-graining
- Lebel's experiment can be well explained by the experiment.

0.7 Planning

0.8 Imagination

0.9 Dream vs non-dream sleep

0.10 Binocular rivalry

0.11 Hallucinations

0.12 Synaesthesia

Theory predictions (Counter-intuitive prediction)

0.13 informational closure

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

0.14 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.

Backup -