

Information Closure Theory of Consciousness

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

ARAYA, Inc., Tokyo, Japan

Word Count:

Total Words :7115

Headers :27

Math Inline :112

Math Display :12

Contents

1	Introduction	3
2	Non-trivial Informational Closure	4
3	Coarse-graining in the Neural System	6
4	Information Closure Theory of Consciousness (ICT)	6
4.1	Level of Consciousness is Equal to the Degree of NTIC of a Process	8
4.2	Conscious Contents Corresponding to States of a NTIC Process	9
4.3	Reconciling the Levels and Contents of Consciousness	10
5	Conscious Versus Unconscious Processing	10
5.1	Unconscious Processing	10
5.2	Conscious Processing	12
6	Comparison with Other Relevant Theories of Consciousness	13
6.1	Multilevel Views on Consciousness and Cognition	13
6.2	Integrated Information Theory	14
6.3	Predictive Processing	15
6.4	Sensorimotor Contingency	16
6.5	Global Workspace Theory	16
7	Limitation and Future Work	16
8	Conclusions	18
	Acknowledgements	18
	Author Contributions Statement	18
	Conflict of Interest Statement	19
	Reference	19

*Corresponding author, acercyc@araya.org

†martin@araya.org

‡yen.yu@araya.org

§kanair@araya.org

34

Abstract

35 Information processing in neural systems can be described and analysed at multiple spa-
36 tiotemporal scales. Generally, information at lower levels is more fine-grained and can be
37 coarse-grained in higher levels. However, information processed only at specific levels seems
38 to be available for conscious awareness. We do not have direct experience of information
39 available at the level of individual neurons, which is noisy and highly stochastic. Neither do
40 we have experience of more macro-level interactions such as interpersonal communications.
41 Neurophysiological evidence suggests that conscious experiences co-vary with information en-
42 coded in coarse-grained neural states such as the firing pattern of a population of neurons.
43 In this article, we introduce a new informational theory of consciousness: Information Clo-
44 sure Theory of Consciousness (ICT). We hypothesise that conscious processes are processes
45 which form non-trivial informational closure (NTIC) with respect to the environment at cer-
46 tain coarse-grained levels. This hypothesis implies that conscious experience is confined due
47 to informational closure from conscious processing to other coarse-grained levels. ICT pro-
48 poses new quantitative definitions of both conscious content and conscious level. With the
49 parsimonious definitions and a hypothesis, ICT provides explanations and predictions of var-
50 ious phenomena associated with consciousness. The implications of ICT naturally reconciles
51 issues in many existing theories of consciousness and provides explanations for many of our
52 intuitions about consciousness. Most importantly, ICT demonstrates that information can be
53 the common language between consciousness and physical reality.

54 **Keywords:**

55 Keywords: theory of consciousness, non-trivial informational closure, NTIC, coarse-graining, level
56 of analysis

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 the neurons in Alice’s supplementary motor area and involved in many motor control processes for Alice’s actions, 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 happens 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 reflect the dynamic of your membrane potential or the action potentials you send to other neurons every second. That is, not all information you have is available to Alice’s conscious mind.

It seems to be true that we are not consciously accessing information processed at every scale in the neural system. There are both more microscopic and more macroscopic levels than the level corresponding to the conscious contents. On the one hand, dynamics of individual neurons are stochastic (Goldwyn & Shea-Brown, 2011; White *et al.*, 2000). However, 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). In addition, some parts of the neural system contribute very little to conscious experience (the cerebellum for example (Lemon & Edgley, 2010)), also suggesting that conscious contents do not have one-to-one mapping to the entire state of the neural system. On the other hand, human conscious experience is more detailed than just a simple (e.g. binary) process can represent, suggesting that the state space of conscious experience is much larger than what a single overly coarse-grained binary variable can represent. These facts suggest that conscious processes occur at a particular scale. We currently lack a theory to identify the scale which conscious processes correspond to. We refer to this notion as **the scale problem of consciousness** (Fig. 1).

In this article, we propose a new information-based theory of consciousness, called Information Closure Theory of Consciousness (ICT). We argue that every process with a positive non-trivial information closure (NTIC) has consciousness. This means that the state of such a process corresponds one-to-one to conscious content.¹ We further postulate that the *level* of consciousness corresponds to the degree of NTIC. (for a discussion of the distinction between level versus content of consciousness see Laureys (2005); Overgaard & Overgaard (2010)).

In the following, we first introduce non-trivial informational closure and argue for its importance to information processing for human scale agents (Sec.2). We next argue that through coarse-graining the neural system can form a high degree of NTIC at a specific coarse-grained level (Sec.3). In the Sec.4, we propose a new theory of consciousness (ICT). We also illustrate how ICT can parsimoniously explain empirical findings from previous consciousness studies (Sec.5) and reconcile several current major theories of consciousness (Sec.6). Finally, we discuss the current theoretical and empirical limitations of ICT and propose the implications from ICT to the current consciousness science (Sec.7).

¹In the following IC stands for "informational closure" or "informationally closed" and NTIC stands for "non-trivial informational closure" or "non-trivially informationally closed".

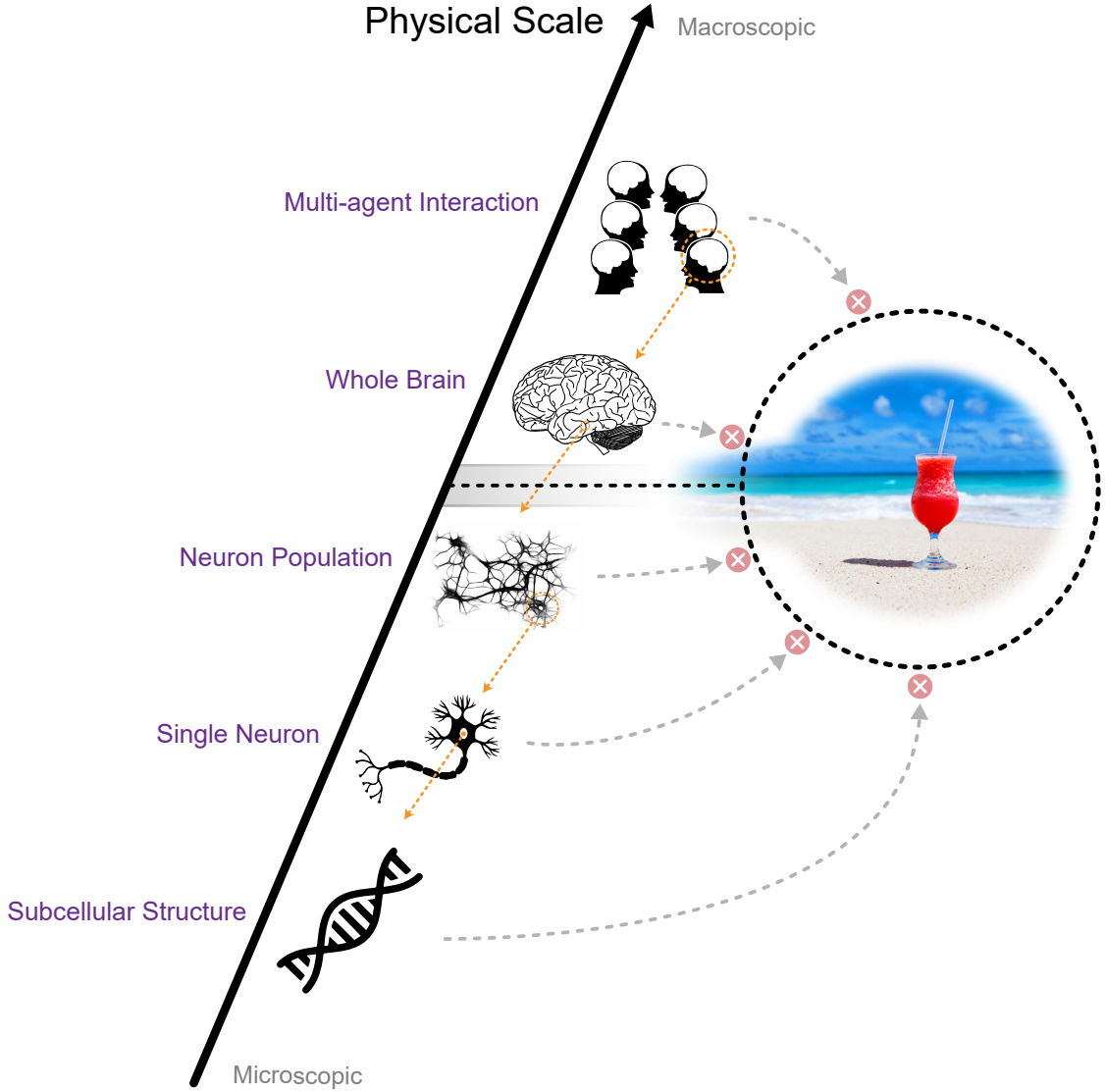


Figure 1: The scale problem of consciousness: Human conscious experience does not reflect information from every scale. Only information at a certain coarse-grained scale in the neural system is reflected in consciousness.

97 2 Non-trivial Informational Closure

98 The notion of non-trivial informational closure (NTIC) is introduced by [Bertschinger et al. \(2006\)](#).
 99 The concept of closure is closely related to system identification in systems theory. One can distinguish a system from its environment by computing the closedness of the system ([Luhmann, 1995](#);
 100 [Maturana & Varela, 1991](#); [Pattee, 2012](#); [Rosen, 1991](#)). The closedness can be further quantified
 101 by information theory.

102 Consider two processes, the environment process $(E_t)_{t \in \mathbb{N}}$ and the system's process $(Y_t)_{t \in \mathbb{N}}$ and
 103 let their interaction be described by the Bayesian network in Fig. 2. Then, information flow J_t from
 104 the environment E to a system S at time t can be defined as the conditional mutual information
 105 I between the current environment state E_t and the future system state Y_{t+1} given the current
 106 system state Y_t

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

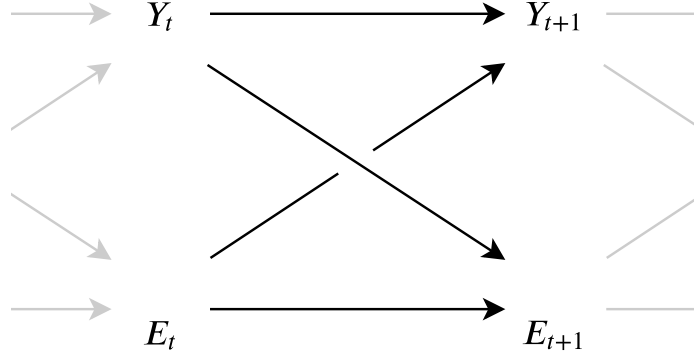


Figure 2: The dependencies between a system and its environment.

108 Bertschinger *et al.* (2006) defines that a system is informationally closed when information flow
 109 from the environment to the system is zero.

$$J_t(E \rightarrow Y) = 0 \quad (2)$$

110 Information closure (minimising J_t) is trivial if the environment and the system are entirely inde-
 111 pendent of each other.

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

112 However, informational closure can be formed non-trivially. In the non-trivial case, even though a
 113 system contains (or encodes) information about the environmental dynamics, the system can still
 114 be informationally closed. In such cases, the mutual information between the current states of the
 115 environment and the future state of the system is larger than zero.

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

116 This also implies

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

And, non-trivial informational closure can be defined as

$$NTIC_t(E \rightarrow Y) := I(Y_{t+1}; Y_t) - I(Y_{t+1}; Y_t | E_t) \quad (6)$$

$$= I(Y_{t+1}; E_t) - I(Y_{t+1}; E_t | Y_t) \quad (7)$$

117 Hence, maximising $NTIC_t(E \rightarrow Y)$ amounts to

$$\begin{aligned} &\text{maximising } I(Y_{t+1}; Y_t) \quad \text{and} \\ &\text{minimising } I(Y_{t+1}; Y_t | E_t) \end{aligned} \quad (8)$$

118 One can also maximise $NTIC_t(E \rightarrow Y)$ by

$$\begin{aligned} &\text{maximising } I(Y_{t+1}; E_t) \quad \text{and} \\ &\text{minimising } I(Y_{t+1}; E_t | Y_t) \end{aligned} \quad (9)$$

119 This implies the system contains in itself all the information about its own future and the self-
 120 predictive information contains the information about the environment. Therefore, to form NTIC,
 121 the system can internalise and synchronise with the dynamics of the environment, e.g., model
 122 the environment. Furthermore, having high degrees of NTIC entails having high predictive power
 123 about the environment. This gives biological agents a great functional and evolutionary advantage.

124 3 Coarse-graining in the Neural System

125 The formation of NTIC with a highly stochastic process is challenging. NTIC requires the pre-
126 dictability of the system state and is therefore impeded by noise in the system. Information
127 processing at the microscopic levels (the cellular levels) in neural systems suffers from multiple
128 environmental noise sources such as sensor, cellular, electrical, and synaptic noises. For exam-
129 ple, neurons exhibit large trial-to-trial variability at the cellular level, and are subject to thermal
130 fluctuations and other physical noises (Faisal *et al.*, 2008).

131 However, it is possible that neural systems form NTIC at certain macroscopic levels through
132 coarse-graining of microscopic neural states. Coarse-graining refers to many-to-one or one-to-one
133 maps which aggregate microscopic states to a macroscopic state. In other words, a number of
134 different micro-states correspond to the same value of the macro-variable (Price & Corry, 2007).
135 Coarse-grainings, can therefore form more stable and deterministic state transitions and more often
136 form NTIC processes. For neural systems this means that a microscopically noisy neural system
137 may still give rise to an NTIC process on a more macroscopic scale.

138 Indeed, empirical evidence suggests that coarse-graining is a common coding strategy to estab-
139 lish robustness against noise at microscopic levels of the neural system. For instance, the inter-spike
140 intervals of an individual neuron are stochastic. This implies that the state of an individual neuron
141 does not represent stable information. However, the firing rate, i.e. the average spike counts over
142 a given time interval, is more stable and robust against noise such as the variability in inter-spike
143 intervals. Using this temporal coarse-graining strategy, known as rate coding (Adrian, 1926; Ger-
144 stner & Kistler, 2002; Maass & Bishop, 2001; Panzeri *et al.*, 2015; Stein *et al.*, 2005), neurons can
145 encode stimulus intensity by increasing or decreasing the firing rate (Kandel *et al.*, 2000). (Stein
146 *et al.*, 2005). The robustness of the rate coding is a direct consequence of the many to one mapping
147 (i.e., coarse-graining).

148 Population coding is another example of encoding information through coarse-graining in neural
149 systems. In this coding scheme, information is encoded by activation patterns of a set of neurons
150 (a neuron population). In the population coding scheme, many states for a neuron population
151 map to the same state of macroscopic variables which encode particular informational contents,
152 thereby reducing the influence of noise in individual neurons. That is, stable representations can
153 be formed through coarse-graining the high dimensional state space of a neuron population to a
154 lower dimensional macroscopic state space (Binder *et al.*, 2009; Kristan Jr & Shaw, 1997; Pouget
155 *et al.*, 2000; Quiari Quiroga & Panzeri, 2009). Therefore, individual neuron states (the microscopic
156 level) are not informative enough about the complete encoded contents at the population level
157 (the macroscopic level). Instead, coarse-grained variables are better substrates for stably encoding
158 information and allow the neural system to ignore noisy interactions at the fine-grained level
159 (Woodward, 2007).

160 These two examples show that the known coding schemes can be viewed as coarse-graining,
161 and provide stochastic neural systems with the ability to form more stable and deterministic
162 macroscopic processes for encoding and processing information reliably. We argue that through
163 coarse-graining the neural systems is able to form NTIC processes at macroscopic levels. Based
164 on the merit of coarse-graining in neural systems, we propose a new theory of consciousness in the
165 next section.

166 4 Information Closure Theory of Consciousness (ICT)

167 In this section, we propose a new theoretical framework of consciousness: Information Closure
168 Theory of Consciousness (ICT). The main hypothesis is that conscious processes are captured by
169 what we call *C-processes*. We first define C-processes, then state our hypothesis and discuss its
170 implications.

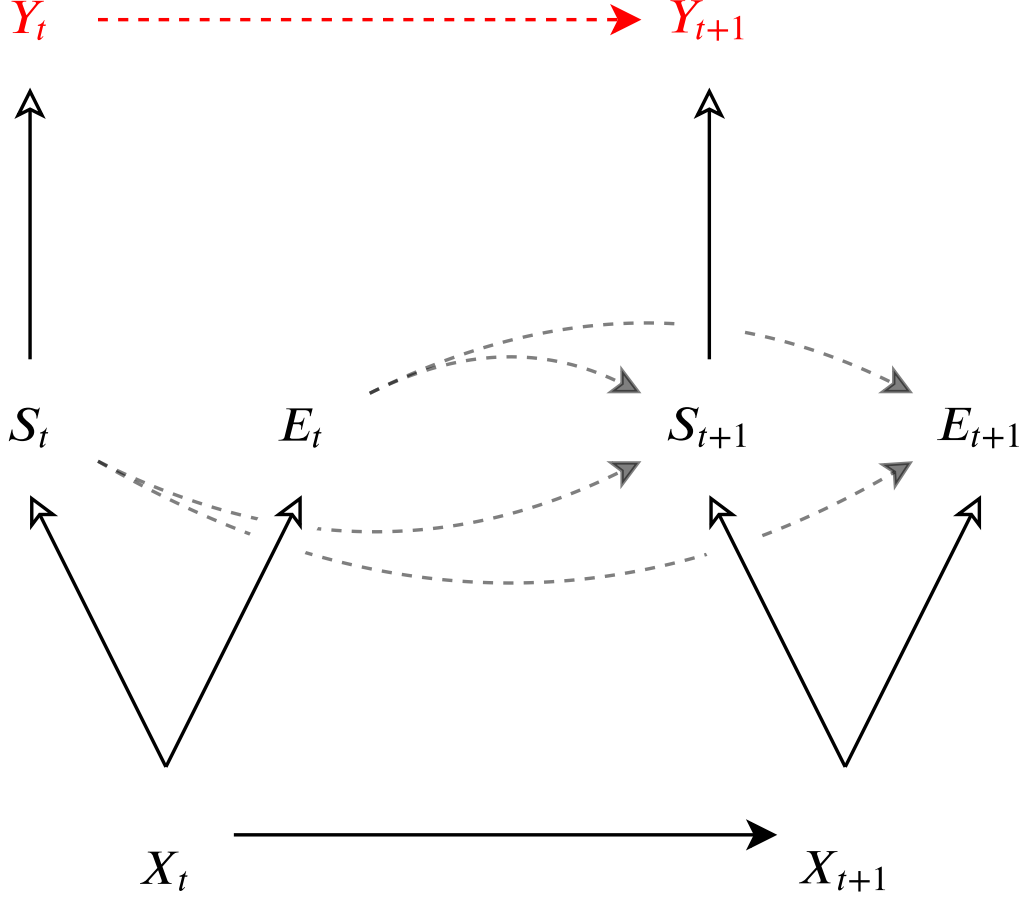


Figure 3: The information flow amounts the universe X , the system S , the environment of the system E , and the coarse-grained process Y of the system S . The solid line with a filled arrow from X_t to X_{t+1} represents the microscopic dynamic of the universe. The solid lines with a empty arrow represent directions of coarse-graining. The dashed lines represents virtual dependencies between two macroscopic variables. The red Y_t , Y_{t+1} , and the red dashed line in between represents a macroscopic process which forms informational closure at a certain coarse-grained level.

171 In order to define C-processes we need to define coarse-grainings first. Every coarse-graining is
 172 characterised by a function that maps the microscopic process to the coarse-grained macroscopic
 173 process. More formally:

174 **Definition 1.** Given a stochastic process X with state space \mathcal{X} , a coarse-graining of X is a
 175 stochastic process Y with state space \mathcal{Y} such that there exists a function ² $f_Y : \mathcal{X} \rightarrow \mathcal{Y}$ with
 176 $Y_t = f_Y(X_t)$.

177 A more general definition of coarse-grainings that maps temporally extended sequences of the
 178 microscopic process to macroscopic states are possible but for his first exposure of our theory the
 179 simpler definition above is sufficient.

180 **Definition 2.** Given a stochastic process X called the universe process, a C-process is a coarse-
 181 graining Y of X such that the following two conditions are satisfied (see Fig. 3):

- 182 1. Y is informationally closed to X
- 183 2. there exists a pair (S, E) of coarse-grainings of X such that
 - 184 • Y is a coarse-graining of S ,

²Functions in the mathematical sense used here are always either one-to-one or many-to-one.

- the state space \mathcal{X} of X is equal to the Cartesian product of the state spaces \mathcal{S} and \mathcal{E} of processes S and E respectively, formally $\mathcal{X} = \mathcal{S} \times \mathcal{E}$, and
- Y is NTIC to E , formally:

$$NTIC_t(E \rightarrow Y) > 0 \quad (10)$$

With the two definitions we can state the main hypothesis of ICT:

Hypothesis. A process Y is conscious if and only if it is a C -process of some process X . Also the content of consciousness $C_t^{Content}$ at time t is the state y_t of the C -process at time t and the level of consciousness C_t^{Level} is the degree of NTIC of the C -process to the environment i.e. $NTIC_t(E \rightarrow Y)$:

$$C_t^{Content} = y_t \quad (11)$$

$$C_t^{Level} = NTIC_t(E \rightarrow Y) \quad (12)$$

A concrete example in the context of neuroscience is that X represents the microscopic level of the universe, S a cellular level process in the neural system, Y a more macroscopic process of the neural system coarse-grained from the cellular level process S , and E the environment which the cellular level process S interacts with. The environment E may include other processes in the neural system, the sensors for perception and interoception, and external physical worlds.

Based on the hypothesis, ICT leads to five core implications:

Implication 1. Consciousness is information. Here, "informative" refers to the resolution of uncertainty. Being in a certain conscious state rules out other possible conscious states. Therefore, every conscious percept resolves some amount of uncertainty and provides information. This implication is also in agreement with the "axiom" of *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)

Implication 2. Consciousness is associated with physical substrates and the self-information of the conscious percept is equal to the self-information of the corresponding physical event. This is a direct implication from our hypothesis that every conscious percept $C_t^{Content}$ corresponds to a physical event y_t .

Implication 3. Conscious processes are self-determining. This is a direct implication of the requirement that Y is informationally closed with respect to X . To be informationally closed with respect to X , no coarse-graining knows anything about the conscious process' future that the conscious process does not know itself. This self-determining characteristics is also in line with our daily life conscious experience which often shows stability and continuity and is ignorant of the stochasticity (e.g., noise) of the cellular levels.

Implication 4. Conscious processes encode the environmental influence on itself. This is due to the non-triviality of the informational closure of Y to E . At the same time all of this information is known to the conscious processes themselves since they are informationally closed with respect to their environments. This also suggests that conscious processes can model the environmental influence without knowing more information from the environment.

Implication 5. Conscious processes can encode environmental information (by forming NTIC), however, be ignorant to part of the information of more microscopic processes (from Implication 3 and 4). This is in line with our conscious experience that information that every conscious percept provides represent rich and structured environmental states without involving all the information about microscopic activities.

4.1 Level of Consciousness is Equal to the Degree of NTIC of a Process

According to Eq. 8, ICT implies that conscious levels are determined by two quantities.

First, to form a high level of NTIC, one can increase the mutual information $I(Y_{t+1}; Y_t)$ between the current internal state Y_t and the future internal state Y_{t+1} . In other words, conscious levels

are associated with the degree of self-predictive information (Bialek *et al.*, 2001). This mutual information term can be further decomposed to two information entropy quantities:

$$I(Y_{t+1}; Y_t) = H(Y_{t+1}) - H(Y_{t+1}|Y_t) \quad (13)$$

This implies that a highly NTIC process must have rich dynamics with self-predictability over time. Another implication is that complex systems can potentially attain higher levels of consciousness due to the larger information capacities needed to attain high mutual information. This outcome is consistent with the common intuition that conscious levels are often associated with the degree of complexity of a system.

Second, one can minimise the conditional mutual information $I(Y_{t+1}; Y_t|E_t)$ to increase the level of NTIC. This quantity suggests that conscious level increases with the amount of information about the environment state E_t that the NTIC process encodes in its own state Y_t . In other words, Y_t should not contain more information about Y_{t+1} than E_t . An important implication is that agents interacting with a complex environment have the chance to build a higher level of NTIC within their systems than those living in a simple environment. In other words, the level of consciousness is associated with environmental complexity.

It is important to note that NTIC can be a non-monotonic function of the scale of coarse-graining. We saw above that not sufficiently coarse-grained variables have low values of NTIC. On the other hand, overly coarse-grained macroscopic variables also result in low values of NTIC. For example, in an extreme scenario, when all microscopic states map to a single macroscopic variable, the macroscopic level does not have any information capacity and thus cannot have high mutual information across time steps. Therefore, only processes at a certain level of coarse-graining in the neural system can form a high degree of NTIC (Fig. 4). ICT indicates that human consciousness occurs at the level of coarse-graining where higher NTIC is formed within the neural system.

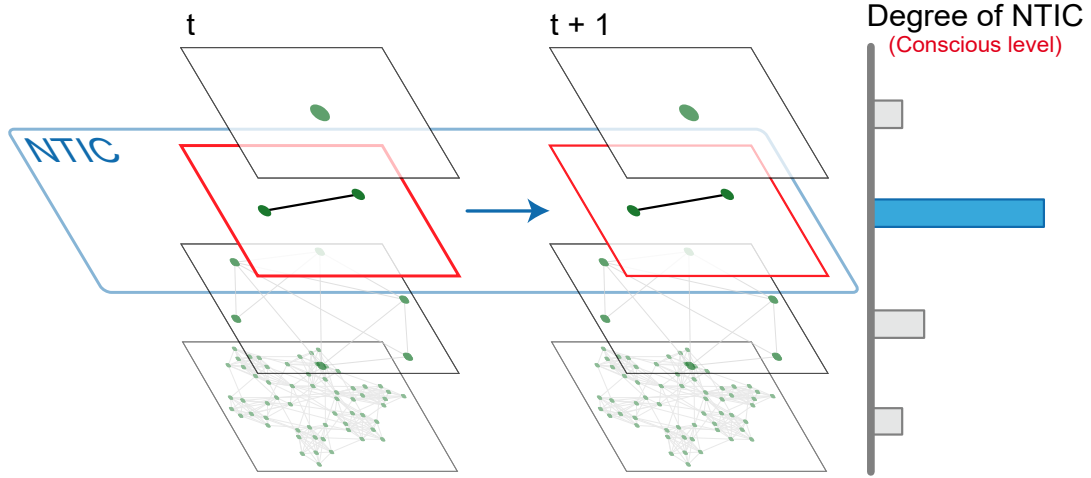


Figure 4: A non-monotonic relationship between Level of coarse-graining and level of consciousness.

4.2 Conscious Contents Corresponding to States of a NTIC Process

ICT proposes that conscious contents correspond to the states of NTIC processes (Eq. 11). This implies that the size of the state space of an NTIC process is associated with the richness of conscious contents that the process can potentially have. Therefore, a complex NTIC process with a high dimensional state space can have richer conscious experience than a simple NTIC process can have. This outcome is consistent with the intuition that the richness of conscious contents is associated with the complexity of a system.

As mentioned above, informational closure can happen between scales of coarse-graining within a single system. Thus, a macroscopic NTIC process can be ignorant to its microscopic states. ICT argues that human conscious contents do not reflect cellular level activity because the conscious process which corresponds to a macroscopic NTIC process is informationally closed to the cellular level in the human neural system. Further more, since NTIC processes are informationally closed,

each of them can be considered as a reality. In the extreme case, when the information flow from its microscopic processes and environment to the informationally closed process is completely zero (Eq. 2), the future states of the process is only determined by its past states.

Importantly, NTIC processes encode environmental information in its state. This suggests that a NTIC process can be considered as a process that simulates the environmental dynamics. This implication fits well with some theories of consciousness (for example, world simulation metaphor (Revonsuo, 2006)). Note that ICT doesn't assume that generative models are necessary for consciousness. The implication is a natural result of processes with NTIC.

Finally, a coarse-graining can be a many to one map from microscopic to macroscopic states and ICT proposes that conscious contents $C^{Content}$ is the state of the NTIC process Y . Therefore, ICT implies multiple realisation thesis of consciousness (Bechtel & Mundale, 1999; Putnam, 1967) which suggests that different physical implementations could map to the same conscious experience.

4.3 Reconciling the Levels and Contents of Consciousness

While it is useful to distinguish the notion of the levels and contents of consciousness, whether they can be clearly dissociated has been a matter of debate (Bayne *et al.*, 2016; Fazekas & Overgaard, 2016). In ICT, conscious levels and conscious contents are just two different properties of NTIC processes, and, therefore, naturally reconciles the two aspects of consciousness. In an NTIC process with a large state space, conscious contents should also consist of rich and high dimensional information. Therefore, this framework integrates the levels and the contents of consciousness in a coherent fashion by providing explicit formal definitions of the two notions.

According to Sec. 4.1 and Sec. 4.2, an important implication from ICT is that both conscious levels and conscious contents are associated with the state space of an NTIC process Y . A large state space of Y contributes conscious levels through the mutual information $I(Y_{t+1}; Y_t)$ and also contributes richer conscious contents by providing more possible states of conscious processes. ICT therefore explains why, in normal physiological states, conscious levels and conscious contents are often positively correlated (Laureys, 2005). This implication is also in line with the intuition in which consciousness is often associated with complex systems.

5 Conscious Versus Unconscious Processing

In this section, we show how ICT can explain and make predictions about what processes are conscious and what are unconscious. ICT is constructed using information theory. Therefore, ICT can provide predictions based on the mathematical definitions.

5.1 Unconscious Processing

Regarding unconscious processing, we highlight two scenarios in which the degree of NTIC is rendered low for a process, and thereby making the process less conscious.

Informational is not Closed

If a process is not informationally closed, the degree of NTIC is low (Eq. 9) resulting in low or no consciousness. In such cases, the current state of a process depends primarily on the environment state (see Fig. 5), but receives little influence from its past state. Reflexive behaviours (Casali *et al.*, 2013) can be considered an example of this scenario. In ICT, if we can view reflexive behaviours as situations in which the internal state Y_t , which triggers reflexive action, is determined by the environment state E_{t-1} overruling the influences from its own past Y_{t-1} . Such interpretation of reflexive behaviour from the viewpoint of ICT naturally explains why reflexes do not involve conscious experience of external stimuli.

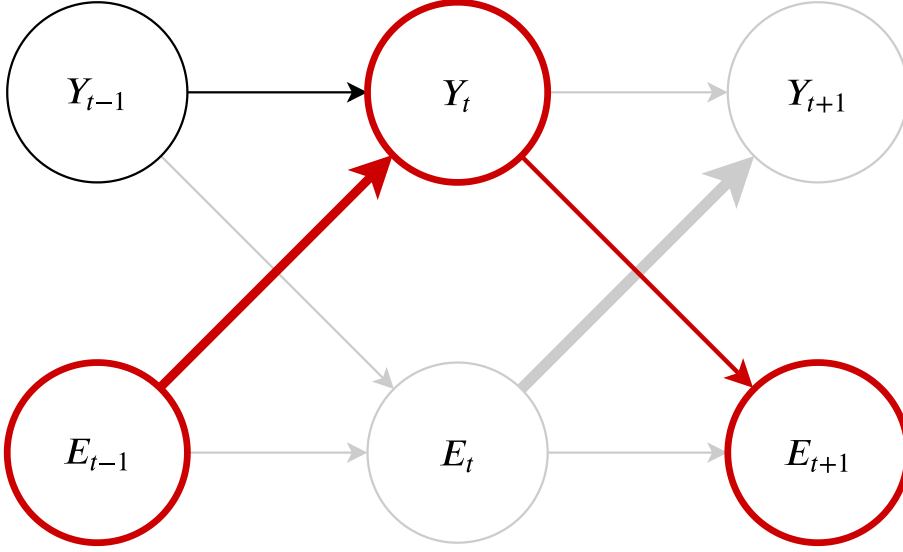


Figure 5: A diagram depicting the information flow in reflexive behaviours (shown by the red nodes and arrows) happening through the interaction between a process Y and its environment E . In such situations, the internal state Y_t is mostly dependent on the environment state E_{t-1} but less on its past state Y_{t-1} . Therefore, the process Y is not informational closed. As a consequence, Y is unable to form high NTIC and, therefore, remain less conscious or unconscious.

303 The same principle can be applied to interpret blindsight (Humphrey, 1970, 1999, 1974) and
 304 procedural memory (Ashby *et al.*, 2010; Doyon *et al.*, 2009) which are often considered as uncon-
 305 scious processes. Blindsight patients are able to track objects, avoid obstacles, and make above
 306 chance-level visual judgements when their visual experiences is degraded or missing (however, they
 307 may still preserve some forms of conscious experience, see Mazzi *et al.* (2016); Overgaard (2011)).
 308 We argue that action outputs of blindsight can be directly guided by sensory inputs through
 309 stimulus-response maps. The neural circuits are not informationally closed and, therefore, uncon-
 310 scious. Similarly, for procedural memory, the state transitions of action controls largely depends
 311 on the concurrent environmental states. This prevents the internal processes of procedure memory
 312 from informational closure and being conscious. ICT also offer an interpretation as to why patients
 313 with visual apperceptive agnosia (James *et al.*, 2003) can perform online motor controls without
 314 visual awareness of action targets (Whitwell *et al.*, 2014).

315 As we have seen in the examples above, a crucial implication of ICT is that a pure feedfor-
 316 ward network cannot produce consciousness because NTIC requires a form of memory $I(Y_{t+1}; Y_t)$.
 317 Without memory, a network’s current state is entirely driven by the input the network receives
 318 from the environment without any influence from its own past states. Therefore, such a network is
 319 incapable of forming an NTIC. In contrast, a network with recurrent loops can maintain informa-
 320 tion about its own past states. This forms an information channel between the past and the future
 321 states of the network and, thus, makes the network capable of being informationally closed. This
 322 result coincides with theories of consciousness emphasising the importance of recurrent circuits to
 323 consciousness (Edelman, 1992; Lamme, 2006; Tononi & Koch, 2008).

324 Information is Trivial

325 According to ICT, when encoded information in a process is trivial, i.e. no mutual information
 326 between the process states and its environment states $I(Y_{t+1}; E_t)$ (Eq. 9), this could lead low
 327 NTIC. In such case, this process is considered to be unconscious due to the low level of NTIC. This
 328 implies an isolated process which is simply informationally closed is insufficient to be conscious.
 329 This mathematical property of ICT provides a natural and intuitive (but only partial, see the
 330 current challenge in Sec. 7) solution to the boundary and the individuality problem of consciousness
 331 ³ (Raymont & Brook, 2006). Consider a NTIC process Y and an isolated informationally closed

³The boundary problem of consciousness refers to identifying physical boundaries of conscious processes and the individuality problem of consciousness refers to identifying individual consciousnesses in the universe.

process \hat{Y} with only trivial information. Adding \hat{Y} to Y can still keep informational closure, but, however, does not increase non-trivial information, i.e. doesn't affect consciousness.

$$\begin{aligned}
I(Y, \hat{Y}; E) &= H(Y, \hat{Y}) - H(Y, \hat{Y}|E) \\
&= H(Y) + H(\hat{Y}|Y) - (H(Y|E) + H(\hat{Y}|Y, E)) \\
&= H(Y) + H(\hat{Y}) - (H(Y|E) + H(\hat{Y})) \\
&= H(Y) - H(Y|E) \\
&= I(Y; E)
\end{aligned} \tag{14}$$

This implies that isolated processes with trivial information do not contribute consciousness and should be considered being outside the information boundary of the conscious processing (for more details of the boundary detection procedure, see [Krakauer et al. \(2014\)](#)). This property also implies that consciousnesses do not emerge from just aggregating informationally closed (isolated) processes which contain trivial information.

5.2 Conscious Processing

According to ICT, we claim that any process, system, or cognitive function which involves any NTIC process should be accompanied by conscious experience.

Previous consciousness research has identified a number of diverse cognitive processes often accompanied by conscious experience. ICT provides an integrated account for the reason why these processes involve conscious experience. As mentioned above, an NTIC process can be seen as an internal simulation engine for the agent-environmental interactions ([Bertschinger et al., 2006](#)). Therefore, information encoded in NTIC processes is essential for several cognitive processes.

One of the most valuable information is the predictions about the environmental states. Cognitive functions requiring agent-scale environmental predictions are likely to recruit NTIC processes and therefore accompanied by conscious experience, for example planning and achieving long term goals.

Second, as a simulation engine, with a given initial state, an NTIC process can self-evolve and simulate the environmental transitions. Cognitive functions involving simulations are expected to involve NTIC processes. Consequently, mental simulation, imagination, computing alternative realities, and generating counterfactuals often come with conscious experience.

Third, as an informationally closed system, an NTIC process can still provide environmental information without new sensory inputs. This is crucial for many types of off-line processing. Therefore, in contrast to reflexive-like behaviours mentioned above (Sec. 5.1), behaviours requiring off-line computations ([Himmelbach & Karnath, 2005](#); [Milner et al., 1999](#); [Revol et al., 2003](#)) often involve conscious experience.

Finally, for agents adapting to complex environments (e.g., human being), any state of the NTIC process can be seen as an integration of high dimensional information. To accurately encode information about the complex environmental states and transitions, the NTIC process requires knowledge about the complex causal dependencies involved in the environment. Therefore, cognitive functions requiring large scale integration are likely to involve NTIC processes and accompanied by conscious experience.

Note that many of the claims above are compatible with several theories of consciousness which highlight the connection between consciousness and internal simulation, predictive mechanism, or generative models inside a system (e.g. world simulation metaphor ([Revonsuo, 2006](#)), predictive processing and Bayesian brain ([Clark, 2013](#); [Hohwy, 2013](#); [Seth, 2014](#)), generative model and information generation ([Kanai et al., 2019](#))). Instead of relating functional or mechanistic aspects of a system to consciousness, ICT captures common informational properties underlying those cognitive functions associated with consciousness. As such, ICT does not assume any functionalist perspective of consciousness, which associate specific functions to consciousness. That is to say, since ICT associates information with consciousness, functional features accompanied by consciousness are collateral consequences of neural systems utilising NTIC processes for adaptive functions.

In sum, we argue that cognitive functions involving the NTIC process are inevitably accompanied by consciousness. Having an NTIC process is potentially an effective approach to increase fitness in the evolution. It is likely that biological creatures evolve NTIC processes at some point in the evolution. Due to the fundamental relation between information and consciousness, biological

380 creatures also evolve different degrees of consciousness depending on the physical scales and the
381 complexity of the environments they adapt to.

382 ICT starts with a non-functional hypothesis, however, it accounts for the association between
383 functional and consciousness. ICT further demonstrates remarkable explanatory power for various
384 findings of conscious and unconscious processing.

385 6 Comparison with Other Relevant Theories of Conscious- 386 ness

387 In this section, we compare ICT with other relevant theories of consciousness.

388 6.1 Multilevel Views on Consciousness and Cognition

389 ICT proposes that conscious processes can occur at any level of coarse-graining which forms NTIC
390 within a system. This suggests that the scale of coarse-graining is critical for searching and iden-
391 tifying the information corresponding to consciousness. A few versions of multilevel views on con-
392 sciousness have previously been (explicitly or implicitly) proposed. To our knowledge, Pennartz's
393 neurorepresentational theory (also called Neurorepresentationalism, (Pennartz, 2015, 2018)) is the
394 proposal closest to the multilevel view of ICT. Similar to Neurorepresentationalism, the concept
395 of levels in ICT is also relevant to Marr's level of analysis (Marr, 1982; Pennartz, 2015, 2018).
396 However, ICT suggests that coarse-graining is necessary only when the microscopic processes are
397 stochastic (e.g. the neural system). An NTIC process can be formed in a noise-free deterministic
398 system without coarse-graining. According to ICT, this NTIC process is sufficient to be conscious.
399 Another fundamental difference between ICT and Neurorepresentationalism is that Neurorepre-
400 sentationalism takes functionalist perspective and suggests consciousness should serve high-level
401 world-modelling and makes a best guess about the interaction between the body and the envi-
402 ronment. However, ICT is grounded by non-functional informational hypothesis. Therefore, ICT
403 provides a more fundamental explanation for the scale problem of consciousness.

404 Another well-known proposal based on multilevel views is the Intermediate Level Theory of
405 Consciousness (Jackendoff, 1987; Prinz, 2007, ILT). ILT proposes that conscious experience is
406 only associated with neural representations at intermediate **levels of the sensory processing**
407 **hierarchy** (e.g., the 2.5D representation of visual processing) rather than lower (e.g., pixel) or
408 higher (e.g., abstract) levels of the sensory hierarchy.

409 Here, we want to make clear that the "level" in ICT refers to the **levels of coarse-graining**
410 instead of the "level" for the cortical anatomy or sensory processing. It is important to note that
411 the coarse-graining direction is an orthogonal dimension irrespective of the level of anatomy or the
412 level of information processing hierarchy in the neural system (see Fig. 6). Because ILT focuses on
413 the levels of the sensory processing hierarchy and ICT focus on informational closure among the
414 levels of coarse-graining, the two theories are fundamentally different.

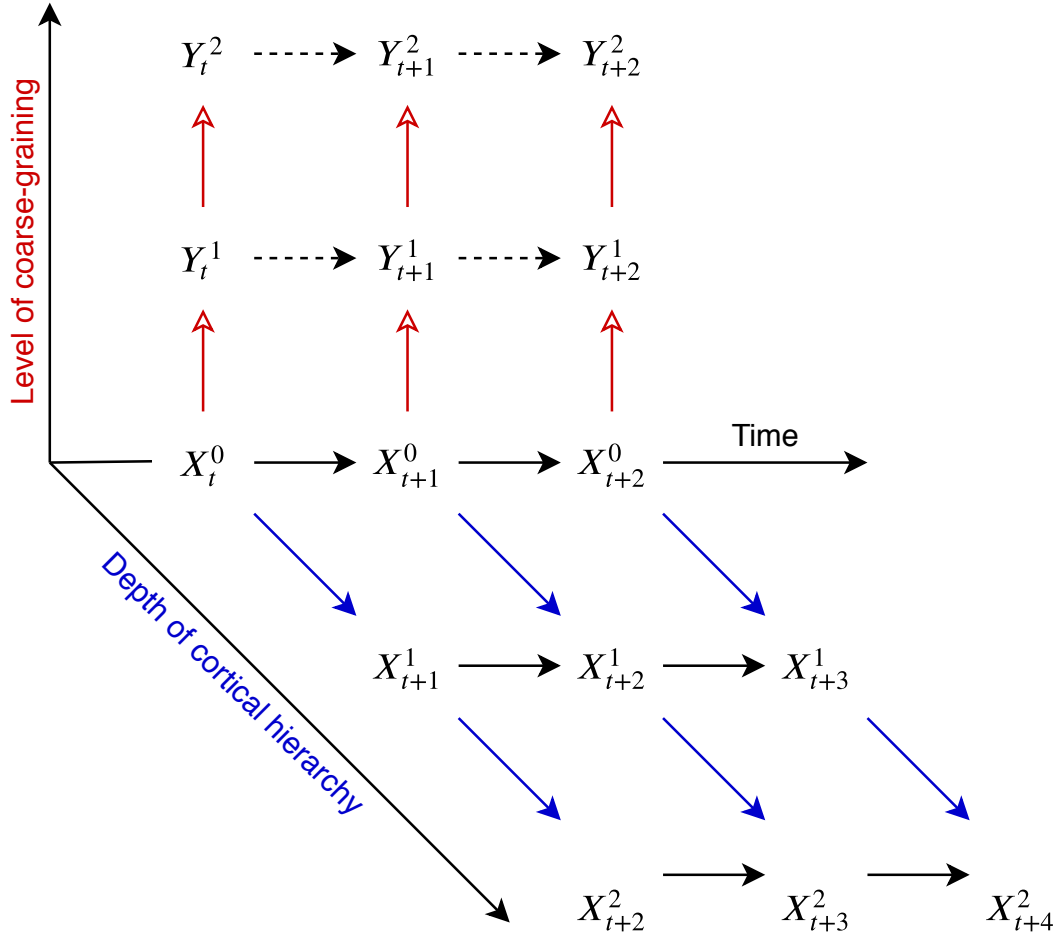


Figure 6: The distinction between the level of coarse-graining and the level of the cortical hierarchy. X and Y represent the microscopic and macroscopic coarse-grained variables, respectively. X^0 represents the microscopic states at the upstream of the cortical hierarchy. The red empty arrows represents the directions of coarse-graining and the blue arrows represent the directions of the physical dependencies in the cortical hierarchy from the upstream to the downstream. (Some variables and dependencies are omitted for clarity.)

6.2 Integrated Information Theory

Integrated information theory (IIT) states that consciousness is integrated information and system's consciousness is determined by its causal properties (Tononi *et al.*, 2016). IIT is in line with IIT in that informational properties are thought to underlie consciousness. In this section, we will discuss IIT in the light of IIT.

The concept of "information": In IIT, information refers to "integrated information": "Information that is specified by a system that is irreducible to that specified by its parts." (Tononi *et al.* 2016) In ICT, information refers to "self-information", i.e. information about the states of conscious experience and the physical states of a process. Therefore, IIT focuses more on the relationships between consciousness and causal interactions among elements within a system, whereas ICT focuses more on the informational relationships between conscious experience and being in a certain state of a process.

The "Exclusion" axiom in IIT: In IIT, the Exclusion axiom claims that of all overlapping sets of elements, only one set with maximal integrated information can be conscious. The exclusion axiom should be applied over elements, space, time, and scales (Hoel *et al.*, 2016; Oizumi *et al.*, 2014). Different from IIT, ICT allows multiple consciousnesses coexist across different levels of coarse-graining within a system if they are informationally closed from each other. The two distinctive predictions decisively pinpoint the core concepts of the two theories.

The concept of "integration": In IIT, integrated information is one of the core concept of defining conscious individuals. In the current paper, we did not include the notion of integrated information within ICT. This, however, results in one of the current weaknesses of ICT that it lacks the ability to individuate NTIC processes in some extreme cases (i.e., the problem of individuality). We discussed this weaknesses in Sec. 7.

Prediction after system damaged: ICT and IIT lead to different predictions when a system suffers from a damage. For example, considering a densely connected networks whose dynamics forms an NTIC process. If we cut the network in half, IIT would predict that this results in two consciousnesses because elements in both networks still maintain high degrees of interactions. In contrast, ICT would predict that this operation could completely destroy NTIC rendering both parts unconscious.

6.3 Predictive Processing

Predictive processing (PP) is a powerful framework which integrates several ideas from neuroscience. This emerging theoretical framework posits that neural systems constantly generate predictions about incoming sensory signals and updates predictions based on prediction errors between predictions and sensory signals. According to PP, neural systems constantly perform unconscious statistical inference about hidden causes in the external environment. The perceptual contents are the "best guess" about the environment states including these hidden causes (Clark, 2013; Hohwy, 2013). PP is well integrated with Bayesian brain hypothesis and has been used to interpret conscious perception in many domains (Hohwy, 2013; Seth, 2014).

PP is a powerful explanatory framework for diverse brain functions. However, to serve as a theory of consciousness, PP is still incomplete due to two explanatory gaps. First, it has been known that the neural system is equipped multiple predictive mechanisms. Apparently, not all the predictive mechanisms are involved in conscious processes (e.g. mismatch negativity, Näätänen et al. (2007)). PP needs to explain the difference between conscious and unconscious predictive mechanisms.

Second, PP can be considered as a sophisticated computation for perceptual inference. It takes von Helmholtz's conception of perception as unconscious inference. Thus, only the most probable outcome computed by the inference processes can be conscious while other details of the computation remain unconscious. PP also needs to explain how unconscious inferences is able to give rise to conscious results. In short, while PP is often discussed in the context of consciousness, these explanatory gaps prevent PP from being a theory of consciousness.

ICT is well compatible with PP. Crucially, ICT further provides natural and fundamental explanations to fills the two explanatory gaps which PP encounters. According to the definition of NTIC, a process with high NTIC can be regarded as a powerful predictive machine which has accurate self-predictive information ($I(Y_{t+1}; Y_t)$, E.q. 6) and concurrently incorporates environmental information into its dynamic ($I(Y_{t+1}; Y_t|E_t)$, E.q. 6). This predictive nature of NTIC processes is in agreement with the core notion of PP in which the conscious contents are always the predicted (inferred) outcome of our predictive mechanisms. Second, due to the informational closure to the environment, the encoded information about its environment in an NTIC process can be seemed as "the best guess" about the external environment in the context of Bayesian inference.

So, eventually, why are some predictive information conscious and some are not? ICT predicts that only the predictions generated from mechanisms involving the NTIC process are conscious. Note that predictive processes are not necessary to involve NTIC processes. A predictive process can make prediction about the future state of its environment based on the current sensory inputs. In this case, the the process is not informationally closed and could not be conscious.

According to ICT, we further propose that we can only be aware of the predictions from predictive processes due to informational closure to computational details of microscopic predictive processes. The macroscopic NTIC process only acquires the coarse-grained summary statistics of the microscopic processes. In other words, we predict that the computation of the statistical inferences of PP is implemented at microscopic (cellular) levels in the neural system.

Finally, we consider PP as an potential empirical implementation of NTIC processes. To maintain accurate information about the environment encoded in an NTIC process, one can open an information channel between the process and the environment for minimal information flow to correct the divergence between them. This proposal is compatible with PP which suggests that PP systems updates (corrects) the current estimations by computing prediction errors between

489 predicted and real sensory inputs.

490 6.4 Sensorimotor Contingency

491 Sensorimotor contingency (SMC) theory of consciousness proposes that different types of SMCs give
492 rise to different characteristics of conscious experience (O'Regan & Noë, 2001). The theory radically
493 rejects the view that conscious content is associated with internal representations of a system.
494 Rather, the quality of conscious experience depends on agents' mastery of SMCs. SMC emphasises
495 that the interaction between a system and its environment determines conscious experience.

496 ICT is not compatible with SMC. As mentioned in Sec. 5, a process directly maps the sensory
497 states to the action states is insufficient to be NTIC. Therefore, learning contingencies between
498 sensory inputs and action outputs does not imply NTIC. Hence, ICT predicts that having senso-
499 rimotor contingencies is neither a necessary nor a sufficient condition for consciousness. In fact,
500 empirically, with extensive training on a sensorimotor task with a fixed contingency, the task can be
501 gradually performed unconsciously. This indicates that strong SMCs do not contribute conscious
502 contents. In contrast, ICT suggests that, with extensive training, the neural system establishes a
503 neural mapping from sensory inputs to action outputs. This decrease the level of informational clo-
504 sure and, as a result, decrease the conscious level of this process. This outcome strongly supports
505 ICT than SMC.

506 Nevertheless, ICT does appreciate the notion that interactions between a process and its envi-
507 ronment is crucial to shape conscious experience. As mentioned above, to form NTIC, a process
508 needs to encode environmental transitions into its own dynamic. Therefore, information of agent-
509 environment interaction should also be encoded in the NTIC process, and therefore, shape conscious
510 contents in a specific way.

511 Different from the classical SMC, a new version of SMC, Predictive Processing of SensoriMo-
512 tor Contingencies (PPSMC), proposed by Seth (2014, 2015) combines SMC and the predictive
513 processing framework together. PPSMC emphasises the important role of generative models in
514 computing counterfactuals, inferring hidden causes of sensory signals, and linking fictive sensory
515 signals to possible actions. According to ICT, if the generative model involving the NTIC process
516 for the computation of counterfactuals, PPSMC will be compatible to our theory and may have
517 strong explanatory power on some specific conscious experience.

518 6.5 Global Workspace Theory

519 Global workspace theory (GWT; Baars (1988, 1997, 2002)) or Global Neuronal Workspace theory
520 (GNWT; Dehaene & Changeux (2011); Dehaene & Naccache (2001); Dehaene *et al.* (1998)) states
521 that the neural system consists of several specialised modules and a central global workspace (GW)
522 which integrates and broadcasts information gathered from those specialised modules. Only the
523 information in the global workspace reaches conscious awareness, and information outside of it
524 remains unconscious. These modules compete with each other to gain the access to the GW and
525 the information from the winner triggers an all-or-none "ignition" in the GW. Information in the
526 GW is broadcasted to other modules. Conscious contents then are associates with the information
527 that gains access to the internal global workspace Dehaene *et al.* (2017).

528 While GWT emphasises the importance of global information sharing as a basis of conscious-
529 ness, the precise meaning of information broadcasting has been somewhat unclear if one tries to
530 describe it more formally in the language of information theory. ICT offers one possible way to con-
531 sider the meaning of broadcasting in GWT. Specifically, one could interpret the global workspace
532 as the network of nodes where information is shared at the scale of NTIC where communication
533 is performed through macro-variables that are linked via mutual predictability. That is, global
534 workspace should be also NTIC. While this link remains speculative at this point, this interpreta-
535 tion encourages empirical studies into the relationship between the contents of consciousness and
536 macrostate neural activities that are mutually predictive of each other.

537 7 Limitation and Future Work

538 As a brand new theory of consciousness, ICT is still far from completion. In the following, we
539 discuss the current limitations and challenges of ICT and point out the potential future research
540 directions.

It's important to clarify that ICT does not intend to completely solve the hard problems of consciousness (Chalmers, 1995). Knowing the state of a conscious process does not allow us to answer "What is it like to be in this state of this process" (Nagel, 1974). Instead, ICT focuses more on bridging consciousness and the physical world using information theory as a common language in between.

The current version of ICT cannot entirely solve the problem of individuality in some extreme circumstances. In common cases, one can identify individual consciousnesses by computing the levels of NTIC of a process. This approach can also be applied to finding the boundaries of individual consciousnesses (for details of the boundary detection procedure, see Krakauer *et al.* (2014)). However, in some specific circumstances, individuality of consciousness is not clear. For instance, we can define a new process Y and also its environment E by recruiting two independent NTIC processes Y^1 & Y^2 and their environments E^1 & E^2 , respectively. So that $Y = \{Y^1, Y^2\}$ and $E = \{E^1, E^2\}$. In such case, the new process Y will also be NTIC to E . Therefore, the current version of ICT cannot determine whether there are two smaller consciousnesses or one bigger consciousness (or 3 coexisting consciousnesses). The problem of individuality is a significant theoretical weakness of the current version of ICT. The notion of integration is a possible remedy for this issue and we will address this issue more explicitly in our future work using the concept of synergy.

The current version of ICT assume that consciousness is only contributed by non-trivial rather than trivial information encoded in a process. In other words, how much information about environmental states and dynamics encoded in a process is a key quantity for consciousness. However, we do not exclude the possibility that environmental information may be just a proxy of other informational quantities. More theoretical work is needed to elucidate the role of environments. This issue will also be discuss in our next theoretical paper.

Empirically, a major challenge to ICT is to find proper coarse-graining functions which map microscopic processes to macroscopic NTIC processes. This will become an imperative issue of finding neurological supporting evidence for ICT. To find proper coarse-graining functions among infinite candidates (Price & Corry, 2007) seem to be very challenging. Nevertheless, there are still theoretical and technical progresses recently that may contribute to solving this issue. For example, the concept of *causal emergence* proposed by Hoel (Hoel, 2018; Hoel *et al.*, 2013) has been further developed recently. Causal emergence is highly relevant to the relationship between informational closure and coarse-graining. In their new study by Klein & Hoel (2019), they started to compare how different coarse-graining functions influence causal emergence at macroscopic levels. Pfante *et al.* (2014a,b) provides thorough mathematical analyses on level identification including informational closure. In neuroscience, the understanding of neural population codes also achieves a tremendous progress due to the advancement of recording technique and data science (Kohn *et al.*, 2016; Panzeri *et al.*, 2015). Gamez (2016) has also systematically described relevant issues in terms of finding data correlates of consciousness amount different levels of abstraction. We believe that interdisciplinary research is required to narrow down the scope of searching the coarse-graining functions and conscious processes at the coarse-grained levels in the neural system and beyond.

In this article, we do not use a state-dependent formulation of NTIC in the current version of ICT. However, we believe that the state-dependent NTIC is essential to describe the dynamics of conscious experience. Therefore, further research using point-wise informational measures to construct state-dependent NTIC is needed in the next version of ICT.

Finally another empirical challenge to ICT is the of empirical supporting evidence. This is understandable because the concept of NTIC is relatively new in the history of information science, not to mention in neuroscience. Very few experiments and data collections are designed for examining NTIC properties in neural systems. To our knowledge, only two studies (Palmer *et al.*, 2015; Sederberg *et al.*, 2018) coincidentally examine relevant properties in salamander retina. They found that the a large group of neural populations of retinal ganglion cells encoded predictive information about external stimuli also had high self-predictive information about their own future states. This result is in line with the characteristic of NTIC. We expect that there will be more empirical studies examining relevant neural properties of NTIC.

595 8 Conclusions

596 In this paper, we introduced **Information Closure Theory of Consciousness (ICT)**, a new
597 informational theory of consciousness. ICT proposes that a process which forms **non-trivial**
598 **informational closure (NTIC)** is conscious and through coarse-graining the neural system can
599 form NTIC processes, i.e., conscious processes, at a certain macroscopic level. ICT considers that
600 information is a common language to bridge the gap between conscious experience and the physical
601 reality. Using information theory, ICT proposes computational definitions for both conscious level
602 and conscious content. This makes ICT be able to generalise to any system beyond the human
603 brains.

604 ICT provides explanation for various findings from research of conscious and unconscious pro-
605 cessing. The implications of ICT point out that the levels of coarse-graining play a critical role
606 in searching for neural substrates of consciousness. Improper measurements, e.g., too fine or too
607 coarse in terms of the scale of measurements, of neurophysiological signals may lead to misleading
608 results and misinterpretations.

609 ICT reconciles several theories of consciousness. ICT indicates that they conditionally coincide
610 with ICT’s implications and predictions but, however, not the fundamental and sufficient condi-
611 tions for consciousness. For example, theories includes the theories emphasising recurrent circuits
612 (Edelman, 1992; Lamme, 2006), the theories highlighting the internal simulation, predictive mech-
613 anism, and generative models (Clark, 2013; Hohwy, 2013; Kanai *et al.*, 2019; Revonsuo, 2006;
614 Seth, 2014, 2015), and theories related to multilevel view of consciousness (Jackendoff, 1987; Pen-
615 nartz, 2015, 2018; Prinz, 2007). Notably, ICT is proposed based on the non-functional hypothesis.
616 Notwithstanding, its implications for the functional aspects of a system well fit several functionalist
617 proposals.

618 Regarding philosophy of mind, ICT connects several distinct arguments together. ICT can be
619 seen as an identity theory because it assumes a fundamental relation between consciousness and
620 information. Second, the implications of ICT tightly link consciousness to several cognitive func-
621 tions in the context of evolution. This explains why people might intuitively have a functionalist
622 point of view of consciousness. ICT emphasises that informational closure between levels of coarse-
623 graining is critical to form NTIC processes in some stochastic systems. In this case, especially for
624 the neural system, forming conscious processes at the macroscopic levels coincide with the per-
625 spective of emergentism. Finally, forming NTIC (conscious) processes through many-to-one maps,
626 i.e., coarse-graining, implies multiple realisability of consciousness. As a result, ICT provides an
627 integrated view for these arguments and is further capable of indicating how and why they are
628 conditionally true.

629 So far, the current version of ICT is still far from completion. Further theoretical and empirical
630 research is indispensably required for ICT to improve and solve several issues in the current version.
631 Nevertheless, ICT offers explanation and prediction for consciousness science. We hope that ICT
632 provides a new way of thinking and understanding neural substrates of consciousness.

633 Acknowledgements

634 A.C., Y.Y, and R.K. are funded by Japan Science and Technology Agency (JST) CREST project.
635 Work by M.B. and R.K. on this publication was made possible through the support of a grant from
636 Templeton World Charity Foundation, Inc. The opinions expressed in this publication are those
637 of the authors and do not necessarily reflect the views of Templeton World Charity Foundation,
638 Inc. This manuscript has been released as a Pre-Print at arXiv (Chang *et al.*, 2019).

639 Author Contributions Statement

640 A.C. conceived and developed the theory. M.B. and A.C. contributed the mathematical formalisa-
641 tion of the theory. A.C., M.B, and R.K wrote the manuscript, based on a first draft by A.C. with
642 extensive comments from Y.Y. All authors contributed to manuscript revision, read and approved
643 the submitted version.

644 Conflict of Interest Statement

645 All authors were employed by the company Araya Inc. The authors declare that the research
646 was conducted in the absence of any commercial or financial relationships that could be construed
647 as a potential conflict of interest.

648 References

- 649 Adrian, Edgar D. 1926. The impulses produced by sensory nerve endings. *The Journal of physiology*,
650 **61**(1), 49–72.
- 651 Ashby, F Gregory, Turner, Benjamin O, & Horvitz, Jon C. 2010. Cortical and basal ganglia
652 contributions to habit learning and automaticity. *Trends in cognitive sciences*, **14**(5), 208–215.
- 653 Baars, Bernard J. 1988. *A cognitive theory of consciousness*. Cambridge University Press.
- 654 Baars, Bernard J. 1997. In the theatre of consciousness. Global workspace theory, a rigorous
655 scientific theory of consciousness. *Journal of Consciousness Studies*, **4**(4), 292–309.
- 656 Baars, Bernard J. 2002. The conscious access hypothesis: origins and recent evidence. *Trends in*
657 *cognitive sciences*, **6**(1), 47–52.
- 658 Bayne, Tim, Hohwy, Jakob, & Owen, Adrian M. 2016. Are there levels of consciousness? *Trends*
659 *in cognitive sciences*, **20**(6), 405–413.
- 660 Bechtel, William, & Mundale, Jennifer. 1999. Multiple realizability revisited: Linking cognitive
661 and neural states. *Philosophy of science*, **66**(2), 175–207.
- 662 Bertschinger, Nils, Olbrich, Eckehard, Ay, Nihat, & Jost, Jürgen. 2006. Information and closure
663 in systems theory. *Pages 9–21 of: Explorations in the Complexity of Possible Life. Proceedings*
664 *of the 7th German Workshop of Artificial Life*.
- 665 Bialek, William, Nemenman, Ilya, & Tishby, Naftali. 2001. Predictability, complexity, and learning.
666 *Neural computation*, **13**(11), 2409–2463.
- 667 Binder, Marc D, Hirokawa, Nobutaka, & Windhorst, Uwe. 2009. Encyclopedia of neuroscience.
- 668 Casali, Adenauer G, Gosseries, Olivia, Rosanova, Mario, Boly, Mélanie, Sarasso, Simone, Casali,
669 Karina R, Casarotto, Silvia, Bruno, Marie-Aurélié, Laureys, Steven, Tononi, Giulio, *et al.* 2013.
670 A theoretically based index of consciousness independent of sensory processing and behavior.
671 *Science translational medicine*, **5**(198), 198ra105–198ra105.
- 672 Chalmers, David J. 1995. Facing up to the problem of consciousness. *Journal of consciousness*
673 *studies*, **2**(3), 200–219.
- 674 Chang, Acer Y. C., Biehl, Martin, Yu, Yen, & Kanai, Ryota. 2019. *Information Closure Theory of*
675 *Consciousness*.
- 676 Clark, Andy. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive
677 science. *Behavioral and Brain Sciences*, **36**(3), 181–204.
- 678 Dehaene, Stanislas, & Changeux, Jean-Pierre. 2011. Experimental and theoretical approaches to
679 conscious processing. *Neuron*, **70**(2), 200–227.
- 680 Dehaene, Stanislas, & Naccache, Lionel. 2001. Towards a cognitive neuroscience of consciousness:
681 basic evidence and a workspace framework. *Cognition*, **79**(1-2), 1–37.
- 682 Dehaene, Stanislas, Kerszberg, Michel, & Changeux, Jean-Pierre. 1998. A neuronal model of a
683 global workspace in effortful cognitive tasks. *Proceedings of the national Academy of Sciences*,
684 **95**(24), 14529–14534.
- 685 Dehaene, Stanislas, Lau, Hakwan, & Kouider, Sid. 2017. What is consciousness, and could machines
686 have it?

687 Doyon, Julien, Bellec, Pierre, Amsel, Rhonda, Penhune, Virginia, Monchi, Oury, Carrier, Julie,
688 Lehericy, Stéphane, & Benali, Habib. 2009. Contributions of the basal ganglia and functionally
689 related brain structures to motor learning. *Behavioural brain research*, **199**(1), 61–75.

690 Edelman, Gerald M. 1992. *Bright air, brilliant fire: On the matter of the mind*. Basic books.

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

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

695 Gamez, David. 2016. Are Information or Data Patterns Correlated with Consciousness? *Topoi*,
696 **35**(1), 225–239.

697 Gerstner, Wulfram, & Kistler, Werner M. 2002. *Spiking neuron models: Single neurons, popula-*
698 *tions, plasticity*. Cambridge university press.

699 Goldwyn, Joshua H., & Shea-Brown, Eric. 2011. The what and where of adding channel noise to
700 the Hodgkin-Huxley equations. *PLoS Computational Biology*, **7**(11).

701 Himmelbach, Marc, & Karnath, Hans-Otto. 2005. Dorsal and ventral stream interaction: contri-
702 butions from optic ataxia. *Journal of Cognitive Neuroscience*, **17**(4), 632–640.

703 Hoel, Erik P. 2018. *Agent Above, Atom Below: How Agents Causally Emerge from Their Underlying*
704 *Microphysics*. Cham: Springer International Publishing. Pages 63–76.

705 Hoel, Erik P., Albantakis, Larissa, & Tononi, Giulio. 2013. Quantifying causal emergence shows
706 that macro can beat micro. *Proceedings of the National Academy of Sciences*, **110**(49), 19790–
707 19795.

708 Hoel, Erik P, Albantakis, Larissa, Marshall, William, & Tononi, Giulio. 2016. Can the macro beat
709 the micro? Integrated information across spatiotemporal scales. *Neuroscience of Consciousness*,
710 **2016**(1), niw012.

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

712 Humphrey, N. K. 1970. What the Frog’s Eye Tells the Monkey’s Brain. *Brain, Behavior and*
713 *Evolution*, **3**(1-4), 324–337.

714 Humphrey, Nicholas. 1999. *A History of the Mind: Evolution and the Birth of Consciousness*.
715 Springer Science & Business Media.

716 Humphrey, Nicholas K. 1974. Vision in a monkey without striate cortex: a case study. *Perception*,
717 **3**(3), 241–255.

718 Jackendoff, Ray. 1987. *Consciousness and the computational mind*. The MIT Press.

719 James, Thomas W, Culham, Jody, Humphrey, G Keith, Milner, A David, & Goodale, Melvyn A.
720 2003. Ventral occipital lesions impair object recognition but not object-directed grasping: an
721 fMRI study. *Brain*, **126**(11), 2463–2475.

722 Kanai, Ryota, Chang, Acer, Yu, Yen, de Abril, Ildefons M, Biehl, Martin, & Guttenberg, Nicholas.
723 2019 (Jul). *Information Generation as a Functional Basis of Consciousness*.

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

727 Klein, Brennan, & Hoel, Erik. 2019. Uncertainty and causal emergence in complex networks. *arXiv*
728 *preprint arXiv:1907.03902*.

729 Kohn, Adam, Coen-Cagli, Ruben, Kanitscheider, Ingmar, & Pouget, Alexandre. 2016. Correlations
730 and Neuronal Population Information. *Annual Review of Neuroscience*, **39**(1), 237–256.

- 731 Krakauer, David, Bertschinger, Nils, Olbrich, Eckehard, Ay, Nihat, & Flack, Jessica C. 2014. The
732 information theory of individuality. *arXiv preprint arXiv:1412.2447*.
- 733 Kristan Jr, William B, & Shaw, Brian K. 1997. Population coding and behavioral choice. *Current*
734 *opinion in neurobiology*, **7**(6), 826–831.
- 735 Lamme, Victor AF. 2006. Towards a true neural stance on consciousness. *Trends in cognitive*
736 *sciences*, **10**(11), 494–501.
- 737 Laureys, Steven. 2005. The neural correlate of (un) awareness: lessons from the vegetative state.
738 *Trends in cognitive sciences*, **9**(12), 556–559.
- 739 Lemon, RN, & Edgley, SA. 2010. Life without a cerebellum. *Brain*, **133**(3), 652–654.
- 740 Luhmann, Niklas. 1995. Probleme mit operativer Schließung. *Soziologische Aufklärung*, **6**, 12–24.
- 741 Maass, Wolfgang, & Bishop, Christopher M. 2001. *Pulsed neural networks*. MIT press.
- 742 Marr, David. 1982. Vision: A Computational Investigation into the Human Representation and
743 Processing of Visual Information. *New York, NY: W.H. Freeman and Company*, **8**(11).
- 744 Mathis, Donald W, & Mozer, Michael C. 1995. On the computational utility of consciousness.
745 *Pages 11–18 of: Advances in neural information processing systems*.
- 746 Maturana, Humberto R, & Varela, Francisco J. 1991. *Autopoiesis and cognition: The realization*
747 *of the living*. Vol. 42. Springer Science & Business Media.
- 748 Mazzi, Chiara, Bagattini, Chiara, & Savazzi, Silvia. 2016. Blind-sight vs. degraded-sight: different
749 measures tell a different story. *Frontiers in psychology*, **7**, 901.
- 750 Milner, AD, Paulignan, Y, Dijkerman, HC, Michel, F, & Jeannerod, M. 1999. A paradoxical
751 improvement of misreaching in optic ataxia: new evidence for two separate neural systems for
752 visual localization. *Proceedings of the Royal Society of London. Series B: Biological Sciences*,
753 **266**(1434), 2225–2229.
- 754 Näätänen, Risto, Paavilainen, Petri, Rinne, Teemu, & Alho, Kimmo. 2007. The mismatch negativ-
755 ity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology*,
756 **118**(12), 2544–2590.
- 757 Nagel, Thomas. 1974. What is it like to be a bat? *The philosophical review*, **83**(4), 435–450.
- 758 Oizumi, Masafumi, Albantakis, Larissa, & Tononi, Giulio. 2014. From the phenomenology to the
759 mechanisms of consciousness: integrated information theory 3.0. *PLoS computational biology*,
760 **10**(5), e1003588.
- 761 O’Regan, J Kevin, & Noë, Alva. 2001. A sensorimotor account of vision and visual consciousness.
762 *Behavioral and brain sciences*, **24**(5), 939–973.
- 763 Overgaard, Morten. 2011. Visual experience and blindsight: a methodological review. *Experimental*
764 *Brain Research*, **209**(4), 473–479.
- 765 Overgaard, Morten, & Overgaard, Rikke. 2010. Neural correlates of contents and levels of con-
766 sciousness. *Frontiers in psychology*, **1**, 164.
- 767 Palmer, Stephanie E., Marre, Olivier, Berry, Michael J., & Bialek, William. 2015. Predictive
768 information in a sensory population. *Proceedings of the National Academy of Sciences*, **112**(22),
769 6908–6913.
- 770 Panzeri, Stefano, Macke, Jakob H, Gross, Joachim, & Kayser, Christoph. 2015. Neural population
771 coding: combining insights from microscopic and mass signals. *Trends in cognitive sciences*,
772 **19**(3), 162–172.
- 773 Pattee, Howard Hunt. 2012. Evolving self-reference: matter, symbols, and semantic closure. *Pages*
774 *211–226 of: Laws, language and life*. Springer.

- 775 Pennartz, Cyriel MA. 2015. *The brain's representational power: on consciousness and the integra-*
776 *tion of modalities*. MIT Press.
- 777 Pennartz, Cyriel MA. 2018. Consciousness, representation, action: the importance of being goal-
778 directed. *Trends in cognitive sciences*, **22**(2), 137–153.
- 779 Pfante, Oliver, Olbrich, Eckehard, Bertschinger, Nils, Ay, Nihat, & Jost, Jürgen. 2014a. Closure
780 measures for coarse-graining of the tent map. *Chaos: An Interdisciplinary Journal of Nonlinear*
781 *Science*, **24**(1), 013136.
- 782 Pfante, Oliver, Bertschinger, Nils, Olbrich, Eckehard, Ay, Nihat, & Jost, Jürgen. 2014b. Compar-
783 ison between different methods of level identification. *Advances in Complex Systems*, **17**(02),
784 1450007.
- 785 Pouget, Alexandre, Dayan, Peter, & Zemel, Richard. 2000. Information processing with population
786 codes. *Nature Reviews Neuroscience*, **1**(2), 125.
- 787 Price, Huw, & Corry, Richard. 2007. *Causation, physics, and the constitution of reality: Russell's*
788 *republic revisited*. Oxford University Press.
- 789 Prinz, Jesse. 2007. The intermediate level theory of consciousness. *The Blackwell companion to*
790 *consciousness*, 257–271.
- 791 Putnam, Hilary. 1967. Psychological predicates. *Art, mind, and religion*, **1**, 37–48.
- 792 Quiñ Quiroga, Rodrigo, & Panzeri, Stefano. 2009. Extracting information from neuronal pop-
793 ulations: Information theory and decoding approaches. *Nature Reviews Neuroscience*, **10**(3),
794 173–185.
- 795 Raymont, Paul, & Brook, Andy. 2006. Unity of Consciousness. *Pages 565–577 of: Beckermann,*
796 *Ansgar, & McLaughlin, Brian P. (eds), The Oxford Handbook of Philosophy of Mind*. Oxford
797 University Press.
- 798 Revol, P, Rossetti, Y, Vighetto, A, Rode, G, Boisson, D, & Pisella, L. 2003. Pointing errors in
799 immediate and delayed conditions in unilateral optic ataxia. *Spatial vision*, **16**(3-4), 347–364.
- 800 Revonsuo, Antti. 2006. *Inner presence: Consciousness as a biological phenomenon*. Mit Press.
- 801 Rosen, Robert. 1991. *Life itself: a comprehensive inquiry into the nature, origin, and fabrication*
802 *of life*. Columbia University Press.
- 803 Sederberg, Audrey J, MacLean, Jason N, & Palmer, Stephanie E. 2018. Learning to make external
804 sensory stimulus predictions using internal correlations in populations of neurons. *Proceedings*
805 *of the National Academy of Sciences*, 201710779.
- 806 Seth, Anil K. 2014. A predictive processing theory of sensorimotor contingencies: Explaining
807 the puzzle of perceptual presence and its absence in synesthesia. *Cognitive neuroscience*, **5**(2),
808 97–118.
- 809 Seth, Anil K. 2015. Presence, objecthood, and the phenomenology of predictive perception. *Cog-*
810 *nitive neuroscience*, **6**(2-3), 111–117.
- 811 Stein, Richard B, Gossen, E Roderich, & Jones, Kelvin E. 2005. Neuronal variability: noise or
812 part of the signal? *Nature Reviews Neuroscience*, **6**(5), 389.
- 813 Tononi, Giulio, & Koch, Christof. 2008. The neural correlates of consciousness: an update. *Annals*
814 *of the New York Academy of Sciences*, **1124**(1), 239–261.
- 815 Tononi, Giulio, Boly, Melanie, Massimini, Marcello, & Koch, Christof. 2016. Integrated information
816 theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, **17**(7), 450.
- 817 White, John A., Rubinstein, Jay T., & Kay, Alan R. 2000. Channel noise in neurons. *Trends in*
818 *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.
- 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.

Figure Legends

Figure 1:	The scale problem of consciousness: Human conscious experience does not reflect information from every scale. Only information at a certain coarse-grained scale in the neural system is reflected in consciousness.	4
Figure 2:	The dependencies between a system and its environment.	5
Figure 3:	The information flow amounts the universe X , the system S , the environment of the system E , and the coarse-grained process Y of the system S . The solid line with a filled arrow from X_t to X_{t+1} represents the microscopic dynamic of the universe. The solid lines with a empty arrow represent directions of coarse-graining. The dashed lines represents virtual dependencies between two macroscopic variables. The red Y_t , Y_{t+1} , and the red dashed line in between represents a macroscopic process which forms informational closure at a certain coarse-grained level.	7
Figure 4:	A non-monotonic relationship between Level of coarse-graining and level of consciousness.	9
Figure 5:	A diagram depicting the information flow in reflexive behaviours (shown by the red nodes and arrows) happening through the interaction between a process Y and its environment E . In such situations, the internal state Y_t is mostly dependent on the environment state E_{t-1} but less on its past state Y_{t-1} . Therefore, the process Y is not informational closed. As a consequence, Y is unable to form high NTIC and, therefore, remain less conscious or unconscious.	11
Figure 6:	The distinction between the level of coarse-graining and the level of the cortical hierarchy. X and Y represent the microscopic and macroscopic coarse-grained variables, respectively. X^0 represents the microscopic states at the upstream of the cortical hierarchy. The red empty arrows represents the directions of coarse-graining and the blue arrows represent the directions of the physical dependencies in the cortical hierarchy from the upstream to the downstream. (Some variables and dependencies are omitted for clarity.)	14