Response to Reviewers Rev^1

April 3, 2020

Reply to Editor

Q1: In addition, I have a question. if we view molecules as coarse grains of elementary particles, would the phenomenon of molecular recognition have high NTIC and thus be conscious?

A: Hi Prof. Hoffman, Thank you for your attention to our manuscript and your interesting question. If we understand correctly, molecular recognition refers to specific types of interactions among molecules to form molecular complexes. There is no question that molecular complexes are coarse-grainings of molecules via molecular recognition. Based on our hypothesis, if the dynamics of a molecular complex is an NTIC process, the information closure theory of consciousness (ICT) would claim that the molecular complex has a certain degree of consciousness. Therefore, to your question, it is theoretically plausible under the framework of ICT. We speculate that it may be easy to find informationally closed molecular complexes (isolated complexes) but may not be easy to find ones forming NTIC processes. Nevertheless, with well-defined mathematical definition, this is an empirical and testable question to ICT. We look forward to learning any self-assembly process forming NTIC at molecular scales in the near future.

Reply to Reviewer 1

Q1: Regarding the NTCI measure:

a) Bertschinger et al. 2006 define informationally closed as $J_t(E- > Y) = 0$, but this is not a requirement for NTIC. Meaning, a positive NTIC measure does not entail that $J_t = 0$. $J_t = 0$ is only true if the NTIC is equal to $I(Y_{t+1}; E_t)$. In what sense does it correspond to non-trivial information closure then? (see also my concern #2, maybe that is related) As defined in eq. 6 a video camera might have very high NTIC if it records a natural scene, even though it is in no way informationally closed from the environment.

b) Is $NTIC_t$ state dependent or not? As capital letters are used I assume it is based on (conditional) mutual information measures across one time step but averaged over the possible states of the system at time t and t+1, so state distributions. But which distributions are used to compute $NTIC_t$? The stationary joint distribution of states at t and t+1? It cannot be just some observed measure, because then the level of consciousness would depend on how long the observation is. Moreover, if $NTIC_t$ is not state dependent, then the Reflex example on p. 10 is not intuitive. Because the system (brain) is the same and awake whether it behaves reflexively or not, so while the content may depend on the type of action, the $NTIC_t$ (level) would stay the same.

A:
a) Considering Eq.7, it is true that the transfer entropy term $J_t(E- > Y) = 0$ is not required for $NTIC > 0$. It is possible to have a high degree of NTIC when $J_t(E- > Y) > 0$, i.e. the process is not completely closed.
However, based on the setting in both Bertschinger et al. 2006 and our Section 2 (at Line 99), Y_{t+1} is determined by both Y_t and E_t. Simply increasing $I(Y_{t+1}; E_t)$ without closedness (e.g, copy the information in sensory signals to the process such as the case in the video camera example), the transfer entropy term $J_t(E- > Y)$ should increase at the same rate and cancel out the increment of

$I(Y_{t+1}; E_t)$ resulting in low NTIC. Namely, keeping high closedness, i.e. minimising $J_t(E->Y)$, is crucial to increase NTIC even though $J_t(E->Y) > 0$.
In our current version of ICT, we consider the level of consciousness corresponds to the degree of NTIC of a process. Therefore, a conscious process does not have to be completely closed. However, a high degree of closedness is crucial to have a high level of consciousness. (Regarding to the video camera example, please see more discussion in our reply to your Q6.)

b) We thank the reviewer for this question. We entirely agree that state-dependent measurements are essential to describe the dynamics of conscious experience. In fact, we now are working on state-dependent formulations for the next version of the ICT. In contrast to the current version, this work involves point-wise informational measures and certainly require more space and discussion. Therefore, we prefer to address the state-dependent version of ICT in future studies and keep the first version of our theory simple.
We appreciated that the reviewer pointed this out. We have added a short paragraph about this issue.

> **Line 588:**
> *" In this article, we do not use a state-dependent formulation of NTIC. 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."*

Q2: Hypothesis and Implication 3: Why is it crucial that there is an underlying X with respect to which Y is a C-process?

A: It is crucial to have an underlying X with respect to Y because this assumption can solve the scale problem of consciousness. It ensures that the information transfer from all coarse-grainings of X to Y vanishes. Therefore, human conscious experiences do not involve information from other scales of coarse-graining. Because X is the micro-level of the universe, the information transfer from any coarse-graining (including E and S) of X to Y must be lower than that from X to Y, i.e. $I(Y_{t+1}; X_t|Y_t) \geq I(Y_{t+1}; S_t|Y_t)$ and $I(Y_{t+1}; X_t|Y_t) \geq I(Y_{t+1}; E_t|Y_t)$. So that Y is informationally closed with respect to X implies that Y is also informationally closed with respect to S, E, and all other coarse-grainings of X.

Q3: Does Y being informationally closed with respect to X imply that Y is also informationally closed with respect to E, i.e. $J_t(E - Y) = 0$? I don't think that could be true for the brain, or part of it really. My brain's next state, at whatever level, certainly depends to some degree on unpredictable sensory inputs from the environment. In the end, is $I(Y_{t+1}; E_t|Y_t) = 0$ required for consciousness or not? (see point 1a). In other points in the manuscript this does not seem required, e.g. p. 10 "If a process is not informationally closed, the degree of NTIC is low resulting in low or no consciousness".

A: Yes. Y being informationally closed with respect to X implies that Y is also informationally closed with respect to E, i.e. $J_t(E - Y) = 0$ (Please see our reply to Q2).
As explained in Question 1a, it is not required for NTIC processes to be completely closed. The information transfer from X to Y $I(Y_{t+1}; X_t|Y_t)$ (if non-zero) serves as the upper bound of the information transfer from any other process to Y. So unpredictable sensory inputs will contribute to a positive information transfer. For high NTIC systems however this transfer must be less than the mutual information between the sensory inputs and the next system state.

Q4: Exclusion or no exclusion? The authors state multiple times that "every process with a positive non-trivial information closure (NTCI) has consciousness." (p.3, and also p.12). a) Yet, Figure 4 is ambiguous in the sense that the maximum somehow seems important, while actually all levels should give rise to separate experiences. I don't see how ICT without something like exclusion (IIT style) would indicate that the maximum should correspond to human consciousness. It could just be any one out of many consciousnesses. b) p. 14: comparison to IIT: what does ICT say about different sets of variables at the same level. There might well be multiple ways to partition X into S and E, with many overlapping S's. Would those all be conscious? They would not be informationally closed from each other but they could all fulfill the requirements in the Hypothesis. It seems like finding the maximum within a level over the possible sets of elements/variables is a necessary step to identify boundaries (I think also Krakauer does that across sets of variables, but not 100% sure)

A: We thank the reviewer for this critical question. We do not impose the exclusion criterion in ICT. This is a critical difference between IIT and ICT.
A key feature of ICT is that, instead of imposing the exclusion criterion, the concept of informational closure itself already involves the notion of system-environment separation and a boundary of a system (Bertschinger *et al.*, 2006). For example, Krakauer *et al.* (2014) proposed an algorithm to add and remove variables in order to find the boundary.
It is true that every subset of a system can be computed and can have a positive value of NTIC. However, following the notion of informational closure, if a subset is not informationally closed, the subset is considered as a part of the system. ICT applies the same notion to consciousness. A subprocess (partition) of a conscious (NTIC) process can have a positive value of NTIC, and therefore, is conscious. However, the subprocess is not "another conscious process". Instead, it is considered as a subspace of the consciousness. Namely, the information contributed from the subprocess corresponds to a part of the whole conscious experience.
This is also why ICT allowed the existence of multiple consciousnesses across the scale of coarse-graining. Because conscious processes are informationally closed from each other across the scale of coarse-graining, they are not able to know the existence of other consciousnesses within the same system (information flow from other processes is zero). This is in line with using informational closure in level identification (Pfante *et al.*, 2014b) .
Therefore, ICT without exclusion criterion can parsimoniously explain the individuality of consciousness within and across the scale of coarse-graining.

Maybe add something in the manuscript

Q5: Feedback and memory: p. 11. If $NTIC > 0$ is sufficient and true information closure is not required, then feedfoward networks can be conscious according to ICT. Also, in that case, memory is not necessary, as the video camera would have $NTIC > 0$ as long as it records from a stable environment. In that case $I(Y_{t+1}, Y_t) > 0$ and $I(Y_{t+1}, Y_t|E_t)$ is small. This issue here is that having information about the next state does not imply that there is any causation, so if Y_t is highly correlated with E_t and E_t causes Y_{t+1} then NTIC is high. (Thus the IIT emphasis on causation). If somehow the fact that Y must be informationally closed wrt X does the heavy-lifting here that should be made more explicit (see above).

A: @ The reviewer is correct about the example of the video camera under the condition that the environment past state E_{t-1} shares information with its current state E_t, i.e., $I(E_{t-1}, E_t) > 0$.

Everything I want to say is here. Just hard to make these pieces into a whole. Feel frustrated.

@ In such case, simply copying the environmental states to another process (assuming that the sensor of the video camera can capture the whole state of the environment) can lead to positive NTIC.

@ This is, in fact, the "Passive adaptation" case mentioned in (Bertschinger *et al.*, 2006, p. 4),

> *"B1) Passive adaptation: The system is driven by the environment and adapts passively to all changes in the environment. In the case of an environment, that appears deterministic to the system, $H(\hat{e}_{n+1}|\hat{e}_n) = 0$, this can be achieved by simply copying the observation of the environment into the system, $S_{n+1} = \hat{e}_n$."*

@ The reviewer is correct about the video camera example under the condition that the environment past state E_{t-1} shares information with its current state E_t, i.e., $I(E_{t-1}, E_t) > 0$.

@ In such case, simply copying environmental states to a process can lead to positive NTIC.

@ In fact, this is one way to achieve NTIC discribed in (Bertschinger *et al.*, 2006, p. 4),

> *"B1) Passive adaptation: The system is driven by the environment and adapts passively to all changes in the environment. In the case of an environment, that appears deterministic to the system, $H(\hat{e}_{n+1}|\hat{e}_n) = 0$, this can be achieved by simply copying the observation of the environment into the system, $S_{n+1} = \hat{e}_n$."*

@ We can also consider this is a special case of modelling the environment when a system has enough sensory and memory capacities, and, therefore, can virtually model the environment by simply copying the environmental states into the system.

@ Because this question point to the core concept of ICT, we would like to clarify and emphasise two critical points here.

@ a) Informational closure and causation: The definition of information closure (Bertschinger *et al.*, 2006) does not involve causation. As described in Bertschinger *et al.* (2006), NTIC can be achieved via "modelling" or "passive adaptation". In the "modelling" case, the future state of a process (Y_{t+1}) can be caused by Y_t, which contain predictive information about how the environment influences the system. In the "passive adaptation" case, the future state of a process (Y_{t+1}) is caused by copying environmental information which involves environmental dynamics.

The latter case may be conceptually counterintuitive because the internal state of the system (e,g, the video camera) is entirely driven by external input. However, it is crucial to know that both cases lead the information flow $J(E \to Y)$ to (virtually) zero. This means that if Y_t contains information about how the environment E_t causes Y_{t+1},

However, this is the most critical consequence of solving the scale problem of consciousness in ICT. In physical reality, there only exists microscopic processes at the finest microscopic level. Any causal dependency at coarse-graining levels should be virtual. For example, when Y is informationally closed to X, even though Y_t

@ b) As described in our manuscript, to form a higher level of NTIC, the informational richness of the environment, i.e. $H(E)$ also has to be higher since $H(E_t)$ and $H(Y_{t+1})$ both are the lower bound of $I(Y_{t+1}; H(E_t))$. Practically, we believe that with limited sensory and memory capacities, it is difficult to give rise to high NTIC via copying sensory signals to a system.

Q6: Does ICT imply that one is unconscious while dreaming? In that case $I(Y_{t+1}; Y_t|E)$ should almost be identical to $I(Y_{t+1}; I_t)$ and thus lead to NTIC approx. 0.

A: We want to emphasise that not all the processes in the neural system are NTIC processes. This is evident since some processes in the neural system are not informationally closed. They only passively react to sensory inputs or other processes of the neural system.

To the conscious (NTIC) process, the rest of the neural system is considered as part of the environment. This notion is shortly indicated at line 187 in our original manuscript. There is no doubt that these processes are still active during sleep and dreaming. We speculate that, during dreaming, the neural system can stably form the NTIC process with respect to its environment, i.e. other parts of the neural system. However, at the current stage, this is mere speculation so we restrained ourself from making the statement in our manuscript. However, we thank the reviewer for bringing up this important question. We believe that this is still worth mentioning. We have added a short paragraph as follows:

> **Line 562:**
> *"Explaining conscious experience during dreaming is always a challenge to all the theories of consciousness. ICT currently does not have a certain answer to dreaming. However, we want to emphasise that not all the processes in the neural system are NTIC. This is trivial since some processes are evidently not informationally closed. They mainly passively react to sensory inputs or other processes in the neural system. To the conscious (NTIC) process, the rest of the neural system is part of the environment, and undoubtedly retains some degree of activity during sleep and dreaming. We speculate that, during dreaming, the neural system stably forms an NTIC process with respect to its environment, i.e. the other parts of the neural system. However, at the current stage, this is mere speculation. Searching for the NTIC process(es) during dreaming is a crucial step to extend the scope of ICT in future research."*

Finally, we believe that this is also an empirical question that can and should be tested in future studies.

Q7: Minor:

1. "contributions to the field" This section is just a copy paste from the abstract and thus not necessarily helpful. Not sure though what the purpose of this section is meant to be by Frontiers.

2. Introduction: "We currently lack a theory... " IIT and the geometric theory of consciousness (Fekete et al.) have proposed solutions. So there are theories. The following paper should be of interest and should probably be cited: Fekete T, van Leeuwen C, Edelman S (2016) System, subsystem, hive: Boundary problems in computational theories of consciousness. Front Psychol 7:1041.

3. y_t corresponds to the content. I would argue that an activation state, without taking the system or relation between elements etc into account is meaningless and cannot capture/explain the structure of an experience. Though this issue can be dealt with at a later time.

A:

1. It's true that we were also confused about it during submission. We have modified this as follows:
   [["contributions to the field"]]

2. We thank the reviewer for the reminder. We have modified our manuscript and cited the relevant references.

   Line 74:
   Original:
   "We currently lack a theory to identify the scale which conscious processes correspond to."

   Revised:
   "We currently have only few theories (e.g., Integrated Information Theory (Hoel et al., 2016) and Geometric Theory of consciousness (Fekete & Edelman, 2011, 2012)) to identify the scale which conscious processes correspond to (please also see the discussion in Fekete et al. (2016))."

3. We agree with the reviewer's comment. In fact, we do not differentiate states of processes between activation and relation. Take neural network as an example; the relations between elements are determined by the topology of a network (structure) and its hyperparameters (e.g. activation functions). We agree that activation and relation of a neural network should be both considered. After all, they are all merely the parameters for the process of the neural network. The only difference between the two groups of the parameters is the temporal scales of changes. States of activation may have more rapid dynamics than states of relation. In ICT, the two sets of parameters do not have any qualitative differences.

Do we need to write something in the main text?

Reply to Reviewer 2

Q1: I had a few problems in reading the paper, which I think should be addressed (especially the first) before the paper is published. For detailed notes on these see below.

1. The coarse-graining idea seems undefined in critical ways, but the paper reads as though it is well-defined. So maybe the authors have compressed too much detail? In the formalism presented it is unclear what coarse-graining (or information closure to other grains) amounts to. With this lack of detail, it is then unclear to me why intermediate maxima in IC are a plausible result (why does IC not just decline with progressive coarse graining).

2. The idea of 'simulation'. This is a more minor point but I would appreciate if the authors could be clearer on this. It is arguable whether or not consciousness simulates anything (e.g. see Hoffman's interface theory); and some of the assertions the authors make re simulation seem unfounded anyways. But it could be that they mean something different or subtle (that consciousness is linked to or represents or etc the environment)

3. There were several other assertions that stood out to me as strange or wrong, but could be a matter of explanation or ignorance on my part (the feedforward system, the cut-apart system). Either way some more explanation would be good.

A: We thank the reviewer for the critical comments. We have replied these comments in details in the following Q&A.

Q2: The English is not perfect and needs some good proofreading, though it was understandable throughout.

A: We have invited a native English speaker for proofreading. Please inform us if it is still not clear or unreadable.

Q3: On line 182, defining C-processes as cases where 'Y is informationally closed to X'. This is now information closure in the sense of coarse-graining, but no such thing has been defined yet? Is it supposed to be so trivial that it is not required to make it explicit? To this point, closure is defined entirely in terms of system with respect to environment.

Basically, I am not sure whether closer between scales is a matter of same-time (e.g. $I(Y_{t+1}; X_{t+1})$) or across-time interactions ($I(Y_{t+1}; X_t)$). Or could it be both?

A: We thank the reviewer for this question. We agree that this point is not clear in our original manuscript. Even though the Bayesian graphs for the system-environment dependency are different from the one for the micro-macro level dependency, we apply the same definitions of information flow ($J_t(X \to Y)$) and informational closure ($J_t(X \to Y) = 0$) to both graphs.

This is a critical assumption in ICT because applying the same definition of informational closure in both scenarios can solve the scale problem of consciousness (informational closure from one level to other levels) and construct our hypothesis about the relation between NTIC and consciousness (informational closed from a process to its environment) in an integrated framework.

In terms of the Bayesian graphs for the micro-macro level dependency and the definition of informational closure between scales, we followed the conventional setting (e.g. Pfante *et al.* (2014b)). It is important to note that coarse-graining does not and should not take time $Y_t = f_Y(X_t)$ (also see our Definition 1). Coarse-graining is merely a function which maps microstates of a system to macrostates, and the microstates and the macrostates are only different level of descriptions of the same system.

We agree that we did not indicate this assumption explicitly and may lead to some confusion for our readers. We have added a footnote to make this point explicit.

> **Line 183:**
> *" Note that, here we applied the same definitions of information flow (Eq. 1) and informational closure (Eq. 2) to the system-environment dependency (e.g. $J_t(E \to Y) = 0$) and the micro-macro level dependency (e.g. $J_t(X \to Y) = 0$), even though the Bayesian graphs are different between the two scenarios. This follows the common settings in previous studies (e.g. Bertschinger* et al. *(2006); Pfante* et al. *(2014b)). "*

Q4: On line 241: '.. not sufficiently coarse-grained variables have low values of NTIC', this is phrased as though it is necessarily true (and also that 'we saw above', though I do not see above where this is justified), but is it?

I am rewriting this part, please wait

At the very finest grain, wouldn't even all the most 'stochastic' dynamics of a system be accounted for by prior states of the system, or by the environment? Why wouldn't we assume that NTIC *only decreases* as the system is coarse-grained (as number of elements/states is decreased), for similar reasons as it must be zero at a 1-element 1-state system as the authors note? This seems an important point to be very clear on...

A:

We thank the reviewer for proposing the question about the relationship between NTIC and scale of coarse-graining. We currently know that NTIC is not a non-monotonic function of the scale of coarse-graining. Some coarse-grainings lead to higher NTICs for macroscopic than microscopic

I think I made a big mistake in the manuscript. When I wrote this part I only considered IC as a function of scales of coarse-graining. But I just realised NTIC is totally a different story that I didn't think about thoroughly. I processes. However, we are just starting the research on this relationship, and have not yet had a systematic understanding of it. This work may involve the technique of partial information decomposition and related research fields. We agree that our original claim about existing a intermediate maxima in NTIC was too strong and have revised the paragraph as follows:

Line 236:
Original:
" 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. "

Revised:
" It is important to note that NTIC can be a non-monotonic function of the scale of coarse-graining. Insufficiently coarse-grained variables could lead to lower values of NTICs at micro than macroscopic processes. 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, processes at certain levels of coarse-graining may have higher values of NTIC than other levels (Fig. 4). ICT suggests that human consciousness occurs at the level of coarse-graining where higher NTIC is formed within the neural system. "

Q5: on line 263: 'NTIC processes encodes environmental information in its state. This suggests that a NTIC process can be considered as a process that simulates the environmental dynamics.'

Why does encoding suggest simulation? An encoding *could* be a simulation, but it could well be nothing like a simulation, if we understand simulation to mean something like an imitation or reconstruction of some structure or dynamics. Two totally different environmental situations could potentially be encoded in exactly the same way (eg. as a string of 1s and 0s).

The 'simulation' notion is brought up again in section 5.2 (line 344), citing Bertschinger et al, though I do not find the idea in that paper.. I think authors need to be clearer on what they mean by 'simulation' to make this point.

A: We thank the reviewer for this mindful comment. Our wording indeed is not clear. The notion of simulation is the same as the 'Modeling' case on Page 4 in Bertschinger *et al.* (2006):

> *"B2) Modeling: The system reaches synchronization and internalizes the correlations observed in the environment by building up own structures."*

To avoid the confusion We have revised the following sentence, changed "simulation" to "modeling", and added the reference.

Line 259:
Original:
" 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. "

Revised:
" Importantly, NTIC processes internalise the environmental dynamics in its states (also see P. 4 Bertschinger et al.*, 2006). This suggests that a NTIC process can be considered as a process that models the environmental dynamics. "*

Q6: line 306: 'Blindsight patients…make above chancel-level visual judgments without having any conscious perception about visual stimuli' I am not sure this characterization of blindsight is correct, it may be a matter of the phrasing. It could be - or probably is - the case that 'blindsight' patients always have some conscious experience of what drives their correct responses, but that the 'visual character' of these experiences is degraded or missing. (Overgaard, Experimental Brain Research 2011) for a specific example (Mazzi, Bagattini, Savazzi; Frontiers in Psychology, 2016)

A: We thank the reviewer for the clarification. Yes, we specifically wanted to address the visual character of their experience. We agree that some patients still preserve some forms of conscious experience. We have changed our sentence and cited the relevant references as follows:

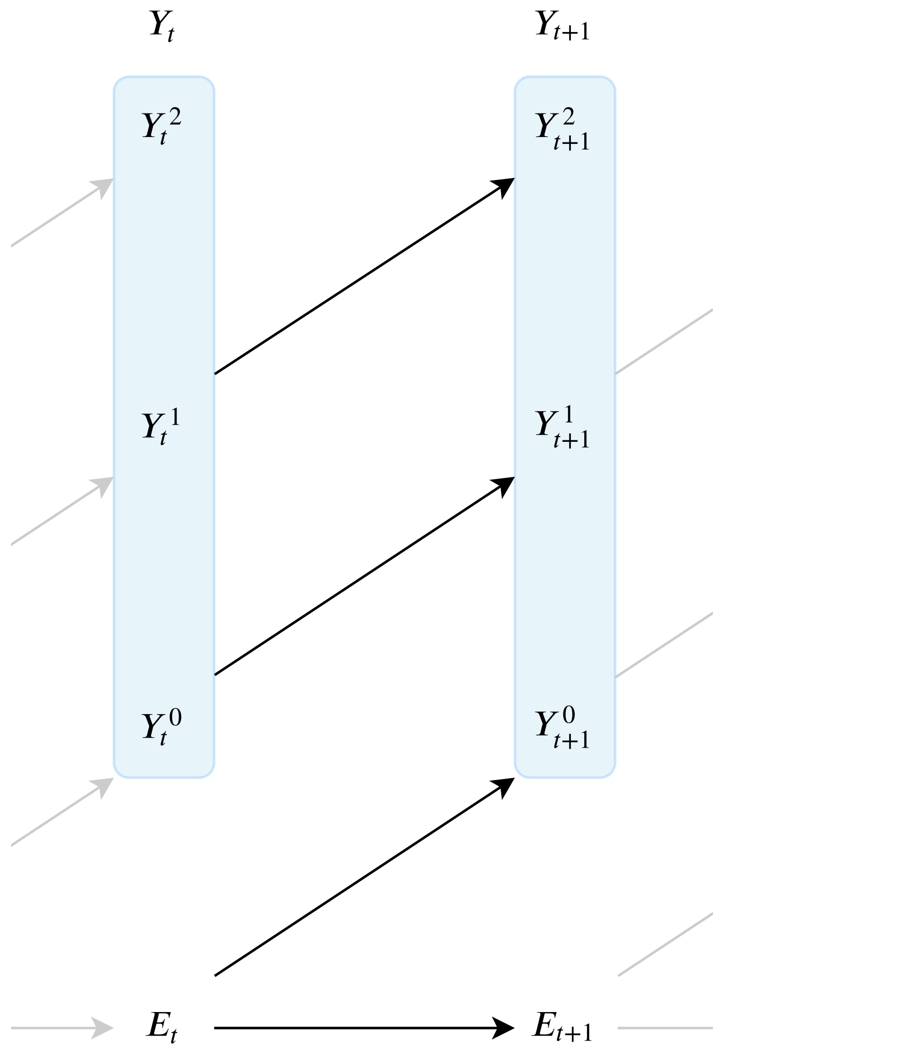
Line 302:
Original:
" Blindsight patients are able to track objects, avoid obstacles, and make above chance-level visual judgements without having any conscious perception about visual stimuli. "

Revised:
" Blindsight patients are able to track objects, avoid obstacles, and make above chance-level visual judgements with degraded or missing visual experience (however, in some cases, they may still preserve some forms of conscious experience, see Mazzi et al. (2016); Overgaard (2011)). "

Q7: on Line 316: On the issue of a feedforward network, the current state of a feedforward network with more than one layer is certainly driven by its own past states! So, mutual information of a feedforward network over a time lag should not be zero, unless I am misunderstanding something here? At the same time, I can see a version of the authors' argument here - for a unit in a feedforward network with depth (distance-from-input) of N, a time lag of N time steps would always account fully for the states of the element. Is this the idea?

Camera thing..., we need to rewrite this reply as well

A: We thank the reviewer for this interesting and valuable comment.
We can consider a simple case in which a feedforward neural network Y consists of three layers, Y^0, Y^1, and Y^2. Y^0 receives sensory input from the environment E. Each layer Y^i_t passes information to the next layer Y^{i+1}_{t+1} with a constant time cost Δt as showed in the following figure.



Because Y_t^0 determines Y_{t+1}^1 and Y_t^1 determines Y_{t+1}^2, it is true that the mutual information $I(Y_t; Y_{t+1}) \geq 0$. However, E_t only influences Y_{t+1}^0 at time $t+1$. Therefore

$$I(Y_t; Y_{t+1}|E_t) = I(Y_t; Y_{t+1}) \quad \text{and}$$
$$NTIC_t(E \to Y) = I(Y_{t+1}; Y_t) - I(Y_{t+1}; Y_t|E_t) = 0$$

Therefore, a feedforward network passively driven by input signals should have zero NTIC even though when time step (temporal coarse-graining) is small.

I don't think I answer this question well. Do we need to add anything in the main text?

Q8: on line 441: under the 'Prediction after system damaged', it is suggested that ICT predicts cutting a system in half would render both halves unconscious. But this would only be the case if neither half contains its own C-processes, yes? Since ICT allows that many C-processes can exist at the same time, it would have to be some special case for this prediction to hold true. So it seems to me the prediction is actually similar to that of IIT.

A: We thank the reviewer for this considerate comment. The reviewer is correct. This prediction is under an assumption which we did not explicitly specify in our last version. We assume that we have one conscious (NTIC) process involving information in both brain hemispheres. Namely, the process is informationally closed only when we consider the information in both hemispheres. We agree that if both hemispheres have their own NTIC proceess ICT should predict no change of conscious experience before and after cutting. Cutting should not make any difference because they are informationally closed with respect to each other. We also agree that this prediction is relatively premature. However, we also acknowledge the strength of all computational theories of consciousness. The reviewer's question is empirical to us under formal mathematical formulations of ICT and IIT. Systematical comparisons between model predictions can be done by rigorous modelling studies in the future.

[[The original paragraph has been changed as follows: lline 435:]]

Q9: Line 540: the theory doesn't really seem to intend to solve the hard problem(s) at all, much less 'completely solve'. I was expecting the problem of dreaming to come up in the last section (maybe along with SMC for which it is a serious problem). When dreaming the information between environment and the system is virtually zero; is this a problem for ICT? Also, more specific phenomena like the eigengrau - if you take away all visual input, for long enough, I do not lose my visual experiences - rather they take on a special state. Does ITC need to accept the possibility that even *trivial* IC can be a conscious process?

A: ICT is our first attempt to approach the hard problem of consciousness. We hope that ICT and our following works can establish an informational perspective on conscious experience. Instead of solving the hard problem of consciousness, we more look forward to mathematically determining whether the hard problem can be solved or not by information theory. However, because the relevant discussion is outside the scope of this article, we prefer not to discuss the hard problem here.

We want to emphasise that not all the processes in the neural system are NTIC processes. This is evident since some processes in the neural system are not informationally closed. They only passively react to sensory inputs or other processes of the neural system.

To the conscious (NTIC) process, the rest of the neural system is considered as part of the environment. This notion is shortly indicated at line 187 in the original manuscript. There is no doubt that these processes are still active during sleep and dreaming. We speculate that, during dreaming, the neural system can stably form the NTIC process with respect to its environment, i.e. other parts of the neural system. The same idea can be also applied to phenomena like Eigengrau. However, at the current stage, this is mere speculation so we restrained ourself from making the statement in our last version of the manuscript. However, we thank the reviewer for bringing up this important question. We believe that this is still worth mentioning. We have added a short paragraph as follows:

> **Line 562:**
> *"Explaining conscious experience during dreaming is always a challenge to all the theories of consciousness. ICT currently does not have a certain answer to dreaming. However, we want to emphasise that not all the processes in the neural system are NTIC. This is evident since some processes are not informationally closed. They mainly passively react to sensory inputs or other processes in the neural system. To the conscious (NTIC) process, the rest of the neural system is part of the environment, and undoubtedly retains some degree of activity during sleep and dreaming. We speculate that, during dreaming, the neural system stably forms an NTIC process with respect to its environment, i.e. the other parts of the neural system. However, at the current stage, this is mere speculation. Searching for the NTIC process(es) during dreaming is a crucial step to extend the scope of ICT in future research."*

Finally, we believe that this is also an empirical question that can and should be tested in future studies.