Is integrated information theory viable as a theory of consciousness?

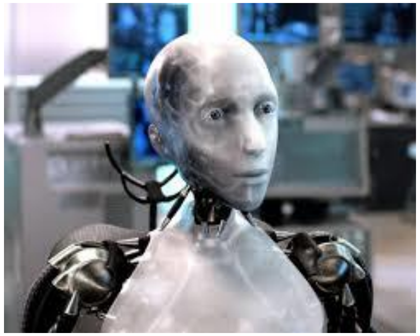
George Deane

Reasonable Scope

We want a theory of consciousness to have **reasonable scope**.

A theory of consciousness that is too **inclusive** or ‘liberal’, would grant consciousness to systems that don’t have consciousness. We intuit that rocks, trees and countries do not have consciousness.

A theory of consciousness that is too **exclusive** or ‘chauvinistic’ would not grant consciousness to systems that do have consciousness.



Multiple Realizability

- IIT is a computationalist theory of consciousness.

- Computationalism entails multiple realizability.

- This means that consciousness is substrate neutral – it can be built out of anything that can implement the computation required for consciousness.

- Multiple realizabiity was initially formulated as an anti-reductionist argument, but has since been used to argue that functionalism and computationalism are **too inclusive**.

China Brain and ‘funny instantiations’

Block (1980) puts forward the China Brain thought experiment:

If each member of the Chinese nation were asked to simulate the actions of neurons in the brain, so the population was functionally equivalent to the human brain, would this system instantiate consciousness?

Block uses this to show that consciousness is a problem for functionalist theories. His intention is to show us that there must be more to consciousness than functional arrangement. Some philosophers, like Daniel Dennett, argue the system would be conscious.

Multiple realizability – possible approaches

Some possible ways to resolve the problem of multiple realizability:

- Simply accept that these systems would instantiate consciousness.

- Constrain what is meant by a computation (Chalmers 1994 makes a start on this)

- Claim that while computation is a necessary condition for consciousness, it is not a sufficient condition. Generation requires other things such as a certain spatiotemporal grain, or a certain substrate.

Observer-relative/observer-independent existence – a distinction

- Observer-independent features of the world are: objective and intrinsic. Natural sciences are generally concerned with these features, eg. mountains, galaxies, electromagnetism.
- Observer-relative features of the world are **extrinsic,** and usually defined in terms of function. Money, wallets, paper weights etc.
- Observer-independent existence is *discoverable*, observer-relative existence is *assigned.*

Consciousness

Consciousness has an observer-independent existence, and so requires an explanation which is also observer-independent.

Does computation have an observer-independent or observer-relative existence?

- For Searle, computation is observer-relative (it does not refer to any intrinsic feature of a system). Therefore to call a brain or consciousness a computer is trivially true but lacks any explanatory power.

Problems with IIT

Silent Units

- According to IIT, the shape in Q space is specified not only by the neurons that are firing, but also by the neurons that are not firing.

- This is a counterintuitive conclusion – how is the mere ***potentiality*** of a neuron firing to contribute to a state of consciousness?

- Fekete and Edelman (2011) see this as a unresolvable flaw with IIT unless it can explain how silent units can contribute to experience.

Silent Units Response

- Tononi treats this problem not as a gap in the explanation but as an empirical prediction: if silent units do indeed contribute to a subject’s experience, then ‘cooling’ them temporarily (thereby removing their potentiality for firing) would cause the qualia shape to collapse.

- Tononi (2008) stops short of explaining how this works and simply states: “consciousness can be characterised extrinsically as a disposition or potentiality – in this case the potential discriminations that a complex can do on its possible states, through all combinations of its mechanisms.”

Is information observer-relative?

- In the traditional sense, information is observer-relative. The rings on a tree stump tells something about the age of a tree only relative to an observer.

- If information is observer relative then it is information only relative to consciousness. Therefore to use information to explain consciousness we would have to invoke consciousness to explain consciousness – a circular definition.

Is information observer-relative?

- Tononi and Koch argue that while in the traditional sense information is observer-relative, integrated information is a novel, non-Shannonian notion of information. Integrated information on this account is an intrinsic property of a system that is observer-independent; a mathematically specified quantity.

- Integrated information is the differences that make a difference to a system from an intrinsic perspective, not relative to an external observer.

- No simple way to settle this dispute, but if it is possible to show information in this sense is an intrinsic feature of a system, then Searle's criticism can be debunked.

The exclusion principle

- The exclusion principle is an anti-nesting principles that prevents multiple consciousnesses from overlapping one another in a system. It states that only a conceptual structure that is maximally irreducible can give rise to consciousness.

- According to IIT, consciousness arises wherever there is integrated information – but if one system of integrated is nested within another, it is only the system that integrates the most information that gives rise to consciousness.

- Tononi's reasoning for this seems to be that it is absurd for multiple consciousnesses to exist in a single system – but it doesn't seem fully explained why it would only be the local maximum that would generate consciousness.

The exclusion principle

- The exclusion principle explains why IIT would not grant consciousness to aggregate systems. For example, two people talking does not create a new third consciousness composite of the two. There is a local maxima in both persons that are separate from one another as the systems do not have direct influence over one another.
- **Thought experiment 1**: IIT entails that if a direct brain to brain interface that made two brains so interconnected that the Φ of the two as a whole exceeded the Φ of each individual, a new consciousness would into existence and the other two consciousnesses would disappear.
- **Thought experiment 2**: If, axon by axon the corpus callosum was blocked, at some point the local Φ present in each hemisphere will exceed the Φ of the whole and the unified consciousness will vanish and be replaced by one consciousness in each hemisphere.

The hard problem

Some philosophers might argue that IIT does not address the hard problems of consciousness:

- Is the experience itself left out?

- Chalmers (1995) claimed that reduction could never explain consciousness because it is not clear “*Why is the performance of these functions accompanied by experience?”* – couldn’t the same question be asked of Integrated Information Theory?

The hard problem

Does this mean the explanation is incomplete? Tononi (2008) argues NO.

- An exhaustive mathematical description of informational relationships completely captures all there is to say about an experience.
- Experience itself is left out of the description, but necessarily so, because *being is not describing.*

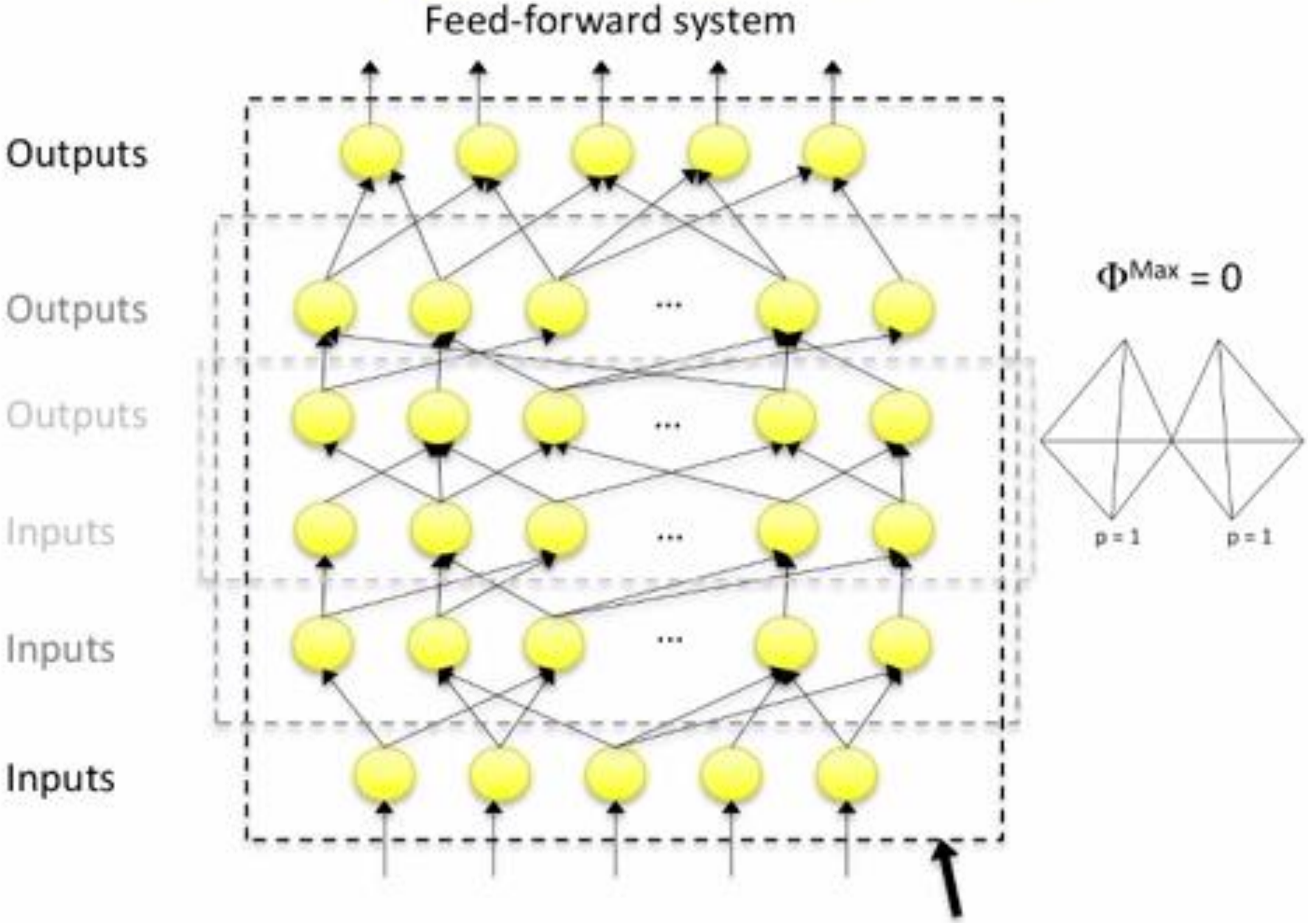
“Experience is a way of being, not a way of knowing”

Q-shapes exhaustively characterise an experience *extrinsically*, from the outside, but experience itself is necessarily *intrinsic.*

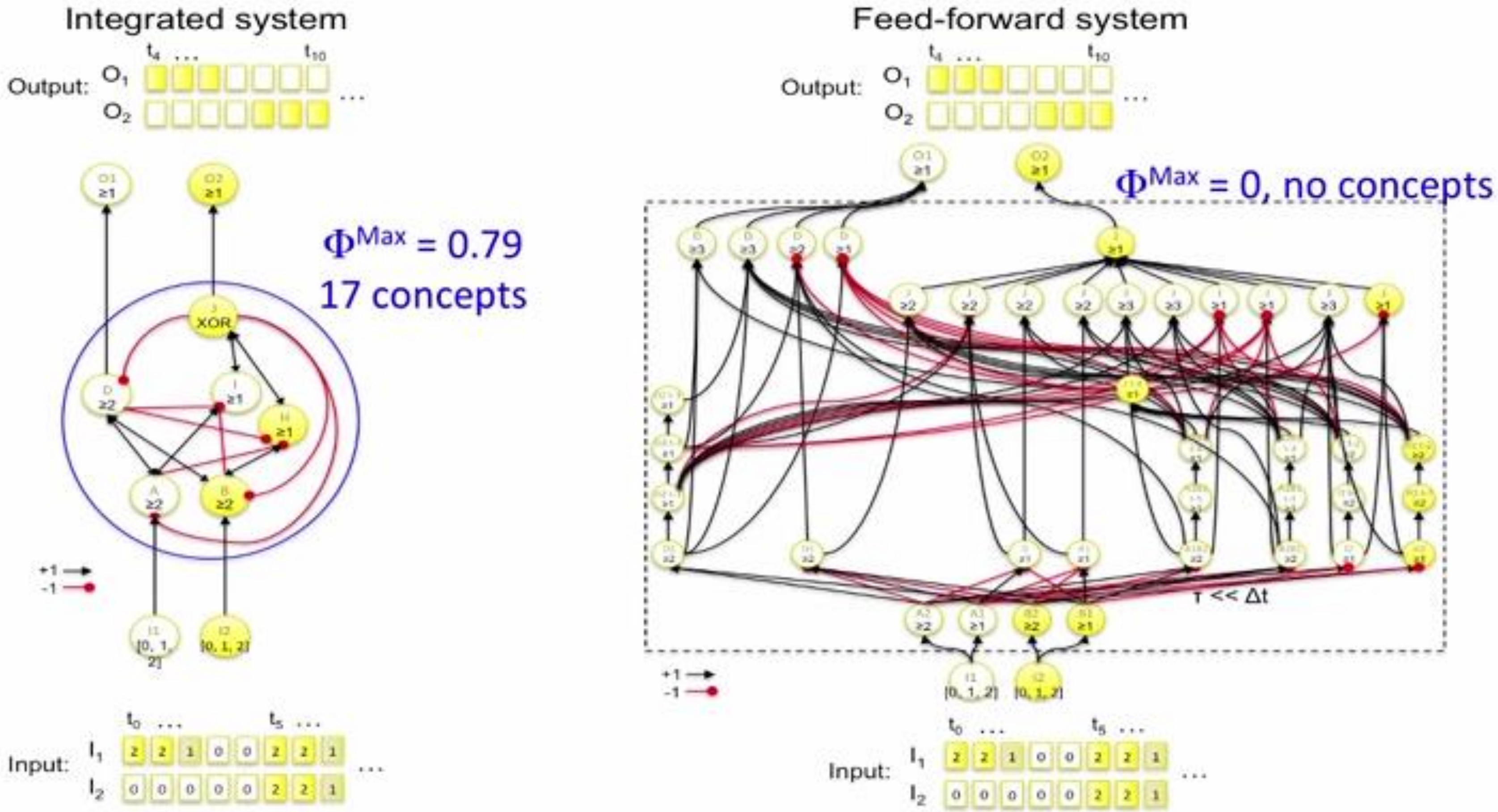
Q-shapes allow for comparison of the similarities between different experiences of conscious beings, but does not substitute for actually being in that experiential state.

IIT 3.0

Complicated systems can be unconscious: feed-forward "zombie" systems do not generate consciousness



Conscious and unconscious systems can be functionally equivalent



Whether a system is conscious cannot be decided by input-output behaviour alone – the processing is important.

Measuring Φ for large systems

- Extremely difficult to measure Φ for large systems, as it very quickly becomes too computationally demanding. Way beyond the specifications of current supercomputers.
- Current computing power can only handle about 16 nodes, a far cry from the billions of neurons in the brain.
- It is unclear whether the computational simulations currently used would scale up to larger networks.
- Nonetheless, empirical observations have so far shown promise for the theory.

georgejwdeane@gmail.com

References

Block, N. (1980). Troubles with functionalism. *Readings in the Philosophy of Psychology*, *1*, 268-305.

Chalmers, D. J. (1994). On implementing a computation. *Minds and Machines*, *4*(4), 391-402.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of consciousness studies*, *2*(3), 200-219.

Fekete, T., & Edelman, S. (2011). Towards a computational theory of experience. *Consciousness and cognition*, *20*(3), 807-827.

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS computational biology*, *10*(5).

Searle, J. R. (1990, November). Is the brain a digital computer?. In *Proceedings and addresses of the american philosophical association* (Vol. 64, No. 3, pp. 21-37).

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin*, *215*(3), 216-242.

Tononi, G. (2012). Integrated information theory of consciousness: an updated account. *Archives italiennes de biologie*, *150*(2-3), 56-90.