What is a Neural Correlate of Consciousness?

David Chalmers

Foundational Questions

(1) What is “consciousness”?

(2) What is a “neural correlate of consciousness”?

(3) How can we find the neural correlate of consciousness?

(4) What will the neural correlate of consciousness explain?

What is Consciousness?

- Consciousness = subjective experience.
- A person or system is conscious if there is something it is like to be that person or system.
- A mental state or brain state is conscious if there is something it is like to be in that state.
- In vision: consciousness corresponds to visual experience, i.e. to what subjects *see*.

Defining the NCC

- First pass: A neural correlate of consciousness is

  a specific system in the brain that correlates directly with states of conscious experience

- I.e. neural system N is an NCC if

  states of N
  correlate directly with
  states of consciousness

States of Consciousness 1

- What is a "state of consciousness" here?
- Interpretation 1: there are two states of consciousness
  - The subject is conscious
  - The subject is not conscious
- Then an NCC will be a neural system whose state determines whether or not the subject is conscious.
- Various thalamocortical proposals: e.g.

  states of intralaminar nucleus (Bogen)

  reticular activating system (Newman).

States of Consciousness 2

- Interpretation 2: background state of consciousness
  - Wakefulness, sleep, dreaming, hypnosis, etc…
- Then an NCC will be a neural system whose state determines whether or not the subject is conscious.
- E.g. neurochemical levels of activation (Hobson)

States of Consciousness 3

- Interpretation 3: Specific contents of consciousness
  - E.g. red vs green, horizontal vs vertical
  - Specific qualitative character of a visual image
- Then an NCC will be a neural representational system N such that the representational content of N determines the representational content of consciousness.
  - Representational content = what a system represents
  - First approximation: receptive field
- This fits with the approach of
  - Logothetis on binocular rivalry
  - Milner and Goodale on ventral and dorsal streams
  - Most work on NCC in visual neuroscience

Direct Correlation

- What does “directly correlate” mean?
  - State of N <---?----> State of consciousness
- Necessary and sufficient?
  - Too strong (allows only one NCC)
- Sufficient
  - Too weak (allows irrelevant processes in NCC)
- Minimally sufficient system
  - Just right (pares NCC down to core)

What Range of Cases?

- Over what ranges of cases must an NCC correlate with consciousness?
- Ordinarily functioning brain, ordinary environments?
  - Too weak: retina might qualify as NCC!
  - We need unusual cases to disentangle correlations among potential NCC (c.f. binocular rivalry, brain-damaged patients)
- All possible cases (any lesions, damage, rewiring, etc)?
  - Too strong
  - If we lesion the entire NCC area, it might not be an NCC any more.
  - So we need to hold something constant.

What Range of Cases 2

- What do we hold constant, and what variation do we allow?
- Ordinary brains, varying inputs?
  - Weak but safe criterion (cf. binocular rivalry)
- Ordinary brains, vary brain stimulation
  - Probably safe but can lead to problems in the extreme
- Vary brain function via lesions, damage, etc
  - Methodologically dangerous, as lesions could change NCC location
  - Q: Can we characterize allowable lesions in NCC studies?

Summary: What is an NCC?

- A neural correlate of consciousness is
  - A minimal neural representational system such that representation of a content in that system is sufficient, under conditions C, for representation of that content in consciousness.

Methodological Consequences

- (1) Be careful with lesion studies
- (2) Expect many NCC's (even within a modality)
- (3) Minimize size of an NCC
- (4) Distinguish NCC for state and content
- (5) Need studies that monitor neural representational content
  (cf. single-cell studies vs. brain imaging)
- (6) Correlation over a few situations is weak(ish) evidence
- (7) We need good criteria for the ascription of conscious states.

Searching for an NCC

- Q: How *can* we search for an NCC?
- Problem: Consciousness is a private, subjective state
- It would be a lot easier if we had a *consciousness meter*…

Bridging Principles

- Q: How can we search for an NCC without a consciousness meter?
- A: With the aid of operational criteria for ascribing consciousness, and other *bridging principles*
  - These bridging principles allow inferences from observed data, e.g. re brain and behavior, to states of consciousness
- The most basic such bridging principles aren't determined by experiments
  - They're pre-experimental assumptions used to interpret experimental results
  - They are usually implicit, so it's helpful to make them explicit

Bridging Principle #1

- The principle of verbal report:
  - When subjects report that they are having a conscious experience, then they are having that experience
- By far the most widely used criterion in experiments on humans
- We cannot prove this assumption (cf. the philosophical problem of other minds)
  - But neither can we prove the existence of the external world
  - The principle serves as a reasonable assumption that makes science possible.

Bridging Principle #2

- We don't want to require language for NCC experiments
- E.g. experiments by Logothetis et al, with monkeys pressing bars to "report" their experiences.
- Criterion is something like:
  - If information is available for an arbitrary voluntary response, it is conscious.
- Underlying principle?
  - When information is directly available for global control, it is conscious (at least within conscious systems)

Refining the Bridging Principle

- For a refined principle, we might put more weight on some aspects of control then others:
  - E.g. direct availability for cognitive/decision functions
  - Also, availability for higher-order thought
- C.f. Milner/Goodale on two visual pathways
  - Ventral stream: conscious information
    - availability for cognitive/decision functions
  - Dorsal stream: unconscious information
    - availability for motor function
- Work on unconscious perception and implicit memory (Merikle, Jacoby, etc)
  - Criterion seems to be forming the higher-order belief that one has the experience and using this belief in report and reasoning (cf. exclusion task).

Global Availability

- Hypothesis: The basic correlation holds between consciousness and *global availability*
- This seems to fit first-person evidence
  - Conscious information is globally available to control behavior
  - Information globally available to control processes is usually conscious
- If something like this is right (and if something like this is guiding research on NCC), then what follows?

Rational Reconstruction

- Premise 1 (pre-experimental): consciousness correlates with global availability
- Premise 2: (experimental): Global availability correlates with neural process N
- Conclusion: Neural process N correlates with consciousness.

Mechanisms of Availability

- If so: A neural correlate of consciousness will be a *mechanism of global availability* in the brain

- This seems to fit with various empirical proposals
  - 40-hertz oscillations
  - Neuronal global workspace
  - …

Multiple NCCs

- There will likely be many neural correlates of consciousness
- As there will likely be many mechanisms of global availability
  - in different modalities
  - at different stages of the processing path
  - at different levels of description
  - …
- So: maybe multiple proposals for the character of an NCC could be correct?

Consciousness Module

- It could turn out that there's a single functionally localizable system that subserves global availability
  - Cf. Baars' global workspace
- If there is, one, it may qualify as a "consciousness module" in the brain
- However, this does not seem especially likely
  - More likely to be multiple nonlocalizable processes?

V1 or Extrastriate Cortex

- An ongoing debate: are neural correlates of visual consciosness in primary visual cortex or in extrastriate cortex.
  - On the view I've outlined, this will depend on which is most directly implicated in global availability
- Crick and Koch: V1 is not an NCC, as it doesn't project to prefrontal cortex and prefrontal cortex co-ordinates control
  - This reasoning is supported by the current methodology
- Block: V1 may be an NCC, as there may be conscious information that is inaccessible in the brain
  - This reasoning is not supported, and it's hard to see what evidence we could have for such information.

Is the NCC the Holy Grail?

- Some hope that isolating the NCC will give a definitive test for consciousness in other systems
- Not so: the primary criterion of correlations remains that between consciousness and global availability, or whatever.
- Dissociate the NCC from this (e.g. in animals or in brain-damaged patients) and all bets are off
- Still, such an NCC may still be a useful guide to consciousness in systems relevantly similar to us.

What Will an NCC Explain?

- Will an account of the NCC *explain* consciousness?
- Arguably it might help explain global availability, and it might specify a *basis* for consciousness.
- But there remains an explanatory gap: Why does the NCC, even if globally available, give rise to consciousness?
- My view: We may need to take certain correlations between physical states and consciousness as primitive principles, not themselves explained.

Conclusion

- Still, the search for an NCC is a central and important project for the science of consciousness
  - It's appears tractable in principle, with good methodologies developing
  - There are numerous obstacles, but there are also ways to overcome them
  - It has the potential for many useful consequences
  - The goal is visible somewhere in the middle distance