# Handout on "The search for invertebrate consciousness" by Jonathan Birch

- Aidan Cafazzo

**Thesis:** Before we can even consider whether invertebrates are conscious, a robust approach for *how* to determine consciousness in invertebrates must first be agreed upon.

Birch examines three possible approaches and proposes what he believes to be the better option in an effort to gain consensus on a methodology to apply to the problem.

"What is needed? At this stage, not more theory, and not more undirected data gathering. What is needed is a systematic search for consciousness-linked cognitive abilities, their relationships to each other, and their sensitivity to masking." (133)

**Question**: Why is it extra difficult to determine consciousness in invertebrates (meaning insects, e.g. bees, ants, decapod crustaceans e.g. crabs, cephalopod molluscs e.g. octopuses) than in mammalian animals?

**Answer**: Invertebrates, like bees, are evolutionary distant from humans. They are so different from us. Eg. Nervous system is dramatically different, contain fewer neutrons (1M vs 100Billion for humans. Just because they do cognitively impressive things, many think that it insufficient to be conscious. Arguing that a robot could do similar things and NOT be conscious.

## Overview

The paper focuses on evaluating the possible methodologies that could be applied to determining consciousness in invertebrates. These are based on existing research into both human determinants of consciousness and what has been applied to other types of animals. He places them into three main categories: Theory Heavy, Theory Neutral, and Theory Light, then examines each approach and any potential issues with each.

Ultimately, he argues for what he believes to be the best methodology to move forward: Theory Light. This option avoids the main drawbacks of the other two options by using a "cluster" of abilities (rather than an exhaustive list or definition) that would be observable only when an animal is conscious in order to determine when an invertebrate is demonstrating consciousness.

**Premise 1: Theory Heavy cannot be used because it relies on an "all or nothing" approach since there is no agreed upon "minimum requirement" to satisfy the theories.**

Take a *definition* of "consciousness" and use it as the measure to compare animals against.

- Requires that first researchers "develop a well-confirmed, complete theory of consciousness in humans, and we take this theory 'off the shelf' and apply it to settle the question of whether animals, in disputed cases, are conscious or not" (134)
- Explores two main theory-based approaches: Global Workspace Theory (Baars, Dehaene et.al.) and the other is midbrain theory (Merker).

**Global Workspace Theory:**

- Says there is a part of the brain (typically thought to be in the cortex) that takes in stimulus for all different systems (perception, memory, affective systems), integrates them and then broadcasts that integrated content onwards to consumer systems (mechanisms of verbal reporting, planning, reasoning, decision-making).
- When representations enter that Workspace, they become phenomenally conscious. (Therefore, anything NOT in the workspace could NOT be consciously reported on).
- **Issue**: there is no way of deciding which parts of the "broadcast system" are indispensable for consciousness. Brain-damaged humans may not have all systems working, but are still capable of consciousness, so it's not all-or-nothing. Without knowing the minimum requirement, the measure is useless for invertebrates (since we know they won't satisfy all)

**Midbrain Theory:**

- The midbrain is a "evolutionary ancient", sub-cortical part of the brain. (136)
- Argues that "the integrative mechanisms of the midbrain are sufficient for consciousness" (136) - even if missing a cortex (e.g. brain damage).
- Argument from analogy then: if invertebrates have an analogous (similar) mechanism, that would suffice.
- **Issue**: Birch feels Merker's midbrain theory doesn't even hold up for humans. Need to find actual evidence of consciousness in humans missing a cortex – which he tried to prove in children born with cortical damage (*hydranencephaly*). ***However***, these children always had some remaining cortical tissue, so could be argued that consciousness was occurring there (global workspace).
- If sceptics don't believe that midbrain consciousness is sufficient in humans, they would not accept it as evidence for insects and other invertebrates. Insects don't have a mid-brain, though they have something functionally similar to it)

- Both suffer from the fact that they only work if ALL conditions are met. No animal is going to possess the FULL human broadcast network, and we can't determine how much is enough. Similarly, midbrain requires less, but we don't know exactly how much less since humans with no midbrain activity (or unconscious e.g. from anaesthesia) also can't report on their experience.

**Question**: Do we have a well-confirmed, complete theory to define consciousness in humans? Is there existing agreement on what determines "consciousness" in humans?

**Answer**: No. That is the key issue. There continue to be many theories as to where consciousness originates or shows up in the brain. It is impossible to extend any of the theories to animals when we cannot definitely say, if we observe X, then it is definite – and sufficient – evidence of consciousness

- Birch rejects Theory Heavy as viable on the basis that researchers have yet to agree on a definitive set of criteria for how to determine consciousness in humans, much less in animals.

**Premise 2: Theory Neutral is insufficient because (similar to Theory Heavy) there is no definition of either the minimum criteria for consciousness nor of any markers that would negate consciousness**

Inference by analogy. Compare animal behaviours to a list of traits that we know to indicate consciousness in humans.

- Rather than start with a theory of human consciousness, we should "build up a list of the behavioural, functional anatomical sim
- Similar to the issue with Theory Heavy, Birch rejects this approach as viable because there is not yet an agreed up list of behaviours or anatomical similarities from which to propose analogies or inferences.

  **Theory Neutral Example:** "same effect, same cause" principle
  - If animal exhibits similar-to-human behaviour in response to a stimuli, then we can infer that they have the same reaction to it (eg. pain), ***unless a defeater exists***.
  - A defeater clause must exist or else the statement could always be true. We could freely infer that every time a creature reacted similarly to a human (in response to stimuli), it was because it was having the same experience that a human does. ***But what counts as a defeater?***
  - Again, we'd need to define the minimum criteria (defeaters) and we can't do that.

**Premise 3: Theory Light provides sufficient flexibility to indicate consciousness and relatively quantify its observation, and could be applied to invertebrates.**

- Birch's own "middle path" between Theory High and Theory Neutral. Essentially proposes determining a *collection* of abilities that only occur during consciousness and using any number of those to determine if an animal meets criteria for consciousness.
- Birch's *facilitation hypothesis: "Phenomenally conscious perception of a stimulus facilitates, relative to unconscious perception, a cluster of cognitive abilities in relation to that stimulus."* (140)
- Using a "cluster" (or set of possible) cognitive abilities is key to his hypothesis because it both avoids having to have a complete, undisputed set of abilities but also allows for variations of what is observed in different animals. **It may not be the exact same cognitive abilities, but it will be a set of abilities that *only* display when consciousness is present.**
- **What abilities go into the cluster?** Birch: focus on those abilities that we believe are "facilitated by conscious perception in humans." (145) No role for neural correlation of those abilities either, since we know invertebrate biology to be quite different to humans – we're looking for general cognitive abilities. (E.g. learning of these types: trace conditioning, interval timing, rapid reversal)

**Trace Conditioning:** There is a stimulus, **followed by a defined time gap**, then a reward. If the subject is aware of the time gap, and anticipates the reward near that defined time, then they are consciously noticing the delay, rather than a knee-jerk, instinctive reaction of stimuli-reward with no consideration of the time gap.

**Rapid Reversal Learning:** Subject learns a pattern (e.g. treat is on red table, yellow table is empty), and then the pattern is switched. Subject quickly reverses their behaviour and begins looking for the treat at the yellow table. This requires conscious perception of the criteria and then conscious adjustment when the criteria switches.

**Cross-modal Learning**: Learning of associations across senses – so using more than one sense to link the stimui together: E.g. A sound is associated with a smell (both experienced consciously). This requires the conscious linking of one stimulus to the other – although it is not impossible to do this unconsciously (just becomes very difficult).

- **How many abilities will signal consciousness?** Birch: Avoid the pitfall of defining a number required. Measure as a percentage or fraction of the cluster. (145) The more abilities shown, the stronger the argument for existence of consciousness. One or two abilities would be a weak argument (which, Birch acknowledges, may turn out to be the case for invertebrates.)

**Question**: Do you agree with Birch that by loosening some of the theory requirements and using his methodology is satisfactory and sufficient to apply to whether animals, and specifically invertebrates exhibit consciousness?

**Bees experiments**

Birch applies his Theory-light option to pre-existing research done on bees. Essentially, he shows that Bumble Bees are better candidates for consciousness than Honey Bees. This is because Bumble Bees did better in Trace Conditioning testing (the honey bees ignored the time interval altogether, whereas the bumble bees seem to learn that there was a delay, even if their timing wasn't quite right).

Bumble bees also did better on Rapid Reversal Learning. Although both types of bees successfully reversed their reaction, the honey bees got worse at it over time (with more switching) while the bumble bees got faster at noticing and adapting to the switch. The speed over repetitions should improve because that's generally a mark of intelligence.

**Conclusion**: The Theory Light approach provides a path forward to the debate of whether invertebrates are conscious. It allows for a systematic search for the cluster of consciousness-linked cognitive abilities that would form the basis for cases of consciousness.

**Question**: Do you agree with Birch that by loosening some of the theory requirements and using his "cluster-based" methodology is satisfactory and sufficient to apply to whether animals, and specifically invertebrates exhibit consciousness?

**Question**: Birch asks, "How should we treat bees, insects and other invertebrates in the absence of strong evidence of consciousness?" (147)

**Answer**: Birch recommends we "err on the side of caution" and safeguard the welfare of these animals.