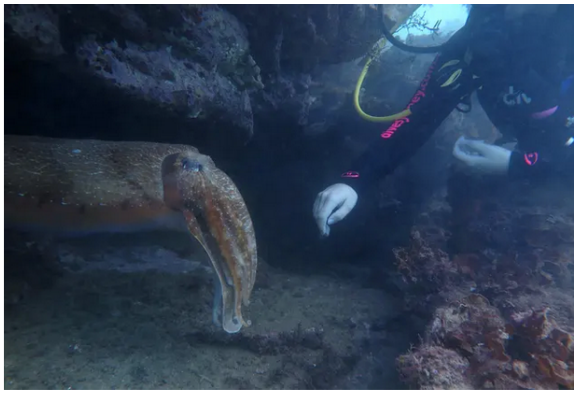
Animal Consciousness (and its Evolution); Robot Consciousness

(Blackmore ch.8, Baars p.31-33)

Detailed elaboration: https://www.cs.rochester.edu/u/schubert/191-291/lecture-notes/animal-and-robot-consciousness.txt

▲ An Australian giant cuttlefish: 'It is no stretch to say they have personalities.' Photograph: Peter Godfrey-Smith



▲ Elle Hunt with an Australian giant cuttlefish at Cabbage Tree Bay, Manly, Sydney. Photograph: Peter Godfrey-Smith

cf. Netflix "My Octopus Teacher"

- **Animal C: from plants to apes to humans**
  - What's behaviorally/anatomically relevant?
  - Marian Stamp Dawkins' chickens
  - Mirror test: "Theory of Mind" (ToM); ToM ←→ self-modelling (a departure from Blackmore)
  - apes, magpies, etc.; devious monkeys; (VIDEO: https://www.youtube.com/watch?v=cKs_iWOQVNY)
  - Language (Koko, etc.)
- **Evolution of C [self-awareness, and phenomenal C – qualia]**
  - Selective advantages & evolution of self-awareness? (planning, social "reasoning"?)
  - Selective advantages & evolution of phenomenal C?? (perception-thought melding?)
- **Take-aways from Blackmore (+ parts of Baars)**
  - Easy Problem, Hard Problem, Explanatory Gap
  - Self-awareness (aspect of access C); phenomenal C (qualia)
  - Physical basis of C (fMRI, drugs, synesthesia, binocular rivalry, split brains, blindsight, damaged minds, unusual states (REM sleep, OBE's, NDE's) & their neural correlates
  - Conscious and unconscious neural activity
  - Illusion of "immediate" perception, conscious will (Libet)
  - Theories of C: dualism, materialism, identity theory, Dennett, Penrose, Chalmers, Zen, Emergentism (Searle), mysteriansm, delusionism, self-modelling HOT's (w. "signal sites" ☺)

Implications for Robots

Blackmore: they'll probably share our delusions (if sufficiently intelligent)

Baars & Franklin: with the right architecture, they'll be conscious

"Consciousness is computational: The LIDA model of global workspace theory"
(https://www.researchgate.net/publication/238423831_Consciousness_is_computational_The_LIDA_model_of_global_workspace_theory)

Self-models with signal sites:

We could create phenomenally conscious or unconscious (Zombie-like) robots, but the latter probably could not "thoughtfully" control their interactions

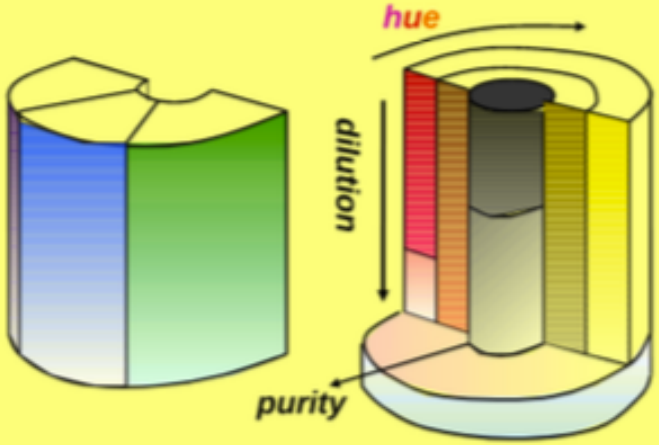
Version 1: rote sensorimotor routines (with affordances) plus separate talk/thought

Version 2: rote sensorimotor routines (with affordances) plus integrated talk/thought (via hyperpropositions in the self-model)

(See Nao28 (that (touching Nao28 (right-hand-of Sue))) )

Perceptual or "Qualia" Geometry

- While qualia are not "absolutely" verifiable, they have a verifiable "logic" (geometry, similarity structure)



Theories of phenomenal C (again – & then some!):

Dualism (Descartes, most people)

Monism (idealistic, materialistic, neutral)

All of the approaches that follow (except perhaps mysterianism) are materialistic

Reductionism (Dennett, Baars, Churchland)

Higher-order theories (McDermott, Rolls, Arbib, …)

Emergentism (John Searle)

Quantum effects (Penrose, Hameroff)

Undiscovered property of information (Chalmers)

Information-theoretic complexity/systematicity tradeoff of a system (Edelman, Tononi, Koch)

Mysterianism (Colin McQuinn)

Counterexamples to "Φ" demonstrated by Scott Aaronson

Delusionism (Blackmore)

Hyperpropositions (with "signal sites")

Property theory (based on Meinard Kuhlman, Sci. Am. Aug. 2013)