Consciousness and Groundedness

15-494 Cognitive Robotics
David S. Touretzky & Ethan Tira-Thompson

Carnegie Mellon
Spring 2010
What is Consciousness?

• A philosophical swamp!

• Phenomenology: what is the sensation “red”?  
  – Qualia: sensations, like “red” or “sweet smelling”.  
  – “The way things seem to us”

• What is it “like” to have mental states, e.g., to see a sunset as “red”?  
  – Explanation in terms of retinal receptors is insufficient.  
  – Nagel: “What is it like to be a bat?” (echo-location)

• The Mind/Body Problem: how can physical matter (the brain) give rise to mental states?
Dualism

- Descartes: mind (spirit) is separate from body.
- Politically expedient: allowed study of the body (including perception and action) without threatening religious leaders concerned with spirit.

Materialism

- The doctrine that mind is just a phenomenon of the body, i.e., mental states = neural states.
- Is it really just that mechanical? Some people hope not.
Aspects of Consciousness

• Awake
  – Altered states of consciousness: sleep, dreaming, trance, ...

• Self-aware
  – All great apes except gorillas pass the mirror test.

• Self-knowledge
  – Able to describe one's own beliefs and motivations.

• Introspection
  – Ability to examine one's own mental states or “thoughts”.
  – Not infallible, but still useful.

• Internal monologue?
  – Having a mental language? (What about animals?)
 Phenomenological vs. Access Consciousness

• Phenomenological consciousness: sensing the environment.

• Access consciousness: having a “thought” about something. The thought can then be referred to in other thoughts.

• P-consciousness without A-consciousness: hearing a sound but paying no attention to it.

• A-consciousness requires thought; P-consciousness does not. (Are animals only P-conscious?)
“Higher Order Thought”  
Theory of Consciousness

• Consciousness as a property of mental states means consciousness *of* mental states.

• Consciousness is the ability to have thoughts about your thoughts.

• But what if some mental states can be experienced but aren't describable by “thoughts”?

• What qualifies as a “thought”? 
Neurophysiological Correlates of Consciousness

• Is consciousness localized in the brain?
  - May be distributed throughout.
  - Lesions to intralaminar nuclei of the thalamus cause loss of consciousness. ILN projects widely to cortex.

• How do anesthetics induce unconsciousness?
  - Decoupling of cortical areas.
  - Reduction in cortical activity.

• Are there “consciousness neurons” in the brain?
  - If yes, where are they?
  - If no, then does every neuron contribute to consciousness?
Unconscious Cognition

- Blindsight

- Tachistoscopic experiments

- Priming effects, e.g., “dealer” → (“deck” = card deck)

- Dorsal visual pathway (“where” stream) may be purely perceptual; ventral (“what”) stream involves cognition.

- Learned fear reaction (amygdala)
Can Robots Be Conscious?

- Similar to another famous question:
  - Could a computer ever “think”? 

- Turing test (the imitation game).
  - Can a human observer reliably discriminate a person from a machine, based on a written conversation?

- Weak vs. strong AI:
  - Weak AI: develop algorithms that allow computers to perform tasks currently considered to require “intelligence”.
  - Strong AI: get computers to be intelligent.
Searle's Chinese Room

- Searle doesn't understand a word of Chinese.
- Does the “Searle + room system” understand Chinese?
- Could the room be “conscious”? 

http://www.unc.edu/~prinz/pictures/c-room.gif
Kuipers' “Trackers” Proposal Concerning Consciousness

- Focuses on phenomenological consciousness.
- Says nothing about access consciousness.

**Basic Idea:**

- We experience the world as a rich high bandwidth stream of sensory impressions.
- A “tracker” monitors some feature of the environment over time. Allows us to be “aware” of the feature.
- Conscious experience is derived from trackers.
- Attention works by controlling trackers.
Kuipers' Trackers

- $x(t) = \text{body state}$
- $w(t) = \text{world state}$
- $z(t) = \text{sensor stream}$
- $u(t) = \text{motor stream}$
- $m(t) = \text{internal symbolic state}$
- $m_k(t) = \text{state of tracker } \tau_k$

- $F(x,w,u) = \text{how the world and body are updated}$
- $G(x,w) = \text{how the world and body are sensed}$
- $H_i(z,m) = \text{i}^{th} \text{ control law}$
Trackers and Searle's
11 Features of Consciousness

• **1. Qualitiveness:**

  Every conscious state has a qualitative feel to it... [This includes] conscious states such as feeling a pain or tasting ice cream... [and also] thinking two plus two equals four.” (Searle 2004)

• “The vividness, intensity, and immediacy of subjective experience are due to the enormous information content of the sensor stream $z(t)$.” (Kuipers 2005)

• Trackers provide structure, and rapid access to parts of the sensory stream.
  - Remembering “red” (rough symbolic label) vs. seeing a particular shade of red in a sunset.
Searle's Features (cont.)

• 2. Subjectivity:
  - “Because of the qualitative character of consciousness, conscious states exist only when they are experienced by a human or animal subject.” (Searle 2004)

• Consciousness is experienced exclusively from a first-person point of view.

• What this means: agent has privileged access to the sensor and motor streams of its own body, \( z(t) \) and \( u(t) \).

• The body is physically embedded in the world, so these streams have causal connections to the world.

• But couldn't a robot have a “point of view”? 
Searle's Features (cont.)

3. Unity:

We experience the audio-visual surround as a single unified field, continuous in space and time.

Our actual sensory stream is not so unified.
  - Visual acuity is low outside of the fovea.
  - Multiple saccades are necessary to “see” a scene.

Dennet's “multiple drafts” model of consciousness: unity and sequentiality are carefully maintained illusions.
“Cartesian Theater” vs. Multiple Drafts Theory

- Daniel Dennett describes conventional theories of conscious experience as being like a “Cartesian theater”:
  - Events play out in strict sequence and are perceived by an inner observer.
  - But who is looking at the play?

- Some psychophysical experiments indicate that sequentiality is **not** always maintained,
  - Color phi effect
  - Flash ring effect

- The mind doesn't “observe” reality, it **constructs** it.
Color Phi Effect

• “Moving” dot appears to change color in mid-flight:

   \[
   \begin{array}{c}
   \text{Red} \\
   \text{Green}
   \end{array}
   \]

• How does the brain know at time $t=75$ ms that the dot will change color at time $t=150$ ms?
Flash Lag Effect

- A flash at the center of a moving ring is perceived to occur offset from the ring.

- Motion channel faster than intensity channel?

- Online demo: www.michaelbach.de/ot/mot_flashlag1
Groundedness

- Percepts aren't arbitrary signals.
- They are about something: the relationship of the perceiver (body and brain) to the world.
- They are causally connected to the world.
- Symbols in the Chinese room are not grounded.
- Some say computers cannot “think” because their symbols are not grounded.
- Is groundedness important for consciousness?
Groundedness (cont.)

- Computers programmed to “notice” certain sensory signals might as well be performing arbitrary operations.

- But can robots, situated in bodies, acquire a repertoire of encodings that reflect their interactions with the world, and are thus grounded in experience?

- Kuipers: to discover abstractions for sensorimotor interactions, need to detect invariants.

- Example: if you turn a full 360°, the world looks the same afterwards.
Spatial Semantic Hierarchy

- Find distinctive places in the world, that can be reached by hill-climbing. Examples: corners, branch points.
- Find control laws that connect distinctive places, e.g., by wall-following.
- Construct topological graph reflecting this.
Selecting Control Laws

Diagram:

- At distinctive state
  - Select TF law
  - Reach fixed-point
  - Hill-climbing
  - Select HC law
  - Enter place neighborhood
  - Termination detected
  - Trajectory-following
Trajectory-Following Laws
Learning Actions

Sensorimotor Level

Control Level

Causal Level

Discrete Abstract Interface

< Vi, Aij, Vk >
Exploring A Simple World

- (a) random wandering
- (b) open-loop homing and path following: use actions that change one feature while keeping another relatively constant
- (c) closed-loop control laws can actively reduce deviations in the constant feature
Implications for Tekkotsu

• The notion of “tracking” would seem to be useful for maintaining continuity of attention across actions.

• Visual target tracking (with the Lookout) is in some ways analogous to Kuipers' tracker notion.

• What's missing?
  – Sensory memory storing recent perceptions (500 msec?) How do we know when things have changed?
  – Thoughts about percepts (access consciousness)
  – Internal language.
  – Goals, plans, etc., etc., etc.