Spatial Awareness and Fringe Consciousness

ASSC 10 Poster

Murray Shanahan
Imperial College London
Department of Computing
Overview

- This work synthesises several ideas in order to account for conscious spatial awareness
  - Global workspace theory (Baars)
  - The simulation hypothesis (Cotterill, Hesslow)
  - Fringe consciousness (James, Baars, Mangan)
  - Distributing spatial structure over time

- Some aspects of the theory have been realised in a computer model
  - See *Consciousness & Cognition* article (Shanahan, 2006)
  - Some newly implemented features making debut here
Spatial Awareness

Our conscious awareness of space is not only of what we can directly see and touch.

It also encompasses regions of space we know are out there but that we cannot perceive directly:
- The space around the next corner
- The space behind our heads
- The other side of a solid object

How can we account for this aspect of our phenomenology?
Fringe Consciousness

- The idea that conscious experience has a focus-fringe structure dates back (at least) to William James.
- The conscious fringe contains *hints* of material that could be brought into focal consciousness.
- Spatial awareness can be viewed in terms of focus and fringe.
  - Surfaces and spaces we can perceive directly appear in focal consciousness.
  - The fringe hints at those regions of space we have no direct perceptual awareness of but know are there.
Trajectories In Sensorimotor Space

- What is this fringe-borne awareness of a region of space outside the range of direct perception?
- We can think in terms of potential trajectories through sensorimotor space
- The fringe hints at possible motor activity and the expected sensory outcome of that activity (cf: Noë & O’Regan’s sensorimotor contingencies)
  - Looking behind you
  - Turning an object around
The Structure of Sensorimotor Space

- Physical space constrains sensorimotor space
  - Reversible actions
  - Cyclic paths

F = front of block
B = back of block
Global Workspace Architecture

- Multiple parallel specialist processes compete and co-operate for access to a global workspace (GW).
- If granted access to the GW, the information a process has to offer is broadcast back to the entire set of specialists.
The Sensorimotor Fringe

To conform to global workspace theory, the conscious fringe must be broadcast.

Here the fringe contains hints of possible trajectories in sensorimotor space.
Space in the Imagination

- Spatial constraints govern not only our conscious awareness of what is *actually* before us, but also our imagination.
- The fringe hints not only at those trajectories through sensorimotor space we could immediately follow, but also at those we could rehearse (*simulate*) prior to following them.
- There are now three elements we need to combine:
  - Global workspace architecture
  - Inner rehearsal (simulation)
  - Fringe consciousness
The Simulation Hypothesis

“Thought is internally simulated interaction with the environment” (Cotterill, 1998)

- Internal simulation is integral to our inner life and facilitates anticipation and planning
- Three assumptions (Hesslow, 2002)
  - The brain’s motor centres can be active without producing overt action
  - The brain’s perceptual apparatus can be active without the presence of external stimuli
  - Internally generated motor activity can elicit internally generated perceptual activity through associative mechanisms
A brain-inspired architecture that combines an internally closed sensorimotor loop (the “core circuit”) with an affect-based mechanism for action selection.
Simulation with a Global Workspace

- Competition and broadcast can be modeled with a cascade of attractor networks
- The GW and the competing cortical populations each define an attractor landscape
- Each attractor abstracts a set of sensory states
- Internal simulation visits a sequence of attractors

Dominant cortical population nudges GW into a new attractor → Pattern of GW activation influences competing cortical populations → Competing cortical populations are nudged into new attractors
The “core circuit” combines an internal sensorimotor loop with mechanisms for broadcast and competition, and thereby marries the simulation hypothesis (Cotterill, Hesslow) with global workspace theory (Baars).
A Computer Model of the Core Circuit

- Based on G-RAM weightless neurons (Aleksander)
- Built with NRM modelling tool (Dunmall)
- Implementation comprises approximately 40,000 neurons with 3,000,000 connections
- Can be embedded in an action selection architecture and used to control a robot
  - Not directly relevant here, but see (Shanahan, 2006)
Weightless Neurons

- One-shot training
- Easily assembled into attractor networks
- Not very biologically plausible at a small scale
- But good for building large-scale network models
Sequences of Broadcast States

An internally generated procession of broadcast imagery, starting from a random initial state, with no retinal input
Encoding Spatial Structure in Time

- What sort of mechanism could realise a fringe-borne awareness of spatially constrained sensorimotor contingencies?
- Spatial structure can be distributed over time
- Instead of *all* the sensorimotor contingencies being present in the fringe together, they can appear *one at a time*
- Phenomenologically this would still feel like seamless awareness of a continuum
A curious phenomenon in my computer model of the core circuit is the transitory appearance of faint subsidiary attractors in the GW overlaid on the dominant attractor.

With the addition of some extra neural pathways, we can obtain a pulsing sequence of such faint attractors.

Since these faint attractors are broadcast alongside the dominant attractor, this is a possible mechanism for realising the fringe.
Conscious thought has the property of *systematicity*

The elements of conscious thought can be arbitrarily recombined
- If I can imagine a red ball to the right of a green cup then I can also imagine a green ball to the left of a red cup

How can we accommodate systematicity within the simulation hypothesis?

It is easier to envisage mechanisms that substitute one element for another in a *temporal series* than in a spatial arrangement
