From Human to Machine Consciousness

DAVID GAMEZ

Department of Computer Science, Middlesex University, UK

10th Polish Philosophical Congress, Poznań
18th September 2015

Talk Overview

• Consciousness and external behaviour.
• Science of human consciousness:
  • Measurement of human consciousness.
  • Measurement of the physical world.
  • Theories of consciousness.
• Machine consciousness.
• Conclusion.

Consciousness and External Behaviour

Attribution of Consciousness

• Recent films on consciousness in artificial systems:
  • Ex_Machina (2015)
  • Transcendence (2014)
  • Her (2013)
  • Chappie (2015)
• We attribute consciousness to these systems because of their external appearance and behaviour.

Recent Films about Machine Consciousness

What are we Attributing Consciousness to?

• The robots’ bodies and the phone’s speaker are just simple mechanics.
• We are attributing consciousness to the systems that control the robots’ bodies and the phone’s speaker:
  • Blue egg thing (Ex_machina).
  • On board computer (Chappie).
  • Cluster of computers (Her, Transcendence).
Conscious Systems

Troubles with Functionalism
- Many different systems can produce external behaviour (Block 2006).
- For example:
  - Babbage’s Analytical Engine.
  - Human brain.
  - Population of China communicating with radios and satellites.
  - Giant lookup table.
  - Etc.

Functionalism and Panpsychism
- Any system can be interpreted as implementing a given computation (Putnam 1988, Bishop 2002, 2009):
  - Map program onto finite state automata.
  - Map finite state automata onto state transitions of physical system.
- The state transitions of any system could be used to produce the behaviour of Ava (Ex_Machina), Chappie and Samantha (Her):
- A given physical system (rock, plant, computer, etc.) would be attributed all possible conscious states.

Failure of Functionalism
- Attributions of consciousness based on external appearance and behaviour do not work.
- Simplistic induction from our own experience: We attribute consciousness to systems that look like us.
- Need a better approach to machine consciousness.
We are interested in the consciousness of machines because we are conscious – we are immersed in a world of colours, smells, sounds etc. So we want to know if other systems are conscious as well.

We are certain that we are conscious. So we could study what is linked to consciousness in humans. And find out whether this is also present in machines.
Which Part of Us is Conscious?
- Our bodies are not conscious:
  - We are conscious when we dream.
  - Locked-in patients are conscious.
- Most people believe that consciousness is linked to brain activity.
- I have discussed sensorimotor theories elsewhere (Gamez 2014a).

Consciousness and the Brain
- Our conscious experiences are linked to patterns, particles, forces, etc. in the brain.
- We could identify these patterns, particles, forces, etc.
- And use this knowledge to make inferences about the consciousness of other systems.

Scientific Experiments on Human Consciousness
- Measure consciousness.
- Measure brain.
- Develop mathematical descriptions of the relationship between these sets of measurements.

Science of Consciousness
Measurement of Human Consciousness

- Consciousness is measured through first-person reports.
- This raises a number of philosophical problems.
- These can be handled with assumptions.
- The science of consciousness is considered to be true given these assumptions.

The Correlates of a Conscious State (CC set)

- A CC set is a set of spatiotemporal structures in the physical world that is correlated with a conscious state.
- Suspend judgement about the relationship between CC sets and conscious states (identity, dualism, epiphenomenalism, etc.).
- CC sets are functionally connected to conscious states.

Assumptions

1. The consciousness associated with a human brain is functionally connected to its reports.
2. All conscious states associated with a human brain are available for report and all aspects of these states can potentially be reported.

Assumptions for the Measurement of Consciousness

1. The consciousness associated with a human brain is functionally connected to its reports.
2. All conscious states associated with a human brain are available for report and all aspects of these states can potentially be reported.

Assumptions

3. The conscious state associated with a CC set nomologically supervenes on the CC set. In our current universe physically identical CC sets are associated with indistinguishable conscious states.
4. CC sets cause a person’s first person reports about consciousness.

Assumptions for the Measurement of Consciousness

- Consciousness is functionally connected to reports about consciousness (R).
- All conscious states are available for reports about consciousness (A).

Consciousness associated with a human brain.
Consciousness nomologically supervenes on the human brain.
"I am conscious of a red hat".
Describing Conscious States

- Consciousness cannot be described in natural language:
  - Context-bound
  - Ambiguous
  - Not applicable to infants, bats, robots, etc.
  - Not mathematically tractable.

C-description

- Need a precise formal way of describing consciousness that is applicable to any system.
- This will be referred to as a c-description.
- Possible methods include:
  - XML/LMNL
  - High dimensional qualia
  - Category theory

Information and Computation

- Correlates of consciousness are objective properties of the physical world.
- Information and computation are subjective properties (Gamez 2014b, Gamez 2014c).
- No informational or computational correlates of consciousness.
- CC sets consist of spatiotemporal patterns in particular physical materials.

Measurement of the Physical World

- The measured object interacts with a calibrated object.
- Observe result and extract a number.

Description of the Physical World

- The number that is extracted through a measurement procedure is attributed to an object in the physical world.
- 3 metres is the height of an elephant.
- Objects are tightly defined in physics and chemistry.
- They are not tightly defined in biology.
**P-description**
- We want a science of consciousness that can make predictions about the consciousness of artificial systems.
- Neurons are not precisely defined.
- A science of consciousness based on biological neurons will not be able to say anything about the consciousness of systems based on synthetic neurons.
- Need a precise formal description of the spatiotemporal structures that form CC sets.
- Will be referred to as a *p-description*.

**Theories of Consciousness**
- A theory of consciousness (c-theory) is a mathematical description of the relationship between c-descriptions and p-descriptions.
- It can generate c-descriptions from p-descriptions.
- It can generate p-descriptions from c-descriptions.

**C-theory**

**Example: Tononi’s Information Integration Theory of Consciousness**
- For example, see Tononi (2008).
- Closest thing to a c-theory that we have so far.
- A mathematical algorithm links a description of the physical world to a description of consciousness.
- A conscious state (a quale) is c-described using a high dimensional mathematical structure.
**Deductions about Consciousness**

- Suppose the science of consciousness has developed a c-theory that can reliably map between:
  - P-descriptions of spatiotemporal patterns in the physical world.
  - C-descriptions of conscious states.
- We could use this c-theory to make deductions about the consciousness of computers and robots.

**C-theory Deduces Machine’s Conscious State**

1. Measure state of robot or computer.
2. Convert measurement into p-description.
3. Use c-theory to convert p-description into c-description of system’s consciousness.

**Construction of a Conscious Machine**

- Similar procedure can be used to design and build an artificial consciousness:
  1. Generate c-description of the consciousness that you want to create in the machine.
  2. Use c-theory to convert c-description into p-description of a CC set.
  3. Realize the CC set in an artificial system.

**Applications**

- Deduce how to build a chip that extends or repairs a person’s consciousness.
- Deduce how to upload a copy of my consciousness into a computer.
Uploading Consciousness into Computer

Transcendence

Conclusion

External Appearance and Behaviour
- Attribution of consciousness based on external appearance and behaviour does not work.
- Almost any system can be used to generate a given set of external behaviour.
- Leads to an untenable panpsychism.

From Human to Machine Consciousness
- Study human consciousness.
- Develop mathematical c-theories that map between p-descriptions and c-descriptions.
- Use c-theories to make predictions about machine consciousness.

Acknowledgements
- Many thanks to Barry Cooper and the John Templeton Foundation for supporting this work (Project ID 15619: Mind, Mechanism and Mathematics: Turing Centenary Research Project).
- I would also like to thank Anil Seth, the Sackler Centre for Consciousness Science and the Department of Informatics at the University of Sussex for hosting me as a Research Fellow during this project.

References
More Information

- Slides: [www.davidgamez.eu/talks](http://www.davidgamez.eu/talks)
- Some papers related to this material: [www.davidgamez.eu/publications](http://www.davidgamez.eu/publications)
- Feel free to contact me if you have any questions: [david@davidgamez.eu](mailto:david@davidgamez.eu)