Lectures on Machine Consciousness – Madrid February 2011

Machine Consciousness is a seminal field placed at the crossing between technical disciplines (AI, Robotics, Computer Science and Engineering), theoretical disciplines (Philosophy of Mind, Linguistic, Logic), and empirical disciplines (Psychology and Neuroscience). Machine consciousness focuses on attempts to apply the methods of AI, robotics and computer science to various ways of understanding consciousness and to examine the possible role of consciousness in AI systems. On one hand there is the hope that facing the problem of consciousness would be a decisive move to design better AI systems, on the other hand the implementations of AI systems could be helpful for understanding consciousness.

Epigenetic robotics, situated and synthetic AI approaches, embodiment, developmental systems, anticipatory systems and human-robot interactions, have pointed out in many ways the importance of the interaction between the brain, the body and the surrounding environment. Machine consciousness is a research theme aimed to an unified view of these approaches.

The lectures serve the theme of interdisciplinary in AI. In particular, the aim of the tutorial lectures is to motivate and explain researches on a new topic of emerging importance for AI.

The lectures will highlight that machine consciousness is a research theme that searches for an unified view of these topics that the AI community deeply discussed during the years.

Well known journals as IEEE Spectrum, AI and Medicine, Neural Networks, Neurocomputing, dedicate issues to topics related with machine consciousness.

The recent 2007 AAAI Fall Symposium on AI and Consciousness (co-organized by the tutorial presenter) received great attention from the audience: more than 50 people attended the lively and focused symposium (AAAI Fall Symposium Reports, AI Magazine 29, 1, pp. 99-100).

Moreover, many papers from the 2008 and 2009 editions of the AAAI Fall Symposium on BICA (the tutorial presenter was in the Core Program Committee of both events) concerns machine consciousness and AI.

The lectures will present the current state of research and will discuss both the theoretical foundations and the experimental result of the emerging field of machine consciousness and their relationships with Artificial Intelligence.

The lectures will be divided in three main parts:
1) theoretical and philosophical issues of consciousness,
2) models of machine consciousness,
3) case studies and implemented systems.

Outline of the lectures

Introduction: features of consciousness

- *Searle, J. (2005): Mind, A Brief Introduction, (Oxford University Press).*

Theoretical and philosophical issues of consciousness

Classification of theories of consciousness: anti-reductionism, reductionism, eliminativism.

Anti-reductionism: consciousness cannot be reduced as a neurobiological entity

Subjective character of experience, not captured by any mental analyses

- *Nagel, T. (1974), 'What is it like to be a bat?' Philosophical Revue, 83, pp. 435–50.*

Phenomenal consciousness (P-Consciousness) and Access consciousness (A-Consciousness)

- *Block, N. (1995), 'On a confusion about a function of consciousness', Behavioral and Brain Sciences, 18, No. 2, pp. 227–87.*

"Easy" and "hard" problems of consciousness; the philosophical zombie.

- *Chalmers, D. J. (1996), The Conscious Mind: In Search of a Fundamental Theory (Oxford University Press).*

Reductionism: consciousness is a neurobiological entity

Local vs. global theories

Neural Correlates of Consciousness (NCC); a framework for consciousness

- *Koch, C. (2004), The Quest for Consciousness (Engewood, CO: Roberts and Co.).*

The reentrant pathways; the dynamic core; information integration

- *Edelman, G. and Tononi, G., (2001), A Universe of Consciousness: How Matter Becomes Imagination, (New York: Basic Books).*
- *Tononi, G. (2004), An information integration theory of consciousness, BMC Neuroscience, 5:42*

The Global Workspace Theory

- *Baars, B., (1997), In the Theater of Consciousness: The Workspace of the Mind, (Oxford University Press)*

Eliminativism: consciousness is intended as a neurobiological entity, but in facts there is no such entity

Dennett's theory: Heterophenomenology; Contrasting the Cartesian theater; Virtual machines in the brain; Multiple draft model.

- *Dennett, D. (1993), Consciousness Explained (London: Penguin).*

Models of Machine Consciousness

Axioms for the presence of consciousness in agents

- *Aleksander, I. & Dunmall, B., (2003), Axioms and Tests for the Presence of Minimal Consciousness in Agents, Journal of Consciousness Studies, 10, No. 4–5, pp. 7–18.*

Virtual machine functionalism; synthetic phenomenology

- *Sloman A. & Chrisley, R. (2003), Virtual Machines and Consciousness, Journal of Consciousness Studies, 10, No. 4–5, pp. 133–72.*
- *Chrisley, R.. (2009), Synthetic Phenomenology, International Journal of Machine Consciousness, 1, pp. 53-70.*

Logical models of machine consciousness

- *McCarthy, J. (1995), Making Robots Conscious of their Mental States, http://www-formal.stanford.edu/jmc/consciousness.html*
- *Schubert, L. (2005), Some Knowledge Representation and Reasoning Requirements for Self-Awareness, in: Anderson, M. & Oates, T., Metacognition in Computation, Menlo Park, Ca: AAAI Press vol. SS-05-04.*

Attention and control for machine consciousness

- *Koch, C. (2004), The Quest for Consciousness (Engewood, CO: Roberts and Co.).*
- *Taylor J. G. (2007) CODAM. A Neural Model of Consciousness. Scholarpedia 2(11):1598*
- *Sanz, R., Lopez, I., Rodriguez, M. & Hernandez C. (2007), Principles for Consciousness in Integrated Cognitive Control, Neural Networks, 20, pp. 938-946.*

Machine consciousness, multiple layers and higher order theories (HOT)

- *Minsky, M. (2006), The Emotion Machine (Simon & Schuster)*
- *McDermott, D. (2001), Mind and Mechanism (Cambridge, MA: MIT Press).*

Machine consciousness and the "hard problem"

- *Kuipers, B. (2008), Drinking from the Firehose of Experience, Artificial Intelligence in Medicine, 44, pp.155-170.*

Case studies and implemented systems

Robots inspired to machine consciousness

- *Holland, O., Knight, R. & Newcombe, R. (2007), The Role of Self Process in Embodied Machine Consciousness, in: Chella, A. & Manzotti, R., Artificial Consciousness. Exeter: Imprint Academic.*
- *Chella, A. & Macaluso, I. (in press)(2008), The Perception Loop in CiceRobot, a Museum Guide Robot, Neurocomputing, vol. 72, pp. 760-766.*

Implementations of the Global Workspace Theory

- *Baars, B., Franklin, S. (2009), Consciousness is Computational: The LIDA Model of Global Workspace Theory, International Journal of Machine Consciousness, 1, pp. 23-32.*
- *Shanahan, M. (2006), A cognitive architecture that combines internal simulation with a global workspace, Consciousness and Cognition 15, pp. 433–449.*

Cognitive architectures for machine consciousness

- *Chella, A., Frixione, M. & Gaglio, S. (2008) A Cognitive Architecture for Robot Self-Consciousness, Artificial Intelligence in Medicine, vol. 44, pp. 147-154.*
- *Haikonen, P. (2007) Robot Brains: Circuits and Systems for Conscious Machines, (Wiley-Interscience).*

Measures of consciousness in humans and artifacts

- *Koch, C. & Tononi, G. (2008) Can Machines Be Conscious? IEEE Spectrum, June 2008.*
- *Tononi, G. (2008) Consciousness as Integrated Information: a Provisional Manifesto, Biological Bulletin 215, pp. 216-242.*
- *Arrabales, R., Ledezma, A. & Sanchis, A. (2010) ConsScale – A Pragmatic Scale for Measuring the Level of Consciousness in Artificial Agents, Journal of Consciousness Studies, 17, pp. 131-164.*

Discussions and perspectives of machine consciousness

- *Manzotti, R. & Tagliasco, V. (2008) Artificial consciousness: A discipline between technological and theoretical obstacles, Artificial Intelligence in Medicine, 44, pp. 105-117.*
- *Harnad, S. & Scherzer, P. (2008) First, Scale Up to the Robotic Turing Test, Then Worry About Feeling, Artificial Intelligence in Medicine, 44, pp. 83-89.*
- *Horgan, J. (2008) The Consciousness Conundrum IEEE Spectrum, June 2008.*

Resources: books, tutorials, journals, conferences and workshops