

An anthropomimetic robot with imagination: one step closer to machine consciousness?

Hugo Gravato Marques, Rob Knight, Richard Newcombe and Owen Holland

*Department of Computer Science
University of Essex
Colchester, CO4 3SQ, UK
hgmarq@essex.ac.uk rrknight@gmail.com
r.a.newcombe@gmail.com owen@essex.ac.uk*

Abstract

The possible role of the self-model in consciousness has received much attention recently, particularly in connection with the theories of Metzinger [11] and Damasio [4]. However, a self-model can also play a variety of useful functional roles within a system; indeed, it may well be that the evolutionary history of the self-model at the core of consciousness may have begun in the context of one of these functional roles. This paper explores the use of a self-model in the plausible candidate context of providing better action selection through the imagination and evaluation of possible self-world interactions; we call this 'functional imagination'.

Functional imagination can roughly be described as the purposeful manipulation of information that is not directly available to the senses - references to imagination always point to something that in reality is not there. One important subjective aspect of imagination is the phenomenon of imagery. This refers to the pseudo-sensory experience we have when we recall a scene from memory, when we imagine something we have never experienced, or when we imagine a future action. This raises a key question: since during imagination the world is not providing us with these sensations, what sort of mechanism is doing it for us? A good candidate for answering the question is the observation that, when imagining or recalling something, humans (and some animals) appear to use the same sensorimotor machinery that they use for real perceptions and actions; the similarity between the real and imagined sensations is therefore quite understandable.

However, in order to be useful, imagination also needs to provide us with information about the likely consequences of actions, and so a mechanism is needed for translating imagined motor actions into sensory-based representations of their consequences. In [6], Hesslow proposes that some associative mechanism performs this function, which he calls anticipation; more generally, the imagined motor actions must serve as the input into a suitable forward model, the output of which is the appropriate activation of the sensory areas. The idea that the brain uses models for prediction is not new [3], but it seems to have been reinvigorated by the recent rise of embodiment-inspired theories of cognition (see [4], [1], [5]).

One of the goals of our research is to develop an understanding of functional imagination through the design and implementation of a variety of different architectures (see [8], [9]). In this talk we describe a successful implementation of a typical architecture on CRONOS, a robot that is probably one of the most extreme attempts to imitate the human skeleto-muscular system [2] [7]. As well as being structurally complex, with 42 degrees of freedom, CRONOS uses elastically driven actuators with similar properties to muscles, and therefore has extremely complex dynamics, making it very difficult to control. The challenge of developing a forward model of CRONOS and its environment that accurately simulated the consequences of a given motor activation, and represented those consequences in sensory-based terms, was dealt with by using physics based techniques to build SIMNOS [10], a detailed model of CRONOS which captures both its geometry and its dynamics. SIMNOS operates within a physics based model of the environment, ensuring the predictive value of any interactions. By visually rendering both SIMNOS and the environment, the outcomes of interactions are made available in sensory-based terms. In the experiments, CRONOS is required to knock down an object placed on a table. Instead of finding the appropriate action through overt trial and error, CRONOS uses the internal model (SIMNOS) for simulating possible actions, and uses its sensory system to evaluate their outcomes. Once a suitable action is found in simulation it is executed overtly.

We can now claim to have successfully implemented a model of functional imagination on a robot with a degree of structural complexity approaching that of a human. It may now be possible to use this platform to

study the further development of the robot's self-model in directions relevant to the emerging discipline of machine consciousness.

References

- [1] Lawrence Barsalou. Perceptual symbol systems. *Behavioral and Brain Sciences*, 22:577–660, 1999.
- [2] Francesco Berton, Lorenzo Jamone, Giulio Sandini, and Giorgio Metta. Accurate control of a human-like tendon-driven neck. In C. Grimes, E. Dickey, and M. Pishko, editors, *The Encyclopedia of Sensors*, volume X, pages 1–16. American Scientific Publishers, 2005.
- [3] Kenneth Craik. *The Nature of Explanation*. Cambridge University Press., Cambridge, England, 1943.
- [4] Antonio Damasio. *The Feeling of What Happens*. Vintage, New York, 1999.
- [5] Rick Grush. The emulation theory of representation - motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27:377–442, 2004.
- [6] Germund Hesslow. Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Science*, 6(6), 2002.
- [7] Owen Holland and Rob Knight. The anthropomimetic principle. In Jeremy Burn and Myra Wilson, editors, *Proceedings of the AISB06 Symposium on Biologically Inspired Robotics*, 2006.
- [8] Hugo Marques and Owen Holland. Architectures for embodied imagination. *Neurocomputing*, 2008. In press.
- [9] Hugo Marques, Owen Holland, and Richard Newcombe. A modelling framework for functional imagination. In *Proceedings of the AISB08 Symposium on Computing and Philosophy*, 2008.
- [10] Hugo Marques, Richard Newcombe, and Owen Holland. Controlling and anthropomimetic robot: A preliminary investigation. In *Proceedings of ECAL2007.*, Lisbon, 2007. Springer Verlag.
- [11] T. Metzinger. *Being No One: The Self-Model Theory of Subjectivity*. MIT-Bradford, 2003.