Towards mental life as it could be: a robot with imagination

Hugo Marques[1], Owen Holland[1], Rob Knight[2] and Richard Newcombe[3]

[1]University of Essex
[2]The Robot Studio
[3]Imperial College
hgmarq@essex.ac.uk

"Artificial life is the study of man-made systems that exhibit behaviors characteristic of natural living systems." So wrote Chris Langton two decades ago, and most people within the discipline still seem happy with his definition. But from the human perspective, the behaviour of our own living system is only half the story, because what matters to most of us is not our behaviour, but our experience — our mental life. Perhaps it is now time for artificial life to embrace the challenge of creating man-made systems that have experiences 'characteristic of natural living systems' and giving an account of 'mental life as it could be'.

This talk will describe some of our recent work within the area known as machine consciousness. Several modern theories of consciousness, notably those of Damasio and Metzinger, stress the importance of embodiment for consciousness, and identify a body-centred self-structure as being a key element in the origin and support of phenomenal experience. In the CRONOS project, we reasoned that human consciousness — the only kind of which we have knowledge — must require both a human-like body, and a human-like body model, and so we set out to construct suitable entities. The physical robot, CRONOS (Holland and Knight 2006), is a new kind of robot: it is anthropomimetic, faithfully copying the human skeleton, and equipped with appropriately placed elastic muscles and tendons. The body model, SIMNOS (Gamez et al. 2006), is an accurate physics based simulation that behaves sufficiently similarly to CRONOS to be controlled by the same control system.

For a self-model to have evolved, it must confer some functional advantage on its host. We are investigating the hypothesis that one useful function might be the ability to support imagination by predicting the outcomes of possible actions through internal simulation, more or less along the lines suggested by Hesslow (2002). Simulating the body and its controller is only the first step; it is also necessary to create an internal representation of the real world that interacts with the body sufficiently accurately to have predictive value. A physics based internal representation meets this requirement, but the problem of creating such a representation goes well beyond the traditional AI/computer vision challenge of achieving correct geometry, since the objects in the modelled world must also have the correct physical properties. We will report on the use of SLAM techniques and cross-modal integration for dealing with this problem.

In order to obtain functional benefit from imagination, it is not enough merely to have predictively valid models of the body and of the world; it is also necessary to embed the models within a suitable architecture to identify when imagination is required, to generate appropriate candidate actions, to evaluate the success, failure, and costs of the actions, to prevent candidate actions from being acted out by the real robot, to select the preferred action, and to execute the preferred action. We have developed a task-dependent taxonomy of these architectures, and have implemented and tested the ones that seem most important (Marques and Holland 2008). For the first time, this talk will describe and demonstrate an anthropomimetic robot using an architecture for embodied imagination, including a physics based internal model of itself and the world, to select and execute an appropriate action; we believe this may be a significant step towards building a robot with a form of mental life.