Melhuish, C.; Ieropoulos, I.; Greenman, J.; and Horsfield, I. 2006. Energetically autonomous robots: Food for thought. *Autonomous Robots*. 21: 187 – 198.

The perception-…-action cycle cognitive architecture and autonomy: the view from the brain

Vassilis Cutsuridis VCUTSURIDIS@GMAIL.COM
Division of Engineering
Kings College London
Strand, WC2R 2LS
U.K.

Autonomous systems with reasoning capabilities are systems able to perceive their environment and act on it by performing complex tasks automatically. Autonomous systems are also able to adapt to unforeseen operating conditions or errors in a robust and predictable manner without the need for human guidance, instructions or programming. To accomplish such complex feats they must master the powers of perception, recognition, attention, learning and memory, cognitive control, reward and motivation, decision making, affordance extraction, action planning and action execution (step 1). Once these powers are successfully mastered, then these systems may be embodied into a robot able to act in the real world (step 2). Their embodiment, however, cannot guarantee that these systems will be able to operate autonomously in the environment as they will still need to solve the issues of the real-time system operation, resource management and meta-learning (step 2).

In their article "Cognitive architectures and autonomy: a comparative review" Thórisson and Helgasson reviewed a number of "autonomous" systems and architectures with general "cognitive" capabilities and compared and contrasted their performance in a hypothetical example of autonomous exploration of an environment by a robot. Instead of their criteria focusing on how the powers of perception, recognition, attention, memory, cognition, decision making and action planning and execution are achieved by these systems (step 1), the authors ignored these powers, and compared and contrasted the systems based on step 2's real-time processing, resource management, learning and meta-learning issues. The authors argued that the former functions (e.g. perception, recognition, attention, memory, etc.) are less important.

I believe that dealing with the issues of real-time processing, resource management, learning and meta-learning first and comparing and contrasting the reasoning capabilities of systems based on them is similar to building a house from the roof down. The systems are forced to solve the real-time system operations of functionalities which they have not deciphered yet, so they will inevitably be dumb, as they will be empty shells not possessing any reasoning powers that will enable them to go beyond the information provided.

Furthermore, though some of the reviewed systems are "biologically inspired" in that they depend on behavioral studies and test themselves by the replication of experimental behavioral data, none of these systems attempt reverse engineering of the brain circuitry that supports these behaviors. Reasoning is the highest faculty of the human brain and it depends on the majority of the brain components (perception, attention, learning and memory, decision making, action, etc.). The brain is a system that has evolved over a million or so years, so it is expected to provide a reasonably optimized solution to many of the cognitive tasks under consideration.

I propose as an alternative to the systems reviewed by the authors a brain-inspired cognitive control architecture for autonomous interaction of a robot with objects situated in its immediate environment (i.e. a form of exploration of an environment by a robot). My approach is based on work done in the EU-DARWIN project. A graphical representation of the cognitive control architecture is given in Figure 1 (Cutsuridis, 2012). The architecture proposes that exploring an environment requires to act upon objects in it, like in the case of vision-guided reaching and grasping of objects. The objects themselves are not to be known a-priori to the system, but their knowledge is built by the system through interaction and experimentation with them. The architecture is multi-modular, consisting of object recognition, object localization, attention, cognitive control, affordance extraction, value, decision making, motor planning and execution modules. The components of the architecture are novel as well as based on previous architectures (Cutsuridis et al., 2011; Cutsuridis, 2009; Taylor et al., 2009) and follow very closely what is currently known of the human and animal brain.

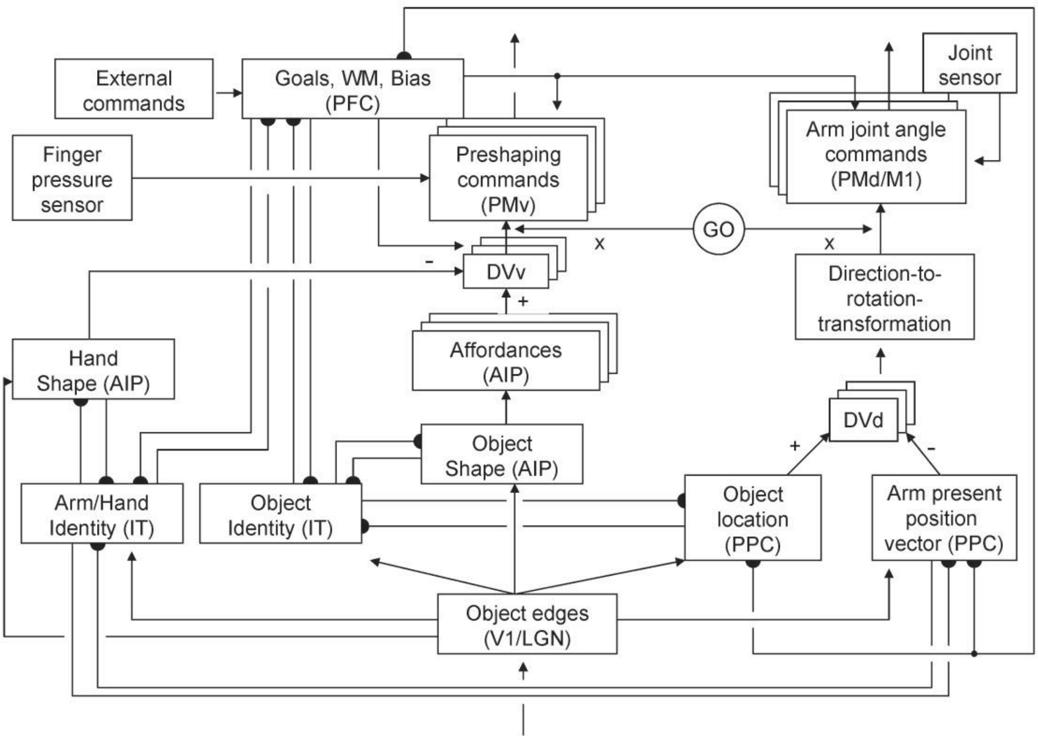

Figure 1. Graphical representation of the cognitive control architecture of object shape and object location recognition, attention reward, decision making, cognitive control, affordances, action planning and execution in reaching and grasping.

Vision-guided reaching and grasping involves two separate visuomotor channels, one for reaching and another one for grasping, which are activated in parallel by specific visual inputs and each channel controls specific parts of limb (arm and hand, respectively). An input image is processed in a bottom-up fashion, providing input to feature detectors, which in turn lead to the formation of visual maps (the *object* map and the *spatial saliency* map). Bidirectional cross-talk between object and spatial maps ensures that the object corresponds to the appropriate spatial location in the environment. The visual maps then activate the *cognitive control* map (goals, motivations, task constraints), which in turn feeds back to amplify the neural representations in the visual maps, which are relevant to the current context, and to suppress the irrelevant ones. Resonance between goals and object, and goals and spatial maps is achieved via a measure of degree of similarity, which depends on the amount of modulation (value) the maps receive from the dopamine system. A winner-take-all competition between resonated neural representations ensures that the object representation and spatial representation that reached resonance first will continue fastest processing first, followed by the second fastest and so on. Once an object and a spatial representation is selected a library of action plans are selected, one for reaching and the other one for grasping. Once again the cognitive control maps will select the action plan most relevant to the current context and suppress the irrelevant ones. The selected reaching and grasping motor plans will be gated by a GO signal (output of the basal ganglia), and form the final motor commands, which will be sent to the motor execution centers for execution. Visual and proprioceptive feedback will update the current arm position and fingers configuration towards the desired ones.

My cognitive control system has been implemented on the iCub robot with considerable success when multiple objects were situated in the environment and the robot had to recognize them, localize them, attend to each of them and reach and grasp them according to an externally dictated sequence of motor actions.

References

Cutsuridis, V.; Hussain, A.; and Taylor, J.G. 2011. *Perception-action cycle: Models, architectures and hardware*. New York, NY: Springer

Cutsuridis, V. 2009. A cognitive model of saliency, attention and active visual search. *Cognitive Computation*. 4(1): 292-299

Cutsuridis, V. 2012. A cognitive control architecture of the perception-action cycle for robots and agents. *Under review*.

Taylor, J.G.; Hartley, M.; Taylor, N.R.; Panchev, C.; and Kasderidis, S. 2009. A hierarchical attention-based neural network architecture, based on human brain guidance, for perception, conceptualisation, action and reasoning. *Image and Vision Computing* 27(11):1641–1657