

The Prince of Holmberg I

Fiora Pirri,
Sapienza, University of Rome, fiora.pirri@dis.uniroma1.it

May 10, 2007

1 Introduction

Alma is a cognitive robotics PhD student, she has been working with the vision system of the P3-DX active media platform and she docked to the land of consciousness studying and implementing a plausible attentional behaviour on her P3-DX. Sitting in the grass with the robot near her she is controlling how the robot is reproducing a scan path to finalize her thesis titled “Reproducing consciousness : from NCC to artificial symbols”. This is just one of the last titles in a sequence of thousands inappropriate titles. She is aware, now, that there is no way to lock up in few words what it is still a deep pond of hopeless hypotheses.

Her starting point is Koch exploration of the neuronal correlates of consciousness (NCC), and the conscious percept. According to Koch consciousness can be associated with elementary feelings and even reproducible in machine. On these grounds the whole research on the perceptual basis of consciousness can be accomplished searching and localizing the NCC, and providing a close understanding of the *minimal neuronal events and mechanisms jointly sufficient for a specific conscious percept. Where the NCC involves the firing activity of the forebrain.* Despite Alma full agree with all of Koch undertakings, the work with robot attentional behaviours led her to some criticisms that she is somehow reluctant to make explicit, so that we will not be able to read about that on her thesis. We shall get only some questioning aspects of her main criticisms concerning what she considers three flaws of the “NCC project”, as she refer to in her thesis.

The three main points that she identify as pivotal considering the NCC project from the cognitive robotics point of view concern the reproducibility of consciousness, that is: representation, reproducibility of experience and the space of experiences.

1.1 Representation

Representation is the ability to transform a sign from one space to another, so that the range space is used as a description of the domain space. Clearly the language is a form of representation, but there can be several forms and unlimited levels of representation. When a visual percept determines a phenomenal state that can be associated with consciousness, then the ability of describing the phenomenon intervenes, that is,

the ability to decipher the correlation between the state and a pre-existent model categorizing boundaries (when yellow becomes orange, far becomes close, an object becomes familiar, air becomes wind etc.) , preferences (interesting elements vs. distractors), causality (e.g. between fixations), and so on. These pre-existent models are learned (both by humans and robots) or simply accumulated along a lifetime, genetically inherited, absorbed by the contemporary culture, like aesthetic feelings, or are a simple train of pulses? What is the relation between the act of feeling and the act of deciphering its outcome? Is representation part of consciousness or it is its epi-phenomenon? After all perception amounts to transporting energy from one state to another and this transformation, allowing to distinguish the self, is a representation.

Alma thinks that observing, measuring and explaining brain functioning for localizing and classifying early consciousness is crucial for understanding the physics of awareness, but there are other aspects grounded on representation levels that need to be undertaken, at least to prove elementary consciousness in robots behaviours. The steps she has taken to make the P3-DX learn how to mimic attention and move its eyes (and head) towards specific targets coherent with the task it is performing, were based on modelling her, or other subjects, attention according to hundreds of training samples. She could not record the activity in the neocortical area but she could record a specific behaviour, that is, the phenomenon and estimate, from several sequences, a model. There are two important questions concerning the estimated behaviour. The first concerns the subject and the second the reproducibility of the behaviour. She has noticed that there are certain regions that are equally fixated by all subjects on static scenes, but the reaction to distractors and moving objects likewise time of fixations are different, both when concerning selective and top-down attention and bottom-up stimuli.

It seems like there are social, epochal and individual characterizations that make the immediate feeling of the environment and the self in it dependent on several factors including space and time, the "dasein und zeit". For example it would be interesting to compare how do humans react to speed and altitude and vertigo now with respect to one hundred years ago. Are these considerations concerned with consciousness? Indeed, since the "Γνωθι Σεαυτόν" the exploration of consciousness has been identified with its representation by philosophy, linguistics, psychology and so on. The recourse to representation was not only due to the lack of knowledge of the brain activity. In other words, the present knowledge of the importance of the neocortex to conscious perception move out all the speculations transcending the factuality of the brain activity, but at the same time it should keep the representation aspects due to the corresponding behaviour. While Koch says that behaviour can occur in the absence of consciousness, the cognitive roboticists refer to it as the privileged expression of the experience. What allows cognitive roboticists to reproduce a certain behaviour in a robot is the observation of the behaviour elicited by the conscious percept, and the model obtained by its representation. It follows that the knowledge of the stimulus comes from both the brain activity and its corresponding behaviour, which is explicitly affected by the representation domain. In fact, also the observation of specific neural functioning is subject to correlation functions whose scientific substratum is determined by current scientific explanations and findings, which could be demonstrated ineffective or disproven after new discoveries. In other words, similarly to Church thesis, which cannot be proven mathematically, consciousness might not be proven within NCC.

1.2 Reproducibility of experience

In agreement with Koch definitions, a robot can be elementarily conscious. What are the minimal conditions to define consciousness in a robot? A robot that can infer from given premises some reasonable, if not common sense based, conclusions, but without any sensory apparatus, and thus locked to any stimuli, is it conscious? A robot endowed with both exteroceptors and proprioceptors all active and receptive but unable to distinguish a percept from another, is it conscious? The answer is not in both cases, because in both cases the robot is unable to both have an experience and reproduce it.

Without the ability to reproduce a specific experience, by associating a representation to it, a robot cannot be conscious. In other words it would not be able to build a model of the sensory abilities allowing it to represent and replicate the experience. Alma has argued in her thesis that a minimal level of consciousness for a robot would be the ability to survive in a non-offensive and nondeterministic environment. Where a non-offensive environment is one in which nobody will attempt to destroy (disassemble) the robot, as far as it is functioning, or impede its survival. Nondeterministic means that survival is not ensured under identical conditions, that is, the reproducibility is an ability of the robot not of the environment. So a robot able to auto-recharge, that is, find out how, where and when to do it, and to switch on and off deliberately, would be an elementary conscious entity. Note that the nondeterministic condition means that for example the charge can change place, and the non-offensive one means that it will be available in some reachable place. Why these requirements would testify a perceptual awareness? Because independently of the environment, and the degree of uncertainty (for example shape, colour and possible locations of the charging place) the perceptual, attentional and recognition like processes should make possible the reproducibility of the experience. For example the robot would need to "feel" that it cannot remain in the state switched off all the time, otherwise it would die, it should "feel" that before recharging itself it has to localize where to do it, and to recognize where to do it, it needs to have some charge left, and so on, and all these are reproducible experiences.

In this sense early consciousness should in general be associated with the ability to reproduce an experience or, analogously, to recognize the same input or the same reaction under different phenomenal states. Further it requires to establish the bounding conditions of the experience, inside which it can determine the reproducibility, for example it should be able to determine what an environment is, and how wide is it, and if it is the same with something changed or a completely different one.

Now, the activation of each of the required processes for a robot to be minimally conscious is possible not because a cortical structure identical to e.g. the macaque or the mice, is implemented in the robot, but because there is a representation of these processes that can be imitated. In fact there might be several cognitive architectures that could effectively achieve the construction of the minimal conscious robot.

Suppose that there are 3 cognitive architectures, say A, B and C underpinning this robot minimal consciousness and such that, under the same conditions, they behave the same. How can we establish which is the one that best interprets the concept of minimal consciousness. Clearly we would prefer the one that is more ductile to scale up to more general requirements. This implies that we tend to prefer (and reproduce) those behaviours alluding to a human like consciousness, therefore we have a pattern of

consciousness which is pre-existent and depend on functional performance, utility, etc. In fact, like for the wheel or for the wings, it might be possible that the best artificial brain is structured in a totally different way with respect to the actual structure of the brain (as we presently know it) but at the same time it is the one that best interpret the epiphenomenal behaviour of perceptual awareness that we can represent.

1.3 Space of experiences

Koch project is factual, strong and, given the current findings, is the best promise, as Alma writes reporting the following phrase from his book: "We live at a unique point in the history of science. The technology to discover and characterize how the subjective mind emerges out of the objective brain is within reach". Still, even supported with all the computational and inspection apparatus needed to explore and map the NCC this approach lacks to account for its counterpart which is cognitive robotics, that is, how to reproduce perceptual awareness with limited computational resources.

Why so? Because any theory can be validated only by prediction. Koch says that there must be an explicit correspondence between a mental event and its neuronal correlate, but does this mean that the spiking activity is the only possible way to replicate consciousness in a machine? Or better if perceptual awareness can be generated only within the NCC it would never be reproducible but by the entity being studied and thus not general.

Alma thinks that reproducibility of perceptual awareness spans a space with varying scales and dimensions. The lower level is given the highest resolution and it allocates the earlier, or most preliminary, detailed experience, the inner experience, the one of the early self through a wild unfolded perception. This early experience is a dense map of energy functions responding to appropriate kernels. What kind of filters perceptual awareness should use to transform itself in communicable representation, that is, to manifest the annotated map of energies? The filters language is based on immediate sign responses. Foucault writes that the "first example of a sign given by the Logique de Port-Royal is not the word, nor the cry, nor the symbol, but the spatial and graphic representation - the drawing as map or picture." This was just to consecrate a long divagation on "Las Meninas" to specify the interplay between consciousness, representation and the observer as emerging from Velasquez classical age. But what is important here is that the cognitive roboticist is the observer of the early visual map of the robot perception like the neurophysiologist is the observer of the fMRI map. But the roboticist has to invent the model of the signs interpretation to reproduce the perceptual awareness, cannot only observe its functioning, because there is no functioning without the model. Now Alma knows that there is a trade off between experience resolution and representation, it is not possible to have both of them at the highest resolution, because of a principle of indeterminacy. So the more details the experience can get the less decipherment its response and representation can have.

On the other hand, at the higher levels of the experience resolution is lower, details are lost, like an awakening from a dream, but the representation is finer, the language is normed, and the sign collection is more and more consistent with the epochal knowledge or universal representation structure, coherently with Benveniste interpretation of Saussure: "linguistics is exclusively a science of form". So at the highest level of

the experience scale space the behaviour can be completely coded and made communicable, like reasoning or planning. This constitutes the strongest reproducibility of experience, the zeit of the epochal consciousness, the now, indeed.

Still this is the problem of experience reproducibility. It is difficult, if not impossible, to distinguish between the resolution of representation and the resolution of the experience. Even accepting the different resolution levels, how is it possible to determine the correct parameters of the responses? The expected perceptual awareness could be hidden despite encodable. False positives are possible when the representation reflects an unknown state of perception.

Looking far, towards the lattice of clouds Alma was pondering on the immense network of interconnected representations, each one positing itself as the sign of what it represents. The role of the cognitive roboticists is to find, interpret, from the artificial signs, the right resolution level of the scale of experiences to reproduce them and to validate, with biologically inspired behaviours, the artificial consciousness. This age of observability and reproducibility is at risk of undistinguishability, we are having a hard time to get the hidden parameters in the experience scales through the different levels of their representation codes.

Suddenly Alma realizes that she is observed "Croa.. Croaaaa.." a big brown toad was croaking from the head of the P3-DX, looking intensely at her.

Alma stared at it for a while "poor little ugly toad with such a cacophonous sound, is it aware of it? Is it envious of the nightingale? What's good about it?". The toad was there with its intermittent "Croaaa,..Croa,Croaaaa,..". She perceived the irksome croak as a paradoxical sign of disgracefulness, "what is it like to be a toad?" She was bewitched by the toad and wanted to find out more. Alma began to record the croaking sounds and to perform some spectral analysis through few scripts she had to integrate visual attention with auditing features. Observing the peeks and amplitude she concluded that the regularity of the frequency suggests mapping the sounds to 0 and 1. Further, to get the most of the observations Alma quickly made a script to transform the audio signals to 0 and 1, according to the frequency between peeks, and get a more and more accurate estimation of the possible sequences, almost in real-time. After a good amount of sampling and parameter re-estimation, she eventually gets persuaded that the length of a meaningful audio signal was 6 digits with which she could interpret the 36 alphanumeric symbols and some more punctuation ones. Not so surprisingly Alma had a good chance. There was a signification in the croaking, or the proposed parameters were well guessed, because she ended with an astonishing discourse:

"I'm a prince coming from a dwarf Galaxy 12 million years far at $b0^\circ, l40^\circ$, earth astronomers call it the Galaxy Holmberg I, a member of the M81 group of galaxies and estimate it at a distance of $\sim 3.6Mpc$. Because of low-mass we are warm entities with heterological shape, that is, our shape does not describe ourselves and is rarefied. To survive the trip to the Earth I had to lower my temperature so when I landed I could only activate the shape of an amphibian with ectothermic regulation. But now I'm stuck into this low temperature body and need immediately to get in contact with a warm entity for activating a reaction for morphing into a warm body, I need a very strict contact. So please kiss me, a real kiss will rescue me."

What might she do with this phrase? She could reject the phrase and kick off the toad or accept it and kiss it. Her sampling was satisfying and she obtained a good

evidence of the decoding algorithm. On the other hand the probability of an entity existing in Holmberg I and travelling 3.6Mpc is zero! But summing the probabilities the outcome is greater than zero. Moreover, mindful of Pascal argument it would certainly be more convenient to believe than to not: if the toad is really coming from this far Galaxy the reward is infinite.

She was puzzled, kissing this ugly and cold amphibian was disgusting and the repulse was stronger than her love for knowing the truth. Thinking about what to do she decided to adjust the algorithm to obtain a more accurate estimation of the transformation of the sound into the code and more evidence for either confirmation or rejection. While adjusting her script strings of 0 and 1, quite long and all ending with 11 and with no 11 before the two 11 in the end began to appear on the screen. Clearly it was a Fibonacci code. Why one more code? What was happening. So she made a new quick script to obtaining the numbers and used the same schema as before to get letters numbers and punctuation symbols. And now the phrase was the following:

” Sorry for the new code I meant to attract your attention, I’m SQC 8001QC23, I’m sending the message through your robot companion. The prince of Holmberg contacted our community as some QC23 specimen is able to interpret his quantum code, that he could transmit using your robot. The QC23 specimen has been able to factorize the data obtained and reconstruct with it a simulation of life in Holmberg that the prince wants you to look at. The QC23 specimen generated the code for the simulation so that you can believe his words. Here is the code.”

To Alma’s surprise the code was extremely long and written in assembler, so she had to embed it into *_asm* blocks in visual C⁺⁺. To end up the whole thing she consumed two batteries and when finally she got the simulation the early evening star was shining in the dark sky. The simulation was incredibly realistic showing entities with shapes varying terribly quickly escaping from this small planet in Holmberg I before the collapse. Everything was apparently very small these far Galaxian entities looked like Lilliputians. She was so moved that she felt the necessity to save the life of the toad that was no more on the head of the P3-DX. She spied out the grass to find the small brown toad, she found its cold body and put it in her hand. But when she looked at it in the light of the robot she understood that the poor toad was dead. It was too late. She could have saved him, the prince of Holmberg, but she did not because she got lost in the entangled streams of possible coding from consciousness to representation, and truth could not be distinguished.