Stuart, S.A.J. (2010) *Conscious machines: memory, melody and muscular imagination.* Phenomenology and the Cognitive Sciences, 9 (1). pp. 37-51. ISSN 1568-7759

http://eprints.gla.ac.uk/40131

Deposited on: 12 October 2010

Conscious Machines: Memory, Melody and Muscular Imagination

Susan A. J. Stuart

August 6, 2010

Abstract

A great deal of effort has been, and continues to be, devoted to developing consciousness artificially[1], and yet a similar amount of effort has gone in to demonstrating the infeasibility of the whole enterprise[2]. My concern in this paper is to steer some navigable channel between the two positions, laying out the necessary pre-conditions for consciousness in an artificial system, and concentrating on what needs to hold for the system to perform as a human being or other phenomenally conscious agent in an intersubjectively-demanding social and moral environment.

By adopting a thick notion of embodiment – one that is bound up with the concepts of the lived body and autopoiesis [Maturana & Varela, 1987; Varela *et al.*, 1991; and Ziemke 2003, 2007a & 2007b] – I will argue that machine phenomenology is only possible within an embodied distributed system that possesses a richly affective musculature and a nervous system such that it can, through action and repetition, develop its tactile-kinaesthetic memory, individual kinaesthetic melodies pertaining to habitual practices, and an anticipatory enactive kinaesthetic imagination. Without these capacities the system would remain unconscious, unaware of itself embodied within a world. Finally, and following on from Damasio's [1991, 1994, 1999, & 2003] claims for the necessity of pre-reflective conscious, emotional, bodily responses for the development of an organism's core and extended consciousness, I will argue that without these capacities any agent would be incapable of developing the sorts of somatic markers or saliency tags that enable affective reactions, and which are indispensable for effective decision-making and subsequent survival.

My position, as presented here, remains agnostic about whether or not the creation of artificial consciousness is an attainable goal.

Introduction

Holland [2003] makes a distinction between weak and strong artificial consciousness. The former is the design and construction of machines that simulate consciousness or conscious cognitive processes. The latter is the design and construction of conscious machines, machines that

[1] A small selection of the many authors writing in this area includes: Cotterill 1995 & Cotterill 1998; Haikonen 2003; Aleksander & Dunmall 2003; Sloman 2004 & Sloman 2005; Aleksander 2005; Holland & Knight 2006; and Chella & Manzotti 2007.

[2] Most notably: Dreyfus 1972/1979, 1992 & 1998; Searle 1980; Harnad 2003; and Sternberg 2007, but there are a great many others.

will have the capacity for subjective conscious experience. Since weak artificial consciousness (WAC) evokes the possibility of conscious inessentialism – the claim that "for any intelligent activity *i*, performed in any cognitive domain *d*, even if we do *i* with conscious accompaniments, *i* can in principle be done without these conscious accompaniments" [Flanagan 1992, p.5] – we might think of it as presenting or, at least, proposing a form of the philosophical zombie; something which appears to have phenomenal experience and which would, from the perspective of an observer, be indistinguishable from the 'real thing', that is, from strong artificial consciousness (SAC).

The possibility of WAC has been with us since the invention of Jacques de Vaucanson's duck[3] and, even though the credulity of the audience has changed very little[4], our ability to produce increasingly complex weakly artificially conscious systems has increased exponentially. However, it is not the creation of something which provides a good simulation of consciousness that is the issue here – the preconditions for such simulations vary widely across the range of possible tasks to be accomplished; the issue here is what needs to be the case if we are to design and create a strongly artificially conscious machine and what, if anything, would be its performance advantage? In short, what do we want a conscious machine to be able to do and what practical requirements would we need to address to make that possible?

Why create a conscious machine?

The question "Why create X?" can be asked of any potential artefact or object and, in the case of tools, there is usually a pretty straightforward answer: "Because X fulfils a particular purpose or function.".[5] This is true whether the object is a knife, a map, or a pocket calculator. The intention behind the tool's creation is that it do the job for which it has been designed more efficiently, that is, some combination of more effectively, more quickly, more easily and more accurately, than the human being who is to use it would otherwise be able to do. In the case of the calculator, just as in the case of a Mars exploration robot[6], or a satellite navigation system, but unlike the case of a butter knife, a bridle, or a jug, we conceive of them incorporating some level of artificial intelligence – they act in a way which, if performed by a human being, would be considered intelligent – however, we do not consider them to be conscious, and we must first ask what would be the purpose of creating a conscious machine?

Assistance with mundane tasks like housework, fixing the plumbing, touching up the paintwork, and cat-sitting when we go away on holiday might be a good place to start. Our lives could be made a great deal simpler by having a conscious machine which could judge when things needed to be done, gauge how to do them, and possess the initiative to embark on the task.[7] And then again we might be interested in the fungibility aspect of artificially created conscious machines and the advantage they offer with respect to their employment in dangerous tasks. In America 118 of every 100,000 fishermen died in 2005 and, in the same year, the logging industry lost 90

[3]See: `http://fr.wikipedia.org/wiki/Jacques_de_Vaucanson`

[4]Our readiness to over-interpret the actions of the 'other' in our midst as intentional is a useful strategy for survival. Even though infants learn to distinguish animate from inanimate, and then to distinguish minded animate from non-minded animate [Stern 1985], it is still wiser to be mistaken occasionally and over-ascribe a mind to some thing rather than under-ascibe and risk becoming that thing's lunch.

[5]The most obvious exceptions from a clear answer of this sort are art objects.

[6]See: `http://marsrover.nasa.gov/home/index.html`

[7]The implications of robot servitude will not be considered in this paper, but they have been discussed in much detail elsewhere, for example, Asimov 1950 and Peterson 2007.

workers from 100,000.[8] But these numbers are nothing compared with the risk of being a soldier in times of conflict. 4000 American soldiers died in the Iraq conflict between 19th March 2003 and 14th February 2009 [9], with the two World Wars providing an even higher ratio of fatalities to combatants.

From housework to warzones, if it were possible to replace these human beings with conscious machines which could act and judge with the concern and compassion we associate with the best of human beings[10] and the efficiency we associate with machines, and in the event of their damage or destruction be repaired or replaced with others of their kind, then we would seem to have a fairly ideal situation. There can be little question that the idea of reducing the risk to human life in these situations has a strong appeal, as does the idea of manufacturing beings capable of calculating these risks far more efficiently than we can, thus becoming less likely to overreach, stand in the wrong place, shoot an innocent bystander, or make a decision that goes catastrophically wrong. But making judgements of this kind, feeling the initiative to act and acting on it, requires strong artificial consciousness, and a basic underpinning of the capacity for such engaged, embodied judgement is the ability to perceive and sense the world, and to feel your own body from the inside as it is positioned in relation to those things within your world. Anything with this kind of consciousness needs more than simply the *potential* to think and to learn, to feel emotions, to build up anticipations, to form beliefs, desires, and memories, it *must* have them. It would not just be indistinguishable from real consciousness, it would be real consciousness, and with the creation of real consciousness the debate teems with moral issues, not least of which is how we might resolve the problem of creating a conscious agent instrumentally as a means to an end rather than treating it as an end in itself.[11] [See, for example, Torrance 2008.]

Machine Phenomenology and the Body

The underpinning for any conscious organism's sensation and feeling – its phenomenology – is its body. It should be noted that the term 'organism' is used judiciously here because it is only in the case of organic systems that we know any phenomenology to exist; this is not meant to imply that it is a sufficient condition for phenomenology, only that it is the most natural place from which to begin our investigation, especially if the judgements made by the artificially conscious system are going to be recognised by us as expedient. Thus, it is to the nature and functioning of the body that I will appeal to support the claim that for a machine to have first-person phenomenal experience it requires a sensory system, a distributed nervous system, a subtle musculature with functionally healthy peripheral nerve cells, and actuators which enable it to react to its experience and bring about further change within its world. I will temper this claim a little to admit that organisms with a more limited nervous system, and which might be exoskeletal rather than endoskeletal, can still be conscious of their environment, but – and, for now, it might only be on the basis of a hunch that I would want to maintain – not to the same

[8]These figures were taken from `http://money.cnn.com`

[9]These figures were taken from `http://www.antiwar.com`

[10]It is not lost on the author that the characteristics of concern, of compassion, and of the possession of moral wisdom (*phronesis*) which is associated with being the best human being, do not sit well with waging war.

[11]This is by no means a full list of the possible reasons for creating an artificially conscious system; there may even be the (possibly) non-instrumental reason of creating it so that we better understand consciousness, or simply the desire to bring another conscious being into existence.

degree of phenomenality that an endoskeletal system would.[12]

It has been argued that the body plays a crucial role in providing "the unified, meaning-giving locus required to support and justify attributions of coherent experience"[13] to the agent – for Merleau-Ponty the body is the locus of the organism's point of view on the world and consciousness is necessarily embodied [Merleau-Ponty 1968] – and it is its dynamically-coupled situatedness that expresses the integration of the system and its world; they seem separate but are inseparable, possessing an interdetermination that is demonstrated through continuous feedback loops of sensation, action and reaction.

A strongly artificially conscious machine

> ... must be able to sense its world, bring about change in its world, and distinguish itself from its world; and for these abilities it will require sensing and actuating systems. The former enables an agent to acquire information about its environment, working as an outer sense, making it possible for the animat to determine its external state. The part of the system that senses the external world links, directly or indirectly, to actuators, making action, and hence interaction with the world, possible. But in more complex agents 'sensing' will comprise an 'inner sense' that not only enables the agent to determine its goal(s) and compare its sensory input with its current internal state(s), but also, if the agent is to make appropriate decisions about its action, to monitor its position in the world, its movement through the world and its actions within the world. Such an agent [a SAC] will maintain a 'body schema' that will provide it with "continually updated, non-conceptual, nonconscious information about [its] body ... [providing] the necessary feedback for the execution of... gross motor programs and their fine-tuning" (Meijsing, 2000, p. 39). [Dobbyn & Stuart 2003, pp.197-8]

The sense of both an inner 'egocentric' space [Brewer 1992] and an affective depiction – the sensation of being 'out-there' [Aleksander & Dunmall 2003] – is formed through the rich interplay of the body's sensory channels that receive information about the environment[14], its actuating system that enables manipulation of that environment, and its proprioceptive mechanisms which make it possible to sense the position, location, orientation and movement of the body and its parts. All of these capacities are evident in the majority of human and non-human animal species, and in some – still rudimentary – machines [See Holland's *Cronos* 2006[15]].

The sort of experiential integration made possible at this level, and which unites experiencer and experienced and synthesizes the contents of experience into a coherent unity of consciousness, is neither necessarily conceptual nor cognitive though it has the potential to become

[12] For a very interesting and detailed discussion of the sensory and proprioceptive capacities of exoskeletal systems and invertebrates in general I recommend Sheets-Johnstone 1998: "hard-bodied invertebrates have external sensilla of various kinds: hairs, exoskeletal plates, epidermal organs, cilia, spines, pegs, slits, and so on. It is these external sensory organs that make possible an awareness of surface events in the double sense noted above: an awareness of the terrain on which and/or the environment through which the animal is moving and an awareness of bodily deformations or stresses occurring coincident with moving on the terrain and/or through the environment" [p.279].

[13] `http://www.consciousness.it/CAI/CAI.htm` *Viz.* Johnson 1990; Chiel & Beer 1997; Clarke 1997; Damasio 1999; Lakoff & Johnson 1999; Seitz 2001; Dobbyn & Stuart 2003; Legrand 2006; Ziemke 2003 & 2007; amongst a great many others.

[14] The concept of 'environment' is used thickly to refer to the system's world and its own variable internal states that are the subject of homeostatic functions.

[15] The CRONOS Project website: http://www.cronosproject.net/

both in the right kind of system. The foundational role in any phenomenally and subjectively conscious system is performed, according to Sheets-Johnstone, by the affective dynamic and tactile-kinesthetic body coupled to its correlative kinetic world [Sheets-Johnstone, 1999, 2000, & 2003]. It is this fully-immersed coupling which makes integration possible and it depends, at a fundamental level, on a kinaesthetic synthesis: the formation of kinaesthetic and nervous patterns established through somatosensory engagement and the repetition of goal-directed muscular actions and reactions, all of which operate at a prereflective, yet possibly still intentional, bodily level.

In a similar vein Cotterill [1995] uses the wonderful word 'plenisentience' to describe the body's being switched on to its world, perceiving, receiving, imagining, anticipating, and actuating. It is an echo of Whitehead's claim that "The essence of an actual entity [conscious agent or, quite possibly, conscious machine] consists solely in the fact that it is a prehending thing." [1929, p.56] Through sensors and receptors the system continuously prehends, grasps or becomes aware of incoming and internally transmitted stimuli and, in its movement and perceptual engagement it asks bodily and kinaesthetic questions about how the world will continue to be for it. This requires a lower, non-self-conscious level of consciousness and expectation which situates the system actively in its world.

In understanding the body's capacity for conscious activity, and what would be required for an artificially created conscious system to be subjectively aware of its experiences as experiences for it, we must first distinguish between the body as subject and the body as the object of consciousness, and then we must examine how the two are mutually necessitated through enaction, that is, through the structural coupling that helps a conscious system to anticipate and enact its prospective states; it is enaction that brings forth its world [Varela, *et al.* 1991]. A bodily consciousness of this kind is the result of an interplay between the agent's body image, its body schema, and its being and activity in its world. The terms 'body image' and 'body schema' were first distinguished by Head and Holmes [1911], and further clarified by Gallagher [1986]. It is a distinction for which there is now neurophysiological evidence [Paillard 2005].

The body image is an "internal representation in the conscious experience of visual, tactile and motor information of corporal origin" [Head and Holmes 1911], it comes directly through looking at or touching parts of our body, or indirectly through our perception of, for example, our reflection in a mirror. On these occasions the body is both the content of our conscious experience and the object of our intentional activity and, as the object of thought, it lacks the transparency – the immediacy and lack of self-conscious awareness – that the body possesses as subject. The body schema as a pre-attentional, non-self-consciously monitored [Merleau-Ponty 1962] "real-time representation of the body in space generated by proprioceptive, somatosensory, vestibular and other sensory inputs" [Schwoebel *et al.* 2001] possesses this transparency.

The body schema is "a combined standard against which all subsequent changes of posture are measured . . . before the changes of posture enter consciousness" [Head and Holmes 1911]. It is extra-intentional, subconscious, subpersonal and unowned, operating through a set of sensorimotor laws which are "constraining and enabling factors that limit and define the possibilities of intentional consciousness" [Gallagher 1995 p.239] making perception and action possible. It remains hidden phenomenologically from the agent, and is set in stark contrast with the sensory images that go to make up the body image. The body schema "organizes the body as it functions in communion with its environment" [Gallagher 1985, p.549] as it actively, and mostly unconsciously, organises its perceptual experience in relation to its pragmatic concerns [Merleau-Ponty 1962; Heidegger 1968].

The essential role of the body and its dynamic goal-directed movement in the generation of consciousness can be established by examining what happens to an agent's consciousness when their body has lost its proprioceptive ability. It is not unusual in cases of such severe deafferentation for a diminution of the sense of self to follow. It is as though the agent's capacity to draw together or integrate their identity has also evaporated, and it is only when they are able to replace their internal feedback system with an external feedback system, most usually their visual sense, that they become able to move – with a great deal of concentration – and regain their sense of self or identity.[16] It seems that just as our eye is not our servant but our ambassador [Ings 2007], so too is our body; that just as our eyes are ceaselessly active, questioning and interrogating our world, so too is our body, and in the absence of afferent stimulation the body loses its integrity and only the eye with its incessant inquiry can provide some compensatory evidence for the agent's bodily perdurance. Thus, it is is not just the passively received information about a changing environment, but the interplay between this information and active self-movement that places the phenomenal agent, as an integrated and coherent unity, firmly at the centre of its world [Meijsing, 2000].

In a very similar way locked-in syndrome, with its lesion to the brain stem and damage to the under side of the pons, provides another example of the evaporation of an integrated and affectively depicted subject. In this short passage Jean-Dominique Bauby quotes a half-imaginary exchange with his friend Florence: "'Are you there, Jean-Do?' she asks anxiously over the air. And I have to admit that at times I do not know any more." [Bauby 1997, pp.49-50]

A normal level of sensory bombardment is also absent in patients with spinal cord injury (SCI). Depending on the location and severity of the damage, SCI can result in restricted or absent sensory and sensory-motor input and can, sometimes, be accompanied by, mostly, visual hallucinations. Thus, in the absence of direct afferent input and efferent feedback from the somatosensory motor cortex, the brain compensates and activation from the visual regions is likely to overlap, even invade, an area of inactivity. However, in some cases the brain, left without sensation, imagines or creates the sensation of pain, often severe and usually situated below the level of the lesion. "Pain, a particular form of 'imagination' of sensory experience without sensory input from the periphery, is a vicious way the brain extracts revenge for being left without sensation" [Cole 2005, p.191]; yet through this revenge the agent once again feels situated in their world.

A still clearer case for the necessity of affection and action is provided by the example of people who have fully-sentient heads but unfelt bodies and who experience themselves as 'floating' unthethered from their body.

> The absence of sensation, both in the immediate aftermath of SCI and later, can be a problem, leading to some odd experiences. Thus lying on a bed or even later in a chair without sensation from the neck down leads people to think of themselves as 'floating', with their sentient heads above unfelt bodies. Several weeks later, trying out a wheelchair for the first time, it can seem an awful long way up. [Cole 2005, p.185]

[16]'Oliver Sacks remarks (in the 'Forward' to Cole, 1991) that the case of IW "shows how such a peripheral disorder can have the profoundest 'central' effects on what Gerald Edelman called the 'primary consciousness' of a person: his ability to experience his body as continuous, as 'owned,' as controlled, as his. We see that a disorder of touch and proprioception, itself unconscious, becomes, at the highest level, a 'disease of consciousness"' (xiii).' [Gallagher & Cole (1995), Footnote 7]

There can be little doubt that whatever subjective, phenomenal consciousness is experienced it appears, at least, to be ungrounded, unsituated, and unembodied.

A much less severe, though nonetheless significant, example is provided by Oliver Sacks and describes the phenomenality of severe nervous and muscular damage that he sustained to his knee. “When I looked straight ahead, I had no idea where my left leg was, nor indeed any definite feeling of its existence. I had to look down, for vision was crucial. And when I did look down I had momentary difficulty in recognizing the ‘object’ next to my right foot as my left foot. It did not seem to ‘belong’ to me in any way.” [Sacks 1984, p.137]

In each of these cases there has been damage to the body in the form of damage to the nervous system and subsequent damage to the individual’s body schema.[17] Each case demonstrates that the body, with its subtleties of sensing and perceiving – its nuanced somatosensory awareness – combined with its capacities for goal-directed movement and actuation, is capable of (i) creating and establishing a feeling of how and where the system *is*, its extent and limitation – its *depiction* [Aleksander & Dunmall 2003] – and also for (ii) anticipating how it might be in its world, that is, how it might act given its awareness and selection of those things that have a particular affordance for it. [Gibson 1968 & 1979]

Enactive engagement is evident in both (i) and (ii), but it is a bodily or kinaesthetic enactivity that produces non-conceptual yet intentional anticipations for how the agent and its world are currently, and how the agent and its world could be should certain potential changes come to pass. None of this engagement need be self-conscious or conceptual though under certain circumstances, perhaps of heightened tension or attention, it might be. Think only of attempting to walk silently across a forest floor. The awareness of every movement is intense. The flexing of each muscle is the subject of minute attention. The vigilance with which we scan our surroundings and place each foot is concentrated and precise, and the judgement of the likelihood of this twig snapping or that twig bending is made with great diligence. But this kind of scenario is rare. In most of our activity and actuation we rely on the body’s natural propensity to automate habituated kinaesthetic activity. The seamless use of any tool, to run and skip, to sign our name, or to throw a curveball is the result of our capacity to memorise and form fluent, non-cognitive melodic movements. [Stuart 2007] As Luria says “with the development of motor skills the individual impulses are synthesized ... into integral kinaesthetic structures or kinetic melodies.” [Luria 1973, p.176] Thus it is through sensation, action and repetition that the agent develops an enactive kinaesthetic imagination, its kinaesthetic memory, and its own unique individual kinaesthetic melodies. Such melodies can correspond to gross motor skills but they can also correspond to subtle and more local bodily affective activity, for example, the unreflexive or ‘gut’ response that provides us with a pre-cognitive ‘feeling’ about potential actions. In terms of effective decision-making, consciousness, even at this bodily level, presents a significant advantage for any situated and embodied agent.

So, fundamental to what we understand about how phenomenal experience arises in conscious systems is their ability to perceive and sense their world, to feel their body as it is positioned in relation to things that constitute the world for it, and through this engagement to feel the initiative to act and to act, when appropriate, on that initiative. The appropriateness of the action, whether the physical response to pursue some prey, to trust someone, or to buy Icelandic

[17]There may also have been some corresponding effect on the individual’s body image, for example, cases of body dysmorphic disorder when someone does not believe that a limb is theirs or apotemnophilia when someone wants a limb amputated because it does not correspond with how they feel themselves to be, might be the result of a faulty or malfunctioning body schema which leaks into how they perceive themselves, but this cannot be taken further in this paper.

stock, is something which has to be judged carefully, and the capacity to make these judgements, whether bodily and perceptual or cognitive and conceptual, requires active intersubjective social engagement.

The Somatic Marker Hypothesis

Over the last two decades Antonio Damasio has presented a consistently forceful case for the role of the body, and specifically for the emotions, in effective decision-making. [*Viz.* Damasio *et al.* 1991; Damasio 1994, 1999, & 2003; and Damasio *et al.* 2000] According to his theory, emotions – defined as spontaneous neural and chemical responses to changes in the agent's physiological state – play a central role in the agent's homeostatic functioning. "[T]he subjective process of feeling emotions is partly grounded in dynamic neural maps, which represent several aspects of the organism's continuously changing internal state". [Damasio *et al.* 2003, p.1049] They are the body's pre-reflective, pre-cognitive affective activity which act to underpin the development of the agent's successful adaptive behaviour.

It is impossible, in this paper, to go into the finer details of Damasio's breakdown of emotions into six primary (universal) emotions [*Viz.* Ekman 1992]; a number of secondary (social) emotions, for example, embarrassment and pride; and background emotions, for example, calm or tension, anticipation or dread; but even simply stating them goes some way to demonstrate the manifest complexity and subtlety of the phenomenally conscious subject.

Decision-making in such a subject isn't simply a cognitive and evaluative process involving a judgement about how best to reach the desired goal; it is also a matter of determining the emotional salience a state of affairs or potential action has for the agent. Thus, it requires sensory and affective elements and, only on some occasions, perhaps involving the secondary social emotions, will it also require a cognitive element. Without an affective element the agent will be unable to rank the items to be judged in order of their significance to her directly, or to her indirectly by the affect they are likely to have on those about whom she is concerned. This ranking is termed 'saliency tagging' and the capacity to tag in this manner is formed through our experience of primary inducers, for example, direct or immediate experience of something dangerous, which establishes in us a network of secondary inducers making it possible for us to recall in similar future situations – though not necessarily consciously for that would hinder the speed of our reactions – the sensation we had on experiencing the primary inducer and now disposing us to react with appropriate caution or incaution to whatever it is we now experience. Thus patterns of bodily affective actions speed up effective decision-making in the same and similar future contexts.

The feelings we have as a response to these emotions can be conscious or unconscious, but when conscious they represent the relation of the emotion to ourselves as agents in the world. In Damasio's language and theory this is expressed through the relation of the proto-self to the core-conscious self, or core-consciousness. So, there is some 'primary mode of being' [Heidegger 1962] which is phenomenologically and ontologically prior to, and necessary for, the formation of a core-consciousness, and the "absence of emotion [the primary mode of being] is [or can be taken to be] a reliable correlate of defective core consciousness". [Damasio 2003, p.100]

Several areas of the brain are crucial for effective decision-making; they include the amygdala, the anterior cingulate cortex, the brain stem, the hypothalamus and the ventromedial prefrontal cortex [vmPFC]. The vmPFC mediates between the limbic system and the cerebral cortex and damage to it can leave the subject intellectually unimpaired yet incapable of making ordinary everyday decisions in real-life situations. Damasio quotes patient EVR as having suffered damage to the vmPFC and to the connection between the amygdala and the hypothalamus. As a result EVR is incapable of making a decision about trivial matters like where he should go for dinner for he becomes overwhelmed by irrelevant information. He is incapable of tagging and prioritising informational criteria as relevant or irrelevant for all criteria carry the same emotional weight.

This kind of day-to-day affective activity and effective decision-making requires a phenomenally conscious agent. But if our system is, in fact, a façade of conscious engagement, a WAC, it will have no capacity for creating and establishing a feeling of how and where *it is*, or for anticipating, pre-cognitively yet affectively, how *it might be*. There can be little doubt that, as a consequence, it will be much less efficient than a phenomenally conscious, subjectively aware system. With no sensation a WAC has no affective somatosensory, motor, and efferent feedback system, it will be unable to develop the sorts of somatic markers or saliency tags that enable affective responses and, thus, the formation and reformation of adaptive behaviours and strategies. Without these a WAC will either take an unconscionable amount of time because it will be unable to decide between equally-weighted items, becoming a modern-day Buridan's ass, or it will make a decision that could impact unfavourably on its continued survival. It will have no natural feel for itself in its world; there won't be *a world* for it; indeed, there won't be an *it*.

Phineas Gage is the best example of someone who falls into the latter category as being perfectly able to continue to make decisions, but which ceased to be effective. [Damasio *et al.* 1994]. Phineas Gage could not be considered equivalent to a WAC, he remained conscious, even through and after his accident, but the damage to his vmPFC left his social action and interaction very seriously impaired. An account from his own doctor reveals that he went from being a hard-working and reliable family man to a confabulator whose decision-making skills were limited to the formation of unrealistic schemes.

> His contractors . . . considered the change in his mind so marked that they could not give him his place again. He is fitful, irreverent, indulging at times in the grossest profanity . . . manifesting but little deference for his fellows, impatient of restraint or advice when it conflicts with his desires, at times pertinaciously obstinate, yet capricious and vacillating, devising many plans of future operation, which are no sooner arranged than they are abandoned . . . . [Harlow 1848]

After the cortical damage Gage's chances of survival outside the protection of his social group would have been extremely limited. Not only had he lost any sensitivity to, or capacity for, informative visceral responses, but in his devising of wild schemes and lack of consideration for others he seemed also to have lost his basic endogenous intersubjectivity [Gallagher 2007], the practical knowledge of oneself that guides one's actions from the inside out.

Damasio's account does not rule out the evaluative capacity that we associate with judgement, it's just that it comes much later on and at a much higher, possibly consciously cognitive, level as the result of reflection and feeling, where feeling is "the realization of a nexus between an object and an emotional body state' [Damasio 1994, p.132]. It would be hoped that with the creation of a SAC the reflexive evaluative capacity would emerge through the ongoing felt dynamics of the agent's social action and interaction but, as Damasio [2003] has shown, damage to certain regions of the frontal lobe in young children can inhibit the development of social emotions, like compassion, shame, and guilt; the very emotions we tend to associate with mature subjective reflection, evaluation and judgement, and the kind we deemed necessary for a machine with a strongly artificial consciousness if it is to exhibit the concern and compassion we associate with the best of human beings. It would now seem that such a machine needs its phenomenology from the outset if it is to manifest any endogenous intersubjectivity and develop the kind of practical knowledge that will guide its actions from the inside out.

Conclusion

We began with the doctrine of conscious inessentialism and the possibility of building a machine that could do everything that a phenomenally conscious subjective agent could but without the phenomenal subjectivity. But conscious inessentialism is a false doctrine; a machine with a weakly artificial consciousness could never succeed in an environment where a capacity for affectively-based effective decision-making is required: only a machine with a strongly artificial consciousness can satisfy the criteria for affective bodily, goal-directed, dynamically-coupled agency which makes possible the development and automatisation of appropriate responses and adaptive behaviours in socially complex and morally diverse environments. Placing a WAC in such an environment, where judgements and decisions have to be made in real-time and where it can neither feel nor think would be fruitless, if not dangerous.[18] Placing a SAC in such an environment is an ideal still some way off.

Bibliography

Aleksander, I. & Dunmall, B. (2003) 'Axioms and Tests for the Presence of Minimal Consciousness in Agents', *Journal of Consciousness Studies*, 10 (4-5), pp.7–18

Aleksander, I. (2005) *The World in My Mind, My Mind In The World: Key Mechanisms of Consciousness in Humans, Animals and Machines*, Imprint Academic, UK

Asimov, I. (1950) *I, Robot*, Fawcett Publications, Inc.

Bauby, J-D. (1997) *The Diving-Bell and The Butterfly*, London: Fourth Estate, Harper Perennial 2004

Brewer, B. (1992) 'Self-Location and Agency', *Mind* 101, pp.17–34

Chella, A. & Manzotti, R. (2007) *Artificial Consciousness*, Imprint Academic, UK

Chiel, H.J. & Beer, R.D. (1997) 'The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment', *Trends in Neurosciences* 20, pp.553–7

Clark, A. (1997) *Being there: putting brain, body, and world together again*, Cambridge, Mass.; London: MIT Press

[18]Having said this, there are times when a phenomenally affective conscious agent can entail ineffective decision-making, and where the operation of a WAC, an unconscious machine, would be preferable. In the case of the mid-air collision of the Bakshirian Airlines Tupolov 154 and the DHL Boeing 757 on 1 July 2002, the pilots gave the conscious human flight controller the benefit of the doubt as he over-ruled the 'unconscious' ACAS II-compliant collision-avoidance system's message; had they not, the disaster might just have been avoided.

Cole, J. (1995) *Pride and a daily marathon*, Cambridge, Massachusetts: MIT Press; orig. 1991 London: Duckworth

Cole, J. (2005) 'Imagination after neurological losses of movement and sensation: The experience of spinal cord injury', *Phenomenology and the Cognitive Sciences*, 4 (2) pp.183–195

Cotterill, R. M. J. (1995) 'On the unity of conscious experience', *Journal of Consciousness Studies*, Imprint Academic, Vol. 2, No. 4, pp.290–311

Cotterill, R. M. J. (1998) *Enchanted Looms: Conscious Networks in Brains and Computers*, Cambridge University Press

Damasio, A.R., Tranel, D. & Damasio, H. (1991) "Somatic markers and the guidance of behaviour: theory and preliminary testing", (pp. 217–229). In H.S. Levin, H.M. Eisenberg & A.L. Benton (Eds.). *Frontal lobe function and dysfunction*. New York: Oxford University Press

Damasio H., Grabowski T,. Frank R., Galaburda AM., Damasio AR. (1994) 'The return of Phineas Gage: clues about the brain from the skull of a famous patient', *Science*, 264 (5162), pp.1102–5

Damasio, A.R. (1994) *Descartes' Error: emotion, reason, and the human brain*, New York: Grosset/Putnam

Damasio, A.R. (1999) *The feeling of what happens : body, emotion and the making of consciousness*, New York: Harcourt Brace

Damasio, A.R., Grabowski, T.J., Bechara, A., Damasio, H., Ponto, L.L.B., Parvizi, J., & Hichwa, R.D.(2000) "Subcortical and cortical brain activity during the feeling of self-generated emotions", *Nature Neuroscience*, 3 (10) pp.1049–56, October 2000

Damasio, A.R. (2003) *Looking for Spinoza : joy, sorrow, and the feeling brain*, London: Harcourt

Dobbyn, C. & Stuart, S.A.J. (2003) 'The Self as an Embedded Agent', *Minds and Machines*, 13 (2) pp.187–201

Dreyfus, H. (1972/1979) *What Computers Can't Do: A Critique of Artificial Reason*, New York: Harper & Row

Dreyfus, H. (1992) *What Computers "Still" Can't Do: A Critique of Artificial Reason*, Revised edition, Cambridge, Mass.: MIT Press

Dreyfus, H. (1998) "Response to My Critics", in *The Digital Phoenix*, Terrell Ward Bynum, ed., Cambridge, Mass.: Blackwell

Ekman, P. (1992) 'Facial expressions of emotions: New findings, New questions', *Psychological Science*, 3 (1), pp.34–8

Flanagan, O. (1992) *Consciousness Reconsidered*, Cambridge, MA: MIT Press

Gallagher, S. (1986) 'Body image and body schema: A Conceptual Clarification', *Journal of Mind and Behavior*, 7 (4), pp.541–554

Gallagher, S. & Cole, J.D. (1995) "Body Schema and Body Image in a Deafferented Subject", *Journal of Mind and Behavior*, 16, pp.369–390

Gallagher, S. (2007) 'Moral Agency, Self-Consciousness ,and Practical Wisdom', *Journal of Consciousness Studies*, 14 (5–6), pp.199-223

Gibson, J.J. (1968) *The Senses Considered as Perceptual Systems*, London, George Allen & Unwin

Gibson, J.J. (1979) *The ecological approach to visual perception*, Boston, Mass.: Houghton Mifflin

Haikonen, P. (2003) *The Cognitive Approach to Conscious Machines*, Imprint Academic, UK

Harlow, J.M. (1848) 'Passage of an iron rod through the head', *Boston Medical and Surgical Journal* 39, pp.389-93. (Republished *Journal of Neuropsychiatry and Clinical Neuroscience*, intro. by T.C. Neylan 'Frontal Lobe Function: Mr Phineas Gage's Famous Injury', (1999) 11 (2), pp.281–83)

Harnad, S. (2003) "Can a Machine Be Conscious? How?", *Journal of Consciousness Studies*, 10 No. 45: 67–75

Head, H. & Holmes, G.M. (1911) 'Sensory disturbances from cerebral lesions', *Brain*, Oxford, Vol. 34: 102–254

Heidegger, M. (1962) *Being and Time*, trans. John Macquarrie & Edward Robinson, London: SCM Press

Holland, O. (2003) *Machine Consciousness*, New York, Imprint Academic

Holland, O. & Knight, R. (2006) "The Anthropomimetic Principle", Department of Computer Science, University of Essex

Ings, S. (2007) *The Eye: A Natural History*, London: Bloomsbury Publishing

Johnson, M. (1990) *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*, London: University of Chicago Press

Lakoff, G. & Johnson, M. (1999) *Philosophy in the Flesh : The Embodied Mind and Its Challenge to Western Thought*, New York: Basic

Legrand, D. (2006) 'The Bodily Self: The Sensori-Motor Roots of Pre-Reflexive Self-Consciousness', *Phenomenology and the Cognitive Sciences*, 5 (1), pp.89–118

Luria, A. R. (1973) *The Working Brain: an introduction to neuropsychology*, trans. Basil Haigh, London: Allen Lane

Maturana, H.R. & Varela, F.J. (1980) *Autopoiesis and Cognition*, Dordrecht: Reidel

Meijsing, M. (2000) 'Self-Consciousness and the Body', *Journal of Consciousness Studies*, 7 (6), pp.34–52

Merleau-Ponty, M. (1962) *Phenomenology of Perception*, trans. by Colin Smith, London: Routledge & Kegan Paul; New York: The Humanities Press

Merleau-Ponty, M. (1968) *The Visible and The Invisible*, Evanston : Northwestern University Press

Paillard, J. (2005) 'Vectorial versus configural encoding of Body Space, A neural basis for a distinc- tion between Body schema and Body image', V.Knockaert & H. De Preester (eds) *Body Image and Body Schema: Interdisciplinary perspectives*, (pp. 89–109), John Benjamin, Amsterdam

Peterson, S. (2007) "The ethics of robot servitude", *Journal of Experimental & Theoretical Artificial Intelligence*, 19(1), pp. 43-54

Sacks, O. (1984) *A Leg to Stand On*, Perennial Library, Harper and Row Publishers

Schwoebel, J., Friedman, R., Duda, N., & Coslet, H. B. (2001) ‘Pain and the body schema: Evidence for peripheral effects on mental representations of movement’, *Brain*, 124 (10), pp.2098–2104

Seitz, J. A. (2000) “The Bodily Basis of Thought”, *New Ideas in Psychology: An International Journal of Innovative Theory in Psychology*, 2000, 18(1), pp.23–40

Sheets-Johnstone, M. (1998) “Consciousness: A Natural History”, *Journal of Consciousness Studies*, 5 (3), pp. 260–94

Sheets-Johnstone, M. (1999) *The Primacy of Movement*, Amsterdam: J. Benjamins

Sheets-Johnstone, M. (2000) “Kinetic Tactile-Kinesthetic Bodies: ontogenetical foundations of apprenticeship learning”, *Human Studies*, 23, pp.343–70

Sheets-Johnstone, M. (2003) “Kinesthetic Memory”, *Theoria et Historia Scientiarum*, 7, pp.69–92

Sloman, A. “Varieties of affect and learning in a complete human-like architecture” accessed online July 2004 at `http://www.cs.bham.ac.uk/research/cogaff/talks/#talk24`

Sloman, A. “What are information-processing machines? what are information-processing vitual machines”, accessed online January 2005 at `http://www.cs.bham.ac.uk/~axs/misc/talks/information.pdf`

Stern, D. (1985) *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*, New York: Basic Books

Sternberg, E. J. (2007) *Are You a Machine? Tha Brain the Mind and What it Means to be Human* Amherst, NY: Prometheus Books

Stuart, S. (2007) “Machine Consciousness: Cognitive and Kinaesthetic Imagination”, *Journal of Consciousness Studies*, 14 (7), pp.141–53

Torrance, S. (2008) “Ethics and consciousness in artificial agents”, *ArtificiaI Intelligence & Society*, 22, pp.495521

Varela, F., Thompson, E., & Rosch, E. (2003) *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press, Cambridge, MA

Whitehead, A. N. (1929) *Process and Reality*, Cambridge: Cambridge University Press

Ziemke, T. (2003) “What’s that thing called embodiment?”, Proceedings of the 25th Annual Meeting of the Cognitive Science Society, Lawrence Erlbaum

Ziemke, T. (2007a) “What’s life got to do with it?”, in: Chella & Manzotti (eds.), *Artificial Consciousness* (pp. 48–66), Exeter: Imprint Academic

Ziemke, T. (2007b) “The Embodied Self: Theories, Hunches and Robot Models”, *Journal of Consciousness Studies*, 14 (7), pp.