Steve Torrance

Two Conceptions of Machine Phenomenality

Abstract: *Current approaches to machine consciousness (MC) tend to offer a range of characteristic responses to critics of the enterprise. Many of these responses seem to marginalize phenomenal consciousness, by presupposing a 'thin' conception of phenomenality. This conception is, we will argue, largely shared by anti-computationalist critics of MC. On the thin conception, physiological or neural or functional or organizational features are secondary accompaniments to consciousness rather than primary components of consciousness itself. We outline an alternative, 'thick' conception of phenomenality. This shows some signposts in the direction of a more adequate approach to MC.*

1. Strong and Weak Machine Consciousness

I shall argue that much existing work in machine consciousness (MC) operates with an inadequate philosophical view of consciousness. This may be called the *thin* (or *shallow*) conception of phenomenality. This conception is, as we'll see, also shared by many people who are, or would naturally be, critics of MC. I will discuss some limitations of the thin conception of phenomenality, and then sketch an alternative conception — *thick* (or *deep*) phenomenality. There are possibly several ways of articulating the notion of 'thick' phenomenality. In the version that I shall defend I shall be taking some cues from certain ways of thinking within the 'enactive' approach (Varela, Rosch and Thompson, 1991; Thompson 2004; 2005; 2007).

One may distinguish between 'weak' and 'strong' MC.[1] Weak MC seeks to model functional analogues to (or aspects of) consciousness. Strong MC aims to develop 'machines' that are (supposedly) *genuinely*

Correspondence: Centre for Research in Cognitive Science, Department of Informatics, University of Sussex, Falmer, Sussex BN1 9QH. *stevet@sussex.ac.uk*

[1] This of course echoes Searle's famous (1980) distinction between 'weak' and 'strong' AI.

conscious.[2] It might be as well to to review briefly the term 'machine' here. The bulk of current discussion about MC is, of course, in the context of computationally-based mechanisms of various sorts (whether robots or virtual agents). Does a (naturally occurring) organism count as a 'machine'? Or a hypothetical nanotechnological construction that replicates the chemical structure of an organism from the molecules up? In fact, the interesting question is, I think: Can a machine which is not a natural organism (or, say, an artificial nanoconstruct of a natural organism), be conscious? Cutting across the machine-organism divide is the natural-artificial divide — which is also problematic, of course. As medical, bionic, and other technologies develop, there may be creatures/beings/entities of which it will become harder and harder to say: 'that's a machine RATHER than an organism', or 'an organism RATHER than a machine'; or indeed 'that's natural RATHER than artificial' or 'artificial RATHER than natural'. (Stock examples like the beaver's dam shows how problematic the natural/artificial divide is even without exotic techno-fantasies.) Nevertheless we want to know if machines as envisaged today could ever be conscious in the way that many organisms *as envisaged today* are. Pointing to organisms nanoconstructed from the molecules up is perhaps an uninteresting answer to that question.

Notwithstanding these complications we can specify how weak and strong MC are distinguished, within the (admittedly limited) domain of currently available technologies. Both will be concerned with the design or creation of agents or agent-models whose mode of functioning essentially involves computational technologies (as these are currently envisaged) of one sort or another. Those who see themselves as engaged in weak MC will describe their activity in terms of modelling various aspects of natural consciousness with the purpose of better understanding the latter. Those who set themselves strong MC goals will be aiming to produce machines which have psychologically real (and perhaps ethically significant — see later) states of consciousness.

2. Functional and Phenomenal Consciousness, and Absent Qualia Arguments

A closely associated distinction to the weak/strong MC distinction is one between 'functional' and 'phenomenal' consciousness (see, for

[2] 'Genuinely conscious' could here mean conscious in the way that humans or other creatures we take to be conscious are; or it could mean conscious in some sui generis way which, though to be sharply contrasted with biological forms of consciousness, nevertheless shares some key features in common with the latter. For some problems in defining what consciousness might be, in the context of a strong MC programme, see Bringsjord (this issue).

example, Franklin, 2003). The distinction can be taken as a rough-and-ready version of Ned Block's (1995) possibly more specialized distinction between 'phenomenal' and 'access' consciousness.

Weak MC could perhaps be described as targeting only functional consciousness, while strong MC seeks to target phenomenality as well. That way of putting things isn't adequate, however, since some supporters of strong MC may deny that there is any sensible distinction between functional and phenomenal consciousness. For those who think the distinction is a valid one, on the other hand, creating a *functionally* conscious mechanism may still be seen as a kind of strong MC, in that it may be thought that such a product would instantiate at least one kind of psychologically real consciousness. Alternatively, producing functional consciousness in a machine might be considered to be kind of midway position between weak and strong MC.

Whatever the merits of the notion of merely functional as opposed to phenomenal consciousness, the idea of phenomenality is often thought not to sit easily within a computational framework. The attempt to explain phenomenality in computational terms is regarded by many as a special instance of the 'explanatory gap' (Levine, 1983) that is thought to affect any attempt to assimilate consciousness to physicalistic frameworks. Many of those who think the explanatory gap can be bridged in some way or other nevertheless believe that there is an explanatory tension between computation and consciousness. As we will see, many enthusiasts of MC tend to deal with that tension by reducing, downgrading or avoiding phenomenality in various ways.

Arguments against strong MC include versions of the absent qualia (AQ) argument. AQ arguments suggest that, for any set of putative computational/functional conditions for phenomenal consciousness, one can always consistently imagine those conditions obtaining but with phenomenal feel absent. To take a classic example, in Ned Block's 'Chinese Nation' argument (Block, 1978), one imagines a scenario meeting our proposed conditions but where the requisite computational operations are performed by some vast population of human operators. Such a scenario may involve much consciousness — all the myriad experiences of the legions of individual participants — but in so doing it leaves no room for the target phenomenal experience supposedly arising out of the computational operations themselves.

AQ-style anti-computationalist arguments in the style of the Chinese Nation describe scenarios where the relevant computational processing is present but where it is very difficult to believe that the relevant (or any) conscious states are present. Another kind of AQ argument deals with scenarios where the computational processing is

present and where it seems plausible to think that conscious states *may* be present, but where *a significant doubt may still remain* about the definitive presence of such conscious states. In such settings the AQ argument will proceed to the conclusion that no fully adequate or conclusive explanatory embedding of phenomenality in computational or cognitive conditions is possible. (For recent versions of AQ-style arguments of that sort see Block, 2002; Prinz, 2003).

There are various kinds of MC response to AQ arguments and to general doubts about the computational realizability of consciousness. All these responses, in some way, try to marginalize phenomenality. Here are three.

(a) *The eliminativist strategy:* Supporters of this strategy claim that notions such as phenomenality, qualia, etc., are conceptually confused, scientifically inadequate and unnecessary to the project of artificially creating genuinely conscious beings (Dennett, 1991; Harvey, 2002; Sloman & Chrisley, 2003; Blackmore, 2003).

(b) *The cognitivist strategy:* This strategy seeks to reconstrue phenomenal consciousness in terms of cognitive (or cognitive-affective) processes, that are more computationally 'friendly'. Examples are theories that associate consciousness with rich self-modelling processes, or with globally shared information-handling, but there are many other variants (Baars, 1988; Sloman & Chrisley, 2003; Holland, 2003; etc.).

(c) *The agnostic strategy:* On this strategy it is conceded that perhaps phenomenal consciousness may not be captured within a computational framework, but the claim is made that an important kind of consciousness — e.g. functional consciousness — may be created nonetheless. The question of whether artificial entities which display only this latter kind of consciousness could ever be 'fully' conscious (or indeed conscious in any genuine sense) is left open (Franklin, 2003).

These different strategies — and others[3] — tend to be combined or to flow into one another. The first two strategies are more easily associated with the strong MC approach, and the third perhaps with the weak MC approach, but this is only a loose grouping.

---

[3] Additional strategies worth mentioning are: (4) The *ascriptivist strategy* (claiming that consciousness is just a matter of people deciding to treat a being as conscious); (5) The *special pleading strategy* (MC is special and so can't be judged in terms of existing ideas of consciousness); (6) The *extreme generosity strategy* (if the system behaves in any way remotely as if it were conscious then we can treat it as conscious). These strategies also tend, I suggest, to marginalize phenomenality in some way.

By associating these various argumentative strategies with a certain conception of consciousness that I wish to criticize, it should not be taken that I think that the authors cited necessarily have a *superficial* or *unreflective* view of consciousness. On the contrary, all the MC-friendly authors cited no doubt offer some very deep insights into aspects of consciousness, artificial and/or natural. However it will be argued that there is a deep difficulty underlying much[4] existing work in the machine consciousness area.

3. Thin Phenomenality

All these strategies rely upon what I have called the 'thin' conception of phenomenal consciousness. The thin conception sees phenomenal consciousness rather in the way one might see the glint on a pair of patent leather shoes. One can imagine someone getting quite tangled up about how the shine gets to be on the shoe, perhaps taking it to be a rarified, evanescent, extra surface, not identifiable with the leather or even with the layer of polish that coats the leather, but which exists rather as a super-layer which somehow sits on top of both. A robust response to such a notion would be to either dismiss the whole idea of the shine as something extra to the shoe or to resort to a 'reductive' physical explanation in terms of the light-reflective properties of particular kinds of surfaces. Putting the point less indirectly: the idea of phenomenal consciousness may often be seen as something extra to all the information-processing going on in the brain; something which, when compared to all the brain activity itself, may be thought of as puzzling, difficult to pin down — even ornamental or epiphenomenal. This is perhaps a caricature, but supporters of the thin conception often seem to make us believe that this is how phenomenality might best be seen. And the challenge of phenomenality, when conceived of in this 'thin' way, can easily be either dismissed as confusion, or defused via a demonstration of how a rich enough information-processing story can capture all the 'specialness' that phenomenality seems to have.

The various strategies adopted to give support to strong MC all tend to rely on a certain view about phenomenal consciousness. But this

[4] Much, but far from all. Several contributions to the present volume, for example (e.g. Ziemke, Holland, Stuart) offer strongly embodied takes on MC which may well go far towards embracing forms of the 'thick' conception described below. Ziemke (this volume) even goes so far as to say 'that there is a convergence towards an *embodied* approach' (emphasis in the original). Nevertheless I think adherence to the thin conception is still very much a characteristic of many workers in the field. Also, as Ziemke himself points out in his excellent contribution (this volume; also 2007) there are significantly different ways that embodied MC can be taken.

view of consciousness — or certain key aspects of it — is also shared by many of those who reject strong MC. Thus, I will claim, both the anti-computationalist critiques of MC *and* the standard MC responses may be seen as based, to a greater or lesser degree, upon a similar, thin conception of phenomenality.

In order to see how critiques of MC draw upon the thin conception of phenomenality, consider AQ arguments of the sort discussed earlier. These all appear to trade on the apparent ease with which phenomenality can apparently be conceptually peeled away in any imagined scenario where that scenario is described in non-phenomenal terms. A common idea in AQ arguments (particularly 'zombie' variants of such arguments) is that a being can be imagined which has all the outward and internal organizational (i.e. functional) characteristics of a paradigmatically conscious being, but which lacks any 'inner life'. In such a picture, the phenomenal 'feel' of consciousness is indeed just like the glint on the patent leather - a special property which obstinately refuses to coalesce with the object's deeper aspects. Small wonder, then, that phenomenality may be so easily problematized and emasculated, as seems to happen within the various MC strategies commonly found.

Such arguments are fed by the idea that 'feel' *is all there is to consciousness* — so that the various physiological or sensorimotor or neural or organizational features investigated by consciousness scientists are secondary accompaniments to the process rather than primary components of the process itself. It is characteristic of the thin conception, then, that phenomenal feel is conceptually divorcible from any other features in an agent. And being so divorcible, it generates these two opposing philosophical camps — reductionists or eliminativists on the one hand and neo-dualists on the other. Neither of these sides is able to offer a convincing refutation of the other's position. It is this conceptual detachability, this 'unbearable lightness,' which may be seen as the objectionable feature of the thin conception of phenomenality — the key reason why it tends to lead to a showdown between computationalists and their opponents.

But there is an alternative, 'thicker' conception of the phenomenal. This can no doubt be characterised in various ways: I will develop it via the notion of *lived embodiment* — that is, in terms of the real, physical properties of organic, embodied beings who experience conscious subjectivity (taken to include environmental and intersubjective aspects as well as 'internal' aspects), as well as in terms of the subjective feeling itself.

4. From Thin to Thick

On a thick conception, a person's consciousness will be seen, not as conceptually detachable from everything else about that person, but rather as a deeply embedded, multidimensional, embodied, part of that person's nature, whose elements are interleaved in a multi-ply-stranded complex phenomenon[5] On this conception, arguments about absent qualia, zombies, and so on, will be harder — perhaps impossible — to state coherently. If phenomenal feel is conceived of as being essentially contextualized in a embodied, living being, then arguments based on supposedly conceivable scenarios where bodily, organic features are all present but the feel is absent, will simply lose their force. (Perhaps arguments feeding from such scenarios will never be subject to a decisive, knock-down refutation. Rather, their persuasive force will simply ebb away, as the alternative, essentially embodied, conception of phenomenality is progressively articulated.)

One source for developing a thick conception of phenomenality may be found within the enactive approach to mind, as developed by Varela, Thompson, and Rosch (1991) The enactive approach centres around the idea of 'lived embodiment' mentioned earlier. Such a conception is derived from the writings of Husserl and of Merleau-Ponty, but is also inspired by writings in theoretical biology, particularly work by Maturana and Varela on the so-called autopoietic mode of existence of organisms (see, for example, Maturana and Varela, 1987).

The relation between mind, body and organism (or animal existence) has been explored in a recent paper by Robert Hanna and Evan Thompson (2003; see also Thompson 2004; 2006; 2007). Hanna and Thompson discuss what they call the 'Mind–body–body problem', which they see as that of reconciling three different ways in which an individual 'I' can be understood. These are:

- as conscious subjectivity (i.e. phenomenality);
- as living, or lived body (*Leib*) with its own perspective or point of view; and
- as a physiological, corporeal, entity investigable within the natural sciences (*Körper*).

How, they ask, can a single individual incorporate all three of these different natures? Their proposed solution is that the lived

[5] Torrance (2004) develops this conception in terms of a 'Grand Inventory' of properties which together make up the 'deep' concept of embodied consciousness. For developments of a conception of consciousness that bear closely on the 'thick' conception as outlined here, see Stuart (this volume), Ziemke (2007; this volume).

embodiment of the individual (*Leib*) is ontologically basic, and that conscious phenomenality and physical corporeality are two aspects of the lived body. On this account subjectivity is essentially embodied, but its embodiment is not that of a merely physical body, but the lived embodiment of organism.

It should be noted that the sense of 'life' which is involved in the notion of 'lived embodiment' is not a purely biological sense (although it relates closely to the latter), but involves selfhood, perspective and purpose. It is a crucial part of the enactive conception of mind and conscious experience, taking its cue from the phenomenology of Husserl and others, that the status of having a mind is intimately related with the process of *living a life* in a teleological, rather than just a merely biological, sense.

Notice how this approach contrasts with traditional approaches to consciousness, as typified by the thin conception. On the thin view, consciousness is radically discontinuous with life. In particular (as we have seen), consciousness, 'thinly' conceived, generates an explanatory gap, in a way that the notion of *living* doesn't. There is thus claimed to be a logical gulf between experiencing and physical functioning, whereas modern biology has (supposedly) closed any such gulf between being alive and physical functioning. However, on the alternative view of consciousness-as-lived, there is a continuity between phenomenal experience, experiencing one's life as an embodied individual, and having a biological, physical existence. There is no more necessity to see a gap in the one case than in the other.

There are other theoretical strands which can be used to explicate the idea of lived embodiment. A central one concerns the idea of what it is to be an autopoietic, or self-recreating, individual. (Varela, 1979; Maturana and Varela, 1987) An autopoietic system — whether a unicellular or a more complex creature — acts to further its existence within its environment, through the appropriate exchange of its internal components with its surroundings, and via the maintenance of a boundary with its environment. In earlier versions of autopoietic theory, an autopoietic system was a special kind of machine — one which was in continuous activity to maintain its own existence. In recent developments of the notion (Weber & Varela, 2002; Thompson, 2004), autopoiesis is closely tied to the notions of sense-making and teleology, in the following way: autopoietic self-maintenance is a source or ground of meaning and purpose *for* that organism (where that meaning or purpose is *intrinsic* to the organism, rather than something which is merely the product of a pragmatically useful *interpretive attribution* on the part of an observer). On this view, autopoietic

entities are radically different from 'mere' mechanisms, since, unlike the latter, they *enact* their own continued existence, and their own purpose or point of view.

It is a matter of some dispute whether the defining properties of autopoiesis can be found outside the realm of the truly biological, and it is thus an open question as to whether there is any sense in which computationally based constructs could ever be seen as being assimilable to an autopoietic framework — that is as original self-enacting loci of meaning and purpose, or indeed of consciousness. (See, for example, Ruiz-Mirazo and Moreno, 2004; McMullin, 2004; Bourgine and Stewart, 2004; Ziemke, this volume.) Clearly, any programme of producing enactive artificial agents would involve a great shift in design philosophy from that which prevails today in most AI or computing science circles. Ezequiel Di Paolo (2003; 2005) is one writer who believes that a programme of developing artificial autopoietic agents, with intrinsic teleology, at least provides a reasonable research objective. If any MC programme is to succeed in its goal of capturing a conception of artificial consciousness compatible with a fully adequate picture of our own human lived experience, then it has to go down a path of this sort.

So could there be a 'strong' MC programme based on a 'thick' conception of phenomenality? If the 'thick' conception takes phenomenal feel to be *deeply embodied,* as conceptually inseparable from the underlying natural organic, living features of biological beings, then what room could there be for the design and development of artificial (non-biological) beings that merited being called 'conscious' in such a sense? Wouldn't the thick conception be taking the MC programme further away from its goal?

Sure, the thick conception doesn't make the strong MC project any easier — quite the reverse. But perhaps it doesn't make it an unrealizable goal (Stuart, this volume; Ziemke, 2007). In building bridges from the human/mammalian consciousness we know to possible artificial forms, our conception of consciousness must necessarily broaden. A Kuhn-style indeterminacy will affect this broadening (the space of discussion isn't, for all that, arbitrary). We shouldn't expect a crisp set of success-conditions for the achievement of 'genuine' (strong) MC. But neither should we expect that such a goal can be ruled out in a peremptory manner by some neat chain of reasoning.[6]

---

[6] Not all those writing about MC can be easily categorized as thick or thin, and many make remarks which could place them in either camp — or with a foot in both. For example, Kiverstein (this volume) draws upon certain strands of the 'enactive' tradition, associated

5. MC and Moral Status

I will end by discussing briefly how this enactively inspired conception of the 'thick' view of consciousness has important consequences for how one views the *moral* status of consciousness (see also Torrance, 2004; forthcoming). Autopoiesis applies to self-maintaining agents of even the most primitive kind, yet it provides an essential element of what is involved in an adequate conception of highly developed, intelligent autonomous moral agency.

Viewing beings as autonomous centres of meaning and purpose, as living and embodied conscious agents that enact their own existence, is, I believe, an important ingredient of building up a *moral* picture of ourselves, and of those beings we wish to create. It necessitates the consideration, of the ethical status of such beings — for surely we will want them to be created in our moral image. On this picture, an agent will be seen as an appropriate source of moral agency only because of that agent's status as a self-enacting being that has its own intrinsic purposes, goals and interests. Such beings will be likely to be a source of intrinsic moral concern, as well as, perhaps, an agent endowed with inherent moral responsibilities. They are likely to enter into the web of expectations, obligations and rights that constitutes our social fabric. It is important to this conception of moral agency that MC agents, if they eventualize, will be our companions — participants with us in social existence — rather than just instruments or tools built for scientific exploration or for economic exploitability.

Thus the MC quest, when understood in terms of a 'thick', or 'rich' conception of consciousness as lived embodiment, has important moral reverberations. One would be guilty of a failure of reflection if one did not see that any genuinely conscious creature that might result from an MC programme informed by such a conception of consciousness, would set us a great deal of moral puzzles — not the least of which is whether such a programme should be even started upon. There is a growing recognition of the inherent moral dimensions of the MC enterprise. Thomas Metzinger, for example (2003), expounds at some length his view that consciousness in a system is bound up with that system's phenomenal self model (PSM). Metzinger writes

---

with an embodied agent's exercise of patterns of sensorimotor knowledge (Hurley,1998; O'Regan and Noë, 2001). However his central claim that patterns of dynamical sensorimotor interaction of the right sort in a robot may be sufficient to allow phenomenal consciousness to be attributed to that robot suggests to me a strong form of functionalism that is much closer to the 'thin' approach; and its not clear to me that these two parts of his view sit together very comfortably. Nevertheless, his paper offers an interesting challenge to the views put forward here.

that the possession of such a PSM will inevitably involve negative affective consequences — suffering — for the system, consequences that have a *moral* weight:

> Suffering starts on the level of PSMs. You cannot consciously suffer without having a globally available self-model. The PSM is the decisive neurocomputational instrument not only in developing a host of new cognitive and social skills but also in forcing any strongly conscious system to functionally and representationally appropriate its own disintegration, its own failures and internal conflicts… .The melodrama, but also the potential tragedy of the ego both start on the level of transparent self-modeling. *Therefore we should ban all attempts to create (or even risk the creation of) artificial and postbiotic PSMs from serious academic research* (Metzinger, 2003, p. 622. My italics).[7]

Metzinger's point may be thought to be somewhat overstated – but it deserves consideration. The fact that so much discussion of machine consciousness has in the past been conducted more or less in a moral vacuum is itself a testimony to the weakness of conceptions of consciousness that have operated in the field. Certainly the moral dimensions of entering into an age of artificially conscious creatures need to be very carefully assessed.

6. Conclusion

The goal of producing a truly conscious machine may be further away than people would like to think To achieve such a goal it is, I have argued, necessary to radically reprogram one's conception of consciousness, in such a way that consciousness is deeply related to lived embodiment. The resulting revised understanding of machine consciousness will need careful analysis: it is not clear that anything (natural or artificial) that could be conscious in this revised sense could count as a ('mere') machine. At the very least the notion of 'machine' that would need to be operative would have to be very closely intertwined with the notion of 'organism'; artificial consciousness as a field would need to take its inspiration from biology in a much deeper sense than is currently envisaged by most in the field.

Also, the considerations proposed here suggest reducing one's confidence in the belief that the strong MC programme might eventually succeed — at least on the basis of the current known technologies. However I am not arguing that it cannot be ruled out in principle. Also, importantly, it can't be ruled out (as many opponents of MC would do currently) on the basis of arguments which, whether

[7] I am grateful to Owen Holland for drawing my attention to this passage from Metzinger's book. See also LeChat (1986).

expressly or no, presuppose a 'thin' conception of phenomenality. Nor, I claim, can arguments to rule it in be successfully launched on the basis of such a conception.

Working out the details of any serious MC programme will involve much further theoretical discussion, which will go hand in hand with actual MC development, but also must centrally include an ongoing assessment of how social and moral attitudes towards AI and artificial agents might evolve.[8]

References

Aleksander, I. and Dunmall, B. (2003), 'Axioms and tests for the presence of minimal consciousness in agents', *Journal of Consciousness Studies*, **10** (4–5), pp. 7–18.

Baars, B. (1988), *A Cognitive Theory of Consciousness* (Cambridge: Cambridge University Press).

Blackmore, S. (2003), 'Consciousness in meme machines', *Journal of Consciousness Studies*, **10** (4–5), pp. 19–30.

Block, N. (1978), 'Troubles with functionalism', in *Minnesota Studies in the Philosophy of Science, IX*, ed. C.W. Savage, pp. 261–325.

Block, N. (1995), 'On a confusion about a function of consciousness', *Behavioral and Brain Sciences*, **18** (2), pp. 227–47

Block, N. (2002), 'The harder problem of consciousness', *Journal of Philosophy*, **XCIX** (8), pp.1–35.

Bourgine, P. and Stewart, J. (2004), 'Autopoiesis and cognition', *Artificial Life*, **20** (3), pp. 327–45.

Bringsjord, S. (this volume), 'Offer: One billion dollars for a conscious robot. If you're honest, you must decline', *Journal of Consciousness Studies*.

Dennett, D. (1991), *Consciousness Explained* (Boston, MA: Little, Brown.)

Di Paolo, E. (2003), 'Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop', in *Dynamical Systems Approach to Embodiment and Sociality*, ed. K. Murase and T. Asakura (Adelaide: Advanced Knowledge International), pp.19–42.

Di Paolo, E. (2005), 'Autopoiesis, adaptivity, teleology, agency', *Phenomenology and the Cognitive Sciences*, **4** (4), pp. 429–52.

Franklin, S. (2003), 'IDA: A conscious artefact?', *Journal of Consciousness Studies*, **10** (4–5), pp. 47–66

Hanna, R. and Thompson, E. (2003), 'The mind-body-body problem', *Theoria et Historia Scientiarum: International Journal for Interdisciplinary Studies*, **7**, pp. 24–44.

Harvey, I. (2002), 'Evolving robot consciousness: The easy problems and the rest', in *Consciousness Evolving*, ed. J. Fetzer (Amsterdam: John Benjamins).

Holland, O. and Goodman, R. (2003), 'Robots with internal models: A route to machine consciousness?', *Journal of Consciousness Studies*, **10** (4–5), pp. 77–109

Holland, O. (this volume), 'A strongly embodied approach to machine consciousness', *Journal of Consciousness Studies*.

---

[8] I have benefited greatly from discussions with many people over issues in this paper, particularly Mike Beaton, Ron Chrisley, Robert Clowes, Ezequiel Di Paolo, Tom Froese, Owen Holland, Susan Stuart and Blay Whitby.

Hurley, S. (1998), *Consciousness in Action* (Harvard, MA: Harvard University Press).
Kiverstein, J. (this volume), 'Could a robot have a subjective point of view?', *Journal of Consciousness Studies*.
LeChat, M. (1986), 'Artificial intelligence and ethics: An exercise in moral imagination', *AI Magazine*, **7** (2), pp. 70–79.
Levine, J. (1983), 'Materialism and qualia: The explanatory gap', *Pacific Philosophical Quarterly*, **64**, pp. 354–61.
McMullin, B.(2004), 'Thirty years of computational autopoiesis: A review', *Artificial Life*, **20** (3), pp. 277–96.
Maturana, H. & Varela, F. (1980) Autopoiesis and cognition. (Boston: Reidel)
Maturana, H.R. and Varela, F.J. (1987), *The Tree of Knowledge: The Biological Roots of Human Understanding* (Boston, MA: Shambala Press/New Science Library).
Metzinger, T. (2003), *Being No One: The Self-model Theory of Subjectivity* (Cambridge, MA: MIT Press).
O'Regan, J. K. and Noë, A. (2001), 'A sensorimotor account of vision and visual consciousness', *Behavioral and Brain Sciences*, **24**, pp. 939–1011.
Prinz, J. (2003), 'Level-headed mysterianism and artificial experience', *Journal of Consciousness Studies*, **10** (4–5), pp. 111–32.
Ruiz-Mirazo, K. & Moreno, A. (2004), 'Basic autonomy as a fundamental step in the synthesis of life', *Artificial Life*, **10** (3), pp. 235–60.
Searle, J. (1980), 'Minds, brains and programs', *Behavioral and Brain Sciences*, **3** (3), pp. 417–57.
Sloman, A. and Chrisley, R. (2003), 'Virtual machines and consciousness', *Journal of Consciousness Studies*, **10** (4–5), pp. 133–72.
Stuart, S. (this volume), 'Machine consciousness: Cognitive and aesthetic imagination', *Journal of Consciousness Studies*, .
Thompson, E. (2004), 'Life and mind: From autopoiesis to neurophenomenology. A tribute to Francisco Varela', *Phenomenology and the Cognitive Sciences*, **3**, pp. 381–98.
Thompson, E. (2005), 'Sensorimotor subjectivity and the enactive approach to experience', *Phenomenology and the Cognitive Sciences*, **4** (4), pp. 407–27.
Thompson, E. (2007), *Mind in Life: Biology, Phenomenology, and the Sciences of Mind* (Cambridge, MA: Harvard University Press.)
Torrance, S. (2004), 'Us and them: Living with self-aware systems', in I. Smit, W. Wallach and G. Lasker (eds), *Cognitive, Emotive and Ethical Aspects of Decision Making In Humans and In Artificial Intelligence*, Vol. III, (Windsor, Ontario: IIAS), pp. 7–14.
Torrance, S. (forthcoming, 2008), 'Ethics and consciousness in artificial agents', *Artificial Intelligence and Society.*
Varela, F. (1979), *Principles of Biological Autonomy* (New York: Elsevier North Holland).
Varela, F., Thompson, E. & Rosch, E. (1991), *The Embodied Mind: Cognitive Science and Human Experience* (Cambridge, MA: MIT Press).
Weber, A. & Varela, F. (2002), 'Life after Kant: Natural purposes and the autopoietic foundations of biological individuality', *Phenomenology and the Cognitive Sciences*, **1** (1), pp. 97–125.
Ziemke, T. (2007), 'What's life got to do with it?', Chella, A. and Manzotti, R. (eds), *Artificial Consciousness* (Exeter: Imprint Academic) pp. 48–66.
Ziemke, T. (this volume),. 'The embodied self: Theories, hunches and robot models', *Journal of Consciousness Studies*,