

Thin Phenomenality and Machine Consciousness

Steve Torrance

Institute for Social and Health Research
Middlesex University
Queensway, Enfield,
Middlesex EN3 4SF UK
s.torrance@mdx.ac.uk

and

Centre for Research in Cognitive Science
University of Sussex
Falmer, Brighton,
Sussex BN1 9QH UK
stevet@sussex.ac.uk

Abstract

Current-generation approaches to machine consciousness (MC) have a number of characteristic responses to arguments levelled against the enterprise. These responses tend to marginalize phenomenal consciousness. They do so by presupposing a ‘thin’ conception of phenomenality which is, in fact, largely shared by anti-computationalist critics of MC. The thin conception sees phenomenality as something that can be easily ‘peeled away’ from the rest of the physical world. On the thin conception, physiological or neural or functional or organizational features are secondary accompaniments to consciousness rather than primary components of consciousness itself. This inadequate conception bedevils much debate about the nature of consciousness. Can there be a more adequate MC programme, operating on an alternative, ‘thick’ conception of phenomenality? Recent ‘enactive’ approaches to consciousness perhaps show some signposts in the right direction.

1 Introduction

In order to prepare the path for next- and future-generation approaches to machine consciousness (MC), I propose to look at some problems in current MC research. Everyone agrees that current MC research has shortcomings – that’s why we’re here at this workshop. But the ones that I will be pointing out may not be the ones that you think you came here to discuss, or ones you recognize! In pointing out certain inadequacies in current work I do not wish to minimize the value of such work both for pushing forward the frontiers of artificial consciousness, and for understanding the nature of natural consciousness. However work that has great value may nevertheless be subject to unrealistic expectations or shaky presuppositions that need to be brought to light to enable fresh directions to be pursued.

Mine is a philosophical exploration. The excuse for philosophers to get involved with practical research in working MC systems is that the broad goals and presuppositions of such practical research

constantly need to be made explicit, evaluated and re-shaped, in the light of the constantly moving horizon of theoretical work in cognitive and consciousness science. The current discussion takes its inspiration from a particular wave in theoretical cognitive science, which has achieved a high profile in the last few years, namely the ‘enactive’ approach (Varela et al 1991, Thompson 2004).

It’s common to hear people who adopt the enactive approach arguing that most previous work in cognitive science has been labouring under various misapprehensions. My aim here is to spell out an argument along those lines, deployed specifically in relation to the field of artificial or machine consciousness.

I shall argue that much existing work in artificial consciousness operates with an inadequate philosophical view of consciousness, which may be called the *thin* (or *shallow*) conception of phenomenality. This conception is in fact also shared by many critics of MC. I will discuss some limitations of thin phenomenality, and then sketch an alterna-

tive conception – *thick* (or *deep*) phenomenality, taking some cues from enactive ways of thinking. I suspect that MC researchers may be rather resistant to the conclusions I come to for they imply that success in achieving machine consciousness may be a lot more remote than is currently thought; and that much of the work to date has been looking for those elusive car-keys under quite the wrong streetlamp. Indeed the right streetlamp may not be on this street or the next, but perhaps in another town or continent!

2 Strong and weak machine consciousness

Echoing Searle (1980), one may distinguish between ‘weak’ and ‘strong’ MC. Weak MC seeks to model functional analogues to (or aspects of) consciousness. Strong MC aims to develop computational mechanisms that are *genuinely* conscious, which have consciousness which is as little distinguishable as possible from our own conscious experience. Hanging on the word ‘genuine’ are, no doubt, a host of begged questions, not to be unduly picked over here. It’s common enough to hear people say that such and such a working system is ‘genuinely’ X - where X is some psychological property - when that system has as little relation to real cases of X-ing as blood oranges¹ have to do with real blood!

Well, it may be easier to get (real) blood out of an orange than out of a stone. And easier to get consciousness out of a machine than out of a stone, perhaps, if the machine is sufficiently elaborately designed? Part of the problem here is that the boundaries of what counts as a machine are intrinsically tentative at any given time, given the continual developments in technology. Turing tried to fix the relevant notion of machinehood in 1950, in a highly restrictive and abstract way. I am not sure how many present-day MC researchers would regard themselves as bound by those restrictions. But unless a clear definition is given of what counts as a machine and what doesn’t (for example, are organisms machines, if non-artificial ones?) it’s difficult to state clearly what strong MC actually amounts to.

Those who would see themselves as engaged in weak MC will avoid a lot of these kinds of difficulties. They will see the MC enterprise in terms of modelling various aspects of natural consciousness with the purpose of better understanding the latter,

¹ Or indeed blood-orange flavoured chocolate – a popular brand of chocolate is currently being promoted in that particular flavour!

rather than duplicating it via a kind of computational trans-substantiation. Those who see their research activity in terms of weak MC goals may nevertheless believe in the realizability of strong MC in principle. What I’m going to say will be relevant to both supporters of strong and of weak MC, but will be particularly relevant to the former.

3 Functional and phenomenal consciousness

A closely associated distinction that may be made is one between ‘functional’ and ‘phenomenal’ consciousness. One recent discussion of the distinction between phenomenal and functional consciousness is to be found in Franklin, 2003. The distinction can be taken as a rough-and-ready version of Ned Block’s (1995) more carefully worked out, but possibly more specialized, distinction between phenomenal and access consciousness.

Weak MC may be represented as targeting only functional consciousness, while strong MC seeks to target phenomenality as well. That way of putting things may not be thought altogether adequate, however: many supporters of strong MC will deny that there is any sensible distinction between functional and phenomenal consciousness. For those who think the distinction is a valid one, creating a merely functionally conscious mechanism may be seen as a kind of strong MC, in that such a product would instantiate at least one kind of ‘genuine’ consciousness. Alternatively it might be considered to be kind of midway between weak and strong MC.

Whatever the merits of the notion of merely functional consciousness as opposed to phenomenal consciousness, the idea of phenomenality is often thought not to sit easily within a computational framework. There is a widely shared feeling that computational processes and phenomenal feel are conceptually disjoint categories. 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. Enthusiasts of MC – particularly strong MC – tend to deal with that tension by reducing, downgrading or avoiding phenomenality in various ways, as we will see.

4 Absent qualia arguments and MC responses

Arguments against the strong MC programme 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 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 inviting to think that conscious states may be present, but where it is nevertheless insisted a significant doubt may still exist about the existence of such conscious states. Hence, the argument goes, no fully adequate 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).

I will discuss three kinds of MC response to AQ arguments and to general doubts about the computational realizability of consciousness: the eliminativist, the cognitivist, and the agnostic strategies. All these responses, in some way, try to marginalize phenomenality. There may be other strategies, but these are the main ones, as far as I can see.

(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 is left open. (Franklin 2003).

These different strategies 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 principle of grouping.

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 have a *superficial* view of consciousness. On the contrary, all the MC-friendly authors cited offer some very deep insights into aspects of consciousness, both natural and artificial (as the latter might be). However I feel that there is a deep difficulty underlying existing work in the MC area, and this is what I'm trying to bring to light.

5 The 'thin' conception of phenomenal consciousness.

All these strategies rely upon what I call the 'thin' conception of phenomenal consciousness. The thin conception sees phenomenal consciousness rather like the glint on a pair of patent leather shoes. One can imagine someone getting quite philosophically tangled up about how the shine gets to be on the shoe, perhaps taking it to be a rarified, evanescent, extra surface, not equatable 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. In a similar way the idea of phenomenal consciousness as something extra to all the information-processing going on in the brain can be either dis-

missed as confusion, or defused by showing how a rich enough information-processing story can capture all the 'specialness' that phenomenality seems to have.

However I would claim that these arguments in defence of strong MC actually buy into a certain view about phenomenal consciousness which is shared by those who reject strong MC. That is, both the anti-computationalist critiques of MC and the standard MC responses are based upon a similar, thin conception of phenomenality.

Thus AQ arguments of the sort discussed earlier 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'. On such a view the phenomenal feel of consciousness is just like the evanescent glint on the patent leather - a special property which obstinately refuses to coalesce with the object's deeper parts. Small wonder, then, that phenomenality may be so easily problematized and emasculated or shelved, as it is 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 essential to 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, neither of which 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 leads to the familiar showdown between computationalists and their opponents.

6 Towards an alternative conception of phenomenality

But is there an alternative conception? What might it consist in? What would a 'thicker' or 'deeper' conception of phenomenality consist of? I suggest

that it would need to be couched in terms of *essential lived embodiment* - in terms of the real, physical properties of organic, embodied beings who experience conscious subjectivity, plus environmental and intersubjective aspects, as well as in terms of the subjective feeling itself. On an alternative, 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 multiply-stranded complex phenomenon. (See Torrance 2004 for a development of this conception in terms of a 'Grand Inventory' of properties which together make up the 'deep' concept of embodied consciousness.)

On the thick conception, arguments about absent qualia, zombies, and so on, would 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 not be subject to a knock-down refutation - rather their persuasive force will simply ebb away, as the alternative, essentially embodied, conception of phenomenality is progressively articulated.)

But could there be a strong machine consciousness programme based on a 'thick' conception of phenomenality? If the 'thick' conception sees phenomenal feel as *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?

I think there are no easy answers to these questions. The thick conception doesn't make the strong MC project any easier - quite the reverse. But it doesn't necessarily make it an unrealizable goal. 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.

7 Lived embodiment

One source for developing a thick conception of phenomenality is, I suggest, to be found in the enactive approach developed by Varela, Thompson, and Rosch (1991). The enactive approach to mind 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, and forthcoming). 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 can a single individual incorporate all three of these different natures? Their proposed solution is that the lived 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 radically embodied, but its embodiment is not that of the 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 to the biological sense), 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 this autobiographical, rather than just merely biological, sense. Notice how this approach contrasts with traditional approaches to consciousness, as typified by the thin conception. On views of the latter sort consciousness is radically discontinuous with life. In particular (as we have seen), consciousness generates an explanatory gap on such views, and in a way that 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, enactive, view there is a continuity between phenomenal experience, living one’s life as an embodied individual, and having a biological, physical existence. There is no necessity to see a gap more in the one case than in the other.

8 Autopoiesis and MC

There are many 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, etc.) 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: that is, 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, Bourguin and Stewart 2004.) 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; and forthcoming) is one writer who believes that a programme of developing artificial autopoietic agents, with intrinsic teleology, at least

provides a reasonable research objective. However even he seems to stop short of proclaiming the possibility of computationally-based consciousness, where the latter is understood in this context. Yet in my view, if any MC programme is to succeed in its goal of capturing a conception of consciousness compatible with a fully adequate picture of our own human lived experience, then it has to go down a path of this sort.

9 MC and moral status

This enactively inspired version of the ‘thick’ conception of consciousness has, I believe, important consequences for how one views the *moral* status of an individual (see Torrance, 2003, 2004). 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 those we wish to create 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.

Clearly, the MC quest, when understood in terms of a ‘thick’, conception of consciousness as lived embodiment, raises important moral questions. 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). (I am sure that possessing a PSM in

something like Metzinger’s sense is a part of what it is to be a ‘lived embodiment’; whether it is sufficient remains to be seen. Metzinger writes that the possession of such a PSM will inevitably involve negative as well as positive 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, 622. My italics)²

Metzinger’s conclusion may be thought somewhat extreme – 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 superficiality of the conception of consciousness that has often operated in the field. Certainly the moral dimensions of entering into an age of artificially conscious creatures need to be very carefully assessed.

10 Conclusion

Machine consciousness research – current and future – has a lot more to do with real consciousness than blood-oranges have to do with real blood. However, 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 am arguing, 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

² I am grateful to Owen Holland for drawing my attention to this passage from Metzinger’s book. See also LeChat 1986, cited by Calverley in his contribution to this symposium.

much more profound 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 it cannot be ruled out in principle. Also, it can't be ruled out (as many opponents of MC would do currently) on the basis of arguments which, whether expressly or no, presuppose a 'thin' conception of phenomenality. Nor 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 with an ongoing assessment of how social and moral attitudes towards AI and artificial agents are evolving.

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), 7-18.
- Baars, B. (1988) *A Cognitive Theory of Consciousness*. Cambridge, England: Cambridge University Press.
- Blackmore, S. (2003) Consciousness in meme machines. *Journal of Consciousness Studies*. 10 (4-5), 19-30
- Block, N. (1978) Troubles with functionalism. In C.W.Savage, ed., *Minnesota Studies in the Philosophy of Science*, IX, 261-325
- Block, N. (1995) On a Confusion about a Function of Consciousness", *Behavioral and Brain Sciences* 18, 2, 227-247
- Block, N. (2002) The harder problem of consciousness. *Journal of Philosophy* XCIX, 8, 1-35
- Bourgine, P. and Stewart, J. (2004) Autopoiesis and cognition. *Artificial Life* 20 (3) 327-345.
- Dennett, D. (1991) *Consciousness Explained*. Boston: Little, Brown.
- Di Paolo, E. (2003) Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In K. Murase and T. Asakura (Eds.) *Dynamical systems approach to embodiment and sociality*. Adelaide: Advanced Knowledge International, pp.19-42.
- Di Paolo, E. (forthcoming) Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*.
- Franklin, S. (2003) IDA: A conscious artefact? *Journal of Consciousness Studies*, 10 (4-5), 47-66
- Hanna, R. and Thompson, E. (2003) The mind-body-body problem. *Theoria et Historia Scientiarum: International Journal for Interdisciplinary Studies* 7.
- Harvey, I. (2002) Evolving robot consciousness: the easy problems and the rest. In J. Fetzer (ed) *Evolving Consciousness*, 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), 77-109
- LeChat, M. (1986) Artificial intelligence and ethics: an exercise in moral imagination. *AI Magazine* 7 (2) 70-79.
- Levine, J. (1983) Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly* 64: 354-361
- McMullin, B. (2004) Thirty Years of Computational Autopoiesis: A Review. *Artificial Life* 20 (3). 277-296.
- 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: Shambala Press/New Science Library.
- Metzinger, T. (2003) *Being No One: The Self-model Theory of Subjectivity*. Cambridge, MA: MIT Press.
- Prinz, J. (2003) Level-headed mysterianism and artificial experience. *Journal of Consciousness Studies*, 10 (4-5), 111-132.

- Ruiz-Mirazo, K., & Moreno, A. (2004) Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life*, 10 (3). 235 – 260.
- Searle, J. (1980) Minds, brains and programs. *The Behavioral and Brain Sciences*. 3. 417-24.
- Sloman, A. and Chrisley, R. (2003) Virtual machines and consciousness. *Journal of Consciousness Studies*. 10 (4-5). 133-172.
- Thompson, E. (2004) Life and mind: from autopoiesis to neurophenomenology. A tribute to Francisco Varela. *Phenomenology and the Cognitive Sciences* 3: 381-398.
- Thompson, E. (forthcoming) Sensorimotor Subjectivity and the Enactive Approach to Experience, *Phenomenology and the Cognitive Sciences*.
- Torrance, S. (2003) Artificial Intelligence and Artificial Consciousness: Continuum or Divide? in I. Smit, W. Wallach and G. Lasker (eds), *Cognitive, Emotive And Ethical Aspects Of Decision Making In Humans And In Artificial Intelligence, Vol. II*, pp. 25-29, Windsor, Ontario: IIAS.
- 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.
- 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, 1991.
- Weber, A., & Varela, F. (2002) Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, 1, 97-125.