

Contemporary Sensorimotor Theory: A Brief Introduction

J. Mark Bishop and Andrew O. Martin

Dept. Computing, Goldsmiths, University of London, UK
m.bishop@gold.ac.uk

1 Background

‘Sensorimotor Theory’ offers a new *enactive approach*¹ to perception that emphasises the role of motor actions and their effect on sensory stimuli. The seminal publication that launched the field is the target paper co-authored by J. Kevin O’Regan and Alva Noë and published in Behavioral and Brain Sciences (BBS) for open peer commentary in 2001 [27].

In the central argument of their paper, O’Regan and Noë suggest radically shifting the nexus of research in visual perception away from analysis of the raw visual patterns of stimulation, to refocus on the law-like changes in visual stimulation brought about as a result of an agent’s actions in the [light-filled] world.

A key consequence of this change is a new way of characterising objects by the unique set of ‘sensorimotor correspondences’ that define the characteristic changes in objective appearance brought about by the agent-object interactions [in the world]. These characteristic correspondences - relating the movement of any object relative to the agent - define its *sensorimotor dependencies* [qua world]; an agent’s practical knowledge of these sensorimotor dependencies constitutes its visual experience.

Thus in O’Regan and Noë’s sensorimotor theory, perhaps for the first time, we have a rich, testable, psychological (and philosophically grounded) theory that accounts for *why* our conscious experience of the world appears as it does. This is a significant achievement and one that, in our opinion, goes a long way to answering at least some of the *hard problems of consciousness*².

This shift of nexus was, in part, informed by experiments in visual perception performed by O’Regan; for example, O’Regan, Rensink and Clark’s discovery

¹ The term ‘enactive approach’ is taken from Noë [25] where he states, “*What I call here the enactive approach was first presented in [27]. We refers to the view as the sensorimotor contingency theory. Hurley and I, in joint work, deploy another term: the dynamic sensorimotor account. I borrow the term enactive from Francisco Varela and Evan Thompson (Varela, Thompson and Roesch 1991 [49]), although I may not use it in exactly their sense. I use the term because it is apt, and to draw attention to the kinship of our view and theirs.*”

² David Chalmers introduced the term ‘hard problem’ to investigate, “Why is all this [neural] processing accompanied by an inner life?” [12]; we deploy the phrase ‘hard problems of consciousness’ to additionally encompass related problems pertaining to Levine’s ‘explanatory gap’ [21].

of *change blindness*³ presents a challenge to the classical characterisation of perception as resulting in the construction of a rich, hi-fidelity mental representation⁴. To account for such experimental data (and many other peculiarities of [visual] consciousness⁵) a new conception of visual perception was outlined by O'Regan and Noë in their BBS article *A sensorimotor account of vision and visual consciousness*:

.. the central idea of our new approach is that *vision is a mode of exploration of the world that is mediated by knowledge of what we call **sensorimotor contingencies***⁶.

Thus O'Regan and Noë's sensorimotor framework for perception shifted the problem of vision away from that of construction of rich internal representations of an 'out there' world to that of active exploration of the environment 'on demand'; conscious experience being *brought forth* via a series of [saccadic] movements that either confirm (or disabuse) the notion that the world *actually is* of the form currently anticipated⁷. In this view the world is effectively used as its own external memory and the 'objects of the world' as their own 'representation'; hence removing at a stroke the absolute requirement to store and update a rich, hi-fidelity internal representation of the rich, out-there world.

This concept of perception as an 'active interrogation' (i.e. the exploration of external visual features as and when they are required) contra 'passive snapshot' (and subsequent rumination over rich internal mental representations) naturally accounts for why some perceptual changes might go unnoticed - *even if they are specifically attended to*⁸ - if they are not in conflict with the agent's [current] anticipation(s).

³ .. where significant portions of a subject's visual field can change and yet not be consciously experienced due either to the transition being gradual, or the subject's attention being distracted by other simultaneous changes; e.g. by the appearance so called 'mud-splat' distractors across the visual field.

⁴ Problems arise in the classical view because, for example, large visual changes would naturally be anticipated to be perceived if there were large-enough differences between the state of the environment and the state of the internal representation.

⁵ For example, the extra-retinal signal; trans-saccadic fusion; saccadic suppression; 'filling in' of the blind spot and other retinal non-homogeneities.

⁶ In this introduction, after Broackes [27], we prefer the term 'sensorimotor dependencies' rather than 'sensorimotor contingencies', as the latter "has perhaps unfortunate connotations of non-necessity".

⁷ O'Regan [29] reports the neuroscientist Donald MacKay as asserting, "*the eye was like a giant hand that samples the outside world*".

⁸ For example, by slowly changing the colour of a merry-go-round, O'Regan and team demonstrate how large-scale changes to a visual scene can go unnoticed by even the most attentive of perceivers. NB. At first sight the phenomena of change blindness seems similar to the phenomenon of 'inattention blindness', where something that is fully in view is not noticed because attention is elsewhere (cf. Daniel Simons' now ubiquitous invisible 'Gorilla in our midst' experiment); but change blindness is conceptually a very different effect since it depends crucially on the occurrence of a brief transitory event in the visual field that *distracts* attention, (instead of simply depending on the fact that attention is elsewhere).

In their response to the BBS peer commentaries [27] O'Regan and Noë particularly emphasize the twin notions of 'bodiliness' and 'grabbiness'. These are, respectively, features of the worldly stimulation that refer to the way in which sensory stimulation changes as bodily actions are performed, and the [involuntary] power of some sensory stimulation(s) to grab the entity's 'attention'.

O'Regan and Noë develop these notions to (a) help distinguish between perceived [real] and conceived [imagined] experience and (b) help illustrate how the sensorimotor conception of visual perception dissolves various degrees of what in 1983 Levine [21] termed 'the explanatory gap' - *the gulf that separates physical processes in the brain from the experienced quality of sensations*.

In their subsequent BBS discussion O'Regan and Noë focus on three different aspects to this 'gap':

Intra-modal: why does an object look *like this* (e.g. spherical), rather than *like that* (e.g. cubical)?

Inter-modal: what makes one experience *visual* whereas another is *tactile*?

Absolute: why is there *any* subjective experience at all?

.. and claim that sensorimotor theory makes a contribution to understanding (and potentially closing) all three of these aspects of the explanatory gap; thus:

- sensorimotor theory helps close the intra-modal gap because the differences between, say, a spherical and a cubical object of sight correspond in part to differences in degrees in which the specific laws of [visual] sensorimotor contingency are exhibited;
- sensorimotor theory helps close the inter-modal gap because of the different types of sensorimotor dependencies at play across the different sensory modalities. For example, 'moving your eyes to the left or right will produce a change in sensory stimulation related to an object if that object is being visually perceived; but not if it is being tactually perceived, or if it is being listened to'.
- however consideration of the third, most fundamental of the explanatory gaps, reveals subtle differences in the way in which Kevin and Alva conceive the extent of their theory ..

To illustrate the latter point of departure between O'Regan and Noë we must first consider an additional 'explanatory gap' - one that sits between the inter-modal and the absolute - and which emphasises the 'perceptual aspect': what is the basis of the difference between perceptual and non-perceptual awareness of a thing? For example O'Regan and Noë suggest to compare the experience of seeing a book on a table in front of you with that of your [non-perceptual] awareness of a book on a bookshelf in the room next door; and describe how two features of the experience serve to disambiguate the two situations:

Bodiliness: movement(s) of your body affect your perception of the book in front of you *but not* your awareness of the book in the room next door.

Grabbiness: sudden changes in the visual field (say a flashing light) when looking at the book in front of you provoke an ‘automatic orientating reflex’ that cause eyes to saccade towards the light; such visual *grabbiness* is entirely absent from your experience of the book on the bookshelf.

As O’Regan and Noë summarise:

Bodiliness and grabbiness explain not only the difference between seeing something and merely thinking about it [...] but they also explain why seeing something has the sort of qualitative characteristics that it does.

In later writings [for example herein] O’Regan also considers a third feature of the real world, its ‘insubordinateness’ - how aspects of the world can change outside of the agent’s command - and contrasts the ‘insubordinateness’ of the real world view of the book in front of the agent, with the [virtual] perception of the book on the shelf in the adjacent room (where no such effects pertain)⁹.

In *Why red doesn’t sound like a bell* [29] Kevin further extends the scope of bodiliness and grabbiness by conceiving them as two dimensions of a ‘phenomenality plot’; a graph which offers insight into the age-old philosophical question of why objects are phenomenally perceived in the way they are¹⁰.

The phenomenality plot positions various experiences (e.g. hunger; an itch; driving etc.) on a two-dimensional graph, with the x-axis defining increasing bodiliness and the y-axis increasing grabbiness, such that *mental states that are located high up on the main diagonal, that is, states that objectively possess high bodiliness and high grabbiness, are precisely those that people will tend to say have a real sensory feel; that is, they have high “phenomenality”, or “something it’s like”,* (ibid).

Thus, where O’Regan [28] and Noë [26] agree [vis a vis sensorimotor theory closing the absolute explanatory gap] is that, after Nagel [24], there is “something it is like” to have a sensory experience and that sensorimotor theory can go a long way in (a) characterising what this likeness is, and (b) shedding insight on the nature of phenomenal perception itself.

⁹ In [29] Kevin also discusses a fourth aspect of sensory interactions in the real world that could explain the presence of raw feel and which helps to distinguish real experience from imagined; its ‘richness’, however this fourth aspect is absent from Kevin’s most recent accounts of sensorimotor theory.

¹⁰ “*The sensorimotor approach to raw feel explains four mysteries: why there’s something it’s like to have a raw sensory feel (presence); why different raw feels feel different; why there is structure in the differences; and why raw feel is ineffable. Of these four mysteries, the one philosophers consider to be the most mysterious is the first one, namely the question of why there’s ‘something it’s like’ to have an experience. If richness, bodiliness, insubordinateness, and grabbiness are the basis for the what it’s like of sensory feels, then we can naturally ask whether these concepts can do more work for us. In particular we would like them to explain, in the case of other, nonsensory types of experiences, the extent to which people will claim these have a ‘something it’s like’*”, O’Regan [29].

First, visual experience is visual, rather than, say, tactual or olfactory. Second, it is forcibly present to your consciousness. Third, it is ongoing, that is, the experience seems to be happening to you in a continuous way: its subjective character lasts while the experience continues. Fourth, the experience strikes us as ineffable, that is, though you experience it as possessing various qualities, the exact qualitative character escapes description in words.

We believe the sensorimotor approach allows us to explain each of these aspects of the quality of the experience. To the extent, then, that the experience itself is constituted by the presence of just these qualities, then the sensorimotor account can explain why the experience occurs at all.

Where O'Regan and Noë begin to part company is in dealing with the apparent promiscuity of the sensorimotor account; a tacit implication of which being that *any medium of interaction with the environment will produce a conscious experience in the agent* with the exact profile of the conscious experience merely contingent upon the profile of sensorimotor dependencies. As Clark and Toribio wryly observed in their response to O'Regan and Noë's magnum opus:

A good ping-pong playing robot, which uses visual input, learns about its own sensorimotor contingencies, and puts this knowledge to use in the service of simple goals (e.g., to win, but not by too many points) would meet all the constraints laid out. Yet it seems implausible to depict such a robot (and they do exist see, e.g., Andersson 1988 [2]) as enjoying even some kind of modest visual experience. Surely someone could accept all that O&N offer, but treat it simply as an account of how certain visual experiences get their contents, rather than as a dissolution of the so-called hard problem of visual qualia.

In their joint response to Clark and Toribo's examination of the internal life of a ping-pong playing robot O'Regan and Noë simply retort that:

.. as it is described, it is simply far too simple to be a plausible candidate for perceptual consciousness of the kind usually attributed to animals or humans [...] once we imagine a robot that not only masters sensorimotor contingencies, but makes use of that mastery to engage with the world in a thoughtful and adaptable way, it becomes necessary to say that it has (at least primitive) visual experiences.

However, in later writings Noë appears to retreat from this position a little; for example in *Action and Perception* [25] he highlights both that:

Nothing in our view committed us to saying that the robot would be perceptually conscious. All we committed ourselves to is the possibility that the robot could be perceptually conscious if it acquired the relevant practical skills, (ibid., chapter 7, footnote 12).

and emphasises, after Thompson [46], a fundamental link between ‘mind and life’..

.. with increasing sensorimotor complexity you get the appearance of a life-form that embodies a measure of sensitivity to the way its own movements change the way the environment stimulates it. In this way, and with a healthy dose of handwaving, we make plausible the idea that the emergence of perceptual consciousness in the biological world is, in effect, a matter of the emergence of cognitive agents with sensorimotor capabilities. There aren’t two stories: the saga of the emergence of cognitive agents, and then that of the appearance of consciousness. Consciousness and cognition are themselves aspects of the development of life”, (ibid., pp. 230)¹¹.

In contrast, in O’Regan’s later writings (and specifically in this volume), to distinguish conscious agents from purely reactive systems (e.g. ping-pong playing robots; missile guidance systems etc) Kevin simply further refines precisely what is required of a [more sophisticated] sensorimotor agent in order for it to genuinely instantiate phenomenal consciousness, by appealing to ..

.. a particular, hierarchical, form of *cognitive access* similar to that used in the higher-order theories (HOT) of consciousness [38] [10]”.

As O’Regan states (ibid):

The trick used in the sensorimotor approach is to try to ‘dissolve’ the hard problem of qualia rather than ‘solving’ it .. [that is, we need to] .. decompose consciousness into two parts: the sensorimotor interactions whose laws constitute experienced quality; and a form of cognitive access which makes that quality conscious.

Thus, for an agent to have the conscious experience of a quality, *say redness*, the agent must - via its interaction with the environment - both instantiate appropriate:

sensorimotor dependencies appropriate to pertaining to *redness*;
cognitive access to the actions being performed, such that the agent may claim, ‘I am doing this’.

Such ‘cognitive access’ introduces additional requirements for the agent to have:

1. *a notion of a self* and thus have knowledge about its own body, mind and social context;
2. *access* to the ‘experienced’ quality.

¹¹ This point is further finessed in private communication (email: 08/10/13) in which Alva further clarified, “I am inclined to think that if you start out with a spark of consciousness, the sensorimotor enactive account can explain how you end up with all the varieties of consciousness. But sms alone won’t tell you how we get started. In my view - and here I do agree with Thompson - you need life to get the ball rolling.”

That is, the agent must exhibit the right sensorimotor dependencies and also (i) be poised to make use of its phenomenal knowledge [of the experiential quality] and (ii) be ‘aware’ that it is poised to make use of this knowledge¹².

As O’Regan summarises [in this volume]:

When I say it feels like something rather than nothing to have a sensory experience, this statement can be decomposed into being conscious of facts like: it’s modal (because it has the properties of touch, not of other senses), it has a particular quality (determined by the particular sensorimotor law), it’s real (because it has the bodiliness, insubordinate-ness and grabbiness of real interactions, and furthermore controls my thoughts the way real things do), etc. The claim is that there is nothing more to phenomenal consciousness than potential conscious access to all these facts.

Thus, in an attempt to close the ‘absolute gap’ via sensorimotor theory it seems one is obliged to follow either Noë and reify a link between mind and life, or tread in O’Regan’s footsteps and insist on the need for additional explicit cognitive access to phenomenal consciousness (and in the process potentially over-complicating, and hence losing some of the elegance of, foundational sensorimotor theory).

However, even aside from its extra complexity, a key implication of O’Regan’s sensorimotor theory (but not Noë’s) is to abstract ‘consciousness’ from any causal connection with its material substrate; as long as the putative conscious agent exercises the appropriate sensorimotor contingencies and concomitant ‘cognitive access’, phenomenal consciousness ‘drops out’ for free. If correct, this conceptual finding is most helpful to those roboters who dream of building a conscious robot; and such pioneers are no straw-men..

For example, as early as 2002 Kevin Warwick (from the University of Reading, UK) stated that he believed machine consciousness was old news; his so called ‘seven dwarf robots’ already “instantiate a machine consciousness, albeit in a very weak form. They are, perhaps, if a comparison is to be drawn in terms of complexity, as conscious as a slug”, [50].

This optimistic view of machine consciousness was subsequently echoed at the very highest stratum of UK research when, in 2004, Owen Holland (from the University of Essex) led a team which won nearly half a million pounds of UK research council funding to realise the goal of instantiating machine consciousness in a humanoid robot called Cronos (through appropriate internal computational modelling of robot-self and robot-external-world)¹³.

¹² As an example consider an agent being poised to press a car brake as it approaches a red traffic light and being aware that it is poised to press the break pedal.

¹³ In their EPSRC funding application Owen’s team wrote that they expected “to enable some of the robots to be regarded as possessing a form of machine consciousness”.

And these dreams of machine consciousness are closely aligned to Kevin's own vision of future robotics¹⁴; for example in chapter 6 of 'Why red doesn't sound like a bell' he remarks:

When I was a child, my dream was to build a conscious robot. At that time in the 1960s, computers were just coming into existence, and they were where I hoped to find inspiration”;

Furthermore, in chapter 7 (*ibid*) Kevin concludes:

Where are we now with respect of my dream of building a conscious robot? So far I have been trying, one by one, to lay aside aspects of consciousness that pose no logical problem for science and that could be - and to a degree, in a few instances, are now - attainable in robots. I started with thought, perception, and communication, and went on to the self. I hope to have shown that even if we are still a long way from attaining human levels of these capacities in a robot, there is no logical problem involved in doing so.

However the formalisation and abstraction of consciousness required to build any such 'tin-can robot'¹⁵ also implies that the instantiation of consciousness is not strongly contingent on the precise physical instantiation of theory; for example, the same conscious experience must arise if systems were instantiated by a [suitably rich] virtual reality (VR) computational simulation of a particular mode of [virtually] 'embodied interactions'.

Thus if one of Kevin Warwick's real 'seven dwarf' robots is as conscious as a slug as it flits about its coral, then a VR simulation of that robot would be as conscious as a slug as it flits about its virtual coral; if Owen Holland's real-world Cronos instantiates a form of machine consciousness as it moves a pint-glass on a table, then a VR simulation of Cronos must instantiate a similar form of machine consciousness as it moves a virtual pint-glass about a virtual table; and if a real robot constructed along O'Regan's sensorimotor principles experienced the particular phenomenal sensation of the 'squidginess of a sponge' [as it squeezed a sponge] then a VR simulation of that robot enacting the same

¹⁴ A summary of Kevin's positive position vis a vis machine consciousness as espoused herein is also presented in: *Why red doesn't sound like a bell* [29]; *How to make a robot that feels*, the talk he gave at the PT-AI Conference in Thessaloniki Greece in 2011, and most recently in the paper *How to Build a Robot that is Conscious and Feels* [30].

¹⁵ A 'tin-can robot' is any robotic device where the 'physical material of its embodiment is relatively unimportant; cf. Tom Ziemke's demarcation of embodiment into strong and weak forms. In a weakly embodied cognitive system the actual material of embodiment doesn't affect the cognitive states of the system and is important only in so far as it enables the system to appropriately support various high level functions (e.g. to be functional a robot arm must at least be strong enough to move itself); in a strongly embodied system, the actual material of embodiment does matter in its giving rise to genuine cognitive states.

sensorimotor dependencies [and cognitive access] of squidding a virtual sponge must experience the same phenomenal quality of ‘squidginess’.

Thus a corollary of O’Regan’s move to close the absolute gap [by a refinement of merely abstract formal processes] implies that it - in common with the other two “VR robot systems” highlighted above - *would be phenomenally conscious purely in virtue of its execution of an appropriate computer program*; and hence that it would be vulnerable to the various critiques of computationalism [7], in so far as these critiques hold at all. E.g.

1. That computation cannot give rise to semantics (cf. Searle’s Chinese room argument [39]).
2. That computation cannot give rise to mathematical insight (cf. Penrose’s Gödelian argument [32]).
3. That computation cannot give rise to consciousness (cf. Bishop’s ‘dancing pixies’ argument [5]).

1.1 The Chinese Room Argument

Perhaps the most well known critic of computational theories of mind is John Searle. His best-known work on machine understanding, first presented in the 1980 paper ‘Minds, Brains & Programs’ [39], has become known as the Chinese room argument (CRA). The central claim of the CRA is that computations alone are not sufficient to give rise to cognitive states, and hence that computational theories of mind cannot fully explain human cognition. More formally Searle stated that the CRA was an attempt to prove the truth of the premise:

- Syntax is not sufficient for semantics;

which, together with the following two axioms ..

- Programs are formal (syntactical);
- Minds have semantics (mental content);

... led him to conclude that ‘programs are not minds’ and hence that computationalism - the idea that the essence of thinking lies in computational processes and that such processes thereby underlie and explain conscious thinking - is false [42].

It is beyond the scope of this introduction to summarise the extensive literature on the CRA¹⁶; however in *A view into the Chinese room* Bishop [6] summarised Searle’s core argument as follows:

.. Searle describes a situation whereby he is locked in a room and presented with a large batch of papers covered with Chinese writing that he does not understand. Indeed, Searle does not even recognise the symbols

¹⁶ For a broad selection of essays detailing these and other critical arguments see Preston and Bishop’s edited collection ‘*Views into the Chinese room*’ [34].

as being Chinese, as distinct from say Japanese or simply meaningless patterns. Later Searle is given a second batch of Chinese symbols, together with a set of rules (in English) that describe an effective method (algorithm) for correlating the second batch with the first, purely by their form or shape. Finally Searle is given a third batch of Chinese symbols together with another set of rules (in English) to enable him to correlate the third batch with the first two, and these rules instruct him how to return certain sets of shapes (Chinese symbols) in response to certain symbols given in the third batch.

Unknown to Searle, the people outside the room call the first batch of Chinese symbols 'the script', the second set 'the story', the third 'questions about the story' and the symbols he returns they call 'answers to the questions about the story'. The set of rules he is obeying they call 'the program'.

To complicate matters further, the people outside the room also give Searle stories in English and ask him questions about these stories in English, to which he can reply in English.

After a while Searle gets so good at following the instructions and "outsiders" get so good at supplying the rules he has to follow, that the answers he gives to the questions in Chinese symbols become indistinguishable from those a real Chinese person might give.

From an external point of view, the answers to the two sets of questions, one in English the other in Chinese, are equally good; Searle, in the Chinese room, has passed the Turing test. Yet in the Chinese language case, Searle behaves 'like a computer' and does not understand either the questions he is given or the answers he returns, whereas in the English case, *ex hypothesi*, he does. Searle contrasts the claim posed by members of the AI community - that any machine capable of following such instructions can genuinely understand the story, the questions and answers - with his own continuing inability to understand a word of Chinese..

However, the thirty plus years since its inception have witnessed many reactions to the Chinese room argument - ranging across communities in cognitive science, artificial intelligence, linguistics, anthropology, philosophy and psychology - with perhaps the most widely held criticism of Searle's position being based on what has become known as the 'Systems Reply'. This concedes that, although the person in the room doesn't understand Chinese, the entire system (of the room, the person and its contents) does.

Not surprisingly Searle finds this response entirely unsatisfactory and responds by allowing the person in the room to internalise everything (the rules, the batches of paper etc) so that there is nothing in the system not internalised within Searle. Now in response to the questions in Chinese and English there are two subsystems, a native monoglot English speaking Searle and an apparently Chinese fluent Searle, busy internalising the Chinese room. All the same he [John Searle] trenchantly continues to insist that he understands nothing of

Chinese, and a fortiori, neither does the internalised system because there isn't anything in the internalised system that is not just a part of him.

Thus, if Searle is told a joke in English, he will laugh and enjoy the experience of finding the joke funny; but if he is told a joke in Chinese, even if his 'internalised Chinese room' dictates he outputs the appropriate 'Chinese ideograph(s) for amusement', he will never experience concomitant laughter as, even with the Chinese room internalised, he cannot ever *get* the joke ..

The fiercesome power of Searle's Chinese room argument is such that, if correct, it demonstrates that no computational systems will 'ever genuinely understand' [Chinese] and hence that *all computational explanations must ultimately fail to provide an adequate model for cognition*; as Michael Wheeler presciently observes [51]:

.. the extent that the Chinese Room argument succeeds in burying the idea that computation is sufficient for mind, it does so by undermining the more general thought that any purely formal process (Searle sometimes says syntactic process) could ever constitute a mind.

Clearly then, if the Chinese room argument holds, it must also hold against any 'tin-can' robot; *even one constructed in strict accord with O'Regan's extended sensorimotor principles*.

1.2 Computations and Understanding: Gödelian Arguments against Computationalism

Gödel's first incompleteness theorem states that *"any effectively generated theory capable of expressing elementary arithmetic cannot be both consistent and complete. In particular, for any consistent, effectively generated formal theory F that proves certain basic arithmetic truths, there is an arithmetical statement that is true, but not provable in the theory."* The resulting true but unprovable statement $G(g)$ is often referred to as 'the Gödel sentence' for the theory, (albeit there are infinitely many other statements in the theory that share with the Gödel sentence the property of being true but not provable from the theory).

Arguments based on Gödel's first incompleteness theorem (initially presented by John Lucas [22] were first criticised by Paul Benacerraf [3] and subsequently extended, developed and widely popularised by Roger Penrose [31] [32] [33]) typically endeavour to show that for any such formal system F , humans can find the Gödel sentence $G(g)$ whilst the computation/machine (being itself bound by F) cannot. In [32] Penrose develops a subtler reformulation of this vanilla argument that purports to show that 'the human mathematician can "see" that the Gödel Sentence is true for consistent F even though the consistent F cannot prove $G(g)$ '.

A detailed discussion of Penrose's own take on the Gödelian argument is outside the scope of this introduction (for critical background see [11] and for Penrose's response see [33]). Nonetheless, it is important to note that Gödelian style arguments, purporting to show computations are not necessary for

cognition, have been extensively¹⁷ and vociferously critiqued in the literature (see [36] for a review); nevertheless interest in them - both positive and negative - continues to surface, (e.g. [45] [9]); and if such Gödelian style arguments do hold, then it is clear that they must also hold against any [virtual] O'Regan robot, *even one constructed in strict accord with sensorimotor principles*.

1.3 The 'Dancing with Pixies' Reductio

Many people hold the view that 'there is a crucial barrier between computer models of minds and real minds: the barrier of consciousness' and thus that computational simulations of mind and 'phenomenal (conscious) experiences' are conceptually distinct [48]. But is consciousness really a prerequisite for genuine cognition and the realisation of mental states? Certainly Searle believes so: "the study of the mind is the study of consciousness, in much the same sense that biology is the study of life" [41] and this observation leads him to postulate a 'connection principle' whereby, "... any mental state must be, at least in principle, capable of being brought to conscious awareness". Hence, if computational machines are not capable of enjoying consciousness, they are incapable of carrying genuine mental states and computational connectionist projects must ultimately fail as an adequate model for cognition.

In the following section I will briefly review a simple reductio ad absurdum argument that suggests there may be serious consequences in granting phenomenal (conscious) experience to any computational system purely in virtue of its execution of a particular program. If correct the 'dancing with pixies' (DwP) reductio [5] entails that either strong computational accounts of consciousness must fail OR that panpsychism is true.

The argument derives from ideas originally outlined by Hilary Putnam [37], Tim Maudlin [23], John Searle [40] and subsequently criticised by David Chalmers [12], Colin Klein [20] and Ron Chrisley [13] [14] amongst others¹⁸.

In what follows, instead of seeking to justify Putnam's claim that "every open system implements every Finite State Automaton" (FSA) and hence that "psychological states of the brain cannot be functional states of a computer", I will simply establish the weaker result that, over a finite time window, every open physical system implements the trace of a Finite State Automata Q on fixed, specified input (I). That this result leads to panpsychism¹⁹ is clear as, equating FSA Q(I) to a specific computational system that is claimed to instantiate phenomenal states as it executes, and following Putnam's procedure, identical computational (and, ex hypothesi, phenomenal) states can be found in every open physical system (OPS).

¹⁷ For example Lucas maintains a web page <http://users.ox.ac.uk/~jrucas/Godel/referenc.html> listing over fifty such criticisms.

¹⁸ For early discussion of these themes see 'Minds and Machines': 4(4), 'What is Computation?'

¹⁹ Panpsychism: the belief that the physical universe is composed of elements each of which is conscious.

Formally DwP is a *reductio ad absurdum* argument that endeavours to demonstrate that:

- IF ‘an appropriately programmed computer really does instantiate genuine phenomenal states’
- THEN ‘panpsychism holds’
 - *However, against the backdrop of our immense scientific knowledge of the closed physical world, and the corresponding widespread desire to explain everything ultimately in physical terms, panpsychism has come to seem an implausible view ..*
- HENCE we are led to reject the assumed claim (that an appropriately programmed computer really does instantiate genuine phenomenal states).

In his 1950 paper, ‘Computing machinery and intelligence’, Turing defined discrete state machines (DSMs) as “machines that move in sudden jumps or clicks from one quite definite state to another”, and explained that modern digital computers fall within the class of them. An example DSM from Turing is one that cycles through three computational states (Q_1, Q_2, Q_3) at discrete clock clicks. Such a device, which cycles through a linear series of state transitions ‘like clockwork’, may be implemented by a simple wheel-machine that revolves through 120° intervals.

By labelling the three discrete positions of the wheel (A, B, C) we can map computational states of the DSM (Q_1, Q_2, Q_3) to the physical positions of the wheel (A, B, C) such that, for example, ($A \rightarrow Q_1; B \rightarrow Q_2; C \rightarrow Q_3$). Clearly this mapping is observer relative: position A could map to Q_2 or Q_3 and, with other states appropriately reassigned, the machine’s function would be unchanged. In general, we can generate the behaviour of any K-state (input-less) DSM, $f(Q) \rightarrow Q'$, by a K-state wheel-machine (e.g. a digital counter), and a function that maps each ‘counter’ state C_n to each computational state Q_n as required.

In addition, Turing’s machine may be stopped by the application of a brake and whenever it enters a specific computational state a lamp will come on. Input to the machine is thus the state of the brake, ($I = ON|OFF$), and its output, (Z), the state of the lamp. Hence the operation of a DSM with input is described by a series of contingent branching state transitions’, which map from current state to next state $f(Q, I) \rightarrow Q'$ and define output (in the Moore form) $f(Q') \rightarrow Z$.

However, clamping the input to the device over a finite time interval entails that such input-sensitive contingent behaviour reverts to ‘mere clockwork’, $f(Q) \rightarrow Q'$. E.g. If Turing’s DSM starts in Q_1 and the brake is OFF for two clicks, its behaviour (execution trace) is fully described by the sequence of state transitions, ($Q_1; Q_2; Q_3$). Hence, over a finite time window, if the input to a DSM is clamped, we can map from each counter state C_n to each computational state Q_n , as required. And similarly, following Putnam, in [5] Bishop demonstrate’s how to map any computational state sequence with defined input onto the [non-repeating] internal states generated by any open physical system (e.g. a rock).

Now, returning to a putative conscious robot: at the heart of such a beast there is a computational system - typically a microprocessor; memory and peripherals. With input clamped such a system is effectively a DSM with no input. Thus, with input to the robot clamped over a finite time interval, we can map its execution trace onto the state evolution of any sufficiently large digital counter or, (ibid), any OPS. Hence, if the state evolution of a robot controlled by DSM instantiates phenomenal experience, then so must the state evolution of any OPS, and we are inexorably led to a panpsychist worldview whereby disembodied phenomenal consciousnesses (aka ‘pixies’) are dancing everywhere ..

In [7] Bishop reviews three arguments (summarised herein) that purport to show that computations are not sufficient for cognition; for example, that the execution of a computational connectionist simulation of the brain cannot instantiate genuine understanding or phenomenal consciousness (qua computation) and hence that there are limits to the use of the computational explanations in cognitive science. But perhaps this conclusion is too strong? E.g. How do the *a priori* arguments discussed herein accommodate the important results being obtained through computational neuroscience to cognition?

There are two responses to this question. The first suggests that there may be principled reasons why it may not be possible to adequately simulate all aspects of mind via a computational system; there are bounds to a [Turing machine based] computational intelligence. Amongst others this position has been espoused by: Penrose; Copeland who claims the belief that “the action of any continuous system can be approximated by a Turing Machine to any required degree of fineness ... is false²⁰”; and Smith who in [44] outlines results from ‘Chaos Theory’ which describe how ‘Shadowing Theorems’ fundamentally limit the set of chaotic functions that a Turing machine can model to those that are ‘well-behaved’; i.e. functions that are not well-behaved cannot, in principle, be accurately described by Turing machine simulation.

However, Gödelian style arguments, purporting to show computations are not necessary for cognition, have been *extensively* criticised in the literature; and are currently endorsed by only a few, albeit in some cases very eminent, scholars. Nonetheless some, for example Hava Siegelmann [43], are confident that even if Gödelian arguments are valid, super Turing computation, in the form of Artificial Recurrent Neural Networks (ARNNs) offer a potential reconciliation between at least one form of [neural] computation and mind:

Our model may also be thought of as a possible answer to Penrose’s recent claim [31] that the standard model of computing is not appropriate for modelling true biological intelligence. Penrose argues that physical

²⁰ Copeland’s argument is detailed, but at heart he follows an extremely simple line of reasoning: consider an idealised analogue computer that can add two reals (a, b) and output one if they are the same, zero otherwise. Clearly either (a) or (b) could be non-computable numbers (in the specific formal sense of non Turing-computable numbers). Hence, clearly there exists no Turing machine that, for any finite precision (k), can decide the general function $F(a = b)$, (see [15] for detailed discussion of the implications of this result).

processes, evolving at a quantum level, may result in computations which cannot be incorporated in Church's Thesis. The analog neural network does allow for non-Turing power while keeping track of computational constraints, and thus embeds a possible answer to Penrose's challenge within the framework of classical computer science.

If Siegelmann is correct, a [O'Regan] 'sensorimotor-robot' (controlled by a suitably configured super-Turing ARNN) would be invulnerable to Penrose's Gödelian style arguments and hence, in the right context, would be capable of 'mathematical insight'.

A second response emerges from the 'Chinese room argument' and the 'dancing with pixies' *reductio*. It acknowledges the huge value that the computational metaphor plays in current cognitive science and concedes, for example, that a future computational neuroscience may be able to simulate *any* aspect of neuronal processing and offer insights into all the workings of the brain. However, although such a computational neuroscience will result in deep understanding of neuronal cognitive processes, it insists on a fundamental ontological division between the *simulation of a thing* and *the thing itself*.

For instance we may simulate the properties of gold using a computer program but such a program does not automatically confer upon us riches (unless of course the simulation becomes duplication; an identity). Hence Searle's famous observation that "*No one supposes that computer simulations of a five-alarm fire will burn the neighbourhood down or that a computer simulation of a rainstorm will leave us all drenched. Why on earth would anyone suppose that a computer simulation of understanding actually understood anything?*" [39]

Both of the above responses accommodate results from computational cognitive science, but the second specifically highlights continued shortcomings of any **purely formal** - vis a vis computational - account of cognition, such as, we suggest, Kevin O'Regan's conception of sensorimotor theory (contra Alva Noë's) is committed to. If this analysis is correct, then it is perhaps time for contemporary sensorimotor theorists to embrace the 'strong-embodiment of brain and body' (and concomitant physical and social context) a little more deeply²¹ ..

2 A Brief Resume of 'Contemporary Sensorimotor Theory'

We believe this survey of contemporary sensorimotor theory offers an interesting selection of current research informed by the sensorimotor account of perception and we will complete this introduction with a brief summary of the contributed works. In this conception we are privileged to open our volume with an introduction to contemporary sensorimotor theory from J. Kevin O'Regan.

In his opening chapter Kevin offers a new overview of extant sensorimotor theory with particular regard to the hard problem of 'phenomenological consciousness'. In so doing, he demonstrates the power of a method - fundamentally

²¹ C.f. Gibson [18], Varela et al [49], Bickhard & Terveen [4], Thompson [46], Deacon [16] etc.

grounded in observable physical phenomena - of describing and potentially explaining features of consciousness that have often been regarded as falling outside of the scope of genuine scientific explanation.

After this resume of contemporary sensorimotor theory, the field is critically scrutinised from two distinct perspectives: (a) ‘the body, enactivism and emotion’ and (b) ‘action as constitutive of perception’.

Firstly Paine describes (a) how sensorimotor theory can be shown to have ‘Heideggerian roots’ (such as its prioritisation of environmental engagement) and (b) the degrees to which it satisfies a range of Heideggerian conditions for the presence of conscious experience; in the process Paine explores key differences between the two approaches and offers suggestions as to how they might be resolved. Paine concludes that although O’Regan’s 2011 sensorimotor account [29] meets Dreyfus’ requirements for a ‘Heideggerian AI’ (and in the process potentially satisfies his objections to ‘good old fashioned - classical - Artificial Intelligence’), it ‘omits room for emotion’; an area where, she suggests, a Heideggerian perspective would offer further insight.

Subsequently, and in a very carefully argued piece, Ken Pepper draws on the philosophy of Maurice Merleau-Ponty to sketch a ‘phenomenological interpretation’ of sensorimotor understanding, appealing to the ‘Phenomenology of Perception’ to show (a) how two of its major operative concepts - the ‘body schema’ and ‘sedimentation’ - can help plug the gaps in Noë’s sensorimotor account and (b) how the notion of ‘body schema’ conditions the way things appear to the agent. In so doing Pepper presents a detailed analysis of the differences between ‘affordances’ and ‘object horizons’, and clarifies Noë’s position with respect to the division.

Then Scarinzi looks at how enactive - with respect to the role of mental representations and embodiment - the sensorimotor approach really is. Scarinzi argues that it is actually ‘semi-enactive’ and that to bring it closer to enactivism proper (in investigation of the agent’s subjective experience of the qualities of interactions) it needs to be more sensitive to (i) the motor and cognitive-emotional role of the lived body and (ii) the agent’s subjective access to it. Thus, although sensorimotor theory and enactivism proper share some important features (such as a fundamental entanglement with the environment) sensorimotor theory merely skirts other experiential features (such as the role of emotional involvement); in this way Scarinzi suggests that in the sensorimotor account *the subjectivity of lived experience* is neglected.

Secondly Rainey (and subsequently Loughlin) examine the so called ‘causal - constitutive objection’:

Noë argues that visual states are not pictorial; he argues that all perception is conceptual; and he argues that the external world makes a constitutive contribution to experience. I am unpersuaded by these arguments .. (Jesse Prinz [35]).

Even if perceptual experience depends causally or counterfactually on movement or another form of activity, it does not follow that perceptual experience constitutively involves movement .. (Ned Block [8]).

Thus Rainey draws on various philosophical traditions to consider the controversial and potentially problematic claim - in the context of sensorimotor theory - that experience is conceptual; i.e. that experience is a complex, conceptually articulated and conceptually very laden affair. If Rainey is correct, and sensorimotor theory is indeed committed to a conceptual view of experience, then he argues, it opens a Pandora's box for critics like Block and Prinz to argue the *causal-constitutive objection*.

Loughlin suggests that sensorimotor theory is 'centrally committed to the claim that visual experience is realized by embodied know-how or skilful engagement' which, Loughlin claims, opens it up to attacks from Aizawa, Block, Clark, Prinz et al. That said, Loughlin concludes that sensorimotor theory can easily accommodate most of their objections, but that the most serious - the causal-constitutive objection - can only be avoided by 'going radical', i.e. shifting to a conceptual stance more closely aligned with Hutto and Myin's 'radical enactivism' [19].

Further scrutiny of the sensorimotor account takes the form of identifying and describing problems in the theory as it stands, and identifying related approaches by which these challenges could potentially be met. In this vein, Wadham identifies an issue in Noë's development of sensorimotor theory, claiming his account of perceptual content as being 'virtual all the way in' is incompatible with his account of perspectival content via 'p-properties'. Wadham suggests that the issue arises from the 'problem of invisible contents'; i.e. Noë's virtual content thesis implies that p-properties must be invisible²² whereas the key role these properties play in Noë's overarching theory of perception requires that they cannot be 'invisible' in this way²³. Wadham concludes by offering a potential resolution to this contradiction, suggesting an alternative to Noë's p-property story: 'appearance-pattern theory'.

Subsequently Lyon suggests that classical sensorimotor accounts of perception, based mainly on vision and touch, are inadequate for other sensory modalities - specifically audition. Lyon examines the effect of including sound in accounts of perception, and suggests that it makes sense to avoid the 'unnecessary strait jacket' of a model based primarily on vision and touch alone. Lyon concludes by suggesting that the sensorimotor approach can be usefully extended to other perceptual modes.

Parthemore considers how sensorimotor theory aligns with and differs from the unified conceptual space theory of concepts (as derived from Gardenfors' canonical conceptual space theory' [17]), showing how the meta-theory resulting from a combination of both introduces a new role for emotional affect and

²² "... while in motion, we cannot see individual p-properties .. if we can't take in any content in an instant, then we can't see a property of the object that is presented to us only for an instant"; hence p-properties are invisible", (Wadham, this volume).

²³ "... as we move, we experience a variety of p-properties presented by the object we are looking at. It is through implicitly understanding the way in which the p-properties we see vary as a function of our movement that we come to see the actual, perspective-independent properties of objects", (Wadham, this volume).

the somatosensory system. These additional aspects are suggested as offering a potential avenue for grounding both ‘salience’ and the ‘proper understanding of mental representation’.

In an echo of Tononi [47] and Aleksander & Morton [1], Gamez examines sensorimotor theory through an information-theoretic lens that views the character of sensation as being constituted by the way in which ‘information is processed and structured’. In this light he offers critical perspectives on three extant issues sometimes viewed as problems for the sensorimotor account: (a) the view that sensation may exist without motor action; (b) conscious sensations can be evoked by directly stimulating the brain, and (c) the claim that visual sensory substitution fails to produce *genuine* visual sensations. Then, in controversial opposition to central tenets of sensorimotor theory, Gamez endorses a weaker form of theory that merely posits a *correlation* between sensory or sensorimotor dependencies and the contents of consciousness²⁴ - effectively “bridging” between classical (correlational) accounts of consciousness and strict sensorimotor accounts - such that (a) conscious sensations are merely correlated with sensory dependencies, and (b) conscious perception is merely correlated with mastery of sensorimotor dependencies.

In the last section of the book we examine how contemporary sensorimotor theory has been successfully applied in other domains. Thus, the key claim of Rucinska’s paper is to develop a theory of basic play grounded in ‘action’ rather than ‘representations’. In so doing Rucinska makes a convincing case that sensorimotor affordances can be seen as a sufficient basis for imaginative play in young children; specifically, Rucinska considers how a sensorimotor account of perception can be used to form a ‘theory of basic pretence’ considering, for example, the possibility that a child may pretend that a banana is a phone through direct on-line capacities, rather than requiring off-line conceptual workings, and furthermore without requiring internal representations. As O’Regan and Noë described ‘perceiving as a way of acting’, so Rucinska suggests ‘pretending as a way of acting’. In shifting focus to applying sensorimotor theory as it stands to a new area of investigation, its usefulness is demonstrated as a force both directing research and informing explanation of results.

Gibbs and Devlin present a survey of sensory substitution and sensory augmentation devices, and enactive interfaces in real, as well as virtual, worlds. They show how a sensorimotor interpretation of existing work on sensory augmentation can be applied to make predictions about the experience of reproducing the experiments in virtual worlds. These predictions, being testable, provide a basis for research into both perception and the explanatory power of the sensorimotor account itself.

Similarly Gillies & Kleinsmith investigate how the sensorimotor account of perception can inform user-interface design, suggesting that learned interactions

²⁴ “While some authors have suggested that there is an identity between a mastery of sensorimotor contingencies and consciousness .. I am only focusing here on the weaker and less contentious claim that there might be a correlation between sensory or sensorimotor contingencies and the contents of consciousness”, Gamez (herein).

can be exploited to allow someone to design with performed actions, rather than representations thereof. Gillies & Kleinsmith illustrate this conception via an innovative application used to design interactions with video game characters.

Hoffmann discusses a number of case studies of robots that detect and exploit sensorimotor regularities in their environment to exhibit more complex behaviour than has previously been achievable. He subsequently broadens his perspective to draw out more general aspects of the relation between body schema, forward models and sensorimotor contingency theory, concluding by highlighting potential limitations in the field of cognitive robotics.

Finally Cowley asks if current conceptual shortcomings in dealing with language result primarily from a fixation with the written word, and recommends a more enactive approach to the subject, rather than, for example, a continuing focus on Chomsky's mentalism. Nonetheless, Cowley warns against a simple reduction to 'sensorimotor dependencies' as explanatory of linguistic behaviour: "*.. in embodied cognition, there is a risk of overplaying work on how bodies (and brains) regulate activity/ system-states. In O'Regan and Noë [27], for example, perceptual modalities are said to exist 'only in the context of the interacting organism'*". A case study - emphasising the dual role of action and language in perception - is examined in detail; it clearly illustrates that such a simple reduction does not apply in human languaging²⁵.

As editors we think this collection of new essays offers a fascinating snapshot of contemporary sensorimotor theory in 2014; early leaves from O'Regan and Noë's 'enactivist approach' have blossomed ..

Acknowledgements. JMB would like to thank his wife Dr. Katerina Koutsantoni for her enduring patience in preparing this volume, his co-editor, Andrew Martin, who has worked miracles in helping to bring this volume about and the reviewers whose careful comments helped both polish the content of the volume and clarify several important issues. He would also like to specifically thank Jan Degenaar, Kevin O'Regan and Alva Noë for their very helpful comments in preparing this introduction.

AOM would like to thank his parents and Norah who contributed much of their time to support the work that went into this volume and the unerring guidance of his co-editor, supervisor and friend 'Bish', who conceived of, and initiated, the project.

Many of the papers herein were originally presented at the "1st AISB Workshop on Sensorimotor Theory" held at Goldsmiths, University of London, on September 26th, 2012, and thus thanks must also be extended to the AISB²⁶,

²⁵ "*.. while the young mans thinking is based in sensorimotor activity, his actions reach into a social domain that lies beyond body-world interaction, Cowley (herein).*"

²⁶ The AISB is the UK society for the study of Artificial Intelligence and the Simulation of Behaviour. It is the longest established such society in the world and celebrates its 50th anniversary at a Convention at Goldsmiths in April 2014.

EUCOG²⁷ and Goldsmiths, University of London for their generous support of that event. In preparing this introduction we have extracted the description of CRA from Bishop [6] and DwP from Bishop [7].

References

1. Aleksander, I., Morton, H.: *Aristotle's laptop: the discovery of our informational mind*. World Scientific Publishing, Singapore (2012)
2. Andersson, R.L.: *Robot ping-pong player*. The MIT Press, Cambridge (1988)
3. Benacerraf, P.: *God, the Devil & Gödel*, *Monist* 51 (1967)
4. Bickhard, M.H., Terveen, L.: *Foundational Issues in Artificial Intelligence and Cognitive Science*. North Holland (1995)
5. Bishop, J.M.: *Dancing with Pixies: strong artificial intelligence and panpsychism*. In: Preston, J., Bishop, J.M. (eds.) *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford University Press, Oxford (2002)
6. Bishop, J.M.: *A view inside the Chinese room*. *Philosopher* 28(4), 47–51 (2004)
7. Bishop, J.M.: *A Cognitive Computation fallacy? Cognition, computations and panpsychism*. *Cognitive Computation* 1(3), 221–233 (2009)
8. Block, N.: *Review of Alva Noë, Action in Perception*. *Journal of Philosophy* 102, 259–272 (2005)
9. Bringsjord, S., Xiao, H.: *A refutation of Penrose' Gödelian case against artificial intelligence*. *J. Exp. Theo. Art. Int.* 12, 307–329 (2000)
10. Carruthers, P.: *Higher-Order Theories of Consciousness*. In: Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* (2011)
11. Chalmers, D.: *Minds, Machines, And Mathematics A Review of Shadows of the Mind by Roger Penrose*. *PSYCHE* 2(9) (1995)
12. Chalmers, D.J.: *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, Oxford (1996)
13. Chrisley, R.: *Why Everything Doesn't Realize Every Computation*. *Minds and Machines* 4, 403–420 (1995)
14. Chrisley, R.: *Counterfactual computational vehicles of consciousness*. In: *Toward a Science of Consciousness*, Tucson Arizona, USA, April 4-8 (2006)
15. Copeland, B.J.: *The broad conception of computation*. *American Behavioral Scientist* 40(6), 690–716 (1997)
16. Deacon, T.: *Incomplete Nature: How Mind Emerged from Matter*. W. W. Norton & Company (2012)
17. Gardenfors, P.: *Conceptual Spaces as a Framework for Knowledge Representation*. *Mind and Matter* 2(2), 9–27 (2004)
18. Gibson, J.J.: *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston (1979)
19. Hutto, D., Myin, D.: *Radicalizing Enactivism: basic minds without content*. The MIT Press, Cambridge (2012)

²⁷ EUCog is a European network of nearly 900 researchers in artificial cognitive systems and related areas who want to connect to other researchers, reflect on the challenges of the discipline and get their research 'out there'. The network funds meetings, workshops, members' participation in academic events, faculty exchanges and other activities that further its aims. It is funded by the Information and Communication Technologies division of the European Commission, Cognitive Systems and Robotics unit, under the 7th Research Framework Programme.

20. Klein, C.: Maudlin on Computation, working paper (2004)
21. Levine, J.: Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly* 64, 354–361 (1983)
22. Lucas, J.R.: Minds, Machines & Gödel. *Philosophy* 36 (1961)
23. Maudlin, T.: Computation and Consciousness. *Journal of Philosophy* 86, 407–432 (1989)
24. Nagel, T.: What Is It Like to Be a Bat? *The Philosophical Review* 83(4), 435–450 (1974)
25. Noë, A.: *Action in Perception*. The MIT Press, Cambridge (2004)
26. Noë, A.: *Out of Our Heads: why you are not your brain, and other lessons from the biology of consciousness*. Hill and Wang, NYC (2009)
27. O'Regan, K., Noë, A.: A sensorimotor account of vision and visual consciousness. *Behaviour and Brain Sciences* 24, 939–1031 (2001)
28. O'Regan, J.K.: Sensorimotor approach to (phenomenal) consciousness. In: Baynes, T., Cleeremans, A., Wilken, P. (eds.) *Oxford Companion to Consciousness*, pp. 588–593. Oxford University Press, Oxford (2009)
29. O'Regan, J.K.: *Why red doesn't sound like a bell: understanding the feel of consciousness*. Oxford University Press, Oxford (2011)
30. O'Regan, J.K.: How to Build a Robot that is Conscious and Feels. *Minds and Machines* 22(2), 117–136 (2012)
31. Penrose, R.: *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford University Press, Oxford (1989)
32. Penrose, R.: *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford University Press, Oxford
33. Penrose, R.: Beyond the Doubting of a Shadow A Reply to Commentaries on *Shadows of the Mind*. *PSYCHE* 2(23) (1996)
34. Preston, J., Bishop, J.M. (eds.): *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford University Press, Oxford (2002)
35. Prinz, J.: Putting the Brakes on Enactive Perception. *PSYCHE* 12(1) (2006)
36. *Psyche: Symposium on Roger Penrose's Shadows of the Mind*. *PSYCHE* 2 (1995)
37. Putnam, H.: *Representation & Reality*. Bradford Books, Cambridge (1988)
38. Rosenthal, D.M.: Varieties of higher-order theory. In: Gennaro, R.J. (ed.) *Higher-Order Theories of Consciousness: An Anthology*. Johns Benjamins Publishing Company, Amsterdam (2004)
39. Searle, J.R.: Minds, brains, and programs. *Behavioral and Brain Sciences* 3(3), 417–457 (1980)
40. Searle, J.: Is the Brain a Digital Computer? *Proceedings of the American Philosophical Association* 64, 21–37 (1990)
41. Searle, J.: *The Rediscovery of the Mind*, p. 227. MIT Press, Cambridge (1992)
42. Searle, J.: *The Mystery of Consciousness*. Granta Books, London (1994)
43. Siegelmann, H.T.: Neural and Super-Turing Computing. *Minds and Machines* 13, 103–114 (1993)
44. Smith, P.: *Explaining Chaos*. Cambridge University Press, Cambridge (1998)
45. Tassinari, R.P., D'Ottaviano, I.M.L.: Cogito ergo sum non machina! About Gödel's first incompleteness theorem and Turing machines. *CLE e-Prints* 7(3) (2007)
46. Thompson, E.: *Mind in Life: biology, phenomenology, and the sciences of mind*. Harvard University Press, Cambridge (2010)
47. Tononi, G.: Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* 215, 216–242 (2008)

48. Torrance, S.: Thin Phenomenality and Machine Consciousness. In: Chrisley, R., Clowes, R., Torrance, S. (eds.) Proc. 2005 Symposium on Next Generation Approaches to Machine Consciousness: Imagination, Development, Intersubjectivity and Embodiment, AISB05 Convention, pp. 59–66. University of Hertfordshire, Hertfordshire (2005)
49. Varela, F.J., Thompson, E., Roesch, E.: *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, Cambridge (1991)
50. Warwick, K.: Alien Encounters. In: Preston, J., Bishop, J.M. (eds.) *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford University Press, Oxford (2002)
51. Wheeler, M.: Change in the Rules: Computers, Dynamical Systems, and Searle. In: Preston, J., Bishop, J.M. (eds.) *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford University Press, Oxford (2002)