International Journal of Machine Consciousness
Vol. 1, No. 2 (2009) 213–223
© World Scientific Publishing Company

THE POTENTIAL FOR CONSCIOUSNESS OF ARTIFICIAL SYSTEMS

DAVID GAMEZ

Department of Electrical Engineering, Imperial College,
London, SW7 2BT, UK
david@davidgamez.eu

The question about the potential for consciousness of artificial systems has often been addressed using thought experiments, which are often problematic in the philosophy of mind. A more promising approach is to use real experiments to gather data about the correlates of consciousness in humans, and develop this data into theories that make predictions about human and artificial consciousness. A key issue with an experimental approach is that consciousness can only be measured using behavior, which places fundamental limits on our ability to identify the correlates of consciousness. This paper formalizes these limits as a distinction between type I and type II potential correlates of consciousness (PCCs). Since it is not possible to decide empirically whether type I PCCs are necessary for consciousness, it is indeterminable whether a machine that lacks neurons or hemoglobin, for example, is potentially conscious. A number of responses have been put forward to this problem, including suspension of judgment, liberal and conservative attribution of the potential for consciousness and a psychometric scale that models our judgment about the relationship between type I PCCs and consciousness.

Keywords: Machine consciousness; science; thought experiment; Chinese Room; correlates of consciousness; type I, type II; synthetic phenomenology.

1. Introduction

One of the key questions in machine consciousness is whether an artificial system, such as a robot controlled by a computer, is capable of phenomenal states. If artificial systems cannot become conscious, then weak artificial consciousness will still be possible [Seth, 2009], but there will be little point in attempting to describe the phenomenology of an artificial system [Chrisley, 2009]. This question about a system's *potential* for consciousness is distinct from the question about whether a system is *actually* conscious at a particular point in time. Living humans are potentially conscious all of the time, but they are only actually conscious for roughly 18 hours a day.

A number of people have addressed this question using thought experiments, such as Searle's Chinese Room [Searle, 1980] or Block's Population of China [Block, 1978], which attempt to demonstrate that certain types of artificial system are incapable of conscious states. However, thought experiments are very poor instruments for studying the potential for consciousness because they have little to say about nomological possibility and because the relationship between the physical and the phenomenal cannot be represented in thought. These limitations of thought experiments are covered in Sec. 2.

A much more promising way of deciding whether machines are potentially conscious is to look for the physical conditions that are correlated with consciousness in humans and determine whether these are present in artificial systems. If the human correlates of consciousness are present in a machine, then there are good grounds for believing that the machine is capable of conscious states. These correlates could be material, neural or cognitive/functional, and an outline of their scientific study is given in Sec. 3.

One difficulty with the scientific search for the correlates of consciousness is that current technology does not have enough temporal or spatial resolution to support detailed experiments on the material or neural correlates of consciousness. Whilst some of these technological limitations may eventually be overcome, there is a more fundamental problem that many experiments on the correlates of consciousness cannot even in principle be carried out. Since consciousness is measured using behavior, it will be impossible to get useful results from experiments that do not affect a system's behavior or interfere with its ability to report phenomenal states. Section 3.3 summarizes these difficulties, which are formalized into a distinction between two types of potential correlates of consciousness (PCCs) in Sec. 4. Type I PCCs cannot be experimentally separated out, and so their actual correlation with consciousness will remain unknown, and it will be impossible to develop a theory of consciousness that includes type I PCCs. This does not affect our predictions about consciousness in natural systems that share the human type I PCCs, but it severely impairs our ability to assess the potential for consciousness in artificial systems, which lack many or all of the type I PCCs.

Section 5 discusses a number of responses to this problem, including suspending judgment, assuming that type I PCCs are essential or inessential for consciousness, and a scale that models our propensity to attribute consciousness to a system based on its type I PCCs. A preliminary draft of this psychometric scale has been set out elsewhere [Gamez, 2008a], and an implementation is available online [Gamez, 2008b].

2. The Limitations of Thought Experiments in Consciousness Research

Some people believe that the question about machines' potential for consciousness can be answered by carrying out thought experiments. A good example is given by [Block, 1978], who attempts to decide whether the brain's functions are associated with consciousness by imagining that these functions are implemented by the population of China communicating with radios and satellites. Since Block cannot imagine that the population of China communicating with radios and satellites could

November 5, 2009 5:02:09pm WSPC/258-IJMC 00019 FA1

be associated with conscious states, he concludes that a functionalist approach to consciousness is incorrect — with the implication that the implementation of the brain's functions in a machine would not be enough to make it conscious. A second influential thought experiment is given by [Searle, 1980], who describes a room in which a person processes Chinese characters according to rules and outputs a response. The person in the room does not understand Chinese, and so they do not have intentional states directed towards the objects represented by the characters, and they are not conscious of the objects represented by the characters. Searle concludes from this thought experiment that the processing of characters according to rules cannot lead to understanding (and consciousness) at the system level. Again, the implication is that a machine processing characters according to rules is not capable of understanding or becoming conscious.

Scientific thought experiments typically highlight inconsistencies in existing theories and identify anomalies in experimental data [Kuhn, 1977], whereas philosophical thought experiments are more commonly used to clarify concepts and establish whether something is possible (or impossible) — following Hume's problematic claim that "*nothing we imagine is absolutely impossible*" ([Hume, 1985], p. 81). To understand this better, it is useful to distinguish three different senses of possibility [Gendler and Hawthorne, 2002]. A statement P is *logically* possible if no contradiction can be deduced from it. So, for example, if a square is defined as a figure with four corners and a circle is defined as a figure with no corners, then the claim that there is a round square is logically impossible because a figure cannot both have and not have four corners. *Metaphysical* possibility refers to the different ways in which the universe might have been created by God, which is often expressed using the terminology of "possible worlds". So, for example, there is a possible world in which unicorns exist and possible worlds in which the laws of nature are completely different from the ones in our world. Finally, a statement is *nomologically* possible if it is consistent with a nomos, or set of laws, such as the laws of nature in our current universe. So the claim that polystyrene floats on water is nomologically possible in our universe, whereas the statement that iron floats on water is not.

In discussions about machine consciousness the main type of possibility at issue is whether machines can be conscious in our current world according to the current laws of nature — in other words, whether it is nomologically possible that artificial systems can be associated with phenomenal states. A first problem with using thought experiments to address nomological questions connected with consciousness is Wilke's argument that successful scientific thought experiments are based on natural kinds, which: "provide, in the main, the central *explananda* and *explanatia* for systematic study: they are the terms for which, and with which, the laws and generalizations of science are framed. Hence 'water', 'mass' and 'tiger' are natural-kind terms; 'fence', 'ashtray', and 'ornament' are not". ([Wilkes, 1988], p. 14). Effective thought experiments can be carried out using natural kinds because we have a good intuitive understanding of how they work, whereas thought experiments about consciousness are extremely problematic because "consciousness" is an extremely badly defined term and subject to a great deal of controversy. It is reasonably easy to use thought experiments to decide what would happen to balls on frictionless planes; it is almost impossible to generate a meaningful thought experiment about something as vague as consciousness, particularly when questions are asked about non-standard cases, such as the potential for consciousness of artificial systems.

A second major limitation of philosophical thought experiments about consciousness is that it is not possible to imagine the physical world. In a typical thought experiment about artificial consciousness, we imagine a robot and then try to imagine whether a phenomenal world (perhaps containing the color red) could be associated with the physical robot. The problem with this type of thought experiment is that the physical robot cannot be imagined by us because it is completely non-phenomenal in character — a problem that often occurs when people are trying to think about the relationship between the brain and consciousness ([Metzinger, 2002], p. 1). Physical reality is by definition non-phenomenal, and so it can only enter our consciousness (and our thought experiments) through a phenomenal representation. We can consciously think about the relationship between a phenomenal representation of the physical world (a phenomenal robot) and another phenomenal representation (the color red), but thought experiments are completely incapable of formulating the real problem of the relationship between a physical robot and phenomenal red.[a]

These and other problems with thought experiments in the philosophy of mind [Wilkes, 1988] suggest that the scientific study of human consciousness would be a much better way of understanding the potential for consciousness of artificial systems.

3. The Science of Consciousness

3.1. *Methodology*

One of the main aims of a science of consciousness is the development of a theory of consciousness that can make falsifiable predictions [Popper, 2002] about the presence of consciousness in humans. As the falsifiable predictions are confirmed by experiments our confidence in the theory will increase, and we will eventually start to use the theory to make predictions about the consciousness of artificial systems.

The starting point for any theory of consciousness is a detailed understanding of the relationship between the phenomenal world of our experiences and the physical world described by science. To identify the correlates of consciousness, we need to take a physical system that is known or assumed to be conscious, such as a human being or monkey, and systematically vary its parameters (individually and in different combinations) to identify the aspects that are correlated with conscious states. For example, if the human brain has attributes W, X, Y and Z, and removing Z and W has no effect on the consciousness of the system, but removing either X or Y individually or in combination leaves the system unconscious, then we can conclude

[a] See [Gamez, 2008a], Sec. 2.4.5 and [Gamez, 2007], pp. 71–83 for more detailed discussion of this point.

that X and Y are necessary for consciousness. However, we can only conclude that X and Y are sufficient for consciousness if the human brain has no other attributes in addition to W, X, Y and Z that might be correlated with consciousness. For example, if the attribute C was left unchanged during the experiments, then it is possible that X + Y is not sufficient for consciousness and C has to be included as well. As each of the system's parameters is varied, the associated consciousness needs to be measured using verbal report or another form of behavior.

Once we have identified the correlates of consciousness in humans, it should be possible to develop a general theory of consciousness, which can be used to make predictions about the phenomenal states associated with artificial systems. For example, if information integration [Balduzzi and Tononi, 2008] turns out to be a good predictor of phenomenal states in humans, and no other aspect of the human brain is systematically correlated with consciousness, then we could measure the information integration of an artificial system and use it to make predictions about its consciousness.

3.2. *Potential correlates of consciousness (PCCs)*

The human brain is the paradigmatic example of a system that is known to be conscious and the most popular theory is that consciousness is correlated with its neurons properties. However, it is also possible that the brain's material properties or higher level functions are relevant to consciousness — an issue that is especially important when questions are asked about consciousness in artificial systems. The material properties of the brain are those that are described by physics and chemistry, for example:

(i) Volume of 1.4 liters.
(ii) Temperature of 310 K.
(iii) Weight of 1350 g.
(iv) Reflects light with a wavelength of 650 nm.
(v) Contains biological amino acids and hemoglobin.

At the neural level it has been shown that activity in biological neurons is strongly correlated with consciousness and a large number of experiments have been carried out that have attempted to distinguish between neural activity that takes place when we are not conscious — in deep sleep or a coma, for example — and neural activity that is correlated with conscious experience. The emerging consensus is that the neural correlates of consciousness are likely to be distributed over many different brain areas — see [Edelman and Tononi, 2000; Dehaene and Naccache, 2001; Zeki, 2003] — and the coordination between these areas might be achieved by synchronization of neural firing [Singer, 2000], NMDA synapses [Flohr, 2000], connections to thalamic nuclei [Newman *et al.*, 1997] or some combination of these mechanisms. Further discussion of the neural correlates of consciousness can be found in [Metzinger, 2002; Chalmers, 1998; Noë and Thompson, 2004].

The human brain can also be analyzed from the perspective of the large number of functions that it carries out, many of which are potential correlates of consciousness. These range from the low level input and output functions of ion channels and neurons, up to higher level functions, such as perception, memory and cross-modal integration. The brain also carries out a number of cognitive functions that have been linked to consciousness, such as emotional evaluation of a situation, representation of the self, imagination and attention.

3.3. *Limitations of experiments on the correlates of consciousness*

A first problem with experiments on the correlates of consciousness is that the accuracy with which we can measure a system's material properties is severely limited. Whilst we have a range of techniques for measuring weight, volume and chemical composition, these break down almost completely at the molecular or subatomic scale. Measurement limitations are also present at the neural level since current non-invasive scanning technologies, such as EEG, fMRI, and PET, have low spatial and/or temporal resolution, and implanted electrodes can only access a few hundred neurons at a time. Our current techniques for measuring the functional and cognitive correlates of consciousness include psychological tests, examining patients with brain damage, and applying transcranial magnetic stimulation to selectively disable parts of the brain. However, these techniques are pretty crude and often interfere with the subject's report of phenomenal states. Many of these measurement limitations may be overcome by technological advances, but it seems unlikely that we are ever going to get enough temporal and spatial resolution to decide whether the brain's subatomic properties, for example, are necessary for consciousness.

A more fundamental limitation of experiments on the correlates of consciousness is that the measurement of conscious states depends on language and possibly memory. If an experiment alters a subject's ability to remember and/or report, then the resulting description (or lack of description) of the phenomenology could be the result of a changed phenomenology or it could be the result of a damaged reporting faculty, and it may be difficult or impossible to decide which is actually the case. For example, a number of theories link consciousness to integration between different parts of the brain, but our ability to report conscious states also presumably depends on integration between brain areas processing conscious information and brain areas responsible for language. An experiment that reduced the brain's information integration might also affect the subject's ability to report their conscious states. In this situation, an alteration in the subject's description of their phenomenology might be due to reduced consciousness, to a reduced ability to report consciousness, or to both, and it could be very difficult to establish which is actually the case.

An equally serious problem is that many PCCs can be changed without affecting the subject's behavior, which makes them impossible to separate out empirically. For example, consider an experiment on the correlation between hemoglobin and consciousness. One way of investigating this would be to replace the blood in a person's brain with artificial blood and measure their consciousness to see if there are any changes. The problem with this experiment is that the neurons will be oxygenated in exactly the same way as before, and so the subject will behave in the same way and report the same phenomenal states. The experiment might have removed the consciousness and preserved the behavior, or it might have had no effect on consciousness at all. An equally ambiguous outcome would occur if we attempted to test whether neurons are correlated with consciousness by replacing part of the brain with a functionally equivalent chip. The global behavior of the brain would be the same before and after the experiment, and so it would be impossible to tell whether the consciousness had been affected.[b]

The last two limitations arise from the fact that consciousness is measured through external behavior. These are hard limits that are set by the nature of the phenomenon — they are not something that might be overcome by better technology. Since behavior is the only way of measuring consciousness in a system, it will always be impossible to carry out behavior-neutral PCC experiments or experiments that interfere with the reporting of conscious states. In the next section this contrast between PCCs that can be experimentally tested and those that cannot is expressed as a distinction between two types of potential correlates of consciousness.

4. Type I and Type II Potential Correlates of Consciousness

4.1. *Type I PCCs*

Type I PCCs are properties of the human brain (and any other paradigmatically conscious system) whose correlation with consciousness cannot be experimentally tested, either because the PCCs are potentially behavior-neutral or because they are strongly linked to the ability to report conscious states. A potential correlate of consciousness is type I if it is impossible to devise an experiment (or series of experiments) that could identify its actual correlation with consciousness.

The material aspects of the human brain are type I PCCs because it is impossible to devise experiments that test their link with consciousness independently of other factors. Somewhat counter-intuitively, many neural aspects of the brain are also type I PCCs because neurons are just one way in which a particular set of functions can be implemented. For example, neural synchronization has been put forward as a PCC [Crick, 1994], but if the same behavioral output could be produced with non-synchronized neurons, then it would be impossible to tell whether synchronization is important or not.

The second class of type I PCCs is linked to our ability to remember and/or report phenomenal experiences. The aspects of a system that are strongly connected to its ability to express its phenomenal experiences (or remember them for later expression) cannot be systematically removed to test for correlations with consciousness because

[b] See [Chalmers, 1996; Moor, 1988; Prinz, 2003; Van Heuveln *et al.*, 1998] for a detailed discussion of this experiment.

this would destroy the measuring instrument that is needed for the experiments. Memory and expressive behavior, such as vocalization, can be removed individually — for example, in short-term memory loss patients or REM sleep — but if both are lost together, then we can no longer measure consciousness in the system.

4.2. *Type II PCCs*

Type II PCCs can be separated out using behavior and there is no overlap with the parts of the system that are used for measuring or reporting consciousness. When a type II PCC is removed or altered, the system's reports of conscious states can change, and this makes it possible to carry out experiments on the link between type II PCCs and consciousness. Many cognitive and functional correlates fall into this category because we can examine patients with reduced functions, such as imagination [Addis *et al.*, 2008], and measure their consciousness through verbal or other behavior.

4.3. *Type I PCCs and artificial systems*

With biological systems the distinction between type I and type II PCCs has few practical implications because mammalian brains have approximately the same type I PCCs as the human brain. Experiments on type II PCCs on humans will enable us to develop theories of consciousness that can be used to make predictions about the phenomenal states of humans and other animals on the basis of third-person data — a procedure commonly known as neuro-phenomenology.

However, the distinction between type I and type II PCCs has substantial implications for our judgments about whether artificial systems are potentially conscious. Artificial systems typically lack many or all of the type I PCCs, and we have no way of identifying which of them are necessarily present in potentially conscious systems. If we cannot prove that hemoglobin, for example, is not a correlate of consciousness, then we cannot tell if an artificial system without hemoglobin is potentially conscious.

The next section outlines a number of possible responses to this problem. The fundamental limits to the experiments on consciousness make the choice of response a pragmatic decision, and not something that might be empirically settled at some point in the future.

5. Responses to the Problem

5.1. *Suspend judgment*

One approach to this problem is to follow Prinz [2003] and suspend judgment about whether robots are capable of phenomenal states. A first difficulty with this approach is that many people have a strong intuition that machines built in a similar way to humans are likely to be phenomenally conscious, and so it may be necessary to take the idea that certain types of machines have conscious experiences seriously. A second

November 5, 2009 5:02:09pm WSPC/258-IJMC 00019 FA1

issue is that as machine consciousness progresses we are likely to start developing machines that exhibit more complex behavior and spend a lot of time being confused and potentially in pain [Metzinger, 2003]. If we suspend judgment about whether machines really feel pain, it would be difficult to address these ethical worries without stifling research. A third problem is that as more sophisticated robots emerge, people are inevitably going to attribute more and more consciousness to them. People are already prepared to attribute emotions to robots as simple as Braitenberg's vehicles [Dautenhahn, 2007], and a systematic way of evaluating the potential for conscious states in a system needs to be in place before this becomes a live public issue. The general public is very interested in the question whether something is really conscious, and it would be helpful if the machine consciousness community could formulate some kind of answer, even if it is based on analogy with human beings.

5.2. *Assume that type I PCCs are correlates of consciousness*

A second option is to assume that *all* of the type I attributes of the human brain are necessary for its association with conscious states — with small variations being allowed to take account of the range of brain sizes and weights. This is a conservative position that denies the potential for consciousness to virtually all of our current and future artificial systems — something we are likely to question as artificial systems are developed with more and more sophisticated forms of behavior.

5.3. *Assume that type I PCCs are not correlates of consciousness*

This is a very liberal position that makes type I PCCs irrelevant to the question about whether a system is capable of consciousness. Consciousness would be possible in any system that implemented the functions of the human brain and could pass Harnad's extended T3 version of the Turing test [Harnad, 1994], in which a human or artificial body is controlled in a way that is functionally indistinguishable from a human for 70 years or more. The same potential for consciousness would be attributed to the population of China collaborating using radios and satellites as to a biological system controlled by neurons.

5.4. *The potential for consciousness scale*

This option gives some weight to the potential association between type I PCCs and consciousness and takes account of the fact that we are more likely to attribute phenomenal states to systems with more type I PCCs. A potential for consciousness scale can be constructed by assigning numbers to each type I PCC and combining them in a systematic way to model our subjective judgments about consciousness. This enables the scale to predict what people would say about a system's potential for consciousness, based solely on its type I PCCs. Systems with few type I PCCs would be lower down on a potential for consciousness scale; systems with more type 1 attributes of the human brain would be higher up. This approach can handle the fact that our belief in the potential for consciousness of different machines is graded, and since it is a psychometric scale, it is not attempting to answer the metaphysical question about which type I PCCs are actually correlated with consciousness. An example of a potential for consciousness scale can be found in [Gamez, 2008a], Chap. 4, and an implementation is available online [Gamez, 2008b].

6. Conclusions

Thought experiments have led to many useful results in science, but their limited applicability to the study of consciousness suggests that questions about artificial systems' potential for consciousness should be addressed through scientific experiments. However, the problem with an empirical approach is that the measurement of consciousness using behavior places fundamental limits on the experiments that can be carried out, and so it cannot be established whether type I PCCs, such as hemoglobin or biological neurons, are actually correlated with consciousness. An artificial system could possess all of the type II PCCs and behave exactly like a human being, but it will be impossible to prove whether it is potentially conscious or not. Possible responses to this problem include suspension of judgment, liberal and conservative attribution of the potential for consciousness to machines, and the development of a psychometric scale that models our attribution of the potential for consciousness on the basis of type I PCCs.

Acknowledgments

Many thanks are due to Owen Holland and Igor Aleksander for their feedback, support and advice about this work. This research was funded by Grants from the Engineering and Physical Science Research Council Adventure Fund (GR/S47946/01) and the *Association for Information Technology Trust*.

References

Addis, D. R., Wong, A. T. and Schacte, D. L. [2008] "Age-related changes in the episodic simulation of future events", *Psychological Science* **19**, 33–41.

Balduzzi, D. and Tononi, G.[2008] "Integrated information in discrete dynamical systems: motivation and theoretical framework", *PLoS Computational Biology* **4**, e1000091.

Block, N. [1978] "Troubles with functionalism, in Minnesota studies in the philosophy of science," in *Perception and Cognition Issues in the Foundations of Psychology*, ed. Savage, C. W., Vol. 9 (University of Minnesota Press, Minneapolis).

Chalmers, D. [1996] *The Conscious Mind* (Oxford University Press, Oxford).

Chalmers, D. [1998] "On the search for the neural correlates of consciousness," in *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*, eds. Hameroff, S., Kaszniak, A. and Scott, A. (MIT Press, Cambridge, Massachusetts).

Chrisley, R. [2009] "Synthetic phenomenology," *International Journal of Machine Consciousness* **1**, 53–70.

Crick, F. [1994] *The Astonishing Hypothesis* (Simon & Schuster, London).

Dautenhahn, K. [2007] "Socially intelligent robots: Dimensions of human–robot interaction," *Phil. Trans. R. Soc. B* **362**, 679–704.

November 5, 2009 5:02:09pm WSPC/258-IJMC 00019 FA1

Dehaene, S. and Naccache, L. [2001] "Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework," *Cognition* **79**, 1–37.
Edelman, G. N. and Tononi, G. [2000] *Consciousness: How Matter Becomes Imagination* (Penguin, London).
Flohr, H. [2000] "NMDA Receptor-mediated computational processes and phenomenal consciousness," in *Neural Correlates of Consciousness*, ed. Metzinger, T. (MIT Press, Cambridge, Massachusetts and London, England), pp. 245–258.
Gamez, D. [2007] *What We Can Never Know* (Continuum, London and New York).
Gamez, D. [2008a] *The Development and Analysis of Conscious Machines*, PhD thesis, University of Essex, Essex.
Gamez, D. [2008b] "Online implementation of a potential for consciousness scale," `http://www.davidgamez.eu/mc-thesis/omc_scale/index.html`.
Gendler, T. S. and Hawthorne, J. [2002] "Introduction," in *Conceivability and Possibility*, eds. Gendler, T. S. and Hawthorne, J. (Clarendon Press, Oxford), pp. 1–70.
Harnad, S. [1994] "Levels of functional equivalence in reverse bioengineering: The Darwinian turing test for artificial life," *Artificial Life* **1**, 293–301.
Hume, D. [1985] *A Treatise of Human Nature* (Penguin, London).
Kuhn, T. [1977] "A function for thought experiments," in *The Essential Tension*, ed. Kuhn, T. (University of Chicago Press, Chicago), pp. 240–265.
Metzinger, T. (ed.) [2000], *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (MIT Press, Cambridge, Massachusetts).
Metzinger, T. [2003] *Being No One* (MIT Press, Cambridge, Massachusetts).
Moor, J. H. [1988] "Testing robots for qualia", in *Perspectives on Mind*, eds. Otto, H. R. and Tuedio, J. A. (D. Reidel Publishing Company, Dordrecht/Boston/Lancaster/Tokyo).
Newman, J., Baars, B. J. and Cho, S.-B. [1997] "A neural global workspace model for conscious attention," *Neural Networks* **10**, 1195–1206.
Noë, A. and Thompson, E. [2004] *Journal of Consciousness Studies* **11**, 3–28.
Popper, K. [2002] *The Logic of Scientific Discovery* (Routledge, London and New York).
Prinz, J. J. [2003] "Level-headed mysterianism and artificial experience," in *Machine Consciousness*, ed. Holland, O. (Imprint Academic, Exeter).
Searle, J. R. [1980] "Minds, brains and programs," *Behavioral and Brain Sciences* **3**, 417–457.
Seth, A. K. [2009] "The strength of weak artificial consciousness," *International Journal of Machine Consciousness* **1**, 71–82.
Singer, W. [2000] "Phenomenal awareness and consciousness from a neurobiological perspective," in *Neural Correlates of Consciousness*, ed. Metzinger, T. (MIT Press, Cambridge, Massachusetts and London, England), pp. 121–137.
Van Heuveln, B., Dietrich, E. and Oshima, M. [1998] "Let's dance! The equivocation in Chalmers' dancing qualia argument," *Minds and Machines* **8**, 237–249.
Wilkes, K. V. [1988] *Real People; Personal Identity without Thought Experiments* (Clarendon Press, Oxford).
Zeki, S. [2003] "The disunity of Consciousness," *Trends in Cognitive Sciences* **7**, 214–218.