THE CO-EVOLUTION OF MATTER AND CONSCIOUSNESS¹

Max Velmans, Goldsmiths, University of London, New Cross, London SE14 6NW; email m.velmans@gold.ac.uk; web address http://www.goldsmiths.ac.uk/psychology/staff/velmans.php

Synthesis Philosophica (2007), Vol. 22, No.44, fasc.2, pp. 273-282.

Keywords. Consciousness, evolution, co-evolution, matter, continuity, discontinuity, complexity, brain

Abstract. Theories about the evolution of consciousness relate in an intimate way to theories about the distribution of consciousness, which range from the view that only human beings are conscious to the view that all matter is in some sense conscious. Broadly speaking, such theories can be classified into discontinuity theories and continuity theories. Discontinuity theories propose that consciousness emerged only when material forms reached a given stage of evolution, but propose different criteria for the stage at which this occurred. Continuity theories argue that in some primal form, consciousness co-evolved, for example into the forms that we now recognise in human beings. Given our limited knowledge of the necessary and sufficient conditions for the presence of human consciousness in human brains, all options remain open. On balance however continuity theory appears to be more elegant than discontinuity theory.

The distribution of consciousness.

Theories about the evolution of consciousness are linked to theories about the distribution of consciousness. Are we the only conscious beings? Or are other animals and other living systems also conscious and, if so, might consciousness extend to non-living systems such as computers? Philosophers and scientists have expressed many different views on these matters. As the data needed to *decide* these matters is not currently available, all views are partly speculative. Why? Because we do not even know the necessary and sufficient conditions for consciousness in our own brains! As John (1976) pointed out we do not know the physical and chemical interactions involved, how big a neuronal system must be to sustain it, nor even whether it is confined to brains. Over 30 years later, little has changed. Given this underdetermination by the data, opinions about the distribution of consciousness have ranged from the ultra-conservative (only humans are conscious) to the extravagantly libertarian (everything that might possibly be construed as having consciousness).

The view that only humans have consciousness has a long history in theology, following naturally from the doctrine that only human beings have souls. Some philosophers and scientists have elaborated this doctrine into a philosophical position. According to Descartes only humans combine *res cogitans* (the stuff of consciousness) with res extensa (material stuff). Non-human animals, which he refers to as "brutes", are nothing more than nonconscious machines. Lacking consciousness, they do not have reason or language. Eccles (in Popper & Eccles, 1976) adopted a similar, dualist position – but argued that it is only through human

language that one can communicate sufficiently well with another being to *establish* whether it is conscious. Without language, he suggests, the only defensible option is agnosticism or doubt. Jaynes (1990) by contrast, argued that human language is a *necessary condition* for consciousness. And Humphrey (1983) adopted a similar view, arguing that consciousness emerged only when humans developed a "theory of mind." He accepts that we might find it useful for our own ethical purposes to treat other animals *as if* they were conscious, but without self-consciousness of the kind provided by a human "theory of mind" they really have no consciousness at all! There are other, modern variants of this position (e.g. Carruthers, 1998) but we do not need an exhaustive survey. It is enough to note that thinkers of very different persuasions have held this view. Early versions of this position appear to be largely informed by theological doctrine; later versions are based on the supposition that higher mental processes of the kinds unique to humans are necessary for consciousness of any kind.

In my book Understanding Consciousness, I argue that this extreme position has little to recommend it when applied to humans, let alone other animals. Phenomenal consciousness in humans is constructed from different exteroceptive and interoceptive resources and is composed of different "experiential materials" (what we see, hear, touch, taste, smell, feel and so on). It is true that our higher cognitive functions also have manifestations in experience, for example, in the form of verbal thoughts. Consequently, without language and the ability to reason, such thoughts would no longer be a part of what we experience (in the form of "inner speech"). But one can lose some sensory and even mental capacities while other capacities remain intact (in cases of sensory impairment, aphasia, agnosia and so on). And there is no scientific evidence to support the view that language, the ability to reason and a theory of mind are necessary conditions for visual, auditory and other sensory experiences. Applied to humans, this view is in any case highly counterintuitive. If true, we would have to believe that, prior to the development of language and other higher cognitive functions, babies experience neither pleasure nor pain, and that their cries and chuckles are just the nonconscious output of small biological machines. We would also have to accept that autistic children without a "theory of mind" never have any conscious experience! To any parent, such views are absurd.

Such views confuse the necessary conditions for the *existence* of consciousness with the added conditions required to support its many *forms*. Consciousness in humans appears to be regulated by global arousal systems, modulated by attentional systems that decide which representations (of the external world, body and mind/brain itself) are to receive focal attention. Neural representations, arousal systems and mechanisms governing attention are found in many other animals (Jerison, 1985). Other animals have sense organs that detect environmental information and perceptual and cognitive processes that analyse and organise that information. Many animals are also able to communicate and live in complex emotional and social worlds (Dawkins, 1998, Panksepp, 2007). Overall, the precise mix of sensory, perceptual, cognitive and social processes found in each species is likely to be species-specific. Given this, it might be reasonable to suppose that only humans can have full *human* consciousness. But it is equally reasonable to suppose that some non-human animals have unique, non-human forms of consciousness.

Given the evidence for the gradual evolution of the human brain, it also seems unlikely that consciousness first emerged in the universe, fully formed, in *homo sapiens*. As the naturalist Thomas Huxley observed in 1874,

"The doctrine of continuity is too well established for it to be permissible to me to suppose that any complex natural phenomenon comes into existence suddenly, and without being preceded by simpler modifications; and very strong arguments would be needed to prove that such complex phenomena as those of consciousness, first make their appearance in man."

Is consciousness confined to complex brains?

One cannot be *certain* that other animals are conscious – or even that other people are conscious (the classical problem of "other minds"). However, the balance of evidence strongly supports it (Dawkins, 1998, Panksepp, 2007). In cases where other animals have brain structures that are similar to humans, that support social behaviour that is similar to humans (agression, sexual activity, pair-bonding and so on), it is difficult to believe that they experience nothing at all! But if one does not place the conscious/nonconscious boundary between humans and non-humans where should one place it?

It might be that consciousness is confined to animals whose brains have achieved some (unknown) critical mass or critical complexity. The contents of human consciousness are constructed from different sense modalities, and within a given sense modality, experiences can be of unlimited variety and be exquisitely detailed. Where such conscious states are complex, the neural states that support them must have equivalent complexity. However it does not follow from this that *only* brains of similar complexity can support *any* experience. Complex, highly differentiated brains are likely to be needed to support complex, highly differentiated experiences. But it remains possible that relatively simple brains can support relatively simple experiences.

Given this, it is tempting to search for the conditions that distinguish conscious from nonconscious processing in our own brains *irrespective* of complexity—for example to isolate neural changes produced by simple stimuli just above and below some threshold of awareness in different sense modalities. This is a sensible strategy that is widely pursued in psychology and associated brain sciences. In the human case, only representations at the focus of attention reach consciousness and then only in a sufficiently aroused state (an awake or dreaming state, but not coma or deep sleep), so it would be useful to learn what happens to such representations to *make* them conscious. Common suggestions are activation of specific consciousness-bearing circuitry (Crick & Koch, 2007; Rees & Frith, 2007), "neural binding" produced by relatively coherent, phase-locked activity of some neural sub-populations relative to the uncoordinated activity of other populations (Singer, 2007), and a transition from modular, restricted forms of information processing to widespread information dissemination throughout the brain (Baars, 2007, Dehaene & Naccache, 2001).

Even if one of these or some combination of these conditions for consciousness turn out to be necessary for consciousness in the human mind/brain, we *still* need to be

cautious about treating such conditions as universal. Under normal conditions, the human mind/brain receives simultaneous information from a range of sense organs that simultaneously monitor the external and internal environment and this information needs to be related to information in long-term memory, and assessed for importance in the light of ongoing needs and goals. In short, there are many things going on at once. But we cannot give everything our full, undivided attention. As Donald Broadbent pointed out in 1958, there is a "bottleneck" in human information processing. The human effector system is also limited—we only have two eyes, hands, legs etc., and effective action in the world requires precise co-ordination of eye-movements, limbs and body posture. As a result, the mind/brain needs to select the most important information, to decide on best strategy and to co-ordinate its activity sufficiently well to interact with the world in a coherent, integrated way.

To achieve this, it is as important to *stop* things happening in the brain as it is to make them happen. As William Uttal observed

"There is an a priori requirement that some substantial portion, perhaps a majority, of the synapses that occur at the terminals of the myriad synaptic contacts of the threedimensional ... (neural) ... lattice must be inhibitory. Otherwise the system would be in a constant state of universal excitement after the very first input signal, and no coherent adaptive response to complex stimuli would be possible" (Uttal, 1978, p192).

This opens up the possibility that selective attention doesn't so much *add* something special to neural representational states at the focus of attention to give them associated consciousness. Consciousness might be a "natural" accompaniment of neural representation (see for example Zeki, 2007). If so, it may just be that for attended to representational states, inhibitory processes don't prevent it. To prevent information overload, not to mention utter confusion, information and awareness of information that is integrated into a representation of the current, "psychological present" might be released from inhibition (Arbuthnott, 1995).

If so, the mechanisms required to select, co-ordinate, integrate and disseminate conscious information in the human brain may not be required for simpler creatures, with simpler brains. If consciousness is a natural accompaniment of neurally encoded information, such creatures might have a simple form of consciousness.

The visual system of the frog, for example, appears to be structured to respond to just four stimulus features: a sustained contrast in brightness between two portions of the visual field, the presence of moving edges, the presence of small moving spots and an overall dimming of the visual field. This is a far cry from the variety and detail provided by the human visual system. But there seems little reason to jump to the conclusion that the frog sees nothing. Rather, as Lettvin, *et. al.* (1959) proposed, the frog may see just four things relating to its survival. A sudden dimming of the light or a moving edge may indicate the presence of a predator and is likely to initiate an escape response. Sustained differences in brightness may allow the frog to see (and catch) a moving fly at tongue's length.

As one continues to descend the evolutionary ladder, the plausibility of extrapolating from human to non-human animal consciousness becomes increasingly remote. There may, for example, be critical transition points in the development of consciousness that accompany critical transitions in functional organisation (Sloman, 1997). Self-awareness, for example, probably occurs only in creatures capable of self-representation. That said, phenomenal consciousness (of any kind) might only require representation. If so, even simple invertebrates might have some rudimentary awareness, in so far as they are able to represent and, indeed, respond to certain features of the world.

Planarians (flat worms) for example, can be taught to avoid a stimulus light if it has been previously associated with an electric shock (following a classical conditioning procedure). And simple molluscs such as the sea-hare Aplysia that withdraw into their shells when touched, respond to stimulus "novelty." For example, they may habituate (show diminished withdrawal) after repeated stimulation at a given site, but withdraw fully if the same stimulation is applied to another nearby site. Habituation in Aplysia appears to be mediated by events at just one centrally placed synapse between sensory and motor neurons. This is very simple learning, and it is very difficult to imagine what a mollusc might experience. But if the ability to learn and respond to the environment were the criterion for consciousness, there would be no principled grounds to rule this out. It might be, for example, that simple approach and avoidance are associated with rudimentary experiences of pleasure and pain.

Is consciousness confined to brains?

It is commonly thought that the evolution of human consciousness is intimately linked to the evolution of the neocortex (e.g. Jerison, 1985)—and it seems likely that cortical structures play a central role in determining the forms of consciousness that we experience. However, whether consciousness first emerged with the emergence of the neocortex or whether there is something special about the nature of cortical cells that somehow "produces" consciousness is less certain. As Charles Sherrington has pointed out, there appears to be nothing special about the internal structure of brain cells that might make them uniquely responsible for mind or consciousness. For,

"A brain-cell is not unalterably from birth a brain-cell. In the embryo-frog the cells destined to be brain can be replaced by cells from the skin of the back, the back even of another embryo; these after transplantation become in their new host brain-cells and seem to serve the brain's purpose duly. But cells of the skin it is difficult to suppose as having a special germ of mind. Moreover cells, like those of the brain in microscopic appearance, in chemical character, and in provenance, are elsewhere concerned with acts wholly devoid of mind, e.g. the knee-jerk, the light-reflex of the pupil. A knee-jerk 'kick' and a mathematical problem employ similar-looking cells. With the spine broken and the spinal cords so torn across as to disconnect the body below from the brain above, although the former retains the unharmed remainder of the spinal cord consisting of masses of nervous cells, and retains a number of nervous reactions, it reveals no trace of recognizable mind.... Mind, as attaching to any unicellular life would seem to be unrecognizable to observation; but I would not feel that permits me to affirm that it is not there. Indeed, I would think, that since mind appears in the developing source that amounts to showing that it is potential in the ovum (and sperm) from which the source spring. The appearance of recognizable

mind in the source would then be not a creation de novo but a development of mind from unrecognizable into recognizable." (Sherrington, 1942)

Indeed, given our current, limited knowledge of the necessary and sufficient conditions for consciousness in humans, we cannot, as yet, rule out even more remote possibilities. If the ability to represent and respond to the world, or the ability to modify behaviour consequent on interactions with the world are the criteria for consciousness then it may be that consciousness extends not just to simple invertebrates (such as Planaria) but also to unicellular organisms, fungi and plants. For example, the leaflets of the Mimosa plant habituate to repeated stimulation, i.e. the leaflets rapidly close when first touched, but after repeated stimulation they reopen fully and do not close again while the stimulus remains the same. Surprisingly, this habituation is stimulus-specific. For example, Holmes & Yost (1966) induced leaflet closure using either water droplets or brush strokes, and after repeated stimulation (with either stimulus) habituation occurred. But, if the stimulus was changed (from water drops to brush strokes or vice-versa) leaflet closure re-occurred.

For many who have thought about this matter, the transition from rudimentary consciousness in animal life to sentience in plants is one transition too far. Perhaps it is. It is important to note however that a criterion of consciousness based on the ability to respond to the world does not prevent it. Nor, on this criterion, can we rule out the possibility of consciousness in systems made of materials other than the carbon-based compounds that (on this planet) form the basis for organic life. Silicon-based computers can in principle carry out many functions that, in humans, we take to be evidence of conscious minds. So how can we be certain that they are not conscious?

One should recognise too, that even a criterion for the existence of consciousness based on the ability to respond or adapt to the world is entirely arbitrary. It might for example be like something to *be* something irrespective of whether one *does* anything! Panpsychists such as Whitehead (1929) have suggested that there is no arbitrary line in the descent from macroscopic to microscopic matter at which consciousness suddenly appears out of nothing. Rather, elementary forms of matter may be associated with elementary forms of experience. And if they encode information they may be associated with rudimentary forms of mind.

Does matter matter?

Many would regard Whitehead's views as extreme (I give my own assessment below). But there is one position that is even more extreme – the view that the nature of matter doesn't matter to consciousness at all. At first glance, it might seem preposterous to claim that matter doesn't matter for consciousness. But, surprising as it might seem, it is a logical consequence of *computational functionalism*—one of the most widely adopted, current theories of mind. As John Searle has noted, it is important to distinguish this position from the view that *silicon robots* might be conscious. For him, human consciousness in spite of its subjectivity, intentionality, and qualia is an emergent *physical* property of the brain. If so, a silicon robot *might* have consciousness. But this would depend not on its programming, but on whether silicon just happens to have the same causal powers (to produce consciousness) as the carbon-based material of brains.

Computational functionalists such as Daniel Dennett take the further step that, apart from providing housing for functioning, material stuff is irrelevant. *Any* system that functions *as-if* it has consciousness and mind *does* have consciousness and mind. If a non-biological system functions exactly like a human mind then it has a human mind, as the only thing that makes a system a "mind" is the way that it functions. In its usual reductionist versions, computational functionalism finesses questions about the distribution of first-person consciousness, routinely translating these into questions about how different systems function (see *Understanding Consciousness*, chapters 4 and 5).

Can one draw a line between things that have consciousness and those that don't?

Where then should one draw the line between entities that are conscious and those that are not? Theories about the distribution of consciousness divide into *continuity* and *discontinuity* theories. Discontinuity theories all claim that consciousness emerged at a particular point in the evolution of the universe. They merely disagree about which point. Consequently, discontinuity theories all face the same problem. What switched the lights on? What is it about matter, at a particular stage of evolution, which suddenly gave it consciousness? As noted above, most try to define the point of transition in functional terms, although they disagree about the nature of the critical function. Some think consciousness "switched on" only in humans, for example once they acquired language or a theory of mind. Some believe that consciousness emerged once brains reached a critical size or complexity. Others believe it co-emerged with the ability to learn, or to respond in an adaptive way to the environment.

In my view, such theories confuse the conditions for the *existence* of consciousness with the conditions that determine the many *forms* that it can take. Who can doubt that verbal thoughts require language, or that full human self-consciousness requires a theory of mind? Without internal representations of the world, how could consciousness be *of* anything? And without motility and the ability to approach or avoid, what point would there be to rudimentary pleasure or pain? However, none of these theories explains what it is about such biological functions that suddenly switches consciousness on.

Continuity theorists do not face this problem for the simple reason that they do not believe that consciousness suddenly emerged at *any* stage of evolution. Rather, as Sherrington suggests above, consciousness is a "development of mind from unrecognizable into recognizable." On this *panpsychist* view, all forms of matter have an associated form of consciousness, although in complex life forms such as ourselves, much of this consciousness is inhibited. In the cosmic explosion that gave birth to the universe, consciousness co-emerged with matter and co-evolves with it. As matter became more differentiated and developed in complexity, consciousness became correspondingly differentiated and complex. The emergence of carbon-based life forms developed into creatures with sensory systems that had associated sensory "qualia." The development of *representation* was accompanied by the development of consciousness that is *of* something. Once conscious states were associated organisms

those states became *organism-centric* and *perspectival*. With the development of motility and the need to approach beneficial stimuli and avoid harmful ones came the beginnings of pleasure and pain. The development of *self-representation* was accompanied by the dawn of differentiated self-consciousness and so on. On this view, evolution accounts for the different *forms* that consciousness takes—and, in this respect, continuity theory does not differ, in principle, from discontinuity theory.

However, consciousness, in some primal form, did not emerge at any particular stage of evolution. Rather, it was there from the beginning. Its emergence, with the birth of the universe, is neither more nor less mysterious than the emergence of matter, energy, space and time.

Most discontinuity theorists take it for granted that consciousness could only have appeared (out of nothing) through some random mutation in complex life forms that happened to confer a reproductive advantage (inclusive survival fitness) that can be specified in third-person functional terms. This deeply ingrained, pre-theoretical assumption has set the agenda for what discontinuity theorists believe they need to explain. Within cognitive psychology, for example, consciousness has been thought by one or another theorist to be necessary for every major phase of human information processing, for example in the analysis of complex or novel input, learning, memory, problem solving, planning, creativity, and the control and monitoring of complex, adaptive response. I have presented extensive analyses of the role of consciousness in human information processing that cast doubt on all these suggestions (Velmans, 1991a,b, 1993, 1996, 2000, 2002a,b, 2003).

It should be apparent that continuity theory shifts this agenda. The persistence of different, emergent biological forms may be governed by reproductive advantage. If each of these biological forms has a unique, associated consciousness, then matter and consciousness co-evolve. However, conventional evolutionary theory does not claim that *matter itself* came into being, or persists through random mutation and reproductive advantage. According to continuity theory, neither does consciousness.

Which view is correct? One must choose for oneself. In the absence of anything other than arbitrary criteria for when consciousness suddenly emerged, I confess that I find continuity theory to be the more elegant. There may be critical transition points in the *forms* of consciousness associated with the development of life, representation, self-representation, and so on. However continuity in the evolution of consciousness favours continuity in the distribution of consciousness.

References

Arbuthnott, K.D. (1995) 'Inhibitory mechanisms in cognition: Phenomena and models', *Cahiers de Psychologie Cognitive* 14(1): 3-45.

Baars, B.J. (2007) The global workspace theory of consciousness. In M.Velmans and S. Schneider (eds.) (2007) *The Blackwell Companion to Consciousness*. Malden, MA: Blackwell, pp 236-246.

Carruthers, P. (1998) 'Natural theories of consciousness', *European Journal of Philosophy*, 6(2): 203-222.

Crick, F. and Koch, C. (2007) A neurobiological framework for consciousness. In In M.Velmans and S. Schneider (eds.) (2007) *The Blackwell Companion to Consciousness*. Malden, MA: Blackwell, pp 567-579.

Dawkins, M.S. (1998) Through Our Eyes Only? The Search for Animal Consciousness, Oxford: Oxford University Press.

Dehaene, S. and Naccache, L. (2001) 'Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework', *Cognition* 79: 1-37.

Holmes, E. and Yost, M. (1966) 'Behavioral studies in the sensitive plant', Worm Runners Digest 8:38.

Humphrey, N. (1983) Consciousness Regained, Oxford: Oxford University Press.

Jaynes, J. (1979) *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, London: Allen Lane.

Jerison, H.J. (1985) 'On the evolution of mind', in D. A. Oakley (ed.) *Brain and mind*, London: Methuen.

John, E.R. (1976) 'A model of consciousness', in G. Schwartz and D. Shapiro (eds.) *Consciousness and Self-Regulation*, New York: Plenum Press.

Lettvin, J.Y., Maturana, H.R, McCulloch, W.S. and Pitts, W.H. (1959) 'What the frog's eye tells the frog's brain', *Institute of Radio Engineer's Proceedings* 47: 1940-1951.

Panksepp, J. (2007) Affective consciousness. In M.Velmans and S. Schneider (eds.) (2007) *The Blackwell Companion to Consciousness*. Malden, MA: Blackwell, pp 114-129.

Popper, K.R. and Eccles, J.C. (1993[1976]) The Self and its Brain, London: Routledge.

Rees, G. and Frith, C. (2007) Methodologies for identifying the neural correlates of consciousness. In M.Velmans and S. Schneider (eds.) (2007) *The Blackwell Companion to Consciousness*. Malden, MA: Blackwell, pp 553-566.

Singer, W. (2007) Large-scale temporal coordination of cortical activity as a prerequisite for conscious experience. In M.Velmans and S. Schneider (eds.) (2007) *The Blackwell Companion to Consciousness*. Malden, MA: Blackwell, pp 605-615.

Sherrington, C.S. (1942) Man on His Nature, Cambridge: Cambridge University Press.

Sloman, A. (1997b) 'What sorts of machine can love? Architectural requirements for human-like agents both natural and artificial'. http://www.sbc.org.uk/literate.htm.

Uttal, W.R. (1978) The Psychobiology of Mind. Hillsdale, NJ: Lawrence Erlbaum.

Velmans, M. (1991a) 'Is human information processing conscious?' *Behavioral and Brain Sciences* 14(4): 651-669. http://cogprints.ecs.soton.ac.uk/archive/00000593/

Velmans, M.(1991b) 'Consciousness from a first-person perspective', *Behavioral and Brain Sciences* 14(4): 702-726. http://cogprints.ecs.soton.ac.uk/archive/00000594/

Velmans, M. (1993) 'Consciousness, causality and complementarity', *Behavioral and Brain Sciences* 16(2): 409-416. http://cogprints.ecs.soton.ac.uk/archive/00000595/

Velmans, Max (1996) Consciousness and the "Causal Paradox". *Behavioral and Brain Sciences* 19(3): 538-542. http://cogprints.ecs.soton.ac.uk/archive/00000596/

Velmans, M. (2000) Understanding Consciousness, London: Routledge/Psychology Press.

Velmans, M. (2002a) How could conscious experiences affect brains? (Target Article for Special Issue) *Journal of Consciousness Studies*, 9 (11): 3-29. http://cogprints.ecs.soton.ac.uk/archive/00002750/

Velmans, M (2002b) Making sense of the causal interactions between consciousness and brain (a reply to commentaries) *Journal of Consciousness Studies*, 9 (11): 69-95. http://cogprints.ecs.soton.ac.uk/archive/00002751/

Velmans, M (2003) *How Could Conscious Experiences Affect Brains*? Exeter: Imprint Academic.

Whitehead, A.N. (1957[1929]). Process and Reality, New York: Macmillan.

Zeki, S. (2007) A theory of microconsciousness. In M.Velmans and S. Schneider (eds.) (2007) *The Blackwell Companion to Consciousness*. Malden, MA: Blackwell, pp 580-588.

¹This paper is adapted from parts of Chapter 12 of Velmans (2000) Understanding Consciousness.