How Virtual Machinery Can Bridge the "Explanatory Gap", In Natural and Artificial Systems*

Aaron Sloman
School of Computer Science, University of Birmingham, UK
http://www.cs.bham.ac.uk/~axs

November 26, 2010

Abstract

We can now show in principle how evolution could have produced the "mysterious" aspects of consciousness if, like engineers in the last six or seven decades, it had to solve increasingly complex problems of representation and control by producing systems with increasingly abstract, but effective, mechanisms, including self-observation capabilities, implemented in non-physical virtual machines which, in turn, are implemented in lower level physical mechanisms. For this, evolution would have had to produce far more complex virtual machines than human engineers have so far managed, but the key idea might be the same. However it is not yet clear whether the biological virtual machines could have been implemented in the kind of discrete technology used in computers as we know them.
Keywords:
Architecture, Body, Causation, Cognition, Consciousness, Darwin, Designer Stance, Evolution, Explanatory Gap, Huxley, Mind, Virtual Machinery

1 Introduction: A Problem for Darwin

A problem that puzzled Darwin and fired up his critics, was how mental phenomena could fit into the theory of evolution by natural selection.

There was evidence for evolution of physical forms, including: fossil records showing gradual changes in skeletal structures, existing species that have been shown to adapt physical features and behaviours to meet changing circumstances, and artificially bred variants of animals and plants. Such evidence convinced Darwin and many of his contemporaries that random mutations and environmental selection pressures could, over time, produce radical changes. Despite gaps in the evidence, most scientists now seem to accept that the whole process, starting from complex molecules and leading to the existence of all the biological diversity now on earth, can be explained in roughly Darwinian terms – though there remain disagreements on some of the mechanisms, e.g. in Jablonka and Lamb (2005).

However, since Darwin's time till the present day, many serious thinkers, including some of his leading supporters, have doubted that there is compelling evidence for the claim that *mental* functioning evolved in the same way, and some think there is no evidence that human minds, or other animal minds, could be products of Darwinian evolution. For some, that is because all the evidence available supports only the hypothesis that evolutionary mechanisms can produce *physical* changes of shape and detailed physiology and to some extent also physical behaviour, such as feeding behaviours, mating behaviours, caring for offspring, and ways of escaping from predators.

If it were clear how physical changes can produce mental changes, that could help to provide an account of how evolution could produce new mental phenomena, and the behaviours that seem to require specifically human mental processes, such as production and enjoyment of poetry, stories, music and dancing, and the

*Published as Sloman(2010a).
Related PDF presentation available at http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#sab2010

advance of mathematics, science and technology. For it could be argued that evolution can produce mental phenomena by producing the required physical mechanisms. But many cannot conceive of physical matter producing mental processes and some even deny that it is possible at all, because of the huge gulf in kind between, on the one hand, behaviours of atoms, molecules, and larger structures composed of those, and, on the other hand, processes of experience, thought, pleasure, pain, and self-awareness.

Even one of Darwin's strongest supporters, T.H. Huxley, is widely reported to have written "How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the Djin when Aladdin rubbed his lamp"[1]. He was not alone. Romanes wrote in (Romanes, 1883) "But we are totally in the dark as to the causal connection, if any, between such a state of turmoil in the ganglion and the occurrence of consciousness."(p75) (quoted in Whittaker's review (Whittaker, 1884)). Moreover, Alfred Wallace, co-inventor of the theory of evolution by natural selection, doubted that evolution could produce anything like *states of consciousness*.

This problem was later labelled the "explanatory gap". Individuals use different names for what it is that they are opposing to physical phenomena. Huxley and Romanes used "consciousness". Some use "sentience". Following Block (Block, 1995), many now refer to "Phenomenal Consciousness" (PC) in contrast with "Access Consciousness" (AC), or, in the terminology of Chalmers (Chalmers, 1995), distinguish the so-called "Hard Problem" of consciousness from a (relatively) "Easy Problem". Such formulations presuppose a dichotomy: a binary divide between things that do and things that do not have the problematic extra feature over and above their physical features. Later we shall challenge the use of a binary division (as Thomas Whittaker did as long ago as 1884).

Debates about evolution of mind echoed and extended older philosophical discussions about the nature of mind and the relations between mind and body. Not only philosophers, but also psychologists, neuroscientists, physicists, anthropologists, biologists and more recently AI researchers, roboticists and cognitive scientists of various sorts have all been exercised about this. Responses to the questions about whether natural selection can produce mental competences and consciousness, and whether physical processes can produce mental processes, vary. The variations include: rejection of the problem as somehow due to a deep muddle (Ryle, 1949), claiming that it is a real problem but lacking any solution that human minds can understand (McGinn, 2004), offering a reformulation of the problem alleged to solve it (Dennett, 1991), resurrecting the problem with a new label (Nagel, 1981; Block, 1995; Chalmers, 1995), proposing a philosophical or scientific research project to solve it (Baars, 1988; Chalmers, 1996), offering specific solutions that appeal to recent advances in physics or mathematics (Penrose, 1989; Ryser, 2009), assembling experimental and observational data about it (Baars, 1988), producing working computer models of various kinds (Shanahan, 2005), developing new technical philosophical concepts in the hope of clarifying it (Kim, 1998), and many more. In 1978, I proposed that the best way to make progress on the philosophical problems was to use new opportunities provided by the development of computers to investigate ways of designing working minds possibly starting with working fragments (Sloman, 1978, Chapter 1). This has been ignored by most philosophers and the majority of psychologists. Those who tried have generally underestimated the problems, expecting success too soon.

Nevertheless, much has been learnt, including a great deal about the diversity of the phenomena involving consciousness and other aspects of mentality. This includes studies of development of various mental competences from infancy, e.g. (Gibson & Pick, 2000; Rochat, 2001), various "disorders of consciousness" caused by brain-damage, physical or mental abuse, and the effects of drugs of various kinds, including local and global anaesthetics used medically.

2 Causes of Difficulty

Despite the vast amount in print, there does not seem to be any clear consensus that one theory of the relationship between mind and brain is right, or even along the right lines. I think there are two main reasons why philosophy has got stuck, and with it theoretical psychology and biology.

[1]The source is alleged to be (Huxley, 1866), though I could not find the words in an online version.

The first hurdle is the difficulty of identifying the problems in a form that genuinely addresses all the main concerns that have arisen in the history pointed to here. Part of the explanation for the difficulty is that there is no *one* problem: rather biological evolution had to solve *many* design problems in the long slow march from microbes to species clearly demonstrating mental competences. This is related to the claim in (Sloman, 2009b, 2010c) that there is no one thing referred to by the noun "consciousness", because the concept of being conscious of something is highly polymorphic – it refers to many different types of phenomena, which vary according to which kind of thing is conscious and what kind of thing it is conscious of. For example being conscious of a pain in your leg is very different from being conscious of your growing unpopularity at work. Both are different from a fly escaping a swatter because it was conscious of the motion. In a biological context this diversity is to be expected: diversity is a pervasive aspect of products of evolution.

The second reason for lack of progress is that most researchers lack conceptual tools adequate for the task of formulating answers with explanatory power. I shall try to show that people working in computer science and software engineering have, over several decades, unintentionally provided new concepts and tools for thinking about, modelling and explaining some of the kinds of phenomena cited by objectors to Darwinism. The key idea is that all organisms use information: living things are *informed control systems* – they use information in making control decisions, i.e. selecting between alternative possible actions, internal or internal. This idea is not new. But it is often ignored by people who ask how matter can produce or influence mind without asking how mind can influence matter and its motion, which it clearly does, e.g. as I type these words.

Early versions of this claim are in books by Craik and Wiener, (Craik, 1943; Wiener, 1961) published in 1943 and 1948. But we have learnt much since then. In particular, whereas early information-based machines (e.g. Jacquard looms) used information to control *physical* actions, the information-processing machinery developed using computers has increasingly been concerned with acting *on information*, and acting *on abstract mechanisms for acting on information* (a possibility Ada Lovelace foresaw). Most of what computers now do is not describable in the language of physics: they run processes in *virtual machines* that are implemented in physical machines though what the virtual machines are and what they do cannot be fully described using the language of the physical sciences. For example, the concepts of "winning" and "losing", required to describe the operation of a computer-based chess program, refer not to physical processes but to more abstract interactions between competing agents (Sloman, 2009a).

3 Towards Better Conceptual Tools

The conceptual tools required for building, testing, debugging, extending, comparing and explaining such virtual machines were developed piecemeal over several decades by hardware and software engineers solving different problems. Portions of the story are in (Dyson, 1997). We need to understand what they have achieved and its relevance to information processing in organisms.

It is sometimes suggested that if you describe a computer as running a certain virtual machine, e.g. a chess playing virtual machine, you are merely hallucinating a pattern onto the physical changes occurring, like choosing to see a rectangular array of dots as composed of horizontal rows of dots, or of vertical columns of dots, implying that virtual machines exist only in the eye (or mind) of the beholder, and cannot *do* anything. But that ignores the *causal* interactions that occur within virtual machines and also between virtual and physical processes. Chess virtual machines really do consider the consequences of certain options and on that basis choose one of them in deciding what move to make next – which in turn causes further changes within the virtual machine and in the computer's memory and possibly also on the screen. For blind human users it may utter the coordinates of the move. The causation is not hallucinated.

We may fancy that one piece of shadow on a forest floor chases another, or that an arrow shaped shadow points at a patch of light, whereas in fact there is no chasing or pointing: the visible patterns are mere by-products of interactions between wind, leaves, branches and rays of sunlight. If an arrow-shaped shadow appears to point at a patch of light that is mere coincidence. In contrast, engineers have learnt how to make enduring, active patterns in computers that really do influence other patterns, which may themselves

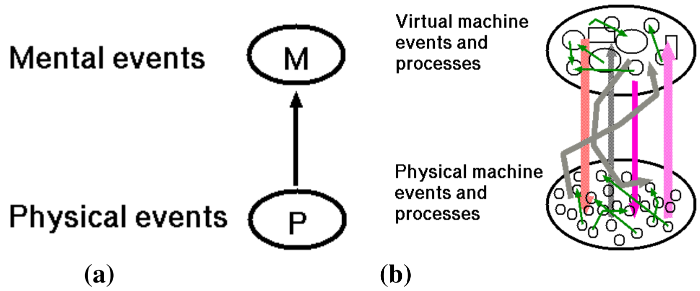

Figure 1: In (a), illustrating epiphenomenalism, physical changes produce patterns visible to a viewer, which do not themselves interact causally, whereas (b) illustrates two-way interaction between physical mechanisms and mechanisms in virtual machines, which interact both with one another and with the physical substrate and environment.

be influencing or referring to other patterns, or even to themselves. We depend on this in flight control systems. Concurrency of interactions is important, as we'll see, and does not match the common view of a computer as a Turing machine.

In Figure 1, (b) schematically indicates a physical system on which is implemented complex virtual machinery composed of enduring, interacting subsystems which can influence and be influenced by other patterns and physical structures and processes. The causal arrows go up and down as well as sideways. In modern computers, such enduring but changing virtual machinery co-exists with, and influences, underlying physical machinery, *which it helps to control,* even though the virtual machinery is all *fully implemented in deterministic physical machinery.* Rules running in a virtual machine can cause changes in the physical memory and attached devices (including screens, motors, speakers, and network links to computers). The use of "cause" there is justified because so much engineering has gone into making a complex collection of conditional statements true, including counter-factual conditionals about what would or would not have happened in various possible situations. A web of connected hardware, software, and firmware subsystems whose overall structure is very complex, and can even change rapidly between and during causal interactions, makes those conditional statements true, by ensuring that the connections are not mere coincidences. That includes not only causation between processes in virtual machinery, but also causation across levels, e.g. using mechanisms that ensure that a decision taken at a high level causes certain changes in physical memory, or causes appropriate signals to go to an attached device. All that uses very complex technology, unimaginable in Darwin's time. However, engineers can make mistakes, and bugs in the virtual machinery are detected and removed, usually by altering a textual specification that drives the creation of the web.

The technology supporting all that functionality includes (in no significant order): *memory management, paging, cacheing, interfaces of many kinds, interfacing protocols, protocol converters, device drivers, interrupt handlers, schedulers, privilege mechanisms, resource control mechanisms, file-management systems, interpreters, compilers, "run-time systems" for various programming languages, garbage collectors, varied types of data-structure and operations on them, debugging tools, pipes, sockets, shared memory systems, firewalls, virus checkers, security systems, network protocols, operating systems, application development systems, name-servers*, and more. Concurrency does not require multiple CPUs, since enduring process records in memory allow a paused process to continue to influence running processes through the latter's memory accesses. However, insofar as a computer has sensors and other interfaces connected with the environment there will be many concurrent processes not wholly under the control of the computer, interacting partly as a result of external interrupts.

Some people find it hard to see how virtual machines can cause anything to happen because they fail to make a three-way distinction, between:

1. *Mathematical Models* (MMs), e.g. numbers, sets, grammars, proofs, ...
2. *Physical Machines* (PMs), including atoms, voltages, chemical processes, ...
3. *Running Virtual Machines* (RVMs), e.g. calculations, games, formatting, proving, checking spelling,

handling email, self-monitoring, ...

MMs are static abstract structures, like proofs, and axiom systems that never do anything. Unfortunately some uses of "virtual machine" refer to MMs, e.g. "the Java virtual machine". These are abstract, inactive, mathematical entities, not RVMs, whereas PMs and RVMs are active and cause things to happen both internally and in external environments. Millions of computer users use RVMs every day, with little knowledge of what they are using. Different computer scientists and software engineers, know about different sub-sets, and the whole system is rarely described adequately. For instance, Pollock's mostly excellent (Pollock, 2008), over-simplifies by frequent references to "the machine table", ignoring the layered implementations used by most application software.

The technology summarised above allows physical machines on our desks to support varying collections of non-physical machinery made up of various kinds of concurrently interacting components whose causal powers operate in parallel with the causal powers of underlying machines, and help to control those physical machines. However, the non-physical (virtual) machinery has *different levels of granularity* and *different kinds of functionality* from the physical machines. The coarser granularity is indispensable for processes of design, testing, debugging, and also for run-time self-monitoring and control, which would be impossible to specify at the level of individual transistors (because of explosive combinatorics, especially with time-sharing). We need to understand the variety of uses of virtual machinery, compared with physical information-processing machinery, including the importance of the coarser granularity, in order to understand the evolutionary pressures that could have produced biological (e.g. brain-based) virtual machines.

4 What Exactly Needs to be Explained?

Thinkers are divided as to the kind of diversity of mental phenomena found in nature. Some believe that there is a major dichotomy separating conscious organisms and the rest. Others claim there are only differences of *degree*, with gradually increasing sophistication and complexity of mental phenomena emerging from gradually increasing complexity of the physical substrate. However, claims regarding continuity (in the mathematical sense) in biological evolution are implausible (a) because all biological phenomena are implemented in chemical mechanisms (e.g. using DNA), and there are only *discrete* ways of transforming one molecular structure to another, and (b) the fact that evolutionary changes occur only across generations implies that between any two evolutionary stages there can be only a finite number of intervening stages: which rules out continuous change. Biological changes must be discontinuous, whether small or large. So, instead of seeking a single major dichotomy between conscious organisms and the rest, or hoping to find continuous variation, we need to understand a large collection of discontinuous design changes, with both small and large differences in structure, behaviour and functionality.

Specifying a good conceptual framework for studying that variety is not easy. Evolution produced widely varying physical structures, and also myriad behavioural capabilities and internal information-processing capabilities, with different subsets of capabilities shared between different species. Even humans differ widely. E.g. new-born infants appear to be unable to perceive, think about, or communicate about most of the things older humans can, and some older humans are also limited in their mental capacities by effects of genetic abnormality, injury, disease, or degeneration. There are also differences in mental functioning that come from cultural influences – e.g. whether people can think in Chinese, or whether they can understand harmonic progressions in music.

Whittaker's review of Romanes (Whittaker, 1884) asks whether mind-like features are present in *all* living things, raising the possibility "that the lowest animals, ... have the beginnings not only of sensibility but also of will and intelligence."[2] His use of "will" implied causal powers. In more neutral language: all organisms are informed control systems. In deploying stored energy, they select between alternatives (e.g. both external and internal behaviours) on the basis of available information. Whittaker also made an important point about the structure of the space of possible minds: "The development of mind is represented

[2] No doubt the idea has occurred to many people. I was unaware of Whittaker's work when I presented a similar idea in "What Has Life Got To Do With Mind? Or vice versa?": http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#maggiefest

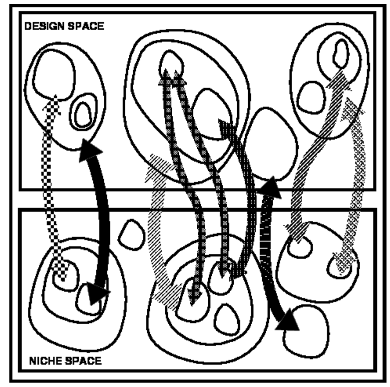

Figure 2: The space of possible sets of requirements (niches), the space of possible designs, and the complex structural relationships between designs and requirements. Relationships between requirements and designs are often represented as numerical fitness functions. But numerical fitness values cannot do justice to the diversity of ways of satisfying or failing to satisfy a set of requirements. Compare descriptive consumer reports on products.

as proceeding only in a single line. .... Nothing is said as to the possibility that at the same level of general intelligence there may be essentially different mental types, ...” (page 294). In a note in *Mind* 1984, Romanes claimed Whittaker had misread his meaning, showing that he too assumed non-linear mental diversity.

That diversity is important when we discuss the evolution of mentality, when we try to design and build working models of mentality and when we try to explain the relationships between matter and mind. If many different things evolved, using many different kinds of working mechanism, then that diversity must be explained by any satisfactory answer to the question of how physical phenomena can produce mental phenomena. The variations in design produced by evolution will be responses to variations in pressures, needs and opportunities. We can summarise that by saying there are many different sets of requirements (i.e. different niches) as well as different designs – satisfying different sets of requirements, as indicated crudely in Figure 2.

We have rejected (a) continuous variation, (b) the use of a dichotomy, and (c) a linear arrangement of types of mentality. What alternatives remain? If we analyse environments in depth, we can systematically develop different combinations of requirements that can arise, for instance, requirements related to the spatial separation of different sources of food and drink, requirements based on different kinds of fixed and changing features of the environment (including obstacles, passage-ways, etc.), different kinds of materials available that can be manipulated to meet various needs, different sorts of food (e.g. food that tries to escape and food that doesn’t) different sorts of predators, and different sorts of competitors for the same resources. These (and many other) differences in requirements[3] entail advantages and disadvantages in both physical design features, e.g. strength, speed of motion, camouflage, types of gripper, etc., and also designs for virtual machinery for processing information – e.g. *factual* information about the environment (including other agents), *control* information about what to do when, *meta-information* about information, and *meta-control information* about good and bad ways to acquire, process and use information. Long before human engineers found the need to develop virtual machinery, could biological evolution have responded to similar pressures, and produced solutions to those problems, which we do not yet understand? Seeking evidence will be difficult, but potentially enormously important.

Conceptual tools and engineering advances in the last half century have made a huge difference to our ability to think about these requirements and design options. But we have not yet developed a biologically adequate theory of types of virtual machinery. On the contrary, we are still a long way from that.[4] But we have made progress that was unimaginable by Darwin and his peers.

5 In Defense, and Explanation, of Qualia

Our task is to produce a new characterisation and explanation of the phenomena that led to views about contents of experience that (a) are private to the individual concerned, (b) have a character that is utterly

[3]http://www.cs.bham.ac.uk/research/projects/cogaff/misc/creativity-boden.html

[4]See `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/`

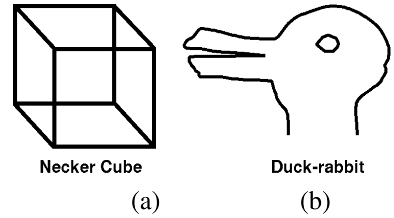

Figure 3: Each of the two figures is ambiguous and flips between two very different views. (a) can be seen as a 3-D wire frame cube. For most people it flips between two different views of the cube, in which the 3-D locations, orientations and other relationships vary. In (b), the flip involves changes in body parts, the facing direction, and likely motion – requiring a very different ontology.

distinct from physical structures and processes, (c) seem to be produced by physical and chemical processes in brains and environments perceived or acted on, yet seem to be of a kind that cannot be produced by physical processes.

Despite deep confusions about consciousness and *qualia*, noted by Ryle, Dennett and others, it is clear that such things exist. Although hard to characterise and to identify in other individuals and other species, we need to study examples, to determine requirements for explanatory mechanisms. Their existence and some of their diversity can be demonstrated using the examples in Figure 3. Stare at each of them for a few minutes.[5] When your interpretation of a picture flips, only changes in you occur, apparently involving states and processes that are private to you, and somehow directly accessible by you. Describing carefully how the content changes when the interpretations flip, reveals that very different vocabularies are required, using a purely geometric ontology for (a) and an ontology of animal parts and capabilities in (b). A cube cannot be experienced as "looking to left or to right", whereas a duck, or rabbit can.

These and other phenomena can be used to demonstrate that there are mental states and processes involving mental entities within us even though they cannot be observed by opening up the skull, using the most powerful physical and chemical measuring instruments. How can biological evolution produce states within an organism that are closely tied to sensory input but can change spontaneously and which have semantic content referring to possible external entities describable only using a complex ontology? Until recently the status of such entities was highly problematic, but I claim that we can now see, at least in outline, how to explain their occurrence: the entities with semantic content occur in a virtual machine that rapidly produces various layers of interpretation of the sensor input using different ontologies, as proposed in (Sloman, 1978, Ch 9).

We already know how to produce computing systems that can observe some of their own internal information processes by recording the existence and properties of abstract data-structures that occur, procedures that are followed, and difficulties and successes encountered. Very often the contents of such self-observations are not the *physical* states of the computer components but the components and contents of *virtual machines*. What now need to collect many examples of the types of qualia that might occur in a human-like robot and develop designs that could explain both the occurrence of those cases and their roles in information-processing systems of various kinds. E.g. one of many ways in which it can be useful to attend to your internal data-structures rather than the full structure of perceived entities, is that you can use your internal data-structures to drive a process of communication with others, perhaps using drawings and gestures, to show them what experiences they can expect if they visit certain terrain, for example. Some of the requirements for such mechanisms have been described in presentations on my web site[6], though most of them still leave many problems unsolved and much work to be done.

One of the important facts relating to the diversity of kinds of mind referred to by Whittaker is that not all organisms that have qualia know that they have them! We can separate the *occurrence* of mental contents in an organism from their *detection* by the organism, which requires additional architectural complexity to support self-observation and self-description mechanisms. Many organisms can (I suspect) create and use such entities without having the meta-semantic mechanisms required to detect and represent that fact, as humans do. It is very likely that the vast majority of organisms have very little self-observation capability, despite having conscious (but not self-conscious) contents, used to guide behaviour. We still need to understand why and how evolution produced those that are capable of introspection. For the others it can be said that they have and use potential contents of experience, but they do not experience them! However it is

[5]See http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#cons09
[6]http://www.cs.bham.ac.uk/research/projects/cogaff/talks/

not likely to be a binary divide but a collection of sub-divisions with more or less functionality, depending on the species. I expect we shall need to experiment with a range of increasingly complicated working examples, using different kinds of mechanism, in order to understand better some of the questions to be asked about about mental phenomena in biological organisms. This is very close to Arbib's research programme described in (Arbib, 2003).

6 What Next?

Long experience of philosophical debates shows that there are thinkers whose belief in an unbridgeable mind/body explanatory gap will be unshaken by all this. As argued in (Sloman, 2010c), some cases of opposition will be based on use of incoherent concepts (e.g. a concept of "phenomenal consciousness" *defined* to involve no causal or functional powers). One way to test these ideas, outlined in (Sloman, 2010a), is to use the ideas to specify a robot design that starts off with the cognitive capacities of a human infant and develops its own architecture in way that produces human-like growth in cognitive sophistication, including development of introspective capabilities that can be focused on experience of things like Figure 3, until it reaches the stage where it becomes puzzled as to how its own internal states and processes, detected at a virtual machine level, can exist in a physical world.

After reading about philosophy of mind, and having philosophical discussions with others, different individual robots that all start from the same design should be able to develop philosophical beliefs about the nature of the mind-body relationship. However, just as the same language learning capability in infant humans can lead to the use of very different languages, depending on the environment in which the infant grows up, so should the same starting design (robot genome) be able to produce "adult" robots whose philosophical views differ as widely as those of human philosophers, such as Renee Descartes, John Searle, Daniel Dennett and the author of this paper. This would not imply that philosophical beliefs about the nature of mind are merely cultural phenomena (like languages): some of them may be, while others are based both on personal (including robotic) experience, and deep scientific and technical knowledge, which other robots starting from the same initial design may never acquire.

If we, who have designed all the robots, can tell which one is right about how it works, and how its mental processes are related to physical processes in and around it, then that would demonstrate at least the possibility that humans with those theories are also right about how *they* work!

Alas: we are nowhere near being able to build such robots, since current achievements in AI vision, motor-control, concept-formation and other forms of learning, language understanding and use, motive-generation, decision-making, plan-formation, problem-solving, and many other areas are still (mostly) far inferior to those of humans. And if we omit the competences that appear unique to humans, current robots are still far inferior to other animals. No easy way to close those gaps is visible on the horizon. But there are many things to try.

6.0.1 Acknowledgements

A lecture at the University of Birmingham by Joe Cain on 13th October 2009 drew my attention to the extent and importance of scientific puzzlement regarding evolution of mental functions even among supporters of Darwin. I owe thanks to many colleagues, over many years, including Jeremy Wyatt and other members of the EU CogX Cognitive Robotics project.
.

References

Arbib, M. A. (2003). Rana computatrix to Human Language: Towards a Computational Neuroethology of Language Evolution. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences,*, *361*(1811), 2345–2379. (http://www.jstor.org/stable/3559127)

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge, UK: Cambridge University Press.

Block, N. (1995). On a confusion about the function of consciousness. *Behavioral and Brain Sciences*, *18*, 227–47.
Chalmers, D. J. (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219. (http://www.imprint.co.uk/chalmers.html)
Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. New York, Oxford: Oxford University Press.
Craik, K. (1943). *The nature of explanation*. London, New York: Cambridge University Press.
Dennett, D. C. (1991). *Consciousness explained*. London and New York: Penguin Press.
Dyson, G. B. (1997). *Darwin Among The Machines: The Evolution Of Global Intelligence*. Reading, MA: Addison-Wesley.
Gibson, E. J., & Pick, A. D. (2000). *An Ecological Approach to Perceptual Learning and Development*. New York: Oxford University Press.
Huxley, T. H. (1866). *Lessons in Elementary Physiology*. New York: MacMillan and Co.
Jablonka, E., & Lamb, M. J. (2005). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge MA: MIT Press.
Kim, J. (1998). *Mind in a Physical World*. Cambridge, MA: MIT Press.
McGinn, C. (2004). *Consciousness and Its Objects*. Oxford: Oxford University Press.
Nagel, T. (1981). What is it like to be a bat. In D. Hofstadter & D.C.Dennett (Eds.), *The mind's i: Fantasies and reflections on self and soul* (pp. 391–403). Penguin Books.
Penrose, R. (1989). *The Emperor's New Mind: Concerning Computers Minds and the Laws of Physics*. Oxford: Oxford University Press.
Pollock, J. L. (2008). What Am I? Virtual machines and the mind/body problem. *Philosophy and Phenomenological Research.*, *76*(2), 237–309. (http://philsci-archive.pitt.edu/archive/00003341)
Rochat, P. (2001). *The Infant's World*. Cambridge, MA: Harvard University Press.
Romanes, G. J. (1883). *Mental evolution in animals*. London: K. Paul, Trench. (http://www.archive.org/details/cihm_16907)
Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
Ryser, P. (2009, February-March). Creative Choice: How the Mind Could Causally Affect the Brain. *Journal of Consciousness Studies*, *16*, 6–29. (2-3)
Shanahan, M. (2005). Consciousness, Emotion, and Imagination: A Brain-Inspired Architecture for Cognitive Robotics. In *Proceedings AISB 2005 Symposium on Next Generation Approaches to Machine Consciousness* (pp. 26–35).
Sloman, A. (1978). *The computer revolution in philosophy*. Hassocks, Sussex: Harvester Press (and Humanities Press). (http://www.cs.bham.ac.uk/research/cogaff/crp)
Sloman, A. (2009a). What Cognitive Scientists Need to Know about Virtual Machines. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1210–1215). Austin, TX: Cognitive Science Society. (http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#901)
Sloman, A. (2009b). *Why the "hard" problem of consciousness is easy and the "easy" problem hard. (And how to make progress)*. University of Birmingham, School of Computer Science. (Online tutorial presentation: http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#cons09)
Sloman, A. (2010a). An Alternative to Working on Machine Consciousness. *Int. J. Of Machine Consciousness*. (http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#910)
Sloman, A. (2010b, August). How Virtual Machinery Can Bridge the "Explanatory Gap", In Natural and Artificial Systems. In S. D. et al. (Ed.), *Proceedings SAB 2010, LNAI 6226* (pp. 13–24). Heidelberg: Springer. (http://www.cs.bham.ac.uk/research/projects/cogaff/10.html#sab)
Sloman, A. (2010c). Phenomenal and Access Consciousness and the "Hard" Problem: A View from the Designer Stance. *Int. J. Of Machine Consciousness*. (http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#906)
Whittaker, T. (1884, April). Review of G.J.Romanes *Mental evolution in animals*. *Mind*, *9*(34), 291–295. (available at http://mind.oxfordjournals.org)

Wiener, N. (1961). *Cybernetics: or control and communication in the animal and the machine*. Cambridge, MA: The MIT Press. (2nd edition)