Virtual Machinery and Evolution of Mind (Part 2)

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

Abstract

The ideas about virtual machinery presented in Part 1 suggest ways in which biological evolution may have taken advantage of virtual machines to produce self-monitoring, self-modifying, self-extending information-processing architectures, some of whose contents would have the defining features of qualia. This could provide a way for Darwin to answer the criticism that natural selection can produce only physical development, not mental states and consciousness. 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.

Key words: Architecture, Causation, Cognition, Consciousness, Darwin, Designer Stance, Evolution, Explanatory Gap, Mind, Virtual Machinery

1. Introduction

Darwin's critics, cited in Sloman (2010a), argued that his evidence supported only the hypothesis that natural selection produces *physical* forms and behaviours. Nobody could understand how physical mechanisms can produce mysterious and externally unobservable mental states and processes: "The explanatory gap". Since Darwin's time the problem has been re-invented and re-labelled several times, e.g. as the problem of "Phenomenal Consciousness" Block (1995) or the "Hard Problem" of consciousness Chalmers (1996). The topic was touched on and side-stepped in Turing (1950). It remains unclear how a genome can, as a result of physical and chemical processes, produce the problematic, apparently non-physical, externally unobservable, personal experiences (qualia) and processes of thinking, feeling and wanting.

Part 1 presented Universal Turing Machines as theoretical precursors of technology supporting networks of interacting running virtual machines

(RVMs) sensing and controlling things in their environment. Such RVMs are *fully implemented* in underlying physical machines (PMs) but the concepts used to describe the states and processes in some RVMs (e.g. "pawn" and "threat" in chess VMs) are not *definable* in the language of the physical sciences. We now develop the biological application of these ideas, explaining how self-monitoring, self-modifying RVMs can include some of the features of consciousness, such as qualia, previously thought to be mysterious, paving the way for a theory of how mind and consciousness might have evolved.

2. Epigenesis: bodies, behaviours, and minds

Turing was interested in both evolution and epigenesis and made some pioneering suggestions regarding the processes of morphogenesis – differentiation of cells to form diverse body parts during development. As far as I know, he did not do any work on how a genome can produce *behavioural competences* of the complete organism, including behaviours with complex conditional structures so that what is done depends on internal and external sensory information, though he briefly considered learning, in Turing (1950)[1].

It is understandable that physical behaviours, such as hunting, eating, escaping predators, and mating, should influence biological fitness and that evolution should select brain and other modifications that produce advantageous behaviours. But there are *internal* non-behavioural competences whose biological uses are not so obvious: thinking, reminiscing, perceiving with enjoyment, finding something puzzling and attempting to understand it. It is not obvious how biological evolution could produce mechanisms that are able to support such mental processes.

Many species develop behavioural and internal competences that depend on the environment during development (e.g. which language a child speaks, and which mathematical problems are understood), so the genome-driven processes must create some innately specified competences partly under the influence of the genome and partly under the influence of combinations of sensorimotor signals during development (Held and Hein (1963); McCarthy (2008)). For humans at least, the internal processes of competence-formation though brain modification must go on long after birth, suggesting that the genome continues producing, or enabling, or constraining effects (including

[1]See Sloman, chapter xxx this collection for a criticism of his suggestion about learning.

changes in sexual and parental motivations and behaviours) long after the main body morphology and sensory-motor mechanisms have developed.

Karmiloff-Smith (1992) presents many examples where *after* achieving behavioural competence in some domain, learners (including some non-human species) re-organise their understanding of the domain in such a way as to give them new abilities to think and communicate about the domain. After children develop linguistic competences based on known phrases they spontaneously switch to using a *generative* syntax that allows *derivation* of solutions to novel problems, instead of having to learn empirically what does and does not work. Craik (1943) pointed out the value of such mechanisms in 1943, suggesting that they could be based on working mental models.[2] Grush (2004) and others suggest that such models could work as simulations or emulations. However, when used for reasoning purposes, as opposed to statistical prediction, a decomposable information structure is required, for instance when proving geometrical theorems (Sloman (1971)).

The mental models we use to explain and predict, include things like gear wheels, bicycles, electric circuits and other mechanisms that are too new to have been part of our evolutionary history. So, at least in humans, the model construction process cannot all be encoded in the genome: the specific models need information obtained after birth from the environment, and, in the case of creative inventors, ideas thought up by the individual.

So, the genome specifies not only physical morphology and physical behavioural competences, but also a multi-functional information-processing architecture developed partly in species-specific ways, over an extended time period, partly under the control of features of the environment, and includes not only mechanisms for interpreting sensory information and mechanisms for controlling external movements, but also mechanisms for building and running predictive and explanatory models of structures and processes, either found in the environment or invented by the individual[3]. How can a genome specify ongoing construction processes to achieve that functionality?

[2] I have not been able to find out whether Craik and Turing ever interacted. Turing must have known about his work, since he was a member of the Ratio club, founded in honour of Craik, shortly after he died in a road accident in 1945.

[3] It is argued in Sloman (1979, 2008) that this requires types of "language" (in a generalised sense of the word, including structural variability and compositional semantics) that evolved, and in young humans develop, initially for *internal* information-processing, not for external communication. We can call these "generalised languages" (GLs).

I don't think anyone is close to an answer, but I'll offer a conjecture: evolution discovered the virtues of virtual machinery long before human engineers.

Part 1 (sections 3 and 4) outlined the benefits of virtual machinery in human-designed computing systems and their advantages compared with specifying, designing, monitoring, controlling and debugging the physical machinery directly, because of the coarser granularity and the use of application-relevant semantics. Perhaps biological evolution also found the use of virtual machinery in animals advantageous for specifying types of competence at a relatively abstract level, avoiding the horrendous complexity of specifying all the physical and chemical details. The initial specification of behavioural competences in the genome might be far more compact and simpler to construct or evolve if a virtual machine specification is used, provided that other mechanisms ensure that that "high level language" is mapped onto physical machinery in an appropriate way. The use of self-monitoring processes required for learning and modifying competences, including de-bugging them, may be totally intractable if the operations of atoms, molecules or even individual neurones are monitored and modified, but more tractable if the monitoring happens at the level of a RVM.

So something like a compiler is required for the basic epigenetic processes creating common features across a design, and something more like an interpreter to drive subsequent processes of learning and development.

3. The evolution of organisms with qualia

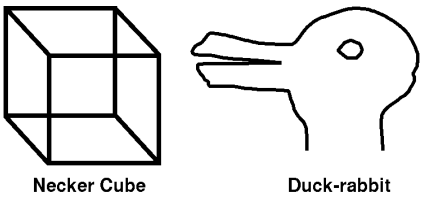
Part 1 showed that virtual machinery can be implemented in physical machinery, and events in virtual machines can be causally connected with other VM events and also with physical events both within the supporting machine and in the environment, as a result of use of complex mixtures of technology for creating and maintaining virtual/physical causal relationships developed over the last seven decades. Some of the events and processes in virtual machines are not identical with the underlying physical machinery and their description requires an ontology that is not definable in terms of the ontology of physics. The use of virtual machinery enormously simplifies the design, debugging, maintenance, and development of complex systems. Finally, and perhaps most importantly, in machines that need to monitor and modify *their own* operations, performing the monitoring and modifications at the level of virtual machinery can be tractable where the corresponding

tasks would he intractably complex and too inflexible if done by monitoring and modifying physical machinery.

So, biological evolution could have gained in power, flexibility, and speed of development by using virtual machine descriptions in the genome for specifying behavioural competences, instead of descriptions of the physical details. Moreover if some of the virtual machinery is not fully specified in the genome, and has to be developed after birth or hatching by making use of new information gained by the individual from the environment, then that post-natal construction process will be much simpler to specify, control and modulate if done at the virtual machine level rather than specifying all the chemical and neuronal changes required. And finally self-monitoring, self-control, and self-modification in a sophisticated information-processing system needs to control virtual not physical, machinery.

For an intelligent organism perceiving, thinking about and acting on a rich and complex environment that contains enduring objects and processes at various locations not all constantly in perceptual range, it will be useful to store information about the environment using one or more appropriate virtual machines. Visual and haptic processes perceiving the same portion of the environment could include overlapping virtual machines dealing with different aspects of the environment processed at different levels of abstraction in parallel Sloman (2009). Data-structures representing visible portions and features of the environment, e.g. visible portions of surfaces with colour, shape, orientation, curvature, speeds of motion or rotation, and relationships to other surface fragments (i.e. not the specific sensory signals), will then be components of virtual machines. If the information structures created during visual perception, are sometimes accessed by self-monitoring processes that attend not to what is in the environment, but to the content of what is currently being perceived, then we potentially have an explanation of the phenomena that have led to philosophical and other puzzles about the existence and nature of sensory qualia, which are often regarded as defining the most difficult aspect of mind to explain in functional terms, and whose evolution and development in organisms Huxley and others found so difficult to explain. See also Sloman and Chrisley (2003).

To illustrate this point: when ambiguous figures, e.g. in Figure 1, are experienced as switching from one view to another, that will involve a change in the contents of some virtual machinery, and those contents will be represented at a virtual machine level, referring to different perceptual contents, including distance, direction of slope, body-parts, direction faced, etc.



(a) (b)

Figure 1: 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.

Ryle, Dennett and others, identified deep confusions in talk about consciousness and *qualia*, but such things clearly exist, though they are hard to characterise and to identify in other individuals and other species. Analysis of examples, including ambiguous figures, such as Figure 1, helps to determine requirements for explanatory mechanisms. Such pictures illustrate the *intentionality* of perceptual experience, i.e. interpreting something as referring to something else and the different *ontologies* used by different experiences. I suggest that that is only possible within running virtual machinery, since concepts like "interpreting", "referring", "intending" and "looking", are no more definable in the language of physics than "pawn" or "threat".

Many organisms can, I suspect, create and use such virtual entities without having the meta-semantic mechanisms required to detect and represent the fact that they do. One of the important facts relating to the diversity of kinds of mind, referred to in Whittaker (1884), 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. 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 mental phenomena in biological organisms. This is very close to Arbib's research programme described in Arbib (2003).

4. What Next?

Experience shows that for many thinkers belief in an unbridgeable mind/body explanatory gap will be unshaken by all this. As argued in Sloman (2010b), 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). Working systems that show how different robot designs correspond to different products of evolution may help. But current achievements in AI vision, motor-control, concept-formation, forms of learning, language

understanding and use, motive-generation, decision-making, plan-formation, problem-solving, and many others, are still (mostly) far inferior to those of humans and other animals, in part because designers typically consider only a small subset of the requirements for biological intelligence. Even if we omit uniquely human competences, current robots are still far inferior to other animals. There is no easy way to close those gaps, but there are many things to try, as long as we think clearly about what needs to be explained. Turing could have made a substantial contribution to this 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.

Block, N., 1995. On a confusion about the function of consciousness. Behavioral and Brain Sciences 18, 227–47.

Chalmers, D. J., 1996. The Conscious Mind: In Search of a Fundamental Theory. Oxford University Press, New York, Oxford.

Craik, K., 1943. The Nature of Explanation. Cambridge University Press, London, New York.

Grush, R., 2004. The emulation theory of representation: Motor control, imagery, and perception. Behavioral and Brain Sciences 27, 377–442.

Held, R., Hein, A., 1963. Movement-produced stimulation in the development of visually guided behaviour. J. of Comparative and Physiological Psychology 56 (5), 872–876.

Karmiloff-Smith, A., 1992. Beyond Modularity: A Developmental Perspective on Cognitive Science. MIT Press, Cambridge, MA.

McCarthy, J., 2008. The well-designed child. Artificial Intelligence 172 (18), 2003–2014. `http://www-formal.stanford.edu/jmc/child.html`

Sloman, A., 1971. Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence. In: Proc 2nd IJCAI.

William Kaufmann, London, pp. 209–226,
http://www.cs.bham.ac.uk/research/cogaff/04.html#200407.

Sloman, A., 1979. The primacy of non-communicative language. In: MacCafferty, M., Gray, K. (Eds.), The analysis of Meaning: Informatics 5 Proceedings ASLIB/BCS Conference, Oxford, March 1979. Aslib, London, pp. 1–15,
http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#43.

Sloman, A., 2008. Evolution of minds and languages. What evolved first and develops first in children: Languages for communicating, or languages for thinking (Generalised Languages: GLs)?
`http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0702`

Sloman, A., 2009. Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress. In: Sendhoff, B., Koerner, E., Sporns, O., Ritter, H., Doya, K. (Eds.), Creating Brain-like Intelligence. Springer-Verlag, Berlin, pp. 248–277.
`http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0804`

Sloman, A., August 2010a. How Virtual Machinery Can Bridge the "Explanatory Gap", In Natural and Artificial Systems. In: Doncieux, S., et al. (Eds.), Proceedings SAB 2010, LNAI 6226. Springer, Heidelberg, pp. 13–24.
`http://www.cs.bham.ac.uk/research/projects/cogaff/10.html#sab`

Sloman, A., 2010b. Phenomenal and Access Consciousness and the "Hard" Problem: A View from the Designer Stance. Int. J. Of Machine Consciousness 2 (1), 117–169.
`http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#906`

Sloman, A., Chrisley, R., 2003. Virtual machines and consciousness. Journal of Consciousness Studies 10 (4-5), 113–172.
`http://www.cs.bham.ac.uk/research/projects/cogaff/03.html#200302`

Turing, A., 1950. Computing machinery and intelligence. Mind 59, 433–460, (reprinted in E.A. Feigenbaum and J. Feldman (eds) *Computers and Thought* McGraw-Hill, New York, 1963, 11–35).

Whittaker, T., April 1884. Review of G.J.Romanes *Mental evolution in animals*. Mind 9 (34), 291–295.