[Extended Abstract for WPE 2008]

Virtual Machines in Philosophy, Engineering & Biology

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

Keywords
Architecture, causation, implementation, information-processing, biology, philosophy, psychology, robots, self-awareness, self-control, supervenience, vertical modularity, virtual machine, virtual machine functionalism .

1.INTRODUCTION
A machine is a complex enduring entity with parts that interact causally with one another as they change their properties and relationships. Most machines are also embedded in a complex environment with which they interact. A *virtual machine* (VM) has non-physical parts, relationships, events and processes, such as parse trees, pattern matching, moves in a game, goals, plans, decisions, predictions, explanations and proofs.

The concept of a virtual machine, invented in the 20th Century, (not to be confused with *virtual reality*) is important (a) for many engineering applications, (b) for theoretical computer science, (c) for understanding some of the major products of biological evolution (e.g. animal minds), and (d) for gaining new insights into several old philosophical problems, e.g. about the mind-body relationship, about qualia, and how to analyse concepts of mind by adopting the design stance in combination with the notion of an information processing *architecture* [1,2]. Analysing relations between different sets of requirements (niches) and designs for meeting the requirements exposes a space of *possible minds* (for animals and artifacts), raising new questions about evolution, about future intelligent machines, and about how concepts of mind should be understood.

Most philosophers, biologists, psychologists and neuro-scientists completely ignore VMs, despite frequently (unwittingly) using them: e.g. for email, spreadsheets, text processing, or web-browsing. Academic philosophers generally ignore or misunderstand the philosophical significance of VMs (in part because many assume VMs are finite state machines). Pollock [3] is a rare exception. Dennett often mentions virtual machines, but claims they are merely a useful fiction [e.g. 4, note 10]. Events in useful fictions cannot cause email to be sent or airliners to crash. The idea of a VM can significantly extend our thinking about problems in several disciplines and pose new problems for future empirical and philosophical research.

2.WHAT ARE VIRTUAL MACHINES?
The idea of a VM had (at least) four sources (a) the demonstrations of universality of certain sorts of machine (e.g. a Universal Turing Machine can implement many other machines as virtual machines), (b) engineering problems related to sharing scarce resources between different processes running on one computer, (c) problems of portability and modularity of code for software systems, and (d) the design of layers of functionality for transmission networks. The common idea is that structures and processes can exist and interact in ways that require physical implementation, where the precise details of the physical implementation can vary from time to time across machines and even within one machine. Often VMs are layered, with VM^1 implemented in VM^2, implemented in VM^3, etc. The existence of causal interactions among VM events and between VM events and physical events (e.g. events in a word processor and events on a computer screen) challenges many (all?) philosophical analyses of supervenience and of causation, but the latter is a topic for another occasion.

Many issues discussed by philosophers (e.g. issues about how mental concepts work and about relations between mind and body, such as supervenience) require adoption of the design stance, using the notion of a VM in which enduring concurrent non-physical (but physically implemented) sub-processes interact with one another and with physical entities. Compare: analyses of concepts like 'iron', 'carbon', 'water', 'rust', 'acidic', 'burning' are much better done using a good theory of the architecture of matter than simply using pre-scientific ideas.

"Virtual Machine Functionalism" (VMF) denotes a type of functionalism that refers to virtual machines that contain many concurrent interacting processes, discrete and continuous, synchronised or asynchronous -- unlike conventional Functionalism, usually explained in terms of a simple finite state machine. See [1,2] and my 'talks' website for more details.

3.SELF MONITORING AND CONTROL
A VM provides a level of abstraction that avoids the need for a designer/maintainer to represent and reason about the vast complexities of the underlying physical mechanisms (molecular, electronic or neural). The same features make VMs important for complex systems that monitor and control themselves: they share some requirements with their designers!

This design strategy works only if: there is a *good* (e.g. reliable, robust, flexible) implementation for the VM, and the VM includes mechanisms enabling relevant states and processes to be sensed and modulated (e.g. blocking email from particular addresses). Identifying requirements for good virtual machines in biological organisms. future robots, and complex control systems (e.g. chemical plants) is a multidisciplinary task for philosophers, engineers (including roboticists), biologists and psychologists.

One requirement is that for organisms reproducing in unpredictably changing environments, some virtual machines need to grow themselves partly under the influence of the environment, rather than being fully specified genetically – see [5]. That's how 3-year olds can play computer games: something none of their ancestors ever did at that age.

Growth of an architecture is different from learning in a fixed architecture with a uniform learning mechanism. Some new mathematics may be required to specify such processes.

4.BIOLOGICAL VIRTUAL MACHINES
Conjecture: Biological evolution 'discovered' the importance of virtual machines long before humans did, and produced many kinds of virtual machine that we have not yet identified or understood.

In doing that, evolution may well have solved far more design problems (=engineering problems) than we have so far identified. Examples we already know about include homeostatic systems, immune systems, perceptual systems, learning systems, many kinds of monitoring, control and repair systems, and social systems. Much work still remains to be done finding out what the problems were, i.e. what the requirements were against which the designs were evaluated (e.g. by natural selection mechanisms), and what solutions were found. A better understanding of the *requirements* may help to direct more fruitful research into the designs and mechanisms.

This can be contrasted with current biologically inspired AI/Robotic research (and some neuropsychology) which often attempts to model supposed mechanisms without finding out what problems biological designs actually solved.

In [6] McCarthy discusses conjectures about the problems evolution solved in producing humans, some of which will also be problems for intelligent machines.

5.LIMITATIONS OF SUCH SYSTEMS

A consequence of the use of virtual machines, important for philosophy and psychology, is that self-monitoring systems that use the design features described above gain practical benefits (from 'vertical' modularity and reduced complexity of control and monitoring). The price is inherently limited self-knowledge and self-control, since implementation details are inaccessible.

These limitations may not matter in most normal conditions (if the design is good) but things can go badly wrong in abnormal conditions.

This sheds new light on philosophical discussions of qualia, their ineffability, their causal powers, the alleged impossibility of being mistaken about them, the nature and limits of introspection, free will, etc. It can also shed light on some possible types of mental/cognitive dysfunction caused by injury, disease, genetic abnormalities or even abuse. In particular it becomes important to distinguish problems with physical causes from problems that exist at the VM level (like software, as opposed to hardware, bugs in machines). This can be very difficult to do. Some genetic abnormalities produce a tangled mixture of hardware (wetware) and VM dysfunctions.

6.VMS FOR INTELLIGENT MACHINES

There are also engineering implications: if use of VMs is needed for sophisticated autonomous machines that monitor and control themselves, and which need to be able to adapt to and cope intelligently with unforeseen situations, and reach practical decisions in reasonable times, then they will have some of the failings that we find in biological systems with such designs (e.g. humans). See http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk51This raises ethical issues that I shall not discuss now, but designers will need to.

7.CONCEPTUAL IMPLICATIONS

We need to understand how VM architectures vary. Concepts that are appropriate for describing such complex systems are different for different virtual machine architectures. E.g. a computer operating system VM that never allows time-sharing or paging can never get into the state described as "thrashing" on a multi-processing system. Similarly an architecture that does not support formation and use of predictions would be incapable of getting into a state of being surprised. (It is very likely that the vast majority of animals are incapable of being surprised, despite apparent 'surprise behaviour' – often an evolved automatic reaction to sudden danger, etc.)

So, philosophers interested in analysing mental concepts need to learn to do new kinds of architecture-informed conceptual analysis, both

(a) to explicate and improve on our existing concepts of mind (e.g. believes, desires, intends, likes, imagines, expects, learns, understands, values, enjoys, dislikes, fears, cares, honest, delusion, self-deception, personality, multiple personality, etc. etc.), and

(b) to work out which sorts of mentalist concepts are relevant to future machines (most of which will, at least in the short run, have far less complex VMs than humans do, which means that the set of concepts that can aptly be used to describe them will be different in important ways – contrary, for instance, to the assumptions of current researchers claiming to build "machines with emotions").

This requires us to extend Ryle's notion of 'logical geography' with a deeper notion of a 'logical topography' that can support different logical geographies, as explained more fully in http://www.cs.bham.ac.uk/research/projects/cogaff/misc/logical-geography.html

8.CONFUSIONS ABOUT EMBODIMENT

The recent emphasis on embodiment in AI, Cognitive Science and Philosophy of mind has mostly involved a failure to understand how the physical morphology and sensorimotor interfaces of an information processing system relate to the variety of virtual machine layers that may coexist in one system, where some layers are far less constrained by the details of their embodiment than by complex features of the whole environment in which they are embedded and which they need to interact with, think about and understand.

That is why seriously physically disabled humans can, with appropriate help, learn to think and communicate like most humans, despite missing limbs, cerebral palsy, blindness, deafness, etc. which seriously limit their physical interactions with the immediate environment. (Examples include: Alison Lapper, Helen Keller, Stephen Hawking, grown up thalidomide babies, etc. Gender differences are not relevant to this point.)

Consequently, machines (robots) with very different physical forms and physical capabilities can, in principle, if their virtual machines are appropriate, share a great many forms of representation, concepts, concerns, values, thoughts, beliefs, hopes, fears, etc. with humans -- and be capable of communicating with them, despite great physical differences.

But before we have any hope of producing such machines, we need a far deeper understanding of (1) the problems evolution solved (the requirements for biological VMs), (2) the design options for solving those problems and the tradeoffs between the options. Philosophers will need to learn to think about tradeoffs and designs as engineers do, and engineers will need to learn to do conceptual analysis in order both to clarify their objectives and to avoid misdescribing what they have achieved, thereby invoking the scorn of McDermott [7]. Self-aware machines will need to use VMs to understand themselves.

9.REFERENCES

[1] Sloman, A. 2002, Architecture-based conceptions of mind. In P. Gardenfors et. Eds., *In the Scope of Logic, Methodology, and Philosophy of Science.* *http://www.cs.bham.ac.uk/research/projects/cogaff/00-02.html#57*

[2] Sloman A., Chrisley, R.L. 2003, Virtual machines and consciousness, *Journal of Consciousness Studies*, http://www.cs.bham.ac.uk/research/projects/cogaff/03.html#200302

[3] 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

[4] Dennett, D.C. 2007, Heterophenomenology reconsidered, *Phenomenology and the Cognitive Sciences*, 6, 1-2, 247—270 DOI 10.1007/s11097-006-9044-9

[5] Sloman & Chappell various papers and presentations in http://www.cs.bham.ac.uk/research/projects/cosy/papers/

[6] McCarthy, J. 1996, The Well Designed Child, Stanford University, http://www-formal.stanford.edu/jmc/

[7] McDermott, D. 1981, Artificial Intelligence meets natural stupidity, in *Mind Design,* Ed. J. Haugeland, MIT Press.