Avant. The Journal of the Philosophical-Interdisciplinary Vanguard
Volume II, Number 2/2011 www.avant.edu.pl
ISSN: 2082-6710

Evolutionist of intelligence

Introduction

Marcin Miłkowski

It would be indeed hard to find a more fervent advocate of the position that computers are of profound significance to philosophy than Aaron Sloman. His 1978 book bore the title *Computer Revolution in Philosophy* (Sloman 1978). He stressed the meaning of computing for understanding cognition:

> *it can change our thinking about ourselves: giving us new models, metaphors, and other thinking tools to aid our efforts to fathom the mysteries of the human mind and heart. The new discipline of Artificial Intelligence is the branch of computing most directly concerned with this revolution. By giving us new, deeper, insights into some of our inner processes, it changes our thinking about ourselves. It therefore changes some of our inner processes, and so changes what we are, like all social, technological and intellectual revolutions.*
>
> (Sloman 1978)

Yet, Sloman is not a stereotypical proponent of AI. Far from it; in his writings, he undermines several popular convictions of functionalists. He stresses that the Universal Turing Machine (UTM) is not really significant for modeling cognition. Real machines are different from abstract ones, and causal complexity of real computers is not reflected in purely abstract structures. A model of cognition based on the UTM is confined to standard digital computations – while physically, if there are random processes intervening, even two asynchronous TMs can compute Turing-uncomputable functions. Moreover, he is not using standard functionalist arguments, like arguments from multiple realizability.

Nonetheless, Sloman's work is far from the standard AI critics: he does not follow Searle in his insistence that computers cannot have real intentionality, and even goes as far as to say that the whole project of solving the symbol grounding problem is ill-conceived. He is not also very impressed, to put it mildly, with theories that deny the role of representation in cognitive systems, and criticizes the radical version of enactivism that turns cognizers into nothing more than complex insects.

Most (if not all) of Sloman's work is available on his website, with numerous presentations, tutorials, papers, and the 1978 book. As can easily be seen, he is more engaged in discussions than in preparing book-length manuscripts; and that makes a slight problem for people that want to cite something else than a draft on a website. Anyway, through his drafts and sometimes very lively polemics, Sloman definitely exerts quite substantial influence on the philosophy of AI.

During the CLMPS 2011 in Nancy, I had an occasion to hear the talk *Evolution of mind as a feat of computer systems engineering...* on which Sloman's paper is based. At first, it seemed very much right, but I could not really agree with some points, as my own conception of implementation of computation makes different assumptions about causality and uses causality as the basis for computational explanations (see Miłkowski forthcoming). Sloman's paper presents a bold hypothesis that the evolution of the human mind actually involved the development of a several dozen of virtual machines that support various forms of self-monitoring. This, in turn, helps explain different features of our cognitive functioning. In passing, he makes multiple points that show that current analytical philosophy does not recognize the complexity of information-processing systems. For example, the standard notion of supervenience seems to be based on heavily simplified cases, as well as naïve notions of causality. I could not agree more. In what follows, I will only focus on what I find problematic, as the paper speaks for itself, and is too rich to comment on in detail. These are quite technical points but I think they were not discussed sufficiently.

I agree that running virtual machines add a really important level of complexity to computers, though I am not so sure as Sloman is that virtual machines are really what is involved in self-monitoring activities. Clearly, the notion of the virtual machine has been seen as important for cognitive science for some time, and Dan Dennett stressed that the stream consciousness might be something like a process on a virtual machine. There are, however, important objections to such an idea:

> *there is nothing specific about VM for the purpose of this analogy [between VM and consciousness – MM], nor anything that makes it more appealing than any other form of software execution as a mental model. It is not plausible to imagine that a machine can be 'thinking within its own VM' any more than with any other software execution, since a VM is no less grounded in machine states than any other process when it is actually implemented …*

> *Whatever metaphorical benefit a VM conveys comes from discussion of software independent of its hardware.*
>
> (Wilks 1992: 263)

It is far from clear for me if this can be easily answered: the notion of the VM in Sloman's use is also a layer of software execution.

Though he stresses the causal complexity due to multiple layers of VMs, which is obviously right, there are several problems with some claims about it. The first problem is that Sloman claims that running VMs exert a non-physical causal influence, which might seem a very radical point. Yet, on closer reading, "non-physical" is just "non definable in physical terms", and the inability of defining the terms in which VM is naturally described is of the same kind that has been traditionally associated with multiple realizability claims. In other words, with VM, Sloman tries to buy theoretical autonomy of cognitive science from lower, physical levels of causality. This might sound nice to non-reductionist ears but is much harder to defend today than in the days of classical functionalism. First, the classical view on reduction as based on logical derivation of theories is based on a proposition-like view on theories, which is no longer treated as sacrosanct in philosophy of science. Second, another way of looking at reduction, namely via mechanisms, seems to be much more prevalent in real science (see Bechtel & Richardson 1993).

Mechanistic explanation, or explanation of the functioning of the whole systems with the causal organization of their parts, relies on causal connections of the same kind, and has no use for definability of terms in the language of physics. Importantly, what it yields are reductionist explanations. So the whole project of defending the autonomy with non-reducibility in the traditional, Nagel-like sense, might be misguided. Reduction by derivation is rare, and from the statistical point of view, it might as well be an outlier; whereas the mechanistic explanation is the everyday activity in neuroscience.

But it's possible that my interpretation makes the claim about non-physical causality too trivial. Another way to understand it is that there is a special kind of causality that relies on information. Sloman stresses that "changes in virtual machines occur they need not all be changes in measurable quantities", and adds: "that's because the processes can include things like construction, transmission, and analysis of complex structured entities". Apparently, "variations in running virtual machines are not all quantitative", and therefore "the causal relations cannot be expressed in algebraic formulae." As a consequence, he claims, such causal relationships are not measurable but only describable. Now, the problem is that sufficiently complex accounts of causality may easily deal with this and make causal claims testable via measurements (though, arguably, not reducible to measurements only). For example, the interventionist conception uses Bayes nets to model causal relationships (Pearl 2000; Spirtes, Glymour & Scheines 2001). Bayes nets can easily be used also to model virtual machines, if you only use sufficient-

ly expressive formalism, like Abstract State Machines (Gurevich 1995). If you don't like state-transition-based formalisms, you could go for abstract string-rewriting, but string-rewriting seems to be as easy modeled on graph-like structures as ASMs. So it is not clear to me if the point is to say that there is a new notion of causality or a complaint against a one-sided, simplified account of it.

Another point that Sloman makes in passing is that symbol grounding problem is over-rated and that the real problem behind it had been actually solved by Carnap years ago. This is one of the favorite claims that he has been making for years, and nobody really replied to them. But Sloman's idea cannot work. Let me elaborate. The symbol grounding problem is how to make symbols in computer systems representational without recourse to any external observers. The “solutions” offered by most authors, as Sloman rightly observers, are simply versions of naïve concept empiricism. This will not work, as concept empiricism is implausible after Kant's critiques. This much is true. What Sloman offers, however, is not a big improvement on that. Instead of grounding, we only need “tethering”, as he calls it. The symbols need only represent in virtue of structural resemblance, and even though multiple things can be said to resemble the same system of symbols in the same degree, only some of them are in standard models. These are the models that are “tethered” via “bridging principles” that do not fully determine the mapping between the system of symbols but only partly reduce the indeterminacy of the meaning of symbols. The problem is that Carnapian “meaning postulates” or “bridging principles” are not really a received solution to the problem of theory meaning in philosophy of science, contrary to what Sloman seems to suggest. Though they are a version of structuralism, which is still in fashion, they rely on the assumptions that cannot make tethering a viable candidate for a theory of representing. The problem is that the bridging principles are principles that relate theoretical terms to observational terms. Now, the observational terms are taken to be meaningful as such, and this is exactly the same assumption of concept empiricism that Sloman does not like. After all, you cannot have your Carnap cake without eating your logical empiricism.

Without observational terms that are taken to be representational by themselves, tethering will not work; but these terms are not tethered to anything by themselves. For the system that contains the symbols, the Carnapian observational terms would not be meaningful at all. They would be just another set of symbols. Unless these terms are meaningful *for* the system, they are just observer-relative, and the representing relationship relies on the knowledge of the observer, and not on the structure of the system that uses the symbols. In other words, how does the system know what the observational terms used in bridging principles *mean?*

What Sloman offers as a solution is therefore no solution at all. It is still another version of a mapping theory of representation: representation is just a matter of mapping. Most philosophical discussions indeed reduce representation to mappings, or encodings, which are derived from some covaration or resemblance relations. As plausible as this is for external representation, as a model for mental representation it cannot work. En-

codiginism, as Mark Bickhard plausibly shows (Bickhard & Terveen 1995), is not a viable theory of representation. Only if the symbol is the representation *for* the system, that is when it plays a role in its functioning as a representation, the impasse is broken. There are further conditions that need to be added, like the ability to misrepresent (which underlies intensionality with an "s"), and the ability of the system to recognize misrepresentation as such. No amount of information-processing or other causal relations will make a symbol into a full-blooded representation when the symbol is not playing a role of representation in the system, rather in the interpretation of the observer. This should be clear to Sloman, who stressed so many times that real computers are causal systems with complex organization whose behavior is not reducible to purely formal modeling, and opposed various observer-relative accounts of computation. The proper theory of representation must rely on this complex organization and causal dynamics of the system in the environment rather than on the external observers.

References:

Bechtel, W., and R.C. Richardson. 1993. *Discovering complexity: Decomposition and localization as strategies in scientific research*. *Discovery*. Princeton: Princeton University Press.

Bickhard, M.H., and L. Terveen. 1995. *Foundational issues in artificial intelligence and cognitive science: Impasse and solution*. North-Holland.

Gurevich, Y. 1995. Evolving algebras 1993: Lipari guide. In *Specification and Validation Methods*, ed. E. Börger, 231-243. Oxford: Oxford University Press.

Miłkowski, M. Forthcoming. *Explaining the Computational Mind.* Cambridge, MA: MIT Press / Bradford Book.

Pearl, Judea. 2000. *Causality: models, reasoning, and inference*. Cambridge: Cambridge University Press.

Sloman, A. 1978. *The Computer Revolution in Philosophy: Philosophy of Science and Models of Mind.* The Harvester Press. Available online at: http://www.cs.bham.ac.uk/research/projects/cogaff/crp/

Spirtes, P., C.N. Glymour, and R. Scheines. 2001. *Causation, prediction, and search*. *Search*. The MIT Press.

Wilks, Y. 1992. Dennett and Artificial Intelligence: On the same side, and if so, of what? in A. Brooks, D. Ross, *Daniel Dennett*, Cambridge: Cambridge University Press.

Web page: http://www.cs.bham.ac.uk/~axs/