CONSCIOUSNESS AND CREATIVITY

J MARK BISHOP
GOLDSMITHS, UNIVERSITY OF LONDON

It is a commonly held view that "*there is a crucial barrier between computer models of minds and real minds: the barrier of consciousness*" and thus that information-processing and the conscious experience of raw sensations[1] are conceptually distinct [12]. Indeed, Cartesian theories typically describe cognition in terms of its objective and subjective aspects, so breaking down the 'problem of mind' into what David Chalmers [7] calls the 'easy' problem of perception - the classification, identification and processing of sensory (and concomitant neural) states - and a corresponding 'hard' problem, which is the realization of the associated raw phenomenal experience of sensation. The difference between the easy and the hard problems - and the apparent lack of a link between theories of the former and an account of the latter - has been termed the 'explanatory-gap'.

But is consciousness experience a necessary prerequisite for the realisation of cognition and genuine mental states in all entities - both natural and artificial? John Searle suggests that it is, "*.. the study of the mind is the study of consciousness, in much the same sense that biology is the study of life*" [11] and this observation leads Searle to outline a 'connection principle' whereby "*... any mental state must be, at least in principle, capable of being brought to conscious awareness*" (ibid).

Yet is such conscious experience also necessary for an entity to be considered 'creative' and, furthermore, can a mere computing machine (qua computation) ever aspire to realise consciousness, in all its beautiful and terrifying grandeur?

Certainly across the realms of science and science fiction the hope is periodically reignited that a computational system will one day be conscious in virtue of its execution of an appropriate program; thus in 2004 the UK funding body EPSRC awarded a substantial 'Adventure Fund' grant to a team of Roboteers and Psychologists at Essex and Bristol led by Owen Holland, with a goal of instantiating consciousness in a humanoid-like robot called Cronos.

But equally, the view that the mere execution of a computer program can bring forth *consciousness* has not gone unchallenged. Indeed, one argument that I have developed, which questions the very possibility of such a *machine consciousness*, is the 'Dancing with Pixies' (DwP) thought experiment [2], [3], [4] &[5].

[1]The term 'consciousness' can imply many things to different people; in the context of this essay I specifically mean that aspect of consciousness Ned Block terms 'phenomenal consciousness' [6] and by this I specifically refer to the first person, subjective phenomenal sensations: pains, smells, the ineffable red of a rose, and so on.

Baldly speaking DwP is a simple *reductio ad absurdum* argument to demonstrate that:

- IF **the assumed claim** – *that an appropriately programmed computer really does instantiate genuine phenomenal states* – ***is true;***
- THEN **panpsychism** – *the view that all matter has consciousness* – **is true.**

However if, against the backdrop of our scientific knowledge of the closed physical world and the corresponding desire to explain everything ultimately in physical terms, we are led to reject panpsychism, then the DwP reductio suggests computational processes cannot instantiate phenomenal consciousness and computational accounts of cognitive processes must, at best, exhibit what John Searle termed weak artificial intelligence; a so called 'weak AI'.

Weak AI does not aim beyond engineering the mere simulation of [human] intelligent behaviour; strong AI, in contrast, takes seriously the idea that one day machines will be built that *really can think* (be conscious, have 'genuine' understanding and other cognitive states) purely in virtue of their execution of a particular computer program [10].

Furthermore, taken alongside the Chinese Room Argument (CRA)[2] - Searle's famous critique of strong AI and machine understanding (ibid) - I suggest the DwP reductio places bounds on the successes of any mere *computationally* powered *creativity* project because, if Searle and I are correct, no purely computational engine can ever genuinely feel or understand anything of the world nor, indeed, anything of its own 'creative response' to that world (nor the world's response to it).

Thus, echoing Searle's taxonomy of Artificial Intelligence, Mohammad Majid al-Rifaie and I have suggested a dual taxonomy of 'computationally creative systems': a **weak** notion, which does not go beyond exploring the *simulation* of [human] creative processes; emphasising that any creativity so exhibited springs forth from the interaction of man and machine (and fundamentally remains the responsibility of the human) and a **strong** notion, in which the expectation is that the underlying creative system is autonomous, autopoietic and conscious, with 'genuine understanding' and other cognitive states [1].

That said, of course there always remains a trivial sense in which every time we run any computer program the machine is in some sense 'computationally creative', as symbols and patterns that, perhaps, have not previously been output together (say as a novel image) are cranked forth into the world; as Newell and Simon [9] famously observed back in 1973:

> Computer science is an empirical discipline. We would have called it an experimental science, but like astronomy, economics and geology, some of its unique forms of observation and experience do not fit a narrow stereotype of the experimental method. None the less, they are experiments. Each new machine that is built is an experiment. Actually constructing the machine poses a question to nature; and we listen for the answer by observing the machine in operation and analyzing it by all analytical and measurement means available.

[2]The Chinese Room Argument is John Searle's (in)famous critique of strong AI and machine understanding [10]; if correct Searle has demonstrated that 'syntax is not sufficient for semantics' and hence that computational systems can never genuinely 'understand' the symbols they so powerfully manipulate.

However, lacking autonomous teleology, contextualisation and intent, even this modest conception of a [computational] creative process is merely analogous to a ballistic *throw of a dice*[3], soliciting only the faintest echo of 'creativity' as the word is more usually employed.

Viewed under a modern conception of creativity - as a process positioned within a reflective historical lineage - such reflections inexorably prompt us to question in what sense any computational system could ever be seriously described as strongly creative (and not simply as a tool, an accelerator, to its programmers own vivid imaginings) ..

Indeed, in his recent address to open AISB50 (the 50th anniversary conference of the UK society for Artificial Intelligence and the Simulation of Behaviour), Harold Cohen, the British-born artist well known as the creator of AARON (a computer program often claimed to produce art 'autonomously') retreated from this very shibboleth by electing to describe his own work merely in terms of interactive collaborations between man and machine.

Hence, in the light of these concerns - and until the challenges of the CRA and DwP have been fully addressed and the role of the mind's embodiment strongly engaged - I suggest a note of caution in labelling any computational system as 'strongly creative'; any creativity displayed therein being simply a projection of its engineer's intellect, aesthetic judgement and desire.

References

[1] al-Rifaie, M.M., & Bishop, J.M., (2014), 'Weak and Strong Computational Creativity' by in Besold & Smaill, "Computational Creativity Research: Towards Creative Machines", (forthcoming 2014)

[2] Bishop J.M. (2002) Dancing with Pixies: strong artificial intelligence and panpyschism. in: Preston J, Bishop JM (Eds), Views into the Chinese Room: New Essays on Searle and Artificial Intelligence, Oxford University Press, Oxford.

[3] Bishop, J.M., (2005), 'Can computers feel?', The AISB Quarterly (199): 6, The AISB, UK.

[4] Bishop, J.M., (2009) Why Computers Can't Feel Pain. Minds and Machines 19(4): 507-516.

[5] Bishop, J.M., (2009) A Cognitive Computation fallacy? Cognition, computations and panpsychism. Cognitive Computation 1(3): 221-233.

[6] Block N. On a Confusion about a Function of Consciousness. In: Block N, Flanagan O, Guzeldere G, editors. The Nature of Consciousness, Cambridge MA: MIT Press; 1997.

[7] Chalmers, D.J., (1996), The Conscious Mind: In Search of a Fundamental Theory, Oxford: Oxford University Press.

[8] Nasuto, S.J., Bishop, J.M., Roesch, E., Spencer, M., (2014), Zombie mouse in a Chinese room. Philosophy & Technology pp. 115. DOI 10.1007/s13347-014-0150-2. URL http://dx.doi.org/10.1007/s13347-014-0150-2.

[9] Newell, A. & Simon, H.A., (1973), Computer Science as Empirical Enquiry: Symbols and Search, Communications of the ACM 19(3): 113-126.

[10] Searle J., (1980), Minds, Brains and Programs. Behavioral and Brain Sciences 3(3): 417-457.

[11] Searle J., (1992), The Rediscovery of the Mind. Cambridge MA: MIT Press.

[3]Tristan Tzara and William S. Burroughs both famously utilised random acts (drawing a series of words from a hat; cutting-up texts and randomly rearranging them, respectively). In the context of this essay it is argued that the creativity here lies more in the artist's decision to deploy a random processes centrally within the creative act, rather than any *Rorschach interpretations* these processes eventually invoke.

[12] Torrance S., (2005), Thin Phenomenality and Machine Consciousness. In: Chrisley R, Clowes R, Torrance S, (Eds). Proceedings of the 2005 Symposium on Next Generation Approaches to Machine Consciousness: Imagination, Development, Intersubjectivity and Embodiment, AISB'05 Convention. Hertfordshire: University of Hertfordshire.