Artificial Intelligence and Extended Cognition

Michael Wheeler
School of Arts and Humanities: Philosophy
University of Stirling

Extended Cognition

- According to the hypothesis of Extended Cognition (ExC), there are actual (in this world) cases of intelligent thought and action in which the thinking and thoughts (more precisely, the material vehicles that realize the thinking and thoughts) are spatially distributed over brain, body and world, in such a way that the external (beyond-the-skin) factors concerned are rightly accorded cognitive status.

- The canonical presentation of the view is the 1998 *Analysis* paper by Clark and Chalmers, 'The Extended Mind'.

Making Space for ExC I

- ExC, as I shall interpret the view, is committed to the multiple realizability of the cognitive

- Functionalism in the philosophy of cognitive science: what matters when one is endeavouring to identify the specific contribution of a state or process *qua cognitive* is not the material constitution of that state or process, but rather the functional role it plays in generating cognitive phenomena, by intervening causally between systemic inputs, systemic outputs and other functionally identified, intrasystemic states and processes.

Making Space for ExC II

- Functionalism has conventionally been interpreted so as to secure the in-the-head multiple realizability of the cognitive.
- However, if it is interpreted without any internalist add-on, functionalism allows that the borders of the cognitive system may fall beyond the sensory-motor interface of the organic body.

- So the possibility of ExC is straightforwardly entailed by a properly formulated functionalism.
- Call this extended functionalism (Clark, 'Pressing the Flesh'; Wheeler, 'Minds, Things and Materiality')

- And note: functionalism has traditionally been the 'house philosophy' of AI

Two Questions for Today

1. How might research in AI bear on the truth or otherwise of ExC?

2. Would the adoption of ExC enable us to do AI better?

Robotics as a Route to ExC?

- The situated robotics strategy: build complete robots capable of integrating perception and action in real time, in real-world environments, so as to generate fluid adaptive behaviour
- Shun detailed inner models, which are difficult and computationally expensive to keep accurate and up to date, in favour of architectures in which the robot regularly senses its environment in order to guide its actions.
- This is “using the world as its own model” (Brooks, ‘Intelligence Without Representation’); what Brooks also calls situatedness.
- Inner (brain-bound) mechanisms, non-neural bodily factors, and scaffolding environmental elements, combine as ‘equal partners’ in the behaviour-generating processes.

The Dynamical Route

- Recent modelling work by Froese, Gershenson and Rosenblueth ('The Dynamically Extended Mind: A Minimal Modeling Case Study') shows that the "change of identity of [an artificial nervous system] ANS from one type of system to another, is only explainable as an emergent outcome of nonlinear coupling between ANS, body and the environment subsystems".
- A continuous-time recurrent neural network (CTRNN), that pre-coupling is a 1-D dynamical system, exhibits properties post-coupling (oscillatory and chaotic dynamics) which establish that it has been transformed into a higher-dimensional dynamical system.
- This establishes that "the phase space of the agent's CTRNN must be explained in terms of the whole brain-body-environment-body-brain system".

The Distributed Cyborg Route

- It might seem that AI has a different positive contribution to make to the case for ExC, by allowing the technological elements in our human-technology couplings to become that much fancier.
- There's something to be said here (e.g. smartphones that implement reasoning routines), but in the end this kind of contribution doesn't seem to be essential to the case for ExC (after all, the argument was canonically made by appeal to written entries in paper notebooks).

Not so Fast…

- So far, in answer to our first question, one might well be tempted to think of certain forms of AI as providing evidence in favour of ExC. But there's a snag…
- The Hypothesis of Embedded Cognition (EmbC): the distinctive adaptive richness and flexibility of intelligent thought and action is regularly, and perhaps sometimes necessarily, causally dependent on (a) non-neural bodily structures and/or movements, and/or on (b) the bodily exploitation of environmental props or scaffolds.
- The key distinction: the 'merely' causal (EmbC) versus the constitutive (ExC) dependence of cognition on external factors (see e.g. Adams and Aizawa, *The Bounds of Cognition*)

The Shape of the Debate

- The embedded theorist seeks to register the important, and sometimes perhaps even necessary, causal contribution made by wider bodily and environmental factors to many cognitive outcomes. That said, the embedded position is that the actual thinking in evidence in such cases remains a purely brain-bound phenomenon (one that is given a performance boost by its embodied context and its technological ecology).

- Given that ExC remains controversial in a way that EmbC mostly doesn't, it is EmbC that currently deserves to be treated as the default position in the debate. So the burden of proof rests with the advocate of ExC.

Policing the EmbC-ExC Divide

- Clark (*Supersizing the Mind*) explains that the phenomenon of cognitive self-stimulation (CSS) occurs when

a) neural systems are causally responsible for producing certain bodily movements and beyond-the-skin structures and events which are then recycled as inputs to those and/or other neural systems, and

b) this feedback process sustains sophisticated brain-body or brain-body-environment loops of exploitation, co-ordination and mutual entrainment, with various problem-solving benefits.

Arguing from CSS to Extended Cognition

- Clark takes cognitive self-stimulation to be indicative of ExC rather than EmbC. Here's his argument:

- "Sometimes, all coupling does is provide a channel allowing externally originating inputs to drive cognitive processing along. But in a wide range of the most interesting cases, there is a crucially important complication. These are the cases where we confront a recognizably cognitive process, running in some agent, that creates outputs (speech, gesture, expressive movements, written words) that, re-cycled as inputs, drive the cognitive process along. In such cases, any intuitive ban on counting inputs as parts of mechanisms seems wrong." (Clark, *Supersizing the Mind*)

Hybrid Mechanisms

- Tellingly, Clark's argument from CSS to ExC doesn't say enough: embedded and extended theorists agree that self-generated inputs that support cognitive self-stimulating loops operate within well-defined mechanisms that turbo-charge thinking.
- For the embedded theorist, the properly cognitive mechanisms in play are sub-systems of larger, performance-enhancing loops, where the latter are not cognitive mechanisms in their own right, even though they contain cognitive mechanisms.
- So a self-generated input in a cognitive self-stimulating loop may make its turbo-charging contribution to thought while remaining non-cognitive in character.

The Moral

- Our reflections on CSS indicate that there may be extended systems that are not themselves cognitive systems, although they may contain (embedded) cognitive systems
- Our AI-based examples of world-involving intelligence can be given an extended reading, but they can also be given this sort of embedded reading
- It is worth pausing to comment on the Froese et al. case. They say: “The non-isolated CTRNN’s output is determined by its input, albeit mediated by its internal activity, while this input is determined by its motor output, albeit mediated by bodily and environmental… activity.”
- In other words, this is a case of CSS, so we have reason to believe that the ExC-driving constitutive claim – held by Froese et al. to be a direct consequence of the claim quoted immediately above – won’t go through

The Appeal to Nonlinearity

- According to Chemero (quoted approvingly by Froese et al.), “when the agent and environment are nonlinearly coupled, they together constitute a nondecomposable system, and when that is the case, the coupling-constitution fallacy [roughly, the move from causal dependence to constitutive dependence] is not a fallacy” (*Radical Embodied Cognitive Science*, pp.31-2).
- The coupling between system A and system B is nonlinear when at least one variable of A is a parameter of B and at least one parameter of B is a variable of A.
- A nondecomposable system is a system whose behaviour “cannot be modeled, even approximately, as a set of separate parts” (Chemero, p.31) or (equivalently), a system whose behaviour can be characterized only using “*collective variables* and/or *order parameters*, variables or parameters… that summarize the behavior of the systems’ components” (Chemero, p.36).

Learning from the Watt Governor

- There's an equation that describes the engine speed and an equation that describes the change in the arm angle of the governor.
- These are nonlinearly coupled.
- Any change in arm angle changes the entire dynamics of the system that describes the speed of the engine.
- Any change in the speed of the engine changes the entire dynamics of the system that describes the change in the arm angle.
- So, for Chemero, Froese and company, steam engine and governor form a nondecomposable system.

EmbC and ‘Nondecomposability’

- Whatever nondecomposability may involve, the governor-engine system still features an engine-side variable/parameter (engine speed) and a governor-side variable/parameter (arm angle)
- So, in a nondecomposable agent-environment system, there may still be agent-side variables/parameters and environment-side variables/parameters
- So, nonlinear coupling may produce a ‘nondecomposable’ agent-environment system, but it is not yet settled which elements within this extended system are cognitive in character.
- So again, the advocate of EmbC will play the EmbC-as-the-default card

Intermediate Conclusion

- From a purely AI-engineering perspective, a perfectly healthy conservatism regarding the conditions for theory change in science would favour EmbC over ExC
- On these grounds, AI could not directly show ExC to be true.
- Note: it could presumably be part of a stronger case against ExC: if our AI-engineered examples of world-involving intelligence consistently failed to be useful, one couldn’t get even get ExC off the ground!

The Mark of the Cognitive

- Nevertheless, AI can have an important indirect effect on the fate of ExC, by helping us to articulate what, in the debate over ExC, is standardly known as the mark of the cognitive.

- A proposal for deciding the ExC issue: first we give a scientifically informed account of what it is for an element to be part of a cognitive system, one that is independent of where any candidate element happens to be spatially located. Then we look to see where cognition falls.
- This is what Adams and Aizawa (*The Bounds of Cognition*) have dubbed a mark of the cognitive.

- With this idea in view, we can see that we have good non-engineering reasons for adopting ExC over EmbC.

Looking for the Mark of the Cognitive

- If we extract our mark of the cognitive from human cognitive psychology, there may be a tendency to beg the question against ExC, by turning accidental aspects of purely organic cognition into defining features of cognition in general.
- For example: one might think that certain robust but accidental characteristics of human memory, such as primacy and recency effects, are essential features of memory. These would plausibly not be replicated in an extended memory system.
- NB: this kind of chauvinistic ‘mark of the cognitive’ has been used in arguments against ExC

AI and Cognitive Science

- Where is the intellectual core of cognitive science?
- One traditional answer: AI.
- It is possible to conceive of AI as the science of intelligence in general: "As such, [the] goal [of AI] is to provide a systematic theory that can explain (and perhaps enable us to replicate) both the general categories of intentionality and the diverse psychological capacities grounded in them. It must encompass not only the psychology of terrestrial creatures, but the entire range of possible minds. It must tell us whether intelligence can be embodied only in systems whose basic architecture is brainlike (involving parallel-processing within networks of associated cells), or whether it can be implemented in some other manner. (Boden, Introduction to *The Philosophy of Artificial Intelligence*)

AI and the Mark of the Cognitive

- With AI (so-conceived) at its core, cognitive science becomes the science of cognition in general, one whose scope is the "entire range of possible minds".
- Indeed, as long as one doesn't understand the notion of a cell in an overly biochemical way, even Boden's final observation that cognition might, in principle, be restricted to 'brain-like' architectures invites an abstract functionalist specification of those architectures.
- A science of cognition in general would seem to be a likely source for precisely the kind of even-handed mark of the cognitive that we need, one inoculated (to a large extent) against a temptation to turn accidental aspects of purely organic cognition into defining features of cognition in general.
- We can give some substance to this idea.

From Symbolic Coupling…

- Bechtel argues that cognitive achievements such as mathematical reasoning, natural language processing and natural deduction, are the result of sensorimotor-mediated interactions between internal neural (connectionist) networks and suites of external symbols.

- Now consider the phenomenon of systematicity
- The "property of systematicity, and the [classical] compositional syntax and semantics that underlie that property, might best be attributed to natural languages themselves but not to the mental mechanisms involved in language use" (Bechtel, 'Natural Deduction in Connectionist Systems')
- Is this a case of cognitive extension?

…to Extended Physical Symbol Systems

- Here is a possible mark of the cognitive: a physical symbol system (PSS), when sufficiently complex and suitably organized, and when placed in the operating context of a complete cognitive architecture, has the necessary and sufficient means for certain aspects of cognition.
- The Bechtel-style network-plus-symbol-system architecture is an instantiation of an extended PSS and thus, if we adopt the above mark of the cognitive, it's an instantiation of an extended cognitive system (or subsystem)
- Notice that questions of revisionism no longer (obviously) favour EmbC
- This (I think) is how to establish ExC.

Answering our Questions

1. How might research in AI bear on the truth or otherwise of ExC?

- We can now give a fuller answer to this question: although AI cannot directly show ExC to be true, it can make a crucial indirect contribution to the issue by helping us to articulate a mark of the cognitive.

2. Would the adoption of ExC enable us to do AI better?

- An answer to this question now suggests itself: if a successful case for ExC cannot be built using the strategy outlined in answer to our first question, then ExC will be of no more than heuristic value to an (embedded) AI researcher who already recognizes the intimate causal dependence of cognition on environmental scaffolding.

Some Relevant Papers by Wheeler

- 'Is Cognition Embedded or Extended? The Case of Gestures', in Radman, Z. ed., *The Hand, an Organ of the Mind: What the Manual tells the Mental*, MIT Press, Cambridge, Mass., 2013, pp. 269-301.
- 'Thinking Beyond the Brain: Educating and Building from the Standpoint of Extended Cognition', *Computational Culture* 1, 2011 (on-line journal).
- 'Embodied Cognition and the Extended Mind', in Garvey, J., ed., *Continuum Companion to the Philosophy of Mind*, Continuum, London, 2011, pp.220-238.
- 'In Defence of Extended Functionalism', in Menary, R., ed., *The Extended Mind*, MIT Press, Cambridge, Mass.. 2010, pp.245-270.
- 'Minds, Things and Materiality', in Renfrew C. and Malafouris L. (eds.), *The Cognitive Life of Things: Recasting the Boundaries of the Mind*, McDonald Institute for Archaeological Research, Cambridge. 2010, pp.29-37. Reprinted in J. Schulkin (ed.) *Action, Perception and the Brain*, Palgrave-Macmillan, 2011, pp.147-63.
- Most of these papers are available from http://rms.stir.ac.uk/converis-stirling/person/11855