Philos Stud (2011) 152:413–416
DOI 10.1007/s11098-010-9597-x

Précis of *Supersizing the mind: embodiment, action, and cognitive extension* (Oxford University Press, NY, 2008)

Andy Clark

Published online: 28 September 2010

Supersizing the Mind (henceforth, SSM) was a book with a double mission. The first mission was to display and discuss the rich and varied landscape of recent work in the area of (broadly speaking) embodied, environmentally embedded, cognitive science. To this end, I canvassed and organized a wide range of examples in which fine details of embodiment, of worldly action, and of worldly resources, could be seen to make diverse and unexpectedly deep contributions to human cognitive achievements. The second mission, building in many ways upon the first, was to pursue the more radical suggestion that is now known as the Hypothesis of Extended Cognition.[1] This was the suggestion[2] that in some such cases, there is a sufficiently dense degree of inter-animation between the neural and the gross-bodily, or even between the organismic and the extra-organismic, for it to become ill-warranted and unproductive to reserve the label of 'cognitive processing' for the inner, neural or organismic contributions alone. Of course, the mere fact of dense inter-animation will not be enough: there may well be dense inter-animation between, say, the sailor and the sailboat, or between the digestive tract and the brain, without either the sailor-sailboat or the brain-digestive tract system counting as an extended *cognitive* system. But where we find dense inter-animation *and* that inter-animation looks to be serving recognizably cognitive (for example, broadly speaking epistemic or knowledge-oriented) ends, then (assuming, see below, that we can also assign "ownership" of the relevant states or processes to a distinct agent) then there is, or so I argued,no good reason to carve the mental cake according to

[1] This idea is introduced in Clark and Chalmers (1998).

[2] Clark and Chalmers (1998).See also, among others, Dennett (1991) (1996), Donald (1991), Haugeland (1998), Hurely (1998), Hutchins (1995), Menary (2007), Rowlands (1999) (2003) (2006), Sutton (In press), Wheler (2005), Wilson (2004).

A. Clark (✉)
Philosophy, University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh EH8 9AD, Scotland (UK)
e-mail: andy.clark@ed.ac.uk

merely metabolic joints: no good reason, that is, to think that all our cognitive processes need be found in the head, or even within the biological organism.

The first mission (that of displaying the broad shape of 'embodied, embedded cognitive science') is important, though it attracts, perhaps understandably, far less critical attention than the second. In pursuit of the first mission, SSM highlighted the notion of 'information self-structuring' (Lungarella and Sporns (2005)). The key idea here was that at multiple time scales, embodied agents structure their own information flows in ways that support richer forms of adaptive and cognitive success. Sometimes, such structuring happens on the short time scale of bodily movement. For example (Ballard et al. (1997)), when we move our heads and eyes as we speak and reason. It also happens on intermediate time scales, for example when we scribble words on a page while attempting to plan just when we need to get the taxi to arrive at the bus station in time to catch the bus to the airport. And it happens on even longer time scales when we structure our persisting environments in ways that will serve future needs, as when we leave a yellow sticky note on the door reminding us to get milk, or post signs on the highways directing motorists to their destinations. In all these cases we (either individually or collectively) structure our worlds and actions in ways that, usually productively, alter the flow of information arriving at the biological brain. It is these abilities of rampant and iterated "cognitive niche construction" that have set us somewhat apart from the other animals with whom we share both a planet and a massive fundamental biological heritage.

SSM was also at pains to distance itself from those versions of embodied cognitive science that reject the use of notions of internal computation and/or internal representation in their explanations of human thought. It was at pains, too, to avoid accounts of thought and experience that (it was argued) give *too strong* a role to the idiosyncrasies of human embodiment and action. More positively, SSM suggested that we now possess the main tools needed to arrive at a mature science of the embodied mind. What is needed is a careful combination of computational, representational, and dynamical sensibilities: one that recognizes that bodily motions (as in the case of eye fixations, gesturing while speaking etc.) may themselves be playing important information processing roles. Given this kind of 'extended functionalist' approach, we might even (see chapter 9 of SSM) one day hope to *quantify* (Lungarella et al. (2005)) the contributions of embodied action to cognitive processing using various information theoretic measures.

Mission 1 phased directly into the rather more contentious (philosophically at least) mission 2: the defense of the claim that, under certain circumstances, the dense interplay between neural and extra-neural factors might be such as to warrant talk of extended cognitive processes or even of an "extended mind". The argument presented comprised three key components. First, there was a general principle (the so-called Parity Principle, more on which below) which is best seen as a heuristic (a rough-and-ready tool) for identifying some plausible cases of cognitive extension. Second, there was a thought experiment (the case of Otto) meant to convince the reader that, under certain conditions, the coarse functional role of a bio-external encoding could be sufficiently similar to that of a persisting internal encoding as to mandate similar treatment, revealing the non-biological resource as part of the

physical machinery underpinning some of an agent's genuine mental states. Third, there was an attempt to display the specific kinds of temporal and computational complexity that might further support the even-more-contentious claim that various relatively transient real-world systems may also be profitably analysed, while up-and-running, so to speak, as unified cognitive wholes, rather than by splitting them into a cognitive (biological, typically neural) component and various (non-cognitive) sources of input, transformation, and storage.

Here is a super-brief sketch of each of these three components, starting with the rule of thumb now known as the Parity Principle. The parity principle, as introduced by Clark and Chalmers (1998, p.8) suggests that "If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process." The idea here was simply to invite the reader to judge various potential cognitive extensions behind a kind of "veil of metabolic ignorance". A good way to do this is ask yourself, concerning some candidate cognitive process P, whether *if* you were to find P (or better, its functional equivalent) occurring inside the head of some alien organism, you would tend to class P as a cognitive process? If so, then the onus, it was claimed, is on the skeptic (the person who wants to deny that the functionally equivalent process P, when bio-externally realized and suitably coupled with the biological agent, is a cognitive process or forms part of the agent's cognitive processing) to make her case.

Alongside this simple rule of thumb, Clark and Chalmers offered a thought experiment meant to show how, for the familiar case of an agent's dispositional beliefs, a bio-external resource might indeed make the grade. The thought experiment concerned Otto, a person with mild memory impairments, who makes extensive use of a notebook to guide his behaviour. I won't rehearse the Otto case yet again in this short précis, except to note that the crucial point here was that by treating the notebook encodings as directly partially determining Otto's some of standing beliefs, we get to grips with the *very same coarse patterns* in Otto's behaviour as we do, in more standard cases, by treating an agent's neural states as such realizers. For example, we lock onto a coarse pattern characteristic of holding the standing belief that MOMA is on 53rd street by treating the inscriptions in Otto's notebook as partially realizing some of *his* standing beliefs, just as we might do with a normal agent ('Inga') by treating some of her neural states as such realizers. Both Otto and Inga, we argued, should be treated as holding the standing belief (that is, as believing even before the moment of conscious recall) that MOMA is on 53rd Street. Unlike Inga, Otto is not a fully normal agent, but courtesy of the extra-biological machinery, much more of his behaviour can be successfully subsumed under our familiar folk psychological kinds: kinds such as 'believing that MOMA is on 53rd street'.

The third and final ingredient in mission 2 was an attempt to plot, in rather more detail than before, something of the kinds of temporal and computational complexity that characterize the best real-world exemplars of extended cognition. The goal here was to highlight the potential complexity of the 'dovetailing' that might be achieved between plastic neural systems and highly practiced bio-external

props and supports. An important upshot of such complexity, I argued (here pursuing ideas that I first encountered in Kirsh and Maglio (1994)) was that there need be no neat, persisting "agent-level" bottleneck mediating the brain's calls to some external resource. Instead, different neural sub-systems would have their own sub-personally mediated ways to "call" or access the external resource, thus building the resource's reliable presence so deep into the information-processing flow chart that it becomes visibly arbitrary and unhelpful to draw a single metabolically determined line dividing the truly "cognitive" (i.e., participating in the cognitive processing of the agent) aspects of that processing from the rest.

SSM presented various bodies of research (from the use of deictic pointers in visual processing, to multiple timescale coupling in Tetris, to recent systematic work by Gray et al. (2006)) on the cognitive control of interactive behaviour) meant to lend support to this vision of complex dovetailing. Such complex dovetailed wholes will genuinely reward, or so SSM argued, understanding in terms of an extended functional organization: an extended functional organization relative to which the metabolically-determined inner-outer boundary is both analytically unhelpful and computationally far less significant than one might have pre-theoretically supposed.

None of this was meant to suggest that in such cases an extended perspective was mandatory, or (worse still) that in such cases the neural contribution to human cognitive success was somehow "less special" than we might have previously imagined. Indeed, a large chunk of SSM is devoted to displaying the human brain as an organ supremely well equipped to thus recruit external structures and opportunities deep into its problem-solving routines. It is only once this important work has been achieved, SSM suggests, that a more egalitarian regime is enabled: one in which no special status accrues to the inner over the outer. In the end, then, the value of the extended perspective was said to depend on the fruitfulness of making a "mental flip" that sees many of our genuinely cognitive unfoldings as running on machinery spread across brain, body, and world. It is always possible at that moment to flip back, and to see only the fine weave of inputs to, and outputs from, a complex inner machine. But to *confine* ourselves to that single perspective is to unreflectively privilege the inner and the biological in ways that, if SSM is on the right track, are both philosophically unmotivated and scientifically unsound.