Intrinsic content, active memory and the extended mind

Andy Clark

Clark and Chalmers (1998) defend the view that the human mind need not be in the human head. To speak more carefully, they defend the view that the material vehicles1 of cognition can be spread out across brain, body and certain aspects of the physical environment itself. Critics of this view (which has become known as the ‘extended mind hypothesis’, henceforth EM) have pointed to the supposedly intrinsic nature of the contents carried by inner biological vehicles. External, non-biological vehicles, such critics claim, are not (at least in any typical human case) similarly endowed, and thus fail to exhibit one of the key ‘marks of the cognitive’ (Adams and Aizawa 2001: 48). I shall argue that this response is mistaken. First, though I shall not dwell upon this, it is unclear that there is any such thing as intrinsic content anyway. Second, in so far as the notion is intelligible at all, there is no reason to believe that external, non-biological structures are incapable of supporting such content. And third, even if they were incapable of so doing, this would not actually compromise the case for the extended mind. The worry about intrinsic content, I conclude, is multiply fatally flawed. Despite all this, there are some related and legitimate worries that may well be informing the unease expressed by Adams, Aizawa and others.2 The paper ends by briefly discussing these worries, pointing to a possible solution, and then gesturing at the larger picture of the extended mind and its place in nature.

The original argument for EM took the form of some thought experiments, a principle meant to command rational assent, and a battery of

1 It is important, in considering the issues to be discussed, to maintain a distinction between vehicles and contents. Possessing a contentual mental state is most plausibly a property of a whole active system (perhaps in some historical and/or environmental context). Within that system, certain enduring material aspects may play a special role in enabling the system to possess (whether occurrencely or dispositionally) a given mental state. These material aspects are the vehicle of the content. The Extended Mind hypothesis is really a hypothesis about extended vehicles, vehicles that may be distributed across brain, body and world. We conflate vehicles and contents, as Dennett (1991) and Hurley (1998) stress, at our philosophical and scientific peril.

2 Some related critical treatments include O’Brian 1998, Dartnall (in press) and Weiskopf (submitted). For a somewhat different critical angle, see Sterelny (in press).
responses to probable objections. One of the thought experiments (the one that the critics typically choose to engage) depicts two agents, Inga and Otto. Inga hears of an intriguing exhibition at MOMA (the Museum of Modern Art in New York). She thinks, recalls it’s on 53rd St., and sets off. Otto suffers from a mild form of Alzheimer’s, and as a result he always carries a thick notebook. When Otto learns useful new information, he always writes it in the notebook. He hears of the exhibition at MOMA, retrieves the address from his trusty notebook and sets off. Just like Inga, we claimed, Otto walked to 53rd St. because he wanted to go to the museum and believed (even before consulting his notebook) that it was on 53rd St. The functional poise of the stored information was, in each case, sufficiently similar (we argued) to warrant similarity of treatment. Otto’s long-term dispositional beliefs just weren’t all in his head.

Supposing you accept (and you might not: see below) the claim about similarity of functional poise, then a very general principle might (we suggested) explain your endorsement of Otto as dispositionally believing (even before opening the notebook) that MOMA was on 53rd St. Let us now call that principle ‘the Parity principle’. It went like this:

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. (Clark and Chalmers 1998: 8)

In other words, for the purposes of identifying the material vehicles of cognitive processes, we should (normatively speaking) ignore the old metabolic boundaries of skin and skull, and attend to the computational and functional organization of the problem-solving whole.\(^3\) We did not argue for this principle, except by providing illustrations and defences against counter-arguments. In the end, it was meant to command rational assent as a means of freeing ourselves from mere bio-chauvinistic prejudices.

To make this case, we anticipated, and replied to, an extensive battery of possible objections. I shall not rehearse these here, save to mention that they cover the most obvious worries. Thus, to the worry that all that Otto really believes before consulting the notebook is that the address might perhaps be in the notebook, we reply that the notebook is consulted automatically and accessed without conscious deliberation, just like Inga’s biological recall. In this way, the notebook (like biological memory) functions as transparent equipment in a broadly speaking Heideggerian (1927/
1961) sense. To the worry that the notebook is fragile and impermanent, we note that biological memory can be compromised by neural trauma and (contrariwise) that Otto may be presumed to reflexively protect the notebook from loss and harm much like a rather fragile biological limb or organ. In response to the more general worry that there is, on our account, just no stopping the leakage of mind into the world (and thus that there would be reductios aplenty in the offing), we pointed to a number of additional constraints, including those just mentioned (fluent and automatic use, transparency in use, general availability as and when needed), as well as others having to do with trusting (ceteris paribus) the information thus stored and retrieved. By these means we sought to erode putatively significant differences from both sides, noting (for example) that the biological is itself vulnerable to loss and intervention, and that the non-biological may become fluently assimilated, typically trusted and jealously protected.

The main argument thus takes the form of a challenge: show us why the case of Otto and his notebook (thus elaborated) is not simply that of an unusual realizer: an extended physical vehicle for a set of dispositional beliefs very much like Inga’s own. Functionalists, as Dan Weiskopf (submitted) usefully points out, have long been open to unusual realizers for mental states, including hybrid realizers that include e.g. silicon and biological parts. All that differs is that our hybrid criss-crosses the ordinary boundaries between brain, body and environment.

Critics such as Adams and Aizawa take up the challenge. They grant us the functionally-inspired Parity Principle, and the various difference-eroding manoeuvres, and then try to block the conclusions downstream. If their argument works, then the external structures (in the typical human case) lack something so crucial that even were they to be located inside the head, they would not count as parts of the cognitive process. This is the argument from intrinsic content, to which I now turn.

Adams and Aizawa label as ‘transcranialist’ a variety of positions attributed to Clark and Chalmers (1998), Merlin Donald (1991), Daniel Dennett (1996) and Ed Hutchins (1995). The transcranialist is one who holds that ‘cognitive processes extend in the physical world beyond the bounds of the brain and the body’ (Adams and Aizawa 2001: 43). But transcranialism, they retort, though it may describe both ‘a logical and a nomological possibility’ is contingently false of human cognition (43). The reason for this, they suggest, is that (as a contingent matter of fact) it is only internal representations, as encoded in neural media, that have intrinsic content. In all other cases (including, especially, the inscriptions in Otto’s notebook) the symbols and representations exhibit only derived content, and thus fail to display a key ‘mark of the cognitive’.
The notion of intrinsic content is not, I think, one of the clearer and more distinct products of mid twentieth-century philosophizing. In so far as the notion has any clear meaning, it seems to mean a state or process that, in some larger context, carries a specific content or meaning in an entirely non-conventional way. Thus the inscription ‘cat’ means CAT in virtue of certain public conventions of use; whereas if some neural state carries the content CAT, it presumably does so in some more direct manner, that underpins, but does not itself depend upon, any history of public use or social practice.

Adams and Aizawa insist that ‘whatever is responsible for non-derived representations seems to find a place only in brains’ (63). I am not at all sure this is true. It seems possible (for example) to ascribe representational contents, in ways that are not obviously conventional or derivative, to the states and processes of artificially evolved creatures (see Pfeifer and Scheier 1999: ch. 8). Or, if simple artificial creatures do not move you, take any inner neural structure deemed (by whatever non-question-begging criteria Adams and Aizawa choose) to be the vehicle of some intrinsic content X. Can we not imagine replacing part or all of that structure with a functionally equivalent silicon part? (As a matter of fact, this kind of replacement has already been done, albeit only with one artificial neuron that functions successfully within a group of 14 biological neurons in a Californian spiny lobster – see Szucs et al. 2000). Unless we question-beggingly assert that only neural stuff can be the bearer of intrinsic content, then surely we should allow that the siliconized vehicle, or at least the hybrid circuit that now includes it, is as capable of supporting intrinsic content as was its biological predecessor? For these kinds of reason, I do not believe that there is any non-question-begging notion of intrinsic content that picks out all and only the neural in any clear and useful fashion. My own view, in fact, is that the very notion of intrinsic content is essentially fuzzy and indistinct, and that (to put it bluntly) content is as content does (see Prinz and Clark 2004).

But since Adams and Aizawa stress that they are defending only a contingent, humans-as-currently-constituted, form of cognitive intracrani-alism, I suspect that they will concede this general point without much argument. The real force of Adams and Aizawa’s worry does not lie in any simple (and surely naive) identification of the neural and the cognitive. Rather, the worry is that the inscriptions in Otto’s notebook (unlike, say,

---

4 It is related, I suppose, to Searle’s (1980, 1990) notion of intrinsic intentionality. But whereas that notion seems designed to apply to (the putative mental states of) whole systems rather than their parts, ‘intrinsic content’ (as Adams and Aizawa use the term) needs to be able to apply to specific sub-states or processes, identifying them as the bearers or vehicles (in systemic context) of genuine (not ersatz or derived) meaning.
the hybrid neural and silicon-based activity that now underlies control of
the oscillatory rhythms in the stomatogastric ganglion in the Californian
lobster) are out-and-out conventional. They are passive representations
(more on this later) that are parasitic, for their meaning, on public prac-
tices of coordinated use. It is at this point that I think the Parity Principle
can be pressed once more into service. Consider, for example, an episode
of in-the-head problem solving during which I imagine the partially over-
lapping circles of a certain Venn diagram. Surely the set-theoretic meaning
of this overlap is a matter of convention? Yet the images figure centrally
in what is surely a cognitive process in good standing. The critics’ retort
will be quick. ‘That shows nothing’, they will say, ‘because the image,
when understood, must be understood only in virtue of its association
with neural goings-on that enjoy the essential luxury of intrinsic about-
ness’. But this won’t yet do, as when Otto’s notebook is accessed and
deployed, neural goings-on with associated intrinsic contents are likewise
triggered. Now the temporal dynamics loom large. For our claim, recall,
concerned precisely that time before the notebook is accessed and
deployed. So the correct question is: ‘What about while the inscription,
“MOMA is on 53rd St.” is simply dormant in the notebook. Inga’s stored
and dispositional beliefs have intrinsic contents too, not just her occurrent
ones. Surely that, at least, is not true of Otto.’

This is indeed a harder question. But suppose (and this is just the Parity
principle at work again) we encountered Martians whose biological mem-
ory allowed them to store, on demand, bit-mapped images of important
chunks of text. What is stored is thus a passive and conventional internal
representation of a passive, and conventional external representation.5
When necessary, they can later access these images, thus perfectly recalling
these chunks of data. Surely, in that case, we would have no hesitation in
counting these goings-on as part and parcel of Martian cognition. But in
that case, it is surely only skin-and-skull based prejudice that stops us
extending the same courtesy to Otto.

One way to understand this proposal is as insisting that all that matters,
for some conventional encoding to count as the vehicle of a dispositional
belief, is that it be appropriately linked, at run-time, to representations
whose content is (as Adams and Aizawa insist) intrinsic. Such linking can
be achieved for conventionally formatted representations both inside and
outside the head. This defence of Otto is the strongest, since it gives away
the most to the critics. Indeed, in a rather telling aside towards the end
of the section on intrinsic content, Adams and Aizawa allow that:

5 For an interesting account of the complexities that can arise when thus dealing with
encodings of encodings, see Haugeland 1998.
Having argued that, in general, there must be non-derived content in cognitive processes, it must be admitted that it is unclear to what extent every cognitive state of each cognitive process must involve non-derived content. (Adams and Aizawa 2001: 50)

But this concession, I submit, removes the entire force of the appeal to intrinsic content as a reason for rejecting EM. For it was no part of EM to claim that one could build an entire cognizer out of Otto-style notebooks!

This, however, immediately suggests a further issue, and one that many critics have seen as the fatal flaw in the original argument. This is the issue of (what I shall dub) active dispositional memory. This worry has been raised in various forms by O’Brian (1998), Dartnall (in press), and Weiskopf (submitted). Central to all these treatments is the observation that biological memory is not a passive encoder so much as a constantly active system that is both integrative and reconstructive. Thus, to borrow an example from Weiskopf, a normal agent who believes that MOMA is on 53rd St. and who later learns that the museum has been torn down to make way for a by-pass will immediately cease to believe that it is on 53rd St. The earlier belief is unconsciously and automatically revised with the receipt of the new information. Moreover, there is what Weiskopf calls ‘active propagation’, such that other related beliefs are similarly revised (for example, the belief that MOMA is a good place to get a latte on 53rd St. and so on). All this, it is claimed, forms part of the normal functional role of belief. Since the inscriptions in Otto’s notebook are not thus subject to automatic, actively propagated, revision, they should not count as the vehicles even of dispositional believings. They are excluded on grounds both of functional dissimilarity (as above) and because the dissimilarity matters, in so far as the lack of automatic integration would otherwise be a source of unacceptable degrees of inconsistency and irrationality, of the kind we would not want to ascribe to a proper cognitive agent.

In response to this (proper and intuitively quite compelling) worry, I ask the reader to consider the following possibility. Suppose it were discovered, by some arcane mixture of experimental design and non-invasive neuro-imaging, that integration and update always occurred not at the moment new information is received but later, at the moment the outdated or otherwise affected information would have been called upon by some process of recall or action-selection? Behaviourally, this system would look just like us. Should we say that it nevertheless fails to be a true believer, just because it uses a routine more akin (in computer science terms) to compilation than interpretation? I see no reason to do so. The system is surely every bit as rational as we are, it simply deploys its resources in a somewhat different temporal sequence. It is rather like the old case of Dennett’s in which he
imagines that we discover that left-handers’ neural organization is in some ways quite unlike that of right handers. In each case, the imagined discovery seems to locate a difference only at a level below that to which the philosophical functionalist should attend.

This is, of course, broadly speaking how it is with Otto. If he has entered two inconsistent inscriptions, one dated after the first and stating that MOMA is no more, he will locate and select the new information, and perhaps delete the old, at run-time. One difference, however, remains, and that concerns the extent of this flow of run-time revision. It still seems quite restricted and local in nature. It is by no means clear, however, that biological memory manages to track and alter all the beliefs that are, logically speaking, affected by the receipt of some new piece of information. I do concede, however, that the notebook’s lack of inbuilt links and pathways makes it unlike biological memory in this regard. This brings me to the final point I want to make in defence of Otto.

It was never our claim that Otto is behaviourally indistinguishable from Inga. Trivially, Otto’s actions involve a whole lot of notebook-manipulation that Inga’s do not. And there may well be deeper and more significant differences too, as we have just seen. The claim is rather that taken as a single, integrated system, Otto-and-the-notebook exhibit enough of the central features and dynamics of a normal agent having (amongst others) the dispositional belief that MOMA is on 53rd St. to warrant treating him as such.

I beg the reader’s forgiveness for deploying one last imaginary case to underline this point. Suppose that it had turned out (as it surely might have) that in a few restricted domains biological memory were not reconstructive, and that, in these cases, what was retrieved was always what was originally laid down. Imagine, for example, that our memory for faces (only) was like this, so that I never merged two faces or made errors of recall due to subsequent learning. This is meant to be somewhat analogous to Otto, whose long-term notebook traces are indeed unusually static. In this counterfactual world, should we say that these passive aspects of memory cannot count as partial determinants of some of the agent’s dispositional beliefs (e.g. about the name of the person who looks like that)? I see no reason, given the role that belief ascription plays in our usual form of life, to be so restrictive. But if an inner mechanism with this functionality would intuitively count as cognitive, then (skin-based prejudices aside) why not an external one? Our central claim was that many of the reasons why we might initially reject such a possibility out of hand are in fact very bad reasons, knee-jerk reactions that simply express our strong prejudices in favour of biology over technology, rather than justifying them in any way. The Parity Principle attempts to make this concrete by asking us to consider how we would react were all the same functions performed by on-board biological apparatus. There is thus a bigger picture
here than the critics’ intensive focus on the (admittedly) strange case of Otto and the notebook might suggest. It concerns the nature of persons, and the relations between biological and (as I shall say) constructive notions of agency and constitution. I would like to end, then, by very briefly (and inadequately) considering this larger picture.

Terry Dartnall, in a recent discussion of EM, comments that ‘If I dig a hole in my garden with a spade ... my-spade-and-I do not get the prize for “best hole in the garden”. I get the prize, even though I could not have done the digging without the spade’ (Dartnall (in press)). For Dartnall, then, it is always an agent using a tool, not an extended agent. This seems very sensible and compelling, until we (ouch) dig deeper concerning the role of the biological arm. Is it really just a tool too? Is his whole body just a tool for the overseeing mind? Are some areas of the brain likewise tools for the executive mind? Where, if anywhere, shall we stop the regress? In the case of the arm versus the spade, it seems right to notice that both the functional and the phenomenological lines between a well-fitted constantly available tool (more like a prosthesis, or Otto’s notebook, than a patchily available spade) and a biological limb must be fuzzy at best. There is also suggestive neuroscientific evidence to support the claim that a tool, even when temporarily in use, is rapidly assimilated into the brain’s body maps and is treated (temporarily) just like a somewhat less sensitive part of the body. When a macaque used a rake for just 30 seconds, visual receptive fields became elongated as if the rake were part of the arm, with the monkey’s fingers extended along the tines of the rake (this work is reported in Iriki et al. 1996).

Indeed, the functional integration of a typical tool in use goes further still. Berti and Frassinetti (2000) note that ‘The brain makes a distinction between “far space” (the space beyond reaching distance) and “near space” (the space within reaching distance) and that “... simply holding a stick causes a remapping of far space to near space. In effect the brain, at least for some purposes, treats the stick as though it were a part of the body.’ Were we to move away from the domain of temporary garden equipment and into that of serious, permanent or semi-permanent assistive technologies (prosthetic arms, thought-controlled wheelchairs and the like), the quip concerning the prize for digging would seem less innocent.

Lying beneath these merely suggestive snippets concerning tool use and neural maps is, I believe, an important scientific fact about the biological brain. Biological brains, I want to say, are by nature open-ended controllers. To deal fluently with bodily change and growth, they have developed ways of computing, pretty much on a moment-to-moment basis, what

---

6 Many readers will recognize the deeply Dennetian themes hereabouts. For more on all this, see Dennett 1984, 1991, 2003; Clark 2002, 2003.
resources are readily available and under direct control. Neural plasticity, exaggerated in our own species, makes it possible for new equipment to be factored deep into both our cognitive and physical problem-solving routines. Just as we can recruit a well-fitted prosthetic arm into the mix for digging the garden, so we can recruit a well-fitted external memory store into the mix for recognizing faces, or (following the spiny lobster) a well-fitted silicon substitute (on- or off-board) for an impaired or compromised neural resource.

One way to think about this is to depict the biological brain as a master of what I shall dub ‘ecological control’. Ecological control is the kind of conscious control that does not micro-manage every detail, but rather allows substantial devolvement of power and (functional) responsibility. It allows dedicated, non-conscious, ‘neural servo-mechanisms’ to deal with the fine details of reach and grasp (Milner and Goodale 1998, Clark 2001). It allows much of our skill at walking to reside in the elastic properties of muscles and tendons (Thelen and Smith 1994). And it allows (I claim) much of our prowess at thought and reason to depend upon the robust and reliable operation, in dense brain-involving loops, of a variety of non-biological epistemic devices (such as pen, paper and sketchbooks – see Clark 2001: 8). Ecological control systems (of which genes are yet another example – see Clark and Wheeler 1999) are essentially opportunistic and exploitative. They take whatever is around, and build it into problem-solving routines, exhibiting a management style delicately poised midway between anarchy and enslavement.

As Wayne Christensen (in press: 22) rather nicely puts it, ‘The high level of regulated plasticity in the brain helps explain how it is so easy to incorporate environmental organization into our cognitive practices.’

We are good at this. It is our biological nature (as I argue at length in Clark 2003) to be open to many forms of physical and cognitive hybridization. Some of these (I claim) may be so intimate as to effectively extend the thinking agent. All of them are crucial parts of the nested, iterated and ongoing process of cognitive self-re-creation that is the characteristic mark of human intelligence. It is important that we develop an understanding of ourselves (both scientific and philosophical) that is adequate to this open-ended process of physical and cognitive self-creation. To do so means

---

7 An MIT based team (led by Alex Pentland) have, for example, developed so-called ‘memory glasses’ that can recognize a face and then flash an associated name in front of the eyes of the wearer at a speed (180th of a second) that is too fast for conscious intrusion. Nonetheless, the unconscious cueing improves memory of mildly Alzheimic patients by about 50%. (Thanks to David Chalmers for drawing this snippet to my attention.)

8 For an attempt to view certain modern business management strategies in essentially these terms, see Clark 1999b.
questioning the notions of the mind and person as essentially biological, and recognizing the very large extent to which the commonplace identification of minds and persons with purely biological structures is itself what Locke (1694) termed a ‘forensic matter’: a matter of legal and moral convenience more than metaphysics, and a convenience, moreover, that must become increasingly inconvenient as science and technology progress.

University of Edinburgh
George Square, Edinburgh EH8 9JX
Scotland, UK
andy.clark@ed.ac.uk

References

9 A better bet, I think, is to identify persons with what Rovane (1998) calls a ‘rational point of view’. Starting there makes it plain that metabolic boundaries are not the issue, and clears the way for more measured reflections that take account of functional role, (loose) integration and (sufficient) coherence. The picture I favour is thus a combination of (a kind of) interpretationism (Dennett 1996) and a weakened version of rational systems theory (Rovane 1998). For some very preliminary forays into this kind of territory, see Clark 2002, 2003.

10 Locke himself was, of course, a prime opponent of the identification of the person (the ‘forensic item’) with the biological animal. Like Locke, I believe that our legal and moral purposes will be better served by a more liberal conception.

11 Thanks to David Chalmers, Dan Weiskopf and Rob Wilson for some useful exchanges concerning early drafts of this paper. Thanks also to Mike Ridge, Matthew Nudds, Thomas Johansen, Richard Holton and Rae Langton for useful comments on a presentation based on some of this material.


Weiskopf, D. Submitted. Patrolling the boundaries: a reply to Clark and Chalmers.