"Author's Reply" to symposium on Natural-Born Cyborgs. METASCIENCE, in press

Author's Response

By Andy Clark

Thought happens. Here I sit, sipping coffee, scribbling on paper, accessing files, reading and re-reading those four wonderful, challenging, yet immaculately constructive reviews. And somewhere, and to my eternal surprise, thought happens. But where, amidst the whirl of organization, should we locate the cognitive process? One possibility is that everything worth counting as (all or part) of any genuinely cognitive process hereabouts is firmly located inside the head, safe behind the ancient fortress of skin and skull. All the rest, according to this surgically neat view, is scene setting: preparing and maintaining the pitch upon which the great thinking organ performs.

Richard Feynman may well have disagreed. Upset by an interlocutor's remark that his extensive notes and scribblings were merely the record of his work, he acidly replied:

> "No, it's not a *record*, not really. It's *working*. You have to work on paper and this is the paper, ok?"[1]

Natural-Born Cyborgs was an extended meditation on this simple theme, leavened (as the title suggests) with a measured dose of techno-futurism. The human mind, I wanted to argue, is naturally designed so as to co-opt a mounting cascade of extra-neural elements as (quite literally) parts of extended and distributed cognitive processes. Moreover (and hence the techno-futurism) this ancient trick looks poised for some new and potent manifestations, fueled by innovative work on human-machine interfaces, swarm intelligence, and bio-technological union.

Let's start, then, by re-visiting that opening gambit. Thought happens. But how?

1. Skin and Out.

Terry Dartnall, in his engaging and inquisitive commentary, seems attracted to the idea that the inner is in some way special and that I make it seem less so only by perpetrating a subtle (or not so subtle) state/content confusion. Were he to confront Richard Feynman on the matter of paper trails and thought, the conversation might go like this:

> Dartnall: Your appeal to a notion of *working* is systematically ambiguous between a claim about the externalization of cognitive content (true but trite) and one about the externalization of cognitive states (interesting but false). The paper trail is a record of the *contents* of your cognitive states. But the *states themselves* never got outside your head.
>
> Feynman: I beg to differ. My thinking itself involved those markings with pen and paper. The loop into paper is far more than an ongoing record of the contents of my mental states. It forms part of the extended dynamic process that *is* the thinking.

Ok, maybe Feynman wouldn't have put it quite like that. But I do think the imaginary reply stays close to the spirit of his earlier remarks. More importantly, it is in any case the kind of view that NBC (Natural-Born Cyborgs) was meant to suggest. Dartnall's resistance to this view has (I think) two sources. One is his suspicion that the biological is in some way special, that it plays a functional role that external elements and media simply do not (currently) play. The other is his restrictive ontology of cognitive states and cognitive contents.

Ontology is a wonderful thing, in moderation. But sometimes, a neat ontology can hide the true complexity of the phenomenon we seek to

understand. If I were forced (presumably at gunpoint) to own up to an ontology hereabouts, it would primarily be one of vehicles and contents, rather than of states and contents. Thus, a perceptual content, such as greenness detected at some location in visual space, might have as its vehicle a brief and complex flurry of neural activity, none of it green. The content is thus one thing, and the vehicle another. As Dennett, Hurley, and others rightly insist, we conflate them at our peril[2]. Not all the contents capable of informing our behavior need, however, be presently active. My long-term memory, for example, enables me (if asked) to answer the question, "where is MOMA (the Museum of Modern Art)?" The vehicle/content distinction obtains here too. There is a content, viz "MOMA is on 53rd St." that has a vehicle even when I am not currently rehearsing that content. When I do mentally rehearse the content, I do so by in some way activating or calling upon[3] some vehicle, much as in the case of perceived greenness. NBC is an extended argument for what Hurley has called 'vehicle externalism': the view that the vehicles of content need not be restricted to the inner biological realm. In fact, the view in NBC is broader even than this, since not just cognitive contents, but cognitive operations (such as the comparing and transforming of representations) can, I argue, be supported by both biological and non-biological structures and processes (vehicles).

Dartnall complains that although (using his ontology) cognitive *states* remain firmly within the head, cognitive *contents* may indeed (but uninterestingly) be external. Thus concerning Otto (the mildly Alzheimic diarist from Clark and Chalmers (1998)), he writes that:

> "Otto's diary contains the content of his cognitive state, but the state itself (once he has read his diary) is in the head"
> Dartnall p.6

The question that we meant to address, however, was not that of the locus of the occurrent state of believing that MOMA is on 53rd St, which (for both Otto and normal subjects) we allow to remain firmly in the head. Rather, our discussion concerned the dispositional state of believing: a state (not a content, a state) that we ordinarily ascribe

even when an agent is not actively rehearsing what they know. It is in this dispositional sense that you may be said to believe that Madrid is in Spain, even when that snippet of world knowledge is not in use. And it is this dispositional belief that (we claimed) might be shared by two agents even if the long-term trace (the vehicle of the content proper to the dispositional state) is in one case internal and biological and in the other case external and non-biological.

There are thus two kinds of cases the argument in NBC is meant to cover. One class of cases concerns the cognitive role, in ongoing problem-solving, of active loops into non-biological media (the artist's use of a sketchpad, Feynman's frantic scribbling). The other concerns the cognitive role, as support for dispositional knowledge and belief, of non-biological forms of data-storage. The two work together in the extended cases, just as they do in the non-extended ones. Information, to be useful, needs to be both stored and deployed. It is our Cyborg nature, I argues, to use non-biological props and aids to turbo-charge both storage and use.

In addition, although Dartnall says that cognitive states (in his sense) are inside the head, it is not really clear what this can mean[4]. He is apparently not talking here about the vehicles of such cognitive states (which, we argue, may sometimes be in the head and sometimes out in the world) but the states themselves. Vehicles can certainly have spatial locations. But the actual states (of believing X, either occurrently or dispositionally) do not seem like good candidates for spatial localization. Indeed, to think otherwise seems to verge on making a kind of category mistake, of the sort that Dartnall himself explicitly warns against.

Part of Dartnall's larger suspicion is that biological forms of storage are intrinsically active (integrative, reconstructive) and that this somehow undermines the claim that non-biological media can (currently at least) serve as the vehicles of dispositional belief. I discuss this worry (and the related worries of Adams and Aizawa,

mentioned by Dartnall) at greater length in Clark (forthcoming). But for now, let me just offer one maximally brief argument. Imagine that it had turned out (as it surely might have) that certain islands of human memory were not reconstructive, and that, in these cases, what was retrieved was always what was originally laid down. Imagine, to be concrete, 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 would 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 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?

2. Technology and Transparency

Adrian Mackenzie, in his trenchant and (dare I say it?) penetrating critique, worries that I may inadvertently over-domesticate the Cyborg vision, robbing it of much of its ideo-erotic boundary-crossing charm.

I am actually sympathetic to this worry. The use of the Cyborg figure in what Mackenzie calls "cultural and feminist studies of science and technology" is in many ways different from (though not, I think, inconsistent with) my own. For it is the very essence of the Cyborg meme, in these uses, to challenge our preconceptions and to make us uneasy in our skins and in our sexual and political identities. By contrast, the key function of the Cyborg meme in NBC is to make us aware of the remarkable extent to which familiar, intuitively non-boundary-crossing, human thought and action is bio-technologically constituted, and thus to accustom us to the idea that hybridization and boundary-blindness is in fact our normal state. Here is my only

defense: the two projects merge and coalesce insofar as this realization (of the domesticity of hybrid being) removes one barrier (but only one barrier) from the exploration of even the most radical wave of near-future options. This matters, for those 'radical near-future' options are perfectly real and pressing. As William Gibson is reputed to have said, "The future is with us, it's just unevenly distributed".

It took me a while to appreciate just how the various pieces of Mackenzie's review, each of them clearly important and appropriate in their own right, actually hung together as a single (and wonderfully from-the-heart) reaction. But I think they do, in the following way. My Cyborg image can seem disappointingly 'domestic' (as above). I lay great stress on the importance of 'transparent technologies' viz those that are effectively invisible in typical daily use (e.g. the pen, which barely intrudes on our conscious awareness as we write). I celebrate a more 'biological' relationship with our best tools, in which they are simply factored in as robustly available problem-solving backdrop. Notably, I don't discuss the tidal wave of hard biotechnology itself, as manifest in designer doping in sports, drug-based enhancement of mental capacities, and the kind of genetic modification that opens the door to the delicate bioengineering of future humans. This is another tidal wave of change that may indeed (as Mackenzie very aptly suggests) augur a more technological relationship with our biology rather than the more biological relationship with our technology advocated in NBC. What NBC presents is indeed, as Mackenzie charges, the image of a rather domesticated Cyborg.

Domesticated but not, I think, toothless. And since domestication was exactly what I intended, I had better just embrace it! I accept, then, that there is a whole bunch of more radical and perhaps worrying stuff waiting in the wings. I accept that what is in this way worrying can also be exciting and liberating. My goal was to show that hybridization, *in and of itself*, is just business as usual for us humans. So if there is something especially exciting or worrying about these

other developments, then let's try to find out what it is, since it isn't (if I am right) simply the fact of hybridization, or the potential mixing of flesh and metal.

In pursuing the image of the non-domesticated Cyborg, Mackenzie rightly notes that transparency, the lynchpin of Cyborg domesticity, has its costs. But there is no disagreement here (see e.g. pages 47-58 of NBC). Sometimes, we certainly do want to bring our tools and technologies into clear focus, so as to de-bug, re-vamp, and even enjoy! And, as Mackenzie rightly points out, just who gets to do this and when is often a socially and politically charged affair. What is transparent to me may be opaque to you, and what is transparent to me today may be rendered opaque tomorrow. As Mackenzie sums it up "transparency and opaqueness are not intrinsic to the technology". I agree. All I meant to argue was that some technologies are better suited to invisibility in use than others. This is consistent with all those important points about "social, cultural, political, personal, economic and sexual power relations" (review p.20)

In sum, I accept the charge of creating a cosified, domesticated Cyborg. But I deny recklessness in so doing. The domestic Cyborg is a device for showing us what we already are, and thus better preparing us for what we might yet become.

3. A *Homo* of our own?

I am always happy to own up to a gaping lacuna, and NBC has plenty to offer in this regard. There is the lacuna, nicely spotted by Mackenzie, of emotion and valence. I don't say anything about these, yet they are central to human life, and surely play some important role in the construction of the self (a topic that NBC does indeed address). That's one very big lacuna indeed. Steven Mithen's

fascinating comments unearth another, and have forced me to think much, much harder about that glib little evolutionary scenario that was slipped under the doormat with the more substantive claims about modern-day *Homo sapiens*.

The scenario, I blush to recall, went pretty much like this. Once upon a time, there were beings whose minds were pretty much locked inside their heads. Then some of them developed (never mind how) the beginnings of human-like language. Cultured in the sea of words, these beings gradually learnt to treat their own thoughts as objects for reflection and study. With the invention of text, this process of building better worlds to think in really took off. We modern humans sit unsteadily atop this careening giant snowball of runaway co-adaptation. Our naturally plastic brains are fired in the developmental furnace of nth generation designer environments for thinking and for learning, and our thoughts are the thoughts of hybrid beings strung out between biology and those transformative waves of culture, technology and learning.

The questions upon which Mithen so wonderfully insists are simply (and profoundly), When *exactly* did this snowball start to roll? And what *exactly* got it in motion? He wants to turn up the magnification on those critical points in human history (and pre-history), so as to identify the hidden wellsprings of cognitive change. I am guilty (like many others I suspect) of sometimes finessing these questions by bluntly insisting on the role of culture, technology and training in constituting the modern mind. But to concede this (as Mithen certainly does) in no way dissolves the tricky questions. In fact, it just makes answering them all the more important. Once we appreciate the true power and reach of material culture, the question of its historical and/or evolutionary origins becomes more pressing and important than ever before.

What I found most exciting about Mithen's speculations was the idea, that I have always found attractive but never dared articulate in public, that among all of these innovations, human speech might *not*

have been the key development. Mithen astutely notes that, in several places in the text, I rather clumsily conjoin reference to speech and to the use of text, as when I say (p.81) that "with speech, text, and the tradition of using them as critical tools under our belts, humankind entered the first phase of its Cyborg existence". Repeatedly, I allude to 'speech and text', and I do so (I now suspect) so as to pay lip-service to the idea that speech might be the key, while deep-down believing (on the basis of no real evidence either way) that something else, perhaps even the use of text, is what marked the real take-off point of our Cyborg existence.

Mithen notes that the development of spoken language some 500,000 years ago did not seem to start any cognitive snowball rolling. But nor does he identify the key moment as the oh-so-recent development, around 3000 BC, of writing. Instead, Mithen's exciting suggestion is that the emergence of art, around 100,000 BC, marked the moment when we humans first began to actively extend, manipulate, and augment our own minds. The spoken word enabled co-ordination, sharing, and the cheap, extensive, non-genetic transmission of acquired knowledge and skill. But it was the practice of inscribing environmentally persisting marks, in the form of cave wall drawings and carving, that was the first scene in the cognitive drama of the modern hybrid mind. In these first artistic acts, Mithen sees the early signature of cognition-enhancing technology: a kind of augmented reality overlay that does indeed begin to blur the boundaries between physical and informational space. With the (much later) invention of text, he suggests, the solid boundary-blurring materiality of art and the informational fluidity of speech combined to yield a truly potent engine of extended cognition. I find this a truly compelling thought.

As an aside, it is interesting (to me at least) to juxtapose Mithen's excitement (at seeing the deep parallels between NBC's "descriptions of modern technology" and the cultural explosion of the Upper Paleolithic revolution) with Mackenzie's worry (p.21) that the notion of technology deployed in NBC is so broad as to risk vacuity. I agree that it is very broad, and deliberately so, so as to raise questions

about the very idea of a tiny inner agent who is the user of the body, or the brain, or the tool. At the same time, I want to stress (and this is what Mithen picks up on) the way *certain* tools and technologies materialize, freeze and externalize biologically generated thoughts and ideas. By keeping this subset in view, we can see that the notion of *kinds of technologies*, at least, is still able to do useful explanatory work.

Mithen asks whether, perhaps, those humans who lived prior to the emergence of idea-materializing artistic practice had complex thoughts (about God, the supernatural etc) but simply lacked the tricks of freezing and offloading onto the stable material environment. I remain officially agnostic on this, though I do believe that highly abstract thought is a product of, much more than a pre-condition for, the use of iterative strategies of freezing thoughts and ideas in material media.

The commentary closes with the million-euro question: do near-future technological innovations mark another major jump in human cognitive evolution? Here's one way in which they might. First wave Cyborg technologies froze thoughts and ideas in material media. New wave Cyborg technologies allow increasingly for more and more dynamic forms of delegation and offloading. For example, we can train personalized software agents to actively seek information, goods or services. This means that our non-biological props and tools are gaining some of the semi-autonomous character of dedicated neural circuitry. I don't think this signals a brand new watershed in human cognitive evolution. But it does suggest a new and exciting twist on the standard Cyborg theme.

4. Who, me?

If there is something radical lurking in the heart of my domesticated Cyborg vision, it is surely the account of the human self. That account is distributed patchily throughout the book[5], and consists, at root, in a kind of no-self (or nearly-no-self) theory, according to which (what we ordinarily think of as) the self is a hastily cobbled together

coalition of biological and non-biological elements, whose membership shifts and alters over time and between contexts.

Alicia Juarrero, in her elegant and unerringly accurate commentary, perfectly captures the spirit of the proposal while raising some of the most fundamental and important questions that it leaves unresolved. Juarrero asks how the concept of 'responsible agency' is to be fleshed out once we allow that it is (as I put it) 'tools all the way down'. If I am just a shifting loose coalition of onboard and offboard devices, how can I (who?) be responsible for my actions? And (relatedly), What holds any such coalition together? What makes any given coalition, at any given moment, count as *me*?

Where Juarrero places a question mark, Terry Dartnall digs a hole and plants a flag. Clearly uncomfortable with the idea that a standard non-biological tool could ever count as a real part of the agent, he writes 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, p.7

For Dartnall, then, it is always an agent using a tool, not an extended agent.

It is worth pushing at this a little. Suppose we ask about the role of Terry's biological arm and hand in the digging. Is this just a tool too? Certainly, it was the burden of much of NBC (as ably rehearsed by Juarrero) to show that we can achieve phenomenologically direct control over non-biological prostheses, and that skilled tool-users are in precisely the same boat. As Wayne Christiansen (forthcoming) nicely notes:

> "From the perspective of the motor control system using a tool is not fundamentally different to controlling an arm"

Christiansen (forthcoming)

There is no time to rehearse Christiansen's detailed argument here. But the main thrust is that a brain that is able to make the most of its own *internal* plasticity, by generating new, context-specific mixes of semi-autonomous modules and integrated processing, is by the very same token a brain that is at least *poised* so as to be able to co-opt *external* structures and processes into the very heart of its problem-solving routines. External and internal resources, as far as the mechanisms of integrative control and learning are concerned, are pretty much on a par. As a simple example, he cites the work of Berti and Frassinetti (2000) who 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"

Spades and sticks are, of course, impermanent parts of our typical physical ensemble, and many of our commonsense judgments about what should count as a tool versus a bodily part are clearly influenced by this. As a tool becomes more robustly available as and when needed, even these first person intuitions shift (see my discussion of Stelarc and his occasional 'third hand' in NBC chapter 5). Overall, then, I think we here confront a wide spectrum of possibilities, rather than any single sharp divide.

But perhaps the idea behind Dartnall's comments is that the whole body is itself really but a tool, and that the locus of the mind and self is smaller still, presumably somewhere in-the-head. It seems to me, however, that the common thrust of much recent work in situated cognition is precisely to reveal the body itself as a genuine player in the cognitive drama, and not just a passive tool that does the brain's bidding. Certainly, much of NBC (like *Being There* before it) aimed to counteract a vision of the brain as a kind of disembodied controller. In NBC, one such argument went like this. Go into the head in search of the (physical vehicles of the) self and you just risk cutting the

cognitive cake ever thinner, until the self vanishes from your grasp. For there is no single circuit in there that makes the decisions, that does the knowing, or that is in any clear sense the seat of the self. At any given moment, lots of neural circuits (but not all) are in play. The mix varies across time and task, as does the mix between bodily and neural activity and all those profoundly participant non-biological props and aids.

But what, Juarrero will rightly insist, holds it all together? I don't have a good answer, but I do have some suspicions. The first is that the commonsense ideas of persons, selves, agents and moral responsibility are all (deeply interanimated) *forensic* notions. That is to say, they are concepts whose application is more a matter of habit and of practical convenience than metaphysical necessity. One lesson of NBC was meant to be that near-future cases[6]will in all likelihood alter those habits and practical balances in ways that increasingly blur the line between tools and bodily parts.

The second suspicion is that the processes of 'soft-assembly' that bind the heterogeneous and distributed elements into temporary, agent-like coalitions will turn out to be scientifically tractable. For example, one intriguing possibility hereabouts[7] may be to extend the notion of a 'dynamic core' (originally developed by Tononi and Edelman 1998 as part of an account of what enters conscious awareness). The dynamic core is a highly integrated functional cluster of neural circuits, defined in such a way that:

> "The term…deliberately does not refer to a unique, invariant set of brain areas..and the core may change in composition over time" Tononi and Edelman (1998)

The core is marked by extremely high integration, rigorously defined in terms of mutual influence, between the contributing parts. Perhaps, then, we may similarly display certain non-biological elements as (at times) suitably causally intertwined with biological ones so as to create profoundly (but temporarily) integrated systems of reasoning, action, and control. Both Mackenzie and Juarrero

suggest, in different ways, that very fine details of timing and feedback loops may be part of the answer to the riddle of soft assembled unity.

An important question would remain, however, concerning the role of the more insulated, semi-autonomous sub-systems (both onboard and biological and offboard and technological) that also play a role in making us who and what we are. These semi-autonomous resources (beautifully captured by Terry Dartnall in his image of the Bioborg) are not densely integrated with the dynamic core (if they were, they'd be part of it), and correspond, in Tononi and Edelman's treatment, to non-conscious neural processing. One of the big unresolved puzzles of NBC is, I think, how best to display these elements as more than simple (internal and external) tools, while respecting their semi-autonomous nature. Wayne Christiansen, in the paper I mentioned earlier, notes (concerning the neural realm) that "a balance of modularity and integration is required in order to produce behavior that is diverse but coherent". I believe that one key to understanding how nature makes cognitive agents is to understand the general principles of this balancing act. When we do so, I suspect we will find many of the same principles at work on larger-scale ensembles, enabling extended cognitive systems to find and occupy the sweet spots between full integration and unstable aggregation. That doesn't answer Juarrero's well-aimed question, alas. But it does show where I am inclined to look.

In closing, I'd like to thank the reviewers for this treasure trove of exciting suggestions and important challenges. They cement my conviction that this is an exciting and transformative time for the sciences of the mind. As technological progress provides new tools and new puzzles at about an equal pace, the time is ripe to begin to put together the many pieces of the puzzle of mind. That means, I firmly believe, seeing our unique cultural and technological scaffoldings as not just aids for understanding the mind, but as key parts of the minds we seek to understand.

Department of Philosophy, and

Program in Cognitive Science,
Indiana University
Bloomington, IN
USA

References

Berti A, Frassinetti F. (2000) When far becomes near: re-mapping of space by tool use. Journal of Cognitive Neuroscience: 12; 415-420.

Christiansen, W (forthcoming) "Self-directedness, integration and higher cognition" *Language Sciences (special issue on Mind and World, D. Spurrett (ed))*

Dennett, D (1991) *Consciousness Explained* (Little Brown and Co, Boston)

Hurley, S (1998) "Vehicles, Contents, Conceptual Structure and Externalism" *Analysis* vol 58, no 1 p 1-6

Tononi, G and Edelman, G (1998) "Consciousness and Complexity" *Science* 12/04/98, vol 282, issue 5395

---

[1] The quote, which was brought to my attention by Galen Strawson, is from James Gleick's excellent biography *Genius: Richard Feynman and Modern Physics* (Abacus, London, 1992)

[2] See eg Hurley (1998), Dennett (1991)

[3] Of course, the active structure and the long-term structure need not be the same. All that matters is that the one somehow lead to the other in appropriate circumstances.

[4] Thanks to David Chalmers for pointing this out.

[5] I wish I could claim this to be a deliberate structural echo of the claim about distributed, soft-assembled selves. But actually, it was at the request of the press, who thought that a whole chapter on 'Soft Selves' would be too heavy-going for a trade book!

[6] As Terry himself clearly sees in his fascinating closing discussion (with which I totally agree) of the prospects for avatar-based presence.

[7] This idea was first suggested to me by Damien Sullivan.