

Coupling, Constitution and the Cognitive Kind: A Reply to Adam and Aizawa

Andy Clark

1. Introduction: Crossed Wires?

Adams and Aizawa, in a series of recent and forthcoming papers ((2001), (In Press), (This Volume)) seek to refute, or perhaps merely to terminally embarrass, the friends of the extended mind. One such paper begins with the following illustration:

"Question: Why did the pencil think that $2+2=4$?

Clark's Answer: Because it was coupled to the mathematician"

Adams and Aizawa (this volume) ms p.1

"That" the authors continue "about sums up what is wrong with Clark's extended mind hypothesis". The example of the pencil, they suggest, is just an especially egregious version of a fallacy said to pervade the literature on the extended mind. This fallacy, which they usefully dub the "coupling-constitution fallacy", is attributedⁱ, in varying degrees and manners, to Van Gelder and Port (1995), Clark and Chalmers (1998), Haugeland (1998), Dennett (2000), Clark (2001), Gibbs (2001), and Wilson (2004). The fallacy, of course, is to move from the causal coupling of some object or process to some cognitive agent, to the conclusion that the object or process is part of the cognitive agentⁱⁱ, or part of the agent's cognitive processing (see e.g. Adams and Aizawa (This volume) ms p.2). Proponents of the extended mind and related theses,

Adams and Aizawa repeatedly assert, are prone to this fallacy in part because they either ignore or fail to properly appreciate the importance of "the mark of the cognitive" viz the importance of an account of "what makes something a cognitive agent" (op cit ms p.3). The positive part of Adams and Aizawa's critique then emerges as a combination of the assertion that this "mark of the cognitive" involves the idea that "cognition is constituted by certain sorts of causal process that involve non-derived contents" (e.g. op cit ms p.3) and that these processes look to be characterized by psychological laws that turn out to apply to many internal goings-on but not currently (as a matter of contingent empirical fact) to any processes that take place in non-biological tools and artifacts.

In what follows, I shall try to show why these arguments display nothing so much as mutual failures of communication: crossed wires concealing a couple of real, important, but much more subterranean, disagreements. In particular, I try to show why the negative considerations advanced by Adams and Aizawa fail to successfully undermine the argument for the extended mind, and why their more radical positive story, *unless supplemented by implausible additional claims*, fails to cast doubt on the claim that minds like ours can (without the need for any radically new techniques, technologies or interventions) extend into the world.

Before embarking on all this, a word about the intended force of the argument. Adams and Aizawa make much of their concession (see e.g. their (in press) ms p.1) that mental extension is *possible*, just not, they claim, actual. Theirs, they insist, is a 'contingent intercranialism' applicable to human

agents in the current state of technology. But they seem to imply that my own view (and that of Clark and Chalmers (1998), must, if it is to stand in contrast to this, be that such extension is *rampant*, and that "in ordinary tool use we have instances in which cognitive processes span the cranial boundary and extend into intercranial space" (op cit ms p.2)). Whatever the truth of such a claim (of rampant extension) it is not one made by the present author, nor by Clark and Chalmers (1998). Rather, that claim was that in imaginable circumstances, ones that involved *no giant leaps of technology or technique*, we would be justified in holding that certain mental and cognitive states extended (in a sense to be explained later) into the non-biological world. This leaves it open whether there are such extensions and (if there are) exactly how widespread they are. But it is still far stronger than the mere claim of 'logical possibility' that Adams and Aizawa suggest as the proper alternative to rampant actual extension.

1. The Odd Coupling

Consider the following exchange, loosely modeled on Adams and Aizawa's opening 'reductio':

Question: Why did the V4 neuron think that there was a spiral pattern in the stimulus?

Answer: Because it was coupled to the monkey.

Now clearly, there is something wrong here. But the absurdity lies not in the appeal to coupling but in the idea that a V4 neuronⁱⁱⁱ (or even a group of V4 neurons, or even a whole parietal lobe...) might *itself* be some kind of self-

contained locus of thinking. It is crazy to think that a V4 neuron thinks, and it is (just as Adams and Aizawa imply) crazy to think that a pencil might think. Yet the thrust of Adams and Aizawa's rhetoric is, again and again, to draw attention to the evident absence of cognition *in the putative part* as a way of 'showing' that coupling (even when properly understood- see below) cannot play the kind of role it plays in the standard arguments for cognitive extension. Thus we read that:

"When Clark *makes an object cognitive* when it is connected to a cognitive agent, he is committing an instance of a "coupling-constitution fallacy"

Adams and Aizawa (this volume, ms p.2) my emphasis

But this talk, of an objects being or failing to be 'cognitive' seems almost unintelligible when applied to some putative *part* of a cognitive agent or of a cognitive system. What would it mean for the neuron *or* the pencil to be, as it were, brute factively 'cognitive'? Nor, I think, is this merely an isolated stylistic infelicity on the part of Adams and Aizawa. For the same issue arose many times during personal exchanges^{iv} concerning the vexed case of Otto and his notebook (the example used, with, importantly, a great many riders and qualifications, in Clark and Chalmers (1998)). And it arises again and again, as we shall later see, in various parts of their more recent challenges concerning 'the mark of the cognitive'

Let us first be clear then about the precise role of the appeal to coupling in the arguments for the extended mind. The appeal to coupling is not intended to make any external

object 'cognitive' (insofar as this notion is even intelligible). Rather, it is intended to make some object, that in and of itself is not usefully (perhaps not even intelligibly) thought of as *either cognitive or non-cognitive*, into a *proper part of some cognitive routine*. It is intended, that is to say, to ensure that the putative part is poised to play the kind of role that *itself* ensures its status as part of the *agent's* cognitive routines.

Now, it is certainly true (and this, I think, is one important fact to which Adams and Aizawa's argument might conceivably draw the reader's attention) that not just any old kind of coupling will achieve even this result. But probably no-one in the literature, and certainly not myself, nor Clark and Chalmers, ever claimed otherwise. Hence the presence of the conditions of (broadly speaking) 'glue and trust' pursued at length in Clark and Chalmers (1998), and briefly summarized in various other places, including the target papers by Adams and Aizawa. There is no need to repeat the conditions, even summarily, here, as the present focus is on the overall shape of the argument and on the issues concerning coupling and the mark of the cognitive, rather than on these aspects of the original content. But it is perhaps worth noting that the bulk of our (1998) treatment was devoted not to some general notion of coupling but to the isolation and defence of a very specific set of features.

The biggest of the crossed wires in the exchange with Adams and Aizawa concerns, I think, the role of those features. For Adams and Aizawa often fail to fully appreciate that the conditions (of 'glue and trust') speak to the question (which we deemed fully intelligible) 'when is some physical object or process acting as part of a larger cognitive routine?' and not to the much murkier (probably unintelligible) question

'when should we say, of some such candidate part, that it is *itself* cognitive?'. The question at issue then, was what kind of coupling makes for *incorporation into* a cognitive system rather than simply and mundanely *use by* a cognitive system?

In outlining an answer, we (Clark and Chalmers) chose to be guided by a set of intuitions derived from reflection on the ordinary use of talk of non-occurrent, dispositional beliefs. In essence, we took these intuitions and systematically showed that the kind of functional poise (poise to guide various forms of behaviour) associated with such dispositional believings *on the part of Otto, a previously identified cognitive agent* might sometimes be partially supported by a highly non-standard physical realization in which a notebook (for example) acted as the physical medium of long-term storage. The right kind of coupling to make the external resource into a part of the cognitive system, we argued, was one that poised the information contained in the notebook for sufficiently easy, reliable and automatic 'use' (deployment would be a better word) in much the same way as is typically (though not always) achieved by biological encoding.

Clark and Chalmers thus offered an argument (which one may accept or reject: that is, of course, another matter) concerning conditions not of 'being cognitive' but for incorporation as part of the physical substrate of a cognitive system. In so doing we were not even close, as far as we can see, to committing any simple coupling-constitution fallacy.

We must be cautious, however, for Adams and Aizawa are nonetheless aware that the conclusion we were aiming for is

that the object or process be part of the agent's cognitive apparatus (see e.g. Adams and Aizawa (this volume) ms p.2). The misunderstanding is thus more complex, and ultimately more interesting, than any simple failure to identify the correct target. The deeper, and more interesting, point is that Adams and Aizawa seem to think that some objects or processes, *in virtue of their own nature* (see section 2 below) are, as I shall now put it, *candidate parts* (for inclusion in a cognitive process). Whereas other objects or processes, still in virtue of their own nature, are not. This, I think, must be the way to give sense to that otherwise baffling question "is some X cognitive?" when asked of some putative part. This then is the link between the skirmish concerning a putative coupling-constitution fallacy and their subsequent positive story concerning the 'mark of the cognitive'. Thus the authors ask:

"..if the fact that an object or process X is coupled to a cognitive agent does not entail that X is part of the cognitive agent's cognitive apparatus, what does? *The nature of X of course*. One needs a theory of what makes a process a cognitive process. One needs a theory of the "mark of the cognitive"."

Adams and Aizawa (this volume) ms p.3. (My italics).

It is to this (vexed and vexing) issue that I now turn.

2. On your marks...

So, let's try this again. Maybe Adams and Aizawa's idea is that that V4 neuron *is*, in some intelligible sense, 'cognitive'? Maybe it is cognitive in the sense just identified above: the sense of being, *in virtue of its own nature*, at least a *candidate* for being or becoming a proper part of a genuinely cognitive process. Such, I am at least tempted to think, has to be the underlying belief driving much of Adams and Aizawa's otherwise rather mystifying critique. This puzzling possibility brings us to the more positive part of Adams and Aizawa's discussion viz their enthusiastic appeal to the idea of the 'mark of the cognitive'.

Notice first that this way of reconstructing the debate, if correct, already suggests a major concession to the role of coupling. For assume we find some such acceptable ('in virtue of its own nature') candidate part. Then what settles the question whether that part belongs to this cognitive system, or to that one, or (currently) to no cognitive system at all? It is hard to see just what, apart from appeal to some kind of coupling (perhaps along with some insufficient but broadly historical considerations), could motivate an answer to this subsequent question.

But put that aside. Let's stick, just as Adams and Aizawa insist we should, to the topic of the 'mark of the cognitive'. What could it be that, as they put it "makes a process a cognitive process" ((this volume) ms p.3)? The question is non-trivial, and has, as Adams and Aizawa (somewhat reluctantly) admit, no well-established answer within cognitive science or philosophy of mind. Nonetheless, they tie their colours to what they depict as "a rather orthodox theory of the nature of the cognitive" (Adams and Aizawa (2001) p.52. According to this theory (op cit p.53) "cognition

involves particular kinds of processes involving non-derived representations". This is the line also pursued in Adams and Aizawa (this volume) and (forthcoming). It comprises two distinct elements, just as presented in the quote viz an appeal to non-derived content and an appeal to "particular kinds of process".

Despite its prominence in their account, Adams and Aizawa really tell us very little about what the first of these (non-derived content) might amount to. We learn that it is content that is in some sense intrinsic ((2001) p. 48). We learn that this is to be *contrasted* with, for example, the way a public language symbol gets its content by 'conventional association' (op cit). We are told, in the same place, that Dretske, Fodor, Millikan and others are (sometimes) in search of an adequate theory of such content, and that the combination of a language of thought with some kind of causal/historical account is a hot contender for such an account. Towards the end of all this, however, the authors make a concession which, I elsewhere argue (Clark (this volume)) takes much of the sting out of the tail of the appeal to non-derived content. This is the concession that:

“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) p.50.

As I (wrongly- see below) understood it, this concession allowed that an external resource, none of whose states or processes or stored representations were themselves

intrinsically contentful (assuming we are able to make sense of that notion in some way) might nonetheless be a proper part of some genuinely cognitive process. Otto's notebook, then, to take the obvious example, might be just such a resource, since it is full of inscriptions written in (let's assume) English. Yet Otto's notebook, in the light of this concession, might still figure as part of the supervenience base for some of Otto's dispositional beliefs even while failing itself to be a repository of states with intrinsic content.

Of course, we do not *have* to think of Otto's notebook this way. A more radical response would be to argue that what makes *any* symbol or representation (internal or external) mean what it does is just something about its behaviour-supporting role (and maybe its causal history) within some larger system. We might then hold that when we understand enough about that role (and, perhaps, history) we will see that the encodings in Otto's notebook are in fact on a par with those in his biological memory. In other words, just because the symbols in the notebook happen to look like words of English, and require some degree of interpretative activity when retrieved and used, that need not rule out the possibility that they have also come to satisfy the demands on being, given their role within the larger system, among the physical vehicles of various forms of intrinsic content.

Be that as it may, there is something quite compelling, I want to admit, about the idea that the notebook encodings are all conventional and derivative, and also about the thought that *some parts of any genuinely cognitive system need to trade in representations that are not thus conventional and derivative*. To accept this, however, is not to give up on the

extended mind claim for Otto, unless one also accepts (what seems to me to be an independent and far less plausible assertion) that *no proper part of a properly cognitive system can afford, at any time, to trade solely in conventional representations*. It was this additional claim that, I thought, was being explicitly rejected (and, I felt, quite rightly so) in the above quoted passage from Adams and Aizawa. No surprise, then, that I thought the concession undermined their own argument!

It seems, however, that I was wrong and that Adams and Aizawa do in fact endorse something like this additional (and I think implausible) claim. Thus (this volume, ms p.5) the authors accuse me of not seriously attempting to understand the point of their actual concession, and hence of (incorrectly) taking it as rendering the appeal to non-derived content argumentatively vacuous, at least in the case of the debate concerning extended cognition. *Mea culpa*.

So what went wrong? The original concession was followed by an example to which I paid insufficient attention. The example involved possible non-representational elements in a language of thought encoding, such as 'punctuation marks' and 'parentheses' (see their (2001) p. 50). Such potential elements, they concede, need not count as 'intrinsic representations' *or even as content-bearing*, yet they would still be proper parts of a properly cognitive process. I confess that I simply did not (and still do not) understand this suggestion regarding a Language of Thought encoding (it is repeated in their (this volume) ms p 4-5, without appearing to me to be any clearer). Nonetheless, it is now clear that whatever it may mean, it was not intended to concede the possibility

concerning Otto's notebook that I scouted above. For the authors now clarify their original claim thus:

"Clearly, we mean that if you have a process that involves no intrinsic content, then the [intrinsic content] condition rules that *the process is non-cognitive*"

Adams and Aizawa (this volume) ms p.5).

My emphasis.

As I now understand it then, their position regarding the role of intrinsic content is this: that there may be a process that is a genuinely cognitive process that has as a proper part, some goings-on (such as, presumably, the tokening of the punctuation mark in the LOT, puzzling as this still sounds to me) that themselves do not themselves involve intrinsic, non-derived contents (presumably because those parts-of-the-part now do not involve contents at all). But such a process (the part, not the part-of-a-part!) must still involve at least *some* intrinsic content on pain of failing to be genuinely 'cognitive'. And Otto's notebook (I presume they must then wish to assert) fails even this very slightly weakened test as here (they think) we *have a process that involves no intrinsic content at all*.

But in what sense do we, in the case of Otto's notebook, confront a *process that involves no intrinsic content at all*? It helps to be careful about timing here. The time at which the notebook looks most clearly to be part of some real *process* is during the retrieval and use phase, and at that point in time, there are clearly plenty of states in play, in the larger notebook-including system, that count as intrinsically contentful, even on the Adams and Aizawa model. At runtime, the *process* is not one that trades *solely* in

representations whose contents are derived or conventionally determined. Instead, it is a hybrid process involving a variety of different kinds of representational element, some having derived contents, others not.

What about at other times? Well, at such other times the claim is just that the notebook is part of the supervenience base for some of Otto's dispositional beliefs. What demands does this make on process? It is very hard to say. Perhaps we can at least say this: the very notion of a dispositional belief already makes *implicit reference* to what would happen in *possible run-time situations*. So there is here implicit reference to everything that those run-time processes would involve. The poise of the encodings in the notebook is such that, in the appropriate whole-system runtime circumstances, those encodings participate in extended processes that involve (let's assume) states with intrinsic contents. It is this functional poise that matters, and that (we claimed) fits the notebook to become part of the physical supervenience base for some of Otto's dispositional beliefs.

But suppose, Adams and Aizawa may insist, we put all that 'implicit involvement of runtime process' talk aside and look solely at the (putative) part itself. Surely here we find a resource all of whose contentful states are derived, and doesn't that contravene the requirement concerning intrinsic content? In Clark (2003) and (2005), I offered a thought experiment meant to show that Adams and Aizawa's requirement, as applied to some storage resource *considered out of the context of its runtime role in a larger system*, was too strong and ought to be rejected. The thought experiment concerned beings ('Martians') endowed with an extra biological routine that allowed them to store *bit-mapped*

images of important chunks of visually encountered text. Later on, at will, they could access (and then interpret) this stored text. Surely, I argued, we would have no hesitation in embracing that kind of bit-mapped storage, even prior to an act of retrieval, as part and parcel of the Martian cognitive equipment. But what is stored is just a bit-mapped image of a fully conventional form of external representation. If we accept the Martian memory into the cognitive fold, surely only skin-and-skull based prejudice stops us extending the same courtesy to Otto.

Despite spending time on what I presented as a preliminary (weaker and more complex) example (the one involving reasoning with imagined Venn Diagrams/ Euler Circles^v), Adams and Aizawa (this volume) do not comment on this case^{vi}. Yet it raises, I still believe, at least some of the right issues. Even if we demand the involvement, in any cognitive process, of *at least some* items that bear their contents intrinsically, it is quite unclear how we should distribute this requirement across time and space. The Martian encodings are poised, here and now, to participate in processes that invoke intrinsic contents. So are those in Otto's notebook. Since it is arguably poised that matters, at least where dispositional believing is concerned, it seems that any reasonably plausible form of the requirement involving intrinsic content can, with a little imagination, be met.

The notebook, I am happy to concede, is not, considered all on its own (and as far as I understand this notion at all) 'intrinsically cognitive'. But it *is* a resource whose encodings, at appropriate run-time moments, inform Otto's behaviour in the way characteristic (Clark and Chalmers claimed) of dispositional beliefs. And this is all that matters. Perhaps it is

indeed essential that any truly cognitive *activity* (and hence any genuinely cognitive *agent*) draw on at least some states with intrinsic content. But we have been given no reason at all to accept the further (and crucial) claim that *no proper part* of such a properly cognitive system, considered now in splendid isolation from those crucial runtime wholes in which it participates, can afford to contain only representations lacking intrinsic content. Indeed, I see no reason why we should accept (or even be tempted by) such a further condition. Thus suppose (to take a very different example) that we think that any genuinely moral agent must be able to reason about the good of others. Still, we should not think that every proper part of such an agent (not even every proper part essential to their moral reasoning) must be capable of so doing. Just so, from the requirement (if it is a requirement) that every truly cognitive agent trade in states with intrinsic contents, it cannot follow that every proper part of the cognitive system of such an agent must so trade.

This, to be sure, cuts both ways. As Adams and Aizawa point out in their (forthcoming, ms p.13) "it does not follow from the fact that one has an 'X system' that every component of the system does X". Consider, they suggest, a sound system:

"Not every component produces sounds. The speakers do, but lasers in CD players, amplifiers, volume controls and tone controls do not. Again, not every component of an X system does X"

(op cit, ms p.14)

But this goes no way at all towards demonstrating what Adams and Aizawa intend. Agreed, the mere fact that the notebook and bio-Otto 'form a system' establishes nothing. Perhaps Otto also forms some kind of a system with his garden tools, but that does not make the garden tools part of Otto. But the way to then proceed is surely *not* by asking, of the candidate part, does it somehow 'possess' the characteristic that we now want to ascribe to the resultant overall system. What the example of the sound systems shows, yet again, is simply the surprising extent to which Adams and Aizawa are committed to the usefulness of pressing a question that looks pretty clearly to be among the reddest of all possible herrings. That is the question whether Otto's notebook is (to put the matter bluntly) 'cognitive'. Since what is at issue is (to repeat) whether the notebook might now be part of the local supervenience base for some of Otto's dispositional beliefs (a putative systems-level fact if ever there was one) the status of the notebook itself, as 'cognitive' or 'non-cognitive' is (to whatever extent that idea is even intelligible) simply irrelevant. By contrast, the *precise nature* of the coupling between the notebook and the rest of the Otto system, and the kinds of behaviour and skilled interaction supported by that larger coupled whole, now seem absolutely crucial to how one ought to conceive of the resulting situation.

3. That Cognitive Kind

Consider now the other major part of Adams and Aizawa's challenge. Recall that their suggestion, concerning the 'mark of the cognitive' was that "cognition involves particular kinds of processes involving non-derived representations"

(Adams and Aizawa (2001) p.53). We have, I think, now said all that needs to be said concerning the appeal to non-derived representation. But what about the other part of the clause, the appeal to "particular kinds of process" involving such representations. It is at this point that a new kind of consideration comes into play. This concerns the possible existence of a *characteristic set of causal processes* found, by painstaking empirical investigation, to pervade the internal, biologically-supported aspects of human cognitive architecture. The operation of these signature causal processes, the authors claim, give rise to a number of laws and regularities that seem to apply to (these known) cognitive processes, but that do not apply elsewhere (for example, to Otto's notebook). In the light of this, Adams and Aizawa ask, shouldn't we judge that the notebook falls outside the class of the cognitive? We should indeed do so, they claim, because "the cognitive must be discriminated on the basis of underlying causal processes" (Adams and Aizawa (2001) p. 52). Here too, then, we must unfortunately grapple with the murky appeal to some kind of apparently self-standing (i.e. non-systemic) notion of "the cognitive". Only this time the notion is linked to the specific laws and regularities characteristic of the internal, biological routines running in (earthly) cognitive agents.

The kinds of law and regularity the authors have in mind here include the pervasiveness, in human (biological) memory systems of effects of chunking, priming, recency etc (Adams and Aizawa (2001) p. 61) and in human perceptual systems of various psychophysical laws (such as Weber's law, *op cit*). Given that science has uncovered these (undeniably important and interesting) regularities, what

does this imply concerning the nature of cognition? According to Adams and Aizawa, the proper conclusion is that:

"the weight of empirical evidence supports the view that, as a matter of empirical fact, there are processes that are (a) recognizably cognitive, (b) take place in the brain, (c) do not take place outside of the brain and (d) do not cross from the brain into the external world"

Adams and Aizawa (this volume) ms p.4

Quite so. Or rather, quite so up until (d) where we again confront the thorny issue of processes, parts and the requirements upon parts. For while specific neural processes and their characteristic properties clearly do not cross over into the non-biological world, there may exist (according to friends of the extended mind) overarching processes that include (some of the) neural ones *and* some non-neural ones, *and* that play the right kind of role in guiding and enabling behaviour to count as part of the physical base for cognition.

Thus recall that opening salvo concerning the mathematician's pencil. The very next sentence read:

"Clark apparently thinks that the nature of the processes internal to a pencil, Rolodex, computer, cell phone, piece of string or whatever, has nothing to do with whether that thing carries out cognitive processing"

Adams and Aizawa (this volume) ms p.1

It is now clearer what is at stake. Adams and Aizawa think that empirical investigations have turned up a number of

features (e.g. priming effects in the case of memory) that reflect the operation, in some parts of the physical universe, of processes internal to those parts. Since these parts support our paradigm cases of terrestrial cognition, we should believe (defeasibly, but justifiably on the basis of current evidence) that these kinds of causal process are essential to the cognitive status (again, just try hard to understand something by this notion) of such parts. It is *the nature of the processes internal to the part* that, so the argument goes, must determine whether it meets the conditions for inclusion into the ranks of the cognitive.

But this is something the extended mind theorist might very reasonably deny. It seems very plausible, for example, that there is no part of the physical universe so devoid of potentially computationally useful properties that that part could not, under some conceivable circumstances, participate as a crucial element in some recognizably computational process upon which some cognitive state of some being supervenes. Whether a candidate part has the 'right nature' seems, in such cases, to have much more to do with the rest of the system (and what it can and can't do in the absence of that part) than with any intrinsic properties of the part itself.

Perhaps Adams and Aizawa will press the question, How do we know the state, in the scenario above, to be cognitive? On their account, we do so (at least in part) by asking to what extent it shares in the casual processes so far identified as characteristic of terrestrial biological cognition. But they surely cannot hold this as a general model of 'cognition-spotting' since it rules out the discovery of *new* signature processes, even of the internal, earth-bound, biological kind.

Nor, I would have thought, can they hold that what goes for internal, biological, earth-bound cognition need be true of cognition tout court. The notion of the cognitive is surely bigger than that. If that special bit-mapped Martian memory, or even the whole of Martian memory, does not exhibit priming and recency effects, should we conclude that it is not memory at all or just that Martian remembering (I assume it is at least some more generic kind of memory) is 'not cognitive'?

But what, Adams and Aizawa will by now be shouting in frustration, *makes* a process cognitive? I haven't said. What makes a process cognitive, it seems to me, is that it supports genuinely intelligent behaviour. This is obviously unhelpful, though it is almost certainly the kind of reply that would be given by, say, the average neuroscientist or cognitive psychologist. Surely no psychologist or neuroscientist would instead assert, for example, that what makes some candidate process cognitive is that it exhibits effects of recency and priming! To identify cognitive processes as those processes, however many and varied, that support intelligent behaviour may be the best we can do. To argue from the other direction, and to identify cognitive processes as those that happen to characterize the neural activity of human agents, is to risk both an unwarranted narrowing of focus (to the neural) and a dangerous and unappealing chauvinism to boot.

One (though certainly not the only) alternative is, to paraphrase Dennett, that cognition is as cognition does. That is to say, we might individuate the cognitive^{vii} by its characteristic effects, not by its characteristic causes. The

notion of a cognitive process, if that were correct, might be best unpacked as the notion of a process that supports certain kinds of behaviour. Such a notion easily allows Otto's notebook, in virtue of its gross behavioral effects, to count as part of the local supervenience base for Otto's cognizings, and, more specifically, for some of his dispositional believings. Adams and Aizawa (2001, p.56) reject such a move as mere behaviorism. But -and this is surely crucial- in between these two extremes (appeal to sameness of gross behavioral effects and appeal to fine-grained sameness of information-processing profile) lie a world of possibilities. Where the ascription of dispositional belief is concerned, it might reasonably be argued, what counts is (amongst other things) the way information is poised to guide reasoning and behavior. This may include the drawing of inferences that never result, and perhaps never could result^{viii}, in overt actions. And this, by most standards at least, is not behaviorism but merely a somewhat coarse-grained kind of functionalism (for some more discussion, see Clark (this volume) section 2).

We are now circling what may be the intractable nub of the problem. For Adams and Aizawa^{ix} are mightily impressed by the clear differences that exist between many of the goings-on found (so far) inside the bounds of skin and skull and the kinds of goings-on found in artifacts such as notebooks. And they invite us, repeatedly, to focus our attention on *the nature of the notebook* (to stick with that tired example) and thus hope to persuade us that it is 'not cognitive', that it fails to partake of the 'mark of the cognitive' and so on. That such differences (between, lets say, the notebook and the neo-cortex) exist, no-one should deny. But some of us are *equally* impressed by our apparent

capacity to form extended computational systems that profoundly factor in both sets of distinctive contributions, creating hybrid wholes that look to support brand new kinds of cognitive capacities. In the case of Otto, the new capacity is just a coarse functional simulacrum of his damaged biological memory capacity. In other cases, the new capacities might be more genuinely novel. But what matters, in every instance, is (1) the degree of complementarity (between the different contributions) and (2) the degree of agent-appropriate integration achieved. Given sufficient complementarity and integration, it becomes plausible (I argue) to treat the resultant system as a cognitive whole, with cognitive properties that supervene on more than the biological components alone. The extended mind is thus a kind of extended functionalism. Or rather, it is a form of normal ('unextended') functionalism taken to its logical conclusion.

As a brief aside, I tend to believe, though nothing in the argument for the extended mind hangs upon it, that large chunks of the internal, biological processing that goes on in us humans (though not in other animals) consists not in the manipulation of items bearing intrinsic content but in the manipulation of a variety of pointers, symbols and markers inherited rather directly from public language itself. These would be image-like neural encodings of public language words and symbols able to act, within the inner realm itself, in many of the same ways as external public language encodings act on, and empower, us. As a result, I believe that the kind of functional complementarity that, on the extended mind model, explains the power of integrated systems of internal and external resources *also* explains much of the apparently unique power of purely internal

human cognition. For more on this idea, see Clark (1998) (in press-a).

Returning to the main matter of complementary internal and external (non-biological) resources, it is important to notice that attention to larger systemic wholes in no way precludes a proper investigation of the special features of various parts, aspects, and components. A useful partial comparison is with the move towards systems-level neuroscience^x. For much of the century, most serious neuroscientific knowledge concerned the responses and behaviors of single cells. Then, with the advent of new techniques of recording, intervention and investigation, attention began to be devoted to understanding the neural dynamics of whole populations of cells, and the distinctive processing styles of different gross anatomical elements (such as the hippocampus). Contemporary neuroscience, courtesy of still-newer techniques of imaging and analysis, and by using increasingly bio-realistic neural network simulations, is just beginning to make progress in understanding some of the key features and properties of even larger-scale neural systems, whole processing cycles that involve the temporally evolving, often highly re-entrant, activity of multiple populations of neurons spanning a variety of brain areas. Note that the advent of true systems-level neuroscience will not (and should not) imply the inappropriateness of investigations that target the special properties and features of distinct cell-types or of distinct populations, or of distinct neural areas. But it must add to these investigations a new sensitivity to the added value created by processing cycles that include multiple complementary operations, performed at various time-scales and using various kinds of neural resource, and whose integrated action is responsible for

much of the power and scope of an individual human intelligence.

The notion of the extended mind is thus really nothing more than the notion of systems-level cognitive (rather than neuro-) science. All it adds to that notion is some discussion, adverting to the details of bio/artifactual coupling, meant to make it plausible to treat some of these larger scale systems as the local supervenience base for the knowledge and cognitive capacities of a specific agent. This added wrinkle was not necessary in the move towards systems-level neuroscience as the old prejudices concerning the bounds of skin and skull there worked in favour of the 'obvious' appropriateness of that larger-scale investigation.

But imagine a world (call it Hippo-world) in which for half a century, all neuroscientific attention was focused on the hippocampus, regarded (for some strange historical reason let's assume) as the obvious locus of human cognitive activity. Specific features of hippocampal processing and encoding are discovered and publicized. One day, a few researchers turn their attention to the rest of the brain. They discover many new and interesting features, and begin to talk about the larger processing circuits that link (for example) hippocampal and neo-cortical processing, and the way certain memory phenomena seem to depend on the complex interactions between the two components. But there is a problem. Some philosophers in Hippo-world believe that in discovering the characteristic causal processes that operate in the hippocampus, they were discovering *the scientific essence of cognition itself*. Better, they now insist, to view *what the hippocampus does as cognitive* and the rest of the brain as merely sending inputs to, or receiving outputs from,

that 'truly cognitive part'. Only the hippocampus, they suggest, exhibits the 'mark of the cognitive'. These other parts, after all, just don't do the same things as the hippocampus, so why regard what they do as cognitive? Others demur, for much of what they see as gross intelligent human behaviour seems to depend just as much upon the special features and properties of the other parts as upon the (important but limited) contribution of the hippocampus itself. Hippo-world begins a public debate on what they clumsily dub "the extended brain". The jury remains out.

One important challenge, for those Hippo-worlders who want to treat the whole brain as a cognitive organ, concerns the question of 'added value'. What do we gain, they are asked, by challenging commonsense and starting to speak of extra-hippocampal activity as part of the physical base for cognition? Can't we explain all that anyone actually does by treating what the hippocampus does as cognitive and the rest as (perhaps instrumentally useful but) non-cognitive? As long as we note what actually gets done, and are sensitive to how information flows through the whole system, this will work fine won't it? Isn't all we need, to paraphrase Rupert (2004) the 'hypothesis of the embedded hippocampus'?

I think the answer to this question must be 'yes'. We could, if we so wished, carve up the contributions in the way suggested. And this may well have the advantage of not challenging commonsense (as it had apparently developed on Hippo-world). But by the same token, if we (standing, admittedly, on Earth, not Hippo-world!) accept the vision of the whole brain as a locus of processing cycles that include multiple complementary operations, performed using

various kinds of neural resource, and whose integrated action is responsible for much of the power and scope of an individual human intelligence, there is a clear case for accusing Hippo-world commonsense of displaying a needlessly restricted vision of cognitive processing. Yet the description just given works every bit as well for the case of the extended mind. There, the idea is that the brain-body-world system is sometimes (when the right coupling conditions are met) the locus of processing cycles that include multiple complementary operations, performed using various kinds of neural resource, and whose integrated action is responsible for much of the power and scope of an individual human intelligence.

The challenge of added-value thus cuts both ways. For what is the added value, one may ask, in *not* embracing these visions of larger systemic wholes? No-one, after all, is suggesting that such larger visions preclude investigation of the special features and properties of any of the parts. Just as systems-level neuroscience should not be seen as a threat to single-cell neuroscience, or indeed to the study of the hippocampus, so systems-level cognitive science should not be seen as a threat to neuroscience or to the study of the special features and properties of the biological brain. In fact, one of those special features and properties, neural plasticity, is visibly crucial to the brain's astounding ability to enter into the most profound forms of cognitive extension^{xi} in the first place.

4. Beneath the Surface

Perhaps we have now had enough of the cut and thrust of argument and counter-argument. More helpful, in the long run, would be a genuine attempt to see (and ideally to accommodate) what moves each side in this debate. Underlying Adams and Aizawa's critique of (the arguments for) extended cognition is, I suspect, a quite reasonable, though subtly misplaced, fear. It is the fear of *undervaluing the cognitive core*. The cognitive core, as I shall use the term, is the set of basic skills of perceiving and learning that we share with many other earthly animals. At the heart of this skill-set lie the rich abilities of subtle pattern-recognition that allow us to learn about important regularities in our environment by exposure to repeated exemplars. In combination with affective and motivational systems, this kind of potent, slow, pattern-based learning enables many animals, ourselves included, to learn to deal with highly complex situations in a remarkably nuanced and efficient manner.

I suspect that the real fear underlying much opposition to the extended mind, and the explanation of the quite palpable unease it causes even in some of its most sensitive critics, is that by celebrating the power of new, hybrid, extended systems we lose sight of that crucial cognitive core^{xii}. The fear would be that to embrace hybrid cognitive forms is to lose sight of the unique importance of the core systems upon whose successful operation the very possibility of such forms depends. But such fears are groundless. It is no part of the agenda of the extended mind^{xiii} to attempt to wash out all the differences between various internal and external contributions, nor to downplay or undervalue the (currently) unique contribution of the cognitive core. Indeed, the actual research program of the extended mind is committed, above

all, to plotting and charting the varied contributions made by a variety of biological and non-biological resources, and the potent and multi-layered interactions between them. The agenda is thus not a negative but a purely positive one: to understand the larger systemic webs that, spun around the common core shared with so many other animals, help to give human cognition its *distinctive* power, character and charm.

Consider, by way of partial analogy, the more mundane fact that human animals, apparently uniquely on the planet, display (in addition to the common core) a second, rather different set of skills. These are the skills of explicit, deliberative, 'language-infected' (see e.g. Dennett (1996)) reason and planning. Working together, these two very different sets of skills make us into especially potent cognitive engines. Nonetheless, if we contemplate these two kinds of cognitive resource, it seems compelling that in some very important sense, it is the skills of basic pattern-recognition, learning and affectively tuned response that are the most fundamental. By this I mean only that without these we would probably be unable to have thoughts at all, and, ipso facto, unable to have linguistically-infected thoughts. But notice that, even for the staunchest cognitive internalist, it does not follow from this that those additional, linguistically-infected modes of thinking are not precisely that: modes of thinking. One can (and should) celebrate the power of explicit, deliberative, linguistically-infected thought without thereby undervaluing the (in one sense) more fundamental contribution of the cognitive core we share with non-linguistic animals.

This very same model (of an essential core with some mind-bogglingly potent add-ons) can then be invoked by the friends of the extended mind. It is surely entirely likely that many of the kinds of extended cognitive system mooted in the literature on the extended mind are *in just the same sense* less fundamental. They are less fundamental in that no genuinely cognitive system could consist *entirely* of the most typical kinds of external resources^{xiv} that (currently) augment the common core. This, I think, is the important grain of truth underlying Adams and Aizawa's arguments concerning derived contents, conventional encodings, the 'non-cognitive' status of notebooks, etc etc. It is a grain of truth, however, that is no more damaging to the vision of the extended mind than it is to the vision of the language-infected mind. In each case, powerful new cognitive wholes are brought into being on the back of some set of more basic, and perhaps even cognitively indispensable, skills and capacities. And in each case, the new integrated systems that result are cognitive systems in their own right. They are, moreover, the cognitive systems whose operation accounts for many of the most unique and characteristic achievements of the human mind.

Conclusions: Watering the Desert

Adams and Aizawa's challenge to the extended mind is rather like a challenge that might be posed to a theorist of irrigation. Take some putative part of a process of irrigation and ask yourself, is that part *irrigative*? To push the question, demand of the theorist of irrigation an account of the 'mark of the irrigative' and then ask whether some putative part of some process of irrigation shares in that mark.

We should not, I think, like to approach the matter of an irrigation system in this way. What we want to know, of some putative proper part of such a system, is whether it contributes to the functional whole. We may ask ourselves, for example, whether it enables that functional whole to irrigate land that it could not otherwise reach. To the extent that the answer is positive, the part (*ceteris paribus* of course) looks to be part of the system of irrigation, regardless of whether water drizzles out of it.

Just so, there is surely little long-term value in pursuing the question, asked of Otto's much-maligned notebook "is it cognitive"? Instead, we must attend (as Clark and Chalmers did attend) to the role of the notebook in the larger organization of which biological Otto is a part. Then we can ask questions such as, Does the notebook enable this larger system to exhibit the kinds of behavioral regularity characteristic of an individual's dispositionally believing that such-and-such? Such a move towards a larger systemic focus is familiar and sometimes helpful. The study of the extended mind presents, I suggest, no *greater* theoretical or practical difficulties than those, significant as they were, that attended the move towards a systems-level neuroscience. And it is justified (or so I believe) in very much the same way. In each case, we confront densely integrated larger-scale organizations that support some of the kinds of intelligent behaviour most characteristic of our species. Systems-level neuroscience, however, had one enormous advantage. It could for the most part simply help itself to the idea of an individual, sufficiently unified, cognizer. Extended mind theorists cannot. Instead, the incorporation of a non-biological resource into the cognitive processing of

an individual requires that certain kinds of coupling between biological and non-biological resources be present. Absent the right kinds of coupling, of course, and even inner biological going-on (replete, let's assume, with all the currently available 'marks of the cognitive') would not count as part of the cognitive activity *of that very agent*. When such couplings are in place, however, the bounds of skin and skull are rendered functionally irrelevant and cognition extends gracefully into the world.

References

Adams, F. and Aizawa, K (This Volume) Defending the Bounds of Cognition

Adams, F. and Aizawa, K (In Press) Why the Mind is Still I the Head in M. Aydede and P. Robbins (eds) *Cambridge Handbook of Situated Cognition* (Cambridge University Press)

Adams, F. and Aizawa, K. 2001. The Bounds of Cognition. *Philosophical Psychology* 14:1: 43-64.

Clark, A (this volume) Memento's Revenge: The Extended Mind, Extended.

Clark, A (In Press-a) Word, Niche and Super-Niche: How Language Makes Minds Matter More *Theoria*

Clark, A (In Press-b) Re-inventing Ourselves: The Plasticity of Embodiment, Sensing, and Mind
Journal Of Philosophy And Medicine.

Clark, A (2005) "Intrinsic Content, Active Memory, and the Extended Mind" *Analysis* 65:1: p.1-11

Clark, A. 2003. *Natural-Born Cyborgs: Minds, Technologies And The Future Of Human Intelligence*. Oxford University Press, NY.

Clark, A (2001) Reasons, Robots, and the Extended Mind
Mind and Language 16, 121-145

Clark, A (1998) Magic Words: How Language Augments Human Computation, in P. Carruthers and J. Boucher (Eds) *Language And Thought: Interdisciplinary Themes* (Cambridge University Press: Cambridge,) P.162-183

Clark, A, and Chalmers, D. 1998. The Extended Mind.
Analysis 58:1: 7–19.

Reprinted *In The Philosopher's Annual* vol 21-1998.
Ridgeview, 2000: 59-74.

Also reprinted in D. Chalmers (ed) *Philosophy Of Mind: Classical and Contemporary Readings*. Oxford University Press, 2002.

Dennett, D (2000) Making Tools for Thinking. In Sperber, D (ed) *Metarepresentations: A multidisciplinary Perspective* (Oxford University Press, Oxford) 43-62

Gibbs, R (2001) Intentions as Emergent Products of Social Interactions in Malle, B, Moses, F., Louis, J., and Bladwin, D. (eds) *Intentions and Intentionality* (MIT Press, Camb. MA) 105-122

Haugeland, J. 1998. Mind Embodied and Embedded, in J. Haugeland (ed) *Having Thought: Essays in the Metaphysics of Mind*. Harvard University Press, MA.

Mundale, J. (2001) Neuroanatomical Foundations of Cognition: Connecting the Neuronal Level with the Study of Higher Brain Areas, In (eds) W. Bechtel, P. Mandik, J. Mundale, and R. S. Stufflebeam *Philosophy and the Neurosciences: A reader* (Basil Blackwell)

Mundale, J (2002) Concepts of Localization: Balkanization in the Brain *Brain and Mind* 3: 1-18

Rupert, R (this volume) Representation in Extended Cognitive Systems: Does the Scaffolding of Language Extend the Mind?

Rupert, R (2004) Challenges to the Hypothesis of Extended Cognition *Journal of Philosophy* 101:8: 389-428

Sutton, J (this volume) Exograms and Interdisciplinarity: History, the Extended Mind and the Civilizing Process

Van Gelder, T and Port, R (1995) (eds) *Mind as Motion* (MIT Press, Camb. MA)

Wheeler, M and Clark, A 1999 Genic Representation: Reconciling Content and Causal Complexity *British Journal For The Philosophy Of Science* 50:1:103-135

Wilson, R. A. (this volume) Meaning Making and the Mind of the Externalist

Wilson, R. A. (2004) *Boundaries of the Mind: The Individual in the Fragile Sciences--Cognition* (Cambridge University Press, Cambridge, UK)

ⁱ These attributions are all fully explicit, but are spread across the three papers ((2001) (in press) (this volume)) mentioned at the start.

ⁱⁱ The need for some kind of 'nontriviality clause' (to borrow a phrase from Rupert (this volume) is actually widely recognized in the literature on the

extended mind. If it were not, then arguments for the extended mind could have been very short indeed! The requirement is made explicit in, for example Wheeler and Clark (1999) p.110

ⁱⁱⁱ Thanks to Rob Wilson for the single-neuron comparison.

^{iv} Thus Ken Aizawa, after a long series of exchanges, asks "so, you really agree with us that the notebook is non-cognitive?", as if an affirmative answer were incompatible with the extended mind thesis. Yet insofar as the question is even intelligible, we would indeed reply that the notebook, considered alone, is 'non-cognitive', *just like a neuron or group of neurons*.

^v Adams and Aizawa (this volume) devote much space to arguing that the case of the Euler Circles) fails to meet their condition, properly understood, and they are right to do so. I offered it only as a case where *some* proper aspects of a genuinely mental process seem to trade in representations whose meanings are conventional. This, after all, was how I saw the case of Otto's notebook (more on which in the text).

^{vi} In a personal communication, Ken Aizawa did offer a tentative response, on behalf of Adams and Aizawa as follows:

"We aren't sure what basis Clark has for saying that these bit-mapped images are part of the Martian cognitive economy, since we aren't sure what he thinks is the mark of the cognitive. Perhaps this is just Clark putting some of his intuitions on the table. Be this as it may, we think it is perfectly reasonable for us to stand by the view that these Martian representational states are not cognitive states. We have a theory of what cognition involves. In brief, we think involves non-derived representations covered by *ceteris paribus* psychological laws".

I hope that the Hippo-world example, developed in the present text, will go some way towards making us uneasy about such appeals to the current state of knowledge. Interestingly, in the same personal communication, Aizawa is clear that "future scientific developments could undermine our theory and force revisions". Perhaps, then, the jury simply remains out until scientific theorizing proves or disproves the value of looking at larger systemic wholes involving biological and non-biological resources.

^{vii} Note that this is a much broader notion than that of the conscious, which may be individuated in ways that appeal to much more than characteristic effects or (what comes to the same thing) characteristic kinds of functional poise.

^{viii} Given some sufficiently complex web of other beliefs and desires.

^{ix} See also Rupert (2004)

^x For a useful survey, see Mundale (2001), and for more discussion, Mundale (2002)

^{xi} See Clark (2003) (In Press)

^{xii} This worry also seems to be at work in Rupert (2004) (this volume).

^{xiii} For some clear and compelling expositions of this kind of view, see Sutton (this volume), Wilson (this volume).

^{xiv} Unless the additional resource is another biological agent of course!