

## **Extending the mind – literally.**

### *Introduction*

Since the appearance of Clark and Chalmer's paper 'The Extended Mind' in 1998 the concept of 'extendedness' has itself been extended and developed in multifarious ways to include other elements of so-called 'nouvelle cognitive science', such as 'embodiedness', 'embeddedness' and 'enactivity'. Clark's role in this development has been significant. However, it might be argued that his distinctive contributions (in Eg. 'Being There' and 'Natural Born Cyborgs' ) have remained rooted in the particular perspective outlined in his and Chalmer's original paper. It therefore seems worthwhile returning to this paper, which provides a clear and succinct account of that perspective, as a prelude to a wider investigation of the concept of 'extendedness' in Clark's work and elsewhere. What follows is a critical assessment of this original paper. In it I concentrate on what I take to be flaws in C&C's approach. It should not be inferred from this that I think their paper only deserving of criticism, but merely that it is more convenient just to concentrate on these aspects here.

I will begin by attempting to summarise what I take to be C&C's main argument.

### *A Summary of the Main Argument*

Let's take as given that cognitive processes are traditionally regarded as occurring "in the head" and only "in the head". Now it transpires that some of the tasks which can be completed using cognitive processes (traditionally so called) can also be completed using a mixture of cognitive processes and manipulation of the external environment. One can, for example, assess the 'fit' of 2 dimensional shapes into 2 dimensional sockets in a game of Tetris either by imagining them rotated in various ways or by actually rotating them using a rotate button. Similarly, when playing Scrabble, one can either 'mentally' rearrange letters to create a new word or one can physically rearrange the tiles.

In cases such as these, say C&C, there is no reason why taking the physical manipulation route, rather than the imagination route, should be seen as choosing the 'non-cognitive' option:

*If as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process. (p.2)*

The traditional conception of the cognitive, then, is wrong. Echoing Putnam, C&C say that "Cognitive processes ain't (all) in the head".

This is more or less all there is to the argument. However, in keeping with the title of their paper, C&C want to extend its scope beyond that of cognitive processes to cover instances of mental states such as beliefs. To this end they devised the now well-known scenario involving 'Otto' and 'Inga'. Otto has Alzheimer's disease, thus impairing his memory. In order to help him remember

things he writes notes down in a notebook which he carries with him. Included amongst his notes are addresses of buildings. Unlike Otto, Inga has normal memory function. She and Otto hear separately of an exhibition at the museum and both decide to go. Inga pauses to recall where the museum is located (53<sup>rd</sup> Street) before walking there. Otto looks up the address in his notebook before walking there. In both cases we can say, according to C&C, that the individual concerned wanted to go to the museum and believed that the museum was on 53<sup>rd</sup> Street. We can say this because “the notebook plays for Otto the same role that memory plays for Inga”(p.7). Moreover, in both cases it is legitimate to characterize the individuals concerned as having had their beliefs before they consulted their respective information sources; we can say of Inga that “the belief was sitting somewhere in memory waiting to be accessed” (p.6) and there is no reason why we cannot say the same of Otto with his notebook.

Having developed this notion of extended belief, C&C conclude matters by indulging in some speculation about further possible applications of their thesis. “Socially extended cognition” is one such possibility. If having a belief can be partly constituted by having a notebook it could also be partly constituted by having a human being who is as readily accessible and reliable as a notebook. “One’s beliefs might naturally be seen to be embodied in one’s secretary”(p.10) for example.

Another such possibility is “the extended self”. Our mental states (such as belief) make up a large part of who we are, say C&C, so if we can talk about these being extended it would seem to follow that we can talk about the self being extended. Two possible consequences of this are that “interfering with someone’s environment will have the same moral significance as interfering with their person” and that “certain forms of social activity might be reconceived as less akin to communication and action and more akin to thought” (p.11)

So much, then, for the summary of C&C’s paper. We might note in passing that certain components of the argument – namely those that deal with what might be called “science fiction” scenarios – have been omitted. I will have something to say about these scenarios later in the paper, but it is my contention that they are not essential to the argument and for this reason I have not given them space here.

### *A Problem with the Logic*

An initial objection to C&C’s argument is that the reasoning would appear to be flawed. The structure of the argument, as I understand it, is:

- 1) Things that I do in my head are sometimes ‘the same’ as things that I do in the external world.
- 2) When I do these things in my head they are called ‘cognitive’.
- 3) Therefore they should be called ‘cognitive’ when I don’t do them in my head.

It is not difficult to see a problem here. Given that we accept 1) and 2) why should 3) follow? The argument seems to be grounded on the general principle that if 2 entities (or processes or whatever) have one thing in common then they must have other things in common too. (Other predicates must be true of them). This is like arguing that: 1) Both my first and second wives

made me unhappy. 2) My first wife was a Catholic. 3) Therefore my second wife must be a Catholic too.

Perhaps some might object here that my first wife's being Catholic played no part in her making me unhappy and that therefore this comparison is not a good one. In the 'cognition' case, it might be said, it was precisely *because* of the shared properties of eg internal and external manipulation of Tetris tiles that we were tempted to attribute a further shared 'cognitive' property. But let's suppose it was exactly and only my first wife's 'Catholicness' that made me unhappy. This still wouldn't give me good grounds for calling my second wife a Catholic.

What is the practical upshot of this logical confusion? Briefly we might say that just as it might not only be 'Catholicness' in wives that makes me unhappy so also it need not only be 'cognitive processes' that help me get a 'cognitive' task done. It is uncontentious that physical manipulation of the environment can aid and abet cognitive functioning. This text is being typed up in 'Word' from a rough hand-written draft. As I type I am making significant alterations to the original and I am enabled to do that partly because seeing my words appear in neat rows and paragraphs on the computer screen helps me get a clearer picture of my own thoughts and acts as a stimulus to further thought. There is, we might say, a fairly fluid interactive process going on between my cognitive processes and my use of the computer. In this way we can say that the computer is important to my cognitive processes. What we are describing here is what Clark, following Vygotsky, has described as 'scaffolding' (eg. In 1997 p.45-47). But to say that the computer is important to my cognitive processes is not to say that it actually *is* part of my cognitive processes. Logically this just doesn't follow. This point is made by Adams and Aizawa (2001) when they assert:

*A process P may actively interact with its environment but this does not mean that P extends into the environment.* (Adams and Aizawa. p.56)

### *A Problem with 'The Same'*

We have said, then, that there is a problem with the logic of the argument put forward to establish that cognitive (and other mental) processes can be 'extended'. However such objections, even if correct, need not be fatal. If it is true that the fact that we use external processes to augment our inner activity is not enough to show that these external processes are part of our inner activity, it is also true that it is not enough to show that they *aren't* part of our inner activity. ( $\sim (A \rightarrow B)$  is not equivalent to  $A \rightarrow \sim B$ ). It may be that C&C have used flawed/inadequate reasoning to establish something that is in fact true anyway.

At this point, though, a second problem emerges. In arguing that the logic or the argument was questionable we had to skate over the fact that it is not really clear what the first premiss of that argument could mean:

- 1) Things that I do in my head are sometimes 'the same' as things that I do in the external world.

The question we want to ask here is "what is meant by 'the same'?" To examine this question we might take Otto and his notebook as a starting point. We are enabled to say that Otto has 'the same' beliefs as Inga because "the notebook plays for Otto the same role that memory plays for

Inga”(p.7), the idea being that just as Inga looks up information in her long term memory so Otto looks up information in his notebook. We will pass by here the highly questionable assumption that we do – in this particular instance - want to attribute the same beliefs to Otto that we attribute to Inga (surely the very fact that he has Alzheimer’s Disease makes Otto’s situation vis-à-vis beliefs a non-standard one?), in order to investigate the question of what the notebook playing “the same” role as Inga’s memory might mean.

A first port of call might be Adams&Azaiwa (2001) (henceforth ‘A&A’) who deny that Otto’s situation with his notebook *is* relevantly the same as that of Inga with her memory. They have two main objections. The first is that “where the symbols written in Otto’s notebook have merely derived content the recollection in Inga’s brain has non-derived content”. The second is “the fact that Inga and Otto carry out distinct processes in coming to arrive at the museum” an example being that “Otto’s ‘memory recall’ involves visual processing for turning to the appropriate page in the notebook” where Inga’s doesn’t. These are both senses in which A&A feel that the notion that Otto’s notebook plays ‘the same’ role as Inga’s memory, can be challenged.

I am not sure that either hold up. With the first objection there seems to be something doubtful about the construction “non-derived content”. It is clear what it might mean to say that the symbols in Otto’s notebook have “derived content” but it is not at all clear that one can create anything coherent out of its opposite and postulate that that’s what cognition involves. Here we are perhaps reminded of the anti-physicalist who driven into a corner by physicalist arguments asserts that mind is an “immaterial substance”. Likewise, A&A, perhaps driven into a corner by Searlean arguments that computationalism must be wrong because symbolic representations aren’t intrinsic, seem to be responding by conjuring up the genie of “intrinsic representations”. As for the second objection, the fact that Otto’s notebook-involving processes are not the same as Inga’s non-notebook-involving processes could surely be countered by the argument that the processes were ‘the same’ in all relevant respects. The notion of ‘relevant’ here of course begs the question, but it seems like a question that could be answered within the parameters that A&A set. One could argue, for instance, that the additional processes which attend the using of a notebook were not part of the core process of ‘recall’. Or if one did not want to take this tack one could perhaps argue for a disjunctive definition of what counts as ‘the same’ cognitive process – remembering that the museum is on 53<sup>rd</sup> Street is either processes a, b and c or it is processes x,y and z.

I am not convinced, then, that A&A’s particular arguments against the ‘sameness’ of Otto’s and Inga’s processes are valid. More significantly, however, I disagree with A&A’s general approach to the question of this ‘sameness’. In keeping with their position of “contingent intracranialism”, the objections that A&A put forward against the sameness of role of Otto’s notebook are merely contingent ones and as such can be challenged at the level of particular detail. They see nothing problematic *in principle* with the idea that physical manipulation of the environment could be “the same” as cognitive processing. My objection, however, is that it is not clear in what sense environmental manipulation could ever be viewed as “the same” as cognitive processing; C&C are *not* saying (pace A&A’s second objection) that the inner accompaniments of Otto’s use of the notebook are the same as Inga’s inner processing, but that the use of the notebook *itself* is the same as something internal that Inga has. How can this be? A passage from Wittgenstein perhaps illustrates the type of bewilderment appropriate here:

*If one has to imagine someone else’s pain on the model of one’s own pain, this is none too easy a thing to do: for I have to imagine pain which I do not feel on the model of pain which I do feel. That is, what I have to do is not simply to make a transition in imagination from one place to another .. (Wittgenstein 1953 p.302)*

Here Wittgenstein is talking about a slightly different problem – the problem presented by the ‘argument from analogy’ – but it has some relevance to our own situation. In both cases we are trying to assimilate internal and (from a first person perspective) external states and processes and in both cases we have problems because they appear to be incommensurate kinds of thing. This problem does not seem to be acknowledged in C&C’s paper where it is asserted:

*A part of the world functions as a process which were it done in the head, we would have no hesitation in recognizing as part of the cognitive process (p.2)*

and again:

*this sort of coupled process counts equally well as a cognitive process whether or not it is wholly in the head. (p.3)*

as though there is as little philosophical bother in moving this “it” in and out of the head as there is physical bother in putting on or taking off a hat.

What might be useful here would be to have (to coin another Wittgensteinian phrase) a ‘criterion’ of sameness – a respect in which one could say that features of both worlds are the same. C&C’s nonchalance perhaps suggests that they are already using one or more such criteria, but these are not anywhere explicitly laid out so some detective work is needed.

One such criterion is perhaps simple similarity in linguistic description. This seems to be the case with the Otto scenario which trades partly on the similarity between “looking up” in memory and looking up in a notebook. As a criterion this seems the most easy to dismiss. One argument to be made here is surely that the sense in which there is any “looking up” in memory is a metaphorical one. Moreover this metaphor (perhaps like many of the constructs in Cognitive Psychology) is itself derived from our everyday practises - such as looking something up in a notebook. We can only “look things up in memory” because we can look things up in a notebook [1]. This has two consequences. Firstly that such metaphors might misrepresent wildly what actually happens when e.g. we remember something. Secondly to take the source of the metaphor as an extension of the metaphor is to view things in a topsy turvy way.

A second possible criterion for sameness is given more explicitly in the ‘Tetris’ example at the very beginning of the paper. C&C compare 3 different cases. In case 1 “a person sits in front of a computer screen which displays images of various two-dimensional geometric shapes and is asked to answer questions concerning the potential fit of such shapes into depicted ‘sockets’. To assess fit, the person must mentally rotate the shapes to align them with the sockets.” In case 2 a person “can choose either to rotate the image on the screen, by pressing a rotate button, or to mentally rotate the image as before.” In case 3, a person “has the benefit of a neural implant which can perform the operation as well as the computer in the previous example. The agent must still choose which internal resource to use” (p.1).

According to C&C cases 2 and 3 both display “(by hypotheses) the same sort of computational structure”(p.2 – my emphasis). C&C then argue from this that “If the latter case is (as it seems to be) a case of cognitive processing, by what right do we count case (2), using the external rotation button, as fundamentally different?”

It is difficult, for me at least, to see how this ‘sameness of computational structure’ argument is meant to work.

Firstly, the relevance of the neural implant case is not immediately clear, and the assertion that its use is likely to be ‘a case of cognitive processing’ seems to me to be beg the question. Partly the problem here is with vagueness of description. How exactly does the neural implant operate? What phenomenology is associated with its use?

One can imagine that using such an implant need not produce any distinctive phenomenology. Neither need there be much distinctively ‘cognitive’ about its use. In (Clark, 2003) an example is given of a neural implant lodged in the motor cortex of a stroke victim, by means of which a cursor on a computer screen is controlled. (p.122). This is not, Clark is clear, a case of “thought control”, for the patient with the implant does get the cursor to move by thinking ‘move cursor right’. Rather the patient has to ‘experiment with their own motor signals’ until they get it right. Then after a period of time it becomes ‘second nature’, in the same way that moving an arm is second nature to an able bodied individual. We are not tempted to call the use of the neural implant in this case ‘cognitive processing’ – at least not in any strong sense. Thus if this is the kind of thing C&C had in mind with the Tetris example, then neither would we be willing to call *that* ‘cognitive processing’. If, on the other hand, C&C were imagining a different kind of set up then more details need to be made available before we can judge whether such a set up would involve cognitive processing. However, as the possibility of any more advanced ‘cognitive style’ implant technology is purely speculative (and highly contentious) it does not seem appropriate to base an argument for the extension of cognitive processes upon it.

There is also a third possibility which is that C&C see what happens in case 3 as ‘the same’ in all relevant respects as that which happens in case 1 and for this reason take it as uncontroversial that the neural implant scenario is a case of cognitive processing. This point will be developed further later. Suffice to say here that this does seem to be part of the underlying perspective of the paper, as though ‘internal’ processes are different from external ones only by virtue of physical location.

So, we have suggested that the neural implant case is a not necessarily useful piece of science fiction, yet it is in comparing this case with case 2 that C&C make the assertion that both have the same kind of ‘computational structure’. Having put to one side the neural implant scenario, we therefore want to know if the ‘computational structure’ criterion can provide us with a means of identifying the essential cognitive sameness between case 1 and case 2 instead.

Again it is not clear what exactly C&C have in mind here. In using the term ‘computational structure’ do they mean the stepwise procedure by means of which the problem with the shapes is solved? We can hypothesize that this could indeed be (verifiably) the same with cases 2 and 3 but can such an identity be meaningfully be applied to 1 and 2? Surely it might be said that whilst there is still a solution to the problem which has the same computational structure in both cases, it is not obvious that there is any strong sense in which the individual in case 1 is adhering to a computational procedure.

This is not to say that sense can’t be made of the idea that the individual in case 1 might be following a stepwise procedure on some interpretation isomorphic to that followed by the individual in case 2. The intention here is not to cast the mental goings-on of the first individual in a mysterious light. Nevertheless there is clearly a problem in a too literalistic interpretation of any identity of procedure between case 1 and case 2. This problem is brought out by making a Rylean

observation to the effect that “mentally rotating” the shapes is not merely rotating the shapes in a parallel domain to the physical.

C&C might well deny this, of course. They make reference to the work of Kirsh & Maglio which shows that in Tetris “the physical rotation of a shape through 90 degrees takes about 100 milliseconds plus about 200 milliseconds to select the rotate button. To achieve the same result by mental rotation takes about 1000 milliseconds.” (p.2). However although much of use can be derived from such research, the fact that a common metric is applied in both situations is not enough to show that any putative ‘sameness’ between the two situations is not radically problematic.

The “sameness of computational structure” criterion, then, at least as I have interpreted it here, does not seem convincing. A third candidate also seems less than satisfactory. This is the criterion of “functional identity”. C&C refer several times to the ‘role’ of the external as part of the mental:

*The notebook plays for Otto the same role that memory plays for Inga (p.7)*

*Beliefs can be constituted partly by features of the environment, when those features play the right sort of role in driving cognitive process. (p.6)*

*What makes some information count as a belief is the role it plays, and there is no reason why the relevant role can be played only from inside the body. (p.8)*

The persistent use of the word ‘role’ suggests an adherence to the ‘role/occupant’ distinction found in functionalist theories of mind. How might such an approach work given C&C’s commitment to (non-distal) externalism? A functional identity theory would seem to work best with processes. One would think, for example, that the gist of a functional identity criterion of “sameness” of internal and external processes would be that if the results achieved by an external process such as doing calculations on paper were reliably the same as those done “in the head”, then the external process could count as “the same process”. Such an account would at least gel with our intuitions that certain pairs of processes (of which mental and written arithmetic are a good example) have a special relationship. However, although C&C do talk about external processes standing in for cognitive processes, they are keen, as the above quotes show, also to talk about externalizable mental *states*, and they stress that a mental role can be played by an individual particular *item* such as the notebook, a feature of the environment or information. This is confusing. For all sorts of reasons it just does not seem plausible that my memory or belief could literally be made manifest in an external object (or ‘category’ such as “information”). Or at least, it is not plausible that it should do so and the rest of the world remain the same. One can of course conceive of other perspectives such as an Hegelian system where everything is viewed as an embodiment of Mind [2], or some kind of semiotic perspective where all my beliefs and memories are constituted by external signs. But such perspectives would, at least ideally, have internal coherence on their side – they would be pictures of different worlds. C&C’s perspective, however, is rooted more or less in *this* world – which we shall describe tentatively as either the world of science or the world of everyday discourse - and in *this* world it is not acceptable to postulate e.g. that “John lost his memory under the sofa”.

A possible counter argument from C&C here might be that this criticism is a ‘linguistic’ one – ie that it centres on what is customarily said in the ‘world of science or the world of everyday

discourse'. What is customarily said, they might argue, is no guide to anything. There is no reason why we might not start saying that "John lost his memory under the sofa". It might seem a strange misuse of language to refer to a notebook as a memory, merely because it plays the same functional role, but, as Churchland points out, "the abuse of accepted modes of speech is often an essential feature of real scientific progress"(Churchland. p.71) That C&C seem to have a criticism like this is suggested by the following passage concerning their novel use of "belief":

*We do not intend to debate what is and is not standard usage..Our broader point is that the notion of belief ought to be used in such a way... (p.8)*

It is difficult to know how to respond to analyses such as these which see attempts to draw attention to conceptual violation as mere quibbling. Certainly – deviating from the Otto example for a moment – C&C's paper is filled with abuses of language that a traditional 'ordinary language' theorist would soon get his teeth into. For instance:

*In a very real sense the rearrangement of tiles on the tray is not part of action it is part of thought. (p.3)*

Here it might be asked *what* "very real sense" is this. We already have a use for the word "action" and it includes things like moving tiles around on the board. "Thought", on the other hand, is meant to denote something separate from actions. We are only enabled to say that "thoughts express themselves through actions" because we conceive the two as separate. A similar example can be found in (Clark, 1997) where approval is given to Carruthers' assertion that "One does not first entertain a private thought and then write it down: rather the thinking *is* the writing" (p.197). Yes, we might reply, thinking and writing may well be dialectically bound in some circumstances, but this internal relatedness still requires that there be two terms – 'thinking' and 'writing'. One cannot reduce thinking to writing, or vice versa.

Likewise with another example from C&C's paper – the suggestion that "one's beliefs might naturally be seen to be embodied in one's secretary..". The notion that slaves and servants in some sense embody their masters beliefs and desires is not a new one. To posit this as a literal possibility however is confusing in the extreme. We cannot say that X has Y's beliefs in the sense that X's believing something is indistinguishable from Y's believing it.

The 'cannot' here, of course, is not the 'cannot' of (what Mundle (1979) called) 'prescriptive linguistics'. In objecting to literal extensions of mind, or to reductions of thought to action, the intention is not to deny new uses of language or to lay down conservative rules about how words should be used. On the contrary the starting point of any kind of linguistic analysis should be an awareness that language is evolving all the time and that this evolution is rooted in the dialectic between use and meaning. Nevertheless, language use has to be consistent. What we want to say with C&C is that the very foundation on which they erect their theory is one that presupposes the separateness of thoughts and actions, the attributability of mental states to conscious entities and not objects, the attributability of my thoughts to me and not someone else and so on [See p.12 for more on this] The notion that an object which plays a functional role identical to a type of mental state is a token (?) of that mental state, must therefore sit very uneasily on such a foundation.

With some side-tracking, then, we have discussed the idea that there is something problematic in the blithe assertion that the ‘same’ process or the ‘same’ state might occur externally or ‘in the head’. We have also argued that criteria for ‘the same’ in terms of 1) linguistic similarity 2) computational structure and 3) functional role, are unconvincing.

[[It could perhaps be added here that Clark and Chalmers are not the only recent philosophers of cognitive science who have failed to see the problems involved in unquestioning application of the concept of ‘the same’ across domains. We might also note similar problems with Gallese’s discussion of the significance of mirror neurons. “We observed,” says Gallese, “that a particular set of neurons, activated during the execution of purposeful, goal-related hand actions, such as grasping, holding or manipulating objects, discharges also when the monkey observes *similar hand actions* performed by another individual. We designated these neurons as *mirror neurons*.” (Gallese 2001 – p.35). The problem here, to labour the point, is that ‘the same’ is not a given. Although, on the face of it, what gives Gallese’s hypothesis its interest is the possibility that it might provide just such an objective physiological basis for the identification of ‘the same’ mental state across first person/third person boundaries, there must be a sense in which this project is misconceived (which is not to say that it isn’t important.) This can be seen from the fact that further investigation could well reveal that the mirror neuron was also activated when the observed monkey did the ‘opposite action’ to the observing monkey, or when the observed monkey made a movement which combined only some of the elements of the observed monkeys movements but was carried out when the sun was in a certain position. If we take the activation of the observing monkey’s ‘mirror’ neuron as our criterion for ‘sameness’, then these actions of the observed monkey automatically become ‘the same’.]]

### *Literalism and Individualism*

The problems alluded to so far – and in particular that of postulating that those external processes deemed ‘the same’ as particular internal ones might therefore be labelled ‘cognitive’ – can be seen as the result of a certain ‘literalism’ which in turn seems to go hand in hand with what might loosely be called ‘individualism’. I will now try to briefly expand on what I mean by these two terms.

Whether the world external to the individual is conceived of as the social world or the physical environment or both, there have throughout history been ongoing disputes between those who think that Man’s nature is intrinsically external and those who conceive of the external as some kind of ‘add on’. As an example of such a dispute we might note the argument between Ayer and Wittgenstein (represented by Malcolm) over the question of whether there could be a ‘private language’.

Wittgenstein’s position was that we learn how to use words like ‘pain’ from practise in a social context, not from our own case. One of his arguments for this position was a *reductio* to the effect that if we *were* to try to define words like ‘pain’ by a kind of private ostensive definition we would face an impossible dilemma, for we would have no way of knowing whether we were using the word ‘pain’ in the same way from one instance to the next. “Whatever is going to seem right to me is right”(para 258). For this reason alone (although there are many others) the notion of a Private Language is incoherent and in disposing of it Wittgenstein hoped to dispose of grounds for philosophical scepticism about others’ pain. Ayer, however, was completely unimpressed by this argument. As far as he was concerned it just *was* possible to construct a first language all by oneself:

*Surely it is not self contradictory to suppose that someone, untutored in the use of any existing languages, makes up a language for himself* (Ayer. p. 259)

and to illustrate this he imagines a Crusoe-like figure brought up without language, unproblematically naming objects that he comes to recognize.

Ayer's view, then, is the 'individualist' view. It is 'individualist' because it projects onto the individual – as intrinsic to the individual - characteristics that many (including myself) would say could really only be developed and have any meaning in relation to a wider (social) world. It is a projection backwards from the world to the individual. We are reminded perhaps of Adam Smith's conception of 'Homo Economicus' who confronts the world as a fully formed free marketeer with a natural "propensity to truck and barter and exchange one thing for another" (Smith. p.117) [3]

Connected to such views is the notion that the addition of the external (physical/social) element adds nothing qualitatively new. In Ayer's case, for example, it may be pretty pointless to invent a private language and talk to oneself in it but such a project is deemed feasible. Moreover, in talking to others I am not thought to be doing anything essentially different to talking (only) to myself. The additional features of a public language are like a 'plug-in' that can interface seamlessly with my private language module. (We might note that elsewhere Clark (1997) propounds a theory of language halfway between Ayer's and Wittgenstein's – this point will be developed in a later paper.)

Most basically such views can be seen as turning on the assumption that mental constituents and external constituents are in some important sense comparable i.e. paradoxically the prior separation of the individual from the social/physical world is only achievable by positing an equivalence of content between both parties. With different theorists and in different contexts this equivalence takes different forms. For Ayer there was a private language just as there was a public language. For Smith there were individual free market propensities just as there was collective free market behaviour. For introspectionists there are private objects just as there are public objects. More generally, for many philosophers and cognitive scientists, there is an inner realm just as there is an outer realm. What is important here is not just that these pairs are similarly named but that in significant respects the names stand for properties in common. In an admittedly vague way we might say that each member of a pair is, in varying degrees and different respects according to context, held to be the same sort of thing as its co member.[4]

How, then, does this relate to C&C's account? For C&C it means taking what W.I.Thompson (2003) calls a 'literalist' approach to mind. There is no difference in kind between the inhabitants of the internal world and the inhabitants of the external world, and it is for this reason that mind should seem to be literally extensible. This homogenisation of mental and worldly constituents is present throughout C&C's account. Indeed they seem to go further than most in positing a qualitative sameness between mental and worldly stuff. It is evident, for instance, in the framing of the opening question: "Where does the mind stop and the rest of the world begin?"(p.1) as though it were only the physical barrier of "skull and skin" that might be thought to stop them intermingling. It is evident also in what I have previously referred to as the 'Sci – Fi Scenarios'. These are three brief speculative passages which prefigure Clark's more extensive account of the man-machine interface given in 'Natural Born Cyborgs', (which itself forms part of a wider literature with a 'mix and match' attitude to man-stuff and machine-stuff. See eg. Moravec (1999))

Sci-Fi Scenario No.1 was outlined earlier. This was the 'Tetris' scenario wherein it was imagined that a neural implant be used for rotating shapes on a screen. Sci-Fi Scenario No.2 concerns an

argument against the claim that Otto's notebook-based memory is a case of true memory. According to the argument Otto's information retrieval has a 'perceptual phenomenology' whereas Inga's doesn't, therefore Otto is not truly remembering. C&C's counter argument to this is that we can imagine a 'Terminator' style scenario wherein an individual has his memories displayed before him like an inner movie. In such cases, say C&C, we still regard the individuals memories as genuine "standing beliefs"(p.9). Sci-Fi Scenario No.3 concerns the argument that so-called "coupled systems", such as the system made up of Otto and his notebook, are too easily decoupled to count as entirely cognitive. One of C&C's responses to this is to argue that it is imaginable that some time in the future we might be able to plug modules into our brain to help us with various tasks such as geometric reasoning. In such a scenario we would not deny that the processes involving these modules were "part of thought"(p.4).

There are a number of individual and general objections to these scenarios . In particular we might ask "Why are such examples thought to be illustrative of anything at all?" It is as though C&C imagine that if we can conceive of a set of circumstances such a set of circumstances must be possible. This does not seem to be the case, as anyone who has read 'Alice in Wonderland' or 'The Time Machine' will testify - what we can conceive of may not even be logically possible. It might be contested, of course that we are taking these scenarios in the wrong way, that they should be regarded as 'thought experiments', but an objection to this would surely be that the scenarios lack the necessary 'experimental' component to be considered legitimate thought experiments. They are more like unargued statements that a certain state of affairs *is* possible.

With respect to our current line of enquiry, however, our main objection to the scenarios is that they make the presupposition that mind stuff is qualitatively indistinct from world stuff. Each scenario takes for granted that (the non extended part of) a cognitive process is a brain process, that such a process is replicable in a silicon entity and that the interface between brain and silicon is unproblematic. There is a seamless continuity between the world of mind and the world of silicon. Given this equation:

$$\text{Cognitive process} = \text{brain process} = \text{hardware realizable process}$$

the move to an 'extended mind' has more or less already been made. It requires no more of a leap of the imagination than the supposition that Turkey could become part of Europe. It may be, of course, that in making this argument we are committing ourselves to the view that all forms of 'functionalism' – insofar as they adhere to the principle of multiple realizability – are committed to mind extension of the C&C variety. This would indeed seem to be the case since, at least in the exposition of such accounts that I'm acquainted with, there is nowhere a proviso made that mental states/processes are indivisible and so cannot be distributed over more than one domain. However, since my position is in no way dependent on the coherence or otherwise of functionalism such a consequence is of little concern here.

The Sci – Fi scenarios are not the only clearly expressed evidence of the literalist basis of C&C's thesis. The Otto scenario provides a further example of the enforced homogeneity of the contents of the 'mental' and the external world. Most of what is relevant here has already been covered earlier. However it's worth noting again how such homogeneity involves the 'projecting backwards' from the environment to features of the mind. Like the writing in the notebook, Inga's belief is information (not an attitude to information, or a use of information, just information) on C&C's analysis, the only difference between the two types of information being that with the writing in the notebook "it just so happens that the information inheres in a physical state located beyond the skin"(p.7).

We also note that, just as the information in Otto's notebook is there, whether viewed or not, so we can say of Inga's belief that it "was sitting somewhere in memory waiting to be accessed"(p.6). Here we might stick our neck out by asserting that this characterization of Inga's belief is unwarranted. It does not seem that we need to postulate that any kind of representation persists in the brain between the act of learning something and the act of remembering that same something. To me, at least, this is an instance of the literalist prejudice which asserts the existence of internal continuities where none need be found. [[A similar example can be found in the literature on 'attention': "In the well known 'cocktail party phenomenon' for example, we ignore all extraneous sounds while taking part in a conversation, but can quickly switch focus if someone else mentions our name. The implication is that we were always 'listening' to ambient sound but not always 'hearing'" (Durie, 2005). This interpretation seems to rest on assumptions as unwarranted in their own way as those made by the early phrenologists]]

I have argued, then, that C&C's account posits a qualitative 'sameness' between mental and 'external' stuff. It is this sameness which – in part - makes the mind appear exportable, although paradoxically that very sameness seems to involve the description of the internal as part of the external, as though the world were being extended inwards rather than the mind outwards. I have also argued that this 'postulation of sameness' is bound up with an individualist perspective, though I will be the first to admit that I haven't developed this connection very clearly.

[[Another point that might be made here – although one that needs to be elaborated on in more detail to do it any justice – is that C&C's literalism seems to hinge on a vagueness about what 'internal' means. This is a difficult issue to grapple with (for me at least!) but it seems plausible to argue that to say that my mental state is 'internal' is to talk on the person level. It is to say that my mental states are internal to me, not to my physical body ie. The 'internal' here is not one of physical space. (See eg Ryle (1950)). C&C's account, with its talk of the barrier of 'skull and skin', often implies that 'internal' merely means 'physically' internal. That Clark (at least) appears to have such a perspective is borne out by other work (eg. 1997) where there is much equivocation between talk of the mind and talk of the brain, (the phrase 'mind brain' often being used), and where sometimes bewildering arguments concerning the internal/external divide are put forward ("The mere fact that we often mentally rehearse sentences in our heads and use these to guide and alter our behaviour means that one cannot and should not treat language and culture as wholly external resources" p.198).

Clearly one's conception of what 'internal' and 'external' mean with respect to mind will inform one's conception of what it is for mind to be extended. If it seems as though Clark (& Chalmers) are making a mountain out of a molehill with their notion of extension it may be because they had a weirdly literalistic conception of non-extended mind prior to their conversion. ]]

### *The Halfway House*

It is perhaps unsurprising that C&C should view the mind and external world as qualitatively indistinct for this is the correct view to take when the minds one is interested in are programmable hardware 'minds' rather than sentient human minds. C&C, however, are allegedly interested in *both* types of minds. Thus in using the one type of description they are inhabiting an

uncomfortable ‘halfway house’ as regards conceptual development/ontological commitment. It is not possible, we might say, to develop a suitable language for describing extended human mind that will also be applicable to the realm of “on board devices”. This point is also relevant to remarks made earlier about C&C’s inconsistent language use. I will now try to develop this idea a little further.

A genuinely philosophical description of the ‘mind’ as extended would seem to involve a thorough going remaking of the world and re orienting of our place in it. It is not enough, and indeed is counterproductive, just to tinker at the edges. C&C seem only partially aware of this. Thus they set in motion a process which they are not willing to complete. They posit the literal existence of an extended mind. They then follow this up with a few separately considered applications of this thesis – that actions might be thoughts, that memories might be objects, that my secretary might be my beliefs, that my garden might be me. These applications if taken seriously would seem to have major theoretical consequences - such as the dissolution of opposites (thought/action, self/other) – but these consequences are not followed through and on most fronts it remains business as normal.

Thus, for example, the twin concepts of “epistemic” and “pragmatic” actions (attributed to Kirsh & Maglio), which form part of the groundwork upon which the extended mind is erected, remain immune to any category dissolving funny business:

*Epistemic actions alter the world so as to aid and augment cognitive processes such as recognition and search. Merely pragmatic actions, by contrast, alter the world because some physical change is desirable for its own sake.(p.2)*

Here we want to say that if the dichotomy between the concepts of ‘thought’ and ‘action’ is to be brought into question, then so surely should that between ‘epistemic’ and ‘pragmatic’ actions. It is just as artificial to suppose that some actions are purely pragmatic while others are purely “cognitive”. It seems not unreasonable to suppose, for example, that every action I take could be viewed as having some cognitive payoff – even having a bath (or maybe *particularly* having a bath). Likewise actions I perform “to aid and augment cognitive processes” might be desirable in themselves. Indeed, on one reading, a string of such actions resulting in a desired consequence could be redescribed as a single action which brought about some physical change “desirable for its own sake”.

Similarly the strictly limited notion of ‘coupling’, derived, as I understand it, from dynamic systems theory, is perhaps indicative of the arrested development of the extended mind. Another theorist (say a Marxist or a Phenomenologist, or even a different Clark) might take a more fluid and integrated approach to the question of mind extension, with self, world and other being reciprocally co determinate. C&C’s ‘coupling’, however, is simply the binding together of two individual, unproblematically given, particulars out on their own in the world. As such it is more a marriage of convenience than a genuine unity. We might add too that this marriage isn’t even democratic because one of the pair has the upper hand. When it is noted, for example, that there is some integral relationship between action and thought the response is not to use the notion of ‘coupling’ to describe the fact in its entirety but is rather to reduce everything to thought. Likewise the relationship between self and other is reduced to self (my secretary is a repository for my beliefs but not vice versa) and that between mind and the world is reduced to mind (though, as I argued in the previous section, ‘mind’ has first to be made like world in order for world to become mind).

C&C's position, then, is a kind of halfway house because although it has opened the doors to the realm of the extended mind it has done so in a fragmented and uncommitted way. It seems possible that this predicament can be partly explained by C&C's parallel commitment to science, that maybe a scientific approach is at odds with the proper development of notions like 'the extended mind'. It could be argued, for instance, that the many references to technology and technology oriented conceptions of mind ("on board devices", "high bandwidth link" etc) posit an objectivism incompatible with the dissolution of opposites that C&C's account of the Extended Mind seems to require. Here we might note Thompson's assertion that there is a "fundamental entre-deux between science and experience, experience and the world" (p.15). From the point of view of experience:

*Our relating to the world, including when we do science, always takes place within  
a matrix whose fundamental structure is I-You-It* (p.21)

but from the point of view of science the 'It' is the fundamental given from which all analysis must start. Thompson thinks the solution to this 'entre-deux' lies in the adoption of *both* approaches – "phenomenology and cognitive science should not be opposed but should be joined together in a relationship of mutual illumination". However it seems arguable that it is just such an attempt at 'joining' which has resulted in the unsatisfactory mishmash found in C&C's account and perhaps this is an indication that the respective approaches are therefore better off left alone i.e. Given that C&C's account *does* posit as non problematic the third person existence of 'out there' technological artefacts, and (as has been argued) conceives of the mental as continuous with these artefacts, it ought at least to be consistent and not get itself involved in speculation (the literal extension of 'mind', actions as thoughts etc) which if developed seriously would undermine its third person perspective [5].

In this connection it seems noteworthy that C&C consider the opposite to be the case i.e. they consider the 'Extended Mind' project in its entirety a thoroughly scientific one. They claim that the theory of Extended Mind is reflected in the theory of situated cognition, studies of real-world-robotics, dynamical approaches to child development and research on the cognitive properties of collectives of agents. In these fields :

*cognition is often taken to be continuous with the environment. Thus in seeing cognition  
as extended one is not merely making a terminological decision; in these cases it makes a  
significant difference to the methodology of scientific investigation.* (p.4)

A reply to this might be that it seems likely that research in any of these fields could get by just as well without any explicit commitment to the belief that action is literally thought or that mind is literally extended i.e. they could get by just as well with the uncontroversial notion that the environment is in a variety of ways integral to our mental states and processes.

---

## Notes

[1] Perhaps this is literally so - the development of memory may have been concurrent with the development of a certain use of symbols. This latter point should not seem obscure to Clark for it parallels certain points he makes elsewhere with respect to language and inner rehearsal. In (1997), for example, he argues that we learn to give ourselves instructions after acquainting ourselves with the process of obeying others instructions, and he quotes approvingly Rumelhart et al. who state:

“Here we have a kind of internalization of an external representational format.” (p.199)

There seems to be no good reason why we shouldn't apply this approach to memory itself.

[2] Since writing this I have come across the following:

“A crucial insight in Hegel's model is that mind exteriorizes itself. Mind makes itself exterior to itself – in language, artefacts and institutions. Its activity is self-exteriorization and self-appropriation. Spirit or mind (Geist) takes place *between* a subject and its world, and between subjects in a more or less shared world. This model could be a fruitful resource for exploring the relation between mind and world in cognitive science. For example, when Andy Clark defines mind as being extended beyond itself (or beyond the brain) into external, physical and social, scaffolding structures (artifacts, institutions etc) that on the one hand inform and guide the daily actions of individuals and that on the other hand are themselves informed and structured by their communicative acts (Clark,1997..), it would be obvious to take the Hegelian model into account.” (Gron, 2004).

Gron's attitude here is that Hegelian perspectives might have something to offer a Clarkian approach, but it seems to me that an acquaintance with the former is more likely to bring to the fore the fragmentary and 'ad hoc' character of the latter.

[3] The opposing view to Smith was put forward by Marx who argued that “the relationship of man to himself first becomes objective to him through his relationship to other men” (Marx p.84)

[4] I have walked into a contradiction here which again seems to call for a 'dialectical' solution. Those who see Man's nature as intrinsically external are opposed to those who think the external is a mere 'add on'. Yet it is the latter who by stealth posit the equivalence of Man and the external world. Some talk of 'sameness in difference' might be useful here perhaps?

[5] This point, as many others, needs to be expanded. Perhaps the desirable solution would be a 'dialectical' one where both poles of the 'entre deux' were 'superseded'. However the phenomenological approach, as I understand it, is already meant to represent the supersession of empiricism and rationalism, so.....?

## References

- Adams, F. & Aizawa, K. (2001). *The Bounds of Cognition*. Philosophical Psychology, Vol 14, 1.
- Ayer, A.J. (1954). *Can There Be a Private Language?* Reprinted in Pitcher(1966) *Wittgenstein*
- Churchland, P.M. (1994). *Matter and Consciousness (Revised Ed)*. MIT.
- Clark, A. (1997). *Being There*. Cambridge, MA: MIT Press.
- Clark, A. & Chalmers, D. (1998). *The Extended Mind*.
- Clark, A. (2003). *Natural Born Cyborgs*. OUP.
- Durie, B. (2005). *Doors of Perception* in New Scientist, 29 Jan 2005, p.34-36.
- Gallese, V.(2001). *The 'Shared Manifold' Hypothesis* in 'Between Ourselves', JCS 8, 5-7.
- Gron, A. (2004). *The Embodied Self* in JCS 11, 10-11.
- Marx, K. (1844) *Economic and Philosophical Manuscripts*. Reprinted in McLellan D. (1988).  
*Karl Marx – Selected Writings*. OUP.
- Moravec, H. (1999). *Robot*. OUP.
- Ryle, G. . (1949). *The Concept of Mind*. Hutchinson.
- Searle, J. (1992). *The Rediscovery of Mind*. MIT Press.
- Smith, A. (1776). *The Wealth of Nations*. Penguin Edition – 1999.
- Thompson, E. (2001). *Empathy and Consciousness*. Journal of Consciousness Studies 8, 5-7.
- Thompson, W.I. (2003). *The Borg or Borges*. In 'Machine Consciousness'. JCS 10, 4-5
- Wittgenstein, L. (1953). *Philosophical Investigations*. Basil Blackwell.