

From:

Michael Wheeler

Extended X: Recarving the Biological and Cognitive Joints of Nature

Draft Book Manuscript

NB: Please do not quote or cite without permission

Chapter 4 On Your Marks

4.1 The Shape of Things to Come

We have been excavating the foundations of ExC, in order to reveal the conditions under which it would be plausible to conclude that cognitive traits may be, or perhaps sometimes are, extended into the environment. In so doing we have discovered that the cogency of the parity-driven case for ExC turns, in part, on our ability to supply a scientifically informed, theory-loaded, locationally uncommitted account of the cognitive. In this chapter we shall explore a number of proposals from the ExC literature that look as if they might conceivably provide such an account. Before that, however, we need to confront a nagging worry. To those versed in the debates over ExC, it will probably not have escaped notice that by placing the idea of a scientifically informed, theory-loaded, locationally uncommitted account of the cognitive at the centre of things, I appear to have argued us into an unholy alliance with two of the friendly hostiles, namely Fred Adams and Ken Aizawa, in the vicinity of what they call the *mark of the cognitive*. So my first task is to clear our way forward, by extracting us from the unholy aspect of this alliance.

4.2 Sleeping with the Enemy (Part One): the Mark of the Cognitive

Over the years, Adams and Aizawa (e.g. 2001, 2008, forthcoming) have argued persistently that ExC fails (by some distance) to discharge an important debt to cognitive theory, namely to provide an adequate mark of the cognitive. What 'adequate' amounts to here will become clear shortly. First, however, a word about exactly what Adams and Aizawa mean by the term 'mark of the cognitive'.

What they are aiming at is not an a priori definition of cognition, but rather an empirical theory of the cognitive that one might reasonably conclude is at work, albeit in a predominantly implicit manner, in a good deal of cognitive-scientific research, where cognitive science is represented principally (and sometimes it seems exclusively) by orthodox human cognitive psychology (more on this last point later). According to Adams and Aizawa (2008, pp.86-7), the need for such a theory of cognition becomes more visible at those times when cognitive science confronts a substantive challenge to its entrenched ways of going on, as when the fan of ExC claims that cognitive traits extend outside the skin. It's at such times that we really need to know what is being meant by the term 'cognition'. Similarly, as Adams and Aizawa note, it might seem that the need for a theory of life would become acute in biology if someone sincerely suggested that there were good reasons to believe that crystals are alive. Closer to home, when Richard Dawkins suggests that phenotypes are extended, he does so on the basis of an account of what a phenotype is, i.e., the "manifested attributes of an organism, the joint product of its genes and their environment during ontogeny" (Dawkins 1999, p.299, quoted by Adams and Aizawa 2008, p.22, footnote 9; more on the extended phenotype in later chapters). Adams and Aizawa expect no less from the fans of extended cognition.

The emerging and (given our diametrically opposed inclinations) somewhat unnerving affinity between Adams and Aizawa's approach and mine becomes even more pronounced once we plug in the fact that their official position is that there should be no *prejudicial* appeal to the inner in the way one formulates one's mark of the cognitive. In other words, not only should we expect our mark of the cognitive to be theoretically loaded (given its relationship with cognitive science), we should also expect it to be locationally uncommitted, in the sense in which we have been using this term. So it should be the case that nothing in our mark of the cognitive rules out, *in principle*, the possibility of cognitive extension. Adams and Aizawa's claim is that once we give the correct content to the mark of the cognitive, and look to see where human cognition falls, we will find that it is a resolutely skin-side phenomenon. But this locational aspect of human cognition is supposed to be a result, not a conceptual starting point, of the approach. Thus (and looking forward to Adams and Aizawa's own favoured mark of the cognitive, more on which below): "We do not maintain that non-derived representations [part of their favoured account] must be found in the head. That cognition involves non-derived representations is one empirical hypothesis; that non-derived representations are to be found in some particular regions of spacetime is another." (Adams and Aizawa 2007, p.55)

Of course, there are structural differences between the two approaches. Most notably, I have been developing a parity-based case for ExC, while Adams and Aizawa ultimately object to parity arguments for reasons that we addressed in chapter 3. However, once parity is reconceived (as I have recommended it should be) not as parity with the inner *simpliciter*, but rather as parity with the inner *with respect to a theory-loaded, locationally uncommitted account of the cognitive*, and once the similarities between Adams and Aizawa's approach and mine are noted, it seems that one might requisition the idea of the mark of the cognitive and, pace its originators, allow it to play the crucial benchmark-setting role within a parity-based account. This explains why an important moral that Adams and Aizawa (2008, p.76) draw from their discussion of the mark of the cognitive transfers directly to my parity-based framework for ExC. That moral is this: if the ExC theorist adopts a weak or promiscuous enough mark of the cognitive, then it will be easy enough for her to secure the result that cognition is extended; but the price of this success will be to welcome into the domain of the cognitive all kinds of wildly unlikely cases in a manner that ultimately casts doubt on the ability of the proposed mark to latch onto only what might be thought of as the proper objects of cognitive science.

Specific applications of this moral will be concern us later. For now the key point is that it gives us another adequacy condition on any proposed mark of the cognitive, in addition to the requirements that it be locationally uncommitted and appropriately related to cognitive science. That new condition is that any proposed mark of the cognitive must exclude from the domain of the cognitive enough of the aforementioned unlikely cases, where what counts as 'enough' is left deliberately vague in classic paradox-of-the-heap-style fashion, and where one needs to recognize that there will be interesting borderline cases to be fought over. We might think of Adams and Aizawa as betting that any account of the cognitive that meets this third adequacy condition will *end up* locating cognition inside the skin, even though in principle it could have turned out to be extended. It seems, then, that a good way in which to begin to extract ourselves from the clutches of these particular friendly hostiles is to see what is wrong with their own suggested mark of the cognitive.

4.3 The Chances of Anything Coming from Mars...

Adams and Aizawa argue for a twofold mark of the cognitive. First, cognition "involves non-derived representations, representations that mean what they do independently of other representational or intentional capacities" (Adams and

Aizawa 2008, p.31). The contrast here is with derived representations, representations whose content “arises from the way in which items are handled or treated by intentional agents... who already have thoughts with meaning” (ibid. p.32). Examples of non-derived representations include thoughts, experiences and perceptions. Examples of derived representations include road signs, ordinance survey maps and flags, items whose representational content depends on those thoughts, experiences and perceptions. The second characteristic feature of the cognitive is that it is “individuated by specific kinds of information processing mechanisms”, namely those identified by human cognitive psychology (ibid. p.31). (Attentive readers will have spotted that we have met this idea already, in chapter 3. We shall revisit the relevant arguments from that chapter shortly.) Having explicated their twofold mark of the cognitive, Adams and Aizawa proceed to suggest that the extended solutions beloved of ExC theorists will systematically fail to display that bipartite characteristic throughout, and so will fail to count as cognitive. The qualification ‘throughout’ is important, since, on Adams and Aizawa’s account, the extended solutions in question may of course *contain* cognitive elements. Thus, to return to one of our stock examples, the environment-involving (pen-and-paper) solution to long multiplication might well contain cognitive elements – elements that involve the manipulation and transformation of non-derived representations by particular kinds of information processing mechanisms. But these will be located in the mathematician’s brain, not the environment. The external factors in this problem-solving system are not themselves (we are told) cognitive, because they do not involve the right kinds of processes or states. Therefore (at best) we have a case of merely embodied, rather than extended, cognition.

How might the ExC theorist react to Adams and Aizawa’s claims? One possible response would be to argue that, to the extent that the notion of a mark of the cognitive is intertwined with that of a theory-loaded, locationally uncommitted account of the cognitive, no genuine substance can be given to the idea. But we have explored this kind of view already, in chapter 3, where we found not only good reasons to reject Clark’s claim that the vehicles of cognition are too much of a motley for anything like such an account to get a grip, but also equally good reasons to believe (as mentioned above) that a parity-driven case for ExC depends on the delivery of precisely this sort of account as a way of establishing a benchmark for parity. A different response is needed. As an alternative, then, the ExC theorist might accept the idea that there is a mark of the cognitive, but challenge, on ExC-independent grounds (so as not to beg any questions), Adams and Aizawa’s specific understanding of that idea. Perhaps the most obvious way to criticize Adams and Aizawa’s account would be to attack the very idea of non-

derived representation. If there is no such thing as non-derived representation, then the property of involving non-derived representation can't be the mark of the cognitive. This sort of strategy is occasionally toyed with by Clark (e.g. 2005, p.4). However, once the concept of non-derived representation is purged of certain unfortunate links with the somewhat murkier notion of intrinsic content, and is given naturalistic flesh by way of theories of content such as Dretske's (1988) indicator function semantics or Fodor's (1987) asymmetric causal dependency theory (see Adams and Aizawa 2007, pp.35-9 for discussion of these and other options), I am inclined to think that it is in good conceptual order, or rather that its conceptual order is at least as good as that of the view that all representation is derived representation (Dennett 1990), a view which has its own problems (see e.g. Newton 1992, Aizawa and Adams 2005). So this response won't do either.

A third response would be to argue that even if Adams and Aizawa are right about how to cash out the mark of the cognitive (and that remains a matter for discussion), they are wrong that this has any negative implications for ExC. It seems to me that the most plausible claim for the ExC theorist to make here is not that certain external elements may carry non-derived content (although see Clark 2005), but that a genuinely cognitive trait may feature elements that carry only derived content, *just so long* as those elements function alongside other elements that carry non-derived content. If we accept that non-derived representations will be found in the brains of some organisms, and that those non-derived representations may combine in the right way with certain externally located derived representations in order to solve cognitive problems, then this would be one way of explicating the idea of organism-centredness within an ExC framework (see chapter 2).

This is one case where a heuristic appeal to counterfactual innards helps to direct our thinking. Recall from chapter 3 Clark's bit-map Martians, creatures whose internally located memory systems involve bit-map images of printed text which are created and internally stored, and then later retrieved in the form of bit-mapped signals sent to and interpreted by the Martian visual cortex. As noted when we first met these alien creatures, pretheoretically speaking, we appear to have no hesitation in counting the bit-map system as a genuinely cognitive mechanism. Nevertheless, as Clark (forthcoming a) observes, it seems clear that the arrangement in question contains representations that have only derived content, which means that it fails to meet Adams and Aizawa's non-derived representation condition. So, on the face of it, Adams and Aizawa confront a dilemma: they must either give up that part of their mark of the cognitive or

stand firm in the face of the seemingly strong intuitions that tell against it. Unsurprisingly, perhaps, it is the latter course that they pursue. “[T]hese Martian representational states are not cognitive states.” (Adams and Aizawa 2007, p.49)

The first thing we need to do here is try to understand the force of Adams and Aizawa’s refusal to count the Martian bit-map system as cognitive. In chapter 3, in the context of a debate that turned on the issue of active versus passive memory systems, I suggested that the realization of the Martian bit-map system in human beings would require a significant transformation in the basic encoding structure of the human brain. I also argued that any pro-ExC argument based on such a transformation-inducing system could at best establish only the general modal version of ExC. Now, because the truth of the human (modal and non-modal) versions of ExC depends on the truth of the general modal version, it seems that any anti-ExC argument that tells *against* the general modal version must tell against the two human versions too. Adams and Aizawa’s refusal to grant cognitive status to the Martian bit-map system looks to be one such argument, since, by way of parity, our tendency to treat these counterfactual innards as cognitive is ultimately supposed to help convince us of the plausibility of (the general modal version of) ExC. However, this is not the only angle from which to approaching things. As an alternative, one might argue, I think, that although, in the previously considered context of active versus passive memory systems, the Martian bit-map system signalled a transformation in the basic encoding structure of the human brain, in the present context of non-derived versus derived representations, the Martian bit-map system signals no such transformation, because it seems likely that the human brain already contains representational states whose content is derived from the content of other, more fundamental, representations. If that is so, then the cognitive status of the Martian bit-map system is directly relevant to the human modal version of ExC, and Adams and Aizawa’s refusal to count that system as cognitive turns into an argument against the more specific version of the target thesis.

Whichever interpretation we place on Adams and Aizawa’s stance here, one might wonder whether the move of excluding the Martian bit-map system is strictly available to them. After all, they understand their own non-derived representation condition as requiring that “*at least some components of cognitive states require some non-derived content*” (ibid. p.51, my emphases). But this would seem to allow for the possibility that some components of cognitive states may possess only derived content, which would, in tune with our pretheoretical intuitions, permit the bit-map Martian system to count as cognitive. This concession would of course re-open the door to the possible existence of

extended cognitive traits via the possible existence of extended systems that have the same functionality as the Martian bit-map system. That's why Adams and Aizawa hold firm, and why, to do justice to their official pronouncements, we need to find a way of reading their seemingly concessive gloss in such a way that our bit-map Martian system remains excluded. The way to do this, I think, is to interpret Adams and Aizawa as taking the state of interest to be nothing 'bigger' than the information-bearing bit-map, a state which does not itself contain any elements with non-derived content. That would restore the claim that the target element fails to meet the non-derived representation condition. However, it is desperately unclear that there is any independent motivation for dividing up the Martian's cognitive architecture in this way, rather than in a way that treats the bit-map as part of a larger cognitive trait with some components that possess non-derived content. So although denying the bit-map Martian system cognitive status seems at first sight to put Adams and Aizawa in a position to resist ExC, it not only flies in the face of some powerful pretheoretical intuitions, but also leaves them vulnerable to the charge of begging the question against ExC, on the grounds that there seems to be no independent justification for dividing up the Martian cognitive architecture in the way that yields the anti-ExC result.

What this tells us, I think, is that the non-derived representation condition is no real threat to ExC. Attention turns, therefore, to the second part of Adams and Aizawa's mark of the cognitive, to the condition, that is, that the cognitive is individuated by specific kinds of information processing mechanisms as identified by cognitive science, where cognitive science is represented by human cognitive psychology. But this spells even more trouble for Adams and Aizawa. As I have argued (again, in chapter 3), any attempt to delineate the domain of the cognitive by way of the specific kinds of information processing mechanisms studied by human cognitive psychology is undermined by the surely plausible observation that, from the fact that some specified psychological phenomenon happens to be of interest to cognitive psychologists, one cannot infer that failing to exhibit that phenomenon is an indication of non-cognitive status. In the wake of all this, it is hard to avoid the conclusion that Adams and Aizawa's mark of the cognitive is seriously flawed.

4.4 Extended Information Processing

Reverting to my less elegant but arguably more descriptive terminology, we might proceed by asking the following question: what sort of theory-loaded, locationally uncommitted account of the cognitive finds favour among the ranks

of ExC theorists? The most prominent answer is highlighted by Clark (2008, p.44), who notes that “[a]rguments in favour of [ExC] appeal mainly, if not exclusively, to the *computational role* played by certain kinds of non-neural events and processes in online problem-solving”. Clark himself tends to work with a very liberal notion of ‘computation’, according to which “showing that a system is computational reduces to the task of showing that it is engaged in the automated processing and transformation of information” (Clark 1997, p.159). So Clark’s observation about ExC and computation amounts to the thought that ExC theorists overwhelmingly conceive of cognition as a matter of *automatic information processing*. (I shall suppress the word ‘automatic’ in what follows. Whenever I write of information processing, I mean automatic information processing.) Construed in the light of this move, the distinctive ExC claim becomes that, in some cases, extra-neural factors may, or perhaps do, implement information processing, and thereby cognitive processing, just as readily as neural tissue. Of course, not any old kind of information processing will do here. As Adams and Aizawa (e.g. 2008, p.77) are fond of pointing out, CD players, DVD players, FM radios, digital computers, mobile phones etc. all do information processing, but they are not cognizers. Indeed, the range of problem items might stretch beyond CD players and mobile phones. As Rowlands (2006, p.32) notes, the concept of information is regularly understood in philosophy by way of notions such as nomic dependence or conditional probability,¹ which means that all sorts of processes (and perhaps even every process) will count as processing information in some sense. If cognition is a matter of information processing, then it is a special kind of information processing, which means that we need to specify the extra constraints that exclude the non-cognitive kinds.

One constraint that might seem to do the job is the notion of contributing to the achievement of a genuinely cognitive task (such as perceiving the environment or reasoning on the basis of remembered facts), as opposed to contributing to the achievement of some other, non-cognitive sort of task (such as receiving and broadcasting radio signals). For example, Rowlands (1999, pp.102-3) claims that “[a] process P is a cognitive process if and only if (i) P is essential to the accomplishing of a cognitive task, T, and (ii) P involves operations on information-bearing structures, where the information carried by such structures is relevant to task T”. Unfortunately this proposal falls short of the mark. In a subsequent treatment, Rowlands (2006, p.32) rightly observes that many processes that are strictly essential to the achievement of cognitive tasks, and which seem to qualify as doing information processing by the conditions identified earlier, are not themselves cognitive. Consider, for example, digestion and respiration. They seem to meet the conditions for doing information

processing, in the undemanding sense that, following Rowlands, we have adopted. Moreover, as long as we shun substance dualism, dead people don't think. So, given the roles of digestion and respiration in supporting organic life, realizing those processes is clearly essential to the achievement of cognitive tasks. Nevertheless, it seems clear that they are not themselves cognitive. Our account of the cognitive needs to exclude such unwanted interlopers.

Rowlands (2006, p.32) endeavours to tighten up his account by adding a third condition, namely that for a process to be cognitive, it must be "of the sort that is capable of yielding a cognitive state", where a cognitive state is a *representational* state whose status as such is determined by the fact that it figures in a mechanism whose *adapted proper function* is precisely to realize that state in some environmental context. Relatedly, Menary (2007, p.15) claims that "[a] process is cognitive when it aims at completing a cognitive task; and it is constituted by manipulating a [representational] vehicle". ('Relatedly', because Menary not only introduces an explicit appeal to *representation*, he writes of *aiming to complete a cognitive task*. In a naturalistic setting, this reference to the *goal* of the process in question is most likely to be cashed out in terms of the adapted proper function of that mechanism; see e.g. Menary 2007, p.117.) The notion of the proper function of an adapted mechanism is now relatively familiar in naturalistic circles, so just a few words of explanation should suffice. In essence, although not in name, the notion goes back at least as far as Wright's (1973) canonical analysis of selective function. The concept was developed further, and was famously given both its moniker and its currency in the philosophy of mind, by Ruth Millikan (1984, 1993). The idea has some complex subtleties (see e.g. essays 1 and 4 in Millikan 1993), but Neander (1995, p.11) gives a tidy definition of the core notion that will do for present purposes: "[s]ome effect (Z) is the proper function of some trait (X) in organism (O) iff the genotype responsible for X was selected for doing Z, because doing Z was adaptive for O's ancestors." Crucially, then, the concept of an adapted proper function is *historical* in character, concerning the job that the mechanism, or more generally the trait, in question was evolutionarily selected to perform in *ancestral* populations, in order to promote survival and reproduction.

Unfortunately, Rowlands' attempt to tighten up his account of the cognitive – by adding in the requirement, that, in order to count as cognitive, a process must realize an adapted proper function by way of a representational state – threatens to make the necessary conditions for being cognitive too strong. For it seems unlikely, in the extreme, that what we understand to be the vehicles of cognition will either always have a selective history that, in the required sense, they can

reasonably call their own, or be universally representational in character. Each of these issues warrants some discussion, especially since both of them will crop up again later. Let's begin with the adapted proper function condition.

Although the terminology of 'proper function' may be useful in various ways, if the notion of *adaptation* is itself defined historically, in terms of ancestral contributions to survival and reproduction, then possessing an adapted proper function is, in all important respects, equivalent to being an adaptation.² As Amundson and Lander (1994, p.447) explain:

...for a trait to be an adaptation (historically defined) is *precisely* for that trait to have a function (selection-effect defined) [i.e. an adapted proper function]. A trait *is* an adaptation when and only when it *has* a function. The two terms are interchangeable. If a law were passed against the [adapted proper function] concept of function, its use in biology could be fully served by the historical concept of adaptation.

What this entails (trivially) is that if realizing an adapted proper function is a necessary condition for a trait to be cognitive, then being an adaptation is a necessary condition for a trait to be cognitive. Now, what is sometimes called *ultra-Darwinism* is the view that almost all phenotypic traits in almost all populations of organisms are adaptations, that is, are the direct product of Darwinian selection. It follows, then, that the person who thinks that realizing an adapted proper function is a necessary condition for a trait to be cognitive is an ultra-Darwinist about the cognitive. Indeed, it seems that he will be an *ultra-ultra-Darwinist* about the cognitive, since, on his view, *all* cognitive traits will be adaptations. So is there anything wrong with that? In a powerful critique, Gould (2000) argues (i) that ultra-Darwinism is on the retreat in evolutionary thinking generally, and (ii) that the human mind looks to be particularly resistant to any ultra-Darwinist treatment. If points (i) and (ii) can be established, then Rowlands' tightened-up account of the cognitive will be undermined by its ultra-Darwinist commitments.

The evidence for point (i) comes from (what Gould takes to be) an increasingly widespread recognition in biology that evolution is a mosaic of many different processes and phenomena, including not only Darwinian selection, but also factors such as contingency, evolutionary spandrels (traits that are not themselves selected for, but rather are by-products of selection for other traits), and punctuated equilibria (according to which the emergence of new species is not a gradual process driven by natural selection acting on geographically

isolated groups in different environments, but rather involves long periods of what is essentially stasis and then moments of abrupt change). The general debate over ultra-Darwinism (or equivalently strong adaptationist positions) in evolutionary biology is far from over (see e.g. Gould and Lewontin 1979, Dennett 1995, Sterelny and Griffiths 1999, Elton 2003), but it does now seem likely that, in the form required by Rowlands' strengthened version of the information processing account of cognition, the view cannot be sustained.

In the face of this difficulty, Rowlands might reply that the domain of the cognitive is a special realm in which ultra-Darwinism reigns supreme. But this strategy would face severe problems, in the form of Gould's point (ii). The notion of evolutionary spandrels is particularly salient here. Gould argues that since all organisms evolve as complex and interconnected wholes, selection-driven change to one feature will typically generate non-adaptive by-products. These side-effects of selection – dubbed *spandrels* by Gould and Lewontin (1979) after the triangular spaces formed as architectural by-products of the decision to mount a dome on rounded arches – may *later* be co-opted by selection to perform some function, but the existence and structure of those by-products are not explained *by* selection. Thus they are not themselves adaptations. Given that the human brain is the most complex and internally interconnected organ around, it is very likely, as Gould puts it, to be “bursting with spandrels that establish central components of what we call human nature but that arose as non-adaptations, and therefore fall outside the compass of evolutionary psychology or any other ultra-Darwinian theory” (ibid, p.104). This argument surely generalizes to the range of ecological vehicles that would provide the machinery for cognitive extension. So wherever one thinks cognition might be located, it looks very likely to be a spandrel-heavy domain.

It might seem that Rowlands has a line of response available to him here, one in which he seeks to establish that, despite what Gould's onslaught suggests, spandrels may be the genuine bearers of adapted proper functions. Here's how the response might go. Commenting on the significance of the spandrels concept in a slightly different context (more on which below), Rowlands (2003, p.168) writes: “[a] structure or mechanism that is maintained because of the role it plays in underwriting certain cognitive processes, even though it was not originally developed for this role, is an evolutionary product no less than a structure that was originally produced for that role”. This is, of course, true. However, to assess the relevance of the claim in the present context, one needs to keep track of two different ways in which a trait might be an evolutionary product. A trait might be an evolutionary product in that, although it originally entered the population

as a by-product of selection and thus qualifies as a spandrel in that sense, it has since been the target of selection in virtue of the fact that it has made some *subsequent* positive contribution to fitness. Thus it has been “maintained *because of* the role it plays in underwriting certain cognitive processes” (my emphasis). A spandrel that is the product of such secondary selection acquires a selective history to call its own. In so doing it seems plausible to suggest, contra Gould, that it becomes an adaptation and thus comes to have an adapted proper function. Thus even if the vehicles of cognition were awash with these sorts of spandrels-that-become-adaptations, that would not undermine Rowlands’ account of the cognitive.

So far so good for Rowlands. However, there is a further sense in which a trait might be an evolutionary product. It might be an evolutionary product in that (a) it was originally produced by one of a range of evolutionary mechanisms other than selection, and (b) although it has subsequently acquired no selective function, it has not been selected out of the population. Consider, for example, the phenotypic results of random genetic drift in neutral fitness landscapes (for discussion see e.g. Sober 2005). Or cases where factors such as epistasis (genetic linkage) and the power of generic self-organization prevent selection from shifting biological form (see e.g. Kauffman 1993, Goodwin 1994). Being an evolutionary product in this further sense is patently not sufficient for a phenotypic trait to have an adapted proper function. After all, the traits in question have not survived *because of* what they contribute to fitness. Instead they are fitness-neutral, or perhaps even mildly deleterious but somehow resistant to pressure from selection. So it is hard to see how the notion of being selected for, and thus of adaptation, and thus of adapted proper function, can get any sort of grip. This makes things difficult for Rowlands, because it seems clear that many of the spandrels that exist in the brain and in the relevant extended systems (if there are any), will not have been the targets of secondary selection. For example, as Schacter and Dodson (2001) observe, it is plausible that misattribution in memory (attributing a recollection or idea to the wrong source) evolved as a by-product, either of a gist-based memory system or of a memory system that does not standardly store all the details of the etiological source of an experience. Against this backdrop it is hard, as Schacter and Donaldson point out, to judge the spandrel of misattribution to be anything other than selectively disadvantageous. (Consider, e.g., the problems that individuals suffer following the ‘recovery’ of false memories in psychotherapy.) The message, then, is that many cognitive spandrels will be evolutionary products in only our first sense of that term. And that isn’t enough to protect Rowlands from the present application of Gould’s arguments.³

If all this is correct, then Rowlands' proposal for how to tighten up the information processing account of cognition is in trouble. For the adapted proper function condition (whether or not that condition is played out representationally) will have the effect of excluding from the class of the cognitive many traits that, it seems, ought to be included (e.g. our second kind of cognitive spandrel). That would be enough to scupper Rowlands' proposal. Nevertheless, just for good measure, let's examine the other part of the suggestion: the claim that the vehicles of cognition are necessarily representational in character.

The obvious objection here – that there are nonrepresentational vehicles of cognition – might be made in a number of ways (see e.g. Varela et al. 1991, Webb 1994, Wheeler 1994, van Gelder 1995, Nöe 2004). Here, however, is one that I find particularly compelling. There is a phenomenon, seemingly important to the functioning of biological nervous systems, that Clark (1997) has dubbed *continuous reciprocal causation* (CRC).⁴ As Clark characterizes it, CRC is causation that involves multiple simultaneous interactions and complex dynamic feedback loops, such that (a) the causal contribution of each systemic component partially determines, and is partially determined by, the causal contributions of large numbers of other systemic components, and, moreover, (b) those contributions may change radically over time. Later in our story, we shall see just how important CRC may be in the production of intelligent behaviour. For now, however, the key point is that CRC undermines representational explanation. It does that because it undermines modularity (Wheeler 1998, 2005a, 2005b).

To explain: In systems that exhibit CRC, the performance of any particular sub-task will be underpinned by a large number of interacting components whose contributions are changing in highly context sensitive ways. With increasing levels and spread of CRC, it becomes progressively more difficult and explanatorily unhelpful to attempt to specify distinct and robust causal-functional roles played by reliably reidentifiable parts of the overall system. So if we take it that a system is modular to the extent that (i) it consists of scientifically identifiable subsystems, each of which performs a particular, well-defined sub-task, and (ii) its global behaviour can be explained in terms of the collective behaviour of an organized ensemble of such subsystems, then what this means is that where CRC is rife, there will be no useful modular explanation of the system.⁵ Although couched in different terms, the same point about CRC and modularity is made by Varela et al. (1991, p.94):

[O]ne needs to study neurons as members of large ensembles that are constantly disappearing and arising through their cooperative interactions and in which every neuron has multiple and changing responses in a context-dependent manner ... The brain is thus a highly cooperative system: the dense interconnections among its components entail that everything going on will be a function of what all the components are doing ... [If] one artificially mobilizes the reticular system, an organism will change behaviorally from, say, being awake to being asleep. This change does not indicate, however, that the reticular system is the controller of wakefulness. That system is, rather, a form of architecture in the brain that permits certain internal coherences to arise. But when these coherences arise, they are not simply due to any particular system.

So CRC undermines modularity. But why exactly should this allow us to infer a lack of representations? It is arguable (Wheeler 2005b) that a particular triad of properties constitutes a set of necessary and jointly sufficient conditions for a target inner state or process in a behaviour-controlling system to rightly be accorded the status of a subagential (i.e., vehicular) representation. The three properties are: (1) being a genuine source of adaptive richness and flexibility; (2) arbitrariness; (3) being part of an homuncular system. Here I shall confine myself to stating, without much explication, the key features of each of these representation-relevant conditions. Further issues of evidence, interpretation and analysis will arise in later chapters (see also Wheeler and Clark, 1999; Wheeler 2005b, 2006). Being a genuine source of adaptive richness and flexibility requires that the target inner state or process be causally correlated upstream with objects and states of affairs, and downstream with behavioural outcomes. In the sense that matters here, arbitrariness may be interpreted as the multiple realizability of the state or process in question, in which the equivalence class of realizers is fixed by informational rather than first-order-physical factors. Finally, a system is homuncular just when it may be analyzed into a set of hierarchically organized, communicating modules, each of which performs a well-defined sub-task that contributes towards the collective achievement of a global systemic solution.⁶

The key point right now concerns the claim that systemic homuncularity is necessary for subagential representation. As defined above, homuncularity is (trivially) a form of modularity (one in which the modules take a particular information-dealing form), so when modular decomposition fails in the face of CRC, so does homuncular decomposition, and so does representational explanation. That's how the modularity-undermining effects of CRC become

representation-undermining effects too. And, once again, this point surely generalizes to putative cases of cognitive extension. Where ecological vehicles of cognition exhibit CRC, the effect will be that modular decomposition, and thus representational explanation, will fail to apply to the extended system in question. This gives us a similar result to the one obtained for Rowlands' adapted proper function condition. If the line of reasoning just presented is correct, then Rowlands' attempt to tighten up the information processing account of cognition by adding a representation condition (adapted-proper-function-based or not) will have the effect of excluding from the class of the cognitive many traits that, it seems, ought to be included. It seems, then, that a different strategy will be required if the information processing approach to cognition is to avoid what we may now characterize as the dual problems of *excessive liberality* (letting in unwanted interlopers) and *disproportionate elitism* (excluding genuinely cognitive traits).

4.5 Glue and Trust

A candidate for just such an alternative strategy may be found in the pages of Clark and Chalmers' (1998) original treatment of the extended mind. Here a set of four extra constraints are placed on a broadly information processing account of cognition. Clark (forthcoming a) has since described these constraints as "a rough-and-ready set of additional criteria to be met by non-biological candidates for inclusion into an individual's cognitive system" (Clark 2008b, chapter 4, p.32, manuscript). As such they are designed to prevent unwanted interlopers, now in the form of certain external resources such as books in a home library or mobile access to Google (Clark's examples), from counting as bona fide parts of an individual's cognitive system. Of course, given the parity considerations that Clark and Chalmers favour (see chapter 3), the constraints in question will need to apply equally to inner and outer elements. And given Clark and Chalmers' appeal to our intuitive folk picture to fix the domain of the cognitive (again see chapter 3), those same constraints, and indeed the broadly information processing framework within which they operate, will need somehow to be interpreted as part of that folk picture, if they are to be consistent with the rest of the philosophical machinery on offer from these authors. By contrast, and in the wake of the arguments developed in chapter 3, we are under the former obligation but not the latter. For us the key point is that Clark and Chalmers' clear intention is to delimit an appropriate sub-class of information processing mechanisms so as to elude the hazard of excessive liberality. But successfully evading that hazard requires that one also steer clear of disproportionate elitism.

The question is, then, are Clark and Chalmers successful in achieving these dual aims?

The specific constraints proposed by Clark and Chalmers have subsequently been dubbed, by Clark (forthcoming a), conditions of *glue and trust*. They are: (i) that the resource be *typically invoked* when the information it contains is relevant; (ii) that the information contained in the resource be *easily accessible* when required; (iii) that any information thus retrieved be *more-or-less automatically endorsed*, just like the information retrieved from ordinary organic memory, rather than being routinely open to the sort of critical scrutiny to which we subject, for instance, the opinions of other people; (iv) that the information be contained in the resource because it has been *consciously endorsed* by the agent at some point in the past (Clark and Chalmers 1998, p.17). Clark and Chalmers are immediately dubious about condition (iv), since it is arguable that one might acquire cognitive information through subliminal perception or memory tampering. Moreover, as Rupert (2004) points out, as long as we maintain the received view that the conscious endorsement of an item of information remains an event in the *inner* cognitive life of an agent, the cognition-determining role assigned to this phenomenon by Clark and Chalmers seems to sit unhappily alongside (although perhaps not in logical conflict with) their support for cognitive extension. On the other hand, as Rupert also notes, in the absence of condition (iv), the excessive liberality problem raises its ugly head again:

the first three criteria imply that virtually every adult, Otto included, with access to a telephone and directory service has true beliefs about the phone numbers of everyone whose number is listed. The directory assistance operator is a constant in Otto's life, easily reached; when the information would be relevant, it guides Otto's behavior; and Otto automatically endorses whatever the operator tells him, about phone numbers, anyway. It is absurd to say that Otto has beliefs about all of the phone numbers available to him through directory assistance (i.e., beliefs of the form, "John Doe's phone number is ###-####"), so long as he remembers how to dial up the operator. (Rupert 2004, pp.17-18)

Unfortunately things only seem to get worse for Clark and Chalmers, because while conditions (i)-(iii), when unembellished by condition (iv), have the effect of allowing in to the domain of the cognitive certain unwanted interlopers, they also have the opposite effect of excluding what certainly look to be indisputable cases of cognitive traits. An argument for the latter conclusion is developed by

Sprevak (manuscript), who highlights (among other examples) not only the cognitive resources deployed in unusual instances of human artistic creativity, resources that are plausibly neither typically invoked when relevant nor easily accessible when required, but also the outputs of human cognitive processes such as imagining, supposing and desiring, outputs that are often subject to critical scrutiny by the thinker herself. If we impose conditions (i)-(iii), these genuinely cognitive resources would be incorrectly categorized as non-cognitive. And notice that this collapse into disproportionate elitism would only be exacerbated if we restored the conscious endorsement condition, since, as we have seen (see previous comment on subliminal perception and memory tampering), there appear to be cases, highlighted by Clark and Chalmers themselves, in which the agent is genuinely in a cognitive state, even though that condition isn't met. So, if we did enforce condition (iv) alongside conditions (i)-(iii) we would end up excluding, from the domain of the cognitive, additional traits that certainly seem to belong there. To drive home the general point here, one might put the argument another way (see Sprevak manuscript, p.13). We have no hesitation in granting cognitive status to various organically realized processes that violate the glue and trust conditions. Why, then, shouldn't the same status be granted to extended processes that violate those conditions? Equality of treatment (the idea at the heart of parity considerations), demands that the same standards be applied to extended processes as are applied to organic ones. But that just shows that the trust and glue conditions ultimately fail to carve the cognitive parts of nature from the non-cognitive parts in a reliable way. In other words, and ironically, a powerful reason for rejecting the glue and trust strategy is provided by the parity principle itself.

One response available to Clark and Chalmers here (a response nicely identified by Drayson 2008) would be to complain that Sprevak's examples only count against the glue and trust conditions, *if we allow those conditions to be wildly over-generalized beyond their intended domain of application*. The conditions in question, so the defence goes, are meant to be rough and ready criteria only for *dispositional belief*, and not for cognition in general. Since Sprevak's examples of genuinely cognitive traits that fail the glue and trust criteria are not instances of dispositional belief, they are strictly irrelevant to the specific ExC claim to which those criteria are supposed to contribute, namely that there may be extended dispositional beliefs. There is textual evidence that supports this interpretation of the disputed criteria. After all, they initially come to the fore within the development of the Inga and Otto thought experiment (see previous discussions), the principal aim of which is to establish that the Otto-notebook system may have dispositional beliefs. Clark and Chalmers (1998, p.17) tag them

accordingly, as helping us to “understand what is involved in ascriptions of extended belief”.

There are worries, however. By playing the scope-restricting card against Sprevak’s putative counter-examples, the ExC theorist inherits the not-inconsiderable task of providing, for all other cognitive traits, local sets of conditions that will enable us to exclude from the domain of the cognitive certain unlikely or absurd cases, while including extended examples of the trait in question, plus all other bona fide instances. Moreover, while Sprevak’s charge of disproportionate elitism may be deflected in this way, Rupert’s example of excessive liberality on the part of conditions (i)-(iii) concerns dispositional beliefs in particular, and so cannot. Finally, Adams and Aizawa (2008, pp.121-5) discuss a series of examples that, in part, suggest that even though Sprevak’s alleged counter-examples may miss their target, the charge of disproportionate elitism does apply to the glue and trust analysis of dispositional belief, because one may have dispositional beliefs that are not typically invoked and that are subject to critical scrutiny. Here is my adaptation of one of Adams and Aizawa’s examples. Imagine that Tom, who has a normal memory but believes he is very bad with names, bumps into Katie White in a bar. Tom might, in fact, reliably recall the information that this person’s name is ‘Katie White’. He might even realize that he should know her name, having seen her on television only the night before. Nevertheless, because Tom believes that he is bad with names, he does not trust his memory. Even when the correct name comes to his mind, he will ask someone nearby to confirm the information. So Tom will not typically invoke the internally stored and reliably triggered information that this person’s name is ‘Katie White’. Moreover, he will subject that information to critical scrutiny. By the glue and trust conditions, Tom does not have the dispositional belief that this person’s name is ‘Katie White’. But that seems wrong. Surely he has the dispositional belief but, to use Adams and Aizawa’s term, is *alienated* from it.

Although there might well be further rounds left in this contest, there are, I think, grounds to doubt the capacity of the glue and trust strategy to successfully tighten up the information processing account of cognition. The upshot of this is that, despite considering a range of options, we still don’t have the scientifically informed, theory-loaded, locationally uncommitted account of the cognitive that the parity-driven argument for cognitive extension so badly needs. In the next chapter I shall endeavour to plug this gap.

Notes

1. Nomic dependence exists where there are lawlike relations. Thus the number of rings in a tree stump carries information about the age of the tree because of the lawlike relationship between the two factors. Interpreting information in terms of conditional property, and borrowing a clear explanation due to Adams (2003), we can say the following: "A signal r carries the information that s is F = The conditional probability of s 's being F , given r (and k), is 1 (but, given k alone, less than 1). K is a variable that takes into account how what one already knows may influence the informational value of a signal. If one knew nothing, k would go to zero." Thus, if I know that Gordon Brown is from Glasgow or Edinburgh, and I learn that he is not from Edinburgh, I thereby glean the information that he is from Glasgow. If I don't know that Gordon Brown is from Glasgow or Edinburgh, and I learn that he is not from Edinburgh, I do not thereby glean the information that he is from Glasgow.

2. The notion of an adaptation may be given a non-historical interpretation, in which case it signals (something like) the result of an agent regulating its intra-lifetime behaviour through development, learning or adaptive plasticity, in order to improve or maintain its situation (see e.g. di Paolo 2000). A phenomenon may be an adaptation in this sense, without having an adapted proper function.

3. One possible reply here might be to argue that, under conditions of competition for resources, selection against and selection for are two ways of looking at the same evolutionary process, such that to the extent that a trait is not selected against, it is selected for. Once upon a time I argued in favour of this view (Wheeler 1995), but it seems to me now that this was a mistake. There needs to be conceptual room for us to make sense of the idea of selectively neutral change in evolution.

4. For further philosophical discussion of continuous reciprocal causation, see Wheeler 1998, 2005a, 2005b especially chapter 10, Clark 2008b, Dreyfus 2008. For what is effectively the same idea discussed before Clark's coining of the term, see Varela et al. 1991, Harvey 1992, Wheeler 1994. For the empirical evidence that this kind of causation may underlie intelligent behaviour in biological systems or, in some cases, in artificial systems inspired by biology, see Husbands, Harvey and Cliff 1995; Husbands, Smith, Jakobi and O'Shea 1998; Freeman 2000. More on the empirical evidence in later chapters.

5. The definition of modularity that I give here is not mandatory. Indeed, the term 'module' is used in widely varying ways in cognitive science. Perhaps the most prominent treatment of the idea is due to Fodor (1983) who specifies some nine conditions --- including, for example, informational encapsulation, domain specificity, speedy processing, and being associated with a fixed neural architecture --- all or most of which a functionally identified subsystem would need to display in order to count as a module. As we shall see in chapter 8, evolutionary psychologists too tend to work with a set of cluster conditions, most or all of which will be in force when some cognitive device is described as a module. Conditions in the evolutionary-psychological frame might include being domain specific, being wholly or perhaps mostly genetically determined, being universal among normally functioning humans, and being computational in character (for discussion, see e.g. Samuels 1998). My preferred account of modularity is far less restrictive than either of these options. It is one with which I suspect most cognitive neuroscientists would be comfortable.

6. It seems that whenever one mentions homuncular explanation in naturalistic philosophy of mind, one receives a torrent of familiar worries about the potential here for an infinite regress of systems, each of which, in order to do what is being asked of it, must literally possess the very sorts of intentional capabilities that the model is supposed to explain. Let's be clear. Homuncular explanation in psychology is in good order, if one is ultimately able to discharge all mentioned intentional capacities. Here is the standard model of how that works (Dennett 1978). According to the homuncular strategy, if we as cognitive scientists wish to understand how a whole agent performs a complex task, we should proceed by analyzing that complex task into a number of simpler sub-tasks (each of which has a well-defined input-output profile), and by supposing that each of these sub-tasks is performed by an internal 'agent' less sophisticated than the actual agent. These internal 'agents' are conceptualized as communicating with each other, and thus as co-ordinating their collective activity so as to perform the overall task. This first-level decomposition is then itself subjected to homuncular analysis. The first-level internal 'agents' are analyzed into committees of even simpler 'agents', and each of these 'agents' is given an even simpler task to perform. This progressive simplification of function continues until, finally, the sort of thing which you are asking each of your 'agents' to do is something so primitive that the explanation is almost certainly going to be a matter for low-level neurobiology rather than psychology. This 'bottoming-out' in low-level neurobiology is what ensures that all talk of 'little people in the head' remains entirely, and non-dangerously, metaphorical.