



Is language the ultimate artefact?

Michael Wheeler

Department of Philosophy, University of Stirling, Stirling FK9 4LA, UK

Abstract

Andy Clark has argued that language is “in many ways the ultimate artefact” (Clark, 1997, p. 218). Fuelling this conclusion is a view according to which the human brain is essentially no more than a pattern-completing device, while language is an external resource which is adaptively fitted to the human brain in such a way that it enables that brain to exceed its unaided (pattern-completing) cognitive capacities, in much the same way as a pair of scissors enables us to “exploit our basic manipulative capacities to fulfil new ends” (Clark, 1997, pp. 193–194). How should we respond to this bold reconceptualization of our linguistic abilities? First we need to understand it properly. So I begin by identifying and unpacking (and making a small “Heideggerian” amendment to) Clark’s main language-specific claims. That done I take a step back. Clark’s approach to language is generated from a theoretical perspective which sees cognition as distributed over brain, body, and world. So I continue my investigation of Clark’s incursion into linguistic territory by uncovering and illustrating those key ideas from the over-all distributed cognition research programme which are particularly relevant in the present context. I then use this analysis as a spring-board from which to examine a crucial issue that arises for Clark’s account of language, namely linguistic inner rehearsal. I argue that while there is much to recommend in Clark’s treatment of this issue, some significant difficulties remain to be overcome. Via this critique of Clark’s position, alongside some proposals for how the revealed problems might be addressed, I hope to edge us that bit closer to a full understanding of our linguistic abilities.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Language; Distributed cognition; Inner rehearsal

E-mail address: m.w.wheeler@stir.ac.uk

1. Introduction

Towards the end of his highly influential 1997 book, *Being There: Putting Brain, Body, and World Together Again*, the philosopher Andy Clark unveils a rich and challenging account of language.¹ My intention in this paper is to interrogate certain key moments in Clark's account. I should say at the outset that I am in deep sympathy with Clark's fundamental approach to the issue. This is unsurprising, perhaps, since he and I share a range of gut intuitions, philosophical assumptions, and science-driven thoughts about the essential nature of mind, cognition and intelligence (more on this in Section 3 below). Nevertheless, one consequence of this large measure of agreement is that any readers out there who find themselves hankering after a blanket demolition of Clark's position will, I wager, be sorely disappointed by what follows. Indeed, it seems to me that something like Clark's view of language must be right, even if, as I shall argue, the details of his own particular treatment stand in need of some significant clarifications and revisions. My ultimate goal then is not to scupper the ship that Clark launches, but rather to enhance its philosophical and scientific sea-worthiness.

2. Cognitive scissors

Let's begin by laying out Clark's account. Clark's first major claim is that language is best viewed not—or, to be more accurate, not merely—as a medium of communication (which is how it is standardly and most readily conceived). Language is equally (perhaps even primarily) “a tool that alters the nature of the computational tasks involved in various kinds of problem solving” (Clark, 1997, p. 193). The idea here is that public language is a resource that enables human beings (or their brains) to restructure certain problems so that (as we shall see later) those problems are rendered amenable to the kinds of systems that we (or our brains) most fundamentally are. Intertwined with this quite radical rethinking of the functional role of language in human cognition is a further claim. Linguistic competence is often paraded as requiring a revolution in psychological innards, with the upshot being that the linguistic-haves and the linguistic-have-nots are held to possess fundamentally different brains. On this view, the brains of the former (but not those of the latter) are typically thought to contain some sort of domain-specific language processing system, one whose elements are organized, at a fundamental level, so as to encode the structural properties of natural language (e.g. a Chomskyan language acquisition device; see e.g. Chomsky, 1986). But Clark rejects the need for any such drastic inner discontinuity at the threshold of language. His striking alternative proposal is that language is “an external resource that complements but does not profoundly alter the brain's own basic modes of representation and computation” (p. 198).

¹ See also Clark (1998, 2001, 2003). Clark would be the first to remind us that his work on language draws and builds on a number of previous treatments by others (for discussion, see Clark, 1997, pp. 194–200).

This further claim has three components. The first two are explicit in the immediately preceding quotation, while the third needs to be teased out. The first (explicit) component, which dovetails neatly with Clark's opening claim that language is a tool which humans exploit, is to promote the externality of language as a phenomenon. Like more familiar tools—hammers, compasses, computers and so on—language is part of the external supporting environment in which our brains and bodies mature and work. The second (explicit) component trades on the thought that the biological brain has certain generic forms of inner state and mechanism (“the brain's own basic modes of representation and computation”) that, from both an evolutionary and a developmental perspective, precede linguistic competence. The idea, then, is that when language comes onto the cognitive scene, it heralds not a transformation in those generic types of inner resource, but rather an augmentation of them.

In developing this picture, Clark explores an analogy with a more mundane external tool, namely a pair of scissors (Clark, 1997, p. 193). Scissors allow human beings to exceed their unaided manipulative capabilities by, for example, enabling us to make straight cuts in paper. It is here that the third component of Clark's external scaffolding claim comes out of the shadows. Part of the reason why scissors are such powerful augmenters of our unaided manipulative capabilities is that scissors are adaptively fitted to the shape and the capacities of the human hand (its ability to grip in a certain way, and so on). This important moral transfers to the case of language. Thus part of the reason why language is so effective in enabling the language-exploiting brain to exceed its unaided psychological capacities is that language is adaptively fitted to that kind of brain. Of course, scissors have been adapted to the hand largely through the conscious and deliberate efforts of human design, whereas (the thought is) language has become adapted to the brain through the blind and unintelligent engine of Darwinian selection, but that difference is not, in and of itself, a difference that makes a difference to the present point, which concerns the end result of the adaptive process.²

² In Clark's treatment, the claim that language is itself a system evolving under selection pressures—pressures established principally by the character of its evolutionary conduit, the human brain—is left somewhat skeletal. However, the idea is given compelling theoretical and experimental flesh by, for example, the Language Evolution and Computation Research Unit at the University of Edinburgh (see <http://www.ling.ed.ac.uk/lec>). As just one example of the exciting work coming out of this group, consider the following: Simon Kirby and his colleagues have used computer simulations to show that if (a) one places language in its learning context, that is as being passed on from one generation to the next by cultural transmission, (b) one begins, as seems evolutionarily likely, with a holistic language (one in which there is no systematic mapping from the structure of the symbols used to the structure of the meanings conveyed), (c) there is, as many have suggested, a transmission bottleneck in the language learning process (such that learners are exposed only to some impoverished subset of the language), and (d) language learners have a rudimentary domain-independent generalization capacity, then the language in question will evolve compositional structure. The reason for this is that compositional languages are generalizable languages, and generalizable languages can be recreated in each generation without exposure to the whole language. This makes such languages more evolutionarily stable, so that once a generalizing learner, by chance, stumbles across compositional structure, such structure will spread throughout the population. (For the details, see e.g. Kirby, 2002; Brighton, 2002; Kirby and Christensen, 2003; Smith et al., 2003) This result is exciting because, among other things, it sees the poverty of the stimulus as a force that drives the evolution of linguistic structure, rather than as a problem to be overcome by some Chomskyan innate language acquisition device.

In sum, then, Clark urges us to take seriously the thought that language is an externally located tool that boosts the unaided psychological capacities of the human brain, in part by being adaptively fitted to that brain. One might say that linguistic systems are like cognitive scissors. This vision bequeaths two immediate interpretative questions: ‘What exactly are the unaided cognitive capacities of the human brain?’ and ‘What exactly are the extra cognitive achievements that language makes possible?’. The answer that Clark gives to the first of these questions (an answer which reveals his broadly connectionist perspective) is that the human brain is essentially a device for pattern-association, pattern-completion and pattern-manipulation. (Henceforth I shall speak simply of the “pattern-completing brain” and assume that we mean to include the other, closely related capacities.) Thus Clark’s claim is that our language-involving behaviour is to be explained by an all-conquering partnership between, on the one hand, a pattern-completing brain and, on the other, an external storehouse of rich linguistic structures.

In order to appreciate just how radical Clark’s position here is, we need to disentangle its evolutionary dimension from its developmental one. And we can do that by noting the way in which Clark distances his position from a view (putatively) held by [Dennett \(1991\)](#). On an interpretation that Clark himself admits is tentative ([Clark, 1997, p. 197](#)), Dennett argues that our innate neural hardware may differ from that of our non-linguistic evolutionary near-neighbours (such as chimpanzees) in only relatively minor ways. Nevertheless, it is precisely these relatively minor hardware differences that constitute the evolutionary source of the human ability to create, learn and use public language. According to Clark, this part of Dennett’s story is correct: there is no mandate to attribute human beings with the kind of innate language processing mechanism whose design would mean that our brains, compared with those of our evolutionary near-neighbours, contain a fundamentally different kind of neural device. However, when we enter the developmental arena, Clark jumps the Dennettian train. Dennett’s further proposal, as Clark explains it, is that developmental exposure to a linguistic environment results in a subtle reprogramming of the computational resources of the human brain, such that our innate pattern-completing neural architecture comes to simulate a kind of logic-like serial processing device. Clark, by contrast, resists the idea of any extensive ontogenetic reprogramming phase driven by language. Thus, we are told, developmental exposure to and use of language brings about no *significant* reorganization of the brain’s processing architecture. In that sense, language remains external and the brain remains a pattern-completing machine (more on this in a moment).

Now let’s turn to our second interpretative question: what are the extra cognitive achievements that language makes possible? This is where Clark’s first major claim re-surfaces and assumes its full importance. We can all agree that language enables sophisticated communication, but, as we have seen, Clark alerts us to a rather different functional contribution that language may make to our lives, namely its power to transform the character of certain problems so that those problems are rendered amenable to the pattern-completing strengths of the biological brain. Clark explicates this thought using a host of examples (see [Clark, 1997, especially](#)

pp. 200–211). For present purposes we can make do with a representative sample of illustrative cases (some of which I have adapted for local effect).

One simple manifestation of the phenomenon of interest is that we sometimes use concrete linguistic artefacts in the world (notebooks, shopping lists, memo boards, etc.) as data stores (especially where the body of data is large and/or complex), thereby reducing the demands on our limited and overworked biological memory. More impressively, we sometimes use linguistic structures to organize actions, of both an individual (writing yourself a note to record Buffy the Vampire Slayer) and a collective (sending group e-mails to organize a celebratory drink) kind. Furthermore, instantiations of language in external media may themselves be non-trivial factors in a distributed reasoning or creative process that flows between and through both the inner resources of the brain and the external resources of the outer environment (more on this sort of thing in the next section). Thus chunks of printed or electronic text may be used to preserve half-baked ideas, or moved around so as to be juxtaposed in newly suggestive ways, or stored for later recall and remanipulation, and so on. And even where the straightforwardly communicative function of language is in the frame, it may still be the case that one thing that public language is doing is enabling individual human minds to overcome certain blockages in their local cognitive trajectories. Thus it seems to be a fact about human cognition that it is heavily path-dependent (i.e., where one can get to in cognitive space depends on where one is and where one has been in that space). But, given the differences between human brains, it is plausible that “one agent’s local minimum [may be] another’s potent building block” (Clark, 1997, p. 206), so communal idea-sharing through linguistic communication may help groups of cognizers to transcend the limitations of individual path-dependence.

The foregoing examples lend powerful support to Clark’s claim that language may extend our unaided cognitive capacities by acting as an external resource that beneficially alters the basic shape of certain problem-solving scenarios. However, other examples that Clark gives of the transformational and augmentational power of language seem, at first sight, to reinvent a vision of language as being, in some sense, inner. For example, he notes that self-directed speech (whether silent or aloud), especially in the form of repeated instructions to oneself, sometimes sets up a control loop that enhances (novice or expert) problem-solving performance. (Think of the squash player silently telling herself over and over again in between shots that she really must get to the ‘T’.) And even more suggestive of an essentially inner aspect to language, one might think, is our introspectively manifest ability to run through sentences ‘in our heads’.

Interlude: Clark suggests that our ability for linguistic inner rehearsal may be a crucial psychological stepping stone, since it may be precisely this ability which enables us to think about our own thinking (in the sense of spotting a problem with one of our judgments or beliefs, uncovering the logical or illogical transitions in our arguments, etc.). The idea (Clark, 1997, p. 209) is that the very process of formulating a thought in words effectively creates that thought as a stable object available for evaluation. That is, thinking in language fashions thoughts as the sort of thing that the thinker can have thoughts about. The fact that words bestow this kind of

stability can be traced, Clark argues, to the communicative function of language, on the grounds that successful communicative interactions (of the kind that language supports) require a code that is (a) context neutral (or largely so), (b) modality independent (e.g., linguistic inputs and outputs may be visual, auditory or tactile), and (c) supportive of the relatively easy learning of simple linguistic structures.

Now, depending on how one thinks about language, one might be tempted to challenge Clark's confident espousal of (a)–(c) as necessary conditions for successful linguistic communication. However, let's allow that at least some forms of ordinary linguistic communication bear these hallmarks. That would be enough for Clark to run his suggestion concerning the linguistic source of second-order thinking (having thoughts about thoughts). A second challenge to this aspect of Clark's account is less easily deflected. One might wonder whether language really is the *only* route to the kind of cognitive stability that, on Clark's view, supports the presence of second-order thinking. Indeed, consider the kind of systematic behaviour observed in many non-linguistic animals, according to which a creature capable of responding selectively to one input is capable of responding selectively to many semantically related inputs. Fodor (1987) has argued that this behavioural systematicity is good evidence for some kind of combinatorial structure in the causally efficacious inner states involved, such that those states are constructed out of the same or overlapping simpler recurrent elements. Let's say that Fodor is right, *in principle*, that is, that it *could in principle* be the case that rats and pigeons have a combinatorially organized system of thought. Given that one standard condition on a system being combinatorial is that for any recurrent element within that system, that element must make approximately the same contribution to each of the many different larger structures in which it figures, there seems to be no obvious reason why, in principle at least, a combinatorial system of non-linguistic inner states couldn't deliver the kind of stability which Clark reserves for language.³ Of course, that isn't to say that combinatorially endowed rats and pigeons would thereby enjoy thoughts about their own thoughts. Such creatures presumably lack the additional cognitive machinery required to turn that impressive psychological trick. However, it does cast Clark's speculation that "public language... is responsible for... the ability to display *second-order cognitive*

³ For Fodor, of course, the behavioural systematicity in play here indicates that animals such as rats and pigeons have a language of thought. I have avoided putting things this way. For one thing, I am not claiming that rats and pigeons *in fact* have a language of thought, only that we can make sense of the idea that they might. More importantly, in the present context, I am trying to distance the kind of combinatorially organized system of inner states that may possibly be present in such animals from public language. So describing such a system as a language of thought would simply muddy the waters (and they're murky enough already). To be clear, however, a Fodorian language of thought is not a public language. Indeed, for Fodor, it couldn't be. This is because, according to Fodor, learning a natural language is a species of concept acquisition, and concept acquisition works by a method of hypothesis and test. This in turn requires the existence of a language-like system in which candidate hypotheses may be expressed. Thus, Fodor concludes, for us to learn our natural language, we must already have a prior (indeed, to avoid an infinite regress, an innate), in-the-head, language-like system in place; hence the language of thought (Fodor, 1975). What all this means is that as long as we reserve the term 'linguistic' for public language, it is correct to speak of non-human animals as, in principle, having a non-linguistic but combinatorially organized system of inner states.

dynamics [second-order thinking]” (Clark, 1997, p. 208) in a different light. Language may very plausibly be *one possible* source of such second-order thinking; but it may not be the *only possible* source. In other words, language isn’t strictly necessary for second-order thinking.

Let’s say that the foregoing argument is sound. Would conceding the conclusion do serious damage to Clark’s position? I don’t think so. It seems to me that Clark could simply retreat to the weaker claim that, as a matter of contingent empirical fact, it’s language that provides the route to second-order thinking in humans. Indeed if, as a matter of contingent empirical fact, no other animals realize second-order thoughts, the claim that language has played such a role would look to be plausible, or at least to be an idea worth pursuing. In any case, for present purposes we need not get overly hung up on this issue, since while the underlying phenomenon of linguistic inner rehearsal will exercise a good deal of our critical attention in what follows, the second-order thinking that such rehearsal may possibly support will be of only secondary interest.

Interlude over: let’s get back to the main plot, which at this point concerns the worry that the very fact of linguistic inner rehearsal (in particular) seems to demand that we think of language as having a robustly internal dimension. This seems to be in conflict with Clark’s claim that language is an external resource. And the whiff of inconsistency here becomes stronger in the wake of the following remark that Clark makes: “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 treat language and culture as wholly external resources” (Clark, 1997, p. 198). What is going on? As it happens, the apparent tension in Clark’s account may be relieved if we interpret him as being implicitly sensitive to two different senses in which language may be inner, only one of which we have met already. According to the first (newly exposed) sense, language is inner just so long as there are private thought processes which are formulated in language.⁴ This is a sense in which Clark would, I think, be happy to say that language is, in part, an inner resource. According to the second sense in which language may be inner (scouted earlier), language is inner just so long as there is, in the brain, a domain-specific language processing system, one whose elements are organized, at a fundamental level, so as to encode the structural properties of natural language. This is a sense in which Clark would, I think, want to reject the claim that language is inner. And crucially, according to Clark, the phenomenon of linguistic inner rehearsal does not force this unwanted vision upon us. As he himself puts the point, citing connectionist research on language processing in support, “it remains possible that... [inner linguistic] rehearsal does not involve the use of any fundamentally different kind of computational device in the brain so much as the use of the same old (essentially pattern-completing) resources to model the special kinds of behaviour observed in the world of public language” (p. 198). So, to recall a quotation from earlier, Clark’s claim is that linguistic competence, even of the mental rehearsal kind, does not require a transformation in “the brain’s own basic modes of

⁴ “Private” here means ‘unvocalized in public earshot by the thinker when thought (plus non-auditory equivalents)’ not ‘inaccessible to others in principle.’

representation and computation.” As we noted last time around, it is this claim which provides a sense in which, for Clark, language remains external.

We shall be returning to the issue of linguistic inner rehearsal and exactly what it tells us about the externality or otherwise of language later. Right now I want to round off my preliminary analysis of Clark’s view with a few critical remarks about where he ends up. The final twist in the plot is Clark’s claim that once we have signed up for the image of our linguistic abilities that he has advocated, the very ideas that have been used to construct that image come under critical pressure. Thus the illuminating power of a term such as ‘tool,’ as well as the distinctive interlocking suggestion that language must be conceived as externally located, are seemingly placed in question when Clark concludes that language is “in many ways the ultimate artefact: so ubiquitous it is almost invisible, so intimate it is not clear whether it is a kind of tool or a dimension of the user” (p. 218). It is important to note here that in claiming (a) that the intimacy between language and user means that there is a mandate to conceive of language as a dimension of the user, Clark is not claiming (b) that the intimacy between language and user means that there is a mandate to conceive of language as inner (in the previously rejected sense of that term). Clark’s point, as I read him, is that for the purposes of conceiving the relationship between the language-user and her language, the boundary between what we have previously been thinking of as two systems may sometimes collapse, such that there is really just a single system, one dimension of which is language. So the ultimate artefact is one that rebels against any separation between it and its user. But it is not as if we should, in the end, shift language from the external to the internal side of some persisting agent–environment boundary. Rather, the local agent–environment boundary is itself in danger of disappearing, which is why terms such as ‘tool’ (conceived as something in the environment which agents use) and ‘external’ become problematic.

Despite appearances, this is not quite a case of early-Wittgensteinian ladder dumping.⁵ To see why, we need only plug in a piece of thinking by a philosopher whom Clark (2003) has described as a subterranean influence on his own work, namely Heidegger. If there’s one bit of Heidegger that’s passed into mass philosophical and cognitive-scientific consciousness, it’s his phenomenological analysis of tool-use (Heidegger, 1926).⁶ According to Heidegger, when skilled tool-use is smooth and uninterrupted, the human agent has no conscious experience of the tools in question as *independent objects*. Thus, to use Heidegger’s most famous example, while engaged in trouble-free hammering, the skilled carpenter has no conscious recognition of the hammer, the nails, or the work-bench, *in the way that one would if one simply stood back and thought about them* (Heidegger, 1926). Considered as independent

⁵ For the uninitiated, the penultimate proposition of Wittgenstein’s *Tractatus Logico-Philosophicus* contains the following, eminently quotable remark: “My propositions are elucidatory in this way: he who understands me finally recognizes them as senseless, when he has climbed out through them, on them, over them. (He must so to speak throw away the ladder, after he has climbed up on it.)” (Wittgenstein, 1922).

⁶ In actual fact, Heidegger was concerned to give an analysis of our skilled dealings with *equipment*, where the term ‘equipment’ has a special, technical meaning, but tool-use is close enough.

objects these tools-in-use become, as it were, phenomenologically transparent. Moreover, Heidegger observes, not only are the hammer, nails, and work-bench in this way not part of the engaged carpenter's phenomenal world, neither, in a sense, is the carpenter! The carpenter becomes absorbed in his activity in such a way that he has no awareness of himself as a subject over and against a world of objects. So, in the domain of smooth and uninterrupted skilled tool-use there are, phenomenologically speaking, no subjects and no objects; there is only the experience of the ongoing task (e.g., hammering).

So what? The message is that hammers too are, or rather under the right circumstances can be, ultimate artefacts. Of course, Heidegger's example turns on phenomenological analysis rather than any hypothesis about the causal mechanisms involved, but given the (in my view) highly plausible principle that phenomenological experience will often reflect its causal underpinnings, not much may hang on that difference.⁷ In any case, we're surely in the same ballpark. Once again the key claim is that the intimacy between user and tool collapses the local agent–environment boundary, leaving behind just one system. But that means that with respect to the present point, and pace Clark, there's nothing special about language when compared with more familiar tools and artefacts. Any tool can meet the condition of intimacy required of an "ultimate artefact." It simply needs to be used skillfully in a hitch-free manner, such that user and tool are best conceived as a single system. This is, of course, a temporary status achieved by a tool when the dynamics of use are of a certain fluid and undisrupted kind. And crucially, for any tool, there will be other contexts of activity in which the condition of intimacy will fail to be met. Heidegger identified two broad categories of such contexts—*disturbances*, in which, for example, a tool breaks, and *detached reflection*, in which, for example, we consider the tool as an object for scientific or philosophical investigation. In such contexts, a kind of distance is established between the agent and the tool.⁸

One might think that this is the moment where my attempt to lump language in with other tools breaks down. "Surely," I hear you say, "there is a dimension of intimacy that is particular to language, namely that we humans are *always* in intimate contact with language." (This might be the way to read Clark's remark that language is ubiquitous.) However, if we continue to draw on a Heideggerian analysis, the force of this objection is blunted. As I interpret him, Heidegger too thought that language was a cognitive tool.⁹ Indeed, he argues explicitly that we encounter words and sentences in all the modes in which we encounter other tools (Heidegger, 1926). Thus, in free-flowing conversation, we employ words in appropriate ways without being

⁷ For an extended discussion of the principle in question, plus a defence of the idea that Heidegger signed up for it, see Wheeler (2004b).

⁸ For much more by way of exegetical discussion, see, for example, Dreyfus (1991) and Wheeler (2004b).

⁹ I should confess that while I believe that this is precisely how Heidegger thought of language, it is a far from uncontroversial interpretation. Guignon (1983), for example, explicitly rejects the tool interpretation in favour of an alternative reading according to which, for Heidegger, language is a constitutive precondition for meaningful experience.

aware of those words as objects to be manipulated. However, there are situations of disturbance in which language-use breaks down and the right words become difficult to find. Finally, we sometimes remove language from its everyday contexts of use, and treat it as an object of scientific or philosophical study, as in formal linguistics or philosophy of language. Where language-use is disturbed, or language becomes the object of detached scientific or philosophical reflection, the condition of intimacy fails to be met.

Given this Heideggerian amendment, the target of my critical pressure in what follows will not be Clark's inference from the conceptualization of language as cognitive scissors to the suspicion that the agent–environment boundary between language-user and language may collapse. If tools in general (cognitive or otherwise) facilitate such boundary-collapsing events, and if language is indeed best conceived as a tool, then language will facilitate such boundary-collapsing events. Given that I accept the boundary-collapsing power of smooth and uninterrupted skilled tool-use, what I think stands in need of careful evaluation here are some of the details of Clark's vision of language as a (cognitive) tool. If language is incorrectly conceived as an artefact, it certainly can't be the, or even an, ultimate one.

3. Distributed cognition

Time to take a step back. Clark's approach to language reflects a more general perspective on mind, cognition and intelligence, one which has been gaining ground recently in cognitive science, and of which Clark is a prominent advocate. In order to clear the way to the difficulty that I want to raise for Clark's account of language, we need to make contact with that wider framework. In the contemporary philosophical and cognitive-scientific literature, it trades under a bewildering plethora of different names, including situated cognition, embodied cognition, embedded cognition, embodied-embedded cognition, active externalism, vehicle externalism, the extended mind and (the moniker we'll be using) distributed cognition. In fact there are almost certainly subtle nuances of approach here which someone with an over-enthusiasm for drawing distinctions and too much time on their hands might exploit to divide these various trends from each other. This is not the place to play that particular game, however, since we'll be concerned with a specific proposal about psychological phenomena that has purchase across the board. The proposal in question is that, under certain conditions, the

organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right. All the components in the system play an active causal role, and they jointly govern behaviour in the same sort of way that cognition usually does. If we remove the external component the system's behavioural competence will drop, just as it would if we removed part of its brain. Our thesis is that this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head (Clark and Chalmers, 1998, p. 7).

Notice that the claim on the table here is that there are conditions under which something which counts as a single cognitive system or as a single cognitive process contains some elements which are agent-internal and some which are agent-external. It is *not* the claim that I deliberately sidelined a little earlier, namely that there are conditions under which the very idea of an agent–environment boundary becomes misleading. What seems likely is that unless we have a case of the former, we wouldn't even be tempted to say that we have a case of the latter; but that doesn't make the two cases equivalent.

For our purposes, Clark and Chalmers' proposal may be unpacked as follows: we have a case of genuinely distributed cognition where (i) the source of intelligent action is to be found not purely in the inner activity of neural states and processes, but rather in complex causal interactions between neural factors and additional (extra-neural) elements located in the non-neural body and the environment, and (ii) the behaviour-generating causal contribution of the additional elements is of the same kind as the corresponding contribution of the brain, in that those additional elements account directly for some of the distinctive adaptive richness and flexibility of the observed behaviour.¹⁰ Elsewhere, Clark and I have played out what amounts to the same idea in terms of what we have dubbed *non-trivial causal spread* (Wheeler and Clark, 1999). One has a case of non-trivial causal spread when (a) some phenomenon of interest (e.g., intelligent behaviour) turns out to depend, in unexpected ways, upon causal factors external to the system previously or intuitively thought responsible, and (b) the newly discovered additional causal factors reveal themselves to be at the root of some distinctive target feature of the phenomenon of interest (e.g., the adaptive richness and flexibility of intelligent behaviour). (If condition (a) alone is met, then the causal spread will be trivial in character. In such cases, while we may find it surprising that certain additional factors turn out to play some supporting causal role in generating the phenomenon of interest, the distinctive character of that phenomenon will still be correctly traced to the system previously or intuitively thought responsible.) Where one's best cognitive science displays a case of non-trivial causal spread, one will have a case of distributed cognition.

So, if we go looking for non-trivial causal spread, where will we find it? Consider the following problem.¹¹ A robot with a control system comprising an artificial neural network and some rather basic visual receptors is placed in a rectangular dark-walled arena. This arena features a white triangle and a white rectangle mounted on one wall. The task is to set up the robot's control system so that, under wildly varying lighting conditions, it will approach the triangle but not the rectangle. The

¹⁰ In another context (Wheeler, 2004a), I've described this proposal as specifying the conditions for *extended cognition*, so that's another label for the list. In the wider philosophical literature, the target position (or something very close to it) is defended, in different ways, by (among others) Haugeland (1995/1998); Hurley (1998); Rowlands (1999) and Wheeler (2004b).

¹¹ I have used this example before. The present analysis draws directly on the treatments to be found in Wheeler (2001), Wheeler (2004b) and Wheeler (2004c).

specific architecture of the neural network, the way in which the network is coupled to the visual receptors, and the field-sizes and spatial positions (within predetermined ranges) of those visual receptors are all to be determined. That's the design specification; so why don't you play along by taking a moment now to think about the general kinds of states and structures that you believe will be needed. . .

Finished? OK. If you are a cognitive-scientificly minded individual who has not been exposed to the delights of distributed cognition, then I'm prepared to bet that your rough design looks something like this: The robot will need a way of internally representing triangles, rectangles and the spatial layout of its environment. Its strategy should be to build the best map it can of its environment, locating the triangle and the rectangle as accurately as possible on that map. The robot should then plan a path to the triangle and follow it. You might have decided to trade in reliability for speed, and thus to compensate for both the fact that the robot's visual receptors are pretty simple and the fact that the lighting conditions will be changing radically. If so, you will probably have built in a number of "stop-and-check" stages in which the robot pauses to assess the accuracy of its map and (if necessary) to adjust that map and revise its path.

That's certainly one way to go. But here's an alternative, computationally cheaper (and thus adaptively more efficient) strategy, revealed when Harvey et al. (1994) presented the self-same problem to a design methodology known as *evolutionary robotics*. In evolutionary robotics, algorithms inspired largely by Darwinian evolution are used to automatically design the control systems for (real or simulated) robots.¹² The solution arrived at in this manner was canny, to say the least. Two visual receptors were positioned geometrically such that visual fixation on the oblique edge of the triangle would typically result in a pair of visual signals (i.e., receptor 1 = low, receptor 2 = high) which was different from such pairs produced anywhere else in the arena (or rather *almost* anywhere else in the arena; more on this shortly). The robot would move in a straight line if the pair of visual signals was appropriate for fixation on the triangle, and in a rotational movement otherwise. Thus if the robot was fixated on the triangle, it would tend to move in a straight line towards it. Otherwise it would simply rotate until it did locate the triangle. Occasionally the robot would fixate, "by mistake," on one edge of the rectangle, simply because, from certain angles, that edge would result in a qualitatively similar pair of visual

¹² Roughly speaking, the evolutionary robotics methodology is to set up a way of encoding robot control systems as genotypes, and then, starting with a randomly generated population of controllers, and some evaluation task, to implement a selection cycle such that more successful controllers have a proportionally higher opportunity to contribute genetic material to subsequent generations, i.e., to be 'parents.' Genetic operators analogous to recombination and mutation in natural reproduction are applied to the parental genotypes to produce 'children,' and (typically) a number of existing members of the population are discarded so that the population size remains constant. Each robot in the resulting new population is then evaluated, and the process starts all over again. Over successive generations, better performing controllers are discovered. For useful introductions to evolutionary robotics, see Husbands and Meyer (1998) and Nolfi and Floreano (2000).

signals being generated as would have been generated by the sloping edge of the triangle. Perturbed into straight line movement, the robot would begin to approach the rectangle. However, the looming rectangle would, unlike a looming triangle, produce a change in the relative values of the visual inputs (receptor 1 would be forced into a high state of activation), and the robot would be perturbed into a rotational movement. During this rotation, the robot would almost invariably refixate on the correct target, the triangle.

This is a demonstration of non-trivial causal spread. The systematic activity of inner maps, route-planning algorithms and so on (as present in our cognitive-scientifically intuitive solution) has been replaced by a suite of organized interactions involving significant causal contributions not only from states and processes in the robot's brain, but also from certain additional bodily factors and from the environment. It is true, of course, that the robot's artificial brain contributes to the observed adaptive success (although it is worth noting that the evolved neural network was, structurally speaking, quite simple). But a further agent-side factor that plays a non-trivial part in the story is the spatial organization of the robot's visual morphology. Indeed, one might well think that it is this geometric fact about the agent's bodily periphery that principally explains the robot's ability to become, and then to remain, locked onto the correct figure. The crucial role played by the environment becomes clear once one realizes that it is the specific ecological niche inhabited by the robot that enables the selected-for strategy to produce reliable triangle-rectangle discrimination. If, for example, non-triangle-related sloping edges were common in the robot's environment, then although the evolved strategy would presumably enable the robot to avoid rectangles, it would no longer enable it to move reliably towards *only* triangles. So adaptive success depends not just on the work done by the agent-side mechanisms, but also on the tight coupling between those mechanisms and certain specific structures in the environment which can be depended upon to be reliably present.

About now, Clark and Chalmers' words should be ringing in our ears: "if we remove the external component the system's behavioural competence will drop, just as it would if we removed part of its brain. . . this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head." With this principle back in full view, it seems clear that we can understand Harvey et al.'s triangle-rectangle discrimination robot as realizing a distributed-cognition-style adaptive solution. Moreover, the robot illustrates an important feature of such solutions, a feature that bears emphasis now because it will turn out to be crucial to my argument later. In paradigmatic cases of distributed cognition, adaptive success ensues because, during the actual run-time of the behaviour, the internal elements become *directly causally locked onto* the contributing external elements. That's the nature of the coupling relationship. And that's why the removal of the external elements results in significant behavioural degradation or collapse. Notice that the same run-time dependence does not exist for more traditional cognitive architectures which deal in detailed (or relatively detailed) internal representations of the environment. In these architectures, once the salient internal surrogate has been built, it's *that* structure (rather than its external source) with which the other

inner elements enter into direct causal commerce, in run-time, in order to guide behaviour.¹³

Still, while Harvey et al.'s robot helps to illuminate the key principles of distributed cognition, the adaptive problem it solves seems to be a long way from anything like language-use. So here's another example of those key principles at work, one which is intuitively closer to our primary target. In a passage written long before the distributed cognition paradigm emerged as a well-formed cognitive-scientific research programme, but in which there appears the prescient phrase "the external environment becomes a key extension to our mind," Rumelhart et al. note that most of us solve difficult multiplication problems by using "pen and paper" as an external resource. This environmental prop enables us to transform a difficult cognitive problem into a set of simpler ones, and to temporarily store the results of those intermediate calculations (Rumelhart et al., 1986, quote from p. 46). Thus, as we might now put it, the externally located pen-and-paper resource makes a non-trivial causal contribution to the observed problem-solving behaviour, and the distributed combination of this resource and certain inner psychological processes constitutes a cognitive system in its own right.

It is by way of Rumelhart et al.'s compelling example that Clark introduces the following thought: the agent-internal contribution to mathematical problem-solving is likely to be 'just' a matter of pattern completion, and so will be amenable to connectionist modelling (Clark, 1997, pp. 61–62). This is the mathematics-specific version of a more general claim that we can understand Clark as making, namely that the agent-internal contribution to the kind of distributed problem-solving that turns on the active exploitation of external symbols is likely to be 'just' a matter of pattern completion, and so will be amenable to connectionist modelling. Of course, as we have seen already, Clark applies this same reasoning to another specific case of external symbol exploitation—language-use. Indeed, we might strengthen the link here further by saying that systems of mathematical symbols, like language, are

¹³ To be clear, I am not intending to suggest that fans of distributed cognition must eschew any concept of internal representation. The status and character of representational explanation in distributed cognition is a complex and subtle issue, and this paper is certainly not the place to explore it. For my own thoughts, see Wheeler and Clark (1999), Wheeler (2001), Wheeler (2004a), Wheeler (2004b) and Wheeler (2004c). However, put briefly (and thus inadequately), one general message is that where a representation-exploiting control strategy is an example of distributed cognition, the kinds of representations paradigmatically deployed will make their contribution to adaptive success not by internally specifying, in any detailed way, the objective properties and relations of the external environment, but rather by coding, sparsely and temporarily, only for certain context-specific properties (often defined in an egocentric manner), and by working in close and ongoing interaction with the environment itself. The reliance on regular sensing here is more fundamental than in the "stop-and-check" strategy mentioned earlier, since there will be no sense in which the distributed solution involves the construction, abandonment or re-construction of an overall plan for achieving the intended goal. Goal-achieving behaviour emerges out of the pattern of ongoing interactions between the inner and outer elements in the distributed cognitive system, without the need for any internally maintained global plan. For a distributed solution that features this kind of representational contribution, see e.g. (Franceschini et al., 1992, discussed in Wheeler, 2001, Wheeler, 2004b).

external cognitive tools that augment and extend the unaided psychological capacities of the human brain.

I could go on all day giving intriguing examples of distributed cognition. There's a lot of it around. (For many more examples see Clark, 1997; Wheeler, 2004b, among other treatments.) For the purposes of the present paper, however, we have now learned enough about the general form of the phenomenon to turn our attention back to Clark's distributed-cognition-style vision of language. So what exactly is wrong with the idea that language is an external tool which augments the pattern-completing activities of the human brain? In the next section I shall explore some problems faced by Clark's account.

4. Off-line language

It seems clear that the natural home of non-trivial causal spread, and thus of distributed cognition, is in the domain of what might be called *on-line intelligence* (Wheeler and Clark, 1999). A creature displays *on-line intelligence* just when it produces a suite of fluid and flexible real-time adaptive responses to incoming sensory stimuli. Examples might include escaping from a predator, catching a prey, tracking a mate, taking a catch in cricket or baseball, manipulating written mathematical symbols to solve a complex multiplication problem, or holding a lively conversation. By contrast, a creature displays *off-line intelligence* just when it disentangles itself in some way from the ongoing perception-action cycle. Examples might include wondering what the weather's like in Durban now, mentally planning that imminent trip to London, or *doing complex multiplication in one's head*.¹⁴ The fact that I have emphasized the final example here is not idle, because the counterbalance to the unquestionable observation that most of us solve difficult multiplication problems by using the external prop of pen and paper is that some lucky souls can solve difficult multiplication problems in their heads without the use of such cognitive scaffolding. In the latter case it seems fair to say that the inner mechanisms involved are functioning in an off-line way. Given the thought that, as we noted above, mathematical symbol systems may, in the present context, be treated as close-cousins of

¹⁴ Recently, some thinkers (e.g., Esther Thelen and Andy Clark, both in seminar discussion) have raised doubts about the on-line/off-line distinction, on the grounds that no intelligent agent is (they claim) ever wholly on-line or wholly off-line. On this view, intelligence is always a dynamic negotiation between on-line and off-line processes. I have no doubt that this interactive view is true for many cases of (at least) human intelligence. Producing this paper, for example, has been an unfolding interplay between off-line reflection and on-line activities such as cutting and pasting text (cf. Clark, 1997, pp. 206–207). However, even if such composite or intermediate cases abound, there are also cognitive achievements that seem to fall squarely into one category or the other. Moreover, even where we judge an observed behaviour to be a composite case—one in which there is an interaction between off-line and on-line components processes—the on-line/off-line distinction is in fact still being used as a way of classifying those component processes and we can still ask meaningful questions about their character as members of one category or the other; so the distinction remains conceptually and explanatorily useful at that level.

language, we can beat a path back to language by examining the distinction between on-line and off-line mathematical reasoning.

What kind of cognitive innards might plausibly support off-line mathematical reasoning? Drawing again on Rumelhart et al.'s rich discussion, Clark suggests that this is a case of learning "to manipulate a mental model in the same way as we originally manipulated the real world" (Clark, 1997, p. 61). In other words it's another instance of inner rehearsal. So how does the mechanics of this process work? Here Clark offers the following argument (again using Rumelhart et al.'s treatment as a source of insights): "experience with drawing and using Venn diagrams allows us to train a neural network which subsequently allows us to manipulate imagined Venn diagrams in our heads... there is no reason to suppose that such training results in the installation of a different *kind* of computational device. It is the same old process of pattern-completion in high-dimensional representational spaces, but applied to the special domain of a specific kind of external representation" (p. 199).

There is something odd about Clark's final remark here, in that it seems to run together the learning and the performance phases of the target behaviour. To be sure, the particular external representations in question must provide the source domain for a process in which certain inner mechanisms learn how to be sensitive to and then how to manipulate certain patterns, in the correct way. However, once that training phase is complete, and we thus have the inner resources to perform off-line mathematical reasoning (conceived as the inner rehearsal of a process that previously involved external symbol manipulation), there is a straightforward mechanical sense in which, during such off-line reasoning, the "process of pattern-completion in high-dimensional representational spaces" must be "applied to" certain inner states and structures. After all, the whole point about off-line cases is the absence of any ongoing interaction with the environment. That's how the representations in which we are interested here earn their adaptive keep, by standing in, within the inner processing economy, for certain absent environmental factors. But Clark's treatment leaves the structure of those inner surrogates awkwardly mysterious.

Why awkwardly so? It is here that our foray into the general character of distributed-cognition-style adaptive solutions yields dividends. There we learned that, in paradigmatic cases of distributed cognition, adaptive success ensues because, during the actual run-time of the behaviour, certain internal elements become directly causally locked onto the contributing external elements. This general principle produces the following local picture. When the mathematical reasoning in which we are interested is a case of on-line distributed cognition, the inner pattern-completing mechanism will be directly causally locked onto certain properties of mathematical symbols located in the environment. But when our mathematical reasoning is off-line, there are, by hypothesis, no such environmental factors onto which the mechanisms concerned could be locked. However, we can easily hold onto the Clark-inspired thought that fundamentally the same kind of processing mechanism (perhaps even the very same pattern-completing mechanism) may still be deployed, just so long as there are in place certain inner surrogates for those missing environmental factors, surrogates which recapitulate certain structural properties of those factors, viz. the ones, whatever they may be, to which the mechanisms concerned are designed so

as to be mechanically keyed. So what we need are inner surrogates that realize certain structural properties of mathematical symbols. That's what the inner models that support off-line mathematical problem-solving will be like.

Clark's account of mathematical reasoning may be summarized as follows: (i) on-line mathematical reasoning essentially involves the manipulation of external symbols; (ii) off-line mathematical reasoning is a case of mathematical inner rehearsal; and (iii) fundamentally the same inner processing mechanisms are active in both cases. This account shapes his view of language-use. To see this, we merely need to restage some of the core aspects of Clark's view of language using the on-line/off-line distinction. Thus (i) on-line language-use essentially involves the manipulation of external symbols (words, sentences, etc.); (ii) off-line language-use is a case of linguistic inner rehearsal (so it requires the internal modelling of certain external symbolic elements); and (iii) fundamentally the same inner processing mechanisms are active in both cases. But, given this, our recent moral ought to transfer too. Thus just as off-line mathematical reasoning requires inner representations that recapitulate certain structural properties of external mathematical symbol-systems, namely those structural properties which are non-trivial contributing factors in on-line distributed mathematical problem-solving, so off-line language-use requires inner representations that recapitulate certain structural properties of external linguistic symbol systems, namely those structural properties which are non-trivial contributing factors in on-line distributed linguistic performance. Exactly what those properties are is a matter for theoretical debate and for interdisciplinary empirical investigation. That they need to be internally recapitulated is an unavoidable consequence of the basic logic of the distributed cognition paradigm.

"So what?" you might be thinking. "Wheeler has agreed, hasn't he, that the computational process active in off-line mathematical problem-solving, and now by extension the computational process active in off-line language-use, may remain one of pattern-completion? Therefore the shortfall identified in Clark's account of the inner models required fails to damage Clark's view that neither the capacity for off-line mathematics nor the capacity for off-line language-use demands the installation of a different kind of computational device." So much is true. But now let's return to one of Clark's position-defining remarks about language. Clark argues that language is "an external resource that complements but does not profoundly alter the brain's own basic modes of *representation* and computation" (p. 198, emphasis added).¹⁵ As far as I can see, while the sub-claim concerning the neutral effects of language on the brain's own basic mode of computation can still be sustained following my critical comments, the sub-claim concerning the neutral effects of language on the brain's own basic mode of representation cannot.

¹⁵ Clark gives the putative persistence of the brain's own basic representational format far less prominence during his discussion of mathematical reasoning. But it is there. Consider, for example, the claim that off-line mathematical reasoning is "the same old process of pattern-completion in *high-dimensional representational spaces*" (Clark, 1997, p. 199, emphasis added). More on high-dimensional representational spaces in a moment.

It is here that we need, perhaps belatedly, to try to say exactly what Clark means by the phrase “the brain’s own basic mode of representation”. We know that the computational counterpart to this notion identifies pattern-completion in the style of connectionist networks. So it seems reasonable to infer (and there is plenty of textual evidence to back up the inference; see e.g. Clark, 1997, p. 141) that, for Clark, the brain’s own basic representational format is broadly connectionist in form. Thus it will be one of distributed (in the connectionist sense of that term), multi-dimensional patterns of activation and similarity metrics. Having said that, it should be noted that Clark sees these patterns and metrics through the lens not of traditional connectionism, but of contemporary neuroscience, dynamical systems approaches to cognitive science and, of course, the distributed cognition paradigm (see e.g. Clark, 1997, p. 174). So the brain’s own basic mode of representation is a substantially recontextualized, broadly connectionist one. The claim of interest, then, is that language-use, on-line or off-line, does not require any radical changes to this format.

At this juncture we need to tread with care. One might be tempted to think that Clark’s persistence claim amounts only to the demand that any new structural properties which we are compelled to introduce into our inner processing story, in order to account for linguistic competence, must be *implemented* in a distinctively brain-like form, that is, in terms of high-dimensional patterns of activation and similarity metrics in connectionist-style networks. This requirement, however, is too weak for what Clark wants. After all, even Fodor and Pylyshyn, those arch enemies of everything connectionist, agree that connectionist-style states and processes may be used to implement classical systems, systems whose fundamental structural properties include a language-like combinatorial syntax (Fodor and Pylyshyn, 1988). So if what Clark’s claim required were nothing weightier than the persistence of the brain’s own basic representational format at an implementational level, his position would be consistent with the following suggestion: the kinds of phenomena that we standardly take to be realized by language—infinite productivity, overwhelming systematicity and so on—can be accounted for only if a classical representational format (one which features a language-like combinatorial syntax) is understood to be a fundamental feature of the neural economy at some ‘higher,’ non-implementational level. But any such suggestion is surely at odds with Clark’s more radical intended message, which (it is now clear) must be that the inner processing story that we need to tell, in order to account for linguistic competence, may be couched *exhaustively* in terms of the structural properties which define the brain’s own characteristic style of representation (at root, high-dimensional patterns of activation and similarity metrics).

From what we have seen, on-line language-use (as conceived within the distributed cognition paradigm) may pass Clark’s more-than-implementation persistence test, since it involves an interactive real-time combination of domain-general inner mechanisms and external linguistic symbol systems. Given such teamwork, the organizational properties (e.g. compositionality) that plausibly count as the source of the distinctive features of language mentioned above may be realized environmentally, in the external symbol systems, *without* being recaptured internally. So the inner states in play will not need to be reconfigured significantly to support successful

performance in the linguistic domain. However, the fundamental logic of the general distributed-cognition framework in which Clark's view is embedded dictates that off-line language-use (linguistic inner rehearsal) requires the presence of inner surrogates that *do* recapture these organizational properties, or at least key aspects of them (see my closing thought below). Thus in the off-line case we confront nothing less than a profound transformation in the brain's own basic mode of representation, and that runs contrary to Clark's avowed view.¹⁶

5. The stings in the tale

I warned you at the outset that this wasn't going to be a blanket demolition of Clark on language. Still, it may seem that I have spent my time focussing on a rather narrow issue in Clark's account, to the exclusion of much that is rich and illuminating in the wider picture. To reassure you that the issue on which I have been concentrating is a far from minor one, I want to conclude by pointing out just how much hangs on it. First I shall identify the direct implications of the preceding reflections for Clark's account of language as a cognitive tool. Then I shall argue that what my argument about off-line language-use demonstrates is that in the case of a creature who has the capacity to engage in off-line thinking, one cannot argue for the distributed cognition perspective on grounds of adaptive efficiency.

Clark's account of language as a cognitive tool turns, in part, on the idea that language is an *external* resource which augments, rather than transforms, our unaided psychological capacities. Recall that a clear sense in which Clark rejects the claim that language is inner is the sense in which there exists some sort of domain-specific language processing system whose elements are organized, at basic level, so as to code for the structural properties of natural language. Now, the view for which I have argued does not herald the return of a Chomskyan in-the-head language device or even a Fodorian language of thought. For one thing there is no reason to think that the implicated inner structures will be innate (see footnote 2). It doesn't even amount to an endorsement of the "reprogramming" view that Clark attributes to Dennett (see above), since, according to the tabled suggestion, (a) on-line language-use heralds no radical change of inner representational format, whereas, on the view of Clark's Dennett, it seems that exposure to language does, and (b) off-line language-use demands a change of inner representational format but not of inner computational process, whereas, on the view of Clark's Dennett, it seems that exposure to language produces changes of both. However, it does open up a conceptual space within which there emerges a robust third sense in which language may be inner. (Recall that Clark can live happily with the position according to which

¹⁶ In this context it is interesting to note that in their treatment of mathematical problem-solving, Rumelhart et al. write of the "internalization of an external representational format" (1986, p. 47). The closest Clark himself comes to opening up this issue is when, as we saw previously, he talks about using "the same old (essentially pattern-completing) resources to model the *special kinds of behaviour observed in the world of public language*" (p. 198, emphasis added).

language is inner just when linguistically formed thoughts are expressed privately.) In the newly emerged sense in which language may be inner, language is inner if there exists, in the brain, representations which recapitulate the structural properties of natural language. Notice that this sense of language being inner doesn't require the presence of any domain-specific language processing system. The mechanisms which deal in the linguistically structured representations may themselves be domain-general in character (e.g. generic connectionist pattern-completers). What matters is the structure of the representations concerned.¹⁷

With this representation-oriented sense of "language as inner" distinguished from its mechanism-oriented cousin, Clark's arguments concerning the nature of language are revealed as constituting a case against (i) the idea that language-use requires a domain-specific language processing mechanism, and not (ii) the idea that (some) language-use requires internal representations with specifically linguistic structure. So they do not constitute an argument against our newly emerged sense in which language may be inner. Indeed, as we have seen, in the particular case of linguistic inner rehearsal, the very logic of the theoretical perspective which Clark endorses demands that off-line language-use is supported by representations which encode linguistic structure. So there is, after all, a robust sense in which language may be inner. And note that Clark cannot accept this newly emerged sense in which language may be inner in the same way that he can accept the sense which turns on the private expression of linguistically formed thoughts. For what makes the sense that turns on the presence of a domain-specific language processing mechanism troublesome is the fact that the relevant language-based structures are to be found in the agent's brain, and that is equally true of our newly emerged sense. So if the key claim that language is a cognitive tool is interpreted as requiring the externality of language (which seems correct, at least within Clark's framework), then that claim is under serious threat.

Let me finish by leaving Clark himself behind, and by mining a general consequence of my critique of his account of language. To bring this consequence into view, we need to note that, according to the position I've defended, linguistic inner rehearsal is just one species of inner rehearsal among others. Of course, we have already met a mathematically oriented kind of inner rehearsal. But there are others too, such as my mentally recalling a favourite walk along the Royal Mile up to Edinburgh Castle. Now notice that the general assumption here (an assumption which Clark shares; see his discussion of reasoning with Venn diagrams) is that we first perform a behaviour (language-use, mathematical reasoning, perceptually guided action, etc.) on-line. Later we may rehearse the same movements and manipulations off-line, in our heads. Now, in each of the different contexts of off-line thinking, inner mechanisms which, in the on-line case, will have been in regular causal commerce with certain external elements are re-targeted on inner surrogates of those external

¹⁷ In effect, this is to exploit a strong distinction between mechanism and information (or computational process and representational state) that Tony Atkinson and I have used elsewhere to shape a critique of the evolutionary-psychological view that the human mind is a collection of domain-specific modules (see [Wheeler and Atkinson, 2001](#); [Atkinson and Wheeler, 2004](#)).

elements, in order to support inner rehearsal. The content realized by the inner surrogates in play here will be determined by the nature of the domain in question.

So where do the crucial surrogates come from? Since, according to the distributed cognition perspective, they aren't causally necessary for on-line intelligent activity, they certainly won't have been built in order to enable such activity at the time of performance. One option here is to hold that such states are constructed in parallel with the action being performed, even though they play no direct causal role in enabling that performance. It is perhaps hard to see how such a system would ever be evolutionarily selected for. After all, if we assume, as we have been, that the on-line behaviour is in place ahead of the capacity for the inner rehearsal of that behaviour, why should Darwinian selection favour the diverting of resources to a model-building strategy when that strategy has no beneficial effects whatsoever on the behaviour itself? Perhaps, however, we are looking with too narrow a field of vision. After all, we know that skilled sportsmen and sportswomen may improve their athletic performance on the day of competition through a prior process in which key moves and techniques are visualized. This imaginative trick is surely based on inner rehearsal, and has adaptive pay-offs.

Whatever the prospects for such evolutionary explanations, however, the fact remains that the state of affairs suggested by the distributed cognition paradigm credits the agent with what seems to be a rather less efficient cognitive set-up than that attributed to her by a more orthodox cognitive science. After all, on a more orthodox story, not only off-line thinking but also real-time intelligent action is guided by, broadly speaking, the kinds of states and structures (structure preserving inner representations) that I have suggested are needed for off-line thinking. So the traditional theorist faces no mystery about how such states and structures get installed: on-line intelligent action demands their presence. Those very same elements are then simply re-used in the off-line case. So while we fans of distributed cognition often claim that the sorts of adaptive strategies we favour are more efficient than their traditional counterparts (since on our view there is no requirement for the agent to build or to maintain computationally costly detailed models of the world in order to generate on-line intelligent action; see e.g. Brooks, 1991), this really does look to be an unportable argument, in those cases where we are dealing with creatures who are also capable of off-line intelligence.

One closing thought: Perhaps this worry about the relative inefficiency of the kind of cognitive architecture that is authorized by the basic logic of the distributed cognition paradigm might be mitigated, if it could be shown that creatures who are capable of off-line cognition may not commit themselves to recapitulating all the relevant environmental and experiential structure present during real-time action. According to this proposal, which applies to linguistic inner rehearsal as much as to post-perceptual recall, at least some off-line thinking may, in truth, be the result of a sly alliance between sparse, just-good-enough memory (laid down as representations in parallel with real-time performance) and creative cognitive reconstruction. Like so many regions of the exciting terrain that is contemporary cognitive science (and the peak that corresponds to an explanation of our linguistic abilities is surely one of the hardest to climb in that terrain), this remains, I think, an open question

ripe for further philosophical and empirical exploration. The distributed cognition paradigm may not have all the answers, but it has, I am convinced, made the problems all the more interesting.

Acknowledgments

Many thanks to the following audiences for invaluable feedback and discussion: the Mind AND World 2003 conference in Durban, the Mental Phenomena VII conference in Dubrovnik, the E-Intentionality Research Seminar at the University of Sussex, and the Language Evolution and Computation Research Seminar at the University of Edinburgh. Also, many thanks to the following individuals for specific critical responses: Henry Brighton, Seth Bullock, Stephen Cowley, Ezequiel di Paolo, Inman Harvey, Simon Kirby, Richard Menary, David Papineau, Don Ross and John Sutton.

References

- Atkinson, A.P., Wheeler, M., 2004. The grain of domains: the evolutionary-psychological case against domain-general cognition. *Mind and Language* 19 (2), 147–176.
- Brighton, H., 2002. Compositional syntax from cultural transmission. *Artificial Life* 8 (1), 25–54.
- Brooks, R.A., 1991. Intelligence without reason. In: *Proceedings of the Twelfth Joint International Conference on Artificial Intelligence*. Morgan Kaufman, San Mateo, CA, pp. 569–595.
- Chomsky, N.A., 1986. *Knowledge of Language*. Praeger, New York.
- Clark, A., 1997. *Being There: Putting Brain, Body, and World Together Again*. MIT Press/Bradford Books, Cambridge, MA/London.
- Clark, A., 1998. Magic Words: How Language Augments Human Computation. In: Carruthers, P., Boucher, J. (Eds.), *Language and Thought: Interdisciplinary Themes*. Cambridge University Press, Cambridge, pp. 162–183.
- Clark, A., 2001. *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford University Press, Oxford.
- Clark, A., 2003. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press, Oxford.
- Clark, A., Chalmers, D., 1998. The Extended Mind. *Analysis* 58 (1), 7–19.
- Dennett, D.C., 1991. *Consciousness Explained*. Little, Brown & Co, Boston.
- Dreyfus, H.L., 1991. *Being-In-The-World: A Commentary on Heidegger's Being and Time, Division 1*. MIT Press, Cambridge, MA.
- Fodor, J.A., 1975. *The Language of Thought*. Thomas Cromwell, New York.
- Fodor, J.A., 1987. Why there still has to be a Language of Thought. In: Fodor, J.A. (Ed.), *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press, Cambridge, MA, pp. 135–167.
- Fodor, J.A., Pylyshyn, Z., 1988. Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition* 28, 3–71.
- Franceschini, N., Pichon, J.-M., Blanes, C., 1992. From Insect Vision to Robot Vision. *Philosophical Transactions of the Royal Society, Series B* 337, 283–294.
- Guignon, C.B., 1983. *Heidegger and the Problem of Knowledge*. Hackett, Indiana.
- Harvey, I., Husbands, P., Cliff, D., 1994. Seeing the light: artificial evolution, real vision. In: Cliff, D., Husbands, P., Meyer, J.-A., Wilson, S.W. (Eds.), *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*. MIT Press/ Bradford Books, Cambridge, MA, pp. 392–401.

- Haugeland, J., 1995/1998. Mind embodied and embedded. In: Haugeland, J. (Ed.), *Having Thought: Essays in the Metaphysics of Mind*. Harvard University Press, Cambridge, MA, pp. 207–237.
- Heidegger, M., 1926. *Being and Time*. Basil Blackwell, Oxford, Translation: J. Macquarrie and E. Robinson.
- Hurley, S., 1998. Vehicles, Contents, Conceptual Structure, and Externalism. *Analysis* 58 (1), 1–6.
- Husbands, P., Meyer, J.-A. (Eds.), 1998. *Evolutionary Robotics: Proceedings of the First European Workshop, EvoRobot98*. vol. 1468 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin.
- Kirby, S., 2002. Learning, Bottlenecks and the Evolution of Recursive Syntax. In: Briscoe, E. (Ed.), *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*. Cambridge University Press, Cambridge, pp. 173–204.
- Kirby, S., Christiansen, M.H., 2003. From language learning to language evolution. In: Christiansen, S., Kirby, S. (Eds.), *Language Evolution*. Oxford University Press, Oxford, pp. 272–294.
- Nolfi, S., Floreano, D., 2000. *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. MIT Press, Cambridge, MA.
- Rowlands, M., 1999. *The Body in Mind: Understanding Cognitive Processes*. Cambridge University Press, Cambridge.
- Rumelhart, D.E., Smolensky, P., McClelland, J.L., Hinton, G., 1986. Schemata and sequential thought processes in PDP models. In: McClelland, J.L., Rumelhart, D. (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 2: *Psychological and Biological Models*, MIT Press, Cambridge, MA, pp. 7–57.
- Smith, K., Kirby, S., Brighton, H., 2003. Iterated learning: a framework for the emergence of language. *Artificial Life* 9 (4), 371–386.
- Wheeler, M., 2001. Two threats to representation. *Synthese* 129, 211–231.
- Wheeler, M., 2004a. How to do things with (and without) representations. In: Menary, R. (Ed.), *The Extended Mind*. John Benjamins, Amsterdam.
- Wheeler, M., 2004b. *Reconstructing The Cognitive World: The Next Step*. MIT Press, Cambridge, MA.
- Wheeler, M., 2004c. Friends reunited? evolutionary robotics and representational explanation. *Artificial Life*.
- Wheeler, M., Clark, A., 1999. Genic representation: reconciling content and causal complexity. *British Journal for the Philosophy of Science* 50 (1), 103–135.
- Wheeler, M., Atkinson, A.P., 2001. Domains, brains and evolution. In: Walsh, D.M. (Ed.), *Naturalism, Evolution and Mind*. Cambridge University Press, Cambridge, pp. 239–266.
- Wittgenstein, L., 1922. *Tractatus Logico-Philosophicus*. Routledge, London, Translation: C.K. Ogden.