Supersizing the Mind

Embodiment, Action,
and Cognitive Extension

Andy Clark

OXFORD
UNIVERSITY PRESS

2008

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford   New York
Auckland   Cape Town   Dar es Salaam   Hong Kong   Karachi
Kuala Lumpur   Madrid   Melbourne   Mexico City   Nairobi
New Delhi   Shanghai   Taipei   Toronto

With offices in
Argentina   Austria   Brazil   Chile   Czech Republic   France   Greece
Guatemala   Hungary   Italy   Japan   Poland   Portugal   Singapore
South Korea   Switzerland   Thailand   Turkey   Ukraine   Vietnam

Foreword

David Chalmers

A month ago, I bought an iPhone. The iPhone has already taken over some of the central functions of my brain. It has replaced part of my memory, storing phone numbers and addresses that I once would have taxed my brain with. It harbors my desires: I call up a memo with the names of my favorite dishes when I need to order at a local restaurant. I use it to calculate, when I need to figure out bills and tips. It is a tremendous resource in an argument, with Google ever present to help settle disputes. I make plans with it, using its calendar to help determine what I can and can't do in the coming months. I even daydream on the iPhone, idly calling up words and images when my concentration slips.

Friends joke that I should get the iPhone implanted into my brain. But if Andy Clark is right, all this would do is speed up the processing and free up my hands. The iPhone is part of my mind already.

Clark is a connoisseur of the myriad ways in which the mind relies on the world to get its work done. The first part of this marvelous book explores some of these ways: the extension of our bodies, the extension of our senses, and, crucially, the use of language as a tool to extend our thought. The second part of the book defends the thesis that, in at least some of these cases, the world is not serving as a mere instrument for the mind. Rather, the relevant parts of the world have become parts of

my mind. My iPhone is not my tool, or at least it is not wholly my tool. Parts of it have become parts of me.

This is the thesis of the extended mind: when parts of the environment are coupled to the brain in the right way, they become parts of the mind. The thesis has a long history: I am told that there are hints of it in Dewey, Heidegger, and Wittgenstein. But no one has done as much to give life to the idea as Andy Clark. In a series of important books and articles—*Being There, Natural-Born Cyborgs*, "Magic Words: How Language Augments Human Computation," and many others—he has explored the many ways in which the boundaries between mind and world are far more flexible than one might have thought. This book is his major statement of the philosophical picture that undergirds the view.

Andy invited me to write this foreword because of my role in coauthoring an article, "The Extended Mind" (included as an appendix to this book), that has come to serve as a sort of flagship philosophical statement for this picture of the mind. This paper was written when Andy and I were colleagues at Washington University in 1995 and was published in *Analysis* in 1998. Now that a decade has passed, Andy suggested that I might offer a retrospective perspective on that article while saying something about my own views on the topic. I am more than happy to do this, though no one should feel any obligation to read what I have to say. Readers who are not already familiar with the issues might want to first look at the rest of the book, or at least at the appendix, before returning to this foreword.

It will come as no surprise to anyone who knows Andy's work to hear that the inspiration behind the original article was all his. In March 1995, Andy handed me a short article that he had written, called "Mind and World: Breaching the Plastic Frontier." This article already contained many of the key theses and arguments that appear in "The Extended Mind." It contained the key thought-experiment comparing Tetris players who rotate images inside and outside the head. It also contained the crucial Parity Principle, which held that if a process in the world works in a way that we should count as a cognitive process if it were done in the head, then we should count it as a cognitive process all the same. I had some ideas about how to further develop and argue for the thesis, and we ended up working together on an expanded and renamed paper.

The original article contained a notorious footnote, saying "The authors are listed in order of their degree of belief in the central thesis." Some have taken this to imply that I reject the extended mind thesis and was working as a sort of hired gun in service of someone else's cause. In fact, I find the thesis very attractive. If anything, I am more confi-

dent of the thesis than I was a decade ago, having seen it survive the many objections that have been raised to it without too much trouble. (I agree with almost everything in Andy's authoritative treatment of these objections in part II of this book.) But I do not think the matter is completely cut and dried.

I am not worried by the most common objections to the thesis: the fact that external cognitive processes work differently from internal cognitive processes, the threat that the mind will extend too far into the world, and the threat that the core role of the brain will be lost, for example. And I certainly do not think that there is anything privileged about skin and skull as boundaries for the mind.

Still, I think that there is one potentially principled place where the opponent of the extended mind can resist. This is an appeal to the dual boundaries of perception and action. It is natural to hold that perception is the interface where the world affects the mind, and that action is the interface where the mind affects the world. If so, it is tempting to hold that what precedes perception and what follows action is not truly mental. And one might use this to draw a principled distinction between the cases of Otto (the Alzheimer's patient who uses a notebook as memory) and Inga (the ordinary subject who uses her brain). To interact with his notebook Otto must read it and write in it, requiring perception and action, where there is no such requirement for Inga. If so, then the boundaries above would place the notebook outside the mind.

We consider this sort of worry briefly in "The Extended Mind," suggesting that Otto's access to the notebook need not be seen as perceptual. But this is surely too quick: there is no denying that Otto sees the notebook and reads from it, just as there is no denying that Otto reaches for the notebook and writes in it. So there is certainly perception and action taking place here. A better reply might be to note that there can be inner perception (when one reads from a mental image, say) and mental action (when one makes a mental note, say). Then Otto's perception and action could be seen as of a piece with these. But an opponent could reasonably respond that Otto's interaction with the notebook involves real perception and real action (where these might be glossed as sensory perception and physical action, or in terms of the right sort of perceptual and agentive experience, or in some other way), of a sort not present in cases of inner perception and mental action. And it could be suggested that real perception and real action mark a plausible boundary for the mental.

Perhaps the best reply for the proponent of the extended mind is to reject the proposed boundary. We touch on this reply in the "The Extended Mind," saying that just because the Terminator retrieves

ing Otto's local-scale interactions with his notebook, it is natural to deny that he has the extended belief and to hold that the relevant actions are explained by internal beliefs. As Clark suggests in his concluding chapter, we can flip back and forth between both ways of looking at things. We have a sort of Necker Cube effect, with mental states counting as extended or not depending on our perspective and our purposes.

As this Necker Cube flips, various things flip along with it. Otto's access to the notebook flips from an act of perception to an act of memory retrieval. His writing in the notebook flips from a physical action to a mental action. We might thus flip from regarding Otto's cognitive system as local to regarding it as extended, and we might even flip our perspective on Otto himself in a similar way. And importantly, Otto's state before opening the notebook flips from ignorance to knowledge.

This duality of perspectives can naturally be accommodated in various stories about the way mental terms like "believe" and "know" function. One such story holds that that ascriptions such as "Otto believes that the museum is on 53rd Street" are sensitive to contextual factors including explanatory purposes. In a context where one is explaining Otto's travels, the ascription is true. In a context where one is explaining his interactions with his notebook, the ascription is false. One can find a similar dependence on explanatory purposes in nonextended cases. Say that someone has to think for a moment before responding that 7 times 8 is 56. If we ask why they hesitated, we might reasonably say that they did not know the answer and had to think about it. But if we are asked why they got all the questions right, we might say that it is because they knew all the values in the multiplication table.

Other semantic stories are possible. One could give an account where the extended belief and knowledge ascriptions are always true, and when we say that Otto reached for his notebook because he did not know the address, we say something that is useful but strictly speaking false. One could also give an account where the extended belief and knowledge ascriptions are always false, so that our explanation of Otto's actions in terms of extended beliefs are at best metaphorically true. Or one could suggest that the reference of terms like "belief" is indeterminate between the two notions. If so, then the original extended belief ascriptions may be neither true nor false, but there will be precisified versions that fall on each side.

Ultimately, however, I think that issues about what really counts as a belief and about how the term "belief" functions are terminological questions that, while interesting, can mask the deeper point. If someone insists that they use the term "believe" in such a way that it picks out states realized in the space between perception and action, then one can

allow them to use the term this way if they like. The deeper point is that extended states can function in explanation in very much the same way that beliefs function, and they should be regarded as sharing a deep and important explanatory kind with beliefs. This explanatory unification is the real underlying point of the extended mind thesis.

In "The Extended Mind," the only extended elements of the mind that we argued for were beliefs and cognitive processes: in particular standing beliefs (like Inga's belief about the museum's location), and cognitive processes such as mental rotation. It is natural to ask whether the extended mind thesis might itself be extended. What about extended desires, extended reasoning, extended perception, extended imagination, and extended emotions? I think that there is something to be said for each of these. Perhaps the camera on my iPhone can serve as an extended perceptual mechanism. And perhaps one might have something akin to an extended mood, if not an extended emotion, when one's environment is always nudging one toward happiness or sadness. Clark discusses many such cases throughout the book, including extended perceptual mechanisms in chapter 2 and extended mechanisms of attention in chapter 3.

But then, what about the big question: extended consciousness? The dispositional beliefs, cognitive processes, perceptual mechanisms, and moods considered above all extend beyond the borders of consciousness, and it is plausible that it is precisely the nonconscious part of them that is extended. I think there is no principled reason why the physical basis of consciousness could not be extended in a similar way. It is probably so extended in some possible worlds: one could imagine that some of the neural correlates of consciousness are replaced by a module on one's belt, for example. There may even be worlds where what is perceived in the environment is itself a direct element of consciousness: my paper "Perception and the Fall from Eden" tells a fable about one such world.

Still, I think it is unlikely that any everyday process akin to Otto's interaction with his notebook will yield extended consciousness, at least in our world. Certainly, relatives of the Otto/Inga argument do not seem to extend to consciousness. The original argument crucially yields a twin case, involving Otto and Twin Otto, who are physical duplicates with different beliefs. An argument for extended consciousness would require twins with different states of consciousness: Olga and Twin Olga are internal duplicates, but what it is like to be Olga differs from what it is like to be Twin Olga. But no matter how hard one tries to construct an Otto-style story that works like this, the story does not seem to succeed. Perhaps part of the reason is that the physical basis of consciousness

requires direct access to information on an extremely high bandwidth. Perhaps some future extended system, with high-bandwidth sensitivity to environmental information, might be able to do the job. But our low-bandwidth conscious connection to the environment seems to have the wrong form as it stands.

In recent years, a few philosophers have argued that the basis of conscious states lies partly outside the head. Some of these arguments, such as those of Dretske (1996), Fisher (2007), and Martin (2004), turn on considerations quite different from the sort of two-way coupling between organism and environment that is at the heart of the extended mind thesis. The resulting views are interesting and challenging, but they are largely independent of the active externalism of the extended mind thesis. Others, including Hurley (1998) and Noë (2006), have argued that the two-way coupling extends to consciousness. But these arguments do not seem to yield a twin case of the sort discussed above, so they do not rule out the supervenience of consciousness on the internal. At best, as Clark suggests in chapter 8, they yield a weaker sort of dependence of consciousness on the environment. I tentatively conclude that the extension of the mind is compatible with retaining an internal conscious core.

What general picture of the mind does the extended mind thesis rest on? It has sometimes been suggested that the thesis requires functionalism about the mental, where all mental states are defined by the causal roles that they play. This cannot be quite right: I think that functionalism about consciousness is implausible, for example, but this implausibility does not affect the arguments for the extended mind thesis. One might support the view by invoking an attenuated functionalism: say, one where certain mental states (such as dispositional beliefs) are defined by their causal relations to conscious states, to behavior, and to other elements of the cognitive network. I find such a picture attractive myself, but strictly speaking even this picture is not required for all the argument to go through. All one needs is the very weak functionalism captured in the Parity Principle: roughly, if a state plays the same causal role in the cognitive network as a mental state, then there is a presumption of mentality, one that can only be defeated by displaying a relevant difference between the two (and not merely the brute difference between inner and outer). Combined with the observation that there are no relevant differences in the relevant cases—an observation that does not require functionalism for its support—the thesis follows.

Likewise, the extended mind thesis is compatible with both physicalism and dualism about the mental. It is compatible with connectionist and classical views, with computational and noncomputational

approaches, and even with internalism and externalism in the traditional debates over mental content (as we suggest in "The Extended Mind"). So I do not think that the extended mind thesis requires much in the way of theoretical presupposition at all. Instead, it is an independently attractive view of the mental.

Ultimately, the proof is in the pudding. The deepest support for the view comes from the explanatory insights that the extended mind perspective yields. And those insights are just what this book provides. In case after case, in domain after domain, Andy Clark brings out the many ways in which the extended view of the mind can productively reconfigure our thinking about the relationship between mind and world. After absorbing this picture, nothing will ever look quite the same way again. And if Clark is right, then the absorption has already started. Just opening his book may have turned you into a smarter, deeper, and more insightful person.

Canberra, November 2007

3

Material Symbols

3.1 *Language as Scaffolding*

Where does language fit into our emerging picture of the plastic, environmentally exploitative, ecologically efficient agent? One useful way to approach this question is to consider language itself as a form of mind-transforming cognitive scaffolding: a persisting, though never stationary, symbolic edifice whose critical role in promoting thought and reason remains surprisingly ill understood.

In this chapter, I examine three distinct but interlocking benefits of the linguistic scaffold. First, the simple act of labeling the world opens up a variety of new computational opportunities and supports the discovery of increasingly abstract patterns in nature. Second, encountering or recalling structured sentences supports the development of otherwise unattainable kinds of expertise. And third, linguistic structures contribute to some of the most important yet conceptually complex of all human capacities: our ability to reflect on our own thoughts and characters and our limited but genuine capacity to control and guide the shape and contents of our own thinking.

3.2   Augmenting Reality

Consider the case of Sheba and the treats as recounted in Boysen et al. (1996). Sheba (an adult female chimpanzee) has had symbol and numeral training: She knows about numerals. Sheba sits with Sarah (another chimp), and two plates of treats are shown. What Sheba points to, Sarah gets. Sheba always points to the greater pile, thus getting less. She visibly hates this result but can't seem to improve. However, when the treats arrive in containers with a cover bearing numerals on top, the spell is broken, and Sheba points to the smaller number, thus gaining *more* treats.

What seems to be going on here, according to Boysen, is that the material symbols, by being simple and stripped of most treat-signifying physical cues, allow the chimps to sidestep the capture of their own behavior by ecologically specific, fast-and-frugal subroutines. The material symbol here acts as a manipulable and, in some sense, a merely "shallowly interpreted" (Clowes 2007) stand-in, able to loosen the bonds between perception and action. Importantly, the presence of the material symbol impacts behavior not in virtue of being the key to a rich inner mental representation (though it may be this also) but rather by itself, qua material symbol, providing a new target for selective attention and a new fulcrum for the control of action. Such effects, as Clowes (2007) argues, do of course depend on the presence of something akin to a system of interpretation. But it is their ability to provide simple, affect-reduced, perceptual targets that (I want to suggest) explains much of their cognitive potency.

In much the same way, the act of labeling creates a new realm of perceptible objects upon which to target basic capacities of statistical and associative learning. The act of labeling thus alters the computational burdens imposed by certain kinds of problems. I have written quite a bit on this elsewhere, so I'll keep this brief. My favorite example (Clark 1998b) begins with the use, by otherwise language-naive chimpanzees, of concrete tags (simple and distinct plastic shapes) for relations such as sameness and difference. Thus, a pair such as cup–cup might be associated with a red triangle (sameness) and cup–shoe with a blue circle (difference). This is not in itself surprising. What is more interesting is that after this training, the tag-trained chimps (and only tag-trained chimps) prove able to learn about the abstract properties of higher order sameness; that is, they are able to learn to judge of two presented pairs (e.g., cup–cup and cup–shoe) that the relationship between the relations is one of higher order difference (or better, lack of higher order sameness)

because the first pair exhibits the sameness relation and the second pair shows the difference relation (Thompson, Oden, and Boysen 1997). The reason the tag-trained chimps can perform this surprising feat is, so the authors suggest, because by mentally recalling the tags, the chimps can reduce the higher order problem to a lower order one: All they have to do is spot that the relation of difference describes the pairing of the two recalled tags (red triangle and blue circle).

This is a nice concrete example of what may well be a very general effect (see Clark 1998b and Dennett 1993). Once fluent in the use of tags, complex properties and relations in the perceptual array are, in effect, artificially reconstituted as simple inspectable wholes. The effect is to reduce the descriptive complexity of the scene. Kirsh (1995b), as we shall see in more detail in chapter 4, describes the intelligent use of space in just these terms. When, for example, you group your groceries in one bag and mine in another, or when the cook places washed vegetables in one location and unwashed ones in another, the effect is to use spatial organization to simplify problem solving by using spatial proximity to reduce descriptive complexity. It is intuitive that once descriptive complexity is thus reduced, processes of selective attention, and of action control, can operate on elements of a scene that were previously too "unmarked" to define such operations over. Experience with tags and labels may be a cheap way of achieving a similar result. Spatial organization reduces descriptive complexity by means of physical groupings that channel perception and action toward functional or appearance-based equivalence classes. Labels allow us to focus attention on all and only the items belonging to equivalence classes (the red shoes, the green apples, etc.). In this way, both linguistic and physical groupings allow selective attention to dwell on all and only the items belonging to the class. And the two resources are seen to work in close cooperation. Spatial groupings are used in teaching children the meanings of words, and mentally rehearsed words may be used to control activities of spatial grouping.

Simple labeling thus functions as a kind of augmented reality trick by means of which we cheaply and open-endedly project new groupings and structures onto a perceived scene.[1] Labeling is cheap because it avoids the physical effort of putting things into piles. And it is open-ended insofar as it can group in ways that defeat simple spatial display—for example, by allowing us to selectively attend to the four corners of a tabletop, an exercise that clearly cannot be performed by physical reorganization! Linguistic labels, on this view, are tools for grouping and in this sense act much like real spatial reorganization. But in addition (and unlike mere physical groupings), they effectively and

open-endedly add new "virtual" items (the recalled labels themselves) to the scene. In this way, experience with tags and labels warps and reconfigures the problem spaces for the cognitive engine.

A related effect may also be observed in recent work on language learning. Thus, in a recent review, Smith and Gasser (2005) ask a very nice question. Why, given that human beings are such experts at grounded, concrete, sensorimotor driven forms of learning, do the symbol systems of public language take the special and rather rarified forms that they do?

> One might expect that a multimodal, grounded, sensorimotor sort of learning would favor a more iconic, pantomime-like language in which symbols were similar to referents. But language is decidedly not like this...there is no intrinsic similarity between the sounds of most words and their referents: the form of the word *dog* gives us no hints about the kind of thing to which it refers. And nothing in the similarity of the forms of *dig* and *dog* conveys a similarity in meaning. (Smith and Gasser 2005, 22)

The question, in short, is: "Why in a so profoundly multimodal sensorimotor agent such as ourselves is language an arbitrary symbol system?" (24).

One possible answer, of course, is that language is like that because (biologically basic) *thought* is like that, and the forms and structures of language reflect this fact. But another answer says just the opposite. Language is like that, it might be suggested, because thought (or rather, biologically basic thought) is *not* like that. The computational value of a public system of essentially context-free, arbitrary symbols, lies, according to this opposing view, in the way such a system can push, pull, tweak, cajole, and eventually cooperate with various nonarbitrary, modality-rich, context-sensitive forms of biologically basic encoding.[2]

3.3   Sculpting Attention

The role of structured language as a tool for scaffolding action has been explored in a variety of literatures, ranging from Vygotskyian developmental psychology to cognitive anthropology (see, e.g., Berk 1994; Hutchins 1995; Donald 2001). Mundane examples of such scaffolding abound and range from memorized instructions for tying one's shoelaces to mentally rehearsed mantras for crossing the road, such as "look right, look left, look right again, and if all is clear, cross with caution"

(that's for UK-style left-hand-drive roads; don't try that in the United States, folks!). In such cases, the language-using agent is able (once the instructions are memorized or, in the written case, visually accessed) to engage in a simple kind of behavioral self-scaffolding, using the phonetic or spatial sequence of symbolic encodings to stand proxy for the temporal sequence of acts. Frequent practice then enables the agent to develop genuine expertise and to dispense with the rehearsal of the helpful mantra.

More interesting than all this, however, is the role of linguistic rehearsal in expert performance itself. In previous work (Clark 1996), I discussed some ways in which linguaform rehearsal enables experts to temporarily alter their own focus of attention, thus fine-tuning the patterns of inputs that are to be processed by fast, fluent, highly trained subpersonal resources. Experts, I argued, are doubly expert. They are expert at the task in hand but also expert at using well-chosen linguistic prompts and reminders to maintain performance in the face of adversity. Sometimes, inner rehearsal here plays a distinctly affective role, as the expert encourages herself to perform at her peak.[3] But in addition to the important cognitive-affective role of inner dialogue, there may also be cases in which verbal rehearsal supports a kind of perceptual restructuring via the controlled disposition of attention (for a nice example, see the discussion of linguistic rehearsal by expert Tetris players in Kirsh and Maglio 1992). The key idea, once again, is that the linguistic tools enable us to deliberately and systematically sculpt and modify our own processes of selective attention. In this regard, Sutton (2007) describes in some detail the value of "instructional nudges" (small strings of words, simple maxims). Such nudges, Sutton argues, are often best employed not by the novice but by the expert, who can use them to tune and modulate highly learned forms of embodied performance.[4]

Direct cognitive benefits from linguaform encodings are also suggested in work by Hermer-Vazquez, Spelke, and Katsnelson (1999). In this study, prelinguistic infants were shown the location of a toy or food in a room and then were spun around or otherwise disoriented and required to try to find the desired item. The location was uniquely determinable only by remembering conjoined cues concerning the color of the wall and its geometry (e.g., the toy might be hidden in the corner between the long wall and the short blue wall). The rooms were designed so that the geometric or color cues were individually insufficient and would yield an unambiguous result only when combined. Prelinguistic infants, though perfectly able to detect and use both kinds of cue, were shown to exploit only the geometric information, searching randomly in each of the two geometrically indistinguishable sites. Yet

adults and older children were easily capable of combining the geometric and nongeometric cues to solve the problem. Importantly, success at combining the cues was not predicted by any measure of the children's intelligence or developmental stage except for the child's use of language. Only children who were able to spontaneously conjoin spatial and color terms in their speech (e.g., who would describe something as, say, to the right of the long green wall) were able to solve the problem. Hermer-Vazquez, Spelke, and Katsnelson (1999) then probed the role of language in this task by asking subjects to solve problems requiring the integration of geometric and nongeometric information while performing one of two other tasks. The first task involved shadowing (repeating back) speech played over headphones. The other involved shadowing (with their hands) a rhythm played over the headphones. The working memory demands of the latter task were at least as heavy as those of the former. Yet subjects engaged in speech shadowing were unable to solve the integration-demanding problem, while those shadowing rhythm were unaffected. Agents' linguistic abilities, the researchers concluded, are indeed actively involved in their ability to solve problems requiring the integration of geometric and nongeometric information.

The precise nature of this linguistic involvement is, however, still in dispute. Hermer-Vazquez, Spelke, and Katsnelson (1999), and following them Carruthers (2002), interpret the results as suggesting that public language provides (or perhaps better, engenders) a unique internal representational medium for the cross-modular integration of information.[5] The linguaform templates of encoded sentences provide, according to Carruthers, special representational vehicles that allow information from otherwise encapsulated resources to interact. This is an attractive and challenging story, and one that I cannot pretend to do justice to here. But it is one that presupposes a specific (and quite contentious; see Fodor 2001) view of the mind as massively (not merely peripherally) modular, requiring linguaform templates to bring multiple knowledge bases into fruitful contact.

Suppose we abandon this presupposition of massive modularity? We may still account, or so I suggest, for the role of language in enabling complex multicued problem solving by depicting the linguistic structures as providing essential scaffolding for the distribution of selective attention to complex (in this case, color–geometry conjunctive) aspects of the scene. According to this alternative account, linguistic resources enable us better to control the disposition of selective attention to ever more complex feature combinations.[6] Attention to a complex conjoined cue, I suggest, requires the (possibly unconscious) retrieval of at least some of the relevant lexical items. This explains the shadowing result.

And it fits nicely with the earlier account of the cognitive impact of simple labels insofar as linguistic activity (in this case, more structured activity) again allows us to target our attentional resources on complex, conjunctive, or otherwise elusive elements of the encountered scene. The idea that language enables new forms of selective attention by, in effect, providing new objects for old (i.e., not specifically linguistic) attentive processes can be further illustrated by the case of arithmetical thought and reason, to which we now turn.

3.4   Hybrid Thoughts?

What is going on when you have the thought that 98 is one more than 97? According to a familiar model, you must have succeeded (if you managed to think the thought at all) in translating the English sentence into something else. The something else might be a sentence of mentalese (e.g., Fodor 1987) or a point in some exotic state space (e.g., Churchland 1989).

But consider a recent account due to Stanislas Dehaene and colleagues (see Dehaene 1997; Dehaene et al. 1999). Dehaene depicts this kind of precise mathematical thought as emerging at the productive intersection of three distinct cognitive contributions. The first involves a basic biological capacity to individuate small quantities: 1-ness, 2-ness, 3-ness, and more-than-that-ness, to take the standard set. The second involves another biologically basic capacity, this time for approximate reasoning concerning magnitudes (discriminating, say, arrays of 8 dots from arrays of 16 but not from more closely matched arrays). The third, not biologically basic but arguably transformative, is the learned capacity to use the specific number words of a language and the eventual appreciation that each such number word names a distinct quantity. Notice that this is not the same as appreciating, in at least one important sense, just what that quantity is. Most of us can't form any clear image of, for example, 98-ness (unlike, say, 2-ness). But we appreciate nonetheless that the number word 98 names a unique quantity between 97 and 99.

When we add the use of number words to the more basic biological nexus, Dehaene argues, we acquire an evolutionarily novel capacity to think about an unlimited set of exact quantities.[7] We gain this capacity not because we now have a mental encoding of 98-ness just like our encoding of 2-ness. Rather, the new thoughts depend directly, but not exhaustively, on our tokening the numerical expressions themselves as symbol strings of our own public language. The actual numerical

thought, on this model, occurs courtesy of the *combination* of this tokening (of the symbol string of a given language) and the appropriate activation of the more biologically basic resources mentioned earlier.

Here is some of the evidence for this view, as presented in Dehaene et al. (1999). First, there are the results of studies of Russian-English bilinguals. In these studies, Russian-English bilinguals were trained quite extensively on 12 cases involving exact and approximate sums of the same pairs of two-digit numbers presented as words in one or other language. For example, in English, subjects might be trained on the problem "Four + Five" and asked to select their answer from "Nine" and "Seven." This is called the exact condition, as it requires exact reasoning because the two candidate numbers are close to each other. By contrast, a problem like "Four + Five" with the choices "Eight" and "Three" belongs to the approximate condition, as it requires only rough reasoning because the candidates are now quite far apart.

After extensive training on the pairs, subjects were later tested on the very same sums in either the original or the other (nontrained) language. After training, performance in the approximation condition was shown to be unaffected by switching the language, whereas in the exact condition, language switching resulted in asymmetric performance, with subjects responding much faster if the test language corresponded to the training language. Crucially, then, there were no switching costs at all for trained approximate sums. Performance was the same regardless of language switching. Training-based speed-up is thus nonlanguage switchable for the exact sums and fully switchable for the inexact sums. Such studies, Dehaene et al. (1999) concluded, provide

> evidence that the arithmetic knowledge acquired during training with exact problems was stored in a language-specific format.... For approximate addition, in contrast, performance was equivalent in the two languages providing evidence that the knowledge was stored in a language-independent form. (973)

A second line of evidence draws on lesion studies in which (to take one example) a patient with severe left-hemisphere damage cannot determine whether 2 + 2 equals 3 or 4 but reliably chooses 3 or 4 over 9, indicating a sparing of the approximation system.

Finally, Dehaene et al. (1999) present neuroimaging data from subjects engaged in exact and approximate numerical tasks. The exact tasks show significant activity in the speech-related areas of the left frontal lobe, while the approximate tasks recruit bilateral areas of the parietal lobes implicated in visuospatial reasoning. These results are

presented as a demonstration "that exact calculation is language dependent, whereas approximation relies on nonverbal visuo-spatial cerebral networks" (970) and that "even within the small domain of elementary arithmetic, multiple mental representations are used for different tasks" (973).

Dehaene (1997) also makes some nice points about the need to somehow establish links between the linguistic labels and our innate sense of simple quantities. At first, it seems, children learn language-based numerical facts *without* such appreciation. According to Dehaene (1997), "for a whole year, children realize that the word 'three' is a number without knowing the precise value it refers to" (107). But once the label gets attached to the simple innate number line, the door is open to understanding that *all* numbers refer to precise quantities, even when we lack the intuitive sense of what the quantity is (e.g., my own intuitive sense of 53-ness is not distinct from my intuitive sense of 52-ness, though all such results are variable according to the level of mathematical expertise of the subject).

Typical human mathematical competence, all this suggests, is plausibly seen as a kind of hybrid, whose elements include:

1. images or encodings of actual words in a specific language,
2. an appreciation of the fact that each distinct number word names a specific and distinct quantity, and
3. a rough appreciation of where that quantity lies on a kind of approximate, analog number line (e.g., 98 is just less than halfway between 1 and 200).

Many of our mathematical thoughts rely, if this is correct, on the coordinated action of various resources. On this view, there is (at least) an internal representation of the numeral, of the word form, and of the phonetics, along with other resources (e.g., the analog number line) to which these become (with learning) roughly keyed via some sense of relative location. What matters for present purposes is that there may be no need to posit (for the average agent[8]), in addition to this coordinated medley, any further content-matching internal representation of, say, 98-ness. Instead, the presence of actual number words in a public code (and of internal representations *of those very public items*) is itself part of the coordinated representational medley that constitutes many kinds of arithmetical knowing.

Consider the thought that there are 98 toys on the table. According to the standard models, to think the thought that there are 98 toys on the table you must have succeeded in translating the English sentence into *a fully content-providing something else*. The something else might be an atom or sentence of mentalese (for Fodor) or a point in some

exotic state space (for Churchland). By contrast, according to this quite radical alternative, the thought that there are 98 toys on the table is (for most of us) dependent on the presence of a hybrid representational vehicle. This is a vehicle that includes, as expected, the activation of a variety of content-relevant internal representations (in neuralese or mentalese, let's assume). But it also includes as a co-opted proper part, a token (let's think of it as an image, very broadly construed) of a conventional public language encoding ("ninety-eight") appropriately linked to various other resources (e.g., some rough position on an analog number line).[9]

3.5   From Translation to Coordination

What general model of language and its relation to thought do these various illustrations suggest? A good place to begin is with the conception of language as complementary to more basic forms of neural processing (for my own explorations of the theme, see Clark 1996, 1998b, 2000b, 2000c, 2006). According to this conception, language works its magic not (or not solely) by means of translation into appropriate expressions of neuralese or the language of thought but also by something more like coordination dynamics. Encounters with words and with structured linguistic encodings act to anchor and discipline intrinsically fluid and context-sensitive modes of thought and reason.

This notion of anchoring is best appreciated in the light of connectionist or artificial neural network models of memory, storage, and processing (for basic overviews, see Clark 1989, 1993; for something closer to the state of the art, see O'Reilly and Munakata 2000). For present purposes, what matters is that such models posit a fundamentally fluid system in which the fine details of recent context color and nuance recall and representation in quite fundamental ways. For systems such as these, the problem of *stabilization* becomes pressing. On the one hand, it is a virtue of these systems that new information automatically impacts similar items that are already "stored" and that information retrieval is highly context sensitive. On the other hand, advanced thought and reason plausibly require the ability to reliably follow trajectories in representational space and to reliably lead others through certain trajectories. All this requires some means to discipline our own, and others', mental spaces in ways that tame (though never eradicate) those biologically more "natural" processes of merging and change. Words and linguistic strings are among the most powerful and basic tools we use to discipline and stabilize dynamic processes of reason and recall. The shift is thus from seeing words and sentences as items apt only

for translation into an inner code to seeing them as inputs (whether externally or internally generated) that drive, sculpt, and discipline the internal representational regime.

Elman (2004) suggests:

> Rather than putting word knowledge into a passive storage (which then entails mechanisms by which that knowledge can be "accessed," "retrieved," "integrated" etc.) words might be thought of in the same way that one thinks of other kinds of sensory stimuli: they act directly on mental states. (301)

"Words," Elman goes on to argue, "do not have meaning, they are cues to meaning" (306).[10] Linguistic inputs, on this model, are quite literally modes of systematic neural manipulation and operate in similar ways both between and within human individuals. Words and sentences act as artificial input signals, often (as in self-directed speech) entirely self-generated, that nudge fluid natural systems of encoding and representation along reliable and useful trajectories. This remarkable display of virtuosic artificial self-manipulation allows language-laden minds to sculpt and guide their own processes of learning, of recall, of representation, and of selective attention (for more on this important theme, see Barsalou 2003). In this way, the symbolic environment (very broadly construed) can impact thought and learning *both* by selectively activating other internal representational resources (the usual suspects) *and* by allowing the material symbols themselves, or shallow imagelike internal representations of them, to act as additional fulcrums of attention, memory, and control. In the maximum strength version, these shallow symbolic objects can even appear as elements in representationally hybrid thoughts.

For quite a few years, I thought this was a radical idea that fans of (to take the most extreme example) the Language of Thought Hypothesis would surely reject out of hand. Their idea, I believed, was that words mean what they do entirely by virtue of being paired with expressively parallel snippets of mentalese and that thinking was all done in mentalese. Imagine my surprise, then, when I found this little snippet hidden away in that review of Carruthers by Fodor (1998):

> I don't think that there are decisive arguments for the theory that all thought is in Mentalese. In fact, I don't think it's even true, in any detail . . . I wouldn't be in the least surprised, for example, if it turned out that some arithmetic thinking is carried out by executing previously memorized algorithms that are defined *over public language symbols for numbers* ("now carry

the '2'" and so forth). It's quite likely that Mentalese co-opts bits of natural language in all sorts of ways; quite likely the story about how it does so will be very complicated indeed by the time that the psychologists get finished telling it. (72, emphasis in original)

Fodor here gestures, it seems to me, at an incredibly potent mechanism of cognitive expansion. Pretty clearly, though, Fodor (1998) himself attaches little importance to the concession, quickly adding that: "For all our philosophical purposes (e.g., for purposes of understanding what thought content is, and what concept possession is, and so forth) *nothing essential is lost* if you assume that all thought is in Mentalese" (72, emphasis added).

By contrast, I am inclined to see the potential for representational hybridity as massively important to understanding the nature and power of much distinctively human cognition. There are, I think, two issues that make the difference between Fodor's assessment of the situation and my own.

First, Fodor has the LOT (Language of Thought) already in place, so the basic biological engine, on his account, comes factory primed with innovations favoring structure, generality, and compositionality.

But what if your vision of the basic biological engine is not one that echoes the properties and features of sentences and propositional attitudes? What if, for example, it is closer to Churchland's vision of a complex but thoroughly connectionist device or to Barsalou's (1999) vision of a "perceptual symbol system"? What if, in short, you don't have what Dennett once called the "walking encyclopedia" view of the basic innards? In such a case, the potential cognitive impact of a little hybridity and co-opting may be much greater than Fodor concedes. It may be essential to such a system's ability to think rather a wide variety of thoughts that the inner goings-on involve, as genuinely constitutive elements, something like images or traces of the public language symbols (words) themselves. Words and sentences, on this view, may be potent real-world structures (material symbols), many of whose features and properties (arbitrary amodal nature, extreme compactness and abstraction, compositional structure, etc.) simply complement, without full replication, the contributions of basic biological cognition. In such a case, it is not clear to me that it would be right to treat the co-opting strategies as marginal for the understanding of thought and concepts.

Second, much of Fodor's insistence upon a deflationary reading of the hybrid option flows directly from his (in)famous views concerning concept learning. For given those views, the meaning of hybrid

representational forms cannot be learned unless the learner already had the resources to represent that very meaning using more biologically basic (indeed, innate) resources.[11] This is not the time or place to engage in this important discussion. But it may be noted that Fodor's skepticism depends (as he himself is the first to admit; see, e.g., Fodor 2004) on detaching our ideas about meaning from any essential (i.e., meaning-constituting) links to action or use. The view I embrace (see Prinz and Clark 2004) is quite different and makes grasp of meaning a function of the range of ways an encoding (ultimately, a mere syntactic structure) poises us to act (where acts include but are not exhausted by acts of thinking and inferring) in the world. On such a view (which is the unremarked norm in much cognitive science) there is, even on Fodor's own account, no need to suppose true (radical, expansive) concept learning impossible. Pretty clearly, such a view leaves room for hybrid representational forms to poise a system to act in new ways and thus to count as understanding brand new kinds of things. The case of mathematical understanding, as rehearsed earlier, looks to be one example of this.

This vision of mind expansion by the use of hybrid representational forms remains visibly close to that of Dennett (1991a, 1996). But Dennett depicts exposure to language as installing a new virtual serial machine in the neural wetware by affecting "myriad microsettings in the plasticity of the brain" (1991a, 219). He thus places most of his bets on the *radically internally transformative* power of our encounters with language and ends up with a story that seems more developmental than genuinely hybrid. On the artifact model, by contrast, words and sentences remain potent real-world structures encountered and used by a basically (though this is obviously too crude) pattern-completing brain. Admittedly, drawing these lines is a delicate task (see, e.g., Densmore and Dennett 1999). For even on this account, the brain sometimes represents these structures so myriad microsettings *must* alter. But perhaps the brain represents these potent real-world items in much the same way it represents anything else. In that case (see Churchland 1995, chap. 10), language need not reorganize neural coding routines in any way that is deeper or more profound than might occur, say, when we first learn to swim or play volleyball.

Wheeler (2004, in press-a) argues that there is an important disanalogy between the volleyball case and the language case. For in learning to represent the structures of volleyball, we do not thereby learn to represent a syntactically structured domain. To represent linguistic structures just is, Wheeler claims, to install brand new modes of representation and processing: It is to install at least a kind of virtual syntactic

engine. I don't think this can be quiet right because, as I argue in Clark (2004), it must be possible to represent syntactically structured language without using syntactically structured representations to do so (just as it is possible to represent green objects without using green representations to do so). But Wheeler's real point (see Wheeler in press-a) is that language presents a very special kind of domain and that experience with language may thus have much more profound effects (from a cognitive point of view) than experience with other domains. I agree with that. It does not follow, though, that experience with language need install radically new forms of processing and encoding and, still less (and here Wheeler might agree), that those forms of encoding and processing would amount to the implementation of a language of thought. Wheeler and I thus agree that minds like ours are transformed by the web of material symbols and epistemic artifacts. But that transformation, at least on the version I favor, may neither require nor result in the installation of brand new internal representational forms. Instead, there may be much underexplored merit in the canny use of the external forms (and internal images of those very forms) themselves. Such forms may help sculpt and modify processes of selective attention and act as elements within hybrid representational wholes.

One immediate merit of such a view is a more nuanced attitude to the vexing question of evolutionary cognitive continuity. Jesse Prinz (2004) makes the point well:

> Researchers who presume that we think in amodal symbols face a dilemma. If they argue that nonhuman animals lack such amodal symbols, they must postulate a radical leap in evolution. If they suppose that animals have amodal thoughts, they must explain why human thought is so much more powerful. Empiricism [Prinz's favorite, though not obligatory in the present context!] when coupled with the assumption that we can think in public language, explains the discrepancy in cognitive capacities without postulating a major discontinuity in evolution. (427)

Needless to say, this radical story leaves many questions unanswered. It would be good to have a clear account of just what attention—that crucial variable that linguistic scaffolding seems so potently to adjust—actually *is*. It would be good to have much more in the way of genuine, implementable, fully mechanistic models of the various ways that internalized language might enhance thought. And it would be good to know just what it is about human brains and/or human history that has enabled structured public language to get such

a comprehensive grip on minds like ours. But shortfalls aside, I hope to have at least brought the strong material symbol model into clearer view and to have shown why it might be attractive to anyone who thinks that language makes a truly *deep* contribution to human thought and reason.

3.6   Second-order Cognitive Dynamics

The augmentation of biological brains with linguaform resources may also shed some light on our ability to display second-order cognitive dynamics (Clark 1998a; and see further discussion in Bermudez 2003). By second-order cognitive dynamics, I mean a cluster of powerful capacities involving reflection on our own thoughts and thought processes. It has recently been suggested, for example, that our capacities for flexible reasoning about others' beliefs depend directly on the linguistic externalization of beliefs using the grammatical structures of embedded complements (de Villiers and de Villiers 2003). Consider also the cluster of powerful capacities that include systematic attempts to train our skills and repair our faults, as well as practices that engender critical self-reflection. Further cases include recognizing a flaw in our own plan or argument and dedicating further cognitive efforts to fixing it, or reflecting on the unreliability of our own initial judgments in certain types of situations and proceeding with special caution as a result. Moving to a kind of meta-meta level, consider the act of thinking about the conditions under which we think best and trying to bring them about. The list could be continued, but the pattern should be clear. In all these cases, we are effectively thinking about our own cognitive profiles or about specific thoughts.

Rather amazingly, we are animals who can think about any aspect of our own thinking and can thus devise cognitive strategies (which may be more or less indirect and baroque) aimed to modify, alter, or control aspects of our own psychology.

All this "thinking about thinking" is a good candidate for a distinctively human capacity and one that may depend on language for its very existence. For (to rehearse a line pursued at length in Clark 1998a) as soon as we formulate a thought in words or on paper, it becomes an object for both ourselves and for others. As an object, it is the kind of thing we can have thoughts about. In creating the object, we need have no thoughts about thoughts, but once it is there, the opportunity immediately exists to attend to it as an object in its own right. The process of linguistic formulation thus creates the stable attendable structure to

which subsequent thinkings can attach. Just such a view concerning the potential role of the inner rehearsal of sentences appears in Jackendoff (1996), who suggests that the mental rehearsal of sentences may be the primary means by which our own thoughts are able to become objects of further attention and reflection.

Linguaform reason, if this is correct, is not just a tool for the novice (e.g., as suggested by Dreyfus and Dreyfus 2000). Instead, it emerges as a key cognitive tool by means of which we are able to objectify, reflect upon, and hence knowingly engage with our own thoughts, trains of reasoning, and cognitive and personal characters. This positions language to act as a kind of cognitive superniche: a cognitive niche, one of whose greatest virtues is to allow us to construct ("with malice aforethought," as Fodor 1994, rather elegantly puts it) an open-ended sequence of new cognitive niches. These may include designer environments in which to think, reason, and perform as well as special training regimes to install (and to make habitual) the complex skills such environments demand.

3.7   Self-made Minds

Coming to grips with our own special cognitive nature demands that we take very seriously the material reality of language: its existence as an additional, actively created, and effortfully maintained structure in our internal and external environment. From sounds in the air to inscriptions on the printed page, the material structures of language both reflect, and then systematically transform, our thinking and reasoning about the world. As a result, our cognitive relation to our own words and language (both as individuals and as a species) defies any simple logic of inner versus outer. Linguistic forms and structures are first encountered as simply objects (additional structure) in our world. But they then form a potent overlay that effectively, and iteratively, reconfigures the space for biological reason and self-control.

The cumulative complexity here is genuinely quite staggering. We do not just self-engineer better worlds to think in. We self-engineer ourselves to think and perform better in the worlds we find ourselves in. We self-engineer worlds in which to build better worlds to think in. We build better tools to think with and use these very tools to discover still better tools to think with. We tune the way we use these tools by building educational practices to train ourselves to use our best cognitive tools better. We even tune the way we tune the way we use our best cognitive tools by devising environments that help build better

environments for educating ourselves in the use of our own cognitive tools (e.g., environments geared toward teacher education and training). Our mature mental routines are not merely self-engineered: They are massively, overwhelmingly, almost *unimaginably* self-engineered. The linguistic scaffoldings that surround us, and that we ourselves create, are both cognition enhancing in their own right and help provide the tools we use to discover and build the myriad *other* props and scaffoldings whose cumulative effect is to press minds like ours from the biological flux.

to mistake contingent features of some contemporary *computer models* for deep features of the much broader explanatory framework in which they have sometimes figured.

CHAPTER 2

1. In fact, it is rather doubtful that these kinds of Gibsonian invariant detection involve truly high-bandwidth coupling at all. But given the extreme difficulty of finding a noncontroversial measure of objective bandwidth, I am willing to grant this for the sake of argument. My point will be that such high-bandwidth coupling, even if present, does not undermine the idea of interfaces located at just those points.

2. A typical description reads: "Computational Grids enable the sharing, selection, and aggregation of a wide variety of geographically distributed computational resources (such as supercomputers, clusters, storage systems, data sources, instruments, people) and presents them as a single, unified resource for solving large-scale and data intensive computing applications" (quote taken from the GRID computing information center at http://www.gridcomputing.com/).

3. See http://www.stelarc.va.com.au and a full discussion in Clark (2003, chap. 5).

4. It is noteworthy, especially in the light of our previous discussion, that "any expansion of the vRF only followed active, intentional use of the tool not its mere grasping by the hand" (Berti and Frassinetti 2000, 81).

5. Gallagher (1998) expresses the difference like this: "Body schema can be defined as a system of preconscious, subpersonal processes that play a dynamic role in governing posture and movement. . . . There is an important and often overlooked conceptual difference between the subpersonal body schema and what is usually called *body image*. The latter is most often defined as a conscious idea or mental representation that one has of one's own body." See also Gallagher (2005, 17–20).

6. For a lovely example of this, see Gray and Fu (2004). This example is discussed later in chapter 6.

7. I explore these near-future themes in Clark (2003).

CHAPTER 3

1. An example of such a display would be the projection, on demand, of green arrows marking the route to a university library onto a glasses-mounted display. The arrows would appear overlaid upon the actual local scene and would update as the agent moves.

2. This picture fits nicely with Barsalou's account of the relation between public symbols and "perceptual symbol systems." See Barsalou (2003) and, for a fuller story about perceptual symbol systems, Barsalou (1999). For an excellent discussion of these issues, and their relation to claims concerning the "extended mind" (see chapter 4 following), see Logan (2007).

3. Here, too, as John Protevi (personal communication) reminds me, the impact is not always positive. We can just as easily derail our own performance by explicit reflection on our own shortcomings.

4. Sutton explores two detailed examples, one involving batting advice for cricketers and the other concerning instructional nudges for piano playing. Concerning the latter, he writes,

> The sociologist and jazz pianist David Sudnow [2001] describes how explicit verbal phrases and maxims actually became more useful as his skills in improvised jazz piano increased…Sudnow explains his [initial] frustration at his teacher's compressed sayings, such as 'sing while you're playing,' 'go for the jazz,' 'get the time into the fingers,' or especially just 'jazz hands.' These at first make no sense, as the novice pianist is all too conscious of the embodied insecurity of his playing: but…what seemed like just vague words to the novice has now become very detailed practical talk, a shorthand compendium of 'caretaking practices' for toning and reshaping the grooved routines. (Sutton 2007)

5. These formulations are not yet precise enough. But I beg the reader's patience. Getting clear about the ways public language might reasonably be said to inform or transform thought and reason is the main topic of the chapter.

6. In partial support of this claim, notice that there is good evidence that children show attentional biases that are sensitive to the language they are learning (or have learned)—see Bowerman and Choi (2001), Lucy and Gaskins (2001), and Smith (2001). Smith (2001, 113) explicitly suggests that learned linguistic contexts come to "serve as cues that automatically control attention."

7. Gordon (2004) presents converging evidence from a tribe in Amazonia that uses only words for one, two, and many. Numerical cognition in this tribe was clearly affected, such that "performance with quantities greater than three was remarkably poor, but showed a constant coefficient of variation, which is suggestive of an analog estimation process" (496). Thanks to Keith Oatley for drawing this to my attention.

8. Those with deeper experience of numbers may come to have a richer mental representation of 98-ness, of course. But even for these agents, there will be thoughts (about other, much larger numbers perhaps) where this deeper grasp fails, and they only succeed in referring to a unique number thanks to the apparatus of the public number system itself.

9. Robert Rupert (personal communication) suggests that the Fodorian account has resources that could also explain the pattern of results we have been discussing. According to Rupert, the Fodorian need only claim that the number words of different languages, when learned by a single subject, receive different mentalese "translations" and that it is these differences in the associated mentalese terms that explain the differences in the subject's reactions and performances. Such a view, while clearly not ruled out, seems

to me less compelling than the alternative that takes shallow, imagistic inner encodings of the natural language items themselves (perhaps in the manner of Barsalou 2003 and Prinz 2004) to be playing the key cognitive role.

10. Elman cites Dave Rumelhart both for this phrasing and for the guiding conception of words as cues to meanings.

11. See Fodor (1981). Fodor energetically reaffirmed this view in a special session at the 2005 Cognitive Science Conference in Stresa, Italy. For a counterview, see Cowie (1999). For an excellent review of the territory, see Samuels (2004).

CHAPTER 4

1. For a host of other examples, see Laland et al. (2000) and Odling-Smee, Laland, and Feldman (2003). See also Dawkins (1982), Lewontin (1983), Odling-Smee (1988), and Turner (2000).

2. The idea of humans as cognitive niche constructors is familiar within cognitive science. Richard Gregory (1981) spoke of "cognition amplifiers," Don Norman (1993a, 1993b) of "things that make us smart," Kirsh and Maglio (1994) of "epistemic actions," and Daniel Dennett (1996) of "tools for thought."

3. In the paper, Tribble explicitly cites Hutchins's (1995) exploration of ship navigation as emergent from the complex interplay of artifacts and social practices as among the main inspirations for her approach.

4. The sense of constitution here in play is not, however, the strong sense of "being conceptually intertwined with" but the weaker (though still interesting and important) sense of "as it happens being part of the apparatus that implements."

5. Examples include Wilson (2004), Clark (1997a), Hurley (1998), Clark and Chalmers (1998), Dennett (1996), Donald (1991), Hutchins (1995), Menary (2007), Wheeler (2005), Sutton (2002b), and Rowlands (2006). See also Rockwell (2005b).

6. I shall focus, however, on a subclass of epistemic actions: the subclass involving "active dovetailing" via the use of multiple nested calls to the world during run-time processing.

7. The experiments used a restricted version of Tetris in which zoids are presented one at a time; see Maglio and Wenger (2000, 2002).

8. Experience is, of course, no *more* than a clue. I do not mean, here or elsewhere, to advance any arguments of the form "it seems to us as if we are/are not cognitively extended; therefore, we are/are not cognitively extended"!

9. It is true, of course, that we can feel like this about our own biological equipment, be it an injured leg or a malfunctioning biomemory. In such cases, it is rather as if we awoke to find some hard-to-use package permanently interfaced to our brain!

10. Thanks to Michael Wheeler for helping to clarify this issue.

11. Kirsh (2004) offers a complementary discussion of the many practical ways metacognition may be spread across neural and extraneural resources.