Supersizing the Mind

Embodiment, Action, and Cognitive Extension

Andy Clark

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Library of Congress Cataloging-in-Publication Data
Clark, Andy, 1957–
Supersizing the mind : embodiment, action, and cognitive extension / Andy Clark.
p. cm. — (Philosophy of mind)
ISBN 978-0-19-533321-3 (Hbk)
978-0-19-977368-8 (Pbk)
1. Philosophy of mind. 2. Mind and body.
3. Distributed cognition. I. Title.
BD418.3.C532 2008
128'.2—dc22 2007051359

OXFORD
UNIVERSITY PRESS

2011

5

Mind Re-bound?

5.1 *EXTENDED Anxiety*

The physical mechanisms of mind, EXTENDED suggests, are simply not all in the head. Is this correct? To raise this question is not necessarily to doubt that heterogeneous mixes of neural, bodily, and environmental elements support much human problem-solving behavior or that understanding such coalitions matters for understanding human thought and reason. It is certainly important, for example, that we appreciate and learn how to analyze the role of epistemic actions in Tetris, of deictic pointers in visual problem solving, and even perhaps of Otto's notebook in his decision making. But should we really count such actions and loops through nonbiological structure as genuine aspects of extended cognitive processes? In this chapter, I consider a range of worries whose starting points concern real or apparent *differences* between what the brain accomplishes and what the other elements in such problem-solving matrices provide.

5.2 *Pencil Me In*

Adams and Aizawa, in a series of recent and forthcoming papers (2001, in press-a, in press-b), seek to refute, or perhaps merely to terminally

embarrass, the proponents of EXTENDED. One such paper begins with the following illustration:

> Question: Why did the pencil think that 2 + 2 = 4?
> Clark's Answer: Because it was coupled to the mathematician.
> (Adams and Aizawa in press-a, 1)

"That," the authors continue, "about sums up what is wrong with Clark's extended mind hypothesis." The example of the pencil, they suggest, is just an especially egregious version of a fallacy said to pervade the literature on distributed cognition and the extended mind. This fallacy, which they usefully dub the "coupling-constitution fallacy," is attributed, in varying degrees and manners, to Van Gelder and Port (1995), Clark and Chalmers (1998), Haugeland (1998), Dennett (2000), Clark (2001a), Gibbs (2001), and Wilson (2002).[1] The fallacy is to move from the causal coupling of some object or process to some cognitive agent to the conclusion that the object or process is part of the cognitive agent or part of the agent's cognitive processing (see, e.g., Adams and Aizawa in press-a, 2).[2]

Proponents of the extended mind and related theses are said to be prone to this fallacy in part because they either ignore or fail to properly appreciate the importance of "the mark of the cognitive"—namely, the importance of an account of "what makes something a cognitive agent" (Adams and Aizawa in press-a). The positive part of Adams and Aizawa's critique then emerges as a combination of the assertion that this "mark of the cognitive"[3] involves the idea that "cognition is constituted by certain sorts of causal process that involve non-derived contents" (in press-a, 3) and that these processes look to be characterized by psychological laws that turn out to apply to many internal goings-on but not currently (as a matter of contingent empirical fact) to any processes that take place in nonbiological tools and artifacts. Let's take these matters in turn.

5.3 *The Odd Coupling*

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

> Question: Why did the V4 neuron think that there was a spiral pattern in the stimulus?
> Answer: Because it was coupled to the monkey.

Now clearly, there is something wrong here. But the absurdity lies not in the appeal to coupling but in the idea that a V4 neuron (or even a group of V4 neurons or even a whole parietal lobe) might *itself* be some kind of self-contained locus of thinking.[4] It is indeed crazy to think that a V4 neuron thinks, and it is (just as Adams and Aizawa imply) crazy to think that a pencil might think. Yet the thrust of Adams and Aizawa's rhetoric is mostly to draw attention to the evident absence of cognition *in the putative part* as a way of "showing" that coupling (even when properly understood; see later) cannot play the kind of role it plays in the standard arguments for cognitive extension. Thus, we read: "When Clark *makes an object cognitive* when it is connected to a cognitive agent, he is committing an instance of a 'coupling-constitution fallacy'" (Adams and Aizawa in press-a, 2, emphasis added).

But this talk of an object's being or failing to be "cognitive" seems almost unintelligible when applied to some putative *part or aspect* of a cognitive agent or of a cognitive system. What would it mean for the pencil *or* the neuron to be, as it were, brute factively "cognitive"? This is not, I think, merely an isolated stylistic infelicity on the part of Adams and Aizawa. For the same issue arose many times during personal exchanges concerning the vexing case of Otto and his notebook.[5] And it arises again, as we shall later see, in various parts of their more recent challenges concerning "the mark of the cognitive."

Let us first be clear, then, about the precise role of the appeal to coupling in the arguments for cognitive extension. The appeal to coupling is not intended to make any external object cognitive (insofar as this notion is even intelligible). Rather, it is intended to make some object, which in and of itself is not usefully (perhaps not even intelligibly) thought of as *either cognitive or noncognitive*, into a *proper part of some cognitive routine.* It is intended, that is to say, to ensure that the putative part is poised to play the kind of role that *itself* ensures its status as part of the *agent's* cognitive routines. Now, it is certainly true (and this, I think, is one important fact to which Adams and Aizawa's argument quite properly draws the reader's attention) that not just any old kind of coupling will achieve this result. But as far as I am aware, nobody in the literature has ever claimed otherwise. It is not the mere presence of a coupling that matters but the effect of the coupling—the way it poises (or fails to poise) information for a certain kind of use within a specific kind of problem-solving routine.

The question that needs to be addressed, then, is: When is some physical object or process acting as part of a larger cognitive routine? It is not the much murkier (probably unintelligible) question: When should we say of some such candidate part, such as a neuron or a notebook,

that it is *itself* cognitive? In the case of Otto, Clark and Chalmers chose to be guided by a set of intuitions derived from reflection on the ordinary "common-sense" use of talk of nonoccurrent dispositional beliefs. In essence, we took these intuitions and systematically showed that the kinds of coarse-grained functional poise (poise to guide various forms of behavior and various conscious states) associated with such dispositional believings on the part of Otto might sometimes be partially supported by a highly nonstandard physical realization in which a mundane, nonmagical notebook acted as the physical medium of long-term storage.

Clark and Chalmers thus offered an argument (which one may accept or reject; that is, of course, another matter) concerning conditions (not of being cognitive) but for recognition as part of the physical substrate of a cognitive system. The key issues concerned coupling only indirectly; what *mattered* was the achieved functional poise of the stored information. In terms of the form of the argument, this is not even close to the commission of a coupling-constitution fallacy. It is better viewed as a simple argumentative extension of at least a subset (see discussion following) of what Braddon-Mitchell and Jackson (2007) describe, and endorse, as "commonsense functionalism" concerning mental states. According to such a view, normal human agents already command a rich (albeit largely implicit) theory of the coarse functional roles distinctive of various familiar mental states—states such as "believing that the MOMA is on 53rd Street." Knowledge of such roles involves knowledge "of the essentials of a certain complex and detailed story about situations, behavioral responses, and mental states" (Braddon-Mitchell and Jackson 2007, 63). This is to be distinguished from the kinds of "empirical functionalism" (Braddon-Mitchell and Jackson 2007, chap. 5) that would use the folk knowledge only as a kind of staging post, going on to identify mental states with further functional role properties as identified by scientific investigation.[6] (Note that Clark and Chalmers's argument concerned only a subset of the folk-identified mental states, since all it requires is a form of common-sense functionalism concerning nonconscious, dispositional states.[7] As such, the argument does not commit us to any sort of functionalism about conscious mental states.[8])

EXTENDED thus involves a kind of double appeal to the functional or systemic role. First, there is an appeal to the common-sense or coarse-grained role implicitly grasped by normal human agents: a broad pattern of flexible, informationally sensitive systemic behavior that underwrites the ascription of some mental state or cognitive activity (dispositional belief, in the case of Otto). Second, we may go on to seek a much more fine-grained description of the actual flow of processing and representation in the (possibly extended) physical array that *realizes* the coarse functional role itself. It is the coarse or common-sense functional role that, on this model (unlike that of empirical functionalism), displays what is essential to the mental state in question. By way of contrast, "distributed functional decompositions," in the sense introduced in section 1.4, are concerned with the second project—namely, the description of how specific systems (perhaps extending across brain, body, and world) realize the common-sense functional role. In laying out the details at this more refined (and cognitive scientifically interesting) level, we display only the *particular way* that a given physical system manages to realize the mental state or activity in question.

5.4 *Cognitive Candidacy*

Adams and Aizawa seem to suggest that some objects or processes, *in virtue of their own nature*, are, as I shall now put it, at least *candidate parts* for inclusion in a cognitive process. And they think that other objects or processes, still in virtue of their own nature, are not even candidates. Or such, I think, is the best way to give sense to that otherwise baffling question "is some X cognitive?" when asked of some putative part of the realizing apparatus. Thus, they ask "if the fact that an object or process X is coupled to a cognitive agent does not entail that X is part of the cognitive agent's cognitive apparatus, what does? *The nature of X of course*. One needs a theory of what makes a process a cognitive process. One needs a theory of the 'mark of the cognitive.'" (Adams and Aizawa in press-a, 3, emphasis added).

What is the mark of the cognitive? The question is nontrivial and has, as Adams and Aizawa (somewhat reluctantly) admit, no well-established answer either within cognitive science or philosophy of mind. Nonetheless, they tie their colors to what they depict as "a rather orthodox theory of the nature of the cognitive" (Adams and Aizawa 2001, 52). According to this theory, "cognition involves particular kinds of processes involving non-derived representations" (53). This is the line also pursued in Adams and Aizawa (in press-a and in press-b). It comprises two distinct elements—namely, an appeal to "non-derived representations" and an appeal to "particular kinds of process."

Despite its prominence in their account, Adams and Aizawa tell us very little about what the first of these (nonderived representations) might amount be. We learn that they are representations whose content is in some sense intrinsic (2001, 48). We learn that this is to be *contrasted* with, for example, the way a public language symbol gets its content

by "conventional association" (48). And we are told, in the same place, that Dretske, Fodor, Millikan, and others are sometimes in search of an adequate theory of such content and that the combination of a language of thought with some kind of causal–historical account is a hot contender for such an account.

Of course, we are not *required* to think of Otto's notebook as contravening some plausible story about intrinsic content. A plausible response would be to argue that what makes *any* symbol or representation (internal or external) mean what it does is just something about its behavior-supporting role (and maybe its causal history) within some larger system. We might then hold that when we understand enough about that role (and perhaps, history), we will see that the encodings in Otto's notebook are in fact on a par with those in his biological memory. In other words, just because the symbols in the notebook happen to look like English words and require some degree of interpretative activity when retrieved and used, that need not rule out the possibility that they have also come to satisfy the demands on being, in virtue of their role within the larger system, among the physical vehicles of various forms of intrinsic content.[9]

Recall that Adams and Aizawa insist that "whatever is responsible for non-derived representations seems to find a place only in brains" (2001, 63). I am not convinced this is true. It seems quite possible, for example, to ascribe representational contents, in ways that are not obviously conventional or derivative, to the states and processes of artificially evolved creatures (see Pfeifer and Scheier 1999, chap. 8). Or if simple artificial creatures do not move you, take any inner neural structure deemed (by whatever nonquestion-begging criteria Adams and Aizawa choose) to be the vehicle of some intrinsic content X. Can we not imagine replacing part or all of that structure with a functionally equivalent silicon part? (As a matter of fact, this kind of replacement has already been done, albeit only with one artificial neuron that functions successfully within a group of 14 biological neurons in a Californian spiny lobster; see Szucs et al. 2000). Unless we question-beggingly assert that only neural stuff can be the bearer of intrinsic content, then surely we should allow that the siliconized vehicle, or at least the hybrid circuit that now includes it, is as capable of supporting intrinsic content as was its biological predecessor. For these kinds of reason, I do not believe that there is any nonquestion-begging notion of intrinsic content that picks out all and only the neural in any clear and useful fashion.

But since Adams and Aizawa stress that they are defending only a contingent, humans-as-currently-constituted, form of cognitive intracranialism, I suspect that they will concede this general point without much argument. The force of Adams and Aizawa's worry does not lie in any simple (and surely naive) identification of the neural and the cognitive. Rather, the real worry is that the inscriptions in Otto's notebook (unlike, say, the hybrid neural and silicon-based activity that now underlies control of the oscillatory rhythms in the stomatogastric ganglion in the Californian spiny lobster) are out-and-out conventional. They are passive representations that are parasitic, for their meaning, on public practices of coordinated use.

Let us agree that there is something quite compelling about the idea that the notebook encodings are all conventional and derivative. Let us agree also, at least for the sake of argument, that some parts of any genuinely cognitive system need to trade in representations that are not thus conventional and derivative. To accept all this, however, is not to give up on the extended mind claim for Otto, unless one *also* accepts (what seems to me an independent and far less plausible assertion) that *no proper part* of a properly cognitive system can afford, at any time, to trade solely in conventional representations.

In Clark (2005b), I offered a thought experiment meant to show that such an additional requirement was too strong and should be rejected. The thought experiment concerned Martians endowed with an extra biological routine that allowed them to store *bitmapped images* of important chunks of visually encountered text. Later on, at will, they could access (and then interpret) this stored text. Surely, I argued, we would have no hesitation in embracing that kind of bitmapped storage, even prior to an act of retrieval, as part and parcel of the Martian cognitive equipment. But what is stored is just a bitmapped image of a fully conventional form of external representation. Upon retrieval, that image, too, would need to be interpreted to yield useful effects. If, courtesy of our common-sense psychological intuitions, we accept this aspect of Martian memory into the cognitive fold, surely only skin-and-skull-based prejudice stops us from extending the same courtesy to Otto. To do so is simply to abide by the Parity Principle as it was meant to be deployed. Thus, even if we demand the involvement, in any cognitive process, of *at least some* items that bear their contents intrinsically, it is quite unclear how we should distribute this requirement across time and space. The Martian encodings are poised, here and now, to participate in processes that invoke intrinsic contents. So are those in Otto's notebook. Since it is arguably poise that matters, at least where dispositional believing is concerned, it seems that any reasonably plausible form of the requirement involving intrinsic content can, with a little imagination, be met. From the requirement, if it is a requirement, that

every truly cognitive agent trade in states that bear intrinsic contents, it cannot follow that every proper part of the cognitive system of an agent must trade (and trade solely) in such contents.

5.5 *The Mark of the Cognitive?*

Consider now the other major part of Adams and Aizawa's challenge. Recall that their suggestion concerning the "mark of the cognitive" was that "cognition involves particular kinds of processes involving non-derived representations" (Adams and Aizawa 2001, 53). We have, I think, just said all that needs to be said concerning the appeal to non-derived representation. But what about the other part of the clause, the appeal to "particular kinds of process" involving such representations? It is at this point that a new kind of consideration comes into play. This concerns the possible existence of a *characteristic set of causal processes* found, by painstaking empirical investigation, to pervade the internal, biologically supported aspects of human cognitive architecture. The operation of these signature causal processes, the authors claim, gives rise to a number of laws and regularities that seem to apply to these known cognitive processes but that do not apply elsewhere (e.g., to Otto's notebook). In the light of this, Adams and Aizawa ask, shouldn't we judge that the notebook falls outside the class of the cognitive? We should indeed do so, they claim, because "the cognitive must be discriminated on the basis of underlying causal processes" (Adams and Aizawa 2001, 52).

The kinds of law and regularity the authors have in mind here include the pervasiveness in human biological memory systems of effects of chunking, priming, recency, and so forth (Adams and Aizawa 2001, 61) and in human perceptual systems of various psychophysical laws (e.g., Weber's law, according to which the change in a stimulus that will be "just noticeable" is a constant ratio of the original stimulus). Given that science has uncovered these undeniably important and interesting regularities, what does this imply concerning the nature of cognition? Adams and Aizawa's argument seems to go like this. Empirical investigations have turned up a number of features (e.g., priming effects in the case of memory) that reflect the detailed operation of processes internal to the brain. Since these clearly pertain to some of our paradigm cases of terrestrial cognition, we should (defeasibly) believe that these kinds of causal process are essential to the "cognitive" status (I use this notion with great discomfort for the reasons mentioned earlier in sec. 5.3) of the neural goings-on.

But this is something we should surely deny. Do Adams and Aizawa really believe that the cognitive status of some target process requires that process to exhibit all the idiosyncratic features of terrestrial neural activity? To insist that some alien mode of storage and retrieval was not cognitive just because it failed to exhibit features such as recency, priming, and crosstalk would be simultaneously to scale new heights of anthropocentrism and neurocentrism, inflating properties of the human neural realizers of certain brainbound cognitive process into requirements that must be met before any process is properly deemed cognitive. Such inflation is both undesirable in itself and question begging in the context of arguments for the extended mind.

One might also reflect that, for all we know, the fine details of the causal role of, say, stored beliefs differ from person to person or (within one person) from hour to hour.[10] This point is merely dramatized by those alien beings whose recall is not subject to recency effects, crosstalk, or error. Do such differences make a difference? Is the mutant human whose recall is fractionally slower, fractionally faster, or much less prone to loss and damage also to be banned from the ranks of true believers and rememberers? To demand identity of fine-grained causal role is surely to set the cognitive bar too high and way too close to home.

5.6 *Kinds and Minds*

In their 2001 paper, Adams and Aizawa also raise a different (though related) kind of worry. This concerns the nature and feasibility of the scientific enterprise implied by taking so-called transcranialism seriously. The worry, in its simplest form, is that "science tries to carve nature at its joints" (51). But they argue that the various types of neural and extraneural goings-on that the transcranialist lumps together as cognitive seem to have little or nothing in common by way of underlying causal processes.

To make this concrete, we are invited to consider once again (see sec. 4.8) the process that physically rotates the image on the Tetris screen. This, they correctly note, is nothing like any neural process. It involves firing electrons at a cathode ray tube! It requires muscular activity to operate the button. Similarly, "Otto's extended 'memory recall' involves cognitive-motor processing not found in Inga's memory recall" (Adams and Aizawa 2001, 55). More generally, they suggest, just look at the range of human memory augmenting technologies (photo albums, Rolodexes, Palm Pilots, notepads, etc.): "what are the chances

of there being interesting regularities that cover humans interacting with all these sorts of things? Slim to none, we speculate" (61).

By contrast, biological memory systems, as noted previously, are said to "display a number of what appear to be law-like regularities, including primacy effects, recency effects, chunking effects and others" (61). And unlike the biological memory processes, "transcranial [extended] processes are not likely to give rise to interesting scientific regularities. There are no laws covering humans and their tool-use over and above the laws of intercranial [inner] human cognition and the laws of the physical tools" (61).

The first thing to say in response to all this is that it is probably unwise to judge, from the armchair, the chances of finding "interesting scientific regularities" in any domain, be it ever so superficially diverse. Consider, for example, the recent successes of complexity theory in unearthing unifying principles that apply across massive differences of scale, physical type, and temporality. There are power laws, it now seems, that compactly explain aspects of the emergent behavior of systems ranging from ant colonies to the World Wide Web. In a similar vein, it is quite possible that despite the bottom-level physical diversity of the processes that write to and read from Otto's notebook, and those that write to and read from Otto's biological memory, there is a level of description of these systems that treats them in a single unified framework (e.g., how about a framework of information storage, transformation, and retrieval?). The mere fact that Adams and Aizawa can find *one* kind of systemic description at which the underlying processes look wildly different says very little, really, about the eventual prospects for an integrated scientific treatment. It is rather as if an opponent of rule and symbol models of mental processing were simply to cite the deep physical differences between brains and von Neumann computers as proof that there could be no proper science that treated processes occurring in each medium in a unified way. Or to take a different kind of case, as if one were to conclude from the fact that chemistry and geology employ distinct vocabularies and techniques, that the burgeoning study of geochemistry is doomed from the outset. But neither of these, I presume, are conclusions that Adams and Aizawa would wish to endorse.

The bedrock problem thus lies with the bald assertion that "the cognitive must be discriminated on the basis of underlying causal processes" (Adams and Aizawa 2001, 52). For it is part of the *job* of a special science to establish a framework in which superficially different phenomena can be brought under a unifying explanatory umbrella. To simply cite radical differences in some base-level physical story goes no

way at all toward showing that this cannot be done. Moreover, it is by no means clear that acceptable forms of unification require that all the systemic elements behave according to the same laws. As long as there is an intelligible domain of convergence, there may be many subregularities of many different kinds involved. Think, for example, of the multiple kinds of factor and force studied by those interested in creating better home audio systems. Even if "home audio" is rejected as any kind of unified science, it certainly names a coherent and proper topic of investigation. The study of mind might, likewise, need to embrace a variety of different explanatory paradigms whose point of convergence lies in the production of intelligent behavior.

Moreover, it seems quite possible that the *inner* goings-on that Adams and Aizawa take to be paradigmatically cognitive themselves will turn out to be a motley crew as far as detailed causal mechanisms go, with not even a family resemblance (at the level of actual mechanism) to hold them together. It is arguable, for example, that conscious seeing and nonconscious uses of visual input to guide fine-grained action involve radically different kinds of computational operation and representational form (see, e.g., Milner and Goodale 1995; Goodale and Milner 2004). And Adams and Aizawa to the contrary, some kinds of mental rehearsal (e.g., watching sports or imagining typing a sentence) do seem to reinvoke distinct motor elements, whereas others (e.g., imagining a lake) do not (see Decety and Grezes 1999). Some aspects of biological visual routines may even use a form of table lookup (Churchland and Sejnowski 1992). In addition, the inner mechanisms of mind seem to include both conscious, controlled, slow processes and fast, automatic, uncontrolled ones, with each of these sets of processes displaying its own characteristic sets of regularities (see Shiffrin and Schneider 1977; and for more recent discussions, Wegner 2005; Bargh and Chartrand 1999). Among such regularities, we may count the finding that controlled processes tend to degrade rapidly under cognitive load, whereas automatic processes do not; that controlled processes are apt for conscious interruption, whereas automatic ones are not; that controlled processes are slow, whereas automatic ones are relatively fast; and so on. With such findings in mind, Levy (in press) concludes that "if it is true that causal regularities pick out natural kinds, then the mind is not a natural kind: it is a compound entity comprised of at least two (and probably many) natural kinds."

In the light of all this, my own suspicion is that the differences between external-looping (putatively cognitive) processes and purely inner ones will be *no greater than those between the inner ones themselves.* But insofar as they all form parts of a flexible and information-sensitive

control system for a being capable of reasoning, of feeling, and of experiencing the world (a "sentient informavore" if you will), the motley crew of mechanisms has something important in common. It may be far less than we would require of any natural or scientific kind. But so what?

The argument from scientific kinds is thus doubly flawed. It is flawed by virtue of its rather limited conception of what makes for a proper scientific or explanatory enterprise. And it is flawed in its assessment of the potential for some form of higher level unification despite mechanistic dissimilarities. It is, above all else, a matter of empirical discovery, not armchair speculation, whether there can be a fully fledged science of the extended mind.

It is also perhaps worth noting that nascent forms of just such a science have been around for quite some time. The field of human-computer interaction (HCI) and its more recent cousins human-centered computing (HCC) and human-centered technologies (HCT) are ongoing attempts to discover unified scientific frameworks in which to treat processes occurring in and between biological and nonbiological information-processing media (see, e.g., Scaife and Rogers 1996; Norman 1999; Dourish 2001).

Adams and Aizawa next attempt to parlay the misconceived appeal to scientific kinds into a kind of dilemma. Either, the argument goes, Clark and Chalmers are radically mistaken about the causal facts, or more likely, they are closet behaviorists. On the one hand, if our claim is that "the active causal processes that extend into the environment are just like the ones found in intracranial cognition" (Adams and Aizawa 2001, 56), we are just plain wrong. On the other hand, if we don't care about that and claim only that "Inga and Otto use distinct sets of capacities in order to produce similar behavior" (56), then we are behaviorists.

This is surely a false dilemma. To repeat, our claim was not that the processes in Otto and Inga are identical, or even similar, in terms of their detailed implementation. It is simply that, with respect to the role that the long-term encodings play in guiding current response, both modes of storage can be seen as supporting dispositional beliefs. It is the way the information is poised to guide reasoning (e.g., conscious inferences that nonetheless result in no overt actions) and behavior that counts. This is not behaviorism but (extended) common-sense functionalism. It is coarse systemic role that matters, not brute similarities in public behavior (though the two are of course related). Perhaps Adams and Aizawa believe that common-sense functionalism just *is* a species of behaviorism. That seems wrong, however, because common-sense functionalism is quite compatible with the assertion that there are *some* internal constraints on being a cognizer. Thus, Braddon-Mitchell and Jackson (2007, chap. 5 and 7) argue that a creature all of whose actions were generated by table lookup would not count, even by the standards of common-sense functionalism, as a thinker. Such coarse architectural requirements flow, they believe, from ordinary intuitions about mind and reason. The issue between the common-sense functionalist and the empirical functionalist is thus not whether there are any internal constraints on being a thinker but "whether it is right to let the *particular* way that we handle the informational problems set by the world dictate what is to count as having a mind" (94). To this question they, and the common-sense functionalist, give a firmly negative response.

A related concern was raised by Terry Dartnall (personal communication). Dartnall worried that the plausibility of the Otto scenario depends on an outmoded image of biological memory itself: the image of biological memory as a kind of static store of information awaiting retrieval and use. This image, Dartnall claimed, cannot do justice to the active nature of real memory. It is somewhat ironic, Dartnall argued, that the present author (in particular) should succumb to this temptation, given his long history of interest in, and support for, the connectionist alternative to classical (text- and rule-based) models of neural processing. By way of illustration (though the illustration may actually raise other issues, too, as we shall see), he offered the following example: Suppose I have a chip in my head that gives me access to a treatise on nuclear physics. That doesn't make it true that *I know* about nuclear physics. In fact, the text might even be in a language I don't understand. "Sterile text," Dartnall concluded, cannot support cognition (properly understood). In a sense, then, the claim once again is that text-based storage is so unlike biological memory that any claim of role parity must fail.

This is an interesting line of objection but one that ultimately fails for reasons closely related to the discussion of intrinsic content in section 5.2. Certainly, biological memory is an active process. And retrieval is to a large extent reconstructive rather than literal: What we recall is influenced by our current mood, by our current goals, and by information stored after the time of the original experience. It is possible, in fact, that biological memory is such an active process as to blur the line between memory systems and reasoning systems. All this I happily accept. But to repeat, the claim is that in the special context of the rest of Otto's information-processing economy, the notebook is co-opted into playing a real cognitive role. And the informal test for this is, just supposing some inner system provided the functionality that Otto derives

from the reliable presence of the notebook, would we hesitate to classify that inner system as part of Otto's cognitive apparatus?

Readers must here rely on their own intuitions. But according to Clark and Chalmers, there would be no such hesitation. To cement the intuition, recall once more (sec. 5.2) the Martians with their additional bitmapped memories or humans with quasi-photographic recall. Or consider the familiar act of rote learning. When we learn a long text by rote, we create a memory object that is in many ways unlike the standard case. For example, to recall the sixth line of the text, we may have to first rehearse the others. Moreover, we can rote learn a text we do not even understand (e.g., a Latin text). Assuming that we count rote learning as the acquisition of some kind of knowledge (even in the case of the Latin text), it seems that we should not be bothered by the consequences that Dartnall unearths. The genuine differences that exist between the notebook-based storage and standard cases of biological memory do not matter because our claim was not one of identity in the first place.

The deeper question is thus how to balance the Parity Principle (which makes no claims about process-level identity at all) against the somewhat stronger claim of "sufficient functional similarity" that underpins treating Otto's notebook as a contributor to Otto's long-term store of dispositional beliefs. Part of the answer emerges as soon as we focus on the role the retrieved information will play in guiding current behavior. It is at that point (and there, of course, all kinds of active and occurrent processing come into play as well) that the common-sense functional similarity becomes apparent. True, that which is stored in Otto's notebook won't shift and alter while stored away. It won't participate in the ongoing underground reorganizations, interpolations, and creative mergers that characterize much of biological memory. But *when called upon*, its immediate contributions to Otto's behavior still fit the profile of a stored belief. Information retrieved from the notebook will guide Otto's reasoning and behavior in the same way as information retrieved from biological memory. The fact that *what* is retrieved may be different is unimportant here. Thus, had Otto stored the information about the color of the car in the auto accident in biological memory, he may be manipulated into a false memory situation by a clever experimenter. The notebook storage is sufficiently different to be immune to that manipulation (though others will be possible). But the information recalled (veridical in one case but not the other) will nonetheless guide Otto's behavior (the way he answers questions and the further beliefs he forms etc.) in exactly the same kind of way. Or simply reflect that for many years the classical "text- and rule-based" image of human cognition was widely accepted. During that time, nobody (to my knowledge) thought that an implication of this was that humans were not cognizers! It might have turned out that all our memory systems operated as sterile storage and that false memory cases and so on were all artifacts of retrieval processes. This shows, again, that there is nothing intuitively noncognitive about less active forms of storage.

Does the stress on similarity of coarse-grained functional role commit us to a merely prosthetic use of nonbiological props and aids? That is, does it commit us to the nonbiological structures merely standing in (as in the case of Otto) for what is normally provided by fully internal means? The many examples sketched in earlier chapters suggest it does not. We should instead be impressed by our remarkable capacity to form extended, densely integrated systems that factor in a variety of distinctive contributions, some of which have no clear internal analogs (a simple example might be an architect whose fluent problem depends in part on the functioning of a fancy software package).[11] Given sufficient complementarity and integration, I want to say, we may sometimes confront hybrid systems displaying novel cognitive profiles that supervene on more than the biological components alone.[12]

Some remain wary of the appeal to complementarity in the nonpathological case. Thus, Michael Wheeler (personal communication) suggests that all the truly persuasive arguments for EXTENDED depend on displaying coarse-grained functional similarities to standard internal cases (e.g., to standing beliefs, as in the case of Otto). Such cases play a key argumentative role but should not be taken as limning the space of extended cognitive circuitry. Rather, they provide the essential first means by which to begin to break the stranglehold of vehicle-internalist intuitions concerning cognition. Once the possibility of vehicle externalism, in humanly possible worlds, is thus established (once, as it were, the hegemony of skin and skull is finally broken), we are free to recognize, as genuinely cognitive and as owned by the human agent, all kinds of process that have no fully biological analog.[13]

5.7 *Perception and Development*

Another common worry, at least about the rather specific test case of Otto (though similar considerations will apply to all manner of actual mind-expanding media and apparatus) is that the role of perception, in "reading in" the information from the notebook, marks a sufficient disanalogy to discount the notebook as part of Otto's cognitive apparatus. We made a few brief comments on this issue in the original paper,

noting that whether the reading in counts as genuinely perceptual or introspective depends, to a large extent, on how one classifies the overall case. From our perspective, the systemic act is more like an act of introspection than one of perception. As a result, each side is here in danger of begging the question against the other.

Thus, Keith Butler complains that

> in the world-involving cases, the subjects have to *act* in a way that demands of them that they perceive their environment [whereas Inga just introspects]... the very fact that the results are achieved in such remarkably different ways suggests that the explanation for one should be quite different from the explanation for the other

and that

> Otto has to look at his notebook while Inga has to look at nothing. (both quotes from Butler 1998, 211)

But from the EXTENDED point of view, Otto's inner processes and the notebook constitute a single cognitive system. Relative to *this* system, the flow of information is wholly internal and functionally akin to introspection (for more on this, see sec. 5.8).

One way to try to push the argument is to seek an independent criterion for the perceptual. With this in mind, Martin Davies (personal communication) has suggested that it is revealing that Otto could misread his own notebook. This opening for error may, Davies suggests, make the notebook seem more like a perceived part of the external world than an aspect of the agent. But parity still prevails: Inga may misremember an event not due to an error in her memory store but because of some disturbance during the act of retrieval. The opening for error does not yet establish that the error is, properly speaking, perceptual. It only establishes that it occurs during retrieval.

A slight variant, again suggested by Davies, is that perception (unlike introspection) targets a potentially public domain. Notebooks and databases are things to which other agents could in principle have access. But, the worry goes, my beliefs are essentially the beliefs to which *I* have a special kind of access unavailable to others.

Notice first that there is, in any case, something special about Otto's relation to the information in the notebook. For as we commented in the original paper, Otto more or less automatically endorses the contents of the notebook. Others, depending on their views of Otto, are less likely to share this perspective. But this is not a special kind of access as much as a special kind of cognitive relationship. But why then

suppose that uniqueness of access is anything more than a contingent fact about standard biological recall? If, in the future, science devised a way for you to occasionally tap into my stored memories, would that make them any less *mine* or part of my cognitive apparatus? Imagine, for that matter, a form of multiple personality disorder (MPD) in which two personalities have equal access to some early childhood memories. Here we have, at least arguably, a case where two distinct persons share access to the same memories. Of course, one may harbor all kinds of reasonable doubts about the proper way to conceptualize MPD in general. But the point is simply that it seems to be at most a contingent fact that I and I alone have a certain kind of access to my own biologically stored memories and beliefs.

Before leaving this topic, I want to briefly mention a very interesting worry raised by Ron Chrisley (personal communication). Chrisley notes that, as children, we do not begin by experiencing our biological memory as any kind of object or resource. This is because we do not encounter our own memory perceptually. Instead, it is just part of the apparatus through which we relate to and experience the world. Might it be this special developmental role that decides what is to count as part of the agent and what is to count as part of the wider world?

Certainly, Otto first experiences notebooks, and even his own special notebook, as objects in his world. But I am doubtful that this genuine point of disanalogy can bear the enormous weight that Chrisley's argument requires. First of all, consider the child's own bodily parts. It is quite possible, it seems to me, that these are first experienced (or at least simultaneously experienced) as objects in the child's world. The child sees its own hand. It may even want to grab a toy and be unable to control the hand well enough to do so. The relation here seems relatively "external," yet the hand is (and is from the start) a proper part of the child.

Perhaps you doubt that there is any moment at which the child's own hand is really experienced, or at any rate conceptualized, as an object for the child. But in that case, we can surely imagine future nonbiological (putatively cognitive) resources being developmentally incorporated in just the same way. Such resources would be provided so early that they, too, are not first conceptualized as objects (perhaps spectacles are like this for some of us already). Contrariwise, as Chrisley himself helpfully points out, we can imagine beings who from a young age are taught to experience even their own *inner* cognitive faculties as objects, courtesy of being plugged into biofeedback controllers and trained to monitor and control their own alpha rhythms and so on.

The developmental issue, though interesting, is thus not conceptually crucial. It points only to a complex of contingent facts about human cognition. What counts in the end is the resource's current role in guiding reasoning and behavior, not its historical positioning in a developmental nexus.

5.8 *Deception and Contested Space*

In a most interesting and constructive critique of the Extended Mind Thesis, Kim Sterelny (2004) worries that Clark and Chalmers underplay the importance of the fact that our "epistemic artifacts" (our diaries, Filofaxes, compasses, and sextants) operate in a "common and often contested" space. By this, he means a shared space apt for sabotage and deception by other agents. As a result, when we store and retrieve information from this space, we often deploy strategies meant to guard against such deception and subversion. More generally still, the development and functional poise of perceptual systems are, for this very reason, radically different from the development and functional poise of biologically internal routes of information flow. The intrusion of acts of perception into Otto's information retrieval routine thus introduces a new set of concerns that justify us in not treating the notebook (or whatever) as a genuine part of Otto's cognitive economy.

Sterelny does not mean to deny the importance of epistemic artifacts (as he calls them; see sec. 4.4) in turbo-charging human thought and reason. Indeed, he offers a novel and attractive coevolutionary account in which our ability to use such artifacts both depends on and further drives a progressive enrichment of our internal representational capacities. In this way, "Our use of epistemic artifacts explains the elaboration of mental representation in our lineage and this elaboration explains our ability to use epistemic artifacts" (Sterelny 2004, 239).

What he does mean to deny, however, is that the use of such artifacts reduces the load on the naked brain and that the brain and the artifacts can coalesce into a single cognitive system. Instead, he sees increased load and a firm boundary between the biological integrated system and the array of props, tools, and storage devices suspended in public space. I tend to differ on both counts but will here restrict my comments to the point about the boundary between the agent and the public space.

Within the biological sheath, Sterelny argues, information flow occurs between a "community of co-operative and co-adaptive parts [that are] under selection for reliability." Over both evolutionary and developmental time, the signals within the sheath should become clearer, less noisy, and less and less in need of constant vetting for reliability and veridicality. As soon as you reach the edge of the sheath, however, things change dramatically. Perceptual systems may be highly optimized for their jobs. But it is still the case that the signals they deliver have their origins in a public space populated in part by organisms under pressure to hide their presence, to present a false appearance, or to otherwise trick and manipulate the unwary so as to increase their own fitness at the other's expense. Unlike internal monitoring, Sterelny (2004, 239) says, "perception operates in an environment of active sabotage by other agents [and] often delivers signals that are noisy, somewhat unreliable and functionally ambiguous."

One result of all this is that we are forced to develop strategies to safeguard against such deceptions and manipulations. The cat moves gingerly across the lawn and may stop and look very hard before trusting even the clear appearance of a safe passage to the other side. While at a higher level by far, we may even deploy the tools of folk logic and consistency checking (here, Sterelny cites Sperber 2001).

The point about vulnerability to malicious manipulation is well taken. Many forms of perceptual input are indeed subject, for that very reason, to much vetting and double-checking. I do not think, however, that we treat all our perceptual inputs in this highly cautious way. Moreover, *as soon as we do not do so*, the issue about extended cognitive systems seems to open up (see below). As a result, I am inclined to think that Sterelny has indeed hit on something important here but something that may in the end be helpful, rather than harmful, to the EXTENDED account.

Take the well-known work on magic tricks and so-called change blindness (for a review, see Simons and Rensink 2005, and further discussion in sec. 7.3). In a typical example of such work, you might be shown a short film clip in which major alterations to the scene occur while you are attending to other matters. Often, these alterations are simply not noticed. Once they are drawn to your attention, however, it seems quite amazing that you ever missed them. The art of the stage magician, it is often remarked, depends on precisely such manipulations. We are, it seems, remarkably vulnerable to certain kinds of deception. But this, I want to suggest, may be grist to the extended mind mill. For on a day-to-day basis, the chances of these kinds of espionage are sufficiently low that they may be traded against the efficiency gains of (for some cognitive purposes) leaving some information "out in the world" and relying on just-in-time access. We may, under certain circumstances, treat a perception-involving loop to the environment as if

it were an inner, relatively safe, and noise-free channel, thus allowing us (with some important qualification; see sec. 7.3) to use the world as a form of "external memory" (O'Regan 1992; O'Regan and Noe 2001).

It is important, in our story about Otto, that he, too, treats the notebook as a typically safe and reliable storage device. He must not feel compelled to check and double-check retrieved information. If this should change (perhaps someone does begin to interfere with his external stored knowledge base) and Otto should notice the change and become cautious, the notebook would at that point cease to unproblematically count as a proper part of his individual cognitive economy. Of course, Otto might wrongly become thus suspicious. This would parallel the case of a person who begins to suspect that aliens are inserting thoughts into his or her head. In these latter cases, we begin to treat biologically internal information flow in the cautious way distinctive of (some) perception. What emerges from the considerations concerning espionage and vigilance is thus not so much an argument *against* the extended mind as a way of further justifying our claim that in some contexts signals routed via perceptual systems are treated *in the way more typical of internal channels* (and vice versa in the case of feared thought insertion). To decide, in any given case, whether the channel is acting more like one of perception or more like one of internal information flow, we must look to the larger functional economy of conscious vigilance and active defenses against deception. The lower the vigilance and defenses, the closer we approximate to the functionality of a typical internal flow.

Sterelny might reply to this by shifting the emphasis from the extent to which agents actually do guard against deception and manipulation to the extent to which they are, as a matter of fact, vulnerable to it. Thus, the fact that we are vulnerable to the magician's art may be said to count for more than the fact that in being thus vulnerable we treat (as I tried to argue) the perceptual route as a quasi-internal one. But this seems unprincipled because, given the right "magician" (say, an alien able to directly affect the flow of energy between my synapses), all routes seem about equally vulnerable. Recall also that false beliefs can (as noted earlier) be generated in biological memory by quite simple psychological manipulations. Or for that matter, consider the many ways in which biological memory and reason can be systematically impaired (e.g., the patients whose memories, like their ongoing experience, exhibit hemispatial neglect; Bisiach and Luzzatti 1978; Cooney and Gazzaniga 2003). What seems to count is not vulnerability as such but rather something like our "ecologically normal" level of vulnerability. And our actual practices of defense and vetting are, I claim, rather

a good guide to this. If Otto doesn't worry about tricksters copying his writing and adding false entries, maybe that is because the channel is *as secure as it needs to be.*

5.9 *Folk Intuition and Cognitive Extension*

Consider the following challenge to the story currently under consideration:

> You invoke our implicit grasp of a common-sense model of mind as part of the case for thinking that (the physical machinery underlying some) mental states and processes extends out into the world. But that latter picture is itself so radically opposed to what common sense believes as to belie the premise. How can our intuitive pretheoretic grip on the notion of mind yield such counterintuitive fruit?

The first point to note is that all the argument requires is an appeal to some notion of the coarse (i.e., unscientifically visible) role associated with some mental state. Given *just that much grip on the mind*, so the argument goes, we can be brought to see (as in the case of Otto) that bioexternal stuff may sometimes help to realize that role. If that comes as something of a surprise, it in no way undermines the form of argument.

Nonetheless, I am also inclined (though nothing in the present treatment depends on this) to dispute the claim that the Extended Mind Model runs so wildly contrary to common sense. For it is only counterintuitive, it seems to me, if we are already in the grip of a form of theoretically loaded neurocentrism. If we subtract the loaded neurocentric intuitions, it is by no means clear that the common-sense grip on mind has any fixed opinion concerning the location of the machinery of mind. Indeed, insofar as one can discern any leanings at all, they may even contain traces of the extended model. For example, ordinary talk about one another's plans and intentions seems already to allow that external media (and often other agents, too) can play the role of physical vehicles for various contents. As Houghton (1997) convincingly argues, it is perfectly in keeping with standard ways of thinking to say that *my* plans for a week's vacation have detailed contents that I never hold, all at once, in my head, let alone before conscious inspection. Similarly, the architect may properly be said to have complex standing intentions, vehicled in drawings and drafts, regarding the shape and structure of the building even though she may never hold, or even have held, the

full sequence and combination of features (the ones that together form the content of those very intentions) in her head or before conscious inspection. To insist that the architect's real intentions are something less (perhaps merely to build whatever the plans she has drafted happen to describe) is surely to do her a serious injustice. The folk grip on mind and mental states, it seems to me, is surprisingly liberal when it comes to just about everything concerning machinery, location, and architecture.

5.10 *Asymmetry and Lopsidedness*

Such liberality is notably absent from Adams and Aizawa's account. The general form of their argument has as a consequence a claim that we may now dub the Dogma of Intrinsic Unsuitability. It goes like this:

> Dogma of Intrinsic Unsuitability
>
> Certain kinds of encoding or processing are intrinsically unsuitable to act as parts of the computational substrate of any genuinely cognitive state or process.

In Adams and Aizawa (2001), the dogma emerged as the claim that certain human neural states, and no extraneural goings-ons, exhibit "intrinsic intentionality," conjoined with the assertion that no proper part of a truly cognitive process can trade solely in representations lacking such intrinsic content (e.g., the conventionally couched encoding in Otto's notebook). The dogma was also at work in their later suggestion that cognitive psychology, in discovering pervasive features of inner biological systems of memory and perception, is uncovering the essential signatures of the kinds of causal process required of all possible forms of cognition.

The Dogma of Intrinsic Unsuitability is, however, just that—a dogma. Moreover, it is one that is ultimately in some tension with a cognitive scientific commonplace that might be dubbed the Tenet of Computational Promiscuity—namely, the idea that pretty much any kind of processing or encoding can form part of an information-based system for flexible adaptive response, just as long as it is properly located in some larger ongoing web of activity. When computational promiscuity meets intrinsic unsuitability, something surely has to give. I think what has to give is pretty clearly the notion of intrinsic unsuitability.

Part of the problem here is that the Dogma of Intrinsic Unsuitability is superficially similar to a quite different and rather more plausible claim—namely:

> Claim of Intrinsic Suitability
>
> Certain kinds of processing and encoding are intrinsically suited to act as the computational substrate of the kinds of fluent, pattern-sensitive engagement characteristic of, and perhaps even essential to, the behavior of intelligent organisms.

Such a claim may well be true. It may, for example, be the case that the action of some kind of interpolating statistical sponge (e.g., a connectionist-style associative learning device) provides the only computationally viable means of supporting some of the basic skills of perceiving and learning that we share with many other earthly animals. At the heart of this skill set lie the rich abilities of subtle pattern recognition that we share with many other animals and that allow us to learn about important regularities in our environment by exposure to repeated exemplars. In combination with affective and motivational systems, this kind of potent, slow, pattern-based learning enables many animals, ourselves included, to learn to deal with highly complex situations in a remarkably nuanced and efficient manner. Since these features are plausibly crucial to the kinds of fluent, adaptable, real-world responses we demand of intelligent beings, it may turn out (purely as a matter of empirical fact) that cognizing systems always incorporate some, very loosely speaking, connectionist kinds of computational underpinning.

Even if this is true, however, it does not follow that, *once such core systems are in place*, other kinds of representational and computational resources may not come to act, either temporarily or permanently, as proper parts of more complex, hybrid, distributed, cognitive wholes. In such cases, it is the very fact that these additional elements trade in modes of representation and processing that are *different* from those of the cognitive core that makes the hybrid organization worthwhile. Tracing and understanding such deep complementarity are surely the most important tasks confronting the sciences of situated cognition. If we embrace the idea of such a cognitive core, we can happily accept, for example, that no genuinely cognitive system will turn out to consist entirely of the kinds of external resources that fans of extended cognition most typically invoke. This is fully compatible, however, with the claim that new integrated and genuinely cognizing wholes are sometimes brought into being on the back of those more basic, perhaps even cognitively indispensable, sets of skills and capacities.

Much opposition to EXTENDED, and the quite palpable unease it causes even in some of its most sensitive critics, may thus be rooted in the mistaken fear that by celebrating the power of new, hybrid, extended systems we lose sight of that crucial cognitive core.[14] The fear would be that to embrace hybrid cognitive forms is to lose sight of the unique importance of the core systems upon whose successful operation *the very possibility of such extended forms depend*. But such fears are groundless. It is not part of the EXTENDED agenda to attempt to wash out all the differences between various internal and external contributions or to downplay or undervalue the potentially unique contribution of the cognitive core. Indeed, the actual research program of distributed cognition is committed, above all, to plotting and charting the varied contributions made by a variety of biological and nonbiological resources and the potent and multilayered interactions between them. The agenda is thus not a negative but a purely positive one: to understand the larger systemic webs that, spun around the common core shared with so many other animals, help to give human cognition its *distinctive* power, character, and charm.

Consider, by way of partial analogy, the more mundane fact that human animals, apparently uniquely on the planet, display (in addition to the common core) a second, rather different set of skills. These are the skills of explicit, deliberative, "language-infected" reason and planning (see, e.g., Dennett 1996, and the more general discussion in chap. 3 of this book). Working together, these two very different sets of skills make us into especially potent cognitive engines. Nonetheless, if we contemplate these two kinds of cognitive resources, it seems compelling that in some very important sense, it is the skills of basic pattern recognition, learning, and affectively tuned response that are the most fundamental. By this I mean only that without these we would probably be unable to have thoughts at all and, ipso facto, unable to have the linguistically infected thoughts. The very same model (depicting an empirically essential core with some mind-bogglingly potent add-ons) may be invoked by the friends of the extended mind. It is surely entirely likely that many of the extended cognitive systems described in this literature are *in just the same sense* less fundamental. They are less fundamental in that no genuinely cognitive system could consist *entirely* of the most typical kinds of external resource (passive notebooks etc.) that currently augment the common core. The contributions are in that sense asymmetrical (Collins in press) or "lopsided" (Rupert in press-a). This, I think, is the important grain of truth underlying Adams and Aizawa's arguments concerning derived contents, conventional encodings, the "noncognitive" status of notebooks, and so forth. It is a grain of truth, however, that is no more damaging to the vision of the extended mind than it is to the vision of the language-infected mind. In each case, powerful new cognitive wholes are brought into being on the back of some set of more basic, and perhaps even cognitively indispensable, skills and capacities. And in each case, the new integrated systems that result are best seen as cognitive systems in their own right. They are, indeed, the cognitive systems whose fluid operation accounts for many of the unique and most characteristic achievements of the human mind.

Notice, finally, that attention to such new and larger systemic wholes in no way precludes a proper investigation of the special features of various parts, aspects, and components. A useful comparison is with the move toward systems-level neuroscience.[15] For much of its history, most serious neuroscientific research concerned the responses and behaviors of single cells. Then, with the advent of new techniques of recording, intervention, and investigation, attention began to be devoted to understanding the neural dynamics of whole populations of cells and the distinctive processing styles of different gross anatomical elements (e.g., the hippocampus and the neocortex). Contemporary neuroscience, courtesy of still newer techniques of imaging and analysis and by using increasingly biorealistic neural network simulations, is just beginning to make progress in understanding some of the key features and properties of even larger scale neural systems: whole processing cycles that involve the temporally evolving, often highly reentrant, activity of multiple populations of neurons spanning a variety of brain areas. The advent of true systems-level neuroscience does not (and should not) imply the inappropriateness of investigations that target the special properties and features of distinct cell types, populations, or neural areas. It simply adds to these investigations a new sensitivity to the value created by processing cycles that include multiple complementary operations, performed at various timescales and using various kinds of neural resources, and whose integrated action is responsible for much of the power and scope of an individual human intelligence. So, too, according to EXTENDED, whole brain-body-world systems can sometimes be the locus of extended processing cycles whose integrated action is responsible for much of what we deem mind and intelligence.

5.11 *Hippo-world*

Imagine a kind of Bizarro-world—call it Hippo-world—in which for half a century, all neuroscientific attention focused on the hippocampus, regarded (for some path-dependent historical reason let's assume) as

the sole and obvious locus of human cognitive activity. Specific features of hippocampal processing and encoding are discovered and publicized. One day, a few researchers turn their attention to the rest of the brain. They discover many new and interesting features and begin to talk about the larger processing circuits that link, for example, hippocampal and neocortical processing and the way certain human memory phenomena seem to depend on the complex interactions between the components. But there is a problem. Some philosophers in Hippo-world believe that in discovering the characteristic causal processes that operate in the hippocampus, they were discovering *the scientific essence of cognition itself*. It is better, they now insist, to view *what the hippocampus does as cognitive* and the rest of the brain as merely sending inputs to, or receiving outputs from, that "truly cognitive part." Only the hippocampus, they suggest, exhibits the "mark of the cognitive." These other parts, after all, just don't do the same things as the hippocampus, so why regard what they do as cognitive? Others demur, for much of what they see as gross intelligent human behavior turns out to depend just as much upon the special features and properties of the other parts as upon the (important but limited) contribution of the hippocampus itself. The study of the extended mind presents no greater theoretical or practical difficulties than those, significant as they were, that might have attended the Hippo-worlders' first tentative moves toward a more inclusive cognitive neuroscience.[16] And it is justified, or so I believe, in very much the same way. In each case, we confront larger scale organizations, defined across a smorgasbord of heterogeneous elements, whose integrated operation makes us the peculiarly successful cognitive agents we are.

6

The Cure for Cognitive Hiccups (HEMC, HEC, HEMC . . .)

6.1 *Rupert's Challenge*

Human cognitive processing, EXTENDED claims, may at times loop into the environment surrounding the organism. Such a view should be contrasted with a nearby, but rather more conservative, view according to which certain cognitive processes lean heavily on environmental structures and scaffoldings but do not thereby include those structures and scaffoldings themselves. This more conservative view, ably championed in a series of papers by Robert Rupert (2004, 2006, in press-a, in press-b) may be claimed to capture all that can be of philosophical or scientific interest in such cases and to avoid some significant methodological dangers in the bargain. What positive value, it may be asked, flows from the adoption of the extended perspective? And isn't there a danger, in embracing such (often transient) larger wholes, of losing our practical and theoretical grip on the very minds—the minds of more or less stable individual agents persisting through time—that we hoped better to understand?

I shall argue, by contrast, that (in the relevant cases) it is the conservative view that threatens to obscure much that is of value and that a robust notion of cognitive extension thus earns its keep as part of the emerging picture of the active embodied mind. To make this case, I first