THE JOURNAL OF PHILOSOPHY
VOLUME CIV, NO. 4, APRIL 2007

CURING COGNITIVE HICCUPS: A DEFENSE OF THE EXTENDED MIND*

Human cognitive processing, according to the *Extended Mind Hypothesis*[1] may at times extend into the environment surrounding the organism.[2] Such a view should be contrasted with a nearby, (but much 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 may be claimed to capture all that can be of philosophical or scientific interest in such cases, and to avoid some significant dangers into the bargain. 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 sketch some quite general responses to the worries that motivate the more conservative view. I then present some new examples

* Thanks to Robert Rupert, Robert Wilson, Susan Hurley, Michael Wheeler, Richard Menary, Mark Rowlands, Dan Hutto, and all who participated in the "Extended Mind 2" conference at the University of Hertfordshire in 2006 for their comments and suggestions on an earlier draft of the present paper.

[1] This is introduced in Andy Clark and David Chalmers, "The Extended Mind," *Analysis*, LVIII, 1 (1998): 7–19.

[2] Other treatments that encompass notions of cognitive extension include Robert Wilson, *Boundaries of the Mind: The Individual in the Fragile Sciences—Cognition* (New York: Cambridge, 2004); John Haugeland, "Mind Embodied and Embedded," in Haugeland, *Having Thought: Essays in the Metaphysics of Mind* (Cambridge: Harvard, 1998), pp. 207–40; Daniel Dennett, *Kinds of Minds* (New York: Basic Books, 1996); Susan Hurley, *Consciousness in Action* (Cambridge: Harvard, 1998); Mark Rowlands, *Externalism: Putting Mind and World Back Together Again* (Montreal: Acumen/McGill-Queen's, 2003); Michael Wheeler, *Reconstructing the Cognitive World* (Cambridge: MIT, 2005); and Richard Menary, *Cognitive Integration: Attacking the Bounds of Cognition* (New York: Palgrave Macmillan, 2007).

0022-362X/07/0404/163–92 © 2007 The Journal of Philosophy, Inc.

and arguments that aim to flesh out the skeleton responses and to illuminate further the nature and importance of cognitive extension itself.

I. SPREADING THE MIND

As neuroimaging studies reveal more and more about the role of the brain in a wide variety of cognitive acts and operations, the temptation grows to locate *all* the physical machinery of mind firmly in the head and central nervous system. Yet we should not conclude, simply from the undoubted importance of such neural unfoldings in all our cognitive endeavors, that the neural machine and the cognitive machine are always and everywhere simply one and the same. Such views seem challenged, for example, by emerging work in "embodied cognitive science"[3] that depicts the nonneural body and the nonbodily world as each capable of making key cognitive contributions. According to the more radical versions of these views,[4] the actual local operations that realize certain forms of human cognizing include inextricable tangles of feedback, feed-forward, and feed-around loops: loops that promiscuously criss-cross the boundaries of brain, body, and world. The local mechanisms of mind, if this is correct, are not all in the head. Cognition leaks out into body and world.

That may sound like a strange idea. But it is hardly stranger, on reflection, than the commonplace idea that the activity of *brain-meat* realizes all that matters about human cognition. For to embrace the extended mind is in no way to question the basic materialist vision of mind as emerging fully and without residue from physical goings-on. Any added strangeness flows merely from the fact that some of the relevant goings-on, if an extended mind story is correct, do not stay neatly in the brain. They do not even stay neatly within the biological body. On the contrary, they prove perfectly (and productively) able to span brain, body, and world.

[3] A representative sample might include Esther Thelen and Linda Smith, *A Dynamic Systems Approach to the Development of Cognition and Action* (Cambridge: MIT, 1994); Thelen, G. Schöner, C. Scheier, and Smith, "The Dynamics of Embodiment: A Field Theory of Infant Perseverative Reaching," *Behavioral and Brain Sciences*, XXIV (2001): 1–86; Hurley, *Consciousness in Action*; Andy Clark, *Being There: Putting Brain, Body and World Together Again* (Cambridge: MIT, 1997); Rowlands, *Body Language: Representing in Action* (Cambridge: MIT, 2006); Wilson, *Boundaries of the Mind*; Michael Wheeler, *Reconstructing the Cognitive World* (Cambridge: MIT, 2005); Alva Noë, *Action in Perception* (Cambridge: MIT, 2004); Lawrence Shapiro, *The Mind Incarnate* (Cambridge: MIT, 2004); Rolf Pfeifer and Josh Bongard, *How the Body Shapes the Way We Think* (Cambridge: MIT, 2007).

[4] For clear examples of the radical tendency (sometimes called "vehicle externalism"), see Hurley; Rowlands; Wilson; and Clark and Chalmers.

But might there be compelling reasons to delimit the realm of the *truly cognitive* in the way more standard approaches suggest? Robert Rupert[5] makes just such a claim. He distinguishes two projects, which he sees as competing proposals for understanding the cognitive processing of embodied and environmentally situated agents. The first (the Extended Mind approach) depicts human cognitive processing as sometimes quite literally extending into the extra-organismic environment. Rupert dubs this the *Hypothesis of Extended Cognition* (HEC) and glosses it like this:

> (HEC)
> According to this view…human cognitive processing literally extends into the environment surrounding the organism, and human cognitive states literally comprise—as wholes do their proper parts—elements in that environment (*ibid.*, p. 389).

Rupert depicts HEC as a radical hypothesis apt (if true) to transform cognitive scientific theory and practice and to impact our conceptions of agency and persons. But it needs to be assessed, Rupert argues, alongside that more conservative (though still interesting and important) competitor. This is the perspective he dubs *Hypothesis of Embedded Cognition* (HEMC) according to which:

> (HEMC)
> Cognitive processes depend very heavily, in hitherto unexpected ways, on organismically external props and devices and on the structure of the external environment in which cognition takes place (*ibid.*, p. 393).

The argument starts from a simple appeal to common sense. Common sense, Rupert suggests, rebels at the vision of extended cognition[6] so we need sound theoretical reasons to endorse it. HEMC, by contrast, is said to be much more compatible with common sense. Two main worries are then raised for HEC.

The first worry, similar to one previously raised by other critics such as Fred Adams and Kenneth Aizawa,[7] concerns the profound differences that appear to distinguish the inner and outer contributions. Thus Rupert suggests that "the external portions of extended

[5] Rupert, "Challenges to the Hypothesis of Extended Cognition," this JOURNAL, CI, 8 (August 2004): 389–428.

[6] Though I do not take issue with this claim here, it is by no means obvious that common sense itself is committed either to in-the-head or in-the-organism cognition.

[7] See Adams and Aizawa, "The Bounds of Cognition," *Philosophical Psychology*, XIV, 1 (2001): 43–64. For some responses, see Clark, "Intrinsic Content, Active Memory, and the Extended Mind," *Analysis*, LXV, 1 (January 2005): 1–11; Clark, "Memento's Revenge: The Extended Mind, Re-visited," to appear in Menary, ed., *The Extended Mind* (Aldershot, UK: Ashgate, forthcoming).

'memory' states (processes) differ so greatly from internal memories (the process of remembering) that they should be treated as distinct kinds" (*op. cit.*, p. 407). Given these differences, there is said to be no immediate pressure to conceive the internal and the external contribution in the same terms. But worse still, there is (allegedly) a significant cost.

The second worry thus concerns the apparent scientific cost[8] of any wholesale endorsement of HEC. For taking all kinds of external props and aids as proper parts of human cognitive processing robs us, he fears, of the standard object of cognitive scientific theorizing, namely, the stable persisting human individual. Even in cases of developmental theorizing, where what is at issue is not so much stability as change, Rupert argues, one still needs to find some persisting, though developing, thinking core. Failure to do so, Rupert fears, will cause cognitive science and cognitive psychology to lose their grip on the very human subjects they aim to study. The price of HEC, if this were so, is nothing less than the loss of much (perhaps all) the progress that cognitive psychology has made thus far. The sciences of the mind, it thus seems, simply cannot afford to identify human cognitive processing with the activity of various "short-lived coupled systems" comprising neural, bodily, and worldly[9] elements.

II. RESPONDING TO THE CHALLENGES

These are important challenges. Nonetheless, the worries are misplaced. First, none of the arguments for extended cognition turn on, or otherwise require, the *fine-grained* similarity of the inner and outer contributions. Second, HEC need not accrue the prohibitive costs that Rupert fears. This section displays the basic form of those replies.

Concerning the (lack of) similarity of the inner and outer contributions, part of the problem stems from a persistent misreading of the so-called "Parity Principle" originally introduced by Andy Clark and David Chalmers. This was the claim that "if, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process."[10] But far from requiring or suggesting the

[8] This worry appears briefly in Rupert and is pursued by Rupert at greater length in "Innateness and the Situated Mind," to appear in Philip Robbins and Murat Aydede, eds., *The Cambridge Handbook of Situated Cognition* (New York: Cambridge, forthcoming).

[9] This is a loose though familiar usage. "Bodily" here means gross-bodily, that is, extra-neural, while "worldly" means "extra-bodily." Clearly, brains are nonetheless parts of bodies, and bodies elements of the world.

[10] See Clark and Chalmers, p. 8.

need for any fine-grained similarity between inner and outer processes, the parity claim was specifically meant to *undermine* any tendency to think that the shape of the (present day, human) inner processes sets some bar on what ought to count as part of a genuinely cognitive process. The parity probe was thus meant to act as a kind of *veil of metabolic ignorance*, inviting us to ask what our attitude would be if currently external means of storage and transformation were, contrary to the presumed facts, found in biology. Thus understood, parity is not about the outer performing just like the (human-specific) inner. Rather, it is about equality of opportunity. It is about avoiding a rush to judgment based on spatial location alone. The parity principle was thus meant to engage our rough sense (based on something like coarse functional role) of what we might intuitively judge to belong to the domain of cognition—rather than, say, that of digestion—but to do so without the pervasive distractions of skin and skull.

This point is nicely recognized by Michael Wheeler[11] who notes that the *wrong* way to assess parity of contribution is:

> [to] fix the benchmarks for what it is to count as a proper part of a cognitive system by identifying all the details of the causal contribution made by (say) the brain [then by looking] to see if any external elements meet those benchmarks (*ibid.*, ms. p. 3).

To do things that way, Wheeler argues, is to open the door to the implausible and highly chauvinistic thought that only systems whose fine-grained causal profile fully matches that of the human brain can be cognitive systems at all. Yet just because some alien neural system failed to match our own biological memory-systems in various ways (perhaps, to borrow an example from Rupert (*op. cit.*) they fail to exhibit the so-called "generation effect" during recall) we should surely not *thereby* be forced to count the action of such systems as noncognitive, or as not an instance of memory at all. The Parity Principle is thus best seen as a demand that we assess the bio-external contributions with the same kind of unbiased vision that even the staunch internalist ought to bring to bear on an alien neural organization. It is wholly misconstrued as a demand for fine-grained sameness of processing and storage. Rather, it is a call for sameness of opportunity, such that bio-external elements *might* turn out to be parts of the machinery of cognition *even if* their contributions are

[11] See Wheeler, "Minds, Things and Materiality," in C. Renfrew and L. Malafouris, eds., *The Cognitive Life of Things* (Cambridge, UK: McDonald Institute for Archaeological Research, forthcoming).

unlike (perhaps because they are deeply complementary to) those of the biological brain.

But even once we lay to rest the mistaken reading of the Parity Principle (as requiring fine-grained identity of causal contribution) there remains an important and closely related question. The question turns on the issue of natural or explanatory kinds. Thus Rupert (*op. cit.*) questions the idea, certainly present in Clark and Chalmers's original treatment, that treating the organism-notebook system as the supervenience base for some of Otto's dispositional beliefs was to be recommended on grounds of *explanatory unity and power*. Rupert's worry then takes as a premise the idea that a kind is natural if it is adverted to by the laws or explanations of a successful science. Biomemory thus meets the requirement as it falls under the laws and explanatory frameworks of a successful science—cognitive psychology or cognitive science more generally. But (the argument continues) "extended memory" does not fit the causal profile of memory as described by this body of successful science and hence should not be subsumed under the heading of "memory" at all.

It is by no means obvious, however, that all acceptable forms of unification require that all systemic elements behave according to the same laws. Indeed, to assume they must do so is simply to beg the question against any science whose target is a genuinely hybrid system (for example, a part-connectionist, part-classical computational organization). In such cases one may, of course, hope to find *additional* principles governing the larger (hybrid) organization itself. At this point it is surely worth remembering that the study of extended cognitive systems is just beginning, and it is no wonder that our best current unified understandings target the inner elements alone: that is where science has so far (mostly) been looking, after all. Nonetheless, it remains the substantive empirical bet of the extended systems theorist that larger hybrid wholes, comprising biological and nonbiological elements, will *also* (and more on this below) prove to be the proper objects of sustained scientific study in their own right.

A further reason to resist the easy assimilation of HEC into HEMC concerns the nature of the interactions between the internal and the external resources themselves. Such interactions, it is important to notice, may be highly complex, nested, and nonlinear. As a result there may, in some cases, be no viable means of understanding the behavior and potential of the extended cognitive ensembles by piecemeal decomposition and additive reassembly. To understand the integrated operation of the extended thinking system created, for example, by combining pen, paper, graphics programs, and a trained mathematical brain, it may be quite insufficient to attempt to understand and then combine (!) the properties of pens, graphics programs, paper, and brains. This may be insufficient for just the same kinds of reason that, within neuroscience itself, invite us to study not *just* the various major neural substructures and their capacities, but *also* their complex nonlinear interactions and the larger-scale activities in which they participate. Here the larger explanatory targets are whole processing cycles, running on temporary (they are sometimes dubbed "soft-assembled") coalitions of neural resources, arising in response to some specific problem-solving purpose. Such soft-assembled neural packages involve the temporally evolving, often highly re-entrant, activity of multiple populations of neurons spanning a variety of brain areas. But why then suppose that the soft-assemblies most relevant to human cognitive achievements are always and everywhere bounded by skin and skull? Why should we not recognize, in our peculiarly structured and artifact-rich world, a succession of similarly complex hybrid ensembles spanning brain, body, and world?

What, finally, of the allegedly high scientific *costs* of such an enlarged perspective? Here it is important to see that there is no need, in taking extended cognition seriously, to lose our grip on the more-or-less stable, more-or-less persisting biological bundle that lies at the heart of each episode of cognitive soft-assembly. Occasionally, under strict and rare conditions, we may confront genuine extensions of that more-or-less persisting core: cases where the persisting, mobile resource bundle is augmented in a potentially permanent manner. But in most other cases, we confront only temporary medleys of information-processing resources comprising an integrated subset of neural activity and bodily and environmental augmentations. The mere fact that such circuits are temporary, however, does not provide sufficient reason to downgrade their cognitive importance. Many purely internal information-processing ensembles are likewise transient creations, generated on-the-spot[12] in response to the particu-

[12] As just one example, consider the account (D.C. Van Essen, C.H. Anderson, and B.A. Olshausen, "Dynamic Routing Strategies in Sensory, Motor, and Cognitive Processing," in Christopher Koch and Joel Davis, eds., *Large Scale Neuronal Theories of the Brain* (Cambridge: MIT, 1994), pp. 271–99) according to which many neurons and neuronal populations serve not as direct encodings of knowledge or information, but as (dumb) middle managers routing and trafficking the internal flow of information between and within cortical areas. These "control neurons" serve to open and close channels of activity, and allow for the creation of a kind of instantaneous, context-sensitive modular cortical architecture. Related proposals include A.R. Damasio and H. Damasio's notion of "convergence zones," which are neuronal populations which likewise initiate and co-ordinate activity in multiple neuronal groups: see Damasio and Damasio, "Cortical Systems for Retrieval of Concrete Knowledge: The Convergence Zone Framework," in Koch and Davis, eds., pp. 61–74.

larities of task and context. As Jerry Fodor once put it, in such cases it is "*unstable instantaneous* connectivity that counts."[13] The resulting soft-wired ensembles, in which information flows and is processed in ways apt to the task at hand, do not cease to be important just because they are transient.

Rupert worries that, by taking seriously the notion of cognitive extension in that special subclass of transient cases where the new organizations span brain, body, and world, we lose our grip on the persisting systems that are meant to be our objects of study. Certainly, much work in cognitive and experimental proceeds by assuming that subjects are "persisting, organismically bound cognitive systems."[14] But the correct response here, it seems to me, is to let a thousand flowers bloom. If our avowed goal is to discover the properties of the neural apparatus alone, we might want to impede subjects from using their fingers as counting buffers during an experiment. Similarly, if our goal is to understand what the persisting biological organism alone can do, we might want to restrict the use of all nonbiological props and aids. But if our goal is to unravel the mechanically modulated flow of energy and information that allows an identifiable agent (a Tom, Dick, Jane, or Janet) to solve a certain kind of problem, we should not simply assume that every biologically-motivated surface or barrier forms a cognitively relevant barrier, or that it constitutes an important interface[15] from an information-processing perspective. That this can be done while still respecting experimental requirements is shown, for example, by the careful studies reviewed in section III following.

Notice also that we do not *find* individual agents by first finding their cognitive mechanisms. Instead, we find an agent by finding (roughly speaking) a reliable, easily identifiable physical nexus of perception and action, apparently driven by a persisting and modestly integrated body of goals and knowledge. Then (and only then) do we ask, of some particular problem-solving performance displayed by *that very agent*, what (and where) are the mechanisms that make possible that performance? It is at that point that we may (sometimes) be surprised to find that the target performance depends upon a far wider variety of factors and forces[16] than we initially imagined. In so

[13] In Fodor, *The Modularity of Mind* (Cambridge: MIT, 1983), p. 118. See also Fodor, *The Mind Doesn't Work That Way* (Cambridge: MIT, 2001).

[14] See Rupert, "Innateness and the Situated Mind," to appear in Robbins and Aydede, eds. (forthcoming).

[15] For more on the questions concerning interfaces, see Haugeland.

[16] The form of reasoning is thus parallel to that which leads Richard Dawkins to describe the web as part of the "extended phenotype" of the spider (see Dawkins,

doing we retain a perfectly good grip on the cognitive agents that are our primary objects of study.

It may also be helpful to distinguish two possible explanatory goals hereabouts. One is to explain the persistingness of specific cognitive agents. The other is to display the machinery that underpins an agent's current mental state or that explains some specific cognitive performance. Thus visual cortex, as Chalmers notes (personal communication), may be quite irrelevant to my persistence as a subject (I would persist without it) while still being the supervenience base for some of my current mental states and performances. What is really at issue, as far as the claims about cognitive extension are concerned, is which bits of the world make true (by serving as the local mechanistic supervenience base for) certain claims about a subject's here-and-now mental states or cognitive processing.

III. COGNITIVE IMPARTIALITY

Let us assume (what is surely uncontroversial) that the biological brain is, currently at least, the essential core element in all episodes of individual human cognitive activity. A question we may then ask is: Does the brain "care about" the nature (biological or nonbiological) or location (organism-bound or organism-external) of the processing and storage resources soft-assembled to tackle some cognitive task?

In an important series of experiments, Wayne Gray and his colleagues have shown (in compelling detail) that it is a mistake to privilege any location or any type of operation in the online soft-assembly of a cognitive routine. In the first set of such experiments[17] subjects were required to program an on-screen simulation of a VCR control panel. The idea was to manipulate the time-costs of accessing the information (concerning channel, start-time, and so on) needed to program the VCR. This information was presented in a window beneath the control panel and was either constantly visible by a flick

---

The Extended Phenotype (New York: Oxford, 1982)) and that leads J. Scott Turner to treat the sound amplifying ("singing") burrows of the Mole Cricket as external physiological organs (see his book, *The Extended Organism: The Physiology of Animal-Built Structures* (Cambridge: Harvard, 2000)). In each case we start with a working sense of some baseline concept (phenotype, organ). We then notice that stuff that we do not ordinarily treat in those terms is playing the right kind of role to be considered as belonging to that class. This is not about the new stuff working just like the old. There is no organ much like a burrow, no animal body much like a web. Nor does it require equal stability and permanence: the web comes and goes in a way the spider-body does not, and singing burrows, unlike inner organs, may be built, destroyed and rebuilt in new locations.

[17] See W.D. Gray and W.t.- Fu, "Soft Constraints in Interactive Behavior," *Cognitive Science*, XXVIII, 3 (2004): 359–82.

of the eyes ("free access group") or available only by moving and clicking the mouse to remove an overlaid opaque cover ("gray box group"). There was also a "memory test" group (run under both the above conditions such as, free access and gray box) who, unlike the others, had previously memorized all the information required.

What the researchers found was that time-costs of information retrieval, measured in milliseconds, appeared to determine the precise mix of resources (bio-memory, motor actions, shifts of attention) recruited to solve the problem. That is, the subjects rapidly settled on whatever strategy yielded (at that phase of the programming) the least-cost (measured by time) information retrieval. In fact, they did this even when the fastest mix of resources sacrificed perfect knowledge in the world for imperfect[18] knowledge in the head. Only when the in-the-world data could be accessed less effortfully (measured by time) than the data stored in biological memory, was it recruited and calls to the external store "built-in" into the dominant strategy.

Gray and Wai-Tat Fu present their results as a challenge to the idea that human cognitive strategies actively favor the use of information in the world over information in the head. The pendulum may have swung a little too far, they fear, in favor of reliance on external cognitive scaffolding. Instead, they argue that their results show that "the time spent retrieving something from memory is weighed the same as time spent in perceptual-motor activity" and that it is therefore a mistake to "presume the privileged status of any location or type of operation" (*ibid.*, pp. 378, 380). They thus argue for a level playing field with time-costs of access playing the key role in determining the mix of resources recruited as part of some cognitive routine. That is:

> The cognitive control of interactive behavior minimizes effort by using a least-effort [measured by time] combination of all the mechanisms available to it. All mechanisms or sub-systems are on the table. There is no reason to think that one mechanism or subsystem has a privileged status in relation to another(*ibid.*, p. 380).

Or as Gray and V.D. Veksler[19] subsequently express the idea:

> The central controller makes no functional distinction between knowledge in-the-head versus in-the-world or the means of acquiring that

[18] This is so in the case of all the non-memory-test subjects, who had to pick up the information piecemeal during the course of the experiment.

[19] Gray and Veksler, "The Acquisition and Asymmetric Transfer of Interactive Routines," in B.G. Bara, L. Barsalou, and M. Bucciarelli, eds., *27th Annual Meeting of the Cognitive Science Society, CogSci2005* (Cognitive Science Society, 2005), pp. 809–14.

information (such as eye movement, mouse movement and click, or retrieval from memory) (*op. cit.*, p. 809).

This model is described as a "soft constraints" account of interactive behavior. Temporal cost-benefit tradeoffs are said to provide a soft constraint (one that may always be over-ridden by various forms of explicit control) on the precise mix of motoric, perceptual, and bio-memory based resources that will, other things being equal, be automatically recruited to perform a given information-processing task on a given occasion. In subsequent work, Gray and colleagues[20] directly compare this to a Minimal Memory Model[21] according to which the resource recruitment process aims to *minimize* the use of bio-memory and to *maximize* the use of environmental support. They thus agree (with D.H. Ballard (*op. cit.*) and others) that the "embodiment level"[22] (the level at which we observe delicate, short time-scale interactions between motoric, perceptual, and bio-memory based resources) is crucial for much of our problem-solving activity, but differ in their account of exactly how the trade-offs are calculated.[23] Where Ballard and others predict a bias towards the use of external encoding and storage, Gray and colleagues depict a level playing field[24] with fine temporal considerations calling the tune: "Milliseconds matter, and they matter the same regardless of the type of activity with which they are filled."[25]

For current purposes, the resolution of this particular dispute is less important than the fact of its existence, and the susceptibility of the issue to systematic empirical investigation. For we have here a

[20] Gray, C.R. Sims, Fu, and M.J. Schoelles, "The Soft Constraints Hypothesis: A Rational Analysis Approach to Resource Allocation for Interactive Behavior," *Psychological Review*, CXIII, 3 (2006): 461–82.

[21] See D.H. Ballard, M.M. Hayhoe, and J.B. Pelz, "Memory Representations in Natural Tasks," *Journal of Cognitive Neuroscience*, VII, 1 (1995): 66–80. See also Hayhoe, "Vision Using Routines: A Functional Account of Vision," *Visual Cognition*, VII (2000): 1–3, 43–64.

[22] See Ballard et al., and see Ballard, Hayhoe, P. Pook, and R. Rao, "Deictic Codes for the Embodiment of Cognition," *Behavioral and Brain Sciences*, XX (1997): 723–67.

[23] This emerging empirical debate between the minimal memory account and the purely time-cost based alternative is evidence, it seems to me, of an important turning point in the study of the embodied mind. For in place of a loose coalition of ideas concerning the cognitive importance of body and world, we now begin to see the first stirrings of a science, complete with nuanced disagreements open to empirical investigation by broadly sympathetic practitioners.

[24] Gray and Veksler depict the level playing field as populated by a set of possible 'interactive routines' where these are "a complex mixture of elementary cognitive, perceptual, and action operations [that] represent basic patterns of interactive behavior" (p. 809).

[25] Gray and Fu, p. 364.

sequence of controlled experiments targeting genuinely hybrid ensembles: soft-assembled coalitions comprising bio-storage, motoric, and perceptual modes of access, and bio-external storage. The work by Gray and colleagues thus provides a clear demonstration of the susceptibility (despite the fears of Rupert, of Adams and Aizawa, and others) of such organizations to quite standard forms of cognitive scientific investigation. Even if Rupert and others are right that key terms such as memory cannot, once extended to the non-biological domain, serve to pick out fully explanatorily unified kinds, this does not mean that the extended organizations in which they participate are not proper objects of scientific inquiry, emerging and dissolving according to determinable principles, and operating in ways that maximize certain properties and features (in this case, speed of access). To the worry that there will be no unified science of heterogeneously constituted systems, we should reply that there not only can be, but already is, a nascent science both of the recruitment (of sets of neural and extra-neural resources) and of the fine-tuned unfolding of activity in just such heterogeneous ensembles.

Gray and colleagues sum up their own preferred model with two claims. The first is that "the control system is indifferent to information source"[26] The second (*ibid.,* p. 478) is that the only bias imposed by biology is that of finding the most cost effective mix of elements available. These very broad conclusions are compatible with a wide variety of possible cost-functions (time taken may not always be the prime or sole determinant). But whatever the cost-function or functions (which may turn out to vary with context and goals) what matters most, for our purposes, is the underlying vision of (what I shall dub) *cognitive impartiality*:

> (Hypothesis of Cognitive Impartiality)
> Our problem-solving performances take shape according to some cost-function or functions that, in the typical course of events, accords no special status or privilege to specific types of operation (motoric, perceptual, introspective) or modes of encoding (in-the-head or in-the-world).

This is, in many ways, a quite natural accompaniment to the Parity Principle itself. It states that the biological control system does not "care about" differences of location or type of resource, but simply uses whatever it can, relative to some cost-benefit trade-off, to get the job done.

[26] Gray, Sims, Fu, and Schoelles, p. 478.

IV. A BRAIN TEASER

Simple as it may sound, the hypothesis of Cognitive Impartiality hides something of a puzzle, at least for those who would depict cognition as (not just embodied but also) extended. For it threatens, unless delicately handled, to undermine the image of cognitive extension[27] in quite a delicious fashion. Thus suppose we now ask: "Just *what is it* that is so potently impartial concerning its sources of order and information?" The answer looks to be "the biological brain." So *have we not ended up firmly privileging the biological brain in the very act of affirming its own impartiality?*

To see past this worry, we must first notice that there are at least two explanatory targets in the immediate vicinity. The first is the recruitment of the extended organization itself. We may ask just how, and according to what principles, the various elements (perhaps some subset of neural operations, "deictic" uses of eye movements, gestures (see below), and scribblings) came to combine into a specific soft-assembled information processing device. In this process of soft-assembly, the brain surely plays a very special role. The second concerns the flow of information and processing in the new soft-assembled extended device. Relative to that device we may ask just how information flows and is processed in ways that (hopefully) solve some problem. HEC helps us to see that, as far as the second of these explanatory projects goes, the bounds of skin and skull are functionally transparent. HEMC, by contrast, both threatens to obscure the scientifically important distinction between the two projects and erects a firm skin-based boundary where the process of recruitment and use marks no boundary at all.

The puzzle concerning cognitive impartiality is thus resolved. Concerning the process of recruitment, it is indeed the biological brain (or perhaps some of its sub-systems) that is in the driving seat. That is to say, it is indeed some neurally-based process of recruitment that (following Gray and colleagues) turns out to be so pointedly unbiased regarding the use of inner versus outer circuits, storage, and operations. But once such an organization is in place,[28] it is the flow and transformation of information in (what is often) an extended, distributed system that provides the actual machinery of on-going thought and reason.

What this suggests is that in rejecting the vision of human cognitive processing as *organism-bound*, we should not feel forced to deny

[27] Thanks to Paul Schweizer for pointing this out.

[28] In most real-world settings, of course, these two stages (though logically distinct) may be fully or partially temporally overlapping.

that it is (in most, perhaps all, real-world cases) *organism-centered.* It is indeed primarily (though not solely) the biological organism that, courtesy especially of its potent neural apparatus, spins and maintains (or more minimally selects and exploits) the webs of additional structure that then form parts of the machinery that accomplishes its own cognizing.[29] Just as it is the spider-body that spins and maintains the web that then (following Richard Dawkins, *op. cit.*) constitutes part of its own extended phenotype, so it is the biological human organism that spins, selects, or maintains the webs of cognitive scaffolding that participate in the extended machinery of its own thought and reason.[30] Individual cognizing, then, is *organism-centered even if it is not organism-bound.*

V. THOUGHTFUL GESTURES

At this stage of our discussion, it will help to consider a worked example of extended cognizing in action. The example I shall take concerns the role of bodily gesture in thought and reason. The case is apt since gesture, though clearly itself an organismic activity, is not merely a *neural* activity. Moreover, bodily gesture turns out to exhibit some key features whose applicability looks to outrun the bounds of the organism itself.

After an extensive enquiry into the nature and organization of human gesture, the psychologist Susan Goldin-Meadow[31] asks an intriguing question. Is gesture simply about the expression of fully-formed thoughts, and thus mainly a prop for inter-agent communication (listeners appreciating meanings through others' gestures) or *might gesture function as part of the actual process of thinking?* Some clues (*ibid.*, pp. 136–49) that it might be more than merely expressive include:

that we do it when talking on the telephone
that we do it when talking to ourselves
that we do it in the dark when no one can see
that gesturing increases with task difficulty
that gesturing increases when speakers must choose between options
that gesturing increases when reasoning about a problem rather than merely describing the problem or a known solution.

[29] This is not to deny, of course, that much of the spinning is done by social groups of organisms spread out over long swathes of history.

[30] One difference is that in the case of the webs of cognitive scaffolding, it is often the human organism acting in concert with existing webs of scaffolding that spins, selects, or maintains new layers of scaffolding, resulting in the powerful process that Kim Sterelny dubs "incremental downstream epistemic engineering": see Sterelny, *Thought in a Hostile World: The Evolution of Human Cognition* (Malden, MA: Blackwell, 2003).

[31] Goldin-Meadow, *Hearing Gesture: How Our Hands Help Us Think* (Cambridge: Harvard, 2003).

A deflationist might suggest that most of these effects are easily explained by mere association: that gesturing without a viewer is just a habit, installed by our experience of gesturing in the normal communicative context. It turns out, however (*ibid.,* pp. 141–44) that speakers blind from birth, who have never spoken to a visible listener, and never seen others moving their hands as they speak, gesture when they speak. Moreover, they do so even when speaking to others they know to be blind.[32]

Supposing (for the sake of argument) that gesture does play some kind of active causal role in thinking, just what role might that be? One way to find out is to see what happens when actual physical gesture is removed from the mix of available resources. To explore the impact of restricting gesture on thought, Goldin-Meadow and colleagues[33] asked two matched groups of children to memorize a list and then to carry out some mathematical problem-solving before trying to recall the list. One group (call it the "free-gesture group") could freely gesture during the intervening math task, the other (call it the "no-gesture group") was told not to gesture. The results were that restricting the use of gesture during the intervening mathematical task had a robust and significant detrimental effect on the separate memory task (remembering the list of words). The best explanation, according to Goldin-Meadow, is that the act of gesturing somehow shifts or lightens aspects of the overall neural cognitive load, thus freeing up resources for the memory task.

Before pursuing this idea, it is necessary to rule out a rather obvious alternative account. According to this alternative account, the effort of remembering *not* to gesture (in the no-gesture group) is *adding* to the load, rather than gesture (in the free gesture group) lessening the load. If this were so, the no-gesture group would indeed perform less well, but not because gesturing lightens the load. Rather, remembering *not* to gesture increases it. As luck would have it, some children and adults spontaneously chose not to gesture during some of the episodes of mathematical problem-solving. This allowed the experimenters to compare the effects of removing gesture by instruction and by spontaneous (hence presumably effortless) inclination. Memory for the initial task turned out to be equally impaired even when

[32] See also J. Iverson and Goldin-Meadow, "Why People Gesture When They Speak," *Nature*, CCCXCVI (1998): 228; and Iverson and Goldin-Meadow, "The Resilience of Gesture in Talk," *Developmental Science*, IV (2001): 416–22.

[33] Goldin-Meadow, *Hearing Gesture*, chapter 11; and Goldin-Meadow, H. Nusbaum, S. Kelly, and S. Wagner, "Explaining Math: Gesturing Lightens the Load," *Psychological Science*, XII (2001): 516–22.

the lack of gesture was a spontaneous choice supporting the claim that the gestures themselves play some active cognitive role.[34]

An important hint as to the nature of this active role emerges, Goldin-Meadow argues, when we look at cases of gesture-speech mismatches.[35] These are cases where what you say and what you gesture[36] are in conflict, for example, you gesture a one-one mapping while failing to appreciate the importance of such a mapping in your simultaneous vocal attempts at solving the problem. Many such cases were found and (importantly) the gestures tended to prefigure the child's consciously finding the right solution in speech at a very slightly later point. Even if the right solution was not shortly found, the presence of the apt gesture turned out to be predictive of that child's being able to learn the right solution more easily than others whose gestures showed no such tacit or nascent appreciation.

After much conjecture and experiment, Goldin-Meadow is led to the following story (drawing also, as she clearly notes, on the groundbreaking work of David McNeill (*op. cit.*)). The physical act of gesturing, Goldin-Meadow suggests, plays an active (not merely expressive) role in learning, reasoning, and cognitive change by providing an alternative (analog, motoric, visuo-spatial) representational format. In this way:

> Gesture…expands the set of representational tools available to speakers and listeners. It can redundantly reflect information represented through verbal formats or it can augment that information, adding nuances possible only though visual or motor formats.[37]

Encodings in that visuo-motor format enter, it is argued, into a kind of ongoing coupled dialectic with encodings in the verbal format. Gesture thus continuously informs and alters verbal thinking which is continuously informed and altered by gesture, that is, the two form a genuinely coupled system. This coupled dialectic creates points of instability (conflict) whose attempted resolutions move forward our thinking, often (though of course not always) in productive ways. The upshot is, according to the developmental psychologists Iverson and

[34] See Goldin-Meadow, *Hearing Gesture*, chapter 11, pp. 155f. There is also rumored to be some older work in which children were simply told to sit on their hands, thus effectively removing the gestural option without adding to the memory load!

[35] Goldin-Meadow, *Hearing Gesture*, chapter 12.

[36] Much of Goldin-Meadow is devoted to the task of systematically attributing meaning to spontaneous free gestures. See also David McNeill, *Hand and Mind* (Chicago: University Press, 1992); and McNeill, *Gesture and Thought* (Chicago: University Press, 2005).

[37] Goldin-Meadow, *Hearing Gesture*, p. 186.

Thelen, is "a dynamic mutuality such that activity in any one component of the system can potentially entrain activity in any other."[38]

But is it (at least partly) the actual physical gestures that matter here, or do they merely reflect the transfer of load between two different neural stores? Does gesturing simply shift the burden from a neural verbal store to a neural visuo-spatial store? If so, then it should be harder to perform a separate spatial memory task when freely gesturing than when not. This was tested[39] by replacing the original word-recall task with a spatial one: that of recalling the location of dots on a grid. The results were unambiguous. The availability of gesture still helps (still yields improved performance on the memory task) even when the second task is itself a spatial one.

The act of gesturing, all this suggests, is not simply a motor act expressive of some fully neurally-realized process of thought. Instead, the physical act of gesturing is part and parcel of a coupled neural-bodily unfolding that is itself usefully seen as an organismically-extended process of thought. In gesture we plausibly confront a cognitive process whose implementation involves machinery that loops out beyond the purely neural realm. This kind of cognitively pregnant unfolding need not stop at the boundary of the biological organism. Something very similar may, as frequently remarked, occur when we are busy writing and thinking at the same time. It is not always that fully formed thoughts get committed to paper. Rather, the paper provides a medium in which, this time via some kind of coupled neural-scribbling-reading unfolding, we are enabled to explore ways of thinking that might otherwise be unavailable to us. Just such a coupled unfolding was eloquently evoked in a famous exchange between Richard Feynman and the historian Charles Weiner:

Weiner once remarked casually that [a batch of notes and sketches] represented "a record of [Feynman's] day-to-day work," and Feynman reacted sharply.

> "I actually did the work on the paper," he said.
> "Well," Weiner said, "the work was done in your head, but the record of it is still here."

[38] Iverson and Thelen, "Hand, Mouth and Brain," in Rafael Núñez and Walter J. Freeman, eds., *Reclaiming Cognition: The Primacy of Action, Intention and Emotion* (Bowling Green, OH: Imprint Academic, 1999), pp. 19–40, p. 37.

[39] Goldin-Meadow and Wagner, "How Our Hands Help Us Learn," *Trends in Cognitive Sciences*, IX, 5 (2004): 234–41; Wagner, Nusbaum, and Goldin-Meadow, "Probing the Mental Representation of Gesture: Is Handwaving Spatial?" *Journal of Memory and Language*, L (2004): 395–407.

> "No, it's not a *record*, not really. It's *working*. You have to work on paper and this is the paper. Okay?"[40]

Feynman is right. If, following McNeill and Goldin-Meadow, we allow that actual gestures (not simply their neural pre- or post-cursors) can form part of an individual's cognitive processing, there seems no principled reason to suddenly stop the spread the moment skin meets air.[41]

VI. MATERIAL CARRIERS

In his most recent work, McNeill[42] offers a clear expression of the view that the physical gestures are actually elements in the cognitive process itself. McNeill's work is grounded in extensive empirical case-studies of the use of gesture in free speech. The key idea, that McNeill uses to understand and organize these studies, is the notion (briefly mentioned above) of an ongoing imagery-language dialectic in which physical gesture acts as a "material carrier." The phrase "material carrier" is due to L.S. Vygotsky[43] and is meant to convey the idea of a physical materialization that has systematic cognitive effects. But once more, we should not be misled by the image of cognitive effects. For according to McNeill, "The concept [of a material carrier] implies that the gesture, *the actual motion of the gesture itself*, is a dimension of thinking."[44] Our free (spontaneous, nonconventional) gestures are not, McNeill argues, merely expressions of or representations of our fully achieved "inner" thoughts but are themselves "thinking in one of its many forms" (*ibid.,* p. 99). Notice that this is not to say that the gestures do not follow from, and lead to, specific forms of neural activity. They do, and McNeill has much to say about the neural systems preferentially involved in the generation and reception of spontaneous gesture (*ibid.*, chapters 7 and 8). Rather, it is to see the physical act of gesturing as *part of a unified thought-language-hand system whose coordinated activity has been selected or maintained for its specifically cognitive virtues.*

[40] The exchange is quoted in James Gleick, *Genius: The Life and Science of Richard Feynman* (New York: Random House, 1993), p. 409. Thanks to Galen Strawson for drawing this material to my attention.

[41] Adams and Aizawa depict the neural as the seat of all truly cognitive activity while Rupert (wisely I suspect) prefers the whole organism. But in so doing he puts a foot (or so I argue) on a slippery slope. For once we allow cognitive processes to become sufficiently hybrid (allowing functional cognitive wholes to be made up of parts as heterogeneous as arm and hand motions plus neural activity) there seems no good reason to stop at the skin.

[42] McNeill, *Gesture and Thought.*

[43] Vygotsky, *Thought and Language,* Alex Kozulin, trans. (Cambridge: MIT, 1986).

[44] McNeill, *Gesture and Thought*, p. 98.

There are important differences between McNeill's account and that of Goldin-Meadow, but they are united in seeing the physical gestures as genuine elements in the cognitive process. McNeill stresses the idea of "growth points," described as "the minimal unit of an imagery-language dialectic" (*ibid.,* p. 105). A growth point is a package of imagistic and linear propositional (linguistic) elements that together form a single idea (for example, both conveying the concept of an antagonistic force as a speaker describes some series of events). The points of productive conflict stressed by Goldin-Meadow are not growth points in this technical sense (see, for example, McNeill, *ibid.*, p. 137). But they are growth points in another, quite routine sense: they are collisions in meaning space, crucially mediated by gestural loops into the physical world, which are able to move our thinking along in productive ways. These differences in emphasis do not amount, as far as I can tell, to any deep inconsistency in the underlying models of the cognitive virtues of gesture. In each case, the loop into gesture creates a material structure that is available to both speaker and listener. And just as that material structure may have a systematic cognitive effect upon the listener, so too it may have a systematic cognitive effect on the speaker. Here, McNeill invokes an evolutionary hypothesis that he dubs "Mead's Loop."[45] The background to McNeill's suggestion is the discovery of so-called "mirror neurons." These are neurons, first discovered in the frontal lobes of macaques, that fire both when an animal performs some intentional action and when it sees another animal performing the same action.[46] McNeill's suggestion is that our own gestures activate mirror-neuron dominated neural resources so that:

> One's own gestures [activate] the part of the brain that responds to intentional actions, including gestures, by someone else, and thus treats one's own gesture as a social stimulus.[47]

Whether this is the correct evolutionary and mechanistic account is unimportant for present purposes. What matters is rather the guiding idea that by materializing imagistic thought in physical gesture we create a stable physical presence that may productively impact and

[45] After George Herbert Mead. See Charles W. Morris, ed., *Mind, Self, and Society* (Chicago: University Press, 1934).

[46] See, for example, G. Rizzolatti, L. Fogassi, and V. Gallese, "Neurophysiological Mechanisms Underlying the Understanding and Imitation of Action," *Nature Review: Neuroscience,* II (2001): 661–70.

[47] McNeill, *Gesture and Thought,* p. 250.

constrain the neural elements of thought and reason. The role of gesture, if this is correct, is closely akin to that of certain forms of self-directed, overt or covert speech or (looping outside the organismic shell) to certain forms of writing-for-thinking (see McNeill, *ibid.*, p. 99).

Shaun Gallagher, in a very rewarding recent discussion of gesture[48] and thought writes that:

> Even if we are not explicitly aware of our gestures, and even in circumstances where they contribute nothing to the communicative process, they may contribute implicitly to the shaping of our cognition (*ibid.*, p. 121).

Gallagher approaches the topic of gesture in the larger framework of his account of the "prenoetic" role of embodiment. This is a term of art that Gallagher uses to signify the role of the body in structuring mind and consciousness. The idea is that facts about the body, and about bodily orientation, and so on, set the scene for conscious acts of perception, memory, and judgment (the "noetic" factors) in various important ways. A prenoetic performance, we are told, is "one that helps to structure consciousness but that does not explicitly show itself in the contents of consciousness" (*ibid.,* p. 32). Thus, to take a very simple example, embodied agents perceive the world from a certain spatial perspective. That perspective shapes what is given to us explicitly in phenomenal experience, but it is not itself part of what we experience. Instead, it "shapes" or "structures" experience (for this example, see Gallagher, *ibid.*, p. 2–3). In this manner, Gallagher speaks of the role of gesture in "shaping" cognition and (following Maurice Merleau-Ponty's usage in describing the cognitive role of speech) in the "accomplishment of thought." Such locutions neatly (though only temporarily) sidestep the thorny issue of whether to see gesture as part of the *actual machinery* of thought and reason. In a footnote to the quoted passage Gallagher is less cautious, suggesting that:

> It may be…that certain aspects of what we call the mind just are in fact nothing other than what we tend to call expression, that is, occurrent linguistic practices ("internal speech"), gesture, and expressive movement (Gallagher, *ibid.*, p. 121).

Gesture, Gallagher here suspects, is both a means by which thought is "accomplished" *and* an aspect of mind—an aspect of the thinking[49] itself.

[48] Gallagher, *How the Body Shapes the Mind* (New York: Oxford, 2005).

[49] There is, unfortunately, substantial ambiguity in the notion of "thinking" invoked in many of these discussions of gesture, since it can sometimes mean (1) "verbal

What finally emerges from our consideration of the role of gesture in thought is that gestures form part of an integrated "language-thought-hand"[50] system, and that it is activity in this whole interlinked system that has been selected for its specifically cognitive virtues. Neural systems coordinate with, help produce, exploit, and can themselves be entrained by, those special-purpose bodily motions that constitute free gestures. In this way speech, as Jana M. Iverson and Esther Thelen noted (*op. cit.*), gesture and neural activity are able to form a single integrated system with clear problem-solving virtues not reducible to the virtues of any of its individual parts.

VII. LOOPS AS MECHANISMS

A single integrated system can have a variety of distinct parts whose contributions to some overall process are hugely different. Some of those parts, moreover, may be cognitive processes in their own right (that is to say, they remain cognitive processes even when considered in isolation from the others) and others not. Just so, a sequence of physical gestures alone could never implement a cognitive state. It is only in coordination with crucial forms of neural activity that the cognitive role of the gestures can emerge and be maintained. By contrast, some set of neural goings-on is often sufficient for the presence of some cognitive state or other. But this (genuine) asymmetry provides no reason to reject the notion that gestures form part of the machinery of cognition. To see this, we need only remind ourselves that the activity of a single neuron is (likewise) never sufficient for the existence of a cognitive state, yet that activity can, in the proper context, still form part of the machinery that implements a cognitive state or process.

It may (or may not) also be true that for any gesture-involving cognitive unfolding, there is a pure sequence of neural events such that *if* they were somehow held in place, or ushered into being without the loop through physical gesture, the cognitive states of the

---

thought" which is conceived, by Goldin-Meadow, as distinct from (though intertwined with) (2) the kinds of holistic, imagistic thinking specifically accomplished by gesture. Finally, there is (3) the overall cognitive state achieved by an agent who has engaged in some ongoing process involving both gestural and verbal elements. To say that gesturing is part of the process that constitutes thinking is thus to say both that it helps mediate and inform the verbal thinkings, and that in so doing it forms part of a larger integrated cognitive system.

[50] The use of "thought" here is misleading (see previous note). It reflects common usage rather than the actual model than McNeill and others develop. There is a similar ambiguity in the use of "language" since gesturing, on McNeill's account, is actually part of language. McNeill is aware of these infelicities but thinks the usage will do no harm: see "Terminological Tango" on page 21 of McNeill.

embodied agent would be the same. It does not follow from this that the gestures play only a causal role and do not help constitute the machinery of cognition. For the same may also be true of a sequence of neural states held together by some internal operation. Achieve that very sequence some other way and the chain of thoughts, let us assume, will come out the same. It does not follow[51] that the inner or the outer operations involved are thereby not (as things actually unfold) genuine aspects of the cognitive process. Thus Susan Hurley[52] usefully cautions against what she dubs "the 'causal-constitutive error' error" where this is:

> ...the error of objecting that externalist explanations give a constitutive role to external factors that are "merely causal" while assuming without independent argument or criteria that the causal/constitutive distinction coincides with some external/internal boundary. To avoid thus begging the question, we should not operate with prior assumptions about where to place the causal/constitutive boundary, but wait on the results of explanation (*ibid.*, ms. p. 5).

In trying to get a grip on these matters we are easily misled by various inessential features of many common cases where bio-external factors and forces impact thought and reason. Thus suppose the rhythmic pulse of rain on my Edinburgh window somehow helps the pace and sequencing of a flow of thoughts. Is the rain now part of my cognitive engine? No. It is merely the backdrop against which my cognizing takes shape. But this, I submit, is not because the rain is outside the bounds of skin and skull. Rather, it is because the rain is not part of (it is not even a side-effect or a "spandrel" within) any system either selected or maintained for the support of better cognizing. It is indeed *mere* (but as it happens helpful) backdrop. Compare this with a robot *designed* to use raindrop sounds to time and pace certain internal operations essential to some kinds of problem-solving. Such a robot would be vulnerable to (non-British) weather. But it is not clear (at least to me) that the whole drop-based timing mechanism is not usefully considered as one of the robot's cognitive routines. Consider finally the Self-Stimulating Spitting Robot. This is a robot that evolved to spit stored water at a plate on its own body for the same purpose, so as to use the auditory signal as a kind of "virtual wire"[53] to

[51] See Hurley, *Consciousness in Action*, chapter 8, for more sophisticated versions of this argument.

[52] Hurley, "The Varieties of Externalism," to appear in Aydede and Robbins, eds. (forthcoming).

[53] This notion can be found in Daniel Dennett's *Consciousness Explained* (Boston: Little Brown, 1991), p. 196.

time other key operations. Those self-maintained cognition-supporting signals are surely part of the cognitive mechanism itself. A neural clock or oscillator would count after all.

What these simple examples show is that (as we might expect) coupling alone is not enough. *Sometimes*, all coupling does is provide a channel allowing externally originating inputs to drive cognitive processing along. But in a wide range of the most interesting cases there is a crucially important complication. These are the cases where we confront a recognizably cognitive process, running in some agent, that creates outputs (speech, gesture, expressive movements, written words) that, recycled as inputs, drive the cognitive process along. In such cases, any intuitive ban on counting *inputs* as parts of *mechanisms* seems wrong. Instead, we confront something rather like the cognitive equivalent of a forced induction system. A familiar example is the turbo-driven automobile engine. The turbocharger uses exhaust flow from the engine to spin a turbine that spins an air pump that compresses the air flowing into the engine. The compression squeezes more air into each cylinder, allowing more fuel to be combined, leading to more powerful explosions (that drive the engine that creates the exhaust flow that powers the turbo). This self-stimulating automotive arrangement provides up to 40% more power on demand. The exhaust flow is an engine output in good standing that also serves as a reliable, self-generated input. There can surely be little doubt that the whole turbocharging cycle should count as part of the automobile's own overall power-generating mechanism! The same is true, I submit, in the case of gesture: gesture is both a systemic output and a self-generated input[54] that plays an important role in an extended neural-bodily cognitive economy. The wrong image here is that of a central reasoning engine that merely uses gesture to clothe or materialize pre-formed ideas. Instead, gesture (and overt or covert speech) emerge as interacting parts of a distributed cognitive engine, participating in cognitively potent self-stimulating loops whose activity is as much an *aspect* of our thinking as its *result*.

VIII. PARTICIPANT MACHINERY AND MORPHOLOGICAL COMPUTATION

The goal of this treatment has been to defend the claim that (non-neural) body and (extra-bodily) world can, and sometimes do, act as what might now be dubbed "participant machinery." Body and

[54] In the case of gesture the relation between the self-created inputs and other processing elements also looks to involve the full complexities of "continuous reciprocal causation" as discussed in my *Being There: Putting Brain, Body and World Together Again* (Cambridge: MIT, 1997), pp. 163–66).

world, if this is correct, sometimes form part of the very machinery by which mind and cognition are physically realized, and hence form part of the local material supervenience base for various mental states and processes. Since this has seemed to many quite an exotic claim, it may be worth ending with one more (just about maximally simple, but nonetheless revealing) illustration.

Chandana Paul describes a toy example designed to demonstrate "that a robot body can be used for computation in addition to merely acting as an effector for the controller."[55] The backdrop to the demonstration involves a very simple class of neural networks known as perceptrons.[56] It is well known that a perceptron, if given two inputs A and B, can compute OR and AND functions (in fact, all linearly separable functions) but not linearly inseparable ones such as exclusive-or. Exclusive-or, normally written XOR, is true if either *but not both* disjuncts are true: that is, if (A or B) is true but – (A and B). Paul's demonstration involves a simple "vehicle," of the kind made famous by Valentino Braitenberg,[57] whose behavior is determined by the activity of two perceptron networks, each with inputs A and B. Perceptron 1 computes OR and controls M1, a forward drive delivered to the single central front wheel of a front-wheel drive vehicle. This means that power is delivered to the single central front wheel if either or both inputs are active (it is thus computing the standard INCLUSIVE OR function). Perceptron 2 computes the standard AND function and controls M2, a lifting device that will raise the single front wheel of the forward drive vehicle off the ground if and only if *both* inputs are active. You can probably see where this is going. When A and B are both reading OFF (zero, false), both nets output zero, the wheel is on the ground, but no power is delivered so the robot stays stationary. When only A is ON, the AND net delivers zero, the wheel stays grounded, and the OR net outputs a one. The wheel turns and the robot goes forward. The same type of scenario occurs when only B is ON. But (the crucial case) when A and B are both ON the OR net causes M1 to move but the AND net lifts the wheel from the ground so the robot stays stationary. The embodied systems response profile to

[55] Paul, "Morphology and Computation," in Stefan Schaal, Auke Jan Ijspeert, Aude Billard, Sethu Vijayakumar, John Hallam, and Jean-Arcady Meyer, eds., *From Animals to Animats, Proceedings of the 8th International Conference on the Simulation of Adaptive Behaviour, Los Angeles, CA, USA, 2004* (Cambridge: MIT, 2004), pp. 33–38, quoted passage p. 33. An expanded version of this paper appears as "Morphological Computation: A Basis for the Analysis of Morphology and Control Requirements," in *Robotics and Autonomous Systems*, LIV (2006): 619–30.

[56] Frank Rosenblatt, *Principles of Neurodynamics* (New York: Spartan Books, 1962).

[57] Braitenberg, *Vehicles: Experiments in Synthetic Psychology* (Cambridge: MIT, 1984).

the different possible values of the A and B inputs thus has the form of the standard XOR truth table, despite the fact that the visible computational controllers are perceptrons, congenitally unable to compute nonlinearly separable functions such as XOR. Lifting the front wheel in response to the conjunction of the two inputs now stands in for the "missing line" of the XOR truth table. The physical vehicle, despite having only perceptrons for controllers, is thus able to behave exactly as if it were controlled by an XOR net, because the active body of the robot is providing the functional equivalent of the missing second layer of neural processing. The overall embodied system thus provides the missing functionality, equivalent to performing a NOT on the first input, followed by an AND. In this way, Paul concludes "the example shows that through its configuration a robot body can perform a quantifiable operation on its inputs."[58]

At this point a skeptic might argue that the XOR computation is in some way unreal: more in the eye of the observer than a true resource for a reasoning robot? And there is (as things stand) some truth in this. For what the robot currently displays is what Paul nicely dubs "latent morphological computation": computation that is visibly (to us) implicit in the response profile of the overall physical device, but not yet available *to the device itself* as a general-purpose problem-solving resource.

A simple (and, as we shall see, biologically unexceptional) tweak, however, makes the new functionality available to the device itself. Thus Paul next describes a "vacuum cleaning robot" (the precise details of which need not concern us here). The vacuum cleaning robot is like the XOR robot except that this time it is augmented with a sensor informing it of the behavioral consequences of its own action. Thus augmented, the robot can learn (or be programmed) so as to incorporate the body-involving XOR circuit into an open-ended set of other routines, routing various A, B signals through the body circuit and reading the XOR result off from a rapid, self-perceived bodily "twitch" of the front wheel: a twitch that need not even persist long enough to cause actual forward motion. The body-involving XOR computation (that may previously have appeared to be merely in the eye of the beholder) is thus scaled up to a general-purpose resource that can be invoked much like a regular logic gate. Quite generally then:

> When a robot with latent morphological computation is augmented with a sensor which can sense the behavioral consequences, it makes the computational function defined by the morphology explicit, such that it

[58] Paul, "Morphology and Computation," p. 34.

> can be used as a standard computational sub-unit at any stage of the processing (Paul, *op. cit.*, p. 36).

It might seem, nonetheless, that this is still just a rather clumsy trick: Why use the robot body to perform a computation that could be so cheaply and easily handled using a simple three layer neural network? This is, however, to miss the true point and force of the demonstration. For the idea is that evolved biological intelligences, unlike the more neatly engineered solutions with which we are still most familiar as designers, are perfectly able to find and exploit unexpected forms of *multiple functionality*. That is to say, they may find and exploit solutions in which a single element (such as a bodily routine or motion) plays many roles, some of them merely practical, others more "epistemic"[59] in nature. The clean division between mechanical (body) design and controller design that characterizes many humanly engineered solutions looks quite unimportant (indeed, often counterproductive) if what we seek is efficiency and maximal exploitation of resources. Paul's demonstration may be compared to the work of Adrian Thompson and colleagues[60] using genetic algorithms to artificially evolve real electronic circuits. The evolved circuits turned out to exploit all manner of physical properties, usually ignored or deliberately suppressed by human engineers.[61] The lesson, according to the authors, was that:

> It can be expected that all of the detailed physics of the hardware will be brought to bear on the problem at hand: time delays, parasitic capacitances, cross-talk, meta-stability constraints and other low-level characteristics might all be used in generating the evolved behavior (Thompson et al., *op. cit.*, p. 21).

What thus goes for the brain (the hardware chip) goes too for the rest of the physical body. It too may be exploited, in all manner of unexpected ways, as an essential part of an information-processing organization. In real-world cases, Paul goes on to suggest, we should expect to find that the computational roles being played by bodily acts are much more complex than the computation of a common binary function, perhaps involving analog functions of quite unex-

[59] For more on this, see David Kirsh and Paul Maglio, "On Distinguishing Epistemic from Pragmatic Action," *Cognitive Science*, XVIII (1994): 513–49.

[60] A. Thompson, I. Harvey, and P. Husbands, "Unconstrained Evolution and Hard Consequences," in Eduardo Sanchez and Marco Tomassini, eds., *Towards Evolvable Hardware* (Boston: Springer, 1996), pp. 136–65.

[61] For further discussion, see chapter 5 of my *Mindware: An Introduction to the Philosophy of Cognitive Science* (New York: Oxford, 2001).

pected degrees of complexity. The case of gesture-for-thought (section VI above) may be an example of just this kind, in which actual hand and arm motions look to implement encoding and processing operations that are (as McNeill suggests) holistic and analog rather than local, symbolic, and discrete.

Once we start to question our simplistic armchair visions of the division of labor between brain, body, and world it rapidly becomes clear that there is no barrier to the realization of key functional and/or computational organizations by complex mixtures of neural, bodily, and environmental structures. The simple robotic illustration also provides further evidence of the power and importance of cognitive self-stimulation (section VII above). For the step from latent to explicit morphological computation depends essentially upon the agent's ability to sense its own bodily states. As embodied agents replete both with systems for efferent copy and for sensing what our own bodies are doing, we are ideally placed to profit from our own bodily actions, and to exploit our own bodily acts for cognitive and computational ends. Daily embodied activity may thus be playing many subtle, yet-to-be-understood cognitive roles.[62]

IX. WHY THE HEC?

This paper began with a double challenge. Show us that HEC (Hypothesis of Extended Cognition) has not priced itself out of the market by depicting us as cognitively extended agents at the cost of identifying persisting subjects for scientific study, and show us that there is real added value in adopting the perspective of HEC rather

[62] To take just one concrete example, there is a growing body of work on the possible role of eye movements in thought, reason, discourse comprehension, and recall. Such work includes Ballard and colleagues's work on deictic (fixation-based) pointers in block-copying (Ballard, Hayhoe, Pook, and Rao (*op. cit.*)), D.C. Richardson and M.J. Spivey's account of the cognitive role of eye movements in recall (Richardson and Spivey, "Representation, Space and Hollywood Squares: Looking at Things That Aren't There Anymore," *Cognition,* LXXVI (2000): 269–95), Richardson and N.Z. Kirkham's exploration of the role of eye movements (in 6 month olds) as spatially indexing auditory information (Richardson and Kirkham, "Multi-modal Events and Moving Locations: Eye Movements of Adults and 6-Month-Olds Reveal Dynamic Spatial Indexing," *Journal of Experimental Psychology: General,* CXXXI, 1 (2004): 46–66), and Richardson and R. Dale's model of the role of coupling between speaker and listener's eye-movements in discourse comprehension (Richardson and Dale, "Looking to Understand: The Coupling between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension," *Proceedings of the Twenty-sixth Annual Meeting of the Cognitive Science Society* (Mawhah, NJ: Erlbaum, 2004), pp. 1143–48. For a useful review, see Spivey, Richardson, and S. Fitneva, "Thinking Outside the Brain: Spatial Indices to Linguistic and Visual Information," in J. Henderson and F. Ferreira, eds., *The Interface of Vision Language and Action* (New York: Psychology Press, forthcoming).

than its more innocuous-seeming cousin HEMC (Hypothesis of Embedded Cognition).

Both challenges have now been met. Concerning the first, we have arrived at a vision of human cognition as organism-centered but not organism-bound. Embracing HEC does not require us to abandon the vision of a persisting biological (and within the biological, a neural) core that is a perfectly proper object of cognitive scientific study. HEC simply asserts that we should also study larger, often temporary, ensembles as units of cognitive activity in their own right.

At that point the second challenge becomes pressing. What is the added value accruing to the choice of HEC over HEMC? Both sides should, I think, concede the in-principle availability of alternative ways of carving the cognitive cake. As with any science, the placement of systemic boundaries reflects both general considerations and specific explanatory ambitions. The added value (for some purposes) of the HEC-style parse emerges, or so I have tried to show, once we recognize a few less familiar facts concerning human cognitive organization. It emerges when we recognize that the brain/CNS is "cognitively impartial": it does not care how and where key operations are performed. And it emerges when we recognize that much human cognizing benefits quite profoundly from cycles of self-stimulating activity ("cognitive turbo-drives") in which we actively create the structures that drive and constrain our own evolving processes of thought and reason.

It seems likely that some opposition to the idea of extended cognitive systems is rooted in the imagined availability of a simple alternative model in which a skull-bound intelligent agent decides to offload[63] certain bits of work and storage onto bodily and environmental structures. In the cases at issue, however, there is no such act of conscious offloading and reloading to be found. We do not, for example, consciously choose to gesture so as to lighten or alter the neuro-cognitive load. In such cases the extended process involves complex *sub-personally integrated* routines that are (typically) either selected or maintained for their peculiarly cognitive virtues. Nor need we imagine, in cases where conscious choice and orchestration is missing, that some highly intelligent, well-informed, though as-it-happens non-conscious, inner executive has made the choice for us.

[63] For example, we decide to place that yellow sticky on the mirror reminding us of the important meeting in the morning. In such cases an identifiable, fully-contained thinking agent offloads onto some environmental structure the semantically well-formed product of some recognizably cognitive activity, only to later re-load it as needed to perform some task.

Instead the "choice" consists simply in the emergence of an effective distributed problem-solving ensemble, where such emergence is guided by principles we are only just beginning (as in the work by Ballard, Gray, and others) to understand. A serious worry about the apparently innocent and conservative HEMC, it seems to me, is thus that it threatens to repeat for outer circuits and elements the mistake that Daniel Dennett (*op. cit.*) famously warns us against with regard to inner circuits and elements. It depicts such bodily or outer resources as doing their work only by parading structure and information in front of some thoughtful neural overseer. In the absence of any such privileged inner component, the outer and the inner operations are free to emerge as well-tuned co-active participants in the construction of thought and reason.

To be sure, this overall vision (of cognition genuinely distributed between brain, body, and world) bequeaths a brand new set of puzzles. It invokes an ill-understood process that soft-assembles a problem-solving whole from a candidate pool that may include neural storage and processing routines, perceptual and motoric routines, external storage and operations, and a variety of self-stimulating cycles involving self-produced material scaffolding. But importantly, this all applies with equal force to the neural economy itself. Here too, a cognitive task will often be addressed by a soft-assembled coalition of distributed (and often highly heterogeneous) neural components and brain areas, temporarily held together by a transient pattern of "functional connectivity."[64] HEC thus gains in plausibility as soon as the *inner* economy is seen aright: as itself hybrid, fragmented, yet vastly empowered by an ill-understood capacity to form and reform into a variety of surprisingly integrated (though temporary) wholes. What HEC allows us to see clearly is that where ongoing human cognitive activity is concerned, there are usually *many* boundaries in play, *many different kinds* of capacity and resource in action, and a complex flux of recruitment, retrieval, and processing defined across these shifting, heterogeneous, multi-faceted wholes.

This turns Rupert's argument on its head. For having allowed that we could, if we so wished, choose to parse our cognitively potent coupled unfoldings according to either HEC or HEMC, we can now see that it is the choice of HEMC that sometimes threatens to obscure much that is of value. We do indeed seek to carve nature at the most

[64] For a nice account of the distinction between persisting structure and short-lived functional complexity in the brain, see O. Sporns, D. Chialvo, M. Kaiser, and C.C. Hilgetag, "Organization, Development and Function of Complex Brain Networks," *Trends in Cognitive Sciences*, VIII (2004): 418–25.

causally relevant joints, a task not accomplished by elevating anatomic and metabolic boundaries into make-or-break cognitive ones. The cure for cognitive hiccups (the unproductive argumentative oscillation from HEC to HEMC to HEC to HEMC…) is thus at hand. For the only real danger from HEC is that it may blind us to the genuine extent to which human cognition, though not organism-bound, remains importantly organism-centered. To guard against that misreading, we may now scout:

> HOC (Hypothesis of Organism-Centered Cognition)
> Human cognitive processing (sometimes) literally extends into the environment surrounding the organism. But the organism (and within the organism, the brain/CNS) remains the core and currently the most active element. Cognition is organism-centered even when it is not organism-bound.

HEC, HEMC, HOC? We should not feel locked into some pale zero sum game. As philosophers and as cognitive scientists we can (and should) practice the art of flipping between these different perspectives, treating each as a lens apt to draw attention to certain features, regularities, and contributions while making it harder to spot others, or to give them their problem-solving due. The cure for cognitive hiccups is to stop worrying and enjoy the ride.

ANDY CLARK

University of Edinburgh