g in the false belief task:
norality have a biological
noral sentiments relating
al Sciences 270: 819–26.
scent with modification,
ature 395: 123–4.
oming auditory motion,
York: Oxford University
iam Morrow.
onent model of working
sychology, Neuroscience
How Culture Transformed
rsistent fear responses in
," Science 136: 982–3.
l processing of auditory
147–51.
ational Journal of Prima-
gist 55: 1233–43.
demiology of representa-
pping the Mind: Domain
w York: Cambridge Uni-
ry. Princeton University
gnition. Cambridge, MA:
ndations of culture, in J.
pted Mind: Evolutionary
-136). Oxford University

14 Embodied, embedded, and extended cognition

Andy Clark

14.1 Introduction: world enough, and flesh

Flesh and world are surely flavors of the moment. Talk of mind as *intimately* embodied and *profoundly* environmentally embedded shimmers at the cusp of the cognitive scientific zeitgeist. But beneath the glamour and glitz lies a still-murky vision. For this is a view of mind that can seem by turns radical and trivial, interestingly true and outrageously false, scientifically important and a mere distraction, philosophically challenging and simply confused. This chapter is an attempt to locate some footholds in this new and at times treacherous landscape.

It is comforting to begin with a seeming truth. Human minds, it can hardly be doubted, are at the very least in deep and critically important contact with human bodies and with the wider world. Human sensing, learning, thought, and feeling are all structured and informed by our body-based interactions with the world around us. Thus when Esther Thelen, a leading proponent of the embodied perspective, writes that "to say that cognition is embodied means that it arises from bodily interactions with the world" (Thelen 2000, p. 4), no sensible person is likely to disagree. But surely that isn't *all* that it means?

Clearly, there is more to this than meets the eye. Here is how the quote continues:

> From this point of view, cognition depends on the kinds of experiences that come from having a body with particular perceptual and motor capacities that are inseparably linked and that together form the matrix within which memory, emotion, language, and all other aspects of life are meshed. The contemporary notion of embodied cognition stands in contrast to the prevailing cognitivist stance which sees the mind as a device to manipulate symbols and is thus concerned with the formal rules and processes by which the symbols appropriately represent the world. (Thelen 2000, p. 4)

Some of the material in this chapter is drawn from *Supersizing the Mind: Embodiment, Action, and Cognitive Extension* (Oxford University Press, 2008). Thanks to the publishers for permission to use this material here. This chapter was prepared thanks to support from the AHRC, under the ESF Eurocores CNCC scheme, for the CONTACT (Consciousness in Interaction) project, AH/E511139/1.

In this much-quoted passage we begin to glimpse some of the key elements of a more radical view. But even here, there are plenty of claims with which no one is likely to take issue. As active sensors of our world, possessed of bodies with specific shapes and characters, it is relatively unsurprising if what we think, do, and perceive all turn out to be *in some sense* deeply intertwined. Nor is it all that surprising if much of higher cognition turns out to be in some sense *built on* a substrate of embodied perceptuo-motor capacities. But the notion of "meshing" that Thelen deploys should give us pause, suggesting as it does a kind of ongoing intermingling of cognitive activity with the perceptuo-motor matrix from which it putatively emerges.

Meshing and intermingling are likewise prominent in John Haugeland's benchmark assertion that

> If we are to understand mind as the locus of intelligence, we cannot follow Descartes in regarding it as separable in principle from the body and the world... Broader approaches, freed of that prejudicial commitment, can look again at perception and action, at skillful involvement with public equipment and social organization, and see not principled separation but all sorts of close coupling and functional unity... Mind, therefore, is not incidentally but *intimately* embodied and *intimately* embedded in its world. (Haugeland 1998, pp. 236–7)

What this passage makes clear is that the core claim at issue is not primarily a claim about development and learning. Nor is it about the undoubted role of body and world in fixing the contents of thought, or in determining the sequence of thoughts, or even in determining what kinds of thing we find it worth thinking about. Rather, what is at issue is something to do with the separability of mind, body, and world, at least for the purposes of understanding mind as the "locus of intelligence." What Haugeland is selling is a radical package deal aimed at undermining a simple, but arguably distortive, model of mind. This is the model of mind as essentially inner and (in our case) neurally realized. It is, to put it bluntly, the model of mind as brain (or perhaps brain and central nervous system): a model increasingly prevalent in a culture where just about everything to do with thinking seems to be accompanied by some kind of image of the brain. Call this model BRAINBOUND.

According to BRAINBOUND the (non-neural) body is just the sensor and effector system of the brain, and the (rest of the) world is just the arena in which adaptive problems get posed and the brain–body system must sense and act. If BRAINBOUND is correct, then all thoughts and feelings, and all cognition properly so called, depend directly upon neural activity alone. The neural activity itself may, of course, in turn depend on worldly inputs and (extra-neural) bodily activity. But that would be merely what Hurley

(1998, pp. 10–11) usefully dubs "instrumental dependence," as when we move our eyes and get a new perceptual experience as a result. BRAINBOUND asserts, seemingly in opposition to the very possibility of non-instrumental forms of bodily and worldly dependence, that all that really matters as far as the actual mechanisms of cognition are concerned is what the brain does: body and world act merely as sources of input and arenas for output.

Maximally opposed to BRAINBOUND is a view according to which thinking, cognizing, and feeling may all (at times) depend directly and non-instrumentally upon the ongoing work of the body and/or the extra-organismic environment. Call this model POROUS. According to POROUS, the actual local operations that make cognizing possible and that give content and character to our mental life include inextricable tangles of feedback, feedforward, and feedaround loops that promiscuously criss-cross the boundaries of brain, body, and world. The local mechanisms of mind, if POROUS is correct, are not all in the head.

Why might anyone think that POROUS expresses a truth about the mind? As a quick and dirty example, consider the familiar practice of writing while problem solving. One way to conceive of this process is in terms of a BRAINBOUND cognitive engine, one that generates ideas that are then stored externally as a hedge against forgetting or as a ploy to enable the communal sharing of information. But while both these roles are real and important, many people feel as if the act of writing is playing some rather more active role, as if the act itself matters in some way that goes beyond the simple offloading of a previously formed thought. Here, for example, is a famous exchange between the physicist 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."
>
> "No, it's not a *record*, not really. It's *working*. You have to work on paper and this is the paper. Okay?" (Quoted in Gleick 1993, p. 409)

Feynman's suggestion is that the loop into the external medium is integral to the intellectual activity, to the *working*, itself. It is not just the contingent environmental outflow of the working, but actually forms part of it. If such loops are indeed integral to certain forms of intelligent activity, we need to understand when and why this can be so, and just what it might mean (if anything) for our general model of minds and agency. Do such examples lend support to a vision such as POROUS or are they better accommodated (as many critics believe) in some much more deflationary way?

14.2 Simple causal spread

In a range of interesting and important cases, there is clear evidence that the problem-solving load is spread out across brain, body, and (sometimes) world. To get the flavor of this, it is helpful to contrast various solutions to a single problem. Take the case of walking (powered locomotion).

Honda's walking robot Asimo is billed, perhaps rightly, as the world's most advanced humanoid robot. Boasting a daunting 26 degrees of freedom (2 on the neck, 6 on each arm, and 6 on each leg) Asimo is able to navigate the real world, reach, grip, walk reasonably smoothly, climb stairs, and recognize faces and voices. The name Asimo stands (a little clumsily perhaps) for "Advance Step in Innovative Mobility." And certainly, Asimo is an incredible feat of engineering: still relatively short on brainpower but high on mobility and maneuverability.

As a walking robot, however, Asimo is far from energy efficient. For a walking agent, one way to measure energy efficiency is by the so-called "specific cost of transport" (Tucker 1975) – viz., "the amount of energy required to carry a unit weight a unit distance" calculated as (energy used)/(weight)(distance traveled). The lower the resulting number, the less energy is required to shift a unit of weight a unit of distance. Asimo rumbles in (see Collins and Ruina 2005) with a specific cost of transport of about 3.2, whereas we humans display a specific metabolic cost of transport of about 0.2. What accounts for this massive difference in energetic expenditure?

Where robots like Asimo walk by means of very precise, and energy-intensive, joint-angle control systems, biological walking agents make maximal use of the mass properties and bio-mechanical couplings present in the overall musculoskeletal system and walking apparatus itself. Wild walkers thus make canny use of so-called "passive dynamics," the kinematics and organization inhering in the physical device alone (McGeer 1990). Pure passive dynamic walkers are simple devices that boast no power source apart from gravity, and no control system apart from some simple mechanical linkages such as a mechanical knee and the pairing of inner and outer legs to prevent the device from keeling over sideways. Yet despite (or perhaps because of) this simplicity, such devices are capable, if set on a slight slope, of walking smoothly and with a very realistic gait. The ancestors of these devices are, as Collins, Wisse, and Ruina (2001) nicely document, not sophisticated robots but children's toys, some dating back to the late nineteenth century: toys that stroll, walk, or waddle down ramps or when pulled by string. Such toys have minimal actuation and no control system. Their walking is a consequence not of complex joint movement planning and actuating, but of basic morphology (the shape of the body, the distribution of linkages and weights of components, etc.).

Behind the passive dynamic approach thus lies the compelling thought that

> Locomotion is mostly a natural motion of legged mechanisms, just as swinging is a natural motion of pendulums. Stiff-legged walking toys naturally generate their comical walking motions. This suggests that human-like motions might come naturally to human-like mechanisms. (Collins *et al.* 2001, p. 608)

Collins *et al.* (2001) built the first such device to mimic human-like walking, by adding curved feet, a compliant heel, and mechanically linked arms to the basic design pioneered by McGeer (1990). In action, the device exhibits good, steady motion and is described by its creators as "pleasing to watch" (2001, p. 613). By contrast, robots that make extensive use of powered operations and joint angle control tend to suffer from "a kind of rigor mortis [because] joints encumbered by motors and high-reduction gear trains... make joint movement inefficient when the actuators are on and nearly impossible when they are off" (2000, p. 607).

What, then, of powered locomotion? Once the body itself is "equipped" with the right kind of passive dynamics, powered walking can be brought about in a remarkably elegant and energy-efficient way. In essence, the tasks of actuation and control have now been massively reconfigured so that powered, directed locomotion can come about by systematically pushing, damping, and tweaking a system in which passive dynamic effects still play a major role. The control design is delicately geared to utilize all the natural dynamics of the passive baseline, and the actuation is consequently efficient and fluid.

Some of the core flavor of such a solution is captured by the broader notion of *ecological control* (see Clark 2007) where an ecological control system is one in which goals are not achieved by micro-managing every detail of the desired action or response, but by making the most of robust, reliable sources of relevant order in the bodily or worldly environment of the controller. In such cases the "matching" (of sensors, morphology, motor system, materials, controller, and ecological niche) yields a spread of responsibility for efficient adaptive response: the details of embodiment take over some of the work that would otherwise need to be done by the brain or the neural network controller, an effect that Pfeifer and Bongard (2007, p. 100) aptly describe as "morphological computation." The exploitation of passive dynamic effects thus exemplifies one of several key characteristics of the embodied, embedded approach: a characteristic that Wheeler and Clark (1999) dubbed *non-trivial causal spread*. Non-trivial causal spread occurs whenever something we might have expected to be achieved by a certain well-demarcated system turns out to involve the exploitation of more far-flung factors and forces. When a Mississippi alligator allows the temperature of the rotting vegetation in which it lays its eggs to determine the sex of its offspring, we encounter some non-trivial causal spread. When the passive dynamics of the actual legs and body take care of many of the demands that we might otherwise have ceded to an energy-hungry joint angle control system, we encounter non-trivial causal spread. One of the big lessons of contemporary robotics is that the co-evolution of morphology (which can include sensor placement, body-plan, and even the choice of basic building materials, etc.) and control yields a truly golden opportunity to spread the problem-solving load between brain, body, and world. (For excellent discussion, see Pfeifer and Scheier 1999; Pfeifer 2000. For the possible importance of bedrock materials, see Brooks 2001.)

14.3 Action as information self-structuring

Ballard *et al.* (1997) describe a task in which you are given a model pattern of colored blocks that you are asked to copy by moving similar blocks from a reserve area to a new workspace. Using the spare blocks in the reserve area, your task is to recreate the pattern by moving one block at a time from the reserve to the new version you are busy creating. The task is to be performed using mouse clicks and drags on a computer screen. As you perform, eye-tracker technology is monitoring exactly where and when you are looking at different bits of the puzzle.

What problem-solving strategy do you think you would use? One neat strategy might be to look at the target, decide on the color and position of the next block to be added, then execute the plan by moving a block from the reserve area. This is, for example, pretty much the kind of strategy you'd expect of a classical Artificial Intelligence planning system. When asked how we would solve the problem, many of us pay lip service to this kind of neat and simple strategy. But the lips tell one story while the hands and eyes tell another. For this is emphatically not the strategy used by most human subjects. What Ballard *et al.* found was that repeated rapid saccades to the model were used in the performance of the task: many more than you might expect. For example, the model is consulted *both before and after* picking up a block, suggesting that when glancing at the model, the subject stores only one piece of information: either the color or the position of the next block to be copied.

To test this hypothesis, Ballard *et al.* used a computer program to alter the color of a block while the subject was looking elsewhere. For most of these interventions, subjects did not notice the changes even for blocks and locations that had been visited many times before, or that were the focus of the current action. The explanation was that when glancing at the model, the subject stores only one piece of information: either the color or the position of the next block to be copied (not both). In other words, even when repeated saccades are made to the same site, very minimal information is retained.

Instead, repeated fixations provide specific items of information "just in time" for use. The experimenters conclude that

> In the block-copying paradigm . . . fixation appears to be tightly linked to the underlying processes by marking the location at which information (e.g., color, relative location) is to be acquired, or the location that specifies the target of the hand movement (picking up, putting down). Thus fixation can be seen as binding the value of the variable currently relevant for the task. (Ballard *et al.* 1997, p. 734)

Two morals matter for the story at hand. The first is that visual fixation is here playing an identifiable computational role. As the authors (p. 725) comment, "changing gaze is analogous to changing the memory reference in a silicon computer." The second is that repeated saccades to the physical model thus allow the subject to deploy what Ballard *et al.* dub "minimal memory strategies" to solve the problem. The idea is that the brain creates its programs so as to minimize the amount of working memory that is required, and that eye motions are here recruited to place a new piece of information into memory. Indeed, by altering the task demands, Ballard *et al.* were also able to systematically alter the particular mixes of biological memory and active, embodied retrieval recruited to solve different versions of the problem. They conclude that, in this kind of task at least, "eye movements, head movements, and memory load trade off against each other in a flexible way" (1997, p. 732). As a result, a Ballard-style approach is able

> To combine the concept that looking is a form of doing with the claim that vision is computation [by] introducing the idea that eye movements . . . allow perceivers to exploit the world as a kind of external storage device. (Wilson 2004, pp. 176–7)

Bodily actions here appear as among the means by which certain (in this case quite familiar) computational and representational operations are implemented. The difference is just that the operations are realized not in the neural system alone, but in the whole embodied system located in the world.

Embodied agents are also able to act on their worlds in ways that conjure cognitively and computationally potent time-locked patterns of sensory stimulation. In this vein Fitzpatrick *et al.* (2003) show, using robot demonstrations, exactly how active object manipulation (the robots are able to push and touch objects in view) can help generate information about object boundaries and affordances. Similarly, in human infants, grasping, poking, pulling, sucking, and shoving create a flow of multi-modal sensory stimulation that has been shown (Lungarella and Sporns 2005) to aid category learning and concept formation. The key to such capabilities is the robot or infant's capacity to maintain coordinated sensorimotor engagement with its environment. Self-generated motor activity, such work suggests, acts as a "complement to neural information-processing" in that

> The agent's control architecture (e.g. nervous system) attends to and processes streams of sensory stimulation, and ultimately generates sequences of motor actions which in turn guide the further production and selection of sensory information. [In this way] "information structuring" by motor activity and "information processing" by the neural system are continuously linked to each other through sensorimotor loops. (Lungarella and Sporns 2005, p. 25)

14.4 Cognitive extensions

So far, we have been seeing evidence of the important roles played by bodily form and bodily action in the solution of basic adaptive problems such as locomotion and learning. But what about mature thought and reason? Does embodiment and environmental embedding play a role here too?

Consider an accountant, Ada, who is extremely good at dealing with long tables of figures. Over the years, Ada has learned how to solve specific classes of accounting problems by rapidly scanning the columns, copying some numbers onto a paper scratchpad, then looking to and from those numbers (carefully arrayed on the page) back to the columns of figures. This is all now second nature to Ada, who scribbles at lightning speed deploying a variety of "minimal memory strategies" (Ballard *et al.* 1997). Instead of attempting to commit multiple complex numerical quantities and dependencies to biological short-term memory, Ada creates and follows trails through the scribbled numbers, relying on self-created external traces every time an intermediate result is obtained. These traces are visited and revisited on a "just-in-time, need to know" basis, briefly shunting specific items of information into and out of short term bio-memory in much the same way as a serial computer shifts information to and from the central registers in the course of carrying out some computation. This process may be analyzed in "extended functional" terms, as a set of problem-solving state-transitions whose implementation happens to involve a distributed combination of biological memory, motor actions, external symbolic storage, and just-in-time perceptual access.

Robert Wilson's notions of "exploitative representation" and "wide computation" (Wilson 1994, 2004) capture some of the key features of such an extended approach. Exploitative representation occurs when a subsystem gets by without explicitly encoding and deploying some piece of information, in virtue of its ability to track that information in some other way. Wilson gives the example of an odometer that keeps track of how many miles a car has traveled not by first counting wheel rotations then multiplying according to the assumption that each rotation = x meters, but by being built so as to record x meters every time a rotation occurs:

> In the first case it encodes a representational assumption and uses this to compute its output. In the second it contains no such encoding but instead uses an existing relationship between its structure and the structure of the world. (Wilson 2004, p. 163)

Wilson's descriptions and central examples can make it seem as if exploitative representation is all about achieving success without representations at all, at least in any robust sense of representation. But this need not be so. Another, very pertinent, range of cases would be those in which a subsystem does not contain within itself a persisting encoding of certain things, but instead leaves that information in the world, or leaves encoding it to some other subsystem to which it has access. Thus Ada's biological brain does not create and maintain persistent internal encodings of every figure she generates and offloads onto the page, though it may very well create and maintain persistent encodings of several other key features (for example, some kind of running approximation that acts to check for gross errors). In much the same way as Ballard's block-puzzlers, Ada's biological brain may thus, via the crucial bridging capacities of available embodied action, key its own internal representational and internal computational strategies to the reliable presence of the external pen-and-paper buffer. Even robustly representational inner goings-on may thus count as exploitative insofar as they merely form one part of a larger, well-balanced process whose cumulative set of state-transitions solves the problem. In this way

> explicit symbolic structures in a cognizer's environment . . . together with explicit symbolic structures in its head [may] constitute the cognitive system relevant for performing some given task. (Wilson 2004, p. 184)

The use of various forms of exploitative representation immediately yields a vision of what Wilson dubs "wide computationalism," according to which "at least some of the computational systems that drive cognition reach beyond the limits of the organismic boundary" (2004, p. 165). Extended functional systems may include coupled motor behaviors as processing devices and more static environmental structures as longer-term storage and encoding devices. The larger systems thus constituted are, as Wilson insists, unified wholes such that "the resulting mind–world computational system itself, and not just the part of it inside the head, is genuinely cognitive" (2004, p. 167).

Extended functionalists thus reject the image of mind as a kind of input–output sandwich with cognition as the filling (for this picture, and many more arguments for its rejection, see Hurley 1998; see also Clark 1997a; Clark and Chalmers 1998). Instead, we confront an image of the local mechanisms of human cognition quite literally bleeding out into body and world. The traditional functionalist was interested in neural goings-on as the contingent means by which human beings manage to implement the specific functional organizations characteristic of the human mind. The extended functionalist takes this one step further. From an extended functionalist perspective, not just the brain, but also the (non-neural) body and world, are apt to provide the physical machinery that implements (some of) the abstract organizations that turn matter into mind.

14.5 Critical reactions

It is the claims concerning cognitive extension, rather than those concerning simple causal spread (which now seems widely accepted in both the philosophical and cognitive scientific communities), that have received the most critical attention. Insofar as the more basic claims (about embodiment and causal spread) have been subject to critical scrutiny, it has mainly consisted in worries about a non-essential accompaniment to those claims, viz., the tendency of some theorists to reject the appeal to internal representation and/or computation in the explanation of adaptive success. Thus Grush (2003) takes issue with what he describes as

> a growing radical trend in current theoretical cognitive science that moves from the premises of embedded cognition, embodied cognition, dynamical systems theory and/or situated robotics to conclusions either to the effect that the mind is not in the head or that cognition does not require representation, or both. (Grush 2003, p. 53)

Grush's stalking horse is, in fact, a view that is in at least one crucial respect much more radical than POROUS itself. It is the view that

> the mind is not essentially a thinking or representing thing: it is a controller, a regulator, an element in a swarm of mutually causally interacting elements that includes the body and environment whose net effect is adaptive behavior. (Grush 2003, p. 55)

POROUS, however, need not deny that the mind is essentially a thinking or representing thing. It is committed only to the weaker claim that the thinking, and even the representing, may in many cases supervene on activities and encodings that criss-cross brain, body, and world. The debate concerning internal representation is thus independent (or so I have argued: see Clark 1997) of many of the key claims concerning causal spread between brain, body, and world.

Concerning the putative extension of (some of) the machinery of mind and reason into the surrounding world, Rupert (2004) worries that not enough has been done to justify talk of genuine cognitive extension. For all that matters, in such cases, is fully captured (Rupert claims) by the more conservative claim that he terms the *hypothesis of embedded cognition* (HEMC). According to 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. (Rupert 2004, p. 393)

In other words, Rupert wants to treat all the cases in the way we (above) treated cases of simple causal spread. One reason for this is that Rupert (see also Adams and Aizawa 2001) is impressed by the profound *differences* that appear to distinguish the inner and outer contributions to human cognitive success. Thus, for example, we read that "the external portions of extended 'memory' states (processes) differ so greatly from internal memories (the process of remembering) that they should be treated as distinct kinds" (Rupert 2004, p. 407).

Part of the problem here may stem from a persistent misreading of the so-called "parity claim" introduced in Clark and Chalmers (1998). 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. But far from *requiring* any deep 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: avoiding a rush to judgment based on spatial location alone. The parity principle was meant to engage our rough sense 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 Wheeler (2010) 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. (Wheeler 2010, p. 3)

To do things that way, Wheeler argues, is to open the door to the highly chauvinistic thought that only systems whose fine-grained causal profile fully matches that of the brain can be cognitive systems at all. Yet, just because some alien neural system failed to match our own in various ways we should surely not *thereby* be forced to count the action of such systems as "non-cognitive." 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 we ought to bring to bear on an alien neural organization. It is 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 unlike (perhaps deeply complementary to) those of the biological brain.

It is also 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, core biological bundle that lies at the heart of each episode of cognitive processing. Occasionally, under strict and rare conditions we may confront genuine extensions of even that more-or-less persisting core: cases where even 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 a dovetailed 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 in response to the particularities of task and context. As just one example, consider Van Essen, Anderson, and Olhausen's (1994) account 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. Control neurons thus weave functional modules "on the hoof," in a way sensitive to the effects of context, attention, and so on. As Jerry Fodor once put it, in such cases it is "*unstable instantaneous* connectivity that counts" (1983 , p. 118; see also Fodor 2001). The resulting soft-wired ensembles, in which information then flows and is processed in ways apt to the task at hand, do not cease to be important just because they are transient creations ushered into being by a preceding wave of "neural recruitment."

Rupert worries that, by taking seriously the notion of cognitive extension in the special subclass of transient cases where the newly recruited organizations span brain, body, and world, we lose our grip on the persisting systems that we ordinarily take to be our objects of study. For indeed, as Rupert (2008) points out, much work in cognitive and experimental psychology proceeds by assuming that subjects are "persisting, organismically bound cognitive systems." Fortunately, however, there is no incompatibility whatsoever between the claims about cognitive extension and the notion of a persisting common biological core. Nor does anything in such treatments threaten to deprive us of that common core as a proper object of scientific study. If our avowed goal is to discover the stand-alone properties of the neural apparatus, 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 non-biological props and aids. But if our goal is to unravel the mechanically modulated flow of energy and information that allows an identifiable agent (a Sally, Johnny, or Terry) 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 from an information-processing perspective.

As philosophers and as cognitive scientists we should, I suggest, practice the black but important art of repeatedly flipping between these different possible perspectives (extended, organismic, neural), treating each as a lens apt to draw attention to certain features, regularities, and contributions while at the same time obscuring others.

14.6 Conclusions

In his famous (1982) treatment, *The Extended Phenotype*, Richard Dawkins (pp. 4–5) encourages readers to try a "mental flip." Where before we saw only whole organisms (albeit replete with smaller parts, and themselves forming and re-forming into larger groups and wholes) we now see transparent bodies and the near-seamless play of replicating DNA. The spider's web appears as a proper part of the spider's extended phenotype, and the organism emerges as no more (and no less) than an adaptively potent non-random concentration of DNA. This perspective, Dawkins (p. 1) concedes, is not compulsory, nor can it be simply proved (or disproved) by experiment. Its virtues lie rather in the new ways of seeing familiar phenomena that it may breed, in that flip of perspective that invites us to view the larger organism–environment system in new and illuminating ways.

Work on embodiment, embedding, and cognitive extension likewise invites us to view mind and cognition in a new and (I believe) illuminating manner. It invites us to cease to unreflectively privilege (as does BRAINBOUND) the inner, the biological, and the neural, while at the same time helping us better to understand the crucial contribution of the whole organism and (within that organism) of neural control systems in the production of intelligent, information-based response. As POROUS cognitive agents we are merciless exploiters of bodily and environmental structure, and inveterate conjurors of our own cognition-enhancing input streams. Somewhat paradoxically, then, sustained attention to embodiment and action renders the bounds of skin and skull increasingly transparent, revealing processes running through body and world as integral parts of the machinery of mind and cognition.

To unravel the workings of these embodied and extended minds requires an unusual mix of neuroscience, computational, dynamical, and information-theoretic understandings, "brute" physiology, ecological sensitivity, and attention to the stacked designer cocoons in which we learn, think, and act. This may seem a daunting prospect, but there is cause for optimism. In learning, development, and evolution, trade-offs between neural control, morphology, action, and the epistemic use of environmental resources and opportunities are regularly and reliably achieved. Since such solutions are reliably found, there is a good chance that they can be systematically understood. Better still, the sciences of the mind are already well on the way to developing frameworks and forms of analysis that make headway with this difficult task. A mature science of the embodied mind will, I have argued, need to combine so-called "dynamical" insights (such as the stress on various forms of coupled organism–environment unfolding) with a much better understanding of the broad space of adaptive trade-offs: an understanding probably best filtered through the more familiar lenses of computational, representational, and information-theoretic tools and constructs.

The appeal to embodiment, embedding, and cognitive extension, if this is correct, marks not so much a radical shift as a natural progression in the maturing of our understanding the mind. It does not call into question all forms of "machine metaphors," and need involve no rejection of (though it is by no means exclusively committed to) accounts couched in terms of representations and computations. Indeed, the most natural way to approach the tough task of understanding *just how* body and world contribute to our cognitive performances is (I have tried to suggest) by the use of what are still broadly speaking functional and information-theoretic perspectives. The hope is rather to add new layers to our functional and information-processing understandings, by revealing the role of complex coupled dynamics, non-neural resources, and embodied action in the very machinery of thought and reason.

Further reading

Adams, F. and Aizawa, K. (2008). *The Bounds of Cognition*. Oxford: Wiley-Blackwell. A concise and accessible critique of the various arguments meant to support the hypothesis of the extended mind.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press. A broad integrative overview, ranging from robotics to language to economic institutions, delivered with a mildly philosophical slant.

(2008).
Oxfor
ing th
Gallagher,
excell
Pfeifer, R.
Camb
Robbins, I
Cogni
entrie
Rowlands,
yet b
Rupert, R.
Philo
novel
the b
of th
Exter
Wheeler,
Press
and c
Wilson, R
ences
exam
defer

Referen

Adams, F.
ogy 1
Ballard, D
imen
Brooks, R
Clark, A.
Caml
(2007).
minc
Clark, A.
Collins, S
hum
Robo
Collins, S
dyna
Robo

(2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension.* Oxford University Press. A detailed reworking of the central claims concerning the "extended mind" and a response to the main critiques.

Gallagher, S. (2005). *How the Body Shapes the Mind.* Oxford University Press. An excellent synthesis of empirical work and phenomenology.

Pfeifer, R. and Bongard, J. (2007). *How the Body Shapes the Way We Think.* Cambridge, MA: MIT Press. An inspiring and rich, robotics-based, overview.

Robbins, P. and Aydede, M. (eds.) (2008). *The Cambridge Handbook of Situated Cognition.* Cambridge University Press. A wonderful and diverse collection of entries concerning the embodied, situated, and extended mind.

Rowlands, M. (1999). *The Body In Mind.* Cambridge University Press. A careful yet broad overview, conducted with a keen philosophical eye.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition, *Journal of Philosophy* 101: 389–428. An important critical treatment, that develops a novel "systems-based" objection to the idea that human minds extend beyond the bounds of the organism. For a fuller and even wider-ranging deployment of this important strategy, see Rupert, R. (2009). *Cognitive Systems and the Extended Mind.* New York: Oxford University Press.

Wheeler, M. (2005). *Reconstructing the Cognitive World.* Cambridge, MA: MIT Press. A wonderfully original take on the debates concerning embodiment and cognitive extension. Embodiment with a Heideggerian twist.

Wilson, R. A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences – Cognition.* Cambridge University Press. A delightfully wide-ranging examination of the role of the individual in the sciences of mind, and a defense of the claim that the boundaries of the mind extend beyond the skin.

References

Adams, F. and Aizawa, K. (2001). The bounds of cognition, *Philosophical Psychology* 14: 43–64.

Ballard, D., Hayhoe, M., Pook, P., and Rao, R. (1997). Deictic codes for the embodiment of cognition, *Behavioral and Brain Sciences* 20: 723–67.

Brooks, R. (2001). The relationship between matter and life, *Nature* 409: 409–11.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again.* Cambridge, MA: MIT Press.

(2007). Re-inventing ourselves: The plasticity of embodiment, sensing, and mind, *Journal of Medicine and Philosophy* 32: 263–82.

Clark, A. and Chalmers, D. (1998). The extended mind, *Analysis* 58: 7–19.

Collins, S. H. and Ruina, A. (2005). A bipedal walking robot with efficient and human-like gait, in *Proceedings of the IEEE International Conference on Robotics and Automation, Barcelona, Spain* (pp. 1983–8).

Collins, S. H., Wisse, M., and Ruina, A. (2001). A three-dimensional passive-dynamic walking robot with two legs and knees, *International Journal of Robotics Research* 20: 607–15.

Dawkins, R. (1982). *The Extended Phenotype.* Oxford University Press.

Fitzpatrick, P., Metta, G., Natale, L., Rao, S., and Sandini, G. (2003). Learning about objects through action: Initial steps towards artificial cognition, in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), May 12–17, Taipei, Taiwan* (pp. 3140–5).

Fodor, J. A. (1983). *The Modularity of Mind.* Cambridge, MA: MIT Press.

(2001). *The Mind Doesn't Work That Way.* Cambridge, MA: MIT Press.

Gleick, J. (1993). *Genius: The Life and Times of Richard Feynman.* New York: Vintage.

Grush, R. (2003). In defence of some "Cartesian" assumptions concerning the brain and its operation, *Biology and Philosophy* 18: 53–93.

Haugeland, J. (1998). Mind embodied and embedded, in J. Haugeland, *Having Thought: Essays in the Metaphysics of Mind* (pp. 207–40). Cambridge, MA: Harvard University Press.

Hurley, S. (1998). *Consciousness in Action.* Cambridge, MA: Harvard University Press.

Lungarella, M. and Sporns, O. (2005). Information self-structuring: Key principle for learning and development, in *Proceedings of the 4th IEEE International Conference on Development and Learning* (pp. 25–30).

McGeer, T. (1990). Passive dynamic walking, *International Journal of Robotics Research* 9: 68–82.

Pfeifer, R. (2000). On the role of morphology and materials in adaptive behavior, in J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, and S.W. Wilson (eds.), *From Animals to Animats 6: Proceedings of the 6th International Conference on Simulation of Adaptive Behavior* (pp. 23-32). Cambridge, MA: MIT Press.

Pfeifer, R. and Bongard, J. (2007). *How the Body Shapes the Way We Think.* Cambridge, MA: MIT Press.

Pfeifer, R. and Scheier, C. (1999). *Understanding Intelligence.* Cambridge, MA: MIT Press.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition, *Journal of Philosophy* 101: 389–428.

(2008). Innateness and the situated mind, in P. Robbins and M. Aydede (eds.), *The Cambridge Handbook of Situated Cognition* (pp. 96–116). Cambridge University Press.

Thelen, E. (2000). Grounded in the world: Developmental origins of the embodied mind, *Infancy* 1: 3–28.

Tucker, V. A. (1975). The energetic cost of moving about, *American Scientist* 63: 413–19.

Van Essen, D. C., Anderson, C. H., and Olshausen, B. A. (1994). Dynamic routing strategies in sensory, motor, and cognitive processing, in C. Koch and J. Davis (eds.), *Large Scale Neuronal Theories of the Brain* (pp. 271–99). Cambridge, MA: MIT Press.

Wheeler, M. (2010). Minds, things and materiality, in L. Malafouris and C. Renfrew (eds.), *The Cognitive Life of Things.* Cambridge: McDonald Institute for Archaeological Research.

Wheeler, M. and Clark, A. (1999). Genic representation: Reconciling content and causal complexity, *British Journal for the Philosophy of Science* 50: 103–35.

Wilson, R. A. (1994). Wide computationalism, *Mind* 103: 351–72.

(2004). *Boundaries of the Mind: The Individual in the Fragile Sciences – Cognition*. Cambridge University Press.