

1

The Active Body

1.1 *A Walk on the Wild Side*

Honda's Asimo (see fig. 1.1) 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 Advanced 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)—namely, “the amount of energy required to carry a unit weight a unit distance.”¹ The lower the number, the less energy is required to shift a unit of weight a unit of distance. Asimo rumbles in 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?

Whereas 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 biomechanical couplings

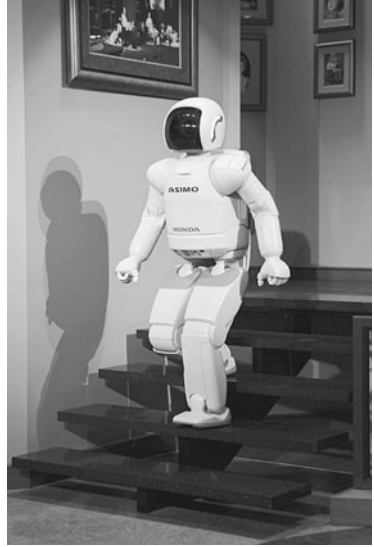


FIGURE 1.1 Honda's Asimo robot.
 (<http://asimo.honda.com/gallery.aspx>;
 by permission of Honda Corporation)

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 19th century. These toys stroll, walk, or waddle down ramps or when pulled by string (see fig. 1.2). 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, Wisse, and Ruina 2001, 608)

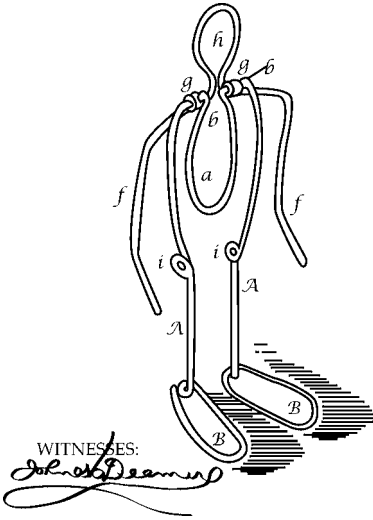


FIGURE 1.2 Fallis's (1888) clever implementation of counterswinging arms. The entire toy is made from two pieces of wire. Each wire makes up a leg, a bearing, an axle, and an arm. One wire also has a head and the other a body of sorts. (S. Collins, M. Wisse and A. Ruina, "A Three-dimensional Passive-dynamic Walking Robot with Two Legs and Knees," *The International Journal of Robotics Research* 20, no. 7 [July 2001]: 607–615, © 2001 Sage Publications, by permission)

Collins, Wisse, and Ruina (2001) built the first such device to mimic humanlike walking by adding curved feet, a compliant heel, and mechanically linked arms to the basic design pioneered by McGeer (1990). In action (see fig. 1.3), the device exhibits good, steady motion and is described by its creators as "pleasing to watch" (McGeer 1990, 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" (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,"² where an ecological control system is one in which goals are not achieved by micromanaging every detail of the desired action or response but by making the most of robust,

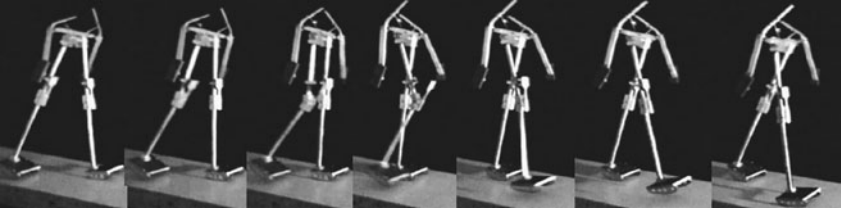


FIGURE 1.3 Pure passive dynamic walker in action. (S. Collins, M. Wisse, and A. Ruina, “A Three-dimensional Passive-dynamic Walking Robot with Two Legs and Knees,” *The International Journal of Robotics Research* 20, no. 7 [July 2001]: 607–615, © 2001 Sage Publications, by permission)

reliable sources of relevant order in the bodily or worldly environment of the controller. In such cases,

part of the “processing” is taken over by the dynamics of the agent-environment interaction, and only sparse neural control needs to be exerted when the self-regulating and stabilizing properties of the natural dynamics can be exploited. (Pfeifer et al. 2006, 7)

A nice example is the use of sparse, well-timed control signals to support the “rolling and rising” motion (see fig. 1.4) of a robot that must raise itself up from a prone position (Kuniyoshi et al. 2004). Another is Iida and Pfeifer’s (2004) work on the running robot Puppy. Puppy has springs (roughly mimicking some of the special properties of a muscle-tendon system) connecting the lower and upper parts of each leg, has pressure sensors on each foot, and benefits from just a few built-in powered oscillatory movements. These simple inbuilt oscillatory movements nonetheless lead, in the special context provided by the sprung body, to fluent running and scampering behavior. Even the simple fact that Puppy has aluminum legs and feet plays an “adaptive” role, for it leads to small amounts of slippage on most surf aces. This might seem like a bad thing, but reducing the slippage by adding rubber pads to the feet caused the robot to begin to fall over: The subtle slippage was actually playing a stabilizing role, effectively enabling the robot to rapidly search for a stable way to proceed (see Pfeifer and Bongard 2007, 96–100, 125–128, for discussion).

In subsequent chapters, we shall encounter ecological control style solutions for problems ranging all the way from perceptuomotor

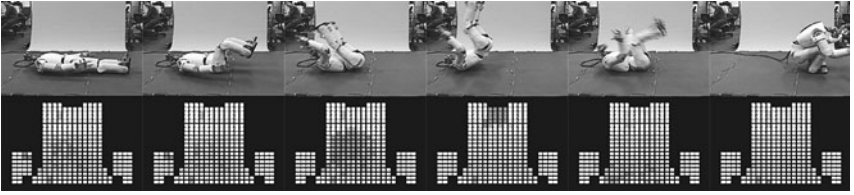


FIGURE 1.4 Sparse but well-timed control signals enable fluent, energy-efficient roll and rise motion. (Work by Kuniyoshi et al. [2004]; figure from Y. Ohmura, by permission)

response to reflection, recall, and deliberation. To capture such effects, Pfeifer and Bongard (2007) invoke the Principle of Ecological Balance.³ This principle states

first... that given a certain task environment there has to be a match between the complexities of the agent's sensory, motor, and neural systems... second... that there is a certain balance or task-distribution between morphology, materials, control, and environment. (123)

The “matching” of sensors, morphology, motor system, materials, controller, and ecological niche yields a spread of responsibility for efficient adaptive response in which “not all the processing is performed by the brain, but certain aspects of it are taken over by the morphology, materials, and environment [yielding] a ‘balance’ or task-distribution between the different aspects of an embodied agent” (see Pfeifer et al. 2006). In such cases, the details of embodiment may 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, 100) aptly describe as “morphological computation.”

The exploitation of passive-dynamic effects exemplifies one of several key characteristics of the embodied, embedded approach that we will encounter as the chapter progresses. This first characteristic has been called *nontrivial causal spread*. Nontrivial causal spread (see Clark 1998b; Wheeler and Clark 1999; Wheeler 2005) occurs whenever *something we might have expected to be achieved by a certain well-demarked system turns out to involve the exploitation of more far-flung factors and forces*.⁴ For the Mississippi alligator, the temperature of the rotting vegetation in which it lays its eggs determines the sex of its offspring. This is an example of nontrivial causal spread. When the passive dynamics of the actual legs and body take care of many of the demands that we

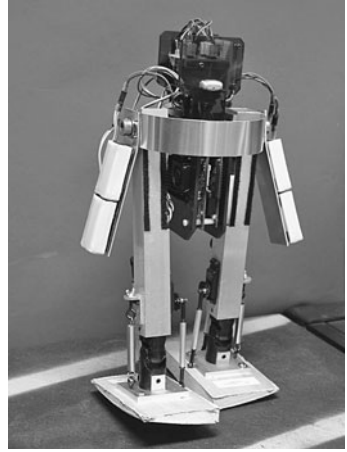


FIGURE 1.5 The Toddler robot, by Russ Tedrake, Teresa Zhang, and H. Sebastian Seung. The robot learns a control policy that exploits the passive dynamics of its own body. (Photo by Teresa Zhang, by permission)

might otherwise have ceded to an energy-hungry joint-angle control system, we likewise encounter nontrivial causal spread. One of the big lessons of contemporary robotics is that the coevolution 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.⁵ Robotics thus rediscovers many ideas explicit in the continuing tradition of J. J. Gibson and of “ecological psychology.”⁶ Thus, William Warren, commenting on a quote from Gibson (1979), suggests that

biology capitalizes on the regularities of the entire system as a means of ordering behavior. Specifically, the structure and physics of the environment, the biomechanics of the body, perceptual information about the state of the agent-environment system, and the demands of the task all serve to constrain the behavioral outcome. (2006, 358)

Such causal spread may be wholly evolved or engineered, wholly learned, or some combination of the two. For example, some control systems are able to actively learn strategies that make the most of passive-dynamic opportunities. An example is the Toddler robot, a walking robot that learns (using so-called actor-critic reinforcement learning) a control policy that exploits the passive dynamics of the body (fig. 1.5). The Toddler robot, which features among the pack of passive-dynamics-based robots described in Collins et al. (2005), can learn to change speeds, go forward and backward, and adapt on the go to different terrains, including bricks, wooden tiles, carpet, and even a variable speed treadmill. And as you’d expect, the use of passive dynamics

cuts power consumption to about one-tenth that of a standard robot like Asimo. The passive-dynamics-based robot described in Collins and Ruina (2005) similarly achieved a specific cost of transport of around 0.20, again around an order of magnitude lower than Asimo and quite comparable to the human case. The discrepancy here is thought not to be significantly reducible by further technological advance using Asimo-style control strategies (i.e., ones that do not exploit passive-dynamic effects). An apt comparison, Collins and Ruina suggest, is with the energy consumption of a helicopter versus airplane or glider. The helicopter, however well designed it may be, will still consume vastly more energy per unit distance traveled.

1.2 *Inhabited Interaction*

Let's switch gears, briefly, to ask what it might be like to be an agent embodied according to these very different sets of principles. What would it feel like to be an intelligent, conscious version of Asimo and, contrariwise, to be an intelligent, conscious version of a fully trained Toddler robot? In the latter case, might it not feel (all other things being equal) as if, with little effort and a simple act of will, directed bodily motion is achieved? In the former, the efforts are large and the body may perhaps be encountered as a complex, resistant object in need of much ongoing energetic micromanagement. Over time, perhaps, control can be streamlined, though energy consumption (as in the case of the helicopter) will still remain high. Nonetheless, the successful exploitation of passive-dynamic effects may well be a major contributing element to what Dourish (2001) nicely calls "inhabited interaction," a way of being in the world that is contrasted with "disconnected control." Here is how Dourish describes the difference, using present-day (i.e., still fairly clunky) virtual-reality systems as a point of comparison:

Even in an immersive virtual-reality environment, users are disconnected observers of a world they do not inhabit directly. They peer out at it, figure out what's going on, decide on some course of action, and enact it through the narrow interface of the keyboard or the data-glove, carefully monitoring the result to see if it turns out the way they expected. Our experience in the everyday world is not of that sort. There is no homunculus sitting inside our heads, staring out at the world through our eyes, enacting some plan of action by manipulating our hands,

and checking carefully to make sure we don't overshoot when reaching for the coffee cup. We inhabit our bodies and they in turn inhabit the world, with seamless connections back and forth. (2001, 102)

It seems unlikely that immersive virtual reality (VR) is *by its very nature* disconnected in this sense. Rather, it is just one more domain in which a skilled agent may act and perceive. But skill matters, and most of us are as yet unskilled in such situations. Moreover, the modes of sensing and interaction supported by current technologies often remain limited and clumsy, and this turns the user experience into that of a kind of alert game player rather than that of an agent genuinely located *inside* the virtual world.

It is worth noticing, however, that to the young human infant, the physical body itself may often share some of this problematic character. The infant, like the VR-exploring adult, must learn how to use initially unresponsive hands, arms, and legs to obtain its goals (for some detailed studies, see Thelen and Smith 1994). In so doing, the infant, like the Toddler robot, learns to make the most of the complex evolved morphology and passive dynamics of its own body. These have been selected so as to dramatically reduce the "gap" that needs to be bridged by the addition of energy and the imposition of control.

With time and practice, enough bodily fluency is achieved to make the wider world itself directly available as a kind of unmediated arena for embodied action. At this point, the extrabodily world becomes poised to present itself to the user not just as a problem space (though it is clearly that) but also as a problem-solving resource. For (as we'll see in more detail in chap. 2–4) the world, especially when encountered via inhabited interaction, is a place in which we can act fluently in ways that simplify or transform the problems that we want to solve. At such moments, the body has become "transparent equipment" (Heidegger 1927/1961): equipment (the classic example is the hammer in the hands of the skilled carpenter) that is not the focus of attention in use. Instead, the user "sees through" the equipment to the task in hand. When you sign your name, the pen is not normally your focus (unless it is out of ink etc.). The pen in use is no more the focus of your attention than is the hand that grips it. Both are transparent equipment.⁷

Doubtless, transparency of this kind may be achieved, with practice, without the large-scale exploitation of passive-dynamic effects.⁸ But one way in which evolved agents truly inhabit, rather than simply

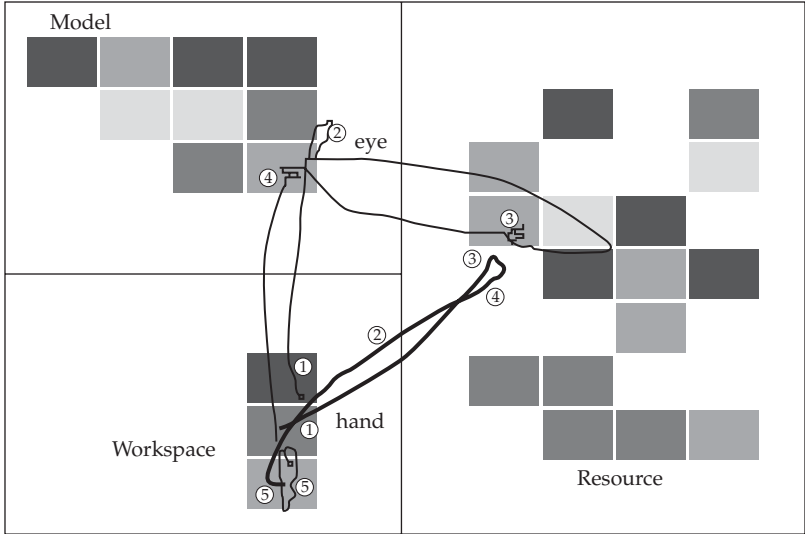


FIGURE 1.6 Copying a single block within the task. The eye-position trace is shown by the cross and the dotted line. The cursor trace is shown by the arrow and the dark line. The numbers indicate corresponding points in time for the eye and hand traces. (From Ballard et al. 2001, by permission)

control, their bodies may be usefully understood in terms of a profound fit between morphology and control. The kind of fit is exhibited by the wild walking systems devised by biological evolution and, in compelling microcosm, by autonomous, passive-dynamics-based walking robots.

1.3 Active Sensing

Suppose you were asked to solve the puzzle shown in figure 1.6. In this task (Ballard et al. 1997), 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 re-create the pattern by moving one block at a time from the reserve to the new version you are busy creating. The task is 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, and 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 (e.g., STRIPS—the Stanford Research Institute Problem Solver) as used by the early mobile robot Shakey; see Nilsson (1984) for a thorough retrospective review.

When asked how we would solve the problem, many of us pay lip service to this 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 (spontaneous scanning eye movements) to the model were used in the performance of the task, and 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. This confirmed 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, 734)

Two morals matter for the story at hand. The first is that visual fixation is here playing an identifiable computational role. As Ballard et al. (1997) comment, “Changing gaze is analogous to changing the memory reference in a silicon computer” (725). (These uses of fixation are thus described using the term “deictic pointers.”) 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” (732).

This is our first example of another important characteristic of embodied, embedded cognition, one that may be called the Principle of Ecological Assembly (PEA). According to the PEA, *the canny cognizer tends to recruit, on the spot, whatever mix of problem-solving resources will yield an acceptable result with a minimum of effort*. The PEA deliberately echoes Pfeifer and Scheier’s Principle of Ecological Balance (see sec. 1.1). Pfeifer and Scheier are, however, most interested in the slowly evolved match among sensory, motor, and neural capabilities and hence between the organismic bundle and its ecological niche. The PEA, by contrast, tracks a kind of near-instantaneous version of such overall balance: the balanced use of a set of potentially highly heterogeneous resources assembled on the spot to solve a given problem. Ecological balance of this latter kind is what a flexible ecological control system seeks to achieve (sec. 1.1).

It is important that, according to the PEA, the recruitment process marks no special distinction among neural, bodily, and environmental resources except insofar as these somehow affect the total effort involved. Though the principle itself seems obvious enough, it is actually far from obvious how best to unpack the notion of effort so as to make sense of the idea of trading off one kind of effort (e.g., recall from biological memory) against another very different kind of effort, such as the production of a head or eye motion that (let’s assume) retrieves the very same information. As our discussion progresses, we will encounter various attempts (see especially chap. 7 and 9) to make quantitative sense of this important but elusive notion of trade-offs among multiple heterogeneous sources of information and order.

1.4 *Distributed Functional Decomposition*

The Ballard et al. model is also our first example of an explanatory strategy that may usefully be called *distributed functional decomposition* (DFD). Distributed functional decomposition is a way of understanding

the capacities of supersized mechanisms (ones created by the interactions of biological brains with bodies and aspects of the local environment) in terms of the flow and transformation of energy, information, control, and where applicable, representations.⁹ The use of the term *functional* in distributed functional decomposition is meant to remind us that even in these larger systems, it is the *roles* played by various elements, and not the specific ways those elements are realized, that do the explanatory work. (This should not be contentious: Even in the case of Puppy's aluminum legs, it is not the material itself that matters as much as the slippage and give that it provides; sec. 1.1.) The goal, familiar enough from traditional internalist approaches, is thus to display some target performance as the outcome of an interacting multitude of unintelligent ("mechanical") interactions and effects but to do so relative to a larger organizational whole. (Imagine, to take a maximally simple case, an algorithm for addition that uses the agent's actual finger positions as a temporary storage buffer for key intermediate results.) Such approaches recognize the important contributions that embodiment and environmental embedding can make to the solution of a problem and then seek to understand those contributions by identifying the role of specific operations (perhaps some gross bodily, some environment involving, and some neural) in real-time performance of the task.

Ballard et al. explicitly recognize this element in their approach, commenting that their model "strongly suggests a functional view of visual computation where different operations are applied at different stages during a complex task" (1997, 735). 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 [integrating the two points by] introducing the idea that eye movements constitute a form of deictic coding... that allow perceivers to exploit the world as a kind of external storage device. (Wilson 2004, 176–177)

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.

Ballard et al. (1997) suggest using the term "the embodiment level" to indicate the level at which functionally critical operations occur at timescales of around one-third second. This corresponds, nonaccidentally, to the observed frequency of saccades and is, the

authors claim, the timescale at which “the natural sequentiality of body movements can be matched to the natural computational economies of sequential decision systems through a system of implicit reference (called *deictic*) in which pointing movements are used to bind objects in the world to cognitive programs” (723). Although this time frame is doubtlessly important, especially for the specific kinds of tasks the authors investigate, I here avoid the identification of (what’s computationally crucial about) embodiment with any specific temporal or spatial window. As we shall see later in the text, body and world play varied and crucial roles at many (often interacting) timescales.

1.5 *Sensing for Coupling*

Finally, it is worth pausing to reflect on the role of sensing in the Ballard et al. block-copying scenario. For sensing here plays an importantly different role to the one associated with classical planning and reasoning. In the classical model, the role of sensing is to get as much information into the system as is needed to solve the problem. For example, a planning agent might scan the environment to build up a problem-sufficient model of what’s out there and where it is located, at which point the reasoning engine can effectively throw away the world and operate instead upon the inner model, planning and then executing a response (perhaps checking now and then during execution to be sure that nothing has changed). In the block-copying scenario, by contrast, the agent does not use sensing to build up a rich inner model sufficient to solve the problem. Rather, sensing is used repeatedly, with the external scene functioning as an information store to be called upon just in time for the task fragment at hand. During all this, the external, screen-based model acts as “its own best model” (to adapt the famous usage from roboticist Rodney Brooks; see, e.g., Brooks 1991). Sensing here acts as a constantly available channel that productively couples agent and environment rather than as a kind of “veil of transduction” whereby world-originating signals must be converted into a persisting inner model of the external scene.

For an even more dramatic illustration of this possibility, consider the now-classic example of running to catch a fly ball in baseball. Giving perception its standard role, we might assume that the job of the visual system is to transduce information about the current position of the ball so as to allow a reasoning system to project its future trajectory. Here, too, however, nature looks to have found a more elegant and efficient

solution: You simply run so that the optical image of the ball appears to present a straight-line constant speed trajectory against the visual background (McBeath, Shaffer, and Kaiser 1995). This solution (the so-called LOT, for Linear Optical Trajectory, model) exploits a powerful invariant in the optic flow, discussed in Lee and Reddish (1981). There is, however, now some debate concerning the precise nature of the simple invariant we lock onto in solving this kind of problem.¹⁰ Thus, McLeod, Reed, and Dienes (2001, 2002) reported data that conflict with the predictions of the simple LOT model and that seem better predicted by an Optical Acceleration Cancellation (OAC) model first suggested by Chapman (1968). Shaffer et al. (2003) offer a mixed model combining uses of both strategies. For present purposes, however, the point is simply that the canny use of data available in the optic flow enables the catcher to sidestep the need to create a rich inner model to calculate the forward trajectory of the ball. In more recent work, multiple uses of the LOT approach seem to offer a better account of how dogs catch Frisbees, a more demanding task due to occasional dramatic fluctuations in the flight path (see Shaffer et al. 2004).

Important for present purposes, such strategies suggest (see also Maturana 1980) a very different role for the perceptual coupling itself. Instead of using sensing to get enough information inside, past the visual bottleneck, so as to allow the reasoning system to “throw away the world” and solve the problem wholly internally, they use the sensor as *an open conduit allowing environmental magnitudes to exert a constant influence on behavior*. Sensing is here depicted as the opening of a channel, with successful whole-system behavior emerging when activity in this channel is kept within a certain range. What is created is thus a kind of new, task-specific agent-world circuit. In such cases, as Randall Beer puts it, “the focus shifts from accurately representing an environment to continuously engaging that environment with a body so as to stabilize appropriate co-ordinated patterns of behavior” (2000, 97).

Interestingly, human subjects are typically unaware of their own deployment of such strategies. Shaffer and McBeath (2005) show that most people, including expert baseball fielders, think that they accurately perceive where the ball is located in physical space at each point in the unfolding trajectory, whereas the strategy actually used is unable, under most conditions, to reveal accurate ball-position information of this kind. That is, “observers seem to confuse or substitute their reasonably accurate semantic knowledge of the physical flight of the ball with the information that is optically available during projectile tracking tasks” (Shaffer and McBeath 2005, 1500).

Summing up the present section, we seem to confront what is really a whole spectrum of cases, ranging from the *classical extreme* (the use of perception to create a rich inner model sufficient to solve the problem) to many intermediate cases (e.g., the blocks-copying task where perception and ongoing bodily engagement are used repeatedly to retrieve and bind fragments of information just in time for use) to the (subjectively unobvious) *nonclassical extreme* (where perception opens a channel such that minimizing energetic variation within some fixed range can directly solve a problem). A third (partially overlapping) characteristic of embodied cognition can thus be added to our list: *The embodied agent is empowered to use active sensing and perceptual coupling in ways that simplify neural problem solving by making the most of environmental opportunities and information freely available in the optic array.*

1.6 Information Self-structuring

Embodied agents are also able to act on their worlds in ways that actively generate cognitively and computationally potent time-locked patterns of sensory stimulation. In this vein, Fitzpatrick et al. (2003; see also Metta and Fitzpatrick 2003), using both the COG and BABYBOT (fig. 1.7) platforms, show how active object manipulation (pushing and touching objects in view) can help generate information about object boundaries. The robot learns about the boundaries by poking and shoving. It uses motion detection to see its own hand–arm moving, but when the hand encounters and pushes an object, there is a sudden spread of motion activity. This cheap signature picks out the object from the rest of the environment.

In human infants, grasping, poking, pulling, sucking, and shoving create a rich flow of time-locked *multimodal* sensory stimulation. Such multimodal input streams have been shown (Lungarella, Sporns, and Kuniyoshi 2008; Lungarella and Sporns 2005) to aid category learning and concept formation. The key to such capabilities is the robot's 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



FIGURE 1.7 BABYBOT learns about object properties and affordances by poking and shoving. (From Metta and Fitzpatrick 2003, by permission)

linked to each other through sensorimotor loops. (Lungarella and Sporns 2005, 25)

An important implication of this focus on the active self-structuring of information flows is that timing (and especially, the time-locked unfolding of multimodal data streams) plays a major functional role in supporting learning and adaptive response. In work implemented on the famous COG robot (Brooks et al. 1999), Fitzpatrick and Arsenio (2004) show that the cross-modal binding of incoming signals that display common rhythmic signatures can aid a robot in learning about objects and, by including proprioception as a modality, about the nature of its own body. The robot first detects rhythmic patterns in the individual modalities (sight, hearing, and proprioception) and then deploys a binding algorithm to associate signals that display the same kind of periodicity. Courtesy of such bindings, COG can learn about its own body parts by binding visual, auditory, and proprioceptive signals. COG's arm is noisy in action, unlike our own, so when a human grabs and moves the robot's arm out of its field of vision it can bind sound and proprioceptive information. With the arm in view binding occurs across three modalities. Thus equipped, COG can even learn to identify

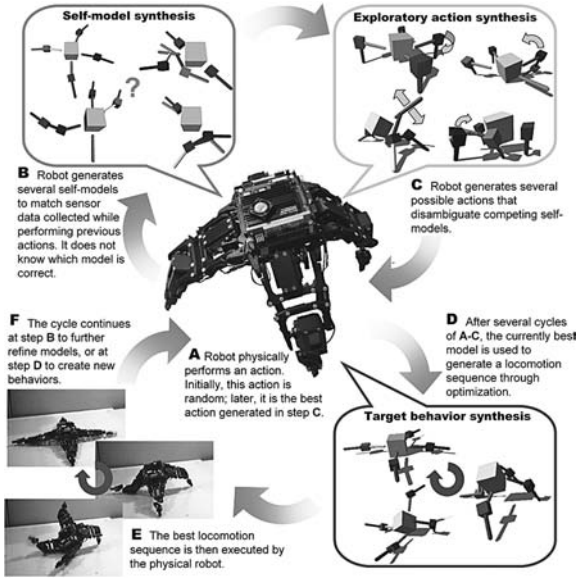


FIGURE 1.8 Outline of the algorithm. (From Josh Bongard, by permission)

its own arm with the moving image seen in a mirror. Summarizing this work, the authors write that

our work is an attempt to build a perceptual system which, from the ground up, focuses on timing just as much as content. This is powerful because timing is truly cross-modal, and leaves its mark on all the robot's senses no matter how they are processed and transformed. (Fitzpatrick and Arsenio 2004, 65)

Here, then, is a nice example of an approach that combines a bedrock computational and information-processing perspective with a potent functional role for timing and environmentally coupled action. We will meet this combination repeatedly in the chapters that follow. Such work depicts intelligent response as grounded in processes of information extraction, transformation, and use, while recognizing the key roles, *in those very processes*, played by timing, action, and coupled unfolding.

Information self-structuring may also play a key role in continuous self-modeling of the kind necessary to regain behavioral competence following bodily injury or change. Bongard, Zykov, and Lipson (2006) describe an algorithm (fig. 1.8) by which a robot continuously learns about its own bodily structure (morphology) by the ongoing generation of competing internal models that are tested by self-generated

actions. In brief, as the robot acts, it records the resulting sensory data and then generates a set (15, in the test case of a four-legged physical robot) of candidate models of its own morphology—models that would be broadly consistent with those data. It next (and this is the important part) finds an action (actuation pattern) that, when executed, will yield the greatest disagreement across the projected sensory consequences of the 15 candidate models. It then performs this action as part of an iterated cycle in which the robot learns about the possibly changing nature of its own body—for example, adapting to damage such as the loss of a limb or change such as the grasping of a tool (for more on this, see chap. 2). The key element in this process is, of course, the robot’s ability to actively produce the kinds of action that will yield the greatest information: a clear case of information self-structuring.

Finally, the active structuring of an information flow is also a potent *between-agent* tool, as demonstrated in striking studies by Yu, Ballard, and Aslin (2005). In these studies, a subject, fitted with eye tracker, head-mounted camera, microphone, and hand and body trackers describes, as if to a child (slowly, with clear enunciations) their current actions (see fig. 1.9). The verbal descriptions, along with the time-locked stream of multimodal training data recorded by the eye, head, hand, and body trackers, are fed to an artificial neural network. The task of the network is to learn visually grounded “meanings” for words for some actions solely by exposure to the time-locked stream of multimodal training data created by the active “caregiver.” In the presence of this critical active structuring, the net can learn image–sound associations using “raw” visual and auditory data (an unsegmented sound stream and an un-preprocessed video stream) and without the benefit of any inbuilt “language model.” The demonstration is compelling to watch as, from this raw but correlated data, the net learns generalizable image–sound pairings (e.g., it learns to produce phonetic strings such as “sta-pling” when shown new video recordings of the same action type). The net has simultaneously learned speech segmentation into meaningful units and “visually grounded meanings” for the units themselves. Key to this success is the information carried by the caregivers “embodied intentions”—that is, their use of eye and body movement to track and isolate salient aspects of the scene (the ones currently being verbally described) from the mass of co-occurring visual data. The added informational punch created by this active structuring of the training data transforms a daunting learning problem into one that is visibly tractable without massive prestructuring or much in the way of prior knowledge.



FIGURE 1.9 The associate training the computational model is wearing ASL eye tracker, CCD camera, microphone, and position sensors. The computational model thus shares multisensory information like a human language learner. This allows the association of coincident signals in different modalities. (From Yu, Ballard, and Aslin 2005, by permission)

In many ways, this is simply the flip side of the work on deictic pointing discussed in the previous section. Deictic pointing allows an agent to exploit the world as external storage. This work allows the learner to exploit another agent's use of deictic pointers (by tracking those very same eye fixations) as a kind of "gating mechanism that determines whether co-occurring data are relevant or not" (Yu, Ballard, and Aslin 2005, 994). As a result, social knowledge transmission is here supported by the very same kinds of embodied strategy (deictic uses of eye, head and body motions, and the active generation of time-locked data flows) that allow the individual learner to simplify her own problem solving and to learn about the world.

Here, then, is another way embodiment seems to matter to human cognition. It matters because *the presence of an active, self-controlled, sensing body allows an agent to create or elicit appropriate inputs, generating good data (for oneself and for others) by actively conjuring flows of multimodal, correlated, time-locked stimulation.* This trick promotes learning, bodily self-modeling, and categorization and may even (deep breath) hold out hope for grounded knowledge acquisition.

1.7 *Perceptual Experience and Sensorimotor Dependencies*

The appeal to action and to active sensing also lies at the heart of a recent, ambitious, and highly influential attempt to give an account of perception and of perceptual experience that centers upon what the agent (implicitly) knows about how sensory stimulation will vary as a result of change or movement.¹¹ This is in terms of our (implicit, nonconscious) knowledge or expectations concerning the many complex ways perceptual stimulations will morph and alter as we move our eyes, heads, and bodies. Such knowledge is dubbed (O'Regan and Noë 2001) “knowledge of sensorimotor dependencies” or of “sensorimotor contingencies”: It is knowledge of the relations between movement or change and resulting patterns of sensory stimulation.

Though superficially similar, this story about perception and perceptual experience goes (as we shall see in much more detail in chap. 8) well beyond the claims made by Ballard et al. (1997) or by most other proponents of so-called active perception (e.g., Churchland, Ramachandran, and Sejnowski 1994). For where the latter depict the active use of bodily motion and just-in-time retrieval as ploys that productively reconfigure the tasks to be performed by the brain and central nervous system, Noë (along with Hurley in press, and others) depicts the sensorimotor-expectation laden cycles as strongly constitutive of the perceptual experiences themselves. By strongly constitutive, I mean they assert a kind of identity such that sameness of active bodies of sensorimotor knowledge (knowledge of sensorimotor dependencies) is required for sameness of perceptual experience.

The central claim is thus that differences in what we perceptually experience correspond to differences in sensorimotor signatures (patterns of association between movements and the sensory effects of movement). If two things look different, they do so because, as we engage them in space and time, we bring to bear (rightly or wrongly) different sets of sensorimotor expectations. As our encounter proceeds, these expectations may or may not be validated. Crucially, it is this whole cycle of (implicit) expecting and subsequent sensory stimulation that is said to determine the content and character of any given perceptual experience. The expectations we have must differ as between, for example, a soccer ball and a rugby ball or an American football. Such differences underwrite the difference in experienced look. But despite such differences, for all *visually* presented objects, there will be some parts of the sensorimotor signatures in common. It is these commonalities that are said to make the experiences visual rather than, say, auditory. For

example, vision (unlike audition or touch) only samples the front or facing sides of objects and so on. The visual attributes of sensed objects are thus that subset of the signature sensorimotor contingencies that pertain to the distinctive ways that the visual sense can sample the real properties of objects. Thus, the very same real property (e.g., size) may be apprehended by vision or sometimes (for small objects) by touch. But the mode of sampling varies dramatically and with it the associated sensorimotor contingencies.

To visually perceive a square object, then, is to bring to bear a body of diverse practical knowledge concerning how movement of the eyes, head, or body would produce sensory change (new sensory inputs) as we inspect or interact with the object. An example is the way a leftward saccade would bring a certain (left-facing) shape of corner into central vision, while a rightward saccade would bring a different (right-facing) shape of corner into central vision. A rich body of such knowledge is said to constitute our visual perception of the square object. One upshot of all this, or so it is claimed, is that “what determines phenomenology is not neural activity set up by stimulation as such, but the way the neural activity is embedded in a sensorimotor dynamic” (Noë 2004, 227). For it is arguably the shape of a whole batch of sensorimotor loops that now determines the nature of the visual experience.

We can now formulate the next feature of recent work that I want to highlight: *attention to the possibility that the substrate (the “vehicles”) of specific perceptual experiences may involve whole cycles of world-engaging activity.*

1.8 Time and Mind

Approaches that foreground embodiment, active sensing, and temporally coupled unfoldings are sometimes rather starkly contrasted with (any or all of) functional, computational, information-processing, and information-theoretic approaches to the study of mind and cognition.¹² The proper explanatory tools, when confronted with apparently intrinsically embodied and richly temporal phenomena, are instead said to be the geometric constructs and differential equations of Dynamical Systems Theory (DST). This polarization (among dynamical and computational and information-theoretic approaches) is, I think, one of the less happy fruits of recent attempts to put brain, body, and world together again. I shall largely refrain (but see chap. 9) in the treatment that follows from re-rehearsing my rather liberal

views on the notions of representation, computation, and dynamical explanation. These views are quite well represented in previous work (especially Clark 1997a, 1997b, and 2001a). Instead, in a more positive vein, the various demonstrations, examples, and thought experiments that populate this book aim to reveal *computational, representational, information-theoretic, and dynamical approaches as deeply complementary elements in a mature science of the mind*. This emerging complementarity is the final feature of recent work that I want to highlight. But to very briefly motivate this more accommodating perspective, it may be worth just pausing to say a few words concerning time, dynamics, and computation (for a much more detailed treatment of these issues, see Clark 1997b).

One challenge that temporal considerations seem to pose to traditional forms of explanation and analysis is to account for cases of what I elsewhere (Clark 1997b) term continuous reciprocal causation. Continuous reciprocal causation (CRC) occurs when some system S is both continuously affecting and simultaneously being affected by activity in some other system O. Internally, we may well confront such causal complexity in the brain since many neural areas are linked by both feedback and feedforward pathways (e.g., Van Essen and Gallant 1994). On a larger canvass, we often find processes of CRC that criss-cross brain, body, and local environment. Think of a dancer, whose bodily orientation is continuously affecting and being affected by her neural states, and whose movements are also influencing those of her partner, to whom she is continuously responding! Or imagine playing improvised jazz in a small combo. Each musician's playing is influencing and being influenced by everyone else. CRC looks, in fact, to pervade the field of natural adaptive intelligence. The delicate dance of predator and prey or of mating animals exhibits the same complex causal structure.

Enter Dynamical Systems Theory. DST is a powerful framework for describing and understanding the temporal evolution of complex systems.¹³ In a typical explanation, the theorist specifies a set of parameters whose collective evolution is governed by a set of differential equations. Such equations always involve a temporal element, and in this way, timing is factored into the heart of the approach. Moreover, such explanations are easily able to span organism and environment. In such cases, the two components are treated as a coupled system in a specific technical sense; that is, the equation describing the evolution of each component contains a term that factors in the other system's current state (technically, the state variables of the first system are also the parameters of the second, and vice versa).

Thus, consider two wall-mounted pendulums placed in close proximity on a single wall. The two pendulums will tend (courtesy of vibrations running along the wall) to become swing synchronized over time. This process admits of an elegant dynamical explanation in which the two pendulums are analyzed as a single coupled system with the motion equation for each one including a term representing the influence of the other's current state (see Salzman and Newsome 1994). A useful way to think of this is by imagining two coevolving state spaces. Each pendulum traces a course through a space of spatial and temporal configurations. But the shape of this space is determined, in part, by the ongoing activity of the other pendulum, which is itself behaving in ways continuously modified by the action of its neighbor.

The crucial upshot of the emphasis on constant mutual interaction is a corresponding emphasis on what Van Gelder and Port (1995, 14) usefully term *total state*. Because we assume that there is widespread and complex interanimation among multiple systemic factors (x influences y and z , and x is itself influenced by y , which also influences z , etc.), the dynamicist chooses to focus on changes in total system state over time. The various geometric devices used to put intuitive flesh on the models (trajectories through state spaces populated by attractors, repellers, etc.; see Clark 2001a, chap. 7, for a brief introduction) thus reflect motion in a space of possible overall system states, with routes and distances defined relative to points each of which assigns a value to all the systemic variables and parameters. This emphasis on total state marks one of the deepest contrasts between (the purest of) dynamical and standard computationalist approaches, and it is both a boon and a burden. It is a boon insofar as it allows the dynamicist to respect the burgeoning complexity of causal webs in which everything (both inner and outer) is continuously influencing everything else. Relative to such cases, the mathematics of a system of interlocking differential equations can (at least in simple cases) accurately capture the way two or more systems engage in a continuous, real-time, and effectively instantaneous dance of mutual codetermining interaction.¹⁴ But it is a burden insofar as it threatens to obscure the specifically intelligence-based route to evolutionary success. That route involves the ability to become apprised of information concerning our surroundings and to use that information as a guide to present and future action. As soon as we embrace the notion of the brain as the principal (though not the *only*) seat of information-processing activity, we are already seeing it as fundamentally different from, say, the flow of a river or the activity of a volcano. And this difference needs to be reflected in our scientific analysis—a difference that typically is reflected when we pursue

the kind of information-processing model associated with computational approaches, but which threatens to be lost if we treat the brain, or any other systemic element engaged in information-based problem-solving activity, in exactly the same terms as the beating of a heart or the unfolding of a basic chemical reaction.¹⁵

The question, in short, is how to do justice to the idea that there is a principled distinction between knowledge-based and merely physical-causal systems. It does not seem likely that the dynamicist will deny that there is a difference (though hints of such a denial are occasionally found).¹⁶ But rather than responding by embracing a different vocabulary for the understanding and analysis of brain events (at least as they pertain to cognition), the dynamicist recasts the issue as the explanation of distinctive kinds of behavioral flexibility and hopes to explain that flexibility using the very same apparatus that works for other physical systems. Such an apparatus, however, may not be intrinsically well suited to explaining the particular way certain neural, and sometimes bodily and extrabodily, processes contribute to behavioral flexibility. This is because (a) it is unclear how it can do justice to the fundamental idea of information-guided choice, and (b) the emphasis on total state may obscure the kinds of rich structural variation especially characteristic of information-guided control systems.

Total state explanations do not fare well as a means of understanding systems in which complex information flow plays a key role. This is because such systems, as Sloman points out, typically depend on multiple, "independently variable, causally interacting sub-states" (1993, 80).¹⁷ Such systems support great behavioral flexibility by being able cheaply to alter the inner flow of information in a wide variety of ways. To understand the operation of a standard computational device, for example, we may appeal to multiple databases, procedures, and operations. The real power of the device consists in its ability to rapidly and cheaply reconfigure the way these components interact. Information-based control systems thus tend to exhibit a kind of complex articulation in which what matters most is the extent to which component processes may be rapidly decoupled and reorganized. This kind of articulation has been depicted as a pervasive and powerful feature of real neural processing.¹⁸ The fundamental idea is that large amounts of neural machinery are devoted not to the direct control of action but to the trafficking and routing of information within the brain. The point, for present purposes, is that to the extent that neural control systems exhibit such complex and information-based articulation (into multiple independently variable information-sensitive subsystems), the *sole* use of total state explanations would tend to obscure explanatorily impor-

tant details, such as the various ways in which substate x may vary independently of substate y and so on.

1.9 Dynamics and “Soft” Computation

The dynamicist should, at this point, reply that the dynamical framework really leaves plenty of room for the understanding of such variability. After all, any location in state space can be specified as a vector comprising multiple elements, and we may then observe how some elements change while others remain fixed and so on. This is true. But notice the difference between this kind of dynamical approach and the radical, total state vision introduced in section 1.8. If the dynamicist is forced (a) to give an information-based reading of various systemic substates and processes and (b) to attend as much to the details of the inner flow of information as to the evolution of total state over time, then it is unclear that we still confront a radical *alternative* to the computational story. Instead, what we seem to end up with is a very powerful and interesting hybrid: a kind of “dynamical computationalism” in which the details of the flow of information are every bit as important as the larger scale dynamics and in which some dynamical features lead a double life as elements in an information-processing economy. Indeed, we have already met one such case. The Ballard et al. model of the role of deictic pointing in the blocks-copying task story analyzed a cognitive task in part by using recognizable computational and information-processing concepts. But it also made coupling and fine temporal coordination crucial and thus applied those familiar computational and information-processing concepts to a larger, essentially embodied dynamic whole.¹⁹ Such work aims to display the *specific contributions* that embodiment and environmental embedding make by identifying what might be termed the *dynamic functional role* of specific bodily and worldly operations in the real-time performance of some task.²⁰

This kind of dynamical “soft” computationalism is surely attractive.²¹ Indeed, it is already the norm in many treatments that combine the use of dynamical tools with notions of internal representation and/or of neural computation (see, e.g., Spencer and Schöner 2003; Elman 1995, 2005). Thus, consider once again those complex loops of reciprocal causal influence. Let us assume for now that some such loop is fully internal and involves some relation of continuous reciprocal causal influence binding the activity of two elements. From this, it does not follow that we could not assign representational and (more broadly) information-processing roles either to the elements or to their coupled unfolding. It might be, for example, that the two elements are still best understood as trading in different kinds of

encoding or information, kinds that nonetheless mutually and continuously modify each other in some useful manner. We shall explore a concrete example of this involving a neural-bodily loop in chapter 6. There we examine a recent account of the role of physical gesture in the unfolding of thought and reason. According to that account, gesture and verbal thinking differ quite radically in the kinds of information they encode, but the gestural and verbal systems are nonetheless depicted as coupled in precisely the manner described earlier.²² In such cases, we need to understand *both* the distinctive individual contributions of the various coupled elements *and* the powerful effects that flow from their coupled unfolding.

It should be admitted, however, that the issues concerning continuous reciprocal causation, and the potential threat it poses to representationalist and computationalist modes of understanding, are complex ones. For *some* forms of CRC may indeed threaten such understandings. This will be so where the nature of the contributions being made by the “parts” is *itself* changing radically over time as a result of the multiple influences from elsewhere in the system.²³ At the extreme limit, such variability may undermine attempts to gloss stable types of systemic events as the bearers or vehicles of specific contents. It is an empirical question where, on this continuum of possibilities, biological information-processing lies (for some discussion, see Clark 1997a, 1997b; Wheeler 2005).

Short of this extreme limit, however, considerations concerning the importance of time and continuous reciprocal causation mandate not an outright rejection of the computational/representational vision²⁴ but rather the addition of a potent and irreducibly dynamical dimension. Such a dimension may manifest itself in several ways, including the use of dynamical tools to recover potential information-bearing states and processes from highly complex (and sometimes bodily and environmentally extended) webs of causal exchange; the recognition that intrinsically dynamical and temporal features may sometimes themselves play identifiable representational and/or computational roles; the (consequent) extension of standard computational ideas to include analog systems that change continuously in time and that exploit continuous state; and the recognition (sec. 1.6) of the importance of information self-structuring (e.g., via the active creation of time-locked flows of multimodal input) in learning and reasoning.

1.10 *Out from the Bedrock*

We have now scouted some of the most fundamental ways in which appeals to the body, to the environment, and to embodied action may

inform our vision and understanding of mind. Firm bedrock is provided by the wide suite of benefits enabled by the coevolution of morphology, materials, and control. Moving into the time frame of lifetime learning, we glimpsed related strategies of “ecological assembly” in which embodied agents exploit the opportunities provided by dynamic loops, active sensing, and iterated bouts of environmental exploitation and intervention. The next three chapters ramp up the complexity, exploring first the surprising lability and negotiability of human sensing and embodiment, then the transformative potential of material artifacts, language, and symbolic culture, and leading finally to the suggestion that *mind itself* leaches into body and world.

2

The Negotiable Body

2.1 *Fear and Loathing*

In a short article in the May 2004 edition of *WIRED* magazine (revelingly subtitled “Fear and Loathing on the Human–Machine Frontier”), the futurist and science fiction writer Bruce Sterling sounds an increasingly familiar alarm. After warning us of the imminent dangers of “brain augmentation,” he adds:

Another troubling frontier is physical, as opposed to mental, augmentation. Japan has a rapidly growing elderly population and a serious shortage of caretakers. So Japanese roboticists...envision walking wheelchairs and mobile arms that manipulate and fetch.

But there’s ethical hell at the interfaces. The peripherals may be dizzyingly clever gizmos...but the CPU is a human being: old, weak, vulnerable, pitifully limited, possibly senile. (116)

But such fears are rooted in a fundamentally misconceived vision of our own humanity: a vision that depicts us as “locked-in agents”—as beings whose minds and physical abilities are fixed quantities, apt (at best) for mere support and scaffolding by their best tools and technologies. In contrast to this view, I believe that human minds and

bodies are essentially open to episodes of deep and transformative restructuring in which new equipment (both physical and “mental”) can become quite literally incorporated into the thinking and acting systems that we identify as our minds and bodies (see, e.g., Clark 1997a, 2001b, 2003). In this chapter, I pursue this theme with special attention to the negotiability of our own embodiment.

It helps to start with the commonplace. Sensing and moving are the spots where the rubber of embodied agency meets the road of the wider world—the world outside the agent’s organismic boundaries. The typical human agent, circa 2008, feels herself to be a bounded physical entity in contact with the world through a variety of standard sensory channels, including touch, vision, smell, and hearing. It is a common observation, however, that the use of simple tools can lead to alterations in that local sense of embodiment. Fluently using a stick, we feel as if we are touching the world at the end of the stick, not (once we are indeed fluent in our use) as if we are touching the stick with our hand. The stick, it has sometimes been suggested, is in some way incorporated, and the overall effect seems more like bringing a temporary whole new agent-world circuit into being rather than simply exploiting the stick as a helpful prop or tool (see Merleau-Ponty 1945/1962 and Gibson 1979; for some more recent explorations of this theme, see Burton 1993; Reed 1996; Peck et al. 1996; Smitsman 1997; Hirose 2002; Maravita and Iriki 2004; Wheeler 2005).

In thinking about the case of stick-augmented perception, there would seem to be two key interfaces at play: the place where the stick meets the hand and the place where the extended system “biological agent + stick” meets the rest of the world. When we read about new forms of human–machine interface, we are again confronted by a similar duality and an accompanying tension. What makes such interfaces *appropriate* as mechanisms for human enhancement is, it seems, precisely their potential role in creating *whole new agent-world circuits*. But insofar as they succeed at this task, the new agent-tool interface itself fades from view, and the proper picture is one of an extended or enhanced agent confronting the (wider) world.

A good place to start, then, is with the notion of an interface itself.

2.2 *What’s in an Interface?*

Haugeland (1998) is, in part, an extended philosophical meditation on the very idea of an interface. The goal is to uncover the underlying principles “for dividing systems into distinct subsystems along

nonarbitrary lines" (211). According to Haugeland, the notions of component, system, and interface are all interdefined and interdefining. Components are those parts of a larger whole that interact through interfaces. Systems are "relatively independent and self-contained" composites of such interfaced components. And an interface itself is "a point of interactive 'contact' between components such that the relevant interactions are well-defined, reliable and relatively simple" (Haugeland 1998, 213).

Haugeland is right to point to the nature of interactions as the key to the location of an interface. We discern an interface where we discern a kind of regimented, often deliberately designed, point of contact between two or more independently tunable or replaceable parts. It does not seem correct, however, to insist that flow across the interface be simple. The idea here seems to be that we find genuine interfaces only where we find energetic or informational bottlenecks, as if an interface must be a narrow channel yielding what Haugeland describes as "low bandwidth" coupling. This is important for Haugeland's argumentative purpose because he means to show that human sensing typically yields very task-variable, high-bandwidth forms of agent-environment coupling and thus to argue that no genuine interface or interfaces separate agent and world. Instead (and see also the longer version of this claim already presented in the Introduction), there is said to be "intimate intermingling of mind, body and world" (Haugeland 1998, 224).

But although agreeing with Haugeland that sensing is at least sometimes best understood in terms of direct agent-environment couplings (as we saw in the previous chapter), his own conclusion that no genuine interfaces then link agent and world seems premature. Haugeland depicts these kinds of "open-channel" solutions as involving "tightly coupled high-bandwidth interaction" (223) and hence as inimical to the very idea of an agent-world interface.¹ But it seems intuitive that there can be genuine interfaces that support extremely high-bandwidth forms of coupling. Think, for example, of multiple computers linked into a network by means of superfast, very high-bandwidth "grid technologies."² There is really no doubt that we here confront a web of distinct intercommunicating component machines. Yet that web, in action, can sometimes function as a single unified resource. Nonetheless, we still think of it as a web of distinct but interfaced devices. And we do so not because the point of each machine's contact with the grid is narrow (it isn't) but because there exist, for each machine on the grid, very well-defined points of potential detachment and reengagement. We discern interfaces at the points at which one machine can be easily disengaged

and another engaged instead, allowing the first to join another grid or to operate in a stand-alone fashion. Grush (2003, 79) calls this the “plug points criterion” according to which “components are entities that can be plugged into, or unplugged from, other components and/or the system at large.”

An interface, I conclude, is indeed a point of contact between two items across which the types of performance-relevant interaction are reliable and well defined. But there is no requirement that such interfaces be narrow-bandwidth bottlenecks. The way to argue for cognitive extensions and blurrings of the mind-world boundary is not by casting doubt on the presence of genuine interfaces (there are plenty of these within the brain, too, and that doesn’t stop us from distinguishing parts and roles) but by displaying special features of the flow of information across those interfaces and by stressing the novel properties of the new systemic wholes that result. It is to these tasks that we now turn.

2.3 *New Systemic Wholes*

Biological systems, from lampreys to primates, display remarkable powers of bodily and sensory adaptability (see Mussa-Ivaldi and Miller 2003; Bach y Rita and Kerzel 2003; Clark 2003). The Australian performance artist Stelarc routinely deploys a “third hand,” a mechanical actuator controlled by Stelarc’s brain through commands to muscle sites on his legs and abdomen.³ Activity at these sites is monitored by electrodes that transmit signals (via a computer) to the artificial hand. Stelarc reports that, after some years of practice and performance, he no longer feels as if he has to actively control the third hand to achieve his goals. It has become “transparent equipment” (recall chap. 1), something through which Stelarc (the agent) can act on the world without first willing an action on anything else. In this respect, it now functions much as his biological hands and arms, serving his goals without (generally) being itself an object of conscious thought or effortful control.

Recent experimental work reveals more about the kinds of mechanisms that may be at work in such cases. A much publicized example is the work by Miguel Nicolelis and colleagues on a brain-machine interface (BMI) that allows a macaque monkey to use thought control to move a robot arm. In the most recent version of this work, Carmena et al. (2003) implanted 320 electrodes in the frontal and parietal lobes of a monkey. The electrodes allowed a monitoring computer to record neural activity across multiple cortical ensembles while the monkey learned to use a joystick to move a cursor across a computer screen

for rewards. As in previous work, the computer was able to extract the neural activity patterns corresponding to different movements, including direction and grip. Next, the joystick is disconnected. But the monkey is still able to use its neural activity, interpreted through the intervening computer, to directly control the cursor for rewards, and it learns to do so. Finally, these commands are diverted to a robot arm whose actual motions are then translated into on-screen cursor movements, including an on-screen equivalent of forceful gripping. This closes the loop. Instead of the monkey merely moving an unseen robot arm by thought control alone, the movement of the distant unseen arm now yields visual feedback in the form of on-screen cursor motion.

When the robot arm was inserted into the control loop, the monkey displayed a striking degradation of behavior. It took two full days of practice to reestablish fluent thought control over the on-screen cursor. The reason was that the monkey's brain now had to learn to factor in the mechanical and temporal "friction" created by the new physical equipment: It had to factor in the mechanical and dynamical properties of the robot arm and the time delays (which were substantial, in the 60–90 millisecond range) caused by interposing the motion of the arm between neural command and on-screen feedback. By the time full fluency was achieved, it is reasonable to conjecture that these properties of the still unseen distant arm were in some sense incorporated into the monkey's own body schema. In support of this, the experimenters were able to track real long-term physiological changes in the response profiles of frontoparietal neurons following use of the BMI, leading them to comment that

the dynamics of the robot arm (reflected by the cursor movements) become incorporated into multiple cortical representations...we propose that the gradual increase in behavioral performance...emerged as a consequence of a plastic reorganization whose main outcome was the assimilation of the dynamics of an artificial actuator into the physiological properties of fronto-parietal neurons. (Carmena et al. 2003, 205)

Creatures capable of this kind of deep incorporation of new bodily (and as we'll later see, also sensory and cognitive) structure are examples of what I shall call "profoundly embodied agents." Such agents are able constantly to negotiate and renegotiate the agent-world boundary itself.

Although our own capacity for such renegotiation is, I believe, vastly underappreciated, it really should come as no great surprise, given the facts of biological bodily growth and change. The human

infant must learn (by self-exploration) which neural commands bring about which bodily effects and must then practice until skilled enough to issue those commands without conscious effort. This process has been dubbed “body babbling” (Meltzoff and Moore 1997) and continues until the infant body becomes transparent equipment (see 1.6). Because bodily growth and change continue, it is simply good design not to permanently lock in knowledge of any particular configuration but instead to deploy plastic neural resources and an ongoing regime of monitoring and recalibration (for some excellent discussion, see Ramachandran and Blakeslee 1998).

2.4 *Substitutes*

As a second class of examples of recalibration and renegotiation, consider the plasticity revealed by work in sensory substitution. Pioneered in the ‘60s and ‘70s by Paul Bach y Rita and colleagues, the earliest such systems were grids of blunt “nails” fitted to the backs of blind subjects and taking input from a head-mounted camera. In response to the camera input, specific regions of the grid became active, gently stimulating the skin under the grid. At first, subjects report only a vague tingling sensation. But after wearing the grid while engaged in various kinds of goal-driven activity (walking, eating, etc.), the reports change dramatically. Subjects stop feeling the tingling on the back and start to report rough, quasi-visual experiences of looming objects and so forth. After a while, a ball thrown at the head causes instinctive and appropriate ducking. The causal chain is “deviant”: It runs via the systematic input to the back. But the nature of the information carried, and the way it supports the control of action, is suggestive of the visual modality. Performance using such devices can be quite impressive. In a recent article, Bach y Rita, Tyler, and Kaczmarek (2003) note that Tactile-Visual Substitution Systems (TVSS) have

been sufficient to perform complex perception and “eye”-hand co-ordination tasks. These have included face recognition, accurate judgment of speed and direction of a rolling ball with over 95% accuracy in batting the ball as it rolls over a table edge, and complex inspection-assembly tasks. (287)

The key to such effective sensory substitution is goal-driven motor engagement. It is crucial that the head-mounted camera be under the subject’s intentional motor control. This meant that the brain could, in effect, experiment through the motor system, giving commands that

systematically varied the input so as to begin to form hypotheses about what information the tactile signals might be carrying. Such training yields quite a flexible new agent-world circuit. Once trained in the use of the head-mounted camera, the motor system operating the camera could be changed (e.g., to a hand-held camera) with no loss of acuity. The touch pad, too, could be moved to new bodily sites, and there was no tactile–visual confusion: An itch scratched under the grid caused no “visual” effects (for these results, see Bach y Rita and Kerzel 2003).

Such technologies, though still experimental, are now increasingly advanced. The back-mounted grid is often replaced by a tongue-mounted coin-sized array and extensions in other sensory modalities. Bach y Rita and Kerzel (2003) give the nice example of a touch-sensor-rich glove that allows leprosy patients to begin to feel again using their hands. The patient is fitted with the glove that transmits signals to a forehead-mounted tactile disc array and rapidly reports feeling sensations of touch *at the fingertips*. This is presumably because the motor control over the sensors runs via commands to the hand, so the sensation is subsequently projected to that site. (See also the discussion of the auditory visual-substitution system known as The Voice in sec. 8.3.)

As an aside, it is worth noticing that the line between these kinds of rehabilitative strategy and wholly new forms of bodily and sensory enhancement is already thin to the point of nonexistence. There is advanced work on night-vision versions of sensory substitution, and at the more dramatic end of this spectrum, it is possible to bypass the existing sensory peripheries, feeding all manner of signals (including commercial TV!) directly to the cortex (see Bach y Rita and Kerzel 2003, and the discussion in Clark 2003, 125). Even without penetrating the existing surface of skin and skull, sensory enhancement and bodily extension are pervasive possibilities. One striking example (see Schrope 2001) is a U.S. Navy innovation known as a tactile flight suit. The suit (a kind of vest worn by the pilot) allows even inexperienced helicopter pilots to perform difficult tasks such as holding the helicopter in a stationary hover in the air. It works by generating bodily sensations (via safe puffs of air) inside the suit. If the craft is tilting to the right or left or forward or backward, the pilot feels a puff-induced vibrating sensation on that side of the body. The pilot’s own responses (moving in the opposite direction to correct the vibrations) can even be monitored by the suit to control the helicopter. The suit is so good at transmitting and delivering information in a natural and easy way that military pilots can use it to fly blindfolded. While the pilot wears the suit, the helicopter behaves very much like an extended body for the pilot: It rapidly links the pilot to the aircraft in the same kind of closed-loop

interaction that linked Stelarc and the third hand, the monkey and the robot arm, or the blind person and the TVSS system. What matters, in each case, is the provision of closed-loop signaling so that motor commands affect sensory input. What varies is the amount of training (and hence the extent of deeper neural changes) required to fully exploit the new agent-world circuits thus created.

It is important, in all these cases, that the new agent-world circuits be trained and calibrated in the context of a whole agent engaged in world-directed (goal-driven) activity. One sign of successful calibration is, as we noted earlier, that once fluency is achieved, the specific details of the (old or new) circuitry by which the world is engaged fall “transparent” in use. The conscious agent is then aware of the oncoming ball, not (usually) of *seeing* the ball or (by the same token) of using a *tactile substitution channel* to detect the ball. In just this way, the tactile-vest-wearing pilot becomes aware of the aircraft’s tilt and slant, not of *the puffs of air*.

In all these diverse ways, humans and other primates are revealed as constantly negotiable bodily platforms of sense, experience, and (as we’ll see in later chapters) reasoning, too. Such platforms are biologically primed so as to fluidly incorporate new bodily and sensory kit, creating brand new systemic wholes. This is just what one would expect of creatures built to engage in what we earlier (sec. 1.1) called “ecological control”: systems evolved so as to constantly search for opportunities to make the most of the reliable properties and dynamic potentialities of body and world.

2.5 *Incorporation Versus Use*

A very natural doubt to raise, at about this point, would be the following:

Critic: “You are making quite a song and a dance out of this, what with talk of brand new systemic wholes and so on. But we all know we can use tools and that we can learn to use them fluently and transparently. Why talk here of new systemic wholes, of extended bodies and reconfigured users, rather than just the same old user in command of a new tool?”

This is the right question to ask. We have already begun to see a hint of the answer in the quoted comments of Carmena et al. concerning the “assimilation of the dynamics of an artificial actuator into the physiological properties of fronto-parietal neurons.” To bring the key idea into focus, it helps next to consider a closely related body of research on tool use by primates.

Recent years have seen the discovery, in primate brains, of a variety of so-called bimodal neurons. These are “pre-motor, parietal and putaminal neurons that respond both to somatosensory information from a given body region (i.e., the somatosensory Receptive Field; sRF) and to visual information from the space (visual Receptive Field; vRF) adjacent to it” (Maravita and Iriki 2004, 79).

For example, some neurons respond to somatosensory stimuli (light touches) at the hand *and* to visually presented stimuli near the hand so as to yield an action-relevant coding of visual space. In a series of experiments, recordings were taken from bimodal neurons in the intraparietal cortex of Japanese macaques while the macaques learned to reach for food using a rake. The experimenters found that after just five minutes of rake use, the responses of some bimodal neurons whose original vRFs picked out stimuli near the hand had expanded to include the entire length of the tool, “as if the rake was part of the arm and forearm” (Maravita and Iriki 2004, 79). Similarly, other bimodal neurons, which previously responded to visual stimuli within the space reachable by the arm, now had vRFs that covered the space accessible by the arm-rake combination.⁴ After surveying a number of other related findings, including some fascinating work in which similar effects are observed after experience of reaching with a virtual arm in an on-screen display, Maravita and Iriki conclude: “Such vRF expansions may constitute the neural substrate of use-dependent assimilation of the tool into the body-schema, suggested by classical neurology” (2004, 80).

In human subjects suffering from unilateral neglect (in which stimuli from within a certain region of egocentrically coded space are selectively ignored), it has been shown that the use of a stick as a tool for reaching actually extends the area of visual neglect to encompass the space now reachable using the tool (see Berti and Frassinetti 2000). Berti and Frassinetti conclude that

the brain makes a distinction between “far space” (the space beyond reaching distance) and “near space” (the space within reaching distance) [and that]... simply holding a stick causes a remapping of far space to near space. In effect the brain, at least for some purposes, treats the stick as though it were a part of the body. (2000, 415)

The plastic neural changes reported by Carmena et al., and now further emphasized by Maravita and Iriki and by Berti and Frassinetti, suggest a real (philosophically important and scientifically well-grounded) distinction between true incorporation into the body schema and mere use. The body schema, it is important to note, is not the same

as the body image, though the two can sometimes be related. As I shall use the terms (see Gallagher 1998), the body image is a conscious construct able to inform thought and reasoning about the body. The body schema, by contrast, names a suite of neural settings that implicitly (and nonconsciously) define a body in terms of its capabilities for action, for example, by defining the extent of “near space” for action programs.⁵

We can certainly *imagine* tool users (perhaps even fluent tool users?) whose brains were not engineered so as to adapt the body schema in these ways. Such beings would *always* use tools the way we typically *begin* to use them: by roughly representing the tool and its features and powers (e.g., its length) and calculating effective uses accordingly. We can probably even imagine beings who were so fast and good at these calculations as to deploy the tools with the same skill and efficacy as an expert human agent. The contrast that would remain, even in the latter kind of case, would be between (a) the skilled agent’s first explicitly representing the shape, dimensions, and powers of the tool and then inferring (consciously or otherwise) that she can now reach such and such and do such and such and (b) agents whose brains were so constituted that experience with the tool results in, for example, a suite of altered vRFs such that objects within tool-augmented reaching range are now *automatically* treated as falling within near space. These are surely distinct strategies. The latter strategy might be especially recommended for beings whose bodies (like our own) are naturally subject to growth and change, as it seems designed to support genuine episodes of integration across change: cases that can now be defined as cases in which plastic neural resources become recalibrated (in the context of goal-directed whole agent activity) so as to automatically take account of new bodily and sensory opportunities. In this way, to paraphrase Varela, Thompson, and Rosch (1991), our own embodied activity *enacts or brings forth new systemic wholes*.

2.6 Toward Cognitive Extension

Could anything like this notion of incorporation (rather than mere use) and the consequent emergence of new systemic wholes get a grip in the more ethereal domain of mind and cognition? Could human *minds* be genuinely extended and augmented by cultural and technological tweaks, or is it (as many evolutionary psychologists, such as Pinker 1997, would have us believe) just the same old mind with a shiny new tool?

Here, the story is murkier by far. My own view, as will become increasingly clear, is that external and nonbiological information-processing

resources are *also* apt for temporary or long-term recruitment and incorporation rather than simply knowledge-based use (see Clark 1997a, 2003; Clark and Chalmers 1998). To whatever extent this holds, we are not just bodily and sensorily but also *cognitively* permeable agents. But whereas we can now begin to point, in the case of basic tool use, to the distinctive kinds of visible neural changes that accompany the genuine assimilation of tools or of new bodily structure, it is harder to know just what to look for in the case of mental and cognitive routines. For the present, we may look for some preliminary hints from the more basic case of physical and sensory augmentation and incorporation.

It may be helpful first to display the bare logical possibility of such cognitive extension. For even the bare possibility, some might feel, is ruled out by a simple argument to the effect that, as an anonymous journal referee once put it, “cognitive enhancement requires that the cognitive operations of the resource be intelligible to the agent.” If this were so, cognitive enhancement would always be in some clear sense superficial: It would provide tools while leaving the user fundamentally untouched. But the argument is flawed because the cognitive operations of much of my own brain (even those elements that mature later during development) are not thus intelligible to me, the conscious agent. Yet those operations surely help make me the cognitive agent I am. It also helps to reflect that biological brains must sometimes change and evolve by coordinating old activities and processes with new ones made available (e.g., by maturation and growth) courtesy of new or subtly altered structures. To insist that such change requires the literal intelligibility of the operations of the new by the old, rather than simply the emergence of appropriate integration and coordination, is to miss the potential for new wholes that are then *themselves* the determiners of what is and is not intelligible to the agent. It must thus be possible, at least in principle, for new *nonbiological* tools and structures to likewise become sufficiently well integrated into our problem-solving activity as to yield new agent-constituting wholes. What might such integration (genuine cognitive incorporation) require?

Consider the case when some existing neural system or systems learn a complex problem-solving routine that makes a variety of deep implicit commitments to the robust bioexternal availability of certain operations and/or bodies of information. This is the cognitive equivalent, I suggest, of the implicit commitments to details of bodily shape and potentials for action made (in the case of the rake) by rapidly retuning the receptive fields of key bimodal neurons and (in the case of the robot arm) by retuning key cortical representations (specifically, populations of frontoparietal neurons).

A quick (though frequently misused; see the critical discussion in sec. 7.3) illustration is provided by recent work on so-called change blindness. In this work (see Simons and Rensink 2005, for a balanced review), simple experimental manipulations, involving the masking of motion transients while various changes are made to a visually presented scene, reveal the surprising sparseness of the change-specifying information easily available to conscious reflection. Subjects seldom spot quite large and important changes, even when the changes are made in focal vision. Subjects are frequently amazed when they realize just how much has changed without their noticing it. How should we reconcile the limitations of such conscious change spotting with our strong sense of rich visual contact with our surroundings? Part of the answer (and see chap. 7 and 8 for more discussion) may be that the strong feeling of rich visual contact is really a reflection of something implicit in the larger overall problem-solving organization in which moment-by-moment vision merely participates. That larger organization “assumes” the (ecologically normal) ability to retrieve, via saccades or head and body movements, more detailed information as and when needed. Given such “availability on demand,” we feel (correctly, in an important sense) that we (qua agents engaged in knowledge-based interactions with the world) are fully in command of the detail (for this idea, see O’Regan and Nöe 2001; Clark 2002).

Or recall the use of visual fixation for binding in the block-copying task described in section 1.3. Here, the brain deploys a problem-solving routine that directly factors in the availability of certain types of information by certain types of embodied action. It is in just this way that nonbiological informational resources can become—either temporarily or more or less permanently—deeply incorporated into a subpersonally defined problem-solving whole. In such cases, a problem-solving routine is delicately geared to automatically exploit, on pretty much an equal footing, both internal and (bio)external forms of information storage.⁶ Rather than drawing a firm line around the inner encodings, we thus expand the relevant forms of storage and retrieval to include inner biological resources, environmental structure, and the data (and operations) made available by cognitive artifacts such as notebooks and laptops. As we move toward an era of wearable computing and ubiquitous information access, the robust, reliable information fields to which our brains delicately adapt their inner cognitive routines will surely become increasingly dense and powerful, perhaps further blurring the boundaries between the cognitive agent and his or her best tools, props and artifacts.⁷

2.7 *Three Grades of Embodiment*

We can now distinguish three grades of embodiment. Let's call them (simply if unimaginatively) mere embodiment, basic embodiment, and profound embodiment. A merely embodied creature or robot is one equipped with a body and sensors, able to engage in closed-loop interactions with its world, but for whom the body is nothing but a highly controllable means to implement practical solutions arrived at by pure reason. A basically embodied creature or robot would then be one (we saw several in chap. 1) for whom the body is not just another problem space, requiring constant micromanaged control, but is rather a resource whose own features and dynamics (of sensor placement, of linked tendons and muscle groups, etc.) could be actively exploited allowing for increasingly fluent forms of action selection and control. Much (though by no means all) work in contemporary robotics has explored this middle ground of modest embodiment. Such systems are, however, congenitally unable to learn new kinds of body-exploiting solution "on the fly," in response to damage, growth, or change. By contrast, as we have seen, biological systems (and especially we primates) seem to be specifically designed to constantly search for opportunities to make the most of body and world, checking for what is available, and then (at various timescales and with varying degrees of difficulty) integrating new resources very deeply, creating whole new agent-world circuits in the process. A profoundly embodied creature or robot is thus one that is highly engineered to be able to learn to make maximal problem-simplifying use of an open-ended variety of internal, bodily, or external sources of order.

Why describe this as profound embodiment rather than as a return to the outdated (or so many of us believe; see Clark 1997a, for a review) image of mind as a truly disembodied organ of control? The answer is that these kinds of minds are not in the least disembodied. Rather, they are *promiscuously* body-and-world exploiting. They are forever testing and exploring the possibilities for incorporating new resources and structures deep into their embodied acting and problem-solving regimes. They are, to use the jargon of Clark (2003), the minds of "natural-born cyborgs"—of systems continuously renegotiating their own limits, components, data stores, and interfaces. On this account, the body is both critically important and constantly negotiable. It is critically important as a key player on the problem-solving stage. It is not simply the point at which processes of transduction pass the real problems (now rendered in rich internal representational formats) to an inner engine of disembodied reason. Instead, much of our successful performance depends

on constant and subtle trade-offs among morphology, real-world action and opportunities, and neural control strategies. But this empowering body is constantly negotiable, constructed moment by moment from the flux of willed action and resulting sensory stimulation.

Those first waves of fear and loathing now give way to something more rewarding. Sterling (sec. 2.1) saw frightening scenes of a merely *superficially augmented* agent within whom “the CPU is a human being: old, weak, vulnerable, pitifully limited, possibly senile.” Such fears play upon a deeply misguided image of who and what we *already* are. They play upon an image of the human agent as doubly locked in: as a fixed mind (one constituted solely by a given biological brain) and as a fixed bodily presence in a wider world. Fortunately for us, human minds are not old-fashioned CPUs trapped in immutable and increasingly feeble corporeal shells. Instead, they are the surprisingly plastic minds of *profoundly* embodied agents: agents whose boundaries and components are forever negotiable and for whom body, sensing, thinking, and reasoning are all woven flexibly and repeatedly from the accommodating weave of situated, intentional action.