8 COGNITIVE TECHNOLOGY
Beyond the Naked Brain

8.1 Sketches

We have come a long way. From the initial image of the mind as a symbol-crunching meat machine, to the delights of vector coding and subsymbolic artificial intelligence, on to the burgeoning complexities of real-world, real-time interactive systems. As the journey continued one issue became ever more pressing: how to relate the insights gained from recent work in robotics, artificial life, and the study of situated cognition to the kinds of capacity and activity associated with so-called higher cognition? How, in short, to link the study of "embodied, environmentally embedded" cognition to the phenomena of abstract thought, advance planning, hypothetical reason, slow deliberation, and so on—the standard stomping grounds of more classical approaches.

In seeking such a link, there are two immediate options:

1. To embrace a deeply hybrid view of the inner computational engine itself. To depict the brain as the locus both of quick, dirty "on line," environment-exploiting strategies and of a variety of more symbolic inner models affording varieties of "off-line" reason.
2. To bet on the basic "bag-of-tricks" kind of strategy *all the way up*—to see the mechanisms of advanced reason as deeply continuous (no really new architectures and features) with the kinds of mechanisms (of dynamic coupling, etc.) scouted in the last two chapters.

In this final section, I investigate a third option—or perhaps it is really just a subtly morphed combination of the two previous options.

3. To depict much of advanced cognition as rooted in the operation of the same basic kinds of capacity used for on-line, adaptive response, but tuned and applied to the special domain of *external and/or artificial cognitive aids*—the domain, as I shall say, of *wideware or cognitive technology.*

It helps, at this point, to abandon all pretence at unbiased discussion. For the interest in the relations between mind and cognitive technology lies squarely at the heart of my own current research program, taking its cue from Dennett (1995, 1996), Hutchins (1995), Kirsh and Maglio (1994), and others.

The central idea is that mindfulness, or rather the special *kind* of mindfulness associated with the distinctive, top-level achievements of the human species, arises at the *productive collision points* of multiple factors and forces—some bodily, some neural, some technological, and some social and cultural. As a result, the project of understanding what is distinctive about human thought and reason may depend on a much broader focus than that to which cognitive science has become most accustomed, one that includes not just body, brain, and the natural world, but the props and aids (pens, papers, PCs, institutions) in which our biological brains learn, mature, and operate.

A short anecdote helps set the stage. Consider the expert bartender. Faced with multiple drink orders in a noisy and crowded environment, the expert mixes and dispenses drinks with amazing skill and accuracy. But what is the basis of this expert performance? Does it all stem from finely tuned memory and motor skills? By no means. In controlled psychological experiments comparing novice and expert bartenders (Beach, 1988, cited in Kirlik, 1998, p. 707), it becomes clear that expert skill involves a delicate interplay between internal and environmental factors. The experts select and array *distinctively shaped glasses* at the time of ordering. They then use these persistent cues so as to help recall and sequence the specific orders. Expert performance thus plummets in tests involving uniform glassware, whereas novice performances are unaffected by any such manipulations. The expert has learned to sculpt and exploit the working environment in ways that transform and simplify the task that confronts the biological brain.

Portions of the external world thus often function as a kind of extraneural memory store. We may deliberately leave a film on our desk to remind us to take it for developing. Or we may write a note "develop film" on paper and leave that on our desk instead. As users of words and texts, we command an especially cheap and potent means of off-loading data and ideas from the biological brain onto a variety of external media. This trick, I think, is not to be underestimated. For it affects not just the quantity of data at our command, but also the kinds of operation we can bring to bear on it. Words, texts, symbols, and diagrams often figure *intimately* in the problem-solving routines developed by biological brains nurtured in language-rich environmental settings. Human brains, trained in a sea of words and text, will surely develop computational strategies that directly "factor-in" the reliable presence of a wide variety of such external props and aids.

Take, for example, the process of writing an academic paper. You work long and hard and at days end you are happy. Being a good physicalist, you assume that all the credit for the final intellectual product belongs to your brain: the seat of human reason. But you are too generous by far. For what really happened was (perhaps) more like this. The brain supported some rereading of old texts, materials, and notes. While rereading these, it responded by generating a few fragmentary ideas and criticisms. These ideas and criticisms were then stored as more marks on paper, in margins, on computer discs, etc. The brain then played a role in reorganizing these data on clean sheets, adding new on-line reactions and ideas. The cycle of reading, responding, and external reorganization is repeated, again and again. Finally, there is a product. A story, argument, or theory. But this intellectual product owes a lot to those repeated loops out into the environment. Credit belongs to the embodied, embedded agent in the world. The naked biological brain is just a part (albeit a crucial and special part) of a spatially and temporally extended process, involving lots of extraneural operations, whose joint action creates the intellectual product. There is thus a real sense (or so I would argue) in which the notion of the "problem-solving engine" is really the notion of the *whole caboodle* (see Box 8.1): the brain and body operating within an environmental setting.

One way to understand the cognitive role of many of our self-created cognitive technologies is as affording *complementary* operations to those that come naturally to biological brains. Thus recall the connectionist image of biological brains as pattern-completing engines (Chapter 4). Such devices are adept at linking patterns of current sensory input with associated information: you hear the first bars of the song and recall the rest, you see the rat's tail and conjure the image of the rat. Computational engines of that broad class prove extremely good at tasks such as sensorimotor coordination, face recognition, voice recognition, etc. But they are not well suited to deductive logic, planning, and the typical tasks of sequential reason (see Chapters 1 and 2). They are, roughly speaking, "Good at Frisbee, Bad at Logic"—a cognitive profile that is at once familiar and alien: familiar, because human intelligence clearly has something of that flavor; alien, because we repeatedly transcend these limits, planning vacations, solving complex sequential problems, etc.

One powerful hypothesis, which I first encountered in McClelland, Rumelhart, Smolensky, and Hinton (1986), is that we transcend these limits, in large part, by combining the internal operation of a connectionist, pattern-completing device with a variety of external operations and tools that serve to reduce the complex, sequential problems to an ordered set of simpler pattern-completing operations of the kind our brains are most comfortable with. Thus, to take a classic illustration, we may tackle the problem of long multiplication by using pen, paper, and numerical symbols. We then engage in a process of external symbol manipulations and storage so as to reduce the complex problem to a sequence of simple pattern-completing steps that we already command, first multiplying 9 by 7 and storing the result on paper, then 9 by 6, and so on.

Box 8.1

The Talented Tuna

Consider, by way of analogy, the idea of a swimming machine. In particular, consider the bluefin tuna. The tuna is paradoxically talented. Physical examination suggests it should not be able to achieve the aquatic feats of which it is demonstrably capable. It is physically too weak (by about a factor of 7) to swim as fast as it does, to turn as compactly as it does, to move off with the acceleration it does, etc. The explanation (according to the fluid dynamicists M. and G. Triantafyllou) is that these fish actively create and exploit additional sources of propulsion and control in their watery environments. For example, the tuna use naturally occurring eddies and vortices to gain speed, and they flap their tails so as to actively create additional vortices and pressure gradients, which they then exploit for quick take-offs, etc. The real swimming machine, I suggest, is thus the fish *in its proper context*: the fish plus the surrounding structures and vortices that it actively creates and then maximally exploits. The *cognitive machine*, in the human case, looks similarly extended (see also Dennett, 1995, Chapters 12 and 13). We humans actively create and exploit multiple external media, yielding a variety of encoding and manipulative opportunities whose reliable presence is then factored deep into our problem-solving strategies. [The tuna story is detailed in Triantafyllou and Triantafyllou (1995) and further discussed in Clark (1997)].

The value of the use of pen, paper, and number symbols is thus that—in the words of Ed Hutchins, a cognitive anthropologist—

> [such tools] permit the [users] to do the tasks that need to be done while doing the kinds of things people are good at: recognizing patterns, modeling simple dynamics of the world, and manipulating objects in the environment. (Hutchins, 1995, p. 155)

A moments reflection will reveal that this description nicely captures what is best about *good* examples of cognitive technology: recent word-processing packages, web browsers, mouse and icon systems, etc. It also suggests, of course, what is wrong with many of our first attempts at creating such tools—the skills needed to use those environments (early VCR's, word-processors, etc.) were *precisely* those that biological brains find hardest to support, such as the recall and execution of long, essentially arbitrary, sequences of operations. See Norman (1999) for discussion.

It is similarly fruitful, I believe, to think of the practice of using words and linguistic labels as *itself* a kind of original "cognitive technology"—a potent add-on to our biological brain that literally transformed the space of human reason. We noted earlier the obvious (but still powerful and important) role of written in-

scriptions as both a form of external memory and an arena for new kinds of manipulative activity. But the very presence of words as *objects* has, I believe, some further, and generally neglected (though see Dennett, 1994, 1996), consequences. A word, then, on this further dimension.

Words can act as potent filters on the search space for a biological learning device. The idea, to a first approximation, is that learning to associate concepts with discrete arbitrary labels (words) makes it easier to use those concepts to constrain future search and hence enables the acquisition of a progressive cascade of more complex and increasingly abstract ideas. The claim (see also Clark and Thornton, 1997) is, otherwise put, that associating a perceptually simple, stable, external item (such as a word) with an idea, concept, or piece of knowledge effectively freezes the concept into a sort of cognitive building block—an item that can then be treated as a simple baseline feature for future episodes of thought, learning, and search.

This broad conjecture (whose statistical and computational foundations are explored in Clark and Thornton, 1997) seems to be supported by some recent work on chimp cognition. Thompson, Oden, and Boyson (in press) studied problem solving in chimps (*pan troglodytes*). What Thompson et al. show is that chimps trained to use an arbitrary plastic marker (a yellow triangle, say) to designate pairs of identical objects (such as two identical cups), and to use a different marker (a red circle, say) to designate pairs of different objects (such as a shoe and a cup), are then able to learn to solve a new class of abstract problems. This is the class of problems—intractable to chimps not provided with the symbolic training—involving recognition of *higher order* relations of sameness and difference. Thus presented with two (different) pairs of identical items (two shoes and two cups, say) the higher order task is to judge the pairs as exhibiting the *same* relation, i.e., to judge that you have two instances of *sameness*. Examples of such higher order judgments (which even human subjects can find hard to master at first) are shown in Table 8.1.

TABLE 8.1 Higher Order Sameness and Difference

Cup/Cup	**Shoe/Shoe**
=	two instances of first-order sameness
=	an instance of higher order sameness
Cup/Shoe	**Cup/Shoe**
=	two instances of first-order difference
=	an instance of higher order sameness
Cup/Shoe	**Cup/Cup**
=	one instance of first-order difference and one of first-order sameness
=	an instance of higher order difference

The token-trained chimps' success at this difficult task, it is conjectured, is explained by their experience with external tokens. For such experience may enable the chimp, on confronting, e.g., the pair of identical cups, to retrieve a mental representation of the *sameness* token (as it happens, a yellow triangle). Exposure to the two identical shoes will likewise cause retrieval of that token. At that point, the higher order task is effectively reduced to the simple, lower order task of identifying the two yellow plastic *tokens* as "the same."

Experience with external tags and labels thus enables the brain itself—by *representing* those tags and labels—to solve problems whose level of complexity and abstraction would otherwise leave us baffled—an intuitive result whose widespread applicability to human reason is increasingly evident (see Box 8.2). Learning a set of tags and labels (which we all do when we learn a language) is, we may thus speculate, rather closely akin to acquiring a new perceptual modality. For like a perceptual modality, it renders certain features of our world concrete and salient, and allows us to target our thoughts (and learning algorithms) on a new domain of basic objects. This new domain compresses what were previously complex and unruly sensory patterns into simple objects. These simple objects can then be attended to in ways that quickly reveal further (otherwise hidden) patterns, as in the case of relations between relations. And of course the whole process is deeply iterative—we coin new words and labels to concretize regularities that we could only originally conceptualize as a result of a backdrop of other words and labels. The most powerful and familiar incarnation of this iterative strategy is, perhaps, the edifice of human science itself.

The augmentation of biological brains with linguaform resources may also shed light on another powerful and characteristic aspect of human thought, an aspect mentioned briefly in the introduction but then abandoned throughout the subsequent discussion. I have in mind our ability to engage in second-order discourse, to think about (and evaluate) our own thoughts. Thus consider a cluster of powerful capacities involving self-evaluation, self-criticism, and finely honed remedial responses.[1] Examples would include recognizing a flaw in our own plan or argument and dedicating further cognitive efforts to fixing it; reflecting on the unreliability of our own initial judgments in certain types of situations and proceeding with special caution as a result; coming to see why we reached a particular conclusion by appreciating the logical transitions in our own thought; thinking about the conditions under which we think best and trying to bring them about. The list could be continued, but the pattern should be clear. In all these cases, we are ef-

[1]Two powerful treatments that emphasize these themes have been brought to my attention. Jean-Pierre Changeux (a neuroscientist and molecular biologist) and Alain Connes (a mathematician) suggest that self-evaluation is the mark of true intelligence—see Changeux and Connes (1995). Derek Bickerton (a linguist) celebrates "off-line thinking" and notes that no other species seems to isolate problems in their own performance and take pointed action to rectify them—see Bickerton (1995).

Box 8.2

Numerical Competence

Stanislas Dehaene and colleagues adduce a powerful body of evidence for a similar claim in the mathematical domain. Biological brains, they suggest, display an innate, but fuzzy and low-level numerical competence: a capacity to represent simple numerocity (1-ness, 2-ness, 3-ness), an appreciation of "more," "less," and of change in quantity. But human mathematical thought, they argue, depends on a delicate interplay between this innate system for low-grade, approximate arithmetic and the new cultural tools provided by the development of language-based representations of numbers. The development of such new tools began, they argue, with the use of body parts as stand-ins for the basic numerical quantities, and was progressively extended so as to provide a means of "pinning down" quantities for which we have no precise innate representation.

More concretely, Dehaene, Sperke, Pinel, Stanescu, and Triskin (1999) depict mature human arithmetical competence as dependent on the combined (and interlocking) contributions of two distinct cognitive resources. One is an innate, parietal lobe-based tool for approximate numerical reasoning. The other is an acquired, left frontal lobe-based tool for the use of language-specific numerical representations in exact arithmetic. In support of this hypothesis, the authors present evidence from studies of arithmetical reasoning in bilinguals, from studies of patients with differential damage to each of the two neural subsystems, and from neuroimaging studies of normal subjects engaged in exact and approximate numerical tasks. In this latter case, subjects performing the exact tasks show significant activity in the speech-related areas of the left frontal lobe, whereas the approximate tasks recruit bilateral areas of the parietal lobes implicated in visuospatial reasoning. These results are together presented as a demonstration "that exact calculation is language dependent, whereas approximation relies on nonverbal visuo-spatial cerebral networks" (Dehaene et al., 1999, p. 970) and that "even within the small domain of elementary arithmetic, multiple mental representations are used for different tasks" (Dehaene et al., 1999, p. 973). What is interesting about this case is that here the additional props and scaffolding (the number names available in a specific natural language) are rerepresented internally, so the process recruits images of the external items for later use. This is similar to the story about the chimps judgments about higher order relations, but quite unlike the case of artistic sketching that I consider later in the chapter.

fectively thinking about either our own cognitive profiles or about specific thoughts. This "thinking about thinking" is a good candidate for a distinctively human capacity—one not evidently shared by the other non-language-using animals who share our planet. As such, it is natural to wonder whether this might be an entire species of thought, in which language plays the generative role, that is not just reflected in, or extended by, our use of words but is directly dependent on language for its very existence.

It is easy to see, in broad outline, how this might come about. For as soon as we formulate a thought in words (or on paper), it becomes an object for both ourselves and for others. As an object, it is the kind of thing we can have thoughts about. In creating the object, we need have no thoughts about thoughts—but once it is there, the opportunity immediately exists to attend to it as an object in its own right. The process of linguistic formulation thus creates the stable structure to which subsequent thinkings attach. Just such a twist on the potential role of the inner rehearsal of sentences has been presented by Jackendoff (1996), who suggests that the mental rehearsal of sentences may be the primary means by which our own thoughts are able to become objects of further attention and reflection. The emergence of such second-order cognitive dynamics is plausibly seen as one root of the veritable explosion of varieties of external technological scaffolding in human cultural evolution. It is because we can think about our own thinking that we can actively structure our world in ways designed to promote, support, and extend our own cognitive achievements. This process also feeds itself, as when the arrival of written text and notation allowed us to begin to fix ever more complex and extended sequences of thought and reason as objects for further scrutiny and attention.

As a final example of cognitive technology (wideware) in action, let us turn away from the case of words and text and symbol-manipulating tools (PCs, etc.) and consider the role of sketching in certain processes of artistic creation. van Leeuwen, Verstijnen, and Hekkert (1999, p. 180) offer a careful account of the creation of abstract art, depicting it as heavily dependent on "an interactive process of imagining, sketching and evaluating [then resketching, reevaluating, etc.]." The question the authors pursue is, why the need to sketch? Why not simply imagine the final artwork "in the mind's eye" and then execute it directly on the canvas? The answer they develop, in great detail and using multiple real case studies, is that human thought is constrained, in mental imagery, in some very specific ways in which it is *not* constrained during on-line perception. In particular, our mental images seem to be more interpretively fixed: less enabling of the discovery of novel forms and components. Suggestive evidence for such constraints includes the intriguing demonstration [Chambers and Reisberg (1985)—see Box 8.3] that it is much harder to discover the second interpretation of an ambiguous figure in recall and imagination than when confronted with a real drawing. It is quite easy, by contrast, to compose imagined elements into novel wholes—for example, to imag-

Box 8.3

Imaginative versus Perceptual "Flipping" of Ambiguous Images

Chambers and Reisberg (1985) asked subjects (with good imagistic capacities) to observe and recall a drawing. The drawing would be "flippable"—able to be seen as either one of two different things, though not as both at once. Famous examples include the duck/rabbit (shown below), the old lady/young lady image, the faces/vase image, and many others.

The experimenters chose a group of subjects ranged across a scale of "image vividness" as measured by Slee's Visual Elaboration scale (Slee, 1980). The subjects, who did not already know the duck/rabbit picture, were trained on related cases (Necker cubes, face/vase pictures) to ensure that they were familiar with the phenomenon in question. They were briefly shown the duck/rabbit and told to form a mental picture so that they could draw it later. They were then asked to attend to their mental image and to seek an alternative interpretation for it. Hints were given that they should try to shift their visual fixation from, e.g., lower left to upper right. Finally, they were asked to draw their image and to seek an alternative interpretation of their drawing. The results were surprising.

> Despite the inclusion of several "high vividness" imagers, none of the 15 subjects tested was able to reconstrue the imaged stimulus. . . . In sharp contrast, all 15 of the subjects were able to find the alternate construal in their own drawings. This makes clear that the subjects did have an adequate memory of the duck/rabbit figure and that they understood our reconstrual task. (Chambers and Reisberg, 1985, p. 321)

The moral, for our purposes, is that the subject's problem-solving capacities are significantly extended by the simple device of externalizing information (*drawing* the image from memory) and then confronting the external trace using on-line visual perception. This "loop into the world" allows the subject to find new interpretations, an activity that (see text) is plausibly central to certain forms of artistic creation. Artistic intelligence, it seems, is not "all in the head."

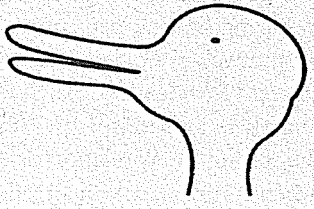

Figure 8.1

inatively combine the letters D and J to form an umbrella ☂ (see Finke, Pinker, and Farah, 1989).

To accommodate both these sets of results, van Leeuwen et al. suggest that our imaginative (intrinsic) capacities do indeed support "synthetic transformations" in which components retain their shapes but are recombined into new wholes (as in the J + D = umbrella case), but lack the "analytic" capacity to decompose an imagined shape into wholly new components (as in the hourglasses-into-overlapping parallelograms case shown in Figure 8.2). This is because (they speculate) the latter type of case (but not the former) requires us to first undo an existing shape interpretation.

Certain forms of abstract art, it is then argued, depend heavily on the deliberate creation of "multilayered meanings"—cases in which a visual form, on continued inspection, supports multiple different structural interpretations (see Figure 8.3). Given the postulated constraints on mental imagery, it is likely that the discovery of such multiply interpretable forms will depend heavily on the kind of trial-and-error process in which we first sketch and then perceptually (not imaginatively) reencounter the forms, which we can then tweak and resketch so as to create an increasingly multilayered set of structural interpretations.

Thus understood, the use of the sketchpad is not just a convenience for the artist, nor simply a kind of external memory, or durable medium for the storage of particular ideas. Instead, the iterated process of externalizing and reperceiving is integral to the process of artistic cognition itself. A realistic computer simulation of the way human brains support this kind of artistic creativity would need likewise to avail itself of one (imaginative) resource supporting synthetic transformations and another, environmentally looping resource, to allow its on-line perceptual systems to search the space of "analytic" transformations.

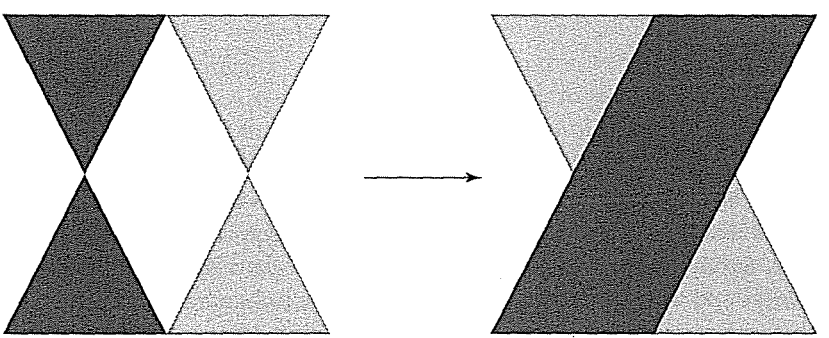

Figure 8.2 Novel decomposition as a form of analytic transformation that is hard to perform in imagery. The leftmost figure, initially synthesized from two hourglasses, requires a novel decomposition to be seen as two overlapping parallelograms. [Reproduced from van Leeuwen et al. (1999) by kind permission of the authors and the publisher, University Press of America.]

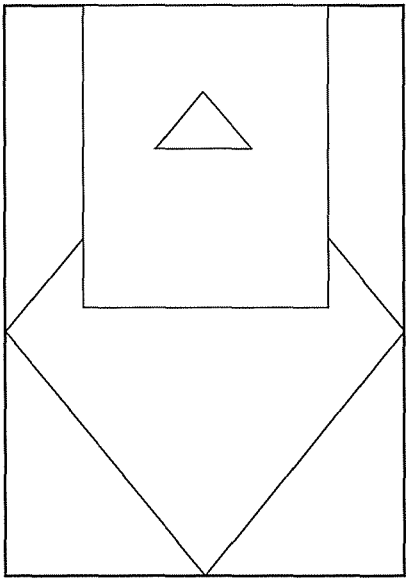

Figure 8.3 A simple example of the kind of multilayered structure found in certain types of abstract art. [Reproduced from van Leeuwen et al. (1999) by kind permission of the authors and the publisher, University Press of America.]

The conjecture, then, is that one large jump or discontinuity in human cognitive evolution involves the distinctive way human brains repeatedly create and exploit wideware—various species of cognitive technology able to expand and reshape the space of human reason. We, more than any other creature on the planet, deploy nonbiological wideware (instruments, media, notations) to *complement* our basic biological modes of processing, creating extended cognitive systems whose computational and problem-solving profiles are quire different from those of the naked brain.

8.2 Discussion

A. THE PARADOX OF ACTIVE STUPIDITY (AND A BOOTSTRAPPING SOLUTION)

The most obvious problem, for any attempt to explain our distinctive smartness by appeal to a kind of symbiosis of brain and technology, lies in the threat of circularity. Surely, the worry goes, only intrinsically *smart* brains could have the knowledge and wherewithal to create such cognitive technologies in the first place. All that wideware cannot come from nowhere. This is what I shall call the "paradox of active stupidity."

There is surely something to the worry. If humans are (as I have claimed) the only animal species to makes such widespread and interactive use of cognitive technologies, it seems likely that the explanation of this capacity turns, in some way,

on distinctive features of the human brain (or perhaps the human brain and body; recall the once-popular stories about tool use and the opposable thumb). Let us be clear, then, that the conjecture scouted in the present chapter is not meant as a denial of the existence of certain crucial neural and/or bodily differences. Rather, my goal is to depict any such differences as the *seed*, rather than the full explanation, of our cognitive capabilities. The idea is that some relatively *small* neural (or neural/bodily) difference was the spark that lit a kind of intellectual forest fire. The brain is, let us assume, wholly responsible (courtesy, perhaps of some quite small tweak of the engineering) for the fulfillment of some precondition of cultural and technological evolution. Thus Deacon (1997) argues that human brains, courtesy of a disproportionate enlargement of our prefrontal lobes relative to the rest of our brains, are uniquely able to learn rich and flexible schemes associating arbitrary symbols with meanings. This, then, is one contender for the neural difference that makes human language acquisition possible, and language (of that type) is, quite plausibly, the fundamental "cognitive technology" (the UR-technology) that got the whole ball rolling. There are many alternative explanations [an especially interesting one, I think, is to be found in Fodor (1994)].[2] But the point is that once the process of cultural and technological evolution is under way, the explanation of our contemporary human achievements lies largely in a kind of iterated bootstrapping in which brains and (first-generation) cognitive technologies cooperate so as to design and create the new, enriched technological environments in which (new) brains and (second-generation) cognitive technologies again conspire, producing the third-generation environment for another set of brains to learn in, and so on.

This idea of a potent succession of cognitive technologies is especially suggestive, I believe, when combined with the (still speculative) neuroscientific perspective known as neural contructivism. The neural contructivist (see Box 8.4) stresses the role of developmental plasticity in allowing the human cortex to actively build and structure itself in response to environmental inputs. One possible result of such a process is to magnify an effect I call "cognitive dovetailing." In cognitive dovetailing, neural resources become structured so as to factor reliable *external* resources and operations into the very heart of their problem-solving routines. In this way, the inner and outer resources come to complement each other's operations, so that the two fit together as tightly as the sides of a precisely dovetailed joint. Thus think, for example, of the way the skilled bartender (see text) combined biological recall and the physical arrangement of differing shaped glasses to solve the cocktail bar problem, or the way the tuna (Box 8.1) swims by creating aquatic

[2]Fodor (1994) locates the principal difference in the capacity (which he thinks is unique to humans) to become aware of the contents of our own thoughts: to not just think that it is raining, but to know that "it is raining" is the content of our thought. This difference could, Fodor argues, help explain our unique ability to actively structure our world so as to be reliably caused to have true thoughts—the central trick of scientific experimentation.

Box 8.4

NEURAL CONSTRUCTIVISM

The neural constructivist depicts neural (especially cortical) growth as experience—dependent, and as involving the actual construction of new neural circuitry (synapses, axons, dendrites) rather than just the fine-tuning of circuitry whose basic shape and form are already determined. The result is that the learning device *itself* changes as a result of organism–environmental interactions—learning does not just alter the knowledge base, it alters the computational architecture itself. Evidence for the neural constructivist view comes primarily from recent neuroscientific studies (especially work in developmental cognitive neuroscience). Key studies include work involving cortical transplants, in which chunks of visual cortex were grafted into other cortical locations (such as somatosensory or auditory cortex) and proved plastic enough to develop the response characteristics appropriate to the new location (see Schlagger and O'Leary, 1991; Roe et al., 1990). There is also work showing the deep dependence of specific cortical response characteristics on developmental interactions between parts of cortex and specific kinds of input signal (Chenn et al., 1997) and a growing body of constructivist work in artificial neural networks: connectionist networks in which the architecture (number of units and layers, etc.) itself alters as learning progresses—see, e.g., Quartz and Sejnowski (1997). The take home message is that immature cortex is surprisingly homogeneous, and that it "requires afferent input, both intrinsically generated and environmentally determined, for its regional specialization" (Quartz, 1999, p. 49). It is this kind of profound plasticity that best underscores the very strongest version of the dovetailing claim made in the text.

vortices that it then exploits. Now picture the young brain, learning to solve problems in an environment packed with pen, paper, PC, etc. That brain may develop problem-solving strategies that factor in these props just as the bartender's brain factors in the availability of differently shaped glasses to reduce memory load. What this suggests, in the rather special context of the neural constructivist's (see Box 8.4) developmental schema, is that young brains may even develop a kind of cortical *architecture* especially suited to promoting a symbiotic problem-solving regime, in which neural subsystems, pen, paper, and PC-based operations are equal partners, performing complementary and delicately orchestrated operations.

The neural constructivist vision thus supports an especially powerful version of the story about cognitive technological bootstrapping. If neural constructivism

is true, it is not just that basic biological brains can achieve more and more as the technological surround evolves. It is that the biological brain literally grows a cortical cognitive architecture suited to the specific technological environment in which it learns and matures. This symbiosis of brain and cognitive technology, repeated again and again, but with new technologies sculpting new brains in different ways, may be the origin of a golden loop, a virtuous spiral of brain/culture influence that allows human minds to go where no animal minds have gone before.

B. CASH VALUE

Some will argue that there is nothing new or surprising in the simple observation that brains plus technology can achieve more than "naked brains." And even the radical "dovetailing" image, in which brains plus reliable props come to act as integrated problem-solving ensembles may seem to have few practical implications for the cognitive scientific project. What, then, is the cash value of treating the human mind as a complex system whose bounds are not those of skin and skull?

One practical, but wholly negative, implication is that there can be no single "cognitive level" (recall Chapter 2) at which to pitch all our investigations, nor any uniquely bounded system (such as the brain) to which we can restrict our interest (*qua* cognitive scientists seeking the natural roots of thought and intelligence). To understand the bartender's skills, for example, we cannot restrict our attention to the bartender's brain; instead we must attend to the problem-solving contributions of active environmental structuring. Nonetheless, it is unrealistic to attempt—in general—to take everything (brain, body, environment, action) into account all at once. Science works by simplifying and focusing, often isolating the contributions of the different elements. One genuine methodological possibility, however, is to use alternate means of focusing and simplifying. Instead of simplifying by dividing the problem space (unrealistically, I have argued) into brain–science, body–science, and culture–science, we should focus (where possible) on the interactions. To keep it tractable we can focus on the interactions in small, idealized cases in which the various elements begin to come together. Work in simple real-world robotics (such as the robot cricket discussed in Chapter 6) provides one window onto such interactive dynamics. Another useful tool is the canny use of multiscale simulations: representative studies here include work that combines artificial evolution with individual lifetime learning in interacting populations (Ackley and Littman, 1992; Nolfi and Parisi, 1991), work that investigates the properties of very large collections of simple agents (Resnick, 1994), and work that targets the relations between successful problem solving and the gradual accumulation of useful environmental props and artifacts (Hutchins, 1995; Hutchins and Hazelhurst, 1991).

The cash value of the emphasis on extended systems (comprising multiple heterogeneous elements) is thus that it forces us to attend to the interactions them-

selves: to see that much of what matters about human-level intelligence is hidden not in the brain, nor in the technology, but in the complex and interated interactions and collaborations between the two. (The account of sketching and artistic creation is a nice example of the kind of thing I have in mind: but the same level of interactive complexity characterizes almost all forms of advanced human cognitive endeavor.) The study of these interaction spaces is not easy, and depends both on new multidisciplinary alliances and new forms of modeling and analysis. The pay-off, however, could be spectacular: nothing less than a new kind of cognitive scientific collaboration involving neuroscience, physiology, and social, cultural, and technological studies in about equal measure.

C. THE BOUNDS OF SELF

One rather problematic area, for those of us attracted to the kind of extended systems picture presented above, concerns the notions of self and agency. Can it be literally true that the physical system whose whirrings and grindings constitute *my* mind is a system that includes (at times) elements and operations that loop outside my physical (biological) body? Put dramatically, am I a dumb agent existing in a very smart and supportive world, or a smart agent whose bounds are simply not those of skin and skull? This is a topic that I have addressed elsewhere (see Clark and Chalmers, 1998), so I shall restrict myself to just a few points here.

We can begin by asking a simple question. Why is it that when we use (for example) a crane to lift a heavy weight, we (properly) do not count the crane as increasing our individual muscle power, whereas when we sit down to fine-tune an argument, using, paper, pen, and diagrams, we are less prone to later "factor out" the contributions of the props and tools and tend to see the intellectual product as purely the results of *our* efforts? My own view, as suggested in the text, is that one difference lies in the way neural problem-solving processes are *themselves* adapted to make deep and repeated use of the cognitive wideware. Another lies, perhaps, in the looping and interactive nature of the interactions themselves. The crane driver and the crane each makes a relatively *independent* contribution to lifting the girders, whereas the patterns of influence linking the artist and the sketches seems significantly more complex, interactive, and reciprocal. It is perhaps no accident that it is in those cases in which the patterns of reciprocal influence uniting the user and tool are most mutually and continuously modulatory (the racing driver and car, windsurfer and rig, etc.) that we are most tempted, in everyday discourse, to speak of a kind of agent–machine unity.

The main point to notice, in any case, is just that the issues here are by no means simple. Consider another obvious worry, that the "extended system" picture, *if* it is meant to suggest (which it need not) a correlative *mental* extension, leads rapidly to an absurd inflation of the individual mind. The worry (discussed in length in Clark and Chalmers, 1998) is thus that allowing (to take the case from

Box 8.5

Cyborgs and Software Agents

Two kinds of technological advance seem ready to extend human mindfulness in radically new kinds of ways.

The first, already familiar but rapidly gaining in ubiquity and sophistication, is exemplified by so-called software agents. A simple example of a software agent would be a program that monitors your on-line reading habits, which newsgroups you frequently access, etc., or your on-line CD buying habits, and then searches out new items that fit your apparent interests. More sophisticated software agents might monitor on-line auctions, bidding and selling on your behalf, or buy and sell your stocks and shares.

Reflect on the possibilities. Imagine that you begin using the web at age 4. Dedicated software agents track and adapt to your emerging interests and random explorations. They then help direct your attention to new ideas, web pages, and products. Over the next 70 years you and your software agents are locked in a complex dance of coevolutionary change and learning, each influencing, and being influenced by, the other. In such a case, in a very real sense, the software entities look less like part of your problem-solving environment than part of you. The intelligent system that now confronts the wider world is biological-you-plus-the-software-agents. These external bundles of code are contributing rather like the various subpersonal cognitive functions active in your own brain. They are constantly at work, contributing to your emerging psychological profile. Perhaps you finally count as "using" the software agents only in the same attenuated and ultimately paradoxical way that you count as "using" your hippocampus or frontal lobes?

Whereas dedicated, coevolving software resources are extending individual cognitive systems outside the local bounds of skin and skull, various forms of bioelectronic implant seem ready to transform the computational architecture from within the biological skin-bag itself. Perceptual input systems are already the beneficiaries of restorative technologies involving the direct linkage of implanted electronics to biological nerves and neurons. Cochlear implants, some of which now bypass the auditory nerve and jack directly into the brain stem (see LeVay, 2000), already help the deaf, and experimental retinal implants are now ready to offset certain causes of adult blindness, such as age-related macular degeneration. The next step in our cyborg future must be to link such implanted electronics evermore directly to the neural systems involved in reason, recall, and imagination. Such a step is already being taken, albeit in a crude and avowedly exploratory way, by

pioneers such as Kevin Warwick, a Reading University professor of Cybernetics. Warwick is experimenting with implants interfacing nerve bundles in his body to a digital computer able to record, replay, and share (via similar implants in others) the signals (see Warwick, 2000). We might imagine, indeed, that the artist's sketchpad, displayed (see text) as a critical external loop in certain processes of artistic creation may one day be replaced, or complemented, by implanted technologies enabling us to deploy our normal perceptual abilities on a kind of secondary visual display, opening the door to an even more powerful symbiosis between biological capacities and the artifactual (but now internalized) support.

In short, human mindfulness is set fast on an explosive trajectory, annexing more and more external and artifactual structures as integral parts of the cognitive machine, while simultaneously reinventing itself from within, augmenting on-board biological systems with delicately interfaced electronics. Just *who* we are, *what* are we, and *where* we are must count among the prime cultural, scientific, and moral puzzles facing the next generations of human (?) life.

the text) the sketchpad operations to count as part of the artist's own mental processes leads inevitably to, e.g., counting the database of the *Encyclopedia Britannica*, which I keep in my garage, as part of my general knowledge. Such intuitively pernicious extension ("cognitive bloat") is not, however, inevitable. It is quite proper to restrict the props and aids that can count as part of *my* mental machinery to those that are, at the very least, reliably available when needed and used (accessed) as automatically as biological processing and memory. Such simple criteria may again allow the incorporation of the artist's sketchpad and the blind-person's cane while blocking the dusty encyclopedia left in the garage. And they positively invite mind-extending depictions of possible future technologies: the cyberpunk neural implant that allows speed-of-thought access to the *Encyclopedia Britannica* database, not to mention the cochlear and retinal implants that already exist and are paving the way for future, more cognitively oriented, kinds of biotechnological explorations (see Box 8.5).

The cyberpunk cases can be misleading, however, for they may seem to support the idea that once equipment lies *inside* the bounds of skin and skull, it can count as part of the physical basis of individual mind, *but not a moment before.* This seems unprincipled. If a functional copy of the implant was strapped to my belt, or carried in my hand, why should *that* make the difference? Easy availability and automatic deployment seem to be what really matter here. Being part of the biological brain pretty well ensures these key features. But it is at most a sufficient, and not a necessary, condition.

Box 8.6

Cognitive Rehabilitation

Consider, as a kind of coda, a case brought to my attention by Carolyn Baum, head of Occupational Therapy at the Washington University School of Medicine. Baum had been puzzled by the capacity of certain Alzheimer's sufferers to live alone in the community, maintaining a level of independent functioning quite out of step with their scores on standard tests designed to measure their capacity to live independently. The puzzle was resolved when Baum and her coworkers (see, e.g., Baum, 1996) observed these patients in their home environments. The environments turned out to be chock full of props and scaffolding able to partially offset the neural deficiency: rooms might be labeled, important objects (bank books, etc.) left in full view so as to be easily found when needed, "memory books" of faces, names, and relations kept available, and specific routines (e.g., bus to Denny's at 11 A.M. for lunch) religiously adhered to. Such cognitive scaffolding might be the work of the patients themselves, put gradually in place as the biological degeneration worsened, and/or set up by family and friends.

Now, when first confronted with such extreme reliance on external scaffolding, it is tempting to see it as underscoring a biocentric view of the individual agent, as deeply psychologically compromised. I submit, however, that this temptation is rooted not in any deep facts about the internal/external boundary, but in a mixture of unfamiliarity (these are not the external props that most of us use) and insufficiency (the external props are currently able to offset only a few of the debilitating effects of the Alzheimer's).

Thus consider, once again, the artist and the sketchpad. In this case we do not find ourselves lamenting the artist's lack of "real" creativity just because the creative process involves repeated and essential episodes of sketching and reperceiving. Nor do we reduce our admiration for the poet, just because the poetry emerges only courtesy of much exploratory activity with pen and paper. To see what I am getting at here, imagine next that *normal* human brains displayed the typical characteristics of the Alzheimer's brains. And imagine that we had slowly evolved a society in which the kinds of props and scaffolding deployed by Baum's Alzheimer's patients were the norm. Finally, reflect that that is exactly (in a sense) what we have done: our PCs, sketchpads, and notebooks complement our basic biological cognitive profile in much the same kind of way. Perhaps seeing the normal deep cognitive symbiosis between human brains and external technologies will prompt us to rethink some ideas about what it *is* to have a cognitive deficit, and to pursue, with increased energy, a vision of full and genuine cognitive rehabilitation using various forms of cognitive scaffolding.

There is also a real danger of erring to the opposite extreme. Once mind is located firmly *inside* the skull, one is tempted to ask whether even finer grained localization might be indicated. Thus consider a view expressed by Herbert Simon. Simon saw, very clearly, that portions of the external world often functioned as a nonbiological kind of memory. But instead of counting those portions (subject to the provisos just rehearsed) as proper parts of the knowing system, Simon chose to go the other way. Regarding biological, on-board memory, Simon invites us to "view this information-packed memory as less a part of the organism than of the environment to which it adapts" (Simon, 1982, p. 65). Part of the problem here no doubt originates from Simon's overly passive (mere storage) view of biological memory—we now know that the old data/process distinction offers precious little leverage when confronting biological computational systems. But the deeper issue, I suspect, concerns the underlying image of something like a "core agent" surrounded by (internal and external) support systems (memories, etc.). This image is incompatible with the emerging body of results from connectionism, neuroscience, and artificial life that we have been reviewing in the past several chapters. Instead of identifying intelligence with any kind of special core process, these recent investigations depict intelligence as arising from the operation of multiple, often quite special-purpose routines, some of which criss-cross neural bodily and environmental boundaries, and which often operate within the benefits of any kind of stable, unique, centralized control. Simon's view makes best sense against the backdrop of a passive view of memory and a commitment to some kind of centralized engine of "real" cognition. To whatever extent we are willing to abandon these commitments, we should be willing to embrace the possibility of genuine systemic extensions in which external processes and operations come to count as integral aspects of individual human intelligence (see Box 8.6 for some further considerations).

8.3 Suggested Readings

For further ideas about *the use of environmental structure* to augment biological cognition, see especially E. Hutchins, *Cognition in the Wild* (Cambridge, MA: MIT Press, 1995), a fantastically rich and detailed account of how multiple external factors contribute to the process of ship navigation (it's a good idea, oddly, to read Chapter 9 of Hutchins' book first). Daniel Dennett has done pioneering conceptual work hereabouts; see especially D. Dennett, *Darwin's Dangerous Idea* (New York: Simon and Schuster, 1995, Chapters 12 and 13) and D. Dennett, "Making Things to Think With," Chapter 5 of his excellent *Kinds of Minds* (New York: Basic Books, 1996). For my own attempts at bringing similar ideas into focus, see A. Clark, *Being There* (Cambridge, MA: MIT Press, 1997, Chapters 9 and 10).

For another (broadly Vygotskian) perspective on *socially and instrumentally mediated action*, see J. Wertsch, *Mind as Action* (New York: Oxford University Press, 1998).

Somewhat more *computationally oriented accounts of the role of environmental structure* include D. Kirsh and P. Maglio, "On Distinguishing Epistemic from Pragmatic Action," *Cognitive Science*, 18, 513–549, 1996, and various papers in P. Agre and S. Rosenschein (eds.),

Computational Theories of Interaction and Agency (Cambridge, MA: MIT Press, 1995), especially the essays by Agre, Beer, Hammond et al., and Kirsh.

For much more *on the possible relations between language and thought*, see the collection by P. Carruthers and J. Boucher (eds.), *Language and Thought* (Cambridge, England: Cambridge University Press, 1998), especially the essays by Carruthers and by Dennett. My paper, A. Clark, "Magic Words: How Language Augments Human Computation," appears there also. For more on the *language/thought/culture connection*, see J. Bruner, *Acts of Meaning* (Cambridge, MA: Harvard University Press, 1990).

For *the interplay between neural differences and the cascade of technological innovation*, see D. Dennett, *Kinds of Minds* (New York: Basic Books, 1996, Chapters 4–6), M. Donald, *Origins of the Modern Mind* (Cambridge, MA: Harvard University Press, 1991, Chapters 6–8), T. Deacon's difficult, but rewarding *The Symbolic Species* (New York: Norton, 1997), and S. Mithen, *The Prehistory of the Mind* (London: Thames and Hudson, 1996, especially Chapters 9–11).

For the specific idea of *language as enabling our own thoughts to become objects of further thought and attention*, see R. Jackendoff, "How language helps us think," published with replies in *Pragmatics and Cognition*, 4(1), 1–34, 1996. See especially the replies by Barnden, Clark, and Ellis.

For a *different, difficult, but very worthwhile* take on such issues, see C. Taylor, "Heidegger, language and ecology." In C. Taylor (ed.), *Philosophical Arguments* (Cambridge, MA: Harvard University Press, 1995).

On the topic *"where does the mind stop and the rest of the world begin?"* try A. Clark and D. Chalmers, "The extended mind." *Analysis*, 58, 7–19, 1998. Also J. Haugeland, "Mind embodied and embedded." In J. Haugeland (ed.), *Having Thought* (Cambridge, MA: Harvard University Press, 1998). For a careful, critical (and negative) appraisal of the "extended mind" idea, see K. Butler, *Internal Affairs* (Dordrecht, The Netherlands: Kluwer, 1998, Chapter 6).

Finally, for a fairly concrete *connectionist proposal about the role of external symbols*, see the chapter "Schemata and sequential thought processes in PDP models" in J. McClelland, D. Rumelhart, and the PDP Research Group, *Parallel Distributed Processing*, Vol. 2 (Cambridge, MA: MIT Press, 1986, pp. 7–58).