distributed memory networks could actually work, though these models are not explicitly discussed in the text.

Several important themes are introduced early in the book and developed throughout. One is the idea that information *storage* and information *processing* are probably not carried out in anatomically distinct areas; rather, memories are probably stored in the very same networks that process the information. For example, visual recognition of objects appears to depend on neurons in a specific area of the temporal lobe (TE) which also process pattern and shape information. That processing and storage can be shared by a network is an idea radically at variance with the design of a serial digital computer, where the memory box is located in one place, and information processing is done somewhere else. This implies that we have to think very differently from traditional AI about both processing and storage if we want to understand how nervous systems accomplish these things.

Another major theme concerns the diversity of kinds of plasticity, with distinct neuronal implementations. For example, lesion studies in humans and monkeys point to the existence of one memory sysem subserving what Squire calls declarative memory—learning of events, facts, propositional knowledge, and images which are accessible to awareness and can be reported by the subject. A distinct system seems to function for the storage of procedures, and involves the acquisition of motor skills, cognitive skills, and perceptual skills, but is independent of the subject's awareness of the acquisition. Although the characterizations of the two categories of memory are as yet only rough and provisional, the evidence for the independence of procedural memory from declarative memory is compelling. Moreover, the hippocampus and amygdala turn out to be essential for learning new declarative knowledge, but are not essential for acquisition of some procedural knowledge. These data have provided the basis for intense research at all levels on the contribution of the hippocampus and related structures to learning, and their relation to the cortical structures that are presumed to store the information. Research is focused at many levels of organization from the level of the molecule (the role of neurotransmitters and the various peptides), the special NMDA receptors at synapses which may have a special role in learning, the level of the circuit in the hippocampus, the system comprising the various structures subserving declarative memory, and finally, the whole brain.

This illustrates Squire's third theme, which is that as a result of research in the last several decades, we are embarked upon a period where we can begin to integrate research at various levels of organization in the brain in order to generate testable models of how the brain learns. Until quite recently, neuroscientists and psychologists have largely tended to ignore one another, since neurobiologists typically focused on cellular mechanisms of plasticity, while psychologists focused on behavioral parameters. But this has begun to change quite dramatically, and the co-evolution of high-level and low-level hypotheses represents the first stages of integration and the emergence of cognitive neurobiology. Learning, or more generally, plasticity, could well be one of the first places where we will figure out how the brain works; that is, where a coherent, integrated account of the biological mechanisms of psychological phenomena may at last be discovered. Since the nature of knowledge is the subject of epistemology, philosophers will want to understand these new developments. *Patricia Smith Churchland, University of California at San Diego.*

ROBERT J. RICHARDS. ***Darwin and the Emergence of Evolutionary Theories of Mind and Behavior.*** Chicago: University of Chicago Press, 1988, xvii + 700 pp. $29.95 (paper).

To some tastes, the system of Darwinian ideas is one of the most beautiful constructions in the universe—elegant, powerful, fecund. Its creation demands explanation, a task that falls properly to the historian of science, just as the explanation of the creation of the marvelous designs of biology is the task for that other sort of historian of science, the evolutionary theorist. Robert J. Richards, in this magnificently researched book, proposes to use the theoretical outlook of the latter sort of researcher as his model for the former task. Lovers of reflexivity will savor the proposal: an account of the evolution by natural selection of the idea of evolution by natural selection. But it is more than just a clever

gimmick; Richards' larger goal is to vindicate the natural selection model of the development of scientific theories in general, and his demonstration of its virtues in his test case is impressive.

Stephen Jay Gould is forever reminding us not to make the tunnel-vision error of looking back down the unbroken line of "successes" in a modern organism's lineage, thereby missing the bushy side branches of lineages without which we cannot make proper sense of evolutionary developments. If, for instance, we succumb to the error of viewing the contemporary organism as perfect, and then view its predecessors as crude approximations of the perfection to come, we create a treacherous illusion of progress (not that progress is never real), and judge the past by an anachronistic standard. The same moral exactly can be applied to the Just So Stories that pass for standard lore about how the Darwinian revolution conquered the intellectual world; Richards' antidote to these myths teems with insights and changes our perspective in many valuable ways.

Darwin's idea is one of the greatest of all time, but it surely wasn't *inspired*: it was created by no miraculous saltation of genius, but rather painfully eked out by Darwin's laborious persistence, in a paradigm of gradualism. Richards charts Darwin's false starts and backsliding, along a trajectory that is no random walk, but no farsighted trek towards the truth, either; sometimes Darwin seems to have been quite blind to the virtues of his own expressions, as he struggled to escape from the false summits of Lamarckian alternatives. This last vision is one of the most arresting in a book with more than its share of surprises: the central Darwinian idea, the progenitor of today's orthodoxy, only gradually emerged from a large pool of Lamarckian alternative genotypes—not one but a dozen, it seems—which persisted, in Darwin's own thought and writing and in that of his students and self-styled converts, well into the twentieth century. One sees hosts of other near neighbors, for instance in the associationist metaphors for gradualist learning, and the lure of sensationalist theories of mind. Distinguishing the best line in this crowd of near kin has been a process with about equal measure of the opposing errors: throwing out the baby with the bathwater, and snatching defeat from the jaws of victory. Again and again we see protagonists refusing to take seriously one good idea or another because of some slight perceived taint (of mechanism or of anti-mechanism, or romanticism or scientism), and on the other hand, sullying the stream of orthodox Darwinism with varieties of mentalism they could not quite bring themselves to foreswear.

As Richards correctly stresses, Darwin saw from the outset that his theory had to include an entirely naturalistic account of the origins of the "mind" and more particularly the "moral sense" of our species, for if Man were to be the golden exception to Darwin's rule, the whole theory would be dismissable. And so the environment in which the ideas were tested has never been—to this day—the relatively cool environment of pure science, but, to use an archaic term that still leaves an important fossil trace in Cambridge University, the hot and stormy region of the "moral sciences," where ethics, religion, anthropology, and psychology come together and vie for control of the territory. Gradually, and not without some local reversals, the ideas got sorted out, and it is eye-opening to learn to what extent all of this was accomplished, by Darwin, Huxley, Spencer, Romanes, Baldwin, William James, and others, in spite of (maybe even because of) the residual mentalism/moralism of the thinkers. Here, for instance, is Baldwin, revealing that the *motivation* behind his discovery of the effect that bears his name was not to find a properly mechanistic alternative to Lamarckian processes but to *secure within the Darwinian framework a place of importance for the mind*: "First, complexity of organism is the reflection and not the cause of complexity of thought, the opposite of the position of materialistic evolution . . . our monism is a monism of mind" (p. 427).

The monism avowed by Darwin and the early Darwinians was supposed to be an acceptable alternative to the shocking reductionism of mechanistic materialism, but in spite of themselves they were mechanists under the skin; they held themselves to mechanistic standards throughout, and Richards shows quite conclusively how their monism did *not* lead them into positing miraculous mind-powers. Still, Richards is obliged to introduce the term "ultra-Darwinian" to refer to that set of strictly non-Lamarckian, mechanistic ideas that Darwin himself couldn't quite embrace, but which we today, with hindsight, might be tempted to call the "essence" or the "true" Darwinian idea. (As good Darwinians,

we should, of course, resist all such essentialist temptations.)

In order to fill in the background, Richards' revisionist scholarship leads us through more than a few deservedly extinct theories and arguments, but along the way we are rewarded with delights and surprises: the convergent evolution of the Baldwin effect, simultaneously discovered by Lloyd Morgan, Baldwin and Osborn; a heroic and persuasive exhumation of Herbert Spencer (who turns out to have been a proto-Marxist, of all things!); a somewhat less compelling exoneration of Konrad Lorenz for his Nazi sympathies; a fascinating comparison of Mill and Darwin, those two admirable and massive Victorians, like ocean liners nudging cautiously close to each other, but keeping their distance. Richards uncovers the charming story of Darwin enlisting the assistance of his son to read over Mill's *Utilitarianism* to help him "get straight on his philosophy".

Evolutionary history teaches us that there is nothing *that* new under the sun, and it is salutary to find some of one's "own ideas" clearly expressed in books a hundred years old. Richards shows that again and again, Darwin (and his critics) saw to the heart of issues that have recently engaged us. For instance, the very arguments that have recently been used (unsuccessfully) against sociobiologists were tried and rebutted in Darwin's own day. Darwin also had the germs of the concept of group selection, and recognized the theoretical importance of (apparent) altruism in the social insects (and *homo sapiens*). And Huxley himself, in his Romanes lecture of 1893, adumbrated all of the major lines of criticism of "evolutionary ethics".

In the course of telling his evolutionary story, Richards provides useful correctives to various similar historiographic visions—those of Mayr, Lakatos, Toulmin, Popper and Kuhn—but also to less well-known hobby-horses of other historians of science. Like Darwin, moreover, he also has a "moral" goal: to rehabilitate not just evolutionary ethics, but a version of it quite close to Darwin's own. And like Darwin, he exhibits more philosophical cunning than most of his philosopher opponents, even if, in the end, his defense shows serious flaws in need of repair. In particular, he and Darwin both seem to fall for subtle versions of the error Skinner exhibits in *Beyond Freedom and Dignity*, where a *summum bonum* is too casually identified with something like long-term "survival of our culture". (So if the survival of our culture were shown to depend, in some crisis, on the enslavement and torture of many innocents, this would be justified?)

One final reflexive twist deserves mention. Richards notes that in order to take evolutionary models of science more seriously than other historians of science have done, we must look closely at the *local* mechanisms of idea-selection. What happens when a theory is replicated in a particular mind? Which books do particular theorists read, which get cited, etc.? It is an undoubted and depressing fact that important truths get published and buried, unread and uncited, while attention-grabbing near-truths replicate like rabbits. Even the wisest and most open-minded theorists can be deflected from taking seriously ideas that, however sound, are ill-equipped to compete in a particular intellectual niche. Richards is keenly aware that this fact applies not only to the ideas whose careers he is charting, but to his own work. He has been careful, for instance, to salt his discussions of such emotionally charged issues as social darwinism and sociobiology with stabs of sardonic wit that efficiently reveal his sympathy with contemporary sensitivities, thus avoiding premature dismissal at the hands of hair-triggered critics.

I conjecture that a similar motive lies behind his otherwise unaccountable omission of any discussion of Richard Dawkins' theory of *memes* (*The Selfish Gene*, 1976), which is to my knowledge the best-developed account of the very mechanisms of cultural selection and replication of ideas that Richards writes about. I think Dawkins' ideas about memes deserve to be taken seriously, and I think Richards thinks so too; he has, in effect, written a detailed account of the Dawkinsian evolution of the Darwinian memes. Why doesn't he discuss Dawkins? For two reasons, I suspect. First, Richards is almost as uncomfortable with "ultra-Darwinism" as Darwin was, and Dawkins is perhaps the most uncompromising ultra-Darwinian around. Second, Dawkins' works are often misperceived to be lightweight or mere journalism, so Richards, I suspect, has succumbed to some perceived social pressure and neglected to acknowledge a controversial ally—a familiar failing he often uncovers, and plausibly excuses, in his heroes.

Richards does not, in the end, convince me of one of his major theses: that the Darwinian

strain of anti-materialism is still viable as anything more than a stalwart refusal to settle for *simplistic* reductions of mind to mechanism. His book does, however, triumphantly achieve the goal of all great scholarship: it not only informs us, but shows us why becoming thus informed is essential to understanding our own issues and projects. *Daniel C. Dennett, Tufts University.*

DANIEL C. DENNETT. ***Elbow Room: The Varieties of Free Will Worth Wanting.*** Cambridge, Mass.: MIT Press, 1984. x + 200 pp. $8.95 (paper).

Elbow Room derives from the 1983 John Locke Lectures at Oxford. It is dedicated to the memory of Gilbert Ryle and is reminiscent of Ryle in its ordinary language style of philosophizing and its plaudits for common sense in opposition to "metaphysical" notions, but quite unRylean in its frequent allusions to scientific literature. The general thesis of *Elbow Room* is that scientific accounts of human nature, however mechanistic or deterministic, pose no threat to our free will. Its content will be generally congenial to compatibilists (with reservations) but will be thought by libertarians to provide yet further evidence that compatibilists are not interested in *real* free will but only in a pale *ersatz* copy.

Two themes run through the book: the first is a rejection of incompatibilism and the arguments used to support it; the second is a positive account, continuing Dennett's work in *Brainstorms* and elsewhere, of how a free, rational, deliberating agent might be constructed out of mechanistic pieces.

The first theme occupies most of the book. Dennett tells us that we are not in the clutches of the Invisible Jailer, the Nefarious Neurosurgeon, and other "bugbears" and "bogeymen" (chap. 1), that we have control over ourselves and are not controlled by our environment or the past (chap. 3), that it is not just a matter of luck that we are who we are but we can take some of the credit (chap. 4), that the future presents us with real opportunities for real deliberation and we must not succumb to fatalistic thinking about future events being inevitable and unpreventable (chap. 5), that free will does not imply that we could have done otherwise in some metaphysical and unverifiable sense but only that as a matter of empirical fact there are other possibilities open to us and we can improve if we want to (chap. 6), that responsibility is essentially forward-looking, a matter of deterrence and reform (chap. 7).

Unfortunately, the second theme occupies less of the book. What Dennett has to say on the first theme, however unfamiliar in presentation, is generally familiar in content (with one notable exception, his epistemic analysis of "prevent" and "can", see below). What he has to say on the second theme is more original. In chapter 1 Dennett describes the bogey of "sphexishness", named after the digger wasp whose apparently intelligent behavior can be unmasked and seen to be "merely mechanical". *Sphex* appeared earlier in *Brainstorms*. Here as there she represents the frightening possibility that we are not really intelligent but merely machines. Here as there Dennett's antidote is to remind us that it is the simplicity of mechanism, not mechanism as such, which is incompatible with intelligence. Dennett argues that reason, by jumping to a meta-level, a process that can be continued indefinitely, provides an escape from sphexishness while not transcending the mechanistic (chap. 2) and that a mechanism can be a self because being a self is a matter of internal organization; the self is not a unitary source or center within us (chap. 4). Dennett says that "if free will and responsibility requires absolute freedom from sphexishness . . ., then we must either overthrow the scientific vision of ourselves altogether or admit defeat" (p. 49). I think this is right, and the compatibilist must reject the antecedent. This leaves the disjuncts in the consequent to the libertarian and hard determinist respectively.

Dennett writes: "my method will be *to go slow where others go fast*" (p. 17). Unfortunately he also goes fast where others go slow, and this makes his defense of compatibilism less persuasive. However, the critical reviewer inevitably appears something of a spoilsport. If only Dennett had said more about this, if only he had considered this objection, if only. . . . Ah, but then his book wouldn't have been nearly so much fun!