An Evolutionary Perspective on Cognition: Through A Glass Lightly

Daniel C. Dennett

Center for Cognitive Studies, Tufts University, Medford, MA 02155-7059, USA

Thought in a hostile world: the evolution of human cognition
Kim Sterelny; Blackwell, Oxford, 2003, pp. 280, Price £50.00 hardback, ISBN 0-631-18886-X, Price £16.99 paperback, ISBN 0-631-18887-8.

For roughly half a century philosophers of mind have gnawed and nibbled on the bones of folk psychology, the informal 'theory of mind' that we articulate in everyday terms: our lore about 'beliefs' and 'desires', 'intentions' and 'thoughts', 'pains' and 'pleasures' and the like. Meanwhile academic psychology has undergone several waves of shifting nomenclature and ideology, from various kinds of behaviorism to various kinds of cognitivism, with neuroscience and evolutionary biology being thrown into the mix-with mixed results. It is depressing to reflect on how little has been truly settled by all this activity. One cannot expect to use *any* particular vocabulary to describe 'the mind' (itself an embattled concept) without provoking a sizeable cohort of objectors prepared to marvel at the fact that one is still in the thrall of *that* deeply problematic way of describing things. And yet there has been progress. It has been won, I think, by the interactions between two more or less opponent processes: opportunistic theory-driven oversimplification on the one hand and hard-won, in-the-trenches, empirical fact-digging on the other. The phenomena are so complicated that breathtaking oversimplification looks like a winning opening strategy here, if only we can find one as fecund as the toy worlds of Galilean mechanics or Mendelian genes or Bohr atoms. Then we can get a bird's-eye view of the task ahead, knowing that further progress will consist largely in conceding, again and again, that things turn out to be much more complicated than that simple picture suggests. But which oversimplifications should we lean on for the time being? Not those of stimulus-response behaviorism, apparently, but

E-mail address: ddennett@turts.edu (D.C. Dennett).

what of the physical symbol systems of GOF AI (Good Old Fashioned AI) or connectionism or neural nets or dynamical systems theory? Or should we cling to good old fashioned folk psychology and try to make its terms and categories line up with whatever we find in the brain?

Since these are questions about how to conceive of extraordinarily complicated interacting bits of cognitive machinery, and since all these had to evolve, it is reasonable to propose that an evolutionary perspective is a good bet to provide us with the largest-scale context in which to reverse-engineer the pieces. And in their own different ways, all the oversimplifiers have acknowledged this, some just paying lip-service to the point, and others trying, opportunistically, to exploit evolutionary ideas in big, simple ways in their theories. Central to Skinner's account of operant conditioning, for instance, was his insight that somehow or other, the brain had to be the site of a redesign process in the individual organism that was fundamentally evolutionary: 'Where inherited behavior leaves off, the inherited modifiability of the process of conditioning takes over' (Skinner, 1953, p. 83). At the other extreme, East Pole cognitivists (Dennett, 1987) such as Chomsky and Fodor have been eager to catalogue the innate gifts we each must have received from natural selection, while postponing indefinitely any attempt to decipher the historical evolutionary process. In between have been a spectrum of specialists, using insights from disciplines as diverse as animal psychology, behavioral ecology, evolutionary game theory, neuroanatomy, Artificial Life, and economics to bring aspects of Darwinian wisdom to bear on the complexities that confront us.

Kim Sterelny has his own ideas about which fruits of evolutionary thinking will help us understand the mind, and he wisely frames them within a larger project, an attempt to unify and consolidate the efforts of others, bringing a more detailed set of evolutionary considerations to bear on this task, analyzing the progress we have made and assessing the near-term prospects for further breakthroughs. This requires him to go slowly and cautiously over ground that others have dashed over. He surefootedly lays out the steps of the arguments, noting the difficulties and unresolved puzzles. He plays a traditional philosophical role, as an abstract agenda-setter and general-purpose argument-critic, and he achieves the traditional philosophical goals of exposing weaknesses in arguments and assumptions, clarifying the concepts, and organizing the issues, but his methods and diction are not all that traditional. First, he has an enviable command of an enormous spread of scientific literature that bears on the topics. A traditional philosopher of mind would be simply unable to undertake this project, since it depends heavily on having a wide and deep familiarity with the empirical research- the presuppositions and methods as well as the proclaimed results. Second (and this is to me the most striking feature of the book), he abandons the standard philosophical modes of assertion in favor of an unsettlingly candid avowal of his *mere opinions.* Again and again, we find him saying 'I conjecture . . .' (24); 'My suspicion…' and 'My bet…' (p. 50); 'My overall line of thought is . . .' (p. 57); 'It is probable, then…' (p. 61); 'I think it is fair to say…' (p. 70); 'I think the best guess is that…' (p. 76); 'I am not sure that . . .' (p. 86); 'I doubt that' and 'my guess is' (p. 92); 'I think this is very suggestive indeed' (p. 115); 'In my view' and 'I conjecture'

(p. 134); 'My own guess' (p. 143); 'my best guess' (p. 205); 'In short, I am far from convinced' (p. 208); '1 am not persuaded ' (p. 214); 'I do not reject this argument. But I do not want to rely exclusively on it . .. ' (p. 228). Is this refreshing humility or outrageous arrogance? Is he uncharacteristically acknowledging that we philosophers in our armchairs can 't figure out all the answers *a priori* but might still make a modest contribution or two? Or is he just a lazy philosopher who can't be bothered to finish the job and *prove* his conjectures ('Thus I refute thee!')? Besides, why should anybody be interested in *his* mere opinion? I think the truth is that he is bravely leading with his chin: he has obviously informed himself assiduously and thought long and well about the issues. Here are his considered views, for what they are worth, and he will not parade them as proofs or discoveries. Take them or leave them. Take them all seriously, say I. This (in *my* opinion) is an honest and attractive way of presenting the fruits of philosophical labors *when they are so well informed,* and I commend both the results and the mode of delivery to others, philosophers and scientists alike. That doesn't mean that I accept all his conjectures and claims, and at times I wish he'd tried harder to finish off a forlorn target of criticism or secure a conclusion. But these are slippery and difficult issues, and it is clear that his own opinions are often still unsettled. Better to admit it than to bluff. At least this way a *caveat* is issued to the reader. It is disconcerting to realize how influential a shoot-from-the-hip verdict from an authority can be in prematurely burying a promising new idea.

Sterelny divides the task he is surveying into two integrative projects. The internal project is strictly scientific: how does the machinery in our brains execute the work of relating us so appropriately to the environment in which we must make a living? (How do the 'wiring facts ' explain the 'connection facts '?) The second project asks where folk psychology fits in: 'the external integrative project, then, is to understand the relationship between the wiring-and-connection facts and the interpretation facts ' (p. 6). For some theorists, of course, the two projects have been seen as inextricably united by what Sterelny calls the Simple Coordination Thesis: the interpretations we bandy about in folk psychology work because they sketchily carve the human brain at its joints; if we look just right we will find those beliefs, desires, thoughts, intentions and pains as salient states and events in the brain's machinery. And if folk psychology 'works' on frogs and birds, we will find the same saliencies in frogs ' brains and birds' brains. But this risks a major distortion of the Whig history sort: anthropocentrically back-projecting into the brains of other species the peculiarities of our particular cognitive mechanisms, which may have evolved for quite different competences. It may be irresistible, when we are in interpretation mode, to wonder what monkeys *believe* about their rivals (or how bats *think* about their *experience* of echolocation) but a wiser course is to review what we know about the conditions under which cognitive mechanisms have arisen over all of evolutionary history. As D'Arcy Thompson famously said, everything is the way it is because it got that way, and as we add detail to our account of the historical journey, adopting a minimalist economy that imputes no cognitive sophistication before its time, we may appreciate more striking differences than similarities when we compare our minds to those of other animals.

The central novelty in Sterelny's account is the distinction between transparent and translucent worlds, a close kin-surprisingly unacknowledged by Sterelny--to the founding insight of game theory: as von Neumann and Morgenstern (1944) recognized, when there are other agents in the environment, an agent has to adopt a more complex way of representing its options, since they can be tracked, and thwarted, by that other, in a feedback loop. In a transparent world there can be problems of dealing with random noise but not of systematic interference by competing agents. As an organism you may need a sun-finder, but you needn't worry about *heliocrypsis;* the sun isn't designed to try to hide from you. (In a transparent *portion* of the world, such as the avenues leading from your eyes to your brain and thence to your effectors, the engineering is similarly less demanding; your eyes may on occasion deceive you, but these are accidents or by-products of trade-offs; your eyes are basically on your side because you and your eyes are in the same boat, evolutionarily.) Sterelny develops a persuasive case that evolution would design quite different systems for dealing with transparent and translucent ('hostile') aspects of the world. 'Hostile agents *pollute* an animal's informational world… they make decision problems more difficult through their *agent-sensitive responses* ... [and] hostile agents impose *costs on epistemic action'* (p. 26). Among the escalating responses to such problems are 'robust tracking' systems that can rely on any of a variety of different cues or modes of perception and, in particular, 'decoupled representations', the opposite, one might say, of what the frog's eye famously tells the frog's brain, a state that may be a representation of sorts, but that is firmly tied at both ends. The frog can neither muse about flies in their visual absence nor refrain from zotting with its tongue when the message is active. This serves the frog admirably until sneaky biologist agents intervene to exploit its traditional reliance on a relatively transparent world.

In between the robotic tropism of a frog and the savvy versatility of a human being lie all manner of hard-to-describe competences that are slowly beginning to expose their boundaries to relentless investigators. 'Some populations of chimps "fish" for termites, pushing twigs into termite mounds, then pulling out the stick and sucking off the termites. Could they adapt their fishing techniques to exploit other social insects?' (p. 47) This question, to which nobody yet knows the answer, highlights a nagging problem in animal studies: the incessant pull of generous overgeneralization by human observers. How could the chimp, so cleverly fishing with its stick, *not* be able to generalize in this simple fashion? And yet, again and again, researchers have stubbed their toes on striking cases of utter cluelessness in cases where 'intuitively' the novel problem should be well nigh identical to the old, solved problem. One would have thought there was a clearly decoupled representation in the chimp's brain, a general-purpose 'fuel for success' (Godfrey-Smith, 1996) available for instant exploitation in a novel setting, but perhaps not. We have to do the experiments. Sterelny reviews this literature well, highlighting the anomalies that have been found (apparently- some of the more anomalous findings are frankly doubted and furiously debated within the field).

Predator-prey relations raise the ante by creating an arms race of deception and detection, but what about the role of a more narrowly social environment (close kin and other conspecific agents, who may cooperate or compete) in driving evolutionary innovation? A popular view, pioneered by Humphrey (1976) and elaborated in several different ways by many others, is that a suite of cognitive elaborations evolved to permit specifically social intelligence, Machiavellian intelligence (Whiten & Byrne, 1988) enabling higher-order intentional states (beliefs about beliefs, desires about desires, and other such metarepresentations). While it is tempting and easy- well nigh irresistible- to impute higher order intentional states to primates (and dolphins and wolves, and, especially, dogs), the evidence needed to *secure* any particular higher-order attribution in a non-speaking animal is mighty hard to come by, with 'romantic' and 'killjoy' interpretations competing for our endorsement (Dennett, 1983). Sterelny rightly sees a need to emphasize the deflationary side of this ongoing tug of war, and his account of the recent campaigns about *imitation* and its confounds is clear and balanced. So is his account of the much discussed suite of experiments with chimps that turned Daniel Povinelli (Povinelli & Eddy, 1996) from a romantic into a killjoy. When philosophers first learn of such experiments in isolation from the rest of the literature, they often jump to the invited interpretation with the enthusiasm of a born-again convert. Sterelny knows better, and nowhere are his best guesses and informed doubts more salutary. At times, however, his insider's knowledge lures him into falling back on shorthand and knowing allusions that will probably baffle the uninitiated. His discussion of Tony Dickinson's proposals about the Garcia effect and 'Norns' (Dickinson & Balleine, 2000), the admirable artificial creatures invented by the British software engineer, Steve Grand (2000), was hard to follow and un persuasive even to me, and I participated with Sterelny in the workshop where Dickinson presented these ideas. A similar lapse later in the book is his treatment of the frame problem. Anyone unfamiliar with the primary literature on this curious and important topic will probably still be in the dark after reading his brief discussion. Much more successful pedagogically is his vivid and clear review of the literature on the evolution of cooperation, especially the emphasis on coalition and enforcement by risky punishment.

By the time evolution has created mechanisms for dealing with tricky conspecifics, we are a long way from what the frog's eye tells the frog's brain: we have some at least partially decoupled representation, equipment for enabling social coalition, and the stage is set for sociality to begin to playa dominant evolutionary role-not just by fostering further cooperation genes or anything like that, but by tidying up and revising the environment in which the individuals live and thrive and compete, so that the *selective* environment 'downstream' (in subsequent generations) becomes radically altered. This is 'niche construction', the idea that in some species, and pre-eminently in ours, the organism actively adapts the environment to its needs instead of passively adapting to its environment. This idea has been a persistent theme among a few evolutionists for years, starting with Lewontin (1982), and is now the subject of a new campaign (see especially Odling-Smee, Laland, & Feldman, 2003) about which Sterelny is cautiously enthusiastic, though he succeeds no better than the chief proponents of the movement in getting clear about whether this is a useful wrinkle in evolutionary thinking or a revolutionary overhaul of evolutionary theory, Just going on the track record of previous presumably revolutionary bandwagons in the area, I suspect that when the dust settles, niche construction will be seen as a minor elaboration of such predecessor ideas as Dawkins's *extended phenotype* (Dawkins, 1982). That's *my* best guess, but I don't want to rain on their parade. The way in which our species has made the world safe for its own kind, buffering most of the selective pressures that shaped our ancestors and hence abruptly shifting the adaptive landscape is at least very dramatic, and perhaps deserves to be focused upon as a major transition (Maynard Smith & Szathmary, 1995) in its own right.

Notice that so far, the evolution of language has not been mentioned. It lies in the background, of course, as a major enhancer of both social interactions and niche construction in our species, but Sterelny rightly, I think, wants to put the brakes on here as well, and see how far we can get before we contaminate all our thinking about other minds by making them speakers like us. This permits him to develop some refreshing perspectives on several well advertised but probably oversold hypotheses about the modularity of language and modularity hypotheses more generally. Language, on Sterelny's analysis, is well suited in *some* regards for a modular design solution, since for instance 'there is no conflict of interest between speaker and listener' with respect to the task of parsing or identifying the speaker's intentions (p. 180). Your likelihood of deceiving your listener depends on your *not* confusing him about which words you are uttering or which speech acts you mean him to take you to be performing. But a 'theory of mind module' in contrast, would seem to be a sitting duck for exploitation in most regards. Fodor's (1983) modules, being encapsulated, obligatory and cognitively impenetrable, 'work because, within particular domains, there are ecological constants that are stable across evolutionarily significant periods of time' (p. 186), This makes them inferior design options for use in volatile social interactions in which heavy feedback effects would be expected. Maybe, but I think this observation ignores a huge difference in time scales. Cuckoos and their unwilling hosts are engaged in a clearly hostile arms race, in which the moves and countermoves presumably could not adapt any faster than they do, but each move is apparently quite a rigid module. Are cuckoo host species more likely than other bird species to develop free-wheeling non-modular brains under the pressure of cuckoo parasitism? That would seem to be an implication of Sterelny's argument here.

Among other bracing buckets of cold water thrown by Sterelny are some good reflections on the poverty of the stimulus argument and several other oversimple analyses that have been held to support innate modules of one sort or another, especially where 'theory of mind' is concerned. None of this is conclusive, as he readily grants, but a burden of proof has been introduced that the other side must now shoulder. It may be that we are simply innately *curious* about other agents, innately biased in our allocation of attention and other cognitive resources to the problems of anticipating their actions, and this may suffice to permit normal people to develop robust interpretive skills with something like a ceiling effect (where just about everybody is as expert as everybody else). Much of the crosscultural uniformity that has been typically interpreted as grounds for a genetic source can be reinterpreted as due to niche construction, which provides an informational superhighway that can easily rival the genes in fidelity and bandwidth, and is much more quickly constructed. Mother nature is not a gene centrist. If valuable information can be transmitted to offspring by cultural means, this will secure the spread of a fitness-enhancing trait well in advance of any genetic modifications that hard-wire the trait. Moreover, there is at least often a positive coupling between interpreter and interpretee: it usually pays to grow into the interpretations others make of you. And that sort of scaffolding by expectation could presumably go a long way in securing the uniformity that strikes others as clear evidence of an innate scheme.

This book contains many arresting observations that defy summary but nevertheless persistently suggest subtle shifts in one's thinking about the topics and theories that currently preoccupy a multi-disciplinary multitude of researchers trying to understand the evolution of cognition. Aside from a few lapses into insider-speak such as those mentioned above, it is very accessibly and vividly written, and should make an enticing guide to neophytes who want to know the state of play today, while battle hardened veterans will be sure to find provocative novelties that will challenge some of their working assumptions about what is important and why.

References

Dawkins, R. (1982). *The extended phenotype: The gene as the unit of selection.* Oxford & San Francisco: Freeman.

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The 'Panglossian paradigm' defended. *Behavioral and Brain Sciences,* 6, 343- 390.

Dennett, D. C. (1987). The logical geography of computational approaches: A view from the east pole. In M. Harnish, & M. Brand (Eds.), *Problems in the representation of knowledge* (pp. 59-79). Tucson: Univ. of Arizona Press.

Dickinson, A., & Balleine, B. W. (2000). Causal cognition and goal directed action. In C. Heyes, & L. Huber (Eds.), *The evolution of cognition* (pp. 185-204). Cambridge, MA: MIT Press.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Godfrey-Smith, P. (1996). *Complexity and the function of mind in nature.* Cambridge: Cambridge University Press.

Grand, S. (2000). *Creation: Life and how to make it.* London: Weidenfeld & Nicolson.

Humphrey, N. (1976). The social function of intellect. In P. P. G. Bateson, & R. A. Hinde (Eds.), *Growing points in ethology* (pp. 303- 317). Cambridge: Cambridge University Press.

Lewontin, R. (1982). Organism and environment. In H. C. Plotkin (Ed.), *Learning, development and culture* (pp. 151 - 170). New York: Wiley.

Maynard Smith, J ., & Szathmary, E. (1995). *The major transitions in evolution.* San Francisco: Freeman.

Odling-Smee, J ., Laland, K., & Feldman, M. (2003). *Niche construction: The neglected process in evolution.* Princeton, NJ: Princeton University Press.

Povinelli, D., & Eddy, T. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development, 61,1 - 152.*

Skinner, B. F. (1953). *Science and human behavior.* New York: Macmillan.

von Neumann, J. , & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton: Princeton University Press.

Whiten, A., & Byrne, R. W. (Eds.) (1988), *Machiavellian intelligence.* Oxford: Clarendon Press.