The trouble with psychological Darwinism

---

*How the Mind Works* by Steven Pinker.
Penguin, 660 pp., £25, 22 January, 0 7139 9130 5

*Evolution in Mind* by Henry Plotkin.
276 pp., £20, 30 October 1997, 0 7139 9138 0

It belongs to the millennial mood to want to sum things up and see where we have gotten and point in the direction that further progress lies. Cognitive science has not been spared this impulse, so here are two books purporting to limn the state of the art. They differ a bit in their intended audience; Plotkin's is more or less a text, while Pinker hopes for a lay readership. Pinker covers much more ground but he takes an unarguably six hundred pages to do it, compared to Plotkin's svelte volume. Both authors are unusually good at exposition, Pinker exceptionally so from time to time. Their general sense of what's going on and of what comes next are remarkably similar, considering that they are writing about a field that is notoriously fractious. Taken severally or together, they present what is probably the best statement you can find in print of a very important contemporary view of mental structure and process.

But how much of it is true? To begin with, Pinker and Plotkin are reporting a minority consensus. Most cognitive scientists still work in a tradition of empiricism and associationism whose main tenets haven't changed much since Locke and Hume. The human mind is a blank slate at birth. Experience writes on the slate, and association extracts and extrapolates whatever trends there are in the record that experience leaves. The structure of the mind is thus an image, made a posteriori, of the statistical regularities in the world in which it finds itself. I would guess that quite a substantial majority of cognitive scientists believe something of this sort; so deeply, indeed, that many hardly notice that they do.

Pinker and Plotkin, by contrast, epitomise a rationalist revival that started about forty years ago with Chomsky's work on the syntax of natural languages and that is by now sufficiently robust to offer a serious alternative to the empiricist tradition. Like Pinker and Plotkin, I think the New Rationalism is the best story about the mind that science has found to tell so far. But I think their version of that story is tendentious, indeed importantly flawed. And I think the cheerful tone that they tell it in is quite unwarranted by the amount of progress that has actually been made. Our best scientific theory about the mind is better than empiricism; but, in all sorts of ways, it's still not very good. Pinker quotes Chomsky's remark the 'ignorance can be divided into problems and mysteries' and continues: 'I wrote this book because dozens of mysteries of the mind, from mental images to romantic love, have recently been upgraded to problems (though there are still some mysteries too!)' Well, cheerfulness sells books, but Ecclesiastes got it right: 'the heart of the wise is in the house of mourning.'

Pinker elaborates his version of rationalism around four basic ideas: the mind as computational system; the mind is massively modular; a lot of mental structure, including a lot of cognitive structure, is innate; a lot of mental structure, including a lot of cognitive structure, is an evolutionary adaptation - in particular, the function of a creature's nervous system is to abet the propagation of its genome (its selfish gene, as one says). Plotkin agrees with all four of these theses, though he puts less emphasis than Pinker does on the minds-are-computers part of the story. Both authors take for granted that psychology should be a part of biology and they are both emphatic about the need for more Darwinian thinking in cognitive science. (Plotkin quotes with approval Theodore Dobzhansky's dictum that 'nothing in biology makes sense except in the light of evolution,' amending it, however, to read 'makes complete sense.') It's their Darwinism, specifically their allegiance to a 'selfish gene' account of the phylogeny of the mind, that most strikingly distinguishes Pinker and Plotkin from a number of their rationalist colleagues (and from Chomsky in particular). All this needs some looking into. I'll offer a sketch of how the four pieces of Pinker-Plotkin's version of rationalism are connected; and, by implication, of what an alternative rationalism might look like. I'm particularly interested in how much of the Pinker-Plotkin consensus turns on the stuff about selfish genes, of which I don't, in fact, believe a word.

**Computation.**

Beyond any doubt, the most important thing that has happened in cognitive science was Turing's invention of the notion of mechanical rationality. Here's a quick, very informal, introduction. (Pinker provides on that's more extensive).

It's a remarkable fact that you can tell, just by looking at it, that any sentence of the syntactic form P and Q ("John swims and Mary drinks", as it might be) is true only if P and Q are both true. 'You can tell just by looking' means: to see that the entailments hold, you don't have to know anything about what either P or Q means and you don't have to know anything about the non-linguistic world. This really is remarkable since, after all, it's what they mean, together with how the nonlinguistic world is, that decide whether P of Q is itself true. This line of thought is often summarised by saying that some inferences are rational in virtue of the syntax of the sentences that enter into them; metaphorically, in virtue of the 'shapes' of these sentences.

Turing noted that, wherever an inference is formal in this sense, a machine can be made to execute the inference. This is because, although machines are awful at figuring out what's going on in the world, you can make them so that they are quite good at detecting and responding to syntactic relations among sentences. Give it an argument that depends just on the syntax of the sentences that it is couched in and the machine will accept the argument if and only if it is valid. To that extent, you can build a rational machine. Thus, in
collide. Many tears will be shed and many hands

Cultural relativism is widely specified; it's just a sort of precursor. In my view, the cognitive science that we've got so far has hardly begun to face this issue. Most practitioners (Pinker and Plotkin included, as far as I can tell) hope that it will resolve itself into lots of small, local problems which will in turn succumb to Turing's kind of treatment. Well, maybe; it's certainly worth the effort of continuing to try. But I'm impressed by this consideration: our best cognitive science is the psychology of perception, and (see just below) it may well be that perceptual processes are largely modular, hence computationally local. Whereas, plausibly, the globality of cognition shows up clearest in the psychology of common sense. Uncoincidentally, as things now stand, we don't have a theory of the psychology of common sense that would survive serious scrutiny by an intelligent five-year-old. Likewise, common sense is egregious what the computers that we know how to build don't have. I think it's likely that we are running into the limits of what can be explained with Turing's kind of computation; and I think we don't have any idea what to do about it.

Suffice it to say, anyhow, that if your notion of computation is exclusively local, then your notion of mental architecture had best be massively modular. That brings us to the second tenet of the Pinker-Plotkin version of Rationalism.

Massive modularity.

A module is a more or less autonomous, special purpose, computational system. It's built to solve a very restricted class of problems, and the information it can use to solve them with is proprietary. Most of the New Rationalists think that at least some human cognitive mechanisms are modular, aspects of perception being among the classical best candidates. For example, the computations that convert a two-dimensional array of retinal stimulations into a stable image of a three dimensional visual world are supposed to be largely autonomous with respect to the rest of one's cognition. That's why many visual illusions don't go away even if you know that they are illusory. Massimo Piatelli, reviewing Plotkin's book in Nature, remarks that the modularity of cognitive processes 'is arguably . . . the single most important discovery of cognitive science.' At a minimum, it's what most distinguishes our current cognitive science from its immediate precursor, the 'New Look' psychology of the Fifties that emphasised the continuity of perception and cognition and hence the impact of what one believes on what one sees.

Both Pinker and Plotkin think the mind is mostly made of modules; that's the massive modularity thesis in a nutshell. I want to stress how well it fits with the idea that mental computation is local. By definition, modular problem-solving works with less than all the information that a creature knows. It thereby minimises the global cognitive effects that are the bane of Turing's kind of computation. If the mind is massively modular, then maybe the notion of computation that Turing gave us is, after all, the only one that cognitive science needs. It would be nice to be able to believe that; Pinker and Plotkin certainly try very hard to do so. But, really, one can't. For, eventually the mind has to integrate the results of all those modular computations and I don't see how there could be a module for doing that. The moon looks bigger when it's on the horizon; but I know perfectly well it's not. My visual perception module gets fooled, but I don't. The question is: who is this 2? And by what -presumably global - computational process doe sit use what I know about the astronomical facts to correct the misleading appearances that my visual perception module insists on computing? If, in short, there is a community of computers living in my head, there had also better be somebody who is in charge; and, by God, it had better be me. The Old Rationalists, like Kant, thought that the integration of information is a lot of what's required to turn it into knowledge. If that's right, then a cognitive science that hasn't faced the integration problem has barely got off the ground. Probably, modular computation doesn't explain how minds are rational; it's just a sort of precursor. It's what you have to work through to get a view of how horribly hard our rationality is to understand.

Innateness.

Rationalists are nativists by definition; and nativism is where cognitive science touches the real world. As both Pinker and Plotkin rightly emphasise, the standard view in current social science - an din what's called 'literary theory' - takes a form of Empiricism for granted: human nature is arbitrarily plastic and minds are social constructs. By contrast, the evidence from cognitive science is that a lot of what's in the modules seems to be there innately. Pinker and Plotkin both review a fair sample of this evidence, including some of the lovely experimental work on infant cognition that psychologists have done in the last couple of decades. There is also, as the linguists have been claiming for years, a lot of indirect evidence that points to much the same conclusion: all human languages appear to e structurally similar in profound and surprising ways. There may be an alternative to the nativist explanation that linguistic structure is genetically specified; but, if there is, nobody has thus far had a glimpse of it. (For a review, see Pinker's earlier book, The Language Instinct.) Cultural relativism is widely held to be politically correct. So, sooner or later, political correctness and cognitive science are going to collide. Many tears will be shed and many hands will be wrung in public. Be that as it may; if there is a human nature, and it is to some
A lot of the fun of Pinker's book is his attempt to deduce human psychology from psychological Darwinism. Pinker and Plotkin both believe that if nativism is the right story about cognition, it follows that much of our psychology must be, in the Darwinian sense, an evolutionary adaptation; that is, it must be intelligible in light of evolutionary selection pressures that shaped it. It's the nativism that makes cognitive science politically interesting. But it's the inference from nativism to Darwinism that is currently divisive within the New Rationalist community. Pinker and Plotkin are selling an evolutionary approach to psychology that a lot of cognitive scientists (myself included) aren't buying. There are two standard arguments, both of which Pinker and Plotkin endorse, that are supposed to underwrite the inference from nativism to psychological Darwinism. The first is empirical, the second methodological. I suspect that both are wrong-headed.

The empirical argument is that, as a matter of fact, there is no way except evolutionary selection for Nature to build a complex, adaptive system. Plotkin says 'neo-Darwinian theory is the central theorem of all biology, including behavioural biology'; 'if behaviour is adaptive, then it must be the product of evolution.' Likewise Pinker: 'Natural selection is the only explanation we have of how complex life can evolve... [so] natural selection is indispensable to understanding the human mind.' One reply to this argument is to say that there is, after all, an alternative to natural selection as the source of adaptive complexity; you could get some by a miracle. But I'm not a Creationist, nor are any of my New Rationalist friends, as far as I know. Nor do we have to be, since there's another way out of the complexity argument. This is a long story, but here's the gist: it's common ground that the evolution of our behaviour was mediated by the evolution of our brains. So, what matters with regard to the question whether the mind is an adaptation is not how complex our behaviour is, but how much change you would have to make in an ape's brain to produce the cognitive structure of a human mind. And about this, exactly nothing is known. That because nothing is known about how the structure of our minds depends on the structure of our brains. Nobody even knows which brain structures it is that our cognitive capacities depend on. Unlike our minds, our brains are, by any gross measure, very like those of apes. So it looks as though relatively small alterations in brain structure must have produced very large behavioural discontinuities in the evolution transition from the ancestral apes to us. If that's right, then you don't have to assume that cognitive complexity is shaped by the gradual action of Darwinian selection on prehuman behavioural phenotypes. Analogies to the evolution of organic structures, though they pervade the literature of psychological Darwinism, don't actually cut much ice here. Make the giraffe's neck just a little longer and you correspondingly increase, by just a little, the animal's capacity to reach he fruit at the top of the tree. So it's plausible, to that extent, that selection stretched giraffes' necks bit by bit. But make an ape's brain just a little bigger (or denser, or more folded, or, who knows, greyer) and it's anybody's guess what happens to the creature's behavioural repertoire. Maybe the ape turns into us. Adaptationists say about the phylogeny of cognition that it's a choice between Darwin and God and they like to parade as scientifically tough-minded about which one of these you should pick. But that misstates the alternatives, so don't let yourself be bullied. In fact, we don't know what the scientifically reasonable view of the phylogeny of behaviour is; nor will we until we begin to understand how behaviour is subserved by the brain. And never mind tough-mindedness; what matters is what's true. Methodology is yet another thing that Pinker and Plotkin agree about. Both believe that the (anyhow, a) proper method of cognitive psychology is 'reverse engineering'. Reverse engineering is inferring how a device must work from, inter alia, a prior appreciation of its function. If you don't know what a can-opener is for, you are going to have trouble figuring out what its parts do. In the case of more complex machines, like for example people, your chance of getting the structure right is effectively nil if you don't know the function. Psychological Darwinism, so the argument goes, gives us the notion of function that the cognitive scientist's reverse engineering of the mind requires: To a first approximation, and with, to be sure, occasional exceptions, the function of a cognitive mechanism is whatever it is that selection it for. Without this evolutionary slant on function, cognitive science is therefore simply in the dark. This, too, is a long story. But if evolution really does underwrite a notion of function, it's a historical notion; and it's far from clear that a historical notion of function is what reverse engineering actually needs. You might think, after all, that what matters in understanding the mind is what ours do now, not what our ancestors' did some millions of years ago. And, anyhow, the reverse-engineering argument is over its head in anachronism. As a matter of fact, lots of physiology got worked out long before there was a theory of evolution. That's because you don't have to know how hands (or hearts, or eyes, or livers) evolved to make a pretty shrewd guess about what they are for. Maybe you also don't have to know how the mind evolved to make a pretty shrewd guess at what it's for; for example, that it's to think with. No doubt, arriving at a 'complete' (sic) explanation of the mind by reverse engineering might require an appreciation of its evolutionary history. But I don't think we should be worrying much about complete explanations at this stage. I'd settle for the merest glimpse of what is going on.

One last point about the status of the inference from nativism to psychological Darwinism. If the mind is mostly a collection of innate modules, then pretty clearly it must have evolved gradually, under selection pressure. That's because, as I remarked above, modules contain lots of specialised information about the problem-domains that they compute in. And it really would be a miracle if all those details got into brains via a relative small, fortuitous alteration of the neurology. To put it the other way around, if adaptationism isn't true in psychology, it must be that what makes our minds so clever is something pretty general; something about their global structure. The moral is that if you aren't into psychological Darwinism, you shouldn't be into massive modularity either. Everything connects. For the sake of the argument, however, let's suppose that the mind is an adaptation after all and see where that leads. It's a point of definition that adaptations have to be for something. Pinker and Plotkin both accept the 'selfish gene' story about what biological adaptations are for. Organic structure is (mostly) in aid of the propagation of the genes. And so is brain structure inter alia. And so is cognitive structure, since how the mind works depends on how the brain does. So there's a route from Darwinism to socio-biology; and Pinker, at least, is keen to take it. (Plotkin seems a bit less so. He's content to argue that some of the notorious problems for the selfish gene theory - the phylogeny of altruism, for example - may be less decisive than one might at first suppose. I think that settling for that is very wise of him.)

A lot of the fun of Pinker's book is his attempt to deduce human psychology from the assumption that our minds are adaptations for transmitting our genes. His last chapters are devoted to this and they range very broadly; including, so help me, one on the meaning of life. Pinker would like to convince us that the predictions that the selfish-gene theory makes about how our minds must be organised are independently plausible. But this project doesn't fare well. Prima facie, the picture of the mind, indeed of human nature in general, that psychological Darwinism suggest is preposterous; a sort of jumped up, down-market version of original sin. Psychological Darwinism is
kind of conspiracy theory; that is, it explains behaviour by imputing an interest (viz. in the proliferation of the genome) that the agent of the behaviour does not acknowledge. When literal conspiracies are alleged, duplicity is generally part of the charge: 'He wasn't making confetti; he was shredding the evidence. He did X in aid of Y, and then he lied about his motive.' But in the kind of conspiracy theories psychologists like best, the motive is supposed to be inaccessible even to the agent, who is thus perfectly sincere in denying the imputation. In the extreme case, it's hardly even the agent to whom the motive is attributed. Freudian explanations provide a familiar example: What seemed to be merely Jones's slip of the tongue was the unconscious expression of a libidinous impulse. But not Jones's libidinous impulse, really; one that his Id had on his behalf. Likewise, for the psychological Darwinist: what seemed to be your, after all, unsurprising interest in your child's well-being turns out to be your genes' conspiracy to propagate themselves. Not your conspiracy, notice, but theirs.

How do you make the case that Jones did X in aid of an interest in Y, when Y is an interest that Jones doesn't own to? The idea is perfectly familiar: you argue that X would have been the rational (reasonable, intelligible) thing for Jones to do if Y had been his motive. Such arguments can be very persuasive. The files Jones shredded were precisely the ones that would have incriminated him; and he shredded them in the middle of the night. What better explanation than that Jones conspired to destroy the evidence? Likewise when the conspiracy is unconscious. Suppose that an interest in the propagation of the genome would rationalise monogamous families in animals whose offspring mature slowly. Well, our offspring do mature slowly; and our species does, by and large, favour monogamous families. So that's evidence that we favour monogamous families because we have an interest in the propagation of our genes. Well, isn't it? Maybe yes, maybe no; this kind of inference needs to be handled with great care. For, often enough, where an interest in X would rationalise Y, so too would an interest in P, Q or R. It's reasonable of Jones to carry an umbrella if it's raining and he wants to keep dry. But, likewise, it's reasonable for Jones to carry an umbrella if he has in mind to return it to its owner. Since either motivation would rationalise the way that Jones behaved, his having behaved that way is compatible with either imputation. This is, in fact, overwhelmingly the general case: there are, most often, all sorts of interests which would rationalise the kinds of behaviour that a creature is observed to produce. What's needed to make it decisive that the creature is interested in Y is that it should produce a kind of behaviour that would be reasonable only given an interest in Y. But such cases are vanishingly rare since, if an interest in Y would rationalise doing X, so too would an interest in doing X. A concern to propagate one's genes would rationalise one's acting to promote one's children's welfare; but so too would an interest in one's children's welfare. Not all of one's motives could be instrumental, after all; there must be some things that one cares for just for their own sakes. Why, indeed, mightn't there be quite a few such things? Why shouldn't one's children be among them?

The literature of Psychological Darwinism is full of what appear to be fallacies of rationalisation: arguments where the evidence offered that an interest in Y is the motive for a creature's behaviour is primarily that an interest in Y would rationalise the behaviour if it were the creature's motive. Pinker's book provides so many examples that one hardly knows where to start. Here he is on friendship:

> Once you have made yourself valuable to someone, the person becomes valuable to you. You value him or her because if you were ever in trouble, they would have a stake - albeit a selfish stake - in getting you out. But now that you value the person, they should value you even more . . . because of your stake in rescuing him or her from hard times . . . This runaway process is what we call friendship.

And here he is on why we like to read fiction: 'Fictional narratives supply us with a mental catalogue of the fatal conundrums we might face someday and the outcomes of strategies we could deploy in them. What are the options if I were to suspect that my uncle killed my father, took his position, and married my mother?' Good question. Or what if it turns out that, having just used the ring that I got by kidnapping a dwarf to pay off the giants who built me my new castle, I should discover that it is the very ring that I need in order to continue to be immortal and rule the world? It's important to think out the options betimes, because a thing like that could happen to anyone and you can never have too much insurance. At one point Pinker quotes H.L. Mencken's wise-crack that 'the most common of all follies is to believe passionately in the palpably not true.' Quite so. I suppose it could turn out that one's interest in having friends, or in reading fictions, or in Wagner's operas, is really at heart prudential. But the claim affronts a robust, and I should think salubrious, intuition that there are lots and lots of things that we care about simply for themselves. Reductionism about this plurality of goals, when not Philistine or cheaply cynical, often sounds simply funny. Thus the joke about the lawyer who is offered sex by a beautiful girl. 'Well, I guess so,' he replies, 'but what's in it for me?' Does wanting to have a beautiful woman - or, for that matter, a good read - really require a further motive to explain it? Pinker duly supplies the explanation that you wouldn't have thought that you needed. 'Both sexes want a spouse who has developed normally and is free of infection . . . We haven't evolved stethoscopes or tongue-depressors, but an eye for beauty does some of the same things . . . Luxuriant hair is always pleasing, possibly because . . . long hair implies a long history of good health.'

Much to his credit, Pinker does seem a bit embarrassed about some of these consequences of his adaptationism, and he does try to duck them.

> Many people think that the theory of the selfish gene says that 'animals try to spread their genes'. This misstates . . . the theory. Animals, including most people, know nothing about genetics and care even less. People love their children not because they want to spread their genes (consciously or unconsciously) but because they can't help it. . . What is selfish is not the real motives of the person but the metaphorical motives of the genes that built the person. Genes 'try' to spread themselves (sic) by wiring animals brains so that animals love their kin . . . and then they[.] get out of the way.

This version sounds a lot more plausible; strictly speaking, nobody has as a motive ('conscious or unconscious') the proliferation of genes after all. Not animals, and not genes either. The only real motives are the ones that everybody knows about: of which love of novels, or women, or kin are presumably a few among many. But, pace Pinker, this reasonable view is not available to a psychological Darwinist. For to say that the genes 'wire' animals brains so that animals love their kin' and to stop there is to say only that loving their kin is innate in these animals. That reduces psychological Darwinism to mere nativism; which, as I remarked above, is common ground to all
of us Rationalists. The difference between Darwinism and mere nativism is the claim that a creature’s innate psychological traits are adaptations; viz. that their role in the propagation of the genes is what they’re for. Take the adaptationism away from a psychological Darwinist and he has nobody left to argue with except empiricists. It is, then, adaptationism that makes Pinker and Plotkin’s kind of rationalism special. Does this argument among nativists really matter? Nativism itself clearly does; everybody cares about human nature. But I have fussed a lot about the difference between nativism and Darwinism, and you might reasonably want to know why anyone should care about that.

For one thing, nativism says there has to be a human nature, but it’s the adaptationism that implies the account of human nature that sociobiologists endorse. If, like me, you find that account grotesquely implausible, it’s perhaps the adaptationism rather than the nativism that you ought to consider throwing overboard. Pinker remarks that ‘people who study the mind would rather not have to think about how it evolved because it would make a hash of cherished theories . . . When advised that [their] claims are evolutionarily implausible, they attack the theory of evolution rather than rethinking the claim.’ I think this is exactly right, though the formulation is a bit tendentious. We know - anyhow we think that we do - a lot about ourselves that doesn't seem to square with the theory that our minds are adaptations for spreading our genes. The question may well come down to which theory we should give up. Well, as far as I can tell, if you take away the bad argument that turns on complexity, and the bad argument from reverse engineering, and the bad arguments that depend on committing the rationalisation fallacy, and the atrociously bad arguments that depend on preempting what's to count as the 'scientific' (and/or the biological) world view, the direct evidence for psychological Darwinism is very slim indeed. In particular, it's arguably much worse than the indirect evidence for our intuitive, pluralistic theory of human nature. It is, after all, our intuitive pluralism that we use to get along with one another. And I have the impression that, by and large, it works pretty well.

Jerry Fodor teaches philosophy at Rutgers and at the Cuny Graduate Center. Concepts: Where Cognitive Science Went Wrong has just been published by Oxford.