

The Darwin College Lectures

What is Intelligence?

EDITED BY JEAN KHALFA



CAMBRIDGE
UNIVERSITY PRESS

Language and intelligence

DANIEL C. DENNETT

We human beings may not be the most admirable species on the planet, or the most likely to survive for another millennium, but we are without any doubt at all the most intelligent. We are also the only species with language. What is the relation between these two obvious facts? Does thought depend on language?

Before going on to consider that question, I must pause briefly to defend my second premise. Don't whales and dolphins, vervet monkeys and honey bees (the list goes on) have languages of sorts? Haven't chimpanzees in laboratories been taught rudimentary languages of sorts? Yes, and body language is a sort of language, and music is the international language (sort of) and politics is a sort of language, and the complex world of odour and olfaction is another, highly emotionally charged language, and so on. It sometimes seems that the highest praise we can bestow on a phenomenon we are studying is the claim that its complexities entitle it to be called a language – of sorts. This admiration for language – real language, the sort only we human beings use – is well-founded. The expressive, information-encoding properties of real language are practically limitless (in at least some dimensions), and the powers that other species acquire in virtue of their use of proto-languages, hemi-semi-demi-languages, are indeed similar to the powers we acquire thanks to our use of real language. These other species do climb a few steps up the mountain

on whose summit we reside, thanks to language. Looking at the vast differences between their gains and ours is one way of approaching the question I want to address.

HOW DOES LANGUAGE CONTRIBUTE TO INTELLIGENCE?

I once saw a cartoon showing two hippopotami basking in a swamp, and one was saying to the other: ‘Funny – I keep thinking it’s Tuesday!’ Surely no hippopotamus could ever think the thought that it is Tuesday. But on the other hand, if a hippopotamus could say that it was thinking any thought, it could probably think the thought that it was Tuesday.

What varieties of thought require language? What varieties of thought (if any) are possible without language? These might be viewed as purely philosophical questions, to be investigated by a systematic logical analysis of the necessary and sufficient conditions for the occurrence of various thoughts in various minds. And in principle such an investigation might work, but in practice it is hopeless. Any such philosophical analysis must be guided at the outset by reflections about what the ‘obvious’ constraining facts about thought and language are, and these initial intuitions turn out to be treacherous.

We watch a chimpanzee, with her soulful face, her inquisitive eyes and deft fingers, and we very definitely get a sense of the mind within, but the more we watch, the more our picture of her mind swims before our eyes. In some ways she is so human, so insightful, but we soon learn (to our dismay or relief, depending on our hopes) that in other ways, she is so dense, so uncomprehending, so unreachably cut off from our human world. How could a chimp who so obviously understands *A* fail to understand *B*? It sometimes seems flat impossible – as impossible as a person who can do multiplication and division but can’t count to ten. But is that really impossible? What about idiot savants who can play the piano but not read music, or children with Williams Syndrome (Infantile Hypercalcemia or IHC) who can carry on hyperfluent, apparently precocious conversations but are so profoundly retarded they cannot clothe themselves?

Philosophical analysis by itself cannot penetrate this thicket of perplexities. While philosophers who define their terms carefully might succeed in proving logically that – let us say – mathematical thoughts are impossible without mathematical language, such a proof might be consigned to irrelevance by the surprising discovery that mathematical *intelligence* does not depend on being able to have mathematical thoughts so defined!

Consider a few simple questions about chimpanzees: could chimpanzees learn to tend a fire – could they gather firewood, keep it dry, preserve the coals, break the wood, keep the fire size within proper bounds? And if they couldn't invent these novel activities on their own, could they be trained by human beings to do these things? I wonder. Here is another question. Suppose you imagine something outlandish – say a man climbing up a rope with a plastic dustbin over his head. An easy mental task for you. Could a chimpanzee do the same thing in her mind's eye? I wonder. I chose the elements – man, rope, climbing, dustbin, head – as familiar objects in the perceptual and behavioural world of a laboratory chimp, but I wonder whether a chimp could put them together in this novel way – even by accident, as it were. You were provoked to perform your mental act by my suggestion, and probably you often perform similar mental acts on your own in response to suggestions you give yourself – not spoken out loud, but definitely verbalised. Could it be otherwise? Could a chimpanzee get itself to perform such a mental act without the help of verbal suggestion? I wonder.

COMPARING OUR MINDS WITH OTHERS

These are rather simple questions about chimpanzees, but neither you nor I know the answers – yet. The answers are not impossible to acquire, but not easy either; controlled experiments could yield the answers, which would shed light on the role of language in turning brains into minds like ours. I think it is very likely that every content that has so far passed through your mind while reading this chapter is strictly off limits to non-language users, be they apes or dolphins, or even non-signing deaf people. If this is true, it is a striking fact, so

striking that it reverses the burden of proof in what otherwise would be a compelling argument: the claim, first advanced by the linguist Noam Chomsky, and more recently defended by the philosophers Jerry Fodor and Colin McGinn, that our minds, like those of all other species, must suffer ‘cognitive closure’ with regard to some topics of inquiry. Spiders can’t contemplate the concept of fishing, and birds – some of whom are excellent at fishing – are not up to thinking about democracy. What is inaccessible to the dog or the dolphin may be readily grasped by the chimp, but the chimp in turn will be cognitively closed to some domains we human beings have no difficulty thinking about. Chomsky and company ask a rhetorical question: What makes us think we are different? Aren’t there bound to be strict limits on what *Homo sapiens* may conceive? This presents itself as a biological, naturalistic argument, reminding us of our kinship with the other beasts, and warning us not to fall into the ancient trap of thinking ‘how like an angel’ we human ‘souls’, with our ‘infinite’ minds are.

I think that on the contrary it is a pseudo-biological argument, one that by ignoring the actual biological details, misdirects us away from the case that can be made for taking one species – our species – right off the scale of intelligence that ranks the pig above the lizard and the ant above the oyster. Comparing our brains with bird brains or dolphin brains is almost beside the point, because our brains are in effect joined together into a single cognitive system that dwarfs all others. They are joined by one of the innovations that has invaded our brains and no others: language. I am not making the foolish claim that all our brains are knit together by language into one gigantic mind, thinking its transnational thoughts, but rather that each individual human brain, thanks to its communicative links, is the beneficiary of the cognitive labours of the others in a way that gives it unprecedented powers. Naked animal brains are no match at all for the heavily armed and outfitted brains we carry in our heads.

A purely philosophical approach to these issues is hopeless. It must be supplemented – not replaced – with researches in a variety of disciplines ranging from cognitive psychology and neuroscience to evolutionary theory and palaeoanthropology. I raised the question

about whether chimps could learn to tend a fire because of its close – but treacherous! – resemblance to questions that have been discussed in the recent flood of excellent books and articles about the evolution of the human mind.

I will not attempt to answer the big questions, but simply explain why answers to them will hinge on answers to the questions raised – and to some degree answered – in this literature. In the terms of the Oxford zoologist Richard Dawkins, my role today is to be a vector of memes, attempting to infect the minds in one niche – my home discipline of philosophy – with memes that are already flourishing in others.

At some point in prehistory, our ancestors tamed fire; the evidence strongly suggests that this happened hundreds of thousands of years – or even as much as a million years – before the advent of language, but of course after our hominid line split away from the ancestors of modern apes such as chimpanzees. What, if not language, gave the first fire-taming hominids the cognitive power to master such a project? Or is fire-tending not such a big deal? Perhaps the only reason we don't find chimps in the wild sitting around campfires is that their rainy habitats have never left enough tinder around to give fire a chance to be tamed. (The neurobiologist William Calvin tells me that Sue Savage-Rumbaugh's pygmy chimps in Atlanta love to go on picnics in the woods, and enjoy staring into the campfire's flames, just as we do.)

NEED TO KNOW VERSUS THE COMMANDO TEAM

If termites can create elaborate, well-ventilated cities of mud, and weaverbirds can weave audaciously engineered hanging nests, and beavers can build dams that take months to complete, couldn't chimpanzees tend a simple campfire? This rhetorical question climbs another misleading ladder of abilities. It ignores the independently well-evidenced possibility that there are two profoundly different ways of building dams: the way beavers do and the way we do. The

differences are not necessarily in the products, but in the control structures within the brains that create them. A child might study a weaverbird building its nest, and then replicate the nest herself, finding the right pieces of grass, and weaving them in the right order, creating, by the very same series of steps, an identical nest. A film of the two building processes occurring side-by-side might overwhelm us with a sense that we were seeing the same phenomenon twice, but it would be a mistake to impute to the bird the sort of thought processes we know or imagine to be going on in the child. There could be very little in common between the processes going on in the child's brain and the bird's brain. The bird is (apparently) endowed with a collection of interlocking special-purpose minimalist subroutines, well-designed by evolution according to the notorious 'need to know principle' of espionage: give each agent as little information as will suffice for it to accomplish its share of the mission.

Control systems designed under this principle can be astonishingly successful – witness the birds' nests, after all – whenever the environment has enough simplicity and regularity, and hence predictability, to favour predesign of the whole system. The system's very design in effect makes a prediction – a wager, in fact – that the environment will be the way it must be for the system to work. When the complexity of encountered environments rises, however, and unpredictability becomes a more severe problem, a different design principle kicks in: the commando team principle illustrated by such films as *The Guns of Navarone* – give each agent as much knowledge about the total project as possible, so that the team has a chance of ad libbing appropriately when unanticipated obstacles arise.

Fortunately, we don't have to inspect brain processes directly to get evidence of the degree to which one design principle or the other is operating in a particular organism – although in due course it will be wonderful to get confirmation from neuroscience. In the meantime, we can conduct experiments that reveal the hidden dissimilarities by showing how bird and child respond to abnormal obstacles and opportunities along the way.

I have a favourite example of such an experiment with beavers. It

turns out that beavers hate the sound of running water and will cast about frantically for something – anything – that will bring relief. As Wilsson has shown, when recordings of running water are played from loudspeakers, beavers will respond by plastering the loudspeakers with mud.

So there is a watershed in the terrain of evolutionary design space; when a control problem lies athwart it, it could be a matter of chance which direction evolution propelled the successful descendants. Perhaps, then, there are two ways of tending fires – roughly, the beaver-dam way, and our way. If so, it is a good thing for us that our ancestors did not hit upon the beaver-dam way. If they had, the woods might today be full of apes sitting around campfires, but we would not be here to marvel at them.

THE TOWER OF GENERATE-AND-TEST

I want to propose a framework in which we can place the various design options for brains, to see where their power comes from. It is an outrageously oversimplified structure, but idealisation is the price one should often be willing to pay for synoptic insight. I will call it the tower of generate-and-test.

In the beginning there was Darwinian evolution of species by natural selection. A variety of candidate organisms were blindly generated by more or less arbitrary processes of recombination and mutation of genes. These organisms were field tested, and only the best designs survived. This is the ground floor of the tower. Let us call its inhabitants *Darwinian creatures*. (Is there perhaps a basement? Recently speculations by physicists and cosmologists about the evolution of universes open the door to such a prospect, but I will not explore it on this occasion. My topic today is the highest stories of the tower.)

This process went through many millions of cycles, producing many wonderful designs, both plant and animal, and eventually among its novel creations were some designs with the property of phenotypic plasticity. The individual candidate organisms were not

wholly designed at birth, or in other words there were elements of their design that could be adjusted by events that occurred during the field tests. Some of these candidates, we may suppose, were no better off than their hard-wired cousins, since they had no way of favouring (selecting for an encore) the behavioural options they were equipped to ‘try out’, but others, we may suppose, were fortunate enough to have wired-in ‘reinforcers’ that happened to favour smart moves, actions that were better for their agents. These individuals thus confronted the environment by generating a variety of actions, which they tried out, one by one, until they found one that ‘worked’. We may call this subset of Darwinian creatures, the creatures with conditionable plasticity, *Skinnerian creatures*, since, as B. F. Skinner was fond of pointing out, operant conditioning is not just analogous to Darwinian natural selection; it is continuous with it. ‘Where inherited behavior leaves off, the inherited modifiability of the process of conditioning takes over.’

Skinnerian conditioning is a fine capacity to have, so long as you are not killed by one of your early errors. A better system involves preselection among all the possible behaviours or actions, weeding out the truly stupid options before risking them in the harsh world. We human beings are creatures capable of this third refinement, but we are probably not alone. We may call the beneficiaries of this third story in the tower *Popperian creatures*, since as Sir Karl Popper once elegantly put it, this design enhancement ‘permits our hypotheses to die in our stead.’ Unlike the merely Skinnerian creatures who survive because they are lucky, we Popperian creatures survive because we are smart – of course we are just lucky to be smart, but that is better than just being lucky.

But how is this preselection in Popperian agents to be done? Where is the feedback to come from? It must come from a sort of inner environment – an inner something-or-other that is structured in such a way that the surrogate actions it favours are more often than not the very actions the real world would also bless, if they were actually performed. In short, the inner environment, whatever it is, must contain lots of information about the outer environment and its regularities.

Nothing else (except magic) could provide preselection worth having. Now here we must be very careful not to think of this inner environment as simply a replica of the outer world, with all its physical contingencies reproduced. (In such a miraculous toy world, the little hot stove in your head would be hot enough to actually burn the little finger in your head that you placed on it!) The information about the world has to be there, but it also has to be structured in such a way that there is a non-miraculous explanation of how it got there, how it is maintained, and how it actually achieves the preselective effects that are its *raison d'être*.

We have now reached the story of the Tower on which I want to build. Once we get to Popperian creatures, creatures whose brains have the potential to be shaped into inner environments with pre-selective prowess, what happens next? How does new information about the outer environment get incorporated into these brains? This is where earlier design decisions – and in particular, choices between need to know and commando team – come back to haunt the designer; for if a particular species' brain design has already gone down the need to know path with regard to some control problem, only minor modifications (fine tuning, you might say) can be readily made to the existing structures, so the only hope of making a major revision of the internal environment to account for new problems, new features of the external environment that matter, is to submerge the old hard-wiring under a new layer of pre-emptive control (a theme developed in the work of the Artificial Intelligence (AI) researcher Rodney Brooks). It is these higher levels of control that have the potential for vast increases in versatility. And it is at these levels, in particular, that we should look for the role of language (when it finally arrives on the scene), in turning our brains into virtuoso preselectors.

We engage in our share of rather mindless routine behaviour, but our important acts are often directed on the world with incredible cunning, composing projects exquisitely designed under the influence of vast libraries of information about the world. The instinctual actions we share with other species show the benefits derived by the harrowing explorations of our ancestors. The imitative actions we

share with some higher animals may show the benefits of information gathered not just by our ancestors, but also by our social groups over generations, transmitted non-genetically by a 'tradition' of imitation. But our more deliberately planned acts show the benefits of information gathered and transmitted by our conspecifics in every culture, including, moreover, items of information that no single individual has embodied or understood in any sense. And while some of this information may be of rather ancient acquisition, much of it is brand new. When comparing the timescales of genetic and cultural evolution, it is useful to bear in mind that we today – every one of us – can easily understand many ideas that were simply unthinkable *by the geniuses* in our grandparents' generation!

The successors to mere Popperian creatures are those whose inner environments are informed by the designed portions of the outer environment. We may call this sub-sub-subset of Darwinian creatures *Gregorian creatures*, since Richard Gregory is to my mind the pre-eminent theorist of the role of information – or more exactly, what Gregory calls potential intelligence – in the creation of smart moves – or what Gregory calls kinetic intelligence. Gregory observes that a pair of scissors, as a well-designed artifact, is not just a result of intelligence, but an endower of intelligence (external potential intelligence), in a very straightforward and intuitive sense: when you give someone a pair of scissors, you enhance their potential to arrive more safely and swiftly at smart moves.

Anthropologists have long recognised that the advent of tool use accompanied a major increase in intelligence. Our fascination with the discovery that chimpanzees in the wild fish for termites with crudely prepared fishing sticks is not misplaced. This fact takes on further significance when we learn that not all chimpanzees have hit upon this trick; in some chimpanzee 'cultures' termites are a present but unexploited food source. This reminds us that tool use is a two-way sign of intelligence; not only does it require intelligence to recognise and maintain a tool (let alone fabricate one), but it confers intelligence on those who are lucky enough to be given the tool. The better designed the tool, the more information is embedded in its fabrica-

tion, the more potential intelligence it confers on its user. And among the pre-eminent tools, Gregory reminds us, are what he calls mind-tools: words. What happens to a human or hominid brain when it becomes equipped with words? I have arrived, finally, back at the question with which I began.

WHAT WORDS DO TO US

There are two related mistakes that are perennially tempting to theorists thinking about the evolution of language and thinking. The first is to suppose that the manifest benefits of communication to humanity (the group, or the species) might themselves explain the evolution of language. The default supposition of evolutionary theory must be that individuals are initially competitive, not cooperative, and while this default can be most interestingly overridden by special conditions, the burden is always to demonstrate the existence of the special conditions. The second mistake is to suppose that mind-tools – words, ideas, techniques – that were not ‘good for us’ would not survive the competition. The best general antidote I know to both these errors is Richard Dawkins’ discussion of *memes* in *The Selfish Gene*. The best detailed discussion I know of the problem of designing communication under the constraint of competitive communicators is by Dan Sperber and Deirdre Wilson in their excellent book, *Relevance: A Theory of Communication*.

One upshot of the considerations raised by these thinkers is that one may usefully think of words – the most effective vehicles for memes – as invading or parasitising a brain, not simply being acquired by a brain. What is the shape of this environment when words first enter it? It is definitely not an even playing field or a *tabula rasa*. Our newfound words must anchor themselves on the hills and valleys of a landscape of considerable complexity. Thanks to earlier evolutionary pressures, our innate quality spaces are species-specific, narcissistic, and even idiosyncratic from individual to individual.

A number of investigators are currently exploring portions of this terrain. The psychologist Frank Keil and his colleagues at Cornell

have evidence that certain highly abstract concepts – such as the concepts of being alive or ownership, for instance – have a genetically imposed head start in the young child's kit of mind-tools; when the specific words for owning, giving and taking, keeping and hiding, and their kin enter a child's brain, they find homes already partially built for them. Ray Jackendoff and other linguists have identified fundamental structures of spatial representation – notably designed to enhance the control of locomotion and the placement of movable things – that underlie our intuitions about concepts like *beside*, *on*, *behind*, and their kin. Nicholas Humphrey has argued in recent years that there must be a genetic predisposition for adopting what I have called the intentional stance, and Alan Leslie and others have developed evidence for this, in the form of what he calls a 'theory of mind module' designed to generate second-order beliefs (beliefs about the beliefs and other mental states of others). Some autistic children seem to be well-described as suffering from the disabling of this module, for which they can occasionally make interesting compensatory adjustments.

We are only just beginning to discern the details of the interactions between such pre-existing information structures and the arrival of language, so theorists who have opportunistically ignored the phenomenon up till now have nothing to apologise for. The time has come, however, to change tactics. In Artificial Intelligence, for instance, even the most ambitiously realistic systems – such as Soar, the star of Allen Newell's *Unified Theories of Cognition* – are described without so much as a hint about which features, if any, are dependent on the system's having acquired a natural language with which to supplement its native representational facilities. The result is that most AI agents, the robotic as well as the bedridden, are designed on the model of the walking encyclopedia, as if all the information in the inner environment were in the form of facts told at one time or another to the system. And in the philosophy of mind, there is a similar tradition of theory-construction and debate about the nature of belief, desire and intention – philosophical 'theories of mental representation' – fed on a diet exclusively drawn from lan-

guage-infected cognitive states. Tom believes that snow is white. Do polar bears believe that snow is white? In the same sense? Supposing one might develop a good general theory of belief by looking exclusively at such specialised examples is like supposing one might develop a good general theory of motor control by looking exclusively at examples of people driving automobiles in city traffic. 'Hey, if that isn't *motor* control, what is?' – a silly pun echoed, I am claiming, by the philosopher who says 'Tom believes snow is white – hey, if that isn't a belief, what is?'

WHAT WORDS DO FOR US

John Holland, a pioneer researcher on genetic algorithms, has recently summarised the powers of the Popperian internal environment, adding a nice wrinkle:

An internal model allows a system to look ahead to the future consequences of current actions, without actually committing itself to those actions. In particular, the system can avoid acts that would set it irretrievably down some road to future disaster ('stepping off a cliff'). Less dramatically, but equally important, the model enables the agent to make current 'stage-setting' moves that set up later moves that are obviously avarageous. The very essence of a competitive advantage, whether it be in chess or economics, is the discovery and execution of stage-setting moves.

But how intricate and long-range can the 'stage-setting' look-ahead be without the intervention of language to help control the manipulation of the model? This is the relevance of my question at the outset about the chimp's capacities to visualise a novel scene. As Merlin Donald points out, Darwin was convinced that language was the prerequisite for 'long trains of thought', and this claim has been differently argued for by several recent theorists, especially Julian Jaynes and Howard Margolis. Long trains of thought have to be controlled, or they will wander off into delicious if futile woolgathering. These authors suggest, plausibly, that the self-exhortations and reminders made possible by language are actually essential to maintaining the

sorts of long-term projects only we human beings engage in (unless, like the beaver, we have a built-in specialist for completing a particular long-term project).

Merlin Donald resists this plausible conjecture, and offers a variety of grounds for believing that the sorts of thinking that we can engage in without language are remarkably sophisticated. I commend his argument to your attention in spite of the doubts about it I will now briefly raise. Donald's argument depends heavily on two sources of information, both problematic in my opinion. First, he makes strong claims about the capabilities of those congenitally deaf human beings who have not yet developed (so far as anyone can tell) any natural language – in particular, signing.

Second, he draws our attention to the amazing case of Brother John, a French Canadian monk who suffers from frequent epileptic seizures that do not render him unconscious or immobile, but just totally aphasic, for periods of a few minutes or hours. During these paroxysms of aphasia, we are told, Brother John has no language, either external or internal. That is, he can neither comprehend nor produce words of his native tongue, not even 'to himself'. At the same time, Brother John can 'still record the episodes of life, assess events, assign meanings and thematic roles to agents in various situations, acquire and execute complex skills, learn and remember how to behave in a variety of settings.'

My doubts about the use to which Donald wants to put these findings are straightforward, and should be readily resolvable in time: both Brother John and the long-term language-less deaf people, are in different ways and to different degrees, still the beneficiaries of the shaping role of language. In the case of Brother John, his performance during aphasic paroxysm relies on 'language-mediated apprenticeships'.

Brother John maintains, for instance, that he need not tell himself the words 'tape recorder', 'magnetic tape', 'red button on the left', 'turn', 'push' and so forth . . . in order to be capable of properly operating a tape recorder . . .

The deaf who lack sign – a group whose numbers are diminishing today, thank goodness – lack Brother John’s specific language-mediated apprenticeships, but we simply don’t know – yet – what structures in their brains are indirect products of the language that most of their ancestors in recent millennia have shared. The evidence that Donald adduces for the powers of language-less thought is thus potentially misleading. These varieties of language-less thought, like barefoot waterskiing, may be possible only for brief periods, and only after a preparatory period that includes the very feature whose absence is later so striking.

There are indirect ways of testing the hypotheses implied by these doubts. Consider episodic memory, for instance. When a dog retrieves a bone it has buried, it manifests an effect on its memory, but must the dog, in retrieving the bone, actually recollect the episode of burying? (Perhaps you can name the current US Secretary of State, but can you recall the occasion of learning his name?) The capacity for genuine episodic recollecting – as opposed to semantic memory installed by a single episode of learning – is in need of careful analysis and investigation. Donald follows Jane Goodall in claiming that chimpanzees in the wild are ‘able to perceive social events accurately and to remember them’ – as episodes in memory. But we have not really been given any evidence from which this strong thesis follows; the social perspicuity of the chimpanzees might be largely due to specialised perceptual talents interacting with specialised signs. Suppose, for instance, that there is something subtle about the posture of a subordinate facing a superior that instantly – visually – tells an observer chimp (but not a human observer) which is subordinate, and how much. Experiments that would demonstrate a genuine capacity for episodic memory in chimpanzees would have to involve circumstances in which an episode was observed or experienced, but in which its relevance as a premise for some social inference was not yet determined – so no ‘inference’ could be drawn at once. If something that transpired later suddenly gave a retrospective relevance to the earlier episode, and if a chimpanzee can tumble to that fact, this would be evidence – but not yet conclusive evidence – of episodic memory.

Another way of testing for episodic memory in the absence of language would be to let a chimpanzee observe – once – a relatively novel and elaborate behavioural sequence that accomplishes some end (for example, to make the door open, you stamp three times, turn in a circle and then push both buttons at once), and see if the chimpanzee, faced with the need to accomplish the same end, can even come close to reproducing the sequence. It is not that there is any doubt that chimpanzee brain tissue is capable of storing this much information – it can obviously store vastly more than is required for such a simple feat – but whether the chimpanzee can exploit this storage medium in such an adaptive way on short notice. And that is the sort of question that no amount of microscopic brain-study is going to shed much light on.

THE ART OF MAKING MISTAKES

This brings me to my final step up the tower of generate-and-test. There is one more embodiment of this wonderful idea, and it is the one that gives our minds their greatest power: once we have language – a bountiful kit of mind-tools – we can use it in the structure of deliberate, foresighted generate-and-test known as science. All the other varieties of generate-and-test are willy-nilly.

The soliloquy that accompanies the errors committed by the lowliest Skinnerian creature might be ‘Well, I mustn’t do *that* again!’ and the hardest lesson for any agent to learn, apparently, is how to learn from one’s own mistakes. In order to learn from them, one has to be able to contemplate them, and this is no small matter. Life rushes on, and unless one has developed positive strategies for recording one’s tracks, the task known in AI as credit assignment (also known, of course, as blame assignment!) is insoluble. The advent of high-speed still photography was a revolutionary technological advance for science because it permitted human beings, for the first time, to examine complicated temporal phenomena not in real time, but in their own good time – in leisurely, methodical backtracking analysis of the traces they had created of those complicated events. Here a technolo-

gical advance carried in its wake a huge enhancement in cognitive power. The advent of language was an exactly parallel boon for human beings, a technology that created a whole new class of objects-to-contemplate, verbally embodied surrogates that could be reviewed in any order at any pace. And this opened up a new dimension of self-improvement – all one had to do was to learn to savour one's own mistakes.

But science is not just a matter of making mistakes, but of making mistakes in public. Making mistakes for all to see, in the hopes of getting the others to help with the corrections. It has been plausibly maintained, by Nicholas Humphrey, David Premack and others, that chimpanzees are natural psychologists – what I would call second-order intentional systems – but if they are, they nevertheless lack a crucial feature shared by all human natural psychologists, folk and professional varieties: they never get to compare notes. They never dispute over attributions, and ask for the grounds for each others' conclusions. No wonder their comprehension is so limited. Ours would be, too, if we had to generate it all on our own.

Let me sum up the results of my rather swift and superficial survey. Our human brains, and only human brains, have been armed by habits and methods, mind-tools and information, drawn from millions of other brains to which we are not closely related. This, amplified by the deliberate use of generate-and-test in science, puts our minds on a different plane from the minds of our nearest relatives among the animals. This species-specific process of enhancement has become so swift and powerful that a single generation of its design improvements can now dwarf the R-and-D efforts of millions of years of evolution by natural selection. So while we cannot rule out the possibility in principle that our minds will be cognitively closed to some domain or other, no good 'naturalistic' reason to believe this can be discovered in our animal origins. On the contrary, a proper application of Darwinian thinking suggests that, if we survive our current self-induced environmental crises, our capacity to comprehend will continue to grow by increments that are now incomprehensible to us.

FURTHER READING

- Calvin, William, *The Ascent of Mind: Ice Age Climates and the Evolution of Intelligence*, New York: Bantam 1990.
- Dawkins, Richard, *The Selfish Gene*, Oxford: Oxford University Press 1976.
- Diamond, Jared, *The Third Chimpanzee: The Evolution and Future of the Human Animal*, New York: Harper 1992.
- Donald, Merlin, *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*, Cambridge, MA: Harvard University Press 1991.
- Gregory, Richard, *Mind in Science*, Cambridge: Cambridge University Press 1981.
- Jackendoff, Ray, *Consciousness and the Computational Mind*, Cambridge, MA: MIT Press 1987.
- Jaynes, Julian, *The Origins of Consciousness in the Breakdown of the Bicameral Mind*, Boston: Houghton Mifflin 1976.
- McGinn, Colin, *The Problem of Consciousness*, Oxford: Blackwell 1990.
- Newell, Allen, *Unified Theories of Cognition*, Cambridge, MA: Harvard University Press 1990.
- Margolis, Howard, *Patterns, Thinking and Cognition*, Chicago: University of Chicago Press 1987.
- Humphrey, Nicholas, *The Inner Eye*, London: Faber and Faber 1986.
- Premack, David, *Gavagai! or the Future History of the Animal Language Controversy*, Cambridge, MA: MIT Press 1986.
- Skinner, B. F., *Science and Human Behavior*, New York: Macmillan 1953.
- Sperber, Dan and Wilson, Deirdre, *Relevance: A Theory of Communication*, Cambridge, MA: Harvard University Press 1986.
- Wilsson, L., 'Observations and Experiments on the Ethology of the European Beaver', *Viltrevy, Swedish Wildlife*, 8 (1974) 115–266.