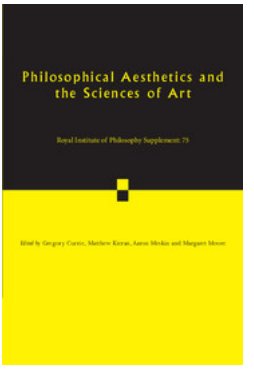

How to do Other Things with Words

Daniel C. Dennett

Royal Institute of Philosophy Supplement / Volume 42 / March 1997, pp 219 - 235
DOI: 10.1017/S1358246100010262, Published online: 12 April 2010

How to do Other Things with Words

DANIEL C. DENNETT

> Nomina si nescis, perit et cognitio rerum.[1]
> (Lucretius, quoted by B. F. Skinner in *Verbal Behavior*, 1957, p. 441).

John Austin's masterpiece, *How to Do Things with Words*, was not just a contribution to philosophy; it has proven to be a major contribution to linguistics, one of the founding documents of pragmatics, the investigation of how we use words to accomplish various ends in the social world. Strangely, not much attention has been paid by philosophers – or by psychologists and linguists – to how we use words in private, you might say, to think. As Wittgenstein (1967, p. 17e) once noted, 'It is very noteworthy that *what goes on* in thinking practically never interests us.'

John Maynard Keynes was once asked if he thought in words or pictures. His reply was 'I think in thoughts.' This is in one way an admirable answer. It abruptly dismisses the two leading mistakes about what thinking might be. But it also encourages that lack of interest that Wittgenstein noted. What, then, are thoughts, if not words or pictures? And more particularly, what is the role of words in thinking? To this overdue question, I want to explore, very tentatively, a few neglected avenues.

Wordless Thought: A Phenomenological Fact

First, I want to sound an alert about an all too familiar line of thought: the idea that thinking goes on in a special functional area of the brain. One of its many guises is what I have called the Cartesian Theatre, but it also appears in the roles of Central Processing, or the Central Executive, or Jerry Fodor's (1983) nonmodular central arena of belief fixation. What is wrong with this idea is not (just) that it (apparently) postulates an *anatomically* discernible central region of the brain – a maximally non-peripheral region, one might say – but that it supposes that there is a functionally identifiable *subsystem* (however located or distributed in the brain) that has some all too remarkable competences achieved by some all too remarkable means. There are many routes to it. Here is

[1] If you are ignorant of names, knowledge of things perishes.

one that starts off in an excellent direction but then veers off. The mistaken fork is not explicitly endorsed by anybody that I can name, but I daresay it has covertly influenced a lot of thinking on the topic.

> (a) One of the things we human beings do is talk to others.
> (b) Another is that we talk to ourselves – out loud.
> (c) A refinement of (b) is to talk silently to oneself, but still in the words of a natural language, often with tone of voice and timing still intact. (For instance, one can often answer such questions as this: 'Are you thinking in English or French as you work on this problem?')

So far, so good, an oft-told tale (going back at least to G. H. Mead 1930, Vygotsky 1934, and Skinner 1957), but watch out for the next step:

> (d) A further refinement of (c) is to drop the auditory/phonemic features (indeed all the features that would tie one's act to a specific natural language) and just think to oneself in *bare propositions*.

Introspection declares that something like this does occur but we must be circumspect in how we describe this phenomenon, since the temptation is to use it to anchor a remarkable set of dubious implications, to wit:

> (e) Since propositions, like numbers, are abstract objects, they would need some vehicles of embodiment in the brain. Moreover, when one thinks in bare propositions, one's thoughts still have one feature of sentences: *logical form*.

(They must have logical form if we are going to explain the phenomenon of reliable deductive inference as the manipulation of these items.)

> (f) So there must be a medium of representation distinct from any natural language (call it the language of thought or Mentalese) that has this feature.

and finally:

> (g) This activity of manipulating formulae of Mentalese is the fundamental variety of thinking – 'thinking proper' or 'real thinking' or even (as somebody recently put it) 'where the understanding happens'.

It has seemed to some that this idea has been properly licensed by an analogy to the central processing unit or CPU of a von Neumann

machine, with its requirement that all instructions be in its proprietary machine language – the only language it 'understands'. But this is a broken-backed analogy, a fantasy that has little or nothing to do with how computers do their work. The 'machine language' understood by the CPU of a traditional computer is entirely composed of imperatives – commands or orders to execute basic tasks which have no external world reference at all. You can't say 'Snow is white' in machine language. You can't even say 'Look for something white' or 'Get some snow' in machine language. What you can say is things like 'Get whatever is is memory location *X* and add it to the number in the accumulator', or 'Output the result in the accumulator' or 'Jump to the instruction in memory location *Y*'. So machine language won't do as a model for Mentalese, even if we want to be computationalists (as I do) about thinking.

We do introspect ourselves thinking, and sometimes it does seem that our thinking is wordless but 'propositional'. This, then, is an indisputable (hetero-)phenomenological fact, bolstered by such widely acknowledged experiences as the tip-of-the-tongue phenomenon (in which we surely do have a particular content 'in mind' and are frustrated in our attempts to find the word that normally clothes it or accompanies it). Publishers have brought out humorous dictionaries of neologisms to fill in the gaps in English: 'sniglets' is proposed, I seem to recall, as a word for those handy little metal sleeves on the ends of shoelaces, and 'yinks' would name the pathetic strands of hair that some men drape carefully but ineffectively over their bald spots. We have all thought of (noticed, reflected upon) these items before – that's why these books are able to strike a humorous chord in us – so it seems obvious that we can have bare or wordless concepts in our consciousness. And so, it seems, we can think bare propositions composed of them.

This kind of thinking is a personal level activity, an intentional activity, something we *do*. It is not just something that happens in our bodies. (Think of the sign that admonishes 'THINK!'). When we think thoughts of this sort, we do, it seems, *manipulate* our thoughts, and it can be difficult or easy work. This too is part of the heterophenomenology of consciousness, and hence forms part of the explicandum of any theory of thinking. But it is *not* given to introspection that these processes, or our perceptual processes, or any of the other cognitive transitions that are not personal actions but do just happen in our bodies, involve the transformation or manipulation of formulae of Mentalese.

Even if the (d) phenomenon occurs quite frequently *in people like us*, professional thinkers, (e) and (f) do not follow, and there are good reasons to resist them. Here is one that is often overlooked. If we

view Mentalese as the lingua franca of all cognition, and view it as 'automatically understood', this apparently secures a 'solution' to the problem of understanding. But in spite of declarations by Fodor and others that Mentalese does not itself require interpretation by the brain that speaks it, this solution-by-fiat is both unsupported (except by an unsound analogy to machine language) and costly: it creates an artefactual problem about the 'access' of consciousness.

How? If Mentalese is the lingua franca of all cognition, it must be the lingua franca of all the *unconscious* cognition in addition to the conscious thinking. Unconscious cognitive processes are granted on all sides, and if it is conducted in Mentalese (as is commonly asserted or assumed by theorists of the language of thought persuasion), getting some content translated into Mentalese cannot be sufficient for getting it into consciousness, even if it is sufficient for getting it understood. There must then be some further translation or transduction, into an even more central arena than central processing, into some extra system – for instance, Ned Block's (1992) postulated consciousness module. Beyond understanding lies conscious appreciation, according to this image, and it needs a place to happen in.[2] This is what I have called the myth of double transduction, and I have criticised it elsewhere (Dennett 1996), so will not pursue it further here.

There are other ways of modelling human thinking, even 'wordless' human thinking, that are much more plausible. Indeed, it is retrospectively astonishing that an idea as bizarrely unbiological as the language of thought has had as robust a career as it has. (What about the genetic code found in DNA? Isn't it a well-attested biological language? Yes, but it too, like machine language, consists of imperatives – orders to create proteins – combined into recipes. The descriptions of body parts and behaviours that emerge from the genetic code's recipes are amazingly indirect; you can't say 'Give this body the belief that snow is white' in DNA, even if something like this belief is innate in polar bears. So the genetic code is also a poor model for the language of thought.)

Contrast the idea of a central processing unit in which orderly transactions are conducted in a common code with the following: a competitive arena in which many different sorts of things happen – Grand Central Station, in which groups of visitors speaking many tongues try to find like-minded cohorts by calling out to each other,

[2] If we follow Jackendoff's idea (1987), this would not be the *central* summit but a sort of ring surrounding that summit – a tempting idea, but not, I think, one to run with. (I'd rather get the benefits of Jackendoff's vision by other routes, other images, but that's a topic for another occasion.)

sweeping across the floor in growing crowds, waving their hands, pushing and shoving and gesturing – and occasionally managing to make contact and co-ordinate activity. How could such an alternative vision work? I have a few hints and hunches to offer.

Making Tools to Think with

Plato, in the *Theaetetus* (196c–d), compares human memory to a huge cage of birds:

> SOCRATES: Now consider whether knowledge is a thing you can possess in that way without having it about you, like a man who has caught some wild birds – pigeons or what-not – and keeps them in an aviary for them at home. In a sense, of course, we might say that he 'has' them all the time inasmuch as he possesses them, mightn't we?
>
> THEAETETUS: Yes.
>
> SOCRATES: But in another sense he 'has' none of them, though he has got control of them, now that he has made them captive in an enclosure of his own; he can take and have hold of them whenever he likes by catching any bird he chooses, and let them go again; and it is open to him to do that as often as he pleases.

The trick is: getting the right bird to come when you need it. How do we do it? By means of technology. We build elaborate systems of mnemonic association – pointers, labels, chutes and ladders, hooks and chains. We refine our resources by incessant rehearsal and tinkering, turning our brains (and all the associated peripheral gear we acquire) into a huge structured network of competences. (No evidence yet unearthed shows that any other animal does anything like that.)

Getting the information into usable form, usable position – that is the task of resource management or resource refinement that faces us and, I want to suggest, we need words for that task. As Goethe once said, 'When ideas fail, words come in very handy.'[3]

Noam Chomsky has often said that birds don't have to learn their feathers and babies don't have to learn their language. I think there is a better parallel between birds and language: a child acquiring language is like a bird building a nest; it is a matter of 'instinctual'

[3] I cited this in Dennett 1995 (p. 370), while doubting the attribution to Goethe, but as several readers informed me, he did indeed have Mephistopheles say it, in *Faust*: 'Denn eben wo Begriffe fehlen, Da stellt ein Wort zur rechten Zeit sich ein'.

or 'automatic' resource-enhancement, taking found objects and constructing something of great biological value – part of what Dawkins (1982) calls the extended phenotype – which blurs the boundary between an organism (or agent) and the environment in which it must act.

Many years ago, Bertrand Russell (1927, pp. 32–3) made a wry observation:

> Animals studied by Americans rush about frantically, with an incredible display of hustle and pep, and at last achieve the desired result by chance. Animals observed by Germans sit still and think, and at last evolve the situation out of their inner consciousness.

Wolfgang Köhler's (1925) early experiments with chimpanzees were the inspiration for Russell's witticism, which helps to perpetuate a common misunderstanding. Köhler's apes did not just sit and think up the solutions. They had to have many hours of exposure to the relevant props – the boxes and sticks, for instance – and they engaged in much manipulation of these items. Those apes that discovered the solutions – some never did – accomplished it with the aid of many hours of trial and error manipulating.

Now were they *thinking* when they were fussing about in their cages? What were they manipulating? Boxes and sticks. It is all too tempting to suppose that their external, visible manipulations were accompanied by, and driven by, internal, covert manipulations – of thoughts about or representations of these objects, but succumbing to this temptation is losing the main chance. What they were attending to, manipulating and turning over and rearranging were boxes and sticks, not thoughts.

They were *familiarising themselves* with objects in their environments. What does that mean? It means that they were building up some sort of perceptuo-locomotor structures tuned to the specific objects, discovering the affordances of those objects, getting used to them, making them salient, etc. So their behaviour was not all that different from the incessant trial and error scrambling of the behaviourists' cats, rats and pigeons. They were acting in the world, rearranging things in the world – without any apparent plan or insight or goal, at least at the outset.

Animals at all levels are designed to tidy up their immediate environments, which are initially messy, confusing, intractable, dangerous, inscrutable, hard to move around in. They build nests, caches, escape tunnels, ambush sites, scent trails, territorial boundaries. They familiarise themselves with landmarks. They do all this to help them keep better track of the things that matter – predators

and prey, mates, etc. These are done by 'instinct': automated routines for improving the environment of action, making a better fit between agent and world.

This wise husbandry of one's own behavioural territory is economically focused on the most important and ubiquitous features. Pre-eminent among these portions of the environment is the agent's own body, of course, always present, its condition always intensely relevant. Animals instinctively groom themselves and engage in behaviours that are apparently designed (although they need not realise this) to repair flaws, to maintain and improve their co-ordination and muscle tone, and, in effect, to familarise themselves with their own bodies. A part of the body that must not be overlooked in this maintenance and improvement schedule is the brain. It, too, can become messy, confusing, inscrutable, an overwhelmingly complex arena of action. So we should expect animals to be instinctually equipped to engage in mindless, automatic routines that tidy up their own brains.

We should especially expect it in *Homo sapiens*, whose huge brains are so plastic, so inundatable, so at the mercy of invading memes and memories. Resource management for a young human brain is, I think, a major task, and we should expect it to be accomplished by activities that are rooted, at the outset, in our biology, in our 'instincts', but which also get enhanced in major ways by techniques that are themselves part of the influx of new resources.

Human infants engage in exploratory activities much like those exhibited by the chimpanzees, handling every object within reach, bringing it to their mouths, looking at it, over and over. One of the fruits of this repetitive activity is a strong capacity to *recognise* and *track* individual things. Tracking and recognition are strictly different phenomena – you can track something you cannot identify at all, except as the whatever-it-is you're tracking. Recognition depends heavily on this familiarisation process. Consider the potent experience of seeing, for the first time since you were, say, five years old, a favourite toy or article of clothing from your childhood; the warm rush of recognition suggests that there has been very little diminution in the associational structures built in your brain so many years before. This strong capacity to recognise does not cover items or individuals or places of mere passing acquaintance with which you have not actively familiarised yourself. And yet active as the process is, it need not be deliberate or directed or thoughtful – it can be unthinking, automatic.

We more mature *Homo sapiens* also engage in a more directed version of this behaviour, purposefully and attentively moving things around in the world in order to solve problems. For instance, most

Scrabble players would be seriously handicapped if they were prevented from sliding the little tiles around on their little shelf. We write words on index cards and slips of paper, doodle and diagram and sketch. These concrete activities are crutches for thinking, and once we get practised at these activities, we can to some degree internalise these manipulations.

Can you alphabetise the words in this sentence in your head?

Yes, probably, but what you do is not easy. It is not easy because you have to *keep track of* things. To perform this stunt, you need to use visual imagery and auditory rehearsal. You work to make each component as vivid as possible, to help you do the required tracking. If we were to expand the task just a little bit by making the words hard to distinguish visually, or by lengthening the sentence, you would find it impossible to alphabetise in your head – and this very sentence is such an instance. Try it.

Your capacity to alphabetise a list of words in your head, or do mental arithmetic, is an advanced trick, not a novice trick, but it doesn't differ in kind in any important way from the capacity to do the same thing with the aid of paper and pencil. You have made a somewhat fancier nest out of your resources – that's all.

What Can We Do with Our Tools?

Even very simple animals are capable of associative learning – what I have called *ABC learning* (for Associative, Behaviourist, Connectionist). We human beings have the capacity for another kind of learning: swift, insightful learning that does not depend on laborious training but is ours as soon as we contemplate a suitable symbolic representation of the knowledge. When psychologists devise a new experimental setup or paradigm in which to test such non-human subjects as rats or cats or monkeys or dolphins, they often have to devote dozens or even hundreds of hours to training each subject on the new tasks. Human subjects, however, can usually just be told what is desired of them. After a brief question-and-answer session and a few minutes of practice, we human subjects will typically be as competent in the new environment as any agent ever could be. Of course, we do have to *understand* the representations presented to us in these briefings, and that's where the transition from ABC learning to our kind of learning is still shrouded in fog. An insight that may help clear the fog is a familiar maxim of artefact making: if you 'make it yourself', you understand it.

Although ABC learning can yield remarkably subtle and power-

ful discriminatory competences, capable of teasing out the patterns lurking in voluminous arrays of data, these competences tend to be anchored in the specific tissues that are modified by training. They are 'embedded' competences, in the sense that they are incapable of being 'transported' readily to be brought to bear on other problems faced by the individual, or shared with other individuals. The philosopher Andy Clark and the psychologist Annette Karmiloff-Smith have recently been exploring the transition from a brain that has only such embedded knowledge to a brain that, as they say, 'enriches itself from within by re-representing the knowledge that it has already represented'. Clark and Karmiloff-Smith note that while there are clear benefits to a design policy that 'intricately interweave[s] the various aspects of our knowledge about a domain in a single knowledge structure', there are costs as well: 'The interweaving makes it practically impossible to operate on or otherwise exploit the various dimensions of our knowledge independently of one another' (Clark and Karmiloff-Smith 1993, pp. 494–5). So opaquely is such knowledge hidden in the mesh of the connections that 'it is knowledge *in* the system, but it is not yet knowledge *to* the system' (ibid., p. 495) – like the wisdom revealed in the precocious single-mindedness with which the newly hatched cuckoo shoulders the competing eggs out of the nest. What would have to be added to the cuckoo's computational architecture for it to be able to appreciate, understand and exploit the wisdom interwoven in its neural nets? (Cf. Dretske 1991, Dennett 1987, pp. 306ff, and Dennett 1992.)

A popular answer to this question, in its many guises, is 'symbols!' The answer is well-nigh tautological, and hence is bound to be right in *some* interpretation. How could it not be the case that implicit or tacit knowledge becomes explicit by being expressed or rendered in some medium of 'explicit' representation? Symbols, unlike the nodes woven into connectionist networks, are movable; they can be manipulated; they can be composed into larger structures in which their contribution to the meaning of the whole can be a definite and generatable function of the structure – the syntactic structure – of the parts. There is surely something right about this, but we must proceed cautiously, since many pioneers have posed these questions in ways that have turned out to be misleading.

To anchor a free-floating rationale (such as the cuckoo's) to an agent in the strong way, so that it is *the agent's own* reason, the agent must 'make' something. A representation of the reason must be composed, designed, edited, revised, manipulated, endorsed. How does any agent come to be able to do such a wonderful thing? Does it have to grow a new organ in its brain? Or could it build this com-

petence out of the sorts of external-world manipulations it has already mastered?

Children enjoy talking to themselves. What might this be doing to their minds?[4] I cannot answer that question yet, but I have some speculative suggestions for further research. Consider what happens early in the linguistic life of any child. 'Hot!' says Mother. 'Don't touch the stove!' At this point, the child doesn't have to know what 'hot' or 'touch' or 'stove' mean – these words are *primarily* just sounds, auditory event-types that have a certain redolence, a certain familiarity, a certain echoing memorability to the child. They come to conjure up a situation-type – stove-approach-and-avoidance – which is not just a situation in which a specific prohibition is typically *heard* but also a situation in which a mimicking auditory rehearsal is encountered. Crudely simplifying, let's suppose that the child acquires the habit of saying to itself (aloud) 'Hot!' 'Don't touch!' without much of an idea what these words means, voicing them merely as an associated part of the drill that goes with approaching and then avoiding the stove – and also as a sort of mantra, which might be uttered at any other time. After all, children are taken with the habit of rehearsing words they have just heard – rehearsing them in and out of context – building up recognition links and association paths between the auditory properties and concurrent sensory properties and other internal states. They are familiarising themselves with these newfound tools.

That's a rough sketch of the sort of process that must go on. This process could have the effect of initiating a habit of what we might call *semi-understood self-commentary*. The child, prompted initially by some insistent auditory associations provoked by its parents' admonitions, acquires the habit of adding a soundtrack to its activities – 'commenting' on them. The actual utterances would consist at the outset of large measures of what psychologists and linguists have called 'scribble' – nonsense talk composed of wordlike sounds – mixed with real words mouthed with much feeling but little or no appreciation of their meaning, and a few understood words. There would be mock exhortation, mock prohibition, mock praise, mock description, and all these would eventually mature into real exhortation, prohibition, praise and description. But the habit of adding the 'labels' would thus be driven into place before the labels themselves were understood, or even partially understood.

I'm suggesting that it's such initially 'stupid' practices – the mere mouthing of labels in circumstances appropriate and inappropriate – that could eventually be turned into the habit of representing

[4] For recent overviews of research in this area, see Diaz and Berk (1992), and Fernyhough (1996).

one's own states and activities to oneself in a new way. As the child lays down more associations between the auditory and articulatory processes on the one hand, and patterns of concurrent activity on the other, this would create nodes of saliency in memory. A word can become familiar even without being understood. And it is these anchors of familiarity that could give a label an independent identity within the system. Without such independence, labels are invisible. For a word to serve as a useful, manipulable label in the refinement of the resources of a brain, it must be a ready *enhancer* of sought-for associations that are already to some extent laid down in the system. Beyond that, words can be arbitrary, and their arbitrariness is actually part of what makes them distinctive. There is little risk of failing to notice the presence of the label; it doesn't just blend into its surroundings. It wears its artificial status on its sleeve.

The habit of semi-understood self-commentary could, I am suggesting, be the origin of the practice of deliberate labelling, in words (or scribble words or other private neologisms), which in turn could lead to a still more efficient practice: dropping all or most of the auditory and articulatory associations and just relying on the *rest* of the associations (and association-possibilities) to do the anchoring. The child, I suggest, can abandon out-loud mouthings and create private, unvoiced neologisms as labels for features of its own activities.

We can take a linguistic object as a found object (even if we have somehow blundered into making it ourselves rather than hearing it from someone else) and store it away for further consideration, offline. Our ability to do this depends on our ability to re-identify or recognise and track such a label on different occasions, and this in turn depends on the label having some feature or features by which to remember it – some guise independent of its meaning.

Once we have created labels and acquired the habit of attaching them to experienced circumstances, we have created a new class of objects that can themselves become the objects of all the pattern-recognition machinery, association-building machinery, and so forth. Like the scientists lingering retrospectively over an unhurried examination of photographs they took in the heat of experimental battle, we can reflect on whatever patterns there are to be discerned in the various labelled exhibits we dredge out of memory.

As we improve, our labels become ever more refined, more perspicuous, ever better articulated, and the point is finally reached when we approximate the near-magical human prowess we began with: the *mere contemplation* of a representation is sufficient to *call to mind* the appropriate lessons. At this point we are well on our way to becoming *understanders* of the objects we have created. We might

call these artefactual nodes in our memories, these pale shadows of articulated and heard words, *concepts*. A concept, then, is a proprietary label which may or may not include among its many associations the auditory and articulatory features of a word (public or private). But words, I am suggesting, are the prototypes or forebears of concepts. Bickerton (1995, p. 111) speaks of 'linguistic concepts (undressed words)', and I am suggesting that the first concepts one can manipulate are 'voiced' concepts or well-dressed words. Only concepts that can be manipulated can become objects of scrutiny for us.

If you have ever tried panning for gold, you know that the right technique of swirling and swishing and draining is not transparently obvious. There is a simple trick, however, that can swiftly lead you to competence or even expertise. Into each panful of muddy sand and gravel you take up, sprinkle a few – half a dozen – tiny lead shots. They are almost as dense as any gold particles present, and will behave almost the same. As you near the end of your swirling, most of the material having been sluiced away, you should be able to see the little group of lead shot emerge once more from hiding. (If they do not appear, you have been too vigorous, and thrown out the baby with the bathwater! Try again, more gently.) Once you get the hang of it, the cluster of lead spheres will chaperone any bits of gold in the pan, which seem to be drawn to their neighborhood as if by magnetism. This activity of seeding a confused mess with something known and reidentifiable, something familiar, and then leaning on the familiar to help track the novel, has a striking parallel in the use of symbols.

This comes out clearly in a recent experiment by Thompson, Oden and Boyson (forthcoming). As Andy Clark explains, in another forthcoming paper (this is fast-breaking news in a new research area),

> The idea...is that learning to associate concepts with discrete arbitrary labels (words) somehow makes it easier to use those concepts to constrain computational search and hence enables the acquisition of a cascade of more complex and increasingly abstract ideas. The claim...is thus that associating a perceptually simple, stable, external item (such as a word) with an idea, concept or piece of knowledge effectively freezes the concept into a sort of cognitive building-block – an item that can then be treated as a simple baseline feature for future episodes of thought, learning and search.

Consider a simple matching task – such as matching one cup to another cup in an array of different objects. Compare that with the

task of matching a relation-between-relations – such as matching a pair of identical items (two shoes) to a different pair of identical items (two cups) amidst a collection of odd pairs (cup and ball, shoe and fork, etc). Or even harder, matching a pair of different items with the only other pair of items that were different. It has been shown that chimps can discriminate the basic relations of similarity and difference, but only language-trained chimps (Premack 1986, Premack and Premack 1983) can master the higher-order judgement required to act on the fact that *this* pair, like *that* pair, consists of a brace of *different* items. (Notice that I did not impute any language of thought to these chimps. They pick up on a higher-order similarity but need not be able to *express* this competence in any verbal way, even to themselves.) As Thompson et al. show, this is due *not* to the minimal syntactic competence such linguistic training has imparted (to Premack's chimps), but to 'simply the experience of associating abstract relations with arbitrary tokens'. The chimps were given familiarity with an arbitrary symbol or label associated with 'SAME' or 'DIFFERENT'; the token, then, is a cognitive crutch, a simplifier of the perceptual world that helps the animal track a pattern that is otherwise just beyond its ken.

So this tracking of higher-order categories is a stunt, not a primitive building-block, and as we learn to throw away the sounds and other associations, we make the stunt ever more sophisticated. In this arena, wordless thought should be seen as a rather exotic specimen instead of the foundation of all cognition. The image I want to explore has it that these rare, hypersophisticated transitions occur when very sketchy images of linguistic representations serve as mnemonic triggers for 'inferential' processes that generate further sketchy images of linguistic representations and so forth. (This sort of short-cut transition can take place in any modality, any system of associations. For a trained musician, the circle of fifths imposes itself automatically and involuntarily on a fragment of heard or imagined music, just the way *modus ponens* intrudes on the perceptions of a trained logician, or the offside line does on a trained soccer player). This way of trying to imagine word-free 'logical inference' makes it look rather like barefoot waterskiing – a stunt that professionals can make look easy, but that is hardly the basic building-block of successful transportation/cognition in the everyday concrete world.

And even when we are very, very good, we sometimes fall back on the old crutches. If your mother tongue is English, you may find that long after you have achieved something approaching fluency in French or German, you have to convert 'vierundzwanzig' or 'quatre-vingt douze' into your more familiar mode of numerical repre-

sentation. Why is this so? Because you have turned your mother-tongue words into highly familiarised objects, anchored by a host of associations. They are birds that not only come when you call, but call forth flocks of further birds.

Consider a familiar human activity that we rely on in many problem-solving circumstances: we ask ourselves explicit questions. This practice has no readily imaginable counterpart in non-linguistic animals, but what does it gain us, if anything? We are calling the birds. To get the right bird to come when you call is a task calling for technical skills if your birds are stored somewhere in your library or on your hard disk, but also if they are stored in your head. You need to develop retrieval strategies, ways of manipulating your own brain. These methods are themselves tools. *You* don't use the tools; they jump into action on their own, bidden by circumstances, and beneath your conscious control. They are tools that have to call themselves into action when needed, beating out the competition.

Have you ever danced with a movie star? Do you know where to buy live eels? Could you climb five flights of stairs carrying a bicycle and a cello? These are questions the answers to which were probably not already formulated and handily stored in your brain (as sentences of Mentalese in your belief box!), and yet these questions are readily – even 'unthinkingly' – answered reliably by most people. How do they do it? By engaging in relatively effortless and automatic 'reasoning'. In the first case, if no recollection of the presumably memorable event is provoked by considering the question, you conclude that the answer is No. The second question initiates a swift survey (pet stores? fancy restaurants? fish markets or live bait dealers?), and the third provokes some mental imagery which reminds you of relevant facts about your own body and its competences. These are facts about oneself that wouldn't 'automatically' come to mind when needed without a little extra provocation, a little calling. The way to get this sort of information into useful position for current purposes is to ask ourselves questions and see what comes back!

Do we need to ask ourselves grammatical questions? Not necessarily. If we look at the history of artificial intelligence, we can detect a series of clashes over the years between the 'neats' and the 'scruffies'. The neats, epitomised by GOFAI (Good Old Fashioned AI – Haugeland 1985), have worked out formal methods for 'querying data-bases', and in these schemes, the logical or grammatical structure of the probes is indeed all-important. The systems work by mechanised formal deduction. This has proven feasible for many real-world applications in which the information stored is itself highly systematic and bounded – airline routes and timetables, stock

market prices and other information, medical records. But such systems are often 'brittle' and hard to revise or expand. They are not at all biological in their shape, at any level, and do not appear remotely applicable to the sorts of open-ended, time-pressured, freewheeling demands of controlling real world engagements. The scruffies, meanwhile, have favoured various less disciplined, less formal ways of teasing the relevant data out of their pigeonholes. At some level of sophistication, and for certain advanced tasks, syntax is surely obligatory. Turning mere 'labels' into names and predicates, nouns and verbs and auxiliaries, is necessary to get the compositionality that permits us – and apparently only us – to extend our comprehension to all topics.

But does this structure have to be in the brain from the outset, part of the underlying computational architecture, or can it be brought in, piecemeal, from the cultural environment, to discipline our brains at a higher level? This is an empirical question, and now that we can dimly see the scruffy alternative, it would be wise to push it as far as we can, knowing that there is plenty of structure in the artefactual world to avail ourselves of, for special purposes. The elementary way to get the birds to come is the way real birds do it, by just calling – yoohoo, in hundreds of varieties, but without syntax. Once the birds have been trained to fly in formation, this will provide further structures out of which to build further competences.

And where does consciousness come into the picture? It is already there, unnoticed, in the activity just described. Mental contents become conscious not by entering some special chamber in the brain, not by being transduced into some privileged and mysterious medium, but by winning the competitions against other mental contents for domination in the control of behaviour, and hence for achieving long-lasting effects – or as we misleadingly say, 'entering into memory'. And since we are talkers, and since talking to ourselves is one of our most influential activities, one of the most effective ways – not the only way – for a mental content to become influential is for it to get into position to drive the language-using parts of the controls. All this has to *happen* in the arena of the brain, in 'central processing', but not under the direction of anything. The Person is the virtual governor, not a real governor; it is the effect of the processes, not their cause.

A common reaction to this suggestion about human consciousness is frank bewilderment, expressed more or less as follows: 'Suppose all these strange competitive processes are going on in my brain, and suppose that, as you say, the conscious processes are simply those that win the competitions. How does *that* make them con-

scious? What happens next to them that makes it true that *I* know about them? For after all, it is *my* consciousness, as I know it from the first-person point of view, that needs explaining!' Such questions betray a deep confusion, for they presuppose that what *you* are is something *else*, some Cartesian *res cogitans* in addition to all this brain-and-body activity. What you are, however, just *is* this organisation of all the competitive activity between a host of competences that your body has developed. You 'automatically' know about these things going on in your body, because if you didn't, it wouldn't be your body!

The acts and events you can tell us about, and the reasons for them, are yours because you made them – and because they made you. What you are is that agent whose life you can tell about. You can tell us, and you can tell yourself. The process of self-description begins in earliest childhood, and includes a good deal of fantasy from the outset. (Think of Snoopy in the *Peanuts* cartoon, sitting on his doghouse and thinking 'Here's the World War I ace, flying into battle...') It continues through life. (Think of the café waiter in Jean Paul Sartre's discussion of 'bad faith' in *Being and Nothingness* (1943), who is all wrapped up in learning how to live up to his self-description as a waiter.) It is what *we* do. It is what *we* are.[5]

References

Bickerton, D. 1995. *Language and Human Behavior*. University of Washington Press

Block, N. J. 1992. 'Begging the Question against Phenomenal Consciousness', (commentary on Kinsbourne and Dennett), *The Behavioral and Brain Sciences* **15**, 205–6

Clark, A. and Karmiloff-Smith, A. 1993. 'The Cognizer's Innards: A Psychological and Philosophical Perspective on the Development of Thought', *Mind and Language* **8**, 487–519

Clark, A. (forthcoming). 'Magic Words: How Language Augments Human Computation' in P. Carruthers and J. Boucher (eds.), *Language and Thought: Interdisciplinary Themes*. Cambridge University Press

Dawkins, R. 1982. *The Extended Phenotype: The Gene as the Unit of Selection*. Oxford and San Francisco: Freeman

Dennett, D. C. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press

—— 1992. 'La Compréhension Artisanale', in *Daniel C. Dennett et les Stratégies Intentionnelles*, in *Lekton* **2**, Université de Québec à Montréal

—— 1993. 'Learning and Labeling' (commentary on Clark and Karmiloff-Smith 1993), *Mind and Language* **8**, 540–47

[5] Parts of this paper are drawn, with slight revisions, from Dennett (1993) and (1996).

—— 1995. *Darwin's Dangerous Idea: Evolution Evaluation and the Meanings of Life*. New York: Simon & Schuster, and London: Allen Lane

—— 1996. *Kinds of Minds*. New York: Basic Books, and London: Weidenfeld & Nicholson

Diaz, R. M. and Berk, L. E. 1992. *Private Speech: From Social Interaction to Self-Regulation*. Hillsdale, NJ: Erlbaum

Dretske, F. I. 1991. 'Replies' in B. McLaughlin (ed.), *Dretske and his Critics*. Oxford: Blackwell

Fernyhough, C. 1996. 'The Dialogic Mind: A Dialogic Approach to the Higher Mental Functions', *New Ideas in Psychology* **14**, 47–62

Fodor, J. A. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press

Haugeland, J. 1985. *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press

Jackendoff, R. 1987. *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press

Köhler, W. 1925. *The Mentality of Apes*. New York: Harcourt Brace and World

Mead, G. H. 1930. *Mind, Self, and Society*. University of Chicago Press

Premack, D. 1986. *Gavagai! Or, the Future History of the Animal Language Controversy*. Cambridge, MA: MIT Press

Premack D. and Premack, A. J. 1983. *The Mind of an Ape*. New York and London: Norton

Russell, B. A. W. 1927. *An Outline of Philosophy*. London: Allen & Unwin

Sartre, J.-P. 1943/1953. *Being and Nothingness*, translated by H. Barnes. New York: Philosophical Library

Skinner, B. F. 1957. *Verbal Behavior*. New York: Appleton-Century-Crofts

Thompson, R., Oden, D. and Boyson, S. (forthcoming). 'Language-Naïve Chimpanzees Judge Relations-between-relations in an Abstract Matching Task', *Journal of Experimental Psychology: Animal Behavior Processes*

Vygotsky, L. S. 1934/1962. *Thought and Language*, edited and translated by E. Hanfmann and G. Vakar. Cambridge, MA: MIT Press

—— 1934/1986. *Thought and Language*, translated by A. Kozulin. Cambridge, MA: MIT Press

Wittgenstein, L. 1967. *Zettel*, edited by G. E. M. Anscombe, G. H. von Wright, and translated by G. E. M. Anscombe. Oxford: Blackwell