THE ROYAL SOCIETY

Rana computatrix to human language: towards a computational neuroethology of language evolution

By Michael A. Arbib

Department of Computer Science, Neuroscience and USC Brain Project, University of Southern California, Los Angeles, CA 90089-2520, USA
(arbib@pollux.usc.edu)

Walter's *Machina speculatrix* inspired the name *Rana computatrix* for a family of models of visuomotor coordination in the frog, which contributed to the development of computational neuroethology. We offer here an 'evolutionary' perspective on models in the same tradition for rat, monkey and human. For rat, we show how the frog-like taxon affordance model provides a basis for the spatial navigation mechanisms that involve the hippocampus and other brain regions. For monkey, we recall two models of neural mechanisms for visuomotor coordination. The first, for saccades, shows how interactions between the parietal and frontal cortex augment superior colliculus seen as the homologue of frog tectum. The second, for grasping, continues the theme of parieto–frontal interactions, linking parietal affordances to motor schemas in premotor cortex. It further emphasizes the mirror system for grasping, in which neurons are active both when the monkey executes a specific grasp and when it observes a similar grasp executed by others. The model of human-brain mechanisms is based on the mirror-system hypothesis of the evolution of the language-ready brain, which sees the human Broca's area as an evolved extension of the mirror system for grasping.

Keywords: biologically inspired robotics; computational neuroethology; mirror system; action-oriented perception; evolutionary modelling; language evolution

1. From *Machina speculatrix* to computational neuroethology

Walter's book *The living brain* (1953; page numbers given below are from the 1961 edition) approaches an understanding of the brain from two very different perspectives: (i) through the analysis of electroencephalograph (EEG) signals of the human brain, monitored by a device of his design, the toposcope, which allowed the simultaneous viewing of waveforms from many electrodes across the scalp; and (ii) through the design of 'biologically inspired' robots, in the form of two electromechanical tortoises, *Machina speculatrix* and *M. docilis*. We here review the robots as a basis for the 'evolutionary' perspective that this article offers on brain and action.

Walter (1953, ch. 5) sets the design criteria for the robots he designed, or their potential successors, as follows.

> Not in looks but in action, the model must resemble an animal. Therefore, it must have these or some measure of these attributes: exploration, curiosity, free-will in the sense of unpredictability, goal-seeking, self-regulation, avoidance of dilemmas, foresight, memory, learning, forgetting, association of ideas, form recognition, and the elements of accommodation.

He dismisses Swiss automata and 1950-vintage computing machines ('a model of nineteenth century predestination'!) as failing to meet these criteria, and then describes two precursors of his work (without bibliographical sources), inventing mock-biological names for each 'species'.

(i) *M. labyrinthea.* Thomas Ross (in 1938), R. A. Wallace (in 1952) and Claude Shannon devised machines that initially find their way through a maze by trial and error, but then store the route. Thereafter, restarted from the same point, the machine will go directly 'home' without error. The machines are, within limits, goal seeking and self-regulating, and possess a form of memory.

(ii) *M. sopora.* He likens Ashby's Homeostat, a network designed to resist perturbations by returning key variables to a designated resting state, to 'a fireside cat or dog which only stirs when disturbed, and then methodically finds a comfortable position and goes to sleep again'.

He then introduces his first 'tortoise' (which is embodied, but its body has wheels not legs, etc.) The 'tortoise' has two miniature valves, two relays, two condensers, two small electric motors and two batteries. It has a steerable photoelectric cell, which makes it sensitive to light, and an electrical contact, which allows it to respond when it bumps into obstacles. The photo-receptor rotates until a light of moderate intensity is registered, at which time the organism orients towards the light and approaches it. However, very bright lights, material obstacles and steep gradients are repellent to the 'tortoise'. The latter stimuli convert the photo-amplifier into an oscillator; 'this causes alternating movements of butting and withdrawal, so that the robot pushes small objects out of its way, goes around heavy ones, and avoids slopes' (p. 114). Walter sees this period of oscillation (a second or so) as providing the machine with 'discernment, since, during that period, its behaviour is not dominated by the light, but, say, the obstacle. The 'tortoise' has a 'hutch', which contains a bright light. When the machine's batteries are charged, this bright light is repellent. When the batteries are low, the light becomes attractive to the machines and 'moderation gives place to appetite and the light continues to exert an attraction until they are well within their quarters', where the machine's circuitry is temporarily turned off until the batteries are recharged, at which time the bright hutch light again exerts a negative tropism (p. 116).

Walter calls his first biologically inspired robot *Machina speculatrix* ('the machine that speculates'), claiming that it exhibits the ability to 'speculate' because it has 'the typical animal propensity... to explore the environment rather than to wait passively for something to happen'. But this is far too behaviourist a view of 'speculation': there is nothing in *M. speculatrix* of the mental gymnastics of human speculation

about possible courses of future events in the endless moving back and forth of *M. speculatrix* under the control of its two 'neurons'.

Walter (1953, ch. 7) presents *M. docilis*, the 'easily taught' machine. This new 'species' is produced by grafting onto *M. speculatrix* a copy of the conditioned reflex analogue (CORA), a circuit designed by Walter to form conditioned reflexes. Its primary role in *M. docilis* is to form associations with inputs from a new sensory channel: a microphone sensitive to sounds. *M. docilis* 'can be taught to come to a whistle by blowing a short blast and showing it a light... light in *M. speculatrix* eliciting an unconditioned reflex of attraction'. In another experiment, Walter 'decided to teach the model that sound means trouble... by connecting CORA to the obstacle avoiding device in *M. speculatrix*, the feedback circuit which made [it]... back and turn away when its shell was touched. Its education by punishment consisted in blowing the whistle and kicking the shell a few times.' Further experiments looked at the effect of having more than one copy of CORA in the brain of *M. docilis*. His ch. 7 further analyses the time-scales of memory involved in the model, and presents some rather preliminary analogies between observations on the circuitry of *M. docilis* and observations of some EEG correlates of human mental activity. These analogies are at a very abstract level, addressing neither the gross anatomy nor the detailed neurophysiology of the human brain.

What most people know of Walter's work is the 'tortoise' *M. speculatrix*, but *The living brain* also contains, as we have seen, a second cyber-creature, *M. docilis*, as well as analyses of actual brain function, including his 'toposcope' for the display of EEG recordings. We thus see two different strands here:

(i) the attempt to find the simplest mechanisms that will yield an interesting class of robot behaviours—and this includes a 'comparative' method in which different additions to *M. speculatrix* yield a variety of different behaviours;

(ii) the attempt to explicate the workings of the human brain.

One can trace a similar duality in the work of Braitenberg. Most people in robotics know him for his 1984 book, *Vehicles*, which is very much in the spirit of *M. speculatrix* and its elaboration. The core argument for *Vehicles* was actually published almost 20 years earlier under the title *Taxis, kinesis, decussation* (Braitenberg 1965) and, even earlier, Braitenberg & Onesto (1960) developed a model of the cerebellum which sought to reconcile Braitenberg's work on neuroanatomy with the role of the cerebellum in the timing of the components of a graceful movement, such as that exemplified in a musical performance. The model was influential, but proved to be wrong, in part because it was developed before Ito's discovery that Purkinje cells, the output cells of cerebellar cortex, are inhibitory (see Eccles *et al*. 1967), but this simply highlights the difference between robotics and neuroscience. In robotics, the test of success is technological—the machine should yield effective performance at reasonable cost. In computational neuroscience, there is a continual dialogue between theory and experiment, as the hypotheses and predictions of models stimulate new experiments, and as new empirical data stimulate the revision or replacement of existing models. Braitenberg (2002) suggests how the very special intrinsic connectivity of the cerebellar cortex may be translated into physiological relations, which lead to propositions about cerebellar function that can be tested experimentally.

In the spirit of Walter and Braitenberg, then, one may focus on one of two approaches:

(a) incremental design of mechanisms that will yield an interesting class of robot behaviours—where the biological inspiration comes from the range of animal behaviours as seen from the 'outside';

(b) the attempt to explicate the workings of the brains of humans and of other creatures.

I add 'and other creatures' because, in his study of the cerebellum, Braitenberg used data from non-human animals to ground his analysis of the structure of cerebellar cortex and has also extensively studied the anatomy of the visual system of the fly, while Walter appealed to studies of Pavlovian conditioned reflexes in dogs and other animals. However, having said that, I think that in both Walter and Braitenberg ones sees somewhat limited interaction between the two approaches (a) and (b). In this article, I wish to emphasize a 'third way':

(c) the attempt to create a comparative computational neuroethology (i.e. a computational analysis of neural mechanisms underlying animal behaviour) in which the brains of humans and other creatures come to be better understood by seeing homologous mechanisms as computational variants which may be related to the different evolutionary history or ecological niche of the creatures that contain them.

Such work can, of course, expand the approach to 'biologically inspired robotics' of (a) by augmenting study of the range of animal behaviours seen externally by the study of the range of brain mechanisms that accompany them. However, the present article will focus on comparative computational neuroethology. The journey into computational neuroethology charted here starts with the paper on 'What the frog's eye tells the frog brain' (Lettvin *et al.* 1959), which treated the frog's visual system from an ethological perspective, showing that certain visual circuitry could be analysed as being specific to the animal's ecological niche, with different cells in the retina and tectum being specialized for detecting predators and prey. In the late 1960s, I turned the question of Lettvin *et al.* into the question 'What does the frog's eye tell the frog?', taking the first step in a career that stresses the embodied nervous system or, perhaps equivalently, an action-oriented view of perception.

In much visually guided behaviour, the animal does not respond to a single stimulus, but rather to some property of the overall configuration. Consider, for example, the snapping behaviour of frogs confronted with one or more fly-like stimuli. Ingle (1968) found that it is only in a restricted region around the head of a frog that the presence of a fly-like stimulus elicits a snap, that is, the frog turns so that its midline is pointed at the stimulus and then lunges forward and captures the 'prey' with its tongue. There is a larger zone in which the frog merely orients itself, and beyond that zone the stimulus elicits no response at all. When confronted with two 'flies' within the snapping zone, either of which is vigorous enough that alone it could elicit a snapping response, the frog exhibits one of three reactions: it snaps at one of the flies; it does not snap at all; or it snaps in between at the 'average fly'. Didday (1976) offered a simple model of this choice behaviour; mathematical analysis of a variant is due to Amari & Arbib (1977). Either may be considered as prototypical for what

has become known as a *winner-takes-all* (WTA) circuit, which receives a variety of inputs and (under ideal circumstances) suppresses the representation of all but one of them; the one that remains is the 'winner', which will play the decisive role in further processing.

In the following years (with the bulk of publications from 1982 to 1995), I worked with a succession of PhD students, starting with R. Lara, to model a variety of aspects of visuomotor coordination in frog and toad. Arbib (1982) is an overview of early progress in this task entitled '*Rana computatrix*: an evolving model of visuo-motor co-ordination in frog and toad'. Although I am not sure if I was consciously aware of my debt at the time, it is clear that the name 'the frog that computes' was inspired by Walter's *M. speculatrix*. The name *R. computatrix* inspired in turn the names of a number of 'species' of 'creatures' developed by computational neuroethologists in the years that followed, including Beer's (1990) computational cockroach *Periplaneta computatrix* and Cliff's (1992) hoverfly *Syritta computatrix*.

With such a background, one may build in two directions, computational neuro-ethology and biologically inspired robotics, with further studies providing cross-links between the two. I close this section with some brief comments on the linkage to robotics. The rest of the paper will then focus on computational neuroethology. It will place studies of frog, rat, monkey and human in an 'evolutionary' framework.

A basic theme of this paper is that the evolutionary study of brains often reveals new functions as emerging through modulation and coordination of existing structures. In other words, to the extent that new circuitry may be identified with the new function, it is not as a module that computes the function autonomously, but rather one that can deploy prior resources to achieve the novel functionality. The discussion of figure 1 below gives an early example of this from the toad (Ewert & von Seelen 1974), showing how circuitry for avoidance behaviour exploits circuitry for approach behaviour. My own approach to schema theory (Arbib 1981) was inspired not only by visuomotor coordination in frog and toad, but also by analysis of human arm and hand movements (the 1979 talk that became Jeannerod & Biguer (1982)). In each case, we see behaviour emerging from the competition and cooperation of perceptual and motor schemas as well as, in more complex behaviours, more abstract coordinating schemas too. Such ideas were, of course, developed independently by a number of authors, and so entered the robotics literature by various routes, of which the best known may be the subsumption architecture of Brooks (1986) and the ideas of Braitenberg cited above; while Arkin's work on behaviour-based robotics (Arkin 1998) is indeed rooted in schema theory (see Arkin (1989), with its explicit linkage of robot behaviours to the model of frog detour behaviour developed by Arbib & House (1987)). Arkin *et al.* (2003) present a recent example of the continuing interaction between robotics and ethology, offering a novel method for creating high-fidelity models of animal behaviour for use in robotic systems based on a behavioural systems approach (i.e. based on a schema-level model of animal behaviour, rather than analysis of biological circuits in animal brains), and describe how an ethological model of a domestic dog can be implemented with AIBO, the Sony entertainment robot. Webb (2001) considers the converse issue: of assessing the extent to which robotics can offer good models of animal behaviours. Similarly, Ruppin (2002) discusses the use of neurally driven evolutionary autonomous agents in neuroscientific investigations. 'Embodied agents' are defined by their semi-autonomous existence in interaction with some environment—whether they are hardware robotic devices or

animals in a physical environment, or software programs that represent key aspects of such a relationship, 'living' in a simulated virtual environment—where they perform tasks such as gathering food, navigating, evading predators and seeking prey and mating partners. Each agent is controlled by an artificial neural network (whether modelled on real brains or defined by some abstract artificial architecture), which receives and processes sensory inputs from the surrounding environment and governs the agent's behaviour by activating the motors that control its actions. What makes them evolutionary is that some of the controlling networks are developed through genetic algorithms that apply computational variants of inheritance and selection to a population of agents in such a way that their neural controllers change automatically from generation to generation as a population of agents are selected and copied with modifications over many generations to survive best in a given environment. What follows, then, can be seen as a contribution to the continuing dialogue between robot behaviour and animal behaviour in which particular emphasis is placed on the search for the neural underpinnings of animal behaviour.

(*a*) *Forms of 'evolution'*

Before proceeding further, I need to delineate the varied forms of 'evolution' that are current in the fields of biologically inspired robotics, neural network modelling and computational neuroethology.

Biological evolution. This is the evolution characterized by Darwin's theory of natural selection and enriched by recent advances in molecular biology and genomics. It must be noted that while many genes have been related to specific components of the phenotype, most genes must interact in a genetic network with many other genes to contribute to a number of features of the phenotype. As a result, the relation between genes and overt patterns of the phenotype may be quite indirect.

Ad hoc evolution. This is the type of 'evolution' exemplified by the transition from *M. speculatrix* to *M. docilis*, and in Braitenberg's vehicles—adding features to a model 'to see what happens'. While essentially unconstrained by biological data, this has the virtue of showing that surprisingly complex behaviours can emerge by putting together relatively few simple mechanisms in such a pseudo-evolutionary sequence (Braitenberg 1984), thus putting into question the simplistic view that every apparently complex behaviour must be characterized by its own explicit and lengthy program.

Genetic algorithms. Here, the inspiration comes from natural selection, but yields a method of parameter optimization in artificial systems. One takes a population of objects with randomly assigned parameters arranged in a 'genotype'. The success of an object determines the likelihood that copies of its genotype will be used in generating (via 'mutation' and 'crossover' operators) the genotypes for the next generation. Over a number of simulated generations, the genotypes that result may come to approach more closely values of the parameters that optimize the objective function. Genetic algorithms have been applied to the optimization of neural nets, but in such studies the 'genotype' has tended to characterize the connection weights of the network rather than the genes that might (indirectly, as noted

above) contribute to network connectivity. Genetic algorithms may also be applied to robots in which the biomechanics, as well as the neural controller, are subject to selection; and even to social systems. It must be stressed again that although genetic algorithms are inspired by the mechanisms of natural selection, the result is a mathematical optimization technique equally applicable to non-biological as to biological systems, and that the 'genotype' for a simulated biological system may bear no resemblance to the biological genotype.

Conceptual neural evolution. This is the approach taken in this paper, seeking to understand complex neural mechanisms through an incremental process. Like ad hoc evolution, the strategy is to add features to a model 'to see what happens'. However, unlike ad hoc evolution, the process *is* constrained by biological data— but data that link behaviour to anatomy and neurophysiology, without a necessary analysis of the underlying genes. The aim is to discover relations between modules (neural circuits at some grain of resolution) that implement basic schemas (functions, as distinct from structures) in simpler species with those that elaborate more elaborate schemas in other species. The approach extends the range of schemas more by module interaction and the addition of modules that modulate old schemas than by implementing schemas in isolation.

Clearly, the 'evolutionary path' described in this last way is *not* necessarily substantiated as the actual path of evolution by natural selection that shaped the brains of the species we can study today. Rather, it is an attempt to show how the methodology of schema theory (introduced in § 3) may help us understand a complex behaviour by the 'evolutionary design' of successively more complex models to better and better approximate the neural realization of that function. The same methodology will be used in relating cortical functions in mammals to the subcortical functions that are more closely homologous to certain non-mammalian forms; it will also be used in refining mammalian models by the successive addition of more brain regions to those that can offer a first approximation to the given function. However, in the latter case, the order of addition will reflect a strategy of modelling or exposition with *no* claim that the regions evolved one after another in the given order. The result definitely leads to an enhanced understanding of the more complex brain. It may also yield hypotheses for the quest for the genetic underpinnings of the evolution of this complexity as the parameters of variation across species become better delimited.

The structure of the paper is as follows. Section 3 gives an account of *R. computatrix* that stresses a quasi-evolutionary framework for relating different subsystems and behaviours. In § 4, for *Rattus computator*, we show how the frog-like taxon affordance model provides a basis for the spatial navigation mechanisms that involve hippocampus and other brain regions in the rat. *Affordances* (Gibson 1966) are features of an object or environment relevant to action—visual processing may exploit them to extract cues on how to interact with an object or move in the environment, complementing processes for categorizing objects or determining their identity. In § 5, we use the term *Macaca computatrix* for models of neural mechanisms of monkey visuomotor coordination and present two specific models. The first, for saccades, shows how interactions between parietal and frontal cortex augment superior colliculus (SC) seen as the homologue of frog tectum; the second, for grasping, continues the theme of parieto–frontal interactions, linking parietal affordances to

motor schemas in premotor cortex. We emphasize extension of the system for grasping to include a mirror system in which neurons are active both when the monkey executes a specific grasp and when it observes a similar grasp executed by others. Finally, § 6 presents *Homo loquens computator* as a conceptual design for a number of planned implementations of subsystems based on the mirror-system hypothesis of the evolution of the language-ready brain. This hypothesis is based on the fact that the mirror system for grasping in monkeys has a homologue located in Broca's area, a key area for language, in the human brain.

The evolutionary strategy of comparative neuroethology (Butler & Hodos 1996) is not, of course, to claim that extant species (in this case, frog, rat, monkey and human) form an evolutionary sequence, but rather to claim that the common ancestors of humans with frog, rat and monkey form an evolutionary sequence in that order, and that many mechanisms of the human brain may have built upon mechanisms already present in the brain of our common ancestor with present-day species X. We thus look to species X for clues about that shared inheritance. However, we must always be on the alert to identify which features of X brains are ancestral, and which reflect evolutionary changes that distinguish species X from the human–X common ancestor.

The research cited below will emphasize models from my own research group, since the aim here is to offer an evolutionary perspective on neuroethology that takes us from action-oriented perception to language, rather than to evaluate alternate models of the relevant phenomena. (The reader may consult Arbib (2003*a*) for many related mini-reviews.) The companion paper (Arbib 2003*b*) considers some of the material reviewed here in a very different perspective, namely that of charting a new approach to computer architecture. Finally, I note that this paper may be read as a sequel to Arbib (1997).

A disclaimer. The reader should not infer from this paper either that the presented models were discovered in the order given here, or that each model is itself the best model of the phenomena (structure and function) which it describes. Some models have stood the test of time, others have seen drastic revision in light of both new experimental data and modelling, at the hands of my group and others. However, all models seem to me to convey important points about brain function and are presented here to provide a viable framework for conceptual neural evolution.

2. *Rana computatrix*

R. computatrix (see, for example, Arbib (1982, 1987, 1989) and Arbib *et al.* (1998) for reviews) is a set of models of visuomotor coordination in frog and toad. We have already noted Didday's model of a prey selector, which can extract the most 'food-like' stimulus from a retinotopic array. Here we offer a brief review of other parts of *R. computatrix*, with a stress on general principles for viewing subcortical visuomotor coordination in an 'evolutionary' framework.

(*a*) *Approach and avoidance in frog and toad*

A schema is a unit of functional analysis, whether in the brain (Arbib 1981) or a robot (Arkin 1998). A given schema, defined functionally, may be distributed across more than one brain region; conversely, a given brain region may be involved in many
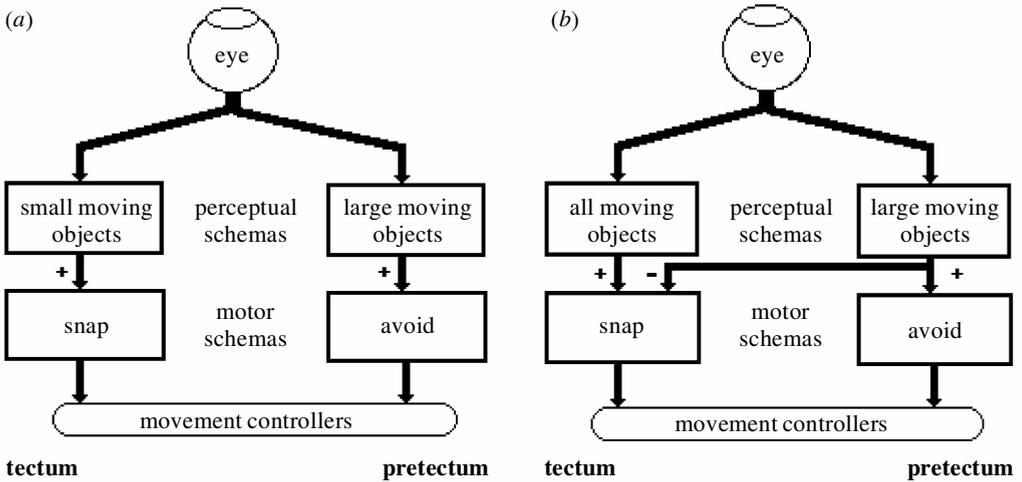
Figure 1. Schema-theoretic models of approach and avoidance behaviour in the toad. (*a*) The initial 'one schema–one brain region' model. (*b*) The refined distributed schema model that is consistent with lesion data.

schemas. A schema model becomes a biological model, as distinct from a purely functional model, when explicit hypotheses are offered as to how the constituent schemas are played over particular regions of the brain.

To exemplify this, consider approach and avoidance in the toad. To simplify rather drastically, we may say that the ability to find food and escape enemies in the frog or toad can be reduced to the ability to tell small moving objects from large moving objects. A frog surrounded by dead flies will starve to death, but the frog will snap with equal 'enthusiasm' at a moving fly or a pencil tip wiggled in a fly-like way. On the other hand, a larger moving object can trigger an escape reaction. Thus a highly simplified model of the functioning of the toad brain has signals from the eye routed to two basic pattern-recognition routines (perceptual schemas): one for recognizing small moving objects (food-like stimuli) and one for recognizing large moving objects (enemy-like stimuli). If the small-moving-object schema is activated, it will, in turn, trigger the motor schema (a system for controlling action) to get the animal to approach what is apparently its prey (this perception–action pair (the left-hand path in figure 1*a*) corresponds to Didday's prey-selector model). If the perceptual schema for a large-moving-object is activated, it will trigger the motor schema for avoidance, causing the animal to escape an apparent enemy (the right-hand path in figure 1*a*).

But to make a biological model, we must relate these schemas to anatomy. Each eye of the frog projects to the opposite half of the brain, especially to the important visual midbrain region called the *tectum*. This projection from the retina to a layered structure of the brain preserves the neighbourhood relationships of the visual field—it is called a *retinotopic* map. Another retinotopic map goes to the *pretectum* (in front of the tectum). If we make the hypothesis that the small-moving-object schema is in the tectum, while the large-moving-object schema is in the pretectum, the above model predicts that animals with a pretectal lesion would continue to approach small moving objects just as the normal animal does, but would not respond at all to large moving objects. However, when Ewert (1987) studied toads in which the pretectum

had been lesioned, he found that not only did the toads respond to small moving objects with approach behaviour, but they also responded to large moving objects with approach behaviour! This observation leads to the new schema-level model shown in figure 1*b* (which reinterprets the work of Ewert & von Seelen (1974)). The new data tell us that, in the absence of the pretectum, the animal must be able to respond to all moving objects with approach. Thus we replace the perceptual schema for *small* moving objects by a perceptual schema for *all* moving objects. On the other hand, in the normal toad, recognition of large moving objects triggers avoidance, and so we leave the right-hand column the way it was. However, although we have now explained the response of the lesioned animal to all moving objects, and the response of the normal animal to large moving objects, it remains to tune the model so that the normal animal will respond to small moving objects with approach but not avoidance. This we can achieve by having an inhibitory pathway running from the perceptual schema for large moving objects (in the pretectum) to the approach schema—or, equivalently, to the schema for all moving objects. This model explains our small database on the behaviour of both normal animals and those with a lesion of the pretectum.

Figure 1*b* provides the first example of 'evolutionary refinement' as we shall study it in this article. In this case, we see on the left a basic system for preying on *all* moving objects. The right-hand system adds functionality, not by adding a disjoint system for escape behaviour, but by *modulating the old system to create a new integrated system that not only adds the new functionality of escape behaviour but also refines the capability of the original prey system.* The other point of this simple exercise is that we have shown how hypotheses about neural localization of subschemas may be tested and refined by lesion experiments. The important point is that biological models can be expressed at the level of a network of interacting 'automata', and that these can really be biological models in the sense that they can be subjected to test at the level of such a coarse-grained network, irrespective of whether or not data or hypotheses are available about the fine-grain implementation of those automata in neural networks.

We may think of each basic schema in figure 1*b* as having its own dedicated neural circuitry. The perceptual schema for recognizing small moving objects is now revealed as more complex than at first appeared. It no longer appears as a module in figure 1*b*, but is instead realized by patterns of activity across the circuits that realize the schemas of the schema assemblage or *coordinated control program* (Arbib 1981) that defines it. If the same basic schema occurs more than once in some coordinated control program, then it must be made clear whether or not the program will require the activity of only one of these instances at any one time. For example, although a more elaborate version (Cobas & Arbib 1992) of the schema for 'prey capture and predator avoidance' contains the 'orient' motor schema in the subschemas for both 'prey capture' and 'predator avoidance', the overall schema is so structured that at most one of those subschemas is active at any one time, and so the circuitry for the 'orient' schema will be activated either with parameters for orienting toward the prey, or with parameters for orienting away from the predator. Moreover, should the competition between the 'prey-capture' and 'predator-avoidance' subschemas be unsuccessful in such a way that both activate the orient schema, it will simply mean that the same circuitry receives simultaneous, conflicting, commands—in which case it might, for example, orient the animal to the 'average' direction. The Cobas–

Arbib model (Cobas & Arbib 1992) generates different motor zones for prey-catching behaviour that match those observed in normal conditions and in studies of lesioned animals, and offers predictions for experiments on both approach and avoidance behaviours.

(*b*) *Detours and path planning*

The motivation for our next example of 'evolutionary refinement' is the behaviour of frogs (Ingle 1976) and toads (Collett 1982) observing a worm through a semi-transparent grating or barrier. Instead of launching directly at its prey as would occur if no barrier were present, the frog often reacts 'appropriately', detouring around the barrier to get its prey. This behaviour already extends the schema picture of figure 1. Now, the perceptual schema for recognizing prey must be augmented by the perceptual schema for recognizing a barrier, and there can no longer be a direct path from prey recognition (which we have already seen to involve both tectum and pretectum) to the triggering of approach behaviour. Rather, there must be some way for this path to be modulated by the recognition of the barrier to yield an indirect detour, rather than the direct response.

In the particular situation under consideration here, the animal must not only recognize prey and barrier, but must locate them in space. If it can recognize that the prey is in front of the barrier or at most a tongue's length behind it, then the animal will indeed snap directly. But if the prey is further behind the barrier, then the animal must use its recognition of where the prey is and where the barrier is to come up with a path that will carry it around the barrier towards the prey.

Epstein (1979) adapted Didday's simple model of the tectum as a row of neurons selecting its maximal stimulus by positing that each visible prey-like stimulus provides a tectal input with a sharp peak at the tectal location corresponding retinotopically to the position of the stimulus in the visual field, with an exponential decay away from the peak. A barrier, on the other hand, provides a trough of inhibition whose tectal extent is slightly greater, retinotopically, than the extent of the barrier in the visual field. Epstein's model can exhibit the choice of a target in the direction of the prey or the barrier edge, but not the spatial structure of the behaviour. To address this, Arbib & House (1987) gave two models for detour behaviour that make use of separate depth maps for prey and barriers. Lara *et al.* (1984) offered an alternative model of detour behaviour in the presence of barriers with gaps in which the recognition of gaps is an explicit step in detour computation. The same paper also offers models—at the level of interacting schemas rather than layers of neuron-like elements—for prey acquisition in environments containing chasms as well as barriers, and for predator avoidance.

Here we discuss the second Arbib–House model, the path-planning model, which associates with each point of the depth map a two-dimensional vector. In place of a single scalar indicating a measure of confidence that there is a target for the first move at the corresponding position in the visual field, the vector is to indicate the preferred direction in which the animal should move were it to find itself at the corresponding position. The model specifies how this vector field is generated and begins to specify how the vector field is processed to determine the appropriate parameters for the coordinated activation of motor schemas. Each prey sets up an attractant field, while each fence post sets up a field for a predominantly lateral movement relative to the
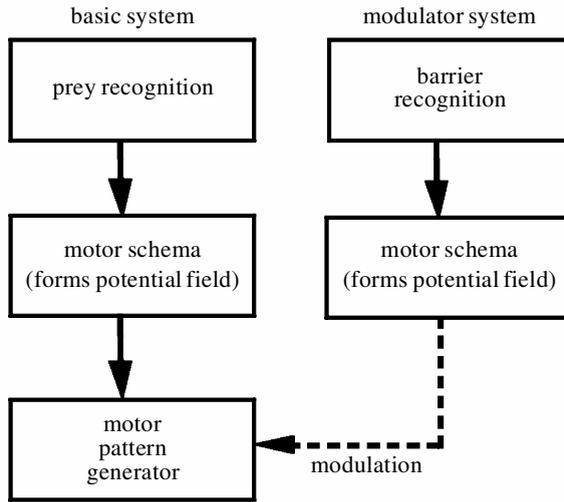
Figure 2. An evolutionary view of brain mechanisms for detour behaviour in frog and toad, in which barrier recognition modulates preexisting schemas for approach (shown here) and avoidance (not shown) behaviour.

position of the post from the viewpoint of the animal. Arbib & House suggest that, in the case of a 'tracking creature', such as the gerbil, the vector field is integrated to yield a variety of trajectories, with a weight factor for each trajectory; whereas, in a 'ballistic creature', such as a frog or toad, processing yields a map of motor targets, appropriately labelled as to type.

I do not claim that the frog brainstem implements the above potential field algorithm in its neural circuitry. Rather, the crucial point is that we have an evolutionary account of how such a system might arise (figure 2). First, the elements of the prey-recognition system (perceptual schema, motor schema and motor pattern generator (MPG)) co-evolve, so that activity in the prey schema can represent a goal in such a way that the approach schema provides the right control signals for the MPG to determine a path to the prey. Then the detour system evolves by combining a perceptual schema for stationary objects with a motor schema that *modulates* the effect that the approach schema has on the MPG.

Figure 1 gave our first example of the evolving subtlety of schema interactions. Here, recognition of small moving objects is not, in fact, localized in any one region, but is rather a system property involving the *modulation* of the tectum by the pretectum. We have now extended the complexity of the environment to which the animal responds—it no longer contains a single prey or a single predator to which the animal may respond with the most basic forms of the 'survival behaviours' of feeding or fleeing, but now contains the more subtle structure of obstacles that can block the animal's path and around which the animal must now be equipped to detour. In evolutionary terms, this corresponds to expansion of the ecological niche in which the animal is well suited to survive.

New schemas often arise as 'modulators' of existing schemas, rather than as new systems with independent functional roles. Further examples of this may be seen in the lesion-based analysis of schemas for approach and avoidance behaviour (Cobas & Arbib 1992). The implications of such work for computational neuroethology and

robot design are reviewed by Cliff (2003) and Arkin (2003), respectively. Corbacho & Arbib (1995) gave a model of how frogs learn to detour as the first step towards a general theory of 'schema-based learning'. This theory attempts to show how the search space for learning can be drastically reduced when learning builds upon partly successful schemas rather than starting *ab initio* with an unstructured neural network.

We close by noting a general property of vertebrate neural control: we see that there may be no place prior to the MPG where the different representation of the animal's situation are brought together. In general, *a multiplicity of different representations must be linked into an integrated whole. However, this may be mediated by distributed processes of competition and cooperation. There need be no one place in the brain where an integrated representation of space plays the sole executive role in linking perception of the current environment to action.*

3. *Rattus computator*

The term *Rattus computator* was introduced by Dean & Redgrave (1989) in an article that developed (in the context of a workshop, many of whose papers were devoted to aspects of *R. computatrix*) the themes of Dean *et al.* (1989). Where most work on mammals had stressed the role of SC (the mammalian homologue of tectum) in approach behaviour (such as the foveation of targets by saccades), Dean & Redgrave were inspired by the studies in frog and toad to discover that there were two response systems in the rat SC—not only one for approach, but also one for avoidance. However, the starting point for the models reported here is the distinction made by O'Keefe & Nadel (1978) between the *taxon* (*behavioural orientation*) *system* for route navigation (as in Braitenberg (1965), a *taxis* is an organism's response to a stimulus by movement in a particular direction, e.g. phototaxis is the drive to move towards light), and the *locale system* for map-based navigation, and proposed that the locale system resides in the hippocampus. Guazzelli *et al.* (1998) qualified the latter assertion, showing how the hippocampus may function as *part of* a cognitive map. The point is that 'place cells' in hippocampus correlate with where the rat is located, but do not correlate with where it wants to be. We thus postulate that the place cells must interact with 'goal cells' and a 'cognitive map' located elsewhere.

But first consider the taxon system. We relate taxis to the notion of *affordances*. Just as a rat may have basic taxes for approaching food or avoiding a bright light, say, so does it have a wider repertoire of affordances for possible actions associated with the immediate sensing of its environment. Such affordances include 'go straight ahead' for visual sighting of a corridor, 'hide' for a dark hole, 'eat' for food as sensed generically, 'drink' similarly, and the various turns afforded by, for example, the sight of the end of the corridor. Since the rat's behaviour depends more on smell than on vision, we should add 'olfactory affordances', but relevant data are sparse.

Both normal and hippocampal-lesioned rats can learn to solve a simple T maze in the absence of any consistent environmental cues other than the T-shape of the maze. If anything, the lesioned animals learn this problem faster than normal animals (O'Keefe 1983). After the criterion was reached, probe trials with an eight-arm radial maze were interspersed with the usual T trials. Animals from both groups consistently chose the side to which they were trained on the T maze. However, many did not choose the 90° arm, but preferred either the 45 or 135° arm, suggesting that

the rats had solved the T maze by learning to rotate within an egocentric orienta-
tion system at the choice point through *ca.* 90°. This leads to the hypothesis of an
orientation vector being stored in the animal's brain, but does not tell us where or
how the orientation vector is stored. One possible model would employ coarse coding
in a linear array of cells, coded for turns from −180 to +180°. From the behaviour,
one might expect that only the cells close to the preferred *behavioural* direction are
excited, and that learning 'marches' this peak from the old to the new preferred
direction. However, it requires a simpler learning scheme to 'unlearn' −90°, say, by
reducing the peak there, while at the same time 'building' a new peak at the new
direction of +90°. If the old peak has 'mass' $p(t)$ and the new peak has 'mass' $q(t)$,
then as $p(t)$ declines towards zero while $q(t)$ increases steadily from zero, the centre
of mass

$$\frac{-90p(t) + 90q(t)}{p(t) + q(t)}$$

will progress from −90 to +90, fitting the behavioural data.

The determination of movement direction is easily modelled by 'rattification' of the
Arbib & House (1987) model of frog detour behaviour. There, prey were represented
by excitation coarsely coded across a population, while barriers were encoded by
inhibition whose extent closely matched the retinotopic extent of each barrier. The
sum of excitation was passed through a WTA circuit to yield the choice of movement
direction. As a result, the direction of the gap closest to the prey, rather than the
direction of the prey itself, was often chosen for the frog's initial movement. The same
model serves for behavioural orientation once we replace the direction of the prey
(frog) by the direction of the orientation vector (rat), while the barriers correspond
to the absence of affordances for movement.

To approach the issue of how a cognitive map can extend the capability of the
affordance system, Guazzelli *et al.* (1998) extended the Lieblich & Arbib (1982)
approach to building a cognitive map as a *world graph*, a set of nodes connected
by a set of edges, where the nodes represent recognized places or situations and the
links represent ways of moving from one situation to another. A crucial notion is
that a place encountered in different circumstances may be represented by multi-
ple nodes, but that these nodes may be merged when the similarity between these
circumstances is recognized. They model the process whereby the animal decides
where to move next, on the basis of its current drive state (hunger, thirst, fear, etc.).
The emphasis is on spatial maps for guiding locomotion into regions not necessarily
current visible, rather than retinotopic representations of immediately visible space,
and yields exploration and latent learning without the introduction of an explicit
exploratory drive. The model shows (i) how a route, possibly of many steps, may
be chosen that leads to the desired goal; (ii) how short cuts may be chosen; and
(iii) through its account of node merging, why, in open fields, place-cell firing does
not seem to depend on direction.

By means of a model employing a competitive Hebbian learning rule responsible
for coding visual and path integration cues, Guazzelli *et al.* (2001) were able to
explain the diversity of place-cell responses observed in a large set of electrophys-
iological experiments with a single fixed set of parameters. Experiments included
changes observed in place fields due to exploration of a new environment, darkness,
retrosplenial cortex inactivation and removal, and rotation and permutation of land-
marks. To code for visual cues for each landmark, we defined two perceptual schemas

representing landmark bearing and distance information over a linear array of cells. The information conveyed by the perceptual schemas is further processed through a network of adaptive layers, which ultimately modulate the resulting activity of our simulated place cells. In path-integration terms, our system is able to dynamically remap a bump of activity coding for the displacement of the animal in relation to an environmental anchor. We hypothesize that path-integration information is computed in the rodent posterior parietal cortex and conveyed to the hippocampus, where, together with visual information, it modulates place-cell activity. It is worth noting that although we place rat before monkey in our pseudo-evolutionary sequence, the modelling went in the reverse order—it was the model by Dominey & Arbib (1992) of dynamic remapping in the saccade system of the monkey that inspired our model of dynamic remapping in the locomotion system of rat. Thus the benefits of conceptual evolutionary modelling can be used to illuminate 'simpler' organisms in the light of available knowledge of 'more complex' organisms, as well as vice versa, as we progress to a unified comparative neuroethology.

The overall structure of the *R. computator* is shown in figure 3. The figure is best understood by 'reading' from right to left.

On the right is the basic motivational system, centred on the hypothalamus, which holds neural representations for basic drives such as fear, hunger and sex. The activity level can depend both on internal state signals (e.g. low blood sugar increases the hunger signal) and by sensory cues (seeing food may provide an 'incentive' signal that increases the hunger signal). As the animal acts, it may or may not be successful in reaching a goal object (e.g. food); as a consequence, the animal may (in the case of food) change its internal state by becoming less or more hungry. Finally, for this block of the figure, the nucleus accumbens is modelled as the locus of reinforcement learning, which yields an adaptive bias signal for action selection dependent on the current internal state.

Before continuing the analysis of the figure, note the following.

(i) Much more is known about the motivational (some call it emotional) system than is included here (see, for example, LeDoux (2000), which notes especially the role of the amygdala, which is not included in figure 3), posing further challenges for the study of *R. computator*.

(ii) The motivational schema shown here has not been incorporated explicitly into the models of monkey and human described below, but the role of nucleus accumbens shown here is paralleled by models that incorporate reinforcement learning in the basal ganglia as a mechanism for the biasing of cortical action selection (see, for example, Dominey *et al.* 1995; Gillies & Arbuthnott 2000; Gurney *et al.* 2001).

In the middle of the figure, we see that sensory inputs determine parietally encoded affordances for the rat's locomotion; these provide the 'menu' for premotor cortex to select (on the basis of a variety of cues) appropriate commands for the MPGs that control the biomechanics of action. The following two points should be borne in mind.

(i) The frog-inspired model might involve midbrain circuitry (including SC, the homologue of tectum). It would thus be useful to return to the data of Dean & Redgrave (1989) and their successors to better understand how to apportion
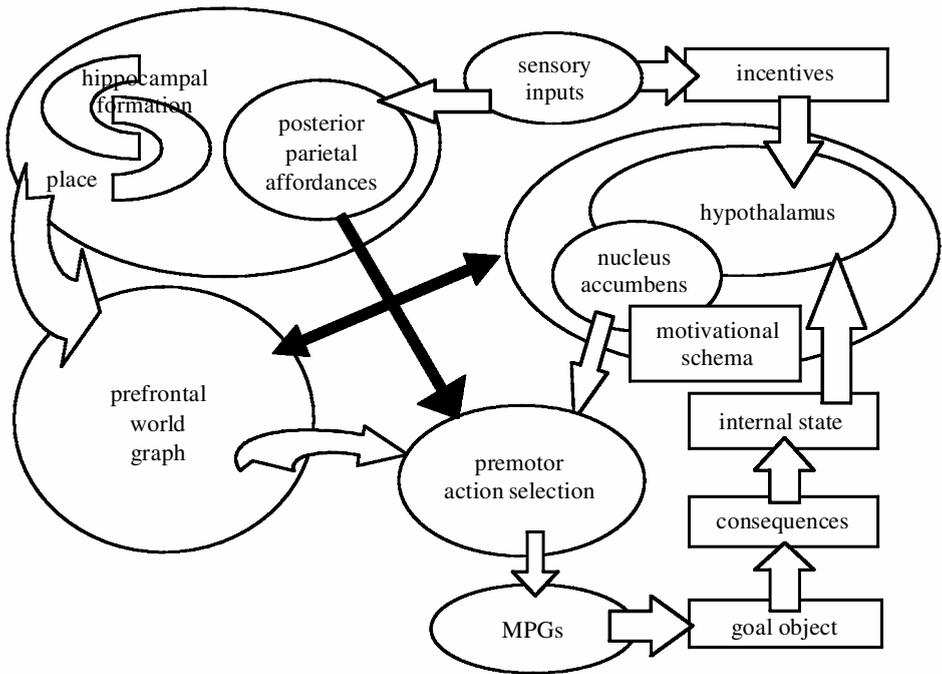
Figure 3. An overview of mechanisms in the current version of *R. computator*. Sensory inputs determine parietally encoded affordances for the rat's locomotion; these provide the 'menu' for premotor cortex to select appropriate commands for MPGs. The right half of the figure illustrates the taxon affordance model. The motivational system, centred on the hypothalamus, holds neural representations for basic drives, which can be updated on the basis of internal state signals and by sensory cues (incentives). The nucleus accumbens is the locus of reinforcement learning, which yields an adaptive bias signal for action selection dependent on the current motivational state. The left of the figure shows the world-graph model. The place ('you are here') system of the hippocampus is augmented by a cognitive map, the world graph, which is posited to be in prefrontal cortex. A dynamic remapping mechanism in parietal cortex allows the animal to update its place-cell encoding on the basis of its recent movements in those case where landmarks encoding the new place are not currently visible.

    affordances and action selection between cortex and midbrain. (Indeed, the model of monkey saccade control in figure 4 does make a more informed attempt at apportionment.)

(ii) More research is required to determine the extent to which action selection really occurs in premotor cortex, and to what extent selection occurs at the level of parietal cortex by restricting the set of 'relevant' affordances.

    Finally, on the left of the figure, we see the systems that augment the *taxon* (*behavioural orientation*) *system* for route navigation with a *locale system* for map-based navigation. Here the place ('you are here') system of the hippocampus is augmented by a cognitive map, the world graph that is posited to be in prefrontal cortex. Mechanisms for updating and using the cognitive map are taken from Lieblich & Arbib (1982)—their implementation in a neural model, and the testing of that model against neurophysiological data, remains a target for future research. The model includes a dynamic remapping mechanism in parietal cortex, which allows the

animal to update its place-cell encoding on the basis of its recent movements in those cases where landmarks encoding the new place are not currently visible.

4. *Macaca computatrix*

Continuing with our quasi-evolutionary progression, we study how, in the monkey, cortex augments SC (the mammalian homologue of frog tectum) in the control of the rapid eye movements known as saccades, and then turn to the visual control of grasping. In both cases, the integration of parietal affordances with premotor motor commands will play a key role (Arbib 1997).

(*a*) *Saccades*

SC is the primary recipient of the projection from the retina to the midbrain, and is a layered structure, with activity of cells in the superficial layers of the SC indicating visual activity but with cells in the deep layers activated only for the target of the upcoming saccade. Both maps are *retinotopic* and in register, i.e. a hypothetical coordinate grid drawn on the retina may be related to a grid on SC such that, given a point on the retina, a vertical penetration at the corresponding point of the SC will yield cells in the upper layers that are activated by stimulation of the retinal locus, and cells in deep SC whose firing correlates with shift of gaze to a target in that retinal direction. It is possible to trigger saccades by stimulation of SC with the length and direction of the saccade encoded retinotopically on the collicular surface. Here, the SC commands an eye movement *relative to the current direction of foveal gaze*, rather than giving commands in a head- or body-centred frame. The projection from SC to neurons called long lead bursters in the brainstem is monosynaptic, while the latency from deep SC activity to motoneurons is polysynaptic. These cells control the brainstem saccade burst generator, which takes as input the parameters of the saccade and provides as output the control signal to the ocular motoneurons—first a burst of activity in burster neurons lasts long enough to bring the eyes to their new position, then maintained activity in tonic neurons causes the oculomotor neurons to fire in such a way as to cause the muscles to hold the eyes in their new position in the orbit (see van Gisbergen & van Opstal (2003) for a current review).

The logic of modelling integrates both structural and functional considerations as, for example, our definition of a schema combines our analysis of some aspect of behavioural or cognitive function with our emerging understanding of the structure of some brain region. As we saw in our discussion of *R. computatrix*, the definition of a schema becomes further refined as we expand it to include neurophysiological and anatomical data on the brain region(s) posited to implement it, and our definition of particular schema will be reshaped as this process of model extension proceeds. The use of the coarse functional description at the schema level, as well as the detailed functional/structural description of the neural level, is to maintain intelligibility as more and more neural details are incorporated into the model to explain a larger and larger body of behavioural/cognitive data. In this spirit, we now turn to a multi-schema view of the saccadic eye-control system. Dominey & Arbib (1992) show how to replace the constituent schemas by neurophysiologically plausible networks in the cerebral cortex and basal ganglia. The discussion here addresses three types of saccade of increasing subtlety.

The simple saccade task. A monkey fixates a spot of light, which later disappears as another spot of light (target point) appears in another location. The monkey is rewarded for making a saccade to the new target at its onset.

The delayed saccade task (Hikosaka & Wurtz 1983). A peripheral target point is briefly illuminated during the display of the fixation point. The monkey is trained to make a saccade to the location of the previously flashed target only after removal of the fixation point, thus showing that it has remembered the location of this target during the period between the removal of the target and the removal of the fixation point.

The double saccade task (Mays & Sparks 1980). Following offset of the initial fixation point (F), targets A and B are successively presented. The total duration of presentation is less than the time required to initiate the first saccade. Reward is contingent on successive saccades from F to A and then from A to B.

The classic models of saccade control address only the simple saccade task. The delayed saccade task requires that the target specification be held in some form of *working memory*, rather than being derived from current retinal input. The double saccade task exhibits a dissociation between the site of retinal stimulation and the metrics of the second saccade it elicits, whereas the initial retinotopic representation of the second target, B, in the colliculus would by itself drive a saccade from F to B, the saccade it elicits starts from A, suggesting that a *dynamic remapping* takes place during the initial saccade to B, effectively subtracting vector FA from vector FB to yield the actual movement specified by the vector AB. Sparks & Mays (1980), using trials in which an intervening saccade changed the position of the eyes after a brief visual target had been extinguished, discovered quasi-visual (QV) cells, the location of whose activity, even if the eyes had moved after the target disappeared, represented the current retinotopic position of the remembered target, while Gnadt & Andersen (1988) observed such remapping in posterior parietal cortex. In the model described here, we hypothesize that the actual remapping is accomplished in PP and that this provides the input to the QV cells of SC.

Dominey & Arbib (1992) used an extensive analysis of the literature to develop the neural model outlined in figure 4. Heavy outlines separate modules representing different brain regions, and each brain region is divided into arrays that model cells with similar neurophysiological responses. For example, there are cells in different regions whose activity is best correlated with the onset of a saccade; other cells have activity best correlated with maintained activity or working memory between the presentation of a stimulus at the beginning of a delay period until the actual response. However, while a great deal of knowledge of the available data went into the construction of the model, the data are not rich enough to get a model that would actually compute and so we, as modellers, had to make a number of hypotheses about missing connections, weights and time constants to get the model to run. The model was then tuned so that the passage from the presentation of visual input to generation of eye movement both matches the external behaviour, and internally, for those populations that were based on cell populations with measured physiological responses, we match, at some level of detail, those responses. In particular, the model detailed how thalamocortical loops could implement target memory (which is thus a schema
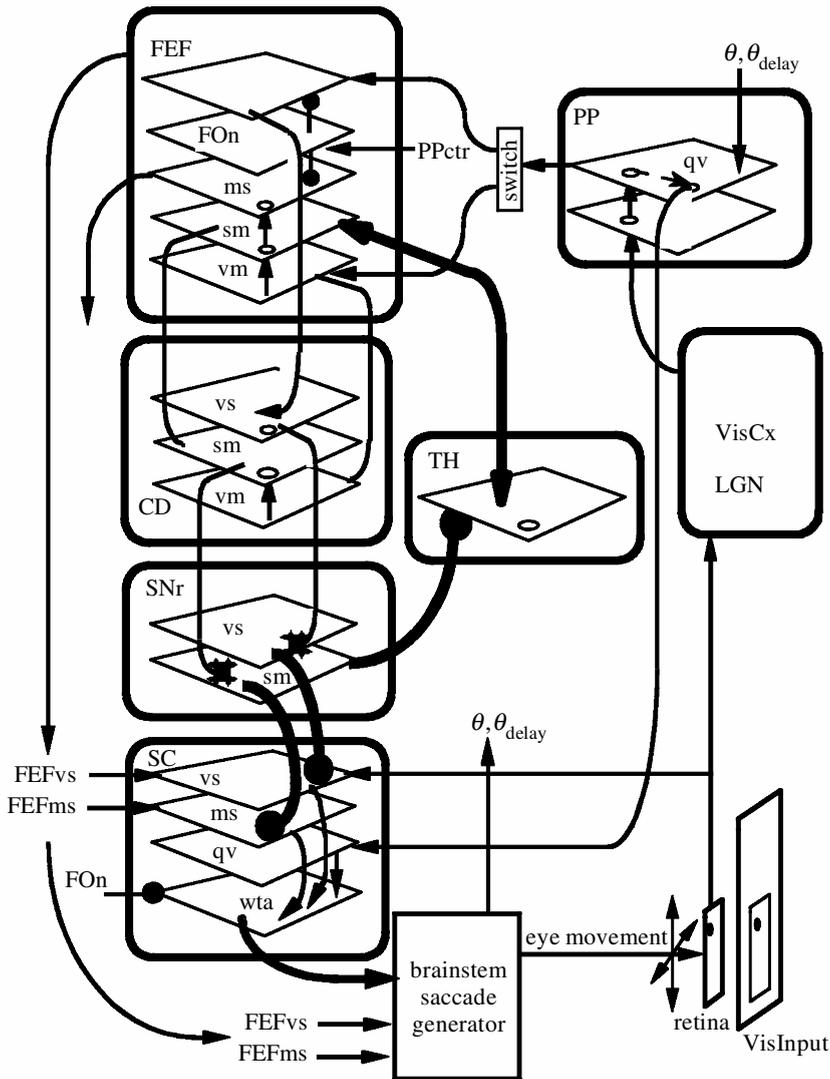
Figure 4. The modular design of the model developed by Dominey & Arbib (1992) to explain the neural mechanisms for memory saccades and double saccades. The model includes SC and brainstem, as well as cortical regions, thalamus and basal ganglia. It represents each brain region by one or more layers of physiologically identified cell types: vs, visual pre-saccade; ms, memory pre-saccade; sm, sustained memory; vm, visual response to memory target; qv, quasi-visual; wta, winner-take-all; PPctr, central element of PPqv; FEF, frontal eye field; PP, posterior parietal cortex; CD, caudate nucleus; SNR, substantia nigra pars reticulata; TH, thalamus (mediodorsal); FOn, fovea on (foveation).

whose implementation is distributed across interacting brain regions), while remapping was conducted by circuitry intrinsic to the region lateral intra-parietal sulcus (LIP) of posterior parietal cortex (shown as 'PP' in figure 4), whose projections then accounted for related activity patterns in other brain regions, including QV activity in SC. In particular, the model then showed how, when a new target representation

was created by remapping, the result would replace the old representation in target memory.

The model again embodies the theme of parietal affordances (here, the encoding of possible saccade targets in the LIP region of PP) controlling premotor action selection (here, the frontal eye fields (FEF)). In this case, the result is that FEF may project one or more peaks of excitation to SC, where it may be combined with concordant or discordant input from retina. We hypothesize (in analogy with the Didday model of frog tectum) that the SC then applies a WTA mechanism to choose one of these peaks to command the brainstem saccade generator. However, the role of the basal ganglia complicates this story. Basically, the portion of the basal ganglia involved in oculomotor control has the caudate ('CD' in figure 4) receive input from cortex, while substantia nigra pars reticulata (SNR) provides output to SC. Intriguingly, the cells in SNR are tonically active, and thus normally will block the SC from triggering a saccade unless the corresponding command signal (either from retina or FEF) can surmount this inhibition. It appears that FEF thus acts on SC in two ways—both by exciting it directly and by pinpointing disinhibition of the SC by the basal ganglia. Briefly, localized activity can excite cells in caudate, which inhibit cells in SNR; since the SNR cells then inhibit the corresponding SC cells less strongly, the latter are more likely to fire. However, it seems strange at first sight that the brain should include this redundant pathway from FEF to SC: one excitatory, the other disinhibitory. Some insight into this was offered by Dominey *et al.* (1995), who used reinforcement learning in the caudate to show how the pathway through basal ganglia could, rather than being redundant, learn to apply appropriate biases on the disinhibition (and thus on target selection) in cases where the choice of target was dependent on visual cues, or where saccades had to be executed in some preassigned sequence. This latter modelling then relates to the role of nucleus accumbens in *R. computator*.

(*b*) *Parietal–premotor interactions in the control of grasping*

The neurophysiological findings of the Sakata group (Taira *et al.* 1990) on the parietal cortex and the Rizzolatti group (Rizzolatti *et al.* 1988) on the premotor cortex indicate that the parietal area anterior intra-parietal (AIP) sulcus and the ventral premotor area F5 in monkeys form key elements in a cortical circuit, which transforms visual information on intrinsic properties of objects into hand movements that allow the animal to grasp the objects appropriately (see Jeannerod *et al.* (1995) for a review). Motor information is transferred from F5 to the primary motor cortex (denoted 'F1' or 'M1'), to which F5 is directly connected, as well as to various subcortical centres for movement execution. Discharge in most F5 neurons correlates with an action; the most common are 'grasping with the hand', 'grasping with the hand and the mouth', 'holding', 'manipulating' and 'tearing'. Rizzolatti *et al.* (1988) thus argued that F5 contains a 'vocabulary' of motor schemas (Arbib 1981). The situation is, in fact, more complex, and 'grasp execution' involves a variety of loops and a variety of other brain regions in addition to AIP and F5. Complementing these data are observations (Ungerleider & Mishkin (1982) in monkey; refined by Goodale & Milner (1992) in human) that we may distinguish a dorsal stream from primary visual cortex (V1) to the parietal cortex, which encodes parameters necessary for interacting with objects (AIP here; LIP for targets for saccades), from a ventral
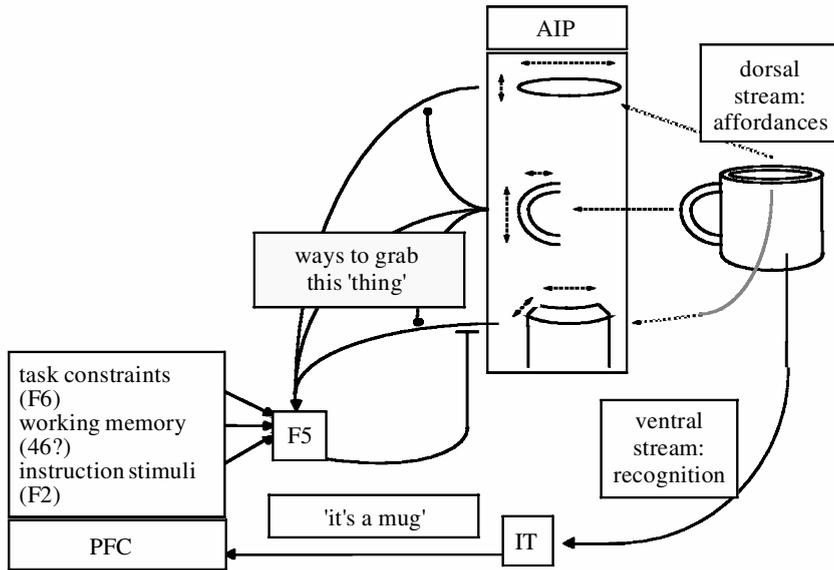
Figure 5. The FARS model of Fagg & Arbib (1998) includes the role of IT and PFC in modulating F5's selection from the menu of affordances sent from region AIP of the parietal cortex. Dorsal stream: extraction of object parameters related to movement (affordances). Ventral stream: object identification.

stream from V1 to the inferotemporal cortex (IT), which serves for recognition of the type or identity of objects.

Within this framework, the Fagg–Arbib–Rizzolatti–Sakata (FARS) model (Fagg & Arbib 1998) provides a computational account of what we call the *canonical system*, centred on the AIP → F5 pathway, showing how it can account for basic phenomena of grasping. Our basic view is that AIP cells encode *affordances for grasping* from the visual stream and send their neural codes on to area F5. Figure 5 shows that F5 may receive signals for multiple affordances, and thus diverse possible actions, through the dorsal stream. Note that the dorsal stream does not know 'what' the object is; it can only see the object as a set of possible affordances. The ventral stream, in contrast, is able to recognize what the object is. The model thus posits that IT, by encoding the nature of the object, can instruct various areas in the prefrontal cortex (PFC) which can help F5 select the appropriate affordance. For example, the appropriate grasp for lifting a mug to drink may be very different for a grasp designed rather to just move it out of the way. Moreover, the brain rarely commands a single movement, but may need to control a sequence of manipulations. To this end, we postulate that, while unit actions may be commanded by F5, each sequence of actions may be stored in the part of the supplementary motor area called pre-SMA (Luppino *et al.* 1990) with administration of the sequence (inhibiting extraneous actions, while priming imminent actions) carried out by the basal ganglia (Bischoff-Grethe *et al.* 2003) using circuitry separate from, but analogous to, that involved in oculomotor control. The main difference is that the output of oculomotor basal ganglia is primarily directed 'downstream' to SC, while the output of skeletomotor basal ganglia is primarily directed 'upstream' to cerebral cortex via regions of the

thalamus. For two different perspectives on the role of basal ganglia, see Bischoff-Grethe *et al.* (2003) and Prescott *et al.* (2003).

We now build on the earlier studies to develop and extend the mirror-system hypothesis of the evolution of the language-ready brain.

(*c*) *The monkey mirror system for grasping*

Rizzolatti *et al.* (1995) found a subset of the grasp-related premotor neurons of F5 that discharge not only when the monkey executes a certain class of actions, as other grasp-related F5 neurons do, but also when the monkey observes more-or-less similar meaningful hand movements made by the experimenter (or another monkey). They call these 'mirror neurons'. Mirror neurons do not discharge in response to simple presentation of objects, even when held by the hand of the experimenter. They require a specific action (whether observed or self-executed) to be triggered. Moreover, mirror neurons do not fire when the monkey sees the hand movement unless it can also see the object—or, more subtly, if the object is not visible but is appropriately 'located' in working memory because it has recently been placed on a surface and has then been obscured behind a screen, towards then behind which the experimenter is reaching (Umilta *et al.* 2001). Thus F5 in the monkey is endowed with an 'observation/execution matching system' and we refer to this system in the monkey brain as the *mirror system for grasping*. We refer to those F5 neurons that are active only during the monkey's own movements as *canonical* F5 neurons. Most current papers seem to view each mirror neuron as coding a specific (more or less broadly tuned) class of actions. However, the available data seem more consistent with the view that mirror neurons encode 'components' of the actions of interest. I suggest that, rather than seek 'the' neuron for each action, we should look at sets of neurons during related movements to try to find the differences and similarities to extract the ensemble code.

The monkey mirror system, as observed neurophysiologically by Rizzolatti *et al.* (1995), is concerned with observation of a single action that is already in the monkey's repertoire. Possible roles for such a mirror system may include the following.

(1) Self-correction, based on the discrepancy between intended and observed self action.

(2) Learning by imitation at the level of a single action.

(3) Social interaction. By anticipating what action another monkey has begun, a monkey can determine how best to compete or cooperate with the other monkey.

To date, analysis of monkey behaviour seems consistent with roles (1) and (3), but monkeys seem poor at imitation in any extended sense. In fact, a major part of our hypothesis on the evolution of language readiness developed in the next section is that evolution of the hominid brain equipped it to support richer and richer forms of imitation.

My hypothesis is that it is function (1) that led to the evolution of the mirror system, i.e. by being able to respond to the relation between hand and object, it laid the basis for generalization from self hand movements to other's hand movements. Function (3) was then a by-product, an 'exaptation', although it is the property that most

people emphasize when discussing mirror neurons. Mirror neurons enable a monkey to see what other monkeys are doing. This relates to the notion of understanding, of not simply seeing a movement as a movement of a hand, but rather recognizing it is a goal-directed action. However, I regard true 'understanding' as involving the interaction of many brain regions, not just activity within the F5 mirror neurons.

The FARS model has two primary components: the recognition of object affordances and the selection by F5 of an appropriate grasp from this menu of affordances. As we move on to the new model, the MNS1 model of the mirror neuron system (MNS) (Oztop & Arbib 2002), we not only have to recognize object affordances, but we also have to recognize how the hand is moving and preshaping. The model incorporates experimental results on other regions of the brain that provide appropriate data for trajectory and preshape. The caudal intraparietal sulcus (cIPS) is not part of this new system, but rather provides information about the shape of the object, which AIP needs to do its job, looking at the orientation, etc., of the surfaces of the object. Two other brain regions are in the parietal cortex: PG, which seems to be particularly good at spatial coding for objects, including motion during interaction of objects as well as self motion, and PF, which seems more related to somatosensory information, touch, etc., but again related to mirror-like responses. Finally, superior temporal sulcus (STS) is in the temporal cortex rather than the parietal cortex, but seems to be very important in detecting biologically meaningful stimuli such as hand movements, as well as having sub-regions that encode motion-related activity. Perrett (1989, 1990) found STS neurons whose properties resemble the visual properties of F5 mirror neurons during action observation, but are not excited during active movements of the observer.

The key criteria for activating an F5 mirror neuron are as follows.

(1) The preshape that the monkey is seeing corresponds to the grasp that the mirror neuron encodes.

(2) The preshape that the observed hand is executing is indeed appropriate to the object that the monkey can see (or remember).

(3) The hand must be moving on a trajectory that will indeed bring it to grasp the object.

In modelling this, we could just have tried to explicitly program the various portions of our overall system to yield appropriate mirror activity. However, the MNS model of Oztop & Arbib (2002) starts with a 'brain' in which the F5 canonical neurons are already controlling an interesting set of grasps and then has the mirror neurons learn to recognize how motion of the hand relative to an object correlates with F5 canonical neuron activity during self-generated movements. We will then see how this can be the basis for the recognition of movements executed by others. An interesting sub-issue is that it will be adaptive for the monkey if it can recognize the other monkey's actions as soon as possible. Thus the ability to activate mirror neurons from smaller and smaller samples of the onset of a trajectory will be an important criterion for further developments in the modelling.

Figure 6 is a diagram of the MNS model of the mirror system. It highlights the components that Oztop & Arbib (2002) see as crucial for operation of the mirror system in the monkey. Along the top diagonal there is a portion of the FARS model
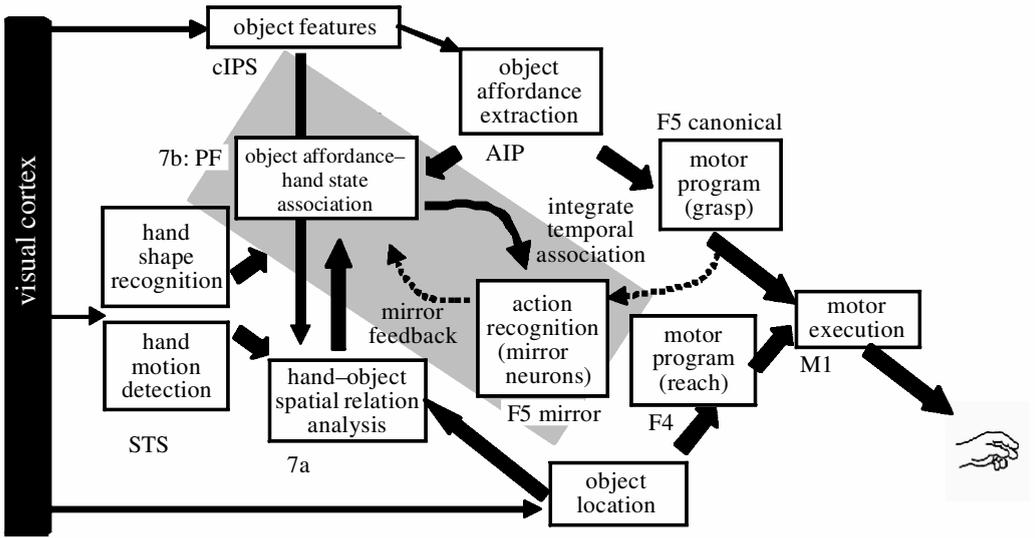
Figure 6. The MNS model of Oztop & Arbib (2002). Top diagonal: a portion of the FARS model. Object features are processed by AIP to extract grasp affordances, and these are sent on to the canonical neurons of F5 that choose a particular grasp. Bottom right: recognizing the location of the object provides parameters to area F4, which programs the reach. The information about the reach and the grasp is taken by the motor cortex M1 to control the hand and the arm. New elements of the MNS model: at the bottom left are two schemas—one to recognize the shape of the hand of the actor being observed by the monkey whose brain we are interested in, and the other to recognize how that hand is moving. Just to the right of these is the schema for hand–object spatial-relation analysis. It takes information about object features, the motion of the hand and the location of the object to infer the relation between hand and object. Just above this is the schema for object associating affordances and hand state. Together with F5 canonical neurons, this last schema provides the input to the F5 mirror neurons.

(cf. figure 5). Object features are processed by AIP to extract grasp affordances, and these are sent on to the canonical neurons of F5 that choose a particular grasp. The bottom right of the figure shows how recognizing the location of the object can provide appropriate parameters to another motor programming area (in this case, F4, which is adjacent to F5 in premotor cortex), which computes the reach. The information about the reach and the grasp is taken by the motor cortex M1 to actually control the hand and the arm.

The other schemas add the new functionality that allows the MNS to do its job. At the bottom left we have two schemas that may be localized in area STS of the monkey brain: one to recognize the shape of the hand of the actor being observed by the monkey whose brain we are interested in, and the other to recognize how that hand is moving. Just to the right of these is the schema for hand–object spatial-relation analysis. It takes information about object features, the motion of the hand and the location of the object to infer the relation between hand and object. Just above this is the schema for associating object affordances and hand state, which may be in area 7b (= PF) of the monkey brain. We want to understand how information coming from the F5 canonical neurons during the monkey's own movements can be used to enable the F5 mirror neurons to learn how to recognize actions. We also want to understand how the mirror neurons are activated not only as an accompaniment

to the monkey's own movement, but also when the monkey observes a similar action by someone else.

We recognize an action when we see that the way in which the hand is moving and shaping is indeed appropriate to one of the affordances of an object. Clearly, in the initial stage of someone else's grasping movement, you may not yet be sure as to whether they will be engaged in a precision pinch or a power grip or whatever. Thus it becomes necessary to monitor the possible movement of a variety of portions of the hand along what may turn out to be the appropriate opposition axis for the movement that will be finally recognized. For this reason, Oztop & Arbib (2002) have defined the *hand state* to characterize the hand in relation to the object to be grasped (see their paper for the particular choice of the seven parameters that define the hand state in the model, and thus the seven-dimensional trajectory $F(t)$ that mirror neurons in the model must learn to classify). By being defined in terms of relationships between hand and object, the hand state has the beauty that it will work just as well for observing how well another monkey's hand is moving to grasp an object (function (3) above) as for measuring how the monkey's own hand is moving to grasp the object (function (1) above). This is the secret that allows self-observation by the monkey to train a system that can be used for observing the actions of others and recognizing just what those actions are.

The details of Oztop & Arbib (2002) are beyond the scope of this article, but the key idea in simulating the 'core of the mirror circuit' (the so-called MNS1 model) is that, starting from grasps that are currently within the repertoire of the system's canonical neurons, the mirror neurons learn to recognize whether the motion of the fingers of the hand is directed towards forming a preshape appropriate to the observed object and whether the motion of the wrist is on a trajectory appropriate to bringing the hand towards the object. This is achieved by a mechanism whereby cells that are initially triggered by F5 canonical activity learn to respond so well to a range of hand-state trajectory input associated with it that eventually these cells can respond to input concerning the object affordance and hand-state trajectory even when the monkey itself is not performing the grasp, i.e. when there is no activity of the F5 canonical neurons. Given the ingenious construction of the hand state, we now have a system that can respond to the actions of others as well as to self-generated actions.

Although MNS1 was constructed as a model of the development of mirror neurons in the monkey, we believe that it serves equally well as a model of the development of mirror neurons in the human infant. A major theme for future modelling, then, will be to clarify which aspects of human development are generic for primates, and which are specific to the human repertoire. The MNS1 model makes the crucial assumption that the grasps that the mirror system comes to recognize are already in the (monkey or human) infant's repertoire. Present modelling within my group postulates the following stages prior to and during the development of the mirror system for grasping in the infant.

(i) The child refines a crude map (SC) to make unstructured reach and 'swipe' movements at objects; the grasp reflex occasionally yields a successful grasp (see Kuperstein (1998) and Baraduc *et al.* (2001) for two examples of related modelling).

(ii) The child develops a set of grasps that succeed by kinesthetic, somatosensory criteria (the infant-learning-to-grasp model (ILGM) of Oztop *et al.* (2003)).

(iii) AIP develops as affordances of objects become learned in association with successful grasps. Grasping becomes visually guided; the grasp reflex disappears.

(iv) The (grasp) MNS develops driven by visual stimuli relating hand and object generated by the actions (grasps) performed by the infant him/herself (see the MNS1 model described above).

(v) The infant (with maturation of visual acuity) gains the ability to map other individual's actions into his internal motor representation.

(vi) Then the infant acquires the ability to create (internal) representations for novel actions observed and develops an action-prediction capability.

We assume that stages (i)–(iv) are much the same in monkey and human, but that stages (v) and (vi) are rudimentary at best in monkeys, somewhat developed in chimpanzees and well developed in human children. In terms of figure 6, we might say that if MNS1 were augmented to have a population of mirror neurons that could acquire population codes for observed actions not yet in the repertoire of self-actions, then in stage (v) the mirror neurons would provide training for the canonical neurons, reversing the information flow seen in the MNSI model. We note that this raises the further possibility that the human infant may come to recognize movements that are not only not within the repertoire, but which never come to be within the repertoire. In this case, the cumulative development of action recognition may proceed to increase the breadth and subtlety of the range of actions that are recognizable but cannot be performed by children.

The important point I wish to emphasize is that such learning models, and the data they address, make it clear that *mirror neurons are not restricted to recognition of an innate set of actions, but can be recruited to recognize and encode an expanding repertoire of novel actions.*

5. *Homo loquens computator*

(*a*) *Human mirror system for grasping*

While our understanding of the mirror system for grasping in the monkey brain rests on neurophysiological recording of the activity of single cells in selected brain regions, the discovery that there is also a mirror system for grasping in the human brain rests on techniques for measuring regional cerebral blood flow to get a measure of how activity in various brain regions differs from task to task. The two main methods are positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). Rizzolatti *et al.* (1996) used PET imaging to show that Broca's region is activated by observation of hand gestures. They compared three experimental conditions: object observation (control condition); grasping observation (the subject observed someone else grasp an object); and object prehension (the subject grasped the object). The most striking result was highly significant activation in the rostral part of Broca's area for both the execution and observation of grasping. Thus a key area in the human language system is a possible mirror system for grasping! Moreover, Matelli (as reported by Rizzolatti & Arbib (1998)) provides an anatomical argument that F5 in the monkey is homologous to (i.e. shares an evolutionary history as a distinguishable brain region with) area 45 in the human (Broca's area = areas 44 + 45). (For further analysis, see Arbib & Bota (2003).)

(b) *The mirror-system hypothesis*

Rizzolatti *et al.* (1995) thus hypothesized that the functional specialization of human Broca's area derives from an ancient mechanism related to the production and understanding of motor acts. The 'generativity', which some see as the hallmark of language, is present in manual behaviour, which can thus supply the evolutionary substrate for its appearance in language. Rizzolatti & Arbib (1998) further developed these ideas in their paper 'Language within our grasp'. Their theory is within the tradition that roots speech in a prior system for communication based on manual gesture—and views the present human capacity for language as based not on speech alone, but rather on communication integrating vocal, facial and manual gestures. What it adds to the manual origins hypothesis is that the mirror system provides a possible neural 'missing link' in the evolution of human language readiness.

The mirror-system hypothesis. Broca's area in humans evolved from a basic mechanism *not* originally related to communication—the *mirror system for grasping in the common ancestor of monkey and human.* The mirror system's capacity to generate *and* recognize a set of actions provides the evolutionary basis for *language parity*, in which an utterance means roughly the same for both speaker and hearer.

It is important to also note that the mirror-system hypothesis does not say that having a mirror system is equivalent to having language. Monkeys have mirror systems but do not have language, and we expect that many species have mirror systems for varied socially relevant behaviours. Moreover, it does not say that the ability to match the perception and production of *single* gestures is sufficient for language.

The monkey mirror system, as observed neurophysiologically by Rizzolatti *et al.* (1995), is concerned with observation of a single action that is already in the monkey's repertoire. 'Beyond the mirror' (as so conceived) lies the imitation of complex behaviours, 'parsing' them into variations of familiar elements and then being able to repeat the observed structure composed from those elements. Arbib (2002) expanded the mirror-system hypothesis by adding two imitation stages to those discussed by Rizzolatti & Arbib (1998) to define seven stages of evolution, from manual grasping through imitation to language. I do not argue for a sharp transition from one stage to the next; rather, I suggest a process of cumulative changes whereby creatures with many of the skills of stages 1 to $n + 1$ gradually emerged from creatures who possessed only the skills of stages 1 to n.

(1) Grasping.

(2) A mirror system for grasping (i.e. a system that matches observation and execution).

(3) A simple imitation system for grasping.

(4) A complex imitation system for grasping.

(5) 'Proto-sign': a manual-based communication system, breaking through the fixed repertoire of primate vocalizations to yield a combinatorially open repertoire.
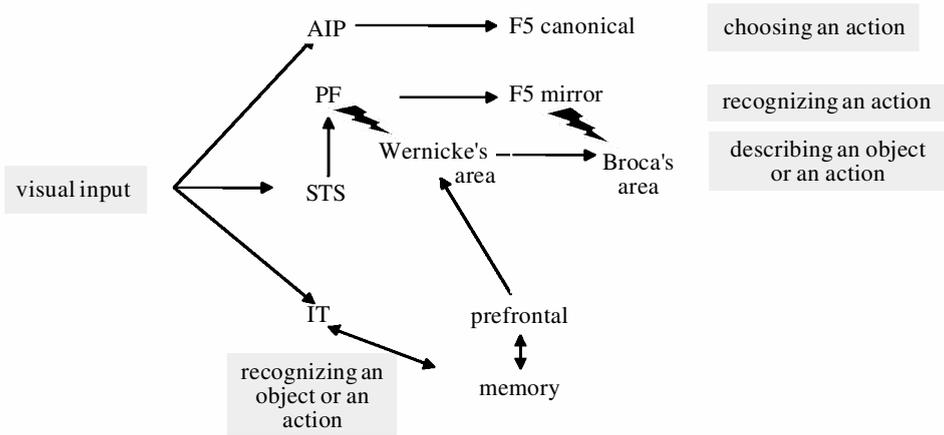
Figure 7. A high-level view of an early pass (Arbib 2001) on a mirror-system-based neurolinguistics. We postulate the cumulative emergence of three fronto-parietal systems: choosing an action (cf. figure 5) → recognizing an action (cf. figure 6) → describing an action (which takes us into the domain of proto-sign and thence to language).

(6) 'Proto-speech': a vocal-based communication system that breaks through the closed nature of primate vocalizations as a result of the 'invasion' of the vocal apparatus by collaterals from the communication system based on F5/Broca's area.

(7) Language: the change from action-object frames to verb-argument structures to syntax and semantics; co-evolution of cognitive and linguistic complexity.

We have already given a full account of stages (1) and (2) and now move on to the later stages. Note the careful use of the term 'proto-speech' in stage (6), rather than the term 'speech'. This is because 'speech' may be taken to mean 'spoken *language*' and I want to emphasize the possibility of an evolving capacity for *proto-speech* as the open-ended production and perception of sequences of vocal gestures, without these sequences *at first* constituting a language. Arbib (2002) spells out these stages, and my current hypothesis is that stages (4)–(6) were present in pre-human hominids, but that the 'explosive' development of proto-speech may have been coupled with further evolution of imitation abilities, and that what we know as language (stage (7)) depended on 'cultural evolution' well after biological evolution had formed modern *Homo sapiens*. This remains speculative, and one should note that biological evolution may have continued to reshape the genome for the human brain even after the skeletal form of *H. sapiens* was essentially stabilized, as it certainly has done for skin pigmentation and other physical characteristics. However, the fact that human children can master any language equally well irrespective of their genetic community shows that these changes are not causal with respect to the structure of language.

To close this discussion, I present figure 7, which provides a crude framework for development of neurolinguistics within the framework offered by the mirror-system hypothesis (see Arbib 2001 for details; an updated version appears in Arbib & Bota (2003)). It shows the cumulative emergence of three fronto-parietal systems: at the top, choosing an action (cf. the FARS model); in the middle, recognizing an action (cf. the MNS model); and, finally, describing an action (which takes us into the

domain of proto-sign and thence to language). When we compare the first two levels as sketched in figure 7 to the greater richness of figures 5 and 6 (which themselves greatly simplify the mechanisms of the monkey brain), we get some sense of the challenges that lie ahead in developing, and carrying out detailed simulations of, *H. loquens computator*. Clearly, this is all too brief, but should provide the key points for our discussion of conceptual neural evolution. Much more detail is provided by Arbib (2003*c*).

To close the section, I mention two examples, which may signal a confluence between robotics and language.

Lowe *et al.* (2002) have constructed a generic human-modelling system to produce human avatars (the graphics equivalent of a robot), and used it as the basis for a tuition system for Auslan, the Australian sign language (Johnston & Wilkin 1998). Using the human-modelling system, the Auslan tuition system provides an interactive extensible interface for learning sign language. Important sign-language features such as hand shape, position, orientation and motion are clearly visible, while user-controllable sign position and viewing angle allow these features to be examined closely by the user. Current work is extending the human-modelling system to display facial expressions and visual speech, which are important components of Auslan. For example, the signs 'angry' and 'complain' differ only in facial expression, whereas 'today' and 'now' differ only by the word that is mouthed while signing.

Billard (2000) has developed a preliminary model of primate imitative ability, to explore hypotheses about the posited role of imitation (stages (3) and (4)) in our theory of language evolution. The aim was to build a comprehensive, though simplified, model of the visuo-motor pathway behind learning by imitation, from processing real video data to directing a complete dynamic simulation of a humanoid. This model is biologically inspired in its function, as its modules have functionalities similar to that of specific brain regions, but the modules are composed of artificial neural networks whose neurons are little constrained by biological data. It is loosely based on neurological findings in primates and incorporates abstract models of some brain areas involved in visuo-motor control, namely STS, the spinal cord, the primary motor cortex (M1), the dorsal premotor area (PMd) and the cerebellum. The model of head, trunk, arms and hands is implemented in a dynamic simulation of a 65-degrees-of-freedom avatar, which allows it to approximate the motions of a humanoid. Learning of movements is done in the PMd and cerebellum modules, which are implemented using the dynamical recurrent associative memory architecture (DRAMA) (Billard & Hayes 1999). The STS–PMd–M1 interconnection is a simplified version of the mirror-neuron model described above. Arbib *et al.* (2000) report on a methodology for inferring predictions from such models for the result of brain imaging using fMRI. fMRI activity of the cerebellum and PMd modules was predicted from the model for imitation against observation tasks in a finger-tapping experiment and compared to the real fMRI activity of the same regions measured by Iacoboni *et al.* (1999). We saw higher ratios for fMRI data than for synthetic data. The point was not to claim that we had achieved the final model of imitation, but rather to highlight the importance of being able to use multiple sources of data in validating a model, and the utility of synthetic fMRI in bringing human-brain imaging to bear in developing a model of 'the supporting neural networks', whether in the control of an avatar or an actual humanoid robot used as a testbed for testing theories in computational neuroscience.

6. Conclusion

With this, we may be closing the circle back to the work of Walter which started this article. The Latin noun *specula* (f.) means a look-out, or watchtower, and is related to the noun *speculator* (m.), which switches to the feminine form *speculatrix* to agree in gender with *Machina* in *M. speculatrix*, 'the machine that speculates'. *Specula* is surely the cognate of *speculum* (m.), which means mirror. Would it be too fanciful, then, to say that the complex computational model of the brain I am developing may indeed give rise to a new *M. speculatrix* as the machine that makes full use of mirror neurons?

In any case, the time comes to briefly recapitulate the path we have trodden.

First, we charted four forms of evolution: biological evolution (cf. Darwin's theory of natural selection); ad hoc evolution (exemplified by the transition from *M. speculatrix* to *M. docilis*, and in Braitenberg's vehicles); genetic algorithms (a method of parameter optimization in artificial systems inspired by natural selection); and conceptual neural evolution (seeking to understand complex neural mechanisms through an incremental process). We then agreed to take an approach to computational neuroethology (computational modelling of the neural mechanisms mediating animal behaviour) that exemplified conceptual neural evolution, though it remains our goal to make the models of conceptual evolution hew more and more closely to studies of biological evolution. Moreover, at various places, we saw convergences between our biological modelling and biologically inspired approaches to the study of robots and intelligent agents.

With this, we offered a conceptual evolution of the vertebrate nervous system that took us through four stages.

(i) *Rana computatrix* (the computational frog). Here we saw basic mechanisms for approach and avoidance in frog and toad and for detours and path planning (the latter an inspiration for work in behaviour-based robotics), which centred on the construction of a retinotopic map in tectum for the visual guidance of behaviour.

(ii) *Rattus computator* (the computational rat). We noted the distinction between the *taxon* (*behavioural orientation*) *system* for route navigation and the *locale system* for map-based navigation, and showed that the taxon system could be modelled using concepts gleaned from our study of *R. computatrix*, but now with the idea that mechanisms of the SC (the mammalian homologue of the frog's tectum) are 'reflected' into cortical structures in which the parietal cortex extracts affordances (visually coded potentialities for action) and premotor cortex (part of the frontal cortex) determines the appropriate motor schemas for action. We then showed how this basic capability may be augmented by a cognitive map that resides *in part* in the hippocampus. Since 'place cells' in the hippocampus correlate with where the rat is located, but do not correlate with where it wants to be, we postulated that the place cells must interact with 'goal cells' and a 'cognitive map' located elsewhere.

(iii) *Macaca computatrix* (the computational monkey). We looked first at the control of saccades, seeing how the SC controls eye movements in a fashion homologous to the way in which tectum controls whole body movement ion the frog,

and then saw how the theme of parietal affordances and frontal motor schemas was carried forward in the relation of posterior parietal cortex and frontal eye fields in the control of saccades. We then saw how this theme was further expressed in the neural mechanisms in the monkey for control of grasping. We then noted a special refinement of this latter system, the monkey mirror system for grasping, in which a subset of the cells of premotor cells involved in grasping are active not only when the monkey performs a certain type of grasp, but also when the monkey observes others (monkey or human) performing a more-or-less similar grasp.

(iv) *Homo loquens computator* (the talking computational human). Here we briefly charted the mirror-system hypothesis for a major evolutionary leap—the ability of humans to learn language whereas other creatures cannot. The key observation was that the human mirror system for grasping appears to be in Broca's area, suggesting that the ability to match the perception and production of hand movements may be crucial to the evolution of human-brain mechanisms that support language, with proto-sign possibly providing the scaffolding for proto-speech en route to the language-ready brain.

All this is but the 'tip of the iceberg' in understanding how the complexity of the human brain may be rooted in the complexities (but simpler complexities) of the brains of our evolutionary cousins. What has made this relevant to the celebration of the legacy of Walter is the growing realization that this exercise in charting conceptual evolution of the vertebrate nervous system not only promises increasing convergence with attempts to probe the biological evolution of brain structures, but may also offer new brain operating principles that will advance the design principles for a new generation of biologically inspired robots.

References

Amari, S. & Arbib, M. A. 1977 Competition and co-operation in neural nets. In *Systems neuroscience* (ed. J. Metzler), pp. 119–165. Academic.

Arbib, M. A. 1981 Perceptual structures and distributed motor control. In *Handbook of physiology—the nervous system. II. Motor control* (ed. V. B. Brooks), pp. 1449–1480. Bethesda, MD: American Physiological Society.

Arbib, M. A. 1982 *Rana computatrix*: an evolving model of visuo-motor coordination in frog and toad. In *Machine intelligence* (ed. J. E. Hayes, D. Michie, & Y. H. Pao), vol. 10, pp. 501–517. New York: Ellis Horwood.

Arbib, M. A. 1987 Levels of modeling of visually guided behavior (with peer commentary and author's response). *Behav. Brain Sci.* **10**, 407–465.

Arbib, M. A. 1989 Visuomotor coordination: neural models and perceptual robotics. In *Visuomotor coordination: amphibians, comparisons, models and robots* (ed. J.-P. Ewert & M. A. Arbib), pp. 121–171. New York: Plenum.

Arbib, M. A. 1997 From visual affordances in monkey parietal cortex to hippocampo-parietal interactions underlying rat navigation. *Phil. Trans. R. Soc. Lond.* B **352**, 1429–1436.

Arbib, M. A. 2001 The mirror system hypothesis for the language-ready brain. In *Computational approaches to the evolution of language and communication* (ed. A. Cangelosi & D. Parisi), ch. 11, pp. 229–254. Springer.

Arbib, M. A. 2002 The mirror system, imitation and the evolution of language. In *Imitation in animals and artifacts* (ed. C. Nehaniv & K. Dautenhahn), pp. 229–280. Cambridge, MA: MIT Press.

Arbib, M. A. (ed.) 2003*a* *The handbook of brain theory and neural networks*, 2nd edn. Cambridge, MA: MIT Press.

Arbib, M. A. 2003*b* Towards a neurally inspired computer architecture. *Natural Computing* **2**, 1–46.

Arbib, M. A. 2003*c* From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav. Brain Sci.* (In the press.)

Arbib, M. A. & Bota, M. 2003 Language evolution: neural homologies and neuroinformatics. *Neural Netw.* (In the press.)

Arbib, M. A. & House, D. H. 1987 Depth and detours: an essay on visually guided behavior. In *Vision, brain and cooperative computation* (ed. M. A. Arbib & A. R. Hanson), pp. 129–163. Cambridge, MA: MIT Press.

Arbib, M. A., Érdi, P. & Szentágothai, J. 1998 *Neural organization: structure, function, and dynamics.* Cambridge, MA: MIT Press.

Arbib, M. A., Billard, A., Iacoboni, M. & Oztop, E. 2000 Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Netw.* **13**, 975–997.

Arkin, R. C. 1989 Neuroscience in motion: the application of schema theory to mobile robotics. In *Visuomotor coordination: amphibians, comparisons, models and robots* (ed. J.-P. Ewert & M. A. Arbib), pp. 649–671. New York: Plenum.

Arkin, R. C. 1998 *Behavior-based robotics.* Cambridge, MA: MIT Press.

Arkin, R. C. 2003 Reactive robotic systems. In *The handbook of brain theory and neural networks* (ed. M. A. Arbib), 2nd edn, pp. 949–951. Cambridge, MA: MIT Press.

Arkin, R., Fujita, M., Takagi, T. & Hasegawa, R. 2003 An ethological and emotional basis for human–robot interaction. *Robot. Auton. Syst.* **42**, 3–4.

Baraduc, P., Guigon, E. & Burnod, Y. 2001 Recoding arm position to learn visuomotor transformations. *Cereb. Cortex* **11**, 906–917.

Beer, R. D. 1990 *Intelligence as adaptive behavior: an experiment in computational neuroethology.* Academic.

Billard, A. 2000 Learning motor skills by imitation: a biologically inspired robotic model. *Cybernetics Syst. J.* **32**, 155–193.

Billard, A. & Hayes, G. 1999 Drama, a connectionist architecture for control and learning in autonomous robots. *Adapt. Behav.* **7**, 35–64.

Bischoff-Grethe, A., Crowley, M. G. & Arbib, M. A. 2003 Movement inhibition and next sensory state prediction in basal ganglia. In *The basal ganglia. VI* (ed. A. M. Graybiel, M. R. Delong & S. T. Kitai), pp. 267–277. New York: Plenum.

Braitenberg, V. 1965 Taxis, kinesis, decussation. *Prog. Brain Res.* **17**, 210–222.

Braitenberg, V. 1984 *Vehicles: experiments in synthetic psychology.* Cambridge, MA: MIT Press.

Braitenberg, V. 2002 In defense of the cerebellum. *Ann. NY Acad. Sci.* **978**, 175–83.

Braitenberg, V. & Onesto, N. 1960 The cerebellar cortex as a timing organ. *Congress Inst. Medicina Cibernetica*, First Naples Atti., pp. 239–255.

Brooks, R. A. 1986 A robust layered control system for a mobile robot. *IEEE J. Robot. Automat.* **2**, 14–23.

Butler, A. B. & Hodos, W. 1996 *Comparative vertebrate neuroanatomy: evolution and adaptation.* Wiley.

Cliff, D. 1992 Neural networks for visual tracking in an artificial fly. In *Towards a Practice of Autonomous Systems: Proc. 1st Euro. Conf. on Artificial Life (ECAL91)* (ed. F. J. Varela & P. Bourgine), pp. 78–87. Cambridge, MA: MIT Press.

Cliff, D. 2003 Neuroethology, computational. In *The handbook of brain theory and neural networks* (ed. M. A. Arbib), 2nd edn, pp. 737–741. Cambridge, MA: MIT Press.

Cobas, A. & Arbib, M. 1992 Prey-catching and predator-avoidance in frog and toad: defining the schemas. *J. Theor. Biol.* **157**, 271–304.

Collett, T. 1982 Do toads plan routes? A study of the detour behavior of *Bufo Viridis. J. Comp. Physiol.* **146**, 261–271.

Corbacho, F. J. & Arbib, M. S. 1995 Learning to detour. *Adapt. Behav.* **4**, 419–468.

Dean, P. & Redgrave, P. 1989 Visual structures and integrated functions. In *Research notes in neural computing* (ed. M. A. Arbib & J.-P. Ewert). Springer.

Dean, P., Redgrave, P. & Westby, G. W. M. 1989 Event or emergency? Two response systems in the mammalian superior colliculus. *Trends Neurosci.* **12**, 138–147.

Didday, R. L. 1976 A model of visuomotor mechanisms in the frog optic tectum. *Math. Biosci.* **30**, 169–180.

Dominey, P. F. & Arbib, M. A. 1992 A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cereb. Cortex* **2**, 153–175.

Dominey, P., Arbib, M. & Joseph, J.-P. 1995 A model of corticostriatal plasticity for learning oculomotor associations and sequences. *J. Cogn. Neurosci.* **7**, 311–336.

Eccles, J. C., Ito, M. & Szentágothai, J. 1967 *The cerebellum as a neuronal machine.* Springer.

Epstein, S. 1979 Vermin users manual. Master's thesis, University of Massachusetts at Amherst, MA, USA.

Ewert, J.-P. 1987 Neuroethology of releasing mechanisms: prey-catching in toads. *Behav. Brain Sci.* **10**, 337–405.

Ewert, J.-P. & von Seelen, W. 1974 Neurobiologie und System-Theorie eines Visuellen Muster-Erkennungsmechanismus bei Kröte. *Kybernetik* **14**, 167–183.

Fagg, A. H. & Arbib, M. A. 1998 Modeling parietal–premotor interactions in primate control of grasping. *Neural Netw.* **11**, 1277–1303.

Gibson, J. J. 1966 *The senses considered as perceptual systems.* London: Allen & Unwin.

Gillies, A. & Arbuthnott, G. 2000 Computational models of the basal ganglia. *Mov. Disord.* **15**, 762–770.

Gnadt, J. W. & Andersen, R. A. 1988 Memory related motor planning activity in posterior parietal cortex of macaque. *Exp. Brain Res.* **70**, 216–220.

Goodale, M. A. & Milner, A. D. 1992 Separate visual pathways for perception and action. *Trends Neurosci.* **15**, 20–25.

Guazzelli, A., Corbacho, F. J., Bota, M. & Arbib, M. A. 1998 Affordances, motivation, and the world graph theory. *Adapt. Behav.* **6**, 435–471.

Guazzelli, A., Bota, B. & Arbib, M. A. 2001 Competitive Hebbian learning and the hippocampal place cell system: modeling the interaction of visual and path integration cues. *Hippocampus* **11**, 216–239.

Gurney, K., Prescott, T. J. & Redgrave, P. 2001 A computational model of action selection in the basal ganglia. I, II. *Biol. Cybern.* **84**, 401–423.

Hikosaka, O. & Wurtz, R. 1983 Visual and oculomotor functions of Monkey Substantia Nigra Pars Reticulata. III. Memory-contingent visual and saccade responses. *J. Neurophysiol.* **49**, 1268–1284.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C. & Rizzolatti, G. 1999 Cortical mechanisms of human imitation. *Science* **286**, 2526–2528.

Ingle, D. 1968 Visual releasers of prey-catching behavior in frogs and toads. *Brain Behav. Evol.* **1**, 500–518.

Ingle, D. 1976 Spatial vision in anurans. In *The amphibian visual system* (ed. K. V. Fite), pp. 119–140. Academic.

Jeannerod, M. & Biguer, B. 1982 Visuomotor mechanisms in reaching within extra-personal space. In *Advances in the analysis of visual behavior* (ed. D. J. Ingle, R. J. W. Mansfield & M. A. Goodale), pp. 387–409. Cambridge, MA: MIT Press.

Jeannerod, M., Arbib, M. A., Rizzolatti, G. & Sakata, H. 1995 Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends Neurosci.* **18**, 314–320.

Johnston, T. A. & Wilkin, P. 1998 *Signs of Australia: a new dictionary of Auslan (the sign language of the australian deaf community)*, 2nd edn. Sydney: North Rocks.

Kuperstein, M. 1988 Neural mode of adaptive hand–eye coordination for single postures. *Science* **239**, 1308–1311.

Lara, R., Carmona, M., Daza, F. & Cruz, A. 1984 A global model of the neural mechanisms responsible for visuomotor coordination in toads. *J. Theor. Biol.* **110**, 587–618.

LeDoux, J. E. 2000 Emotion circuits in the brain. *A. Rev. Neurosci.* **23**, 155–184.

Lettvin, J. Y., Maturana, H., McCulloch, W. S. & Pitts, W. H. 1959 What the frog's eye tells the frog brain. *Proc. Inst. Radio Engrs* **47**, 1940–1951.

Lieblich, I. & Arbib, M. A. 1982 Multiple representations of space underlying behavior. *Behav. Brain Sci.* **5**, 627–659.

Lowe, N., Strauss, J., Yeates, S. & Holden, E.-J. 2002 Auslan jam: a graphical sign language display system. In *Proc. 6th Annual Conf. on Digital Image Computing Techniques and Applications*, pp. 98–103.

Luppino, G., Matelli, M. & Rizzolatti, G. 1990 Cortico-cortical connections of two electrophysiologically identified arm representations in the mesial agranular frontal cortex. *Exp. Brain Res.* **82**, 214–218.

Mays, L. E. & Sparks, D. L. 1980 Dissociation of visual and saccade related responses in superior colliculus neurons. *J. Neurophysiol.* **43**, 207–232.

O'Keefe, J. 1983 Spatial memory within and without the hippocampal system. In *Neurobiology of the hippocampus* (ed. W. Seifert), pp. 375–403. Academic.

O'Keefe, J. & Nadel, L. 1978 *The hippocampus as a cognitive map*. Oxford: Clarendon.

Oztop, E. & Arbib, M. A. 2002 Schema design and implementation of the grasp-related mirror neuron system. *Biol. Cybern.* **87**, 116–140.

Oztop, E., Bradley, N. & Arbib, M. A. 2003 Learning to grasp. I. The Infant Learning to Grasp Model (ILGM). (Submitted.)

Perrett, D. I. 1989 Frameworks of analysis for the neural representation of animate objects and actions. *J. Exp. Biol.* **146**, 87–113.

Perrett, D. I. 1990 Understanding the visual appearance and consequence of hand actions. In *Vision and action: the control of grasping* (ed. M. A. Goodale), pp. 163–180. Norwood, NJ: Ablex.

Prescott, T. J., Gurney, K. & Redgrave, P. 2003 *The handbook of brain theory and neural networks* (ed. M. A. Arbib), 2nd edn, pp. 147–151. Cambridge, MA: MIT Press.

Rizzolatti, G. & Arbib, M. A. 1998 Language within our grasp. *Trends Neurosci.* **21**, 188–194.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G. & Matelli, M. 1988 Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Exp. Brain Res.* **71**, 491–507.

Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. 1995 Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* **3**, 131–141.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Perani, D. & Fazio, F. 1996 Localization of grasp representations in humans by positron emission tomography. 1. Observation versus execution. *Exp. Brain Res.* **111**, 246–252.

Ruppin, E. 2002 Evolutionary autonomous agents: a neuroscience perspective. *Nature Rev. Neurosci.* **3**, 132–141.

Sparks, D. L. & Mays, L. E. 1980 Movement fields of saccade-related burst neurons in the monkey superior colliculus. *Brain Res.* **190**, 39–50.

Taira, M., Mine, S., Georgopoulos, A. P., Murata, A. & Sakata, H. 1990 Parietal cortex neurons of the monkey related to the visual guidance of hand movement. *Exp. Brain Res.* **83**, 29–36.

Umilta, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C. & Rizzolatti, G. 2001 I know what you are doing: a neurophysiological study. *Neuron* **31**, 155–165.

Ungerleider, L. G. & Mishkin, M. 1982 Two cortical visual systems. In *Analysis of visual behavior* (ed. D. J. Ingle, M. A. Goodale & R. J. W. Mansfield). Cambridge, MA: MIT Press.

van Gisbergen, J. A. M. & van Opstal, A. J. 2003 Collicular visuomotor transformations for gaze control. In *The handbook of brain theory and neural networks* (ed. M. A. Arbib), 2nd edn. Cambridge, MA: MIT Press.

Walter, W. G. 1953 *The living brain.* London: Duckworth. (Pelican edn, 1961. Harmondsworth: Penguin Books.)

Webb, B. 2001 Can robots make good models of biological behaviour? *Behav. Brain Sci.* **24**, 1033–1094.