LANGUAGE BEYOND OUR GRASP: WHAT MIRROR NEURONS CAN, AND CANNOT, DO FOR LANGUAGE EVOLUTION

James R Hurford,
Language Evolution and Computation Research Unit,
Linguistics Department, University of Edinburgh

(**Note:** To appear in Kimbrough Oller, D., U. Griebel and K. Plunkett (eds.) (2002) *The Evolution of Communication Systems: A Comparative Approach*. The Vienna Series in Theoretical Biology, MIT Press, Cambridge MA. [in press]. In case of any discrepancy with the eventually printed version, the printed version will be the 'authorized' version.)

> ... before trying to construct scenarios of language origin and evolution based on MNS we must take care to analyse properly the nature of MNS itself. (Stamenov, 2002) [MNS = mirror neurons or mirror neuron system]

And, I would add, we must also take care to analyse properly the nature of language itself. Several recent papers (Rizzolatti and Arbib, 1998; Arbib, 2001; Arbib, 2002) suggest that the discovery of mirror neurons helps us to understand in some more detail how human language evolved. The present paper tries to explore and continue further the development of ideas in these papers(1). Two main issues raised by mirror neurons are addressed, in Sections 1 and 2, respectively.

Section 1 aims to correct a possibly widespread misunderstanding of the significance of mirror neurons, as reflected in the following journalistic passages.

> USC's Michael A. Arbib, Ph.D., says ''the neurons, located in the premotor cortex just in front of the motor cortex, are a mechanism for recognizing the meaning of actions made by others ...
>
> For communication to succeed, both the individual sending a message and the individual receiving it must recognize the significance of the sender's signal. Mirror neurons are thus the missing link in the evolution of language. They provide a mechanism for the sharing of meaning.'' (*ScienceDaily*, 8/20/1998, repeated in USC *Trojan Family* alumni magazine, Spring 1999.)
>
> Rizzolatti and Arbib think that mirror neurons may have provided the bridge from "doing" to "communicating". The relationship between actor and observer may have developed into one involving the sending and receiving of a message. In all communication the sender and receiver have to have a common understanding about what's passing between them. Could mirror neurons explain how this is achieved? Rizzolatti and Arbib think the answer is yes. (*New Scientist*, vol 169 issue 2275, 27/01/2001, p22)

Popular reporters often exaggerate and oversimplify. In academic publications, authors reporting work on mirror neurons are usually more cautious, avoiding direct claims about meaning and communication. But even in the academic literature one can find claims that mirror neurons provide access to meaning. "[A]udiovisual mirror neurons code abstract contents -- the meanings of actions" (Kohler et al. 2002:846). "[T]he mirror neuron system is preprogrammed (via values and biases) to be able to read [i.e. extract the meanings of] these emotional expressions [of loneliness and sadness]" (Wolf et al. 2001:108). Section 1 first outlines areas of agreement with the main mirror neuron researchers on the significance of mirror neurons for the understanding of language. But I then argue that mirror neurons cannot give us any new insight into one of the most crucial features of language, namely the meanings of signs.

Section 2 also seeks to correct a widespread misconception, which attributes very special status to the discovery of mirror neurons, as reflected in the "missing link" remark above, and in such claims, by a distinguished neuroscientist, as the following:

> "I predict that mirror neurons will do for psychology what DNA did for biology," he [V.S.Ramachandran] says. "They will provide a unifying framework and help explain a host of mental abilities that have hitherto remained mysterious." (*New Scientist*, vol 169 issue 2275, 27/01/2001, p22)

Section 2 attempts to locate the concept of mirror neurons within the wider context of behavioural and brain mechanisms, only some of which may be involved in communication. It is argued that mirror neurons are simply a special case of mechanisms that are widespread and well known.

1. Mirror neurons and linguistic signs

Definition: A mirror neuron is a neuron which fires both when performing an action and when observing the same action(2) performed by another (possibly conspecific) creature.

The classic case is that of neurons in a macaque which fire both when the monkey grasps a nut and when it sees a human grasp a nut. The discovery of mirror neurons(3) has important implications for the evolution of language, suggesting pre-existing brain structure which could have provided a basis for human language. I concede for present purposes that mirror neurons can be shown plausibly to provide a basis for linguistically exploitable representations of both sounds (or gestures) and meanings, although a great deal of work is still needed in fleshing out such a claim satisfactorily. But my argument in this section is that mirror neurons cannot, by their very nature, provide a basis for the central, essential structural relation in human language, namely the bidirectional arbitrary mapping between sounds and meanings inherent in the Saussurean sign, as traditionally diagrammed as in Figure 1.

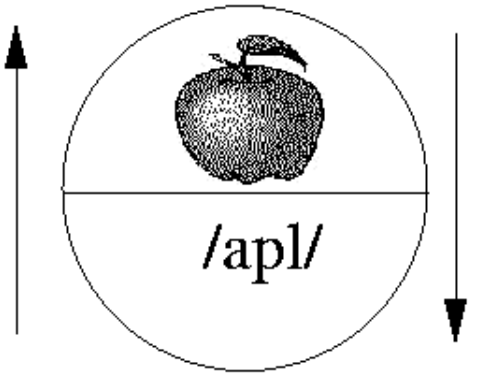

Figure 1: An example of a Saussurean sign, a bidirectional mapping between a concept and a sound image. The upper part represents the concept of an apple, and the lower part the sound image of the word ‘apple’.

The relation between the signified concept and the signifying sound image is arbitrary; there is nothing in the pronunciation of the word that in any way resembles the denoted concept. Both ends of the sign relation are internal mental representations. The ‘meaning’ end of the relation is a mental entity (a concept or a ‘sense’), not a referent object, action or event in the real world. And the ‘sound’ end of the relation is not an actual utterance, or articulatory/acoustic event located in space-time, but a schematic representation of a class of such events.

> Language is naturally homogenous: it is a system of signs in which the sole essential is the union of meaning and sound image, and in which both these parts of the sign are equally psychological in nature. (Saussure, 1916, my translation(4), JRH)

Descriptions in non-neural terms such as ‘meaning’, ‘concept’, ‘sense’, ‘representation’ and ‘sound image’ can be interpreted in neural terms as follows.

1.2 Mirror neurons may explain speech imitation

First, concerning the sound image, the existence of mirror neurons is consistent with two somewhat similar theories of speech perception, the Motor Theory of speech perception (Liberman, 1957; Liberman et al., 1967; Liberman and Mattingley, 1985) and the ‘articulatory filter hypothesis’, proposed by Vihman (1993, 2002).

The motor theory of speech perception, originating long before mirror neurons were discovered, holds that the mental representation of perceived speech is in terms of motor articulatory categories, as opposed to acoustic categories (which might seem more likely, as the input to the ear is acoustic). If perception of grasping involves some neurons which are also involved in the performance of grasping, it lends some plausibility to the idea that perception of a spoken particular sound involves neurons which are also involved in the performance of speaking that sound. In these terms, the ‘sound’ end of the sign relation can be conceived as the intersection of a motor schema and a sensory (auditory) schema. Motor schemata are configurations of neurons that, when activated, produce recognizable specific bodily movements. And a sensory schema is a configuration which, when activated, produces an image of something in the mind. And activation can be half-hearted, as when we just imagine hearing a word, or imagine pronouncing it. So the motor theory of speech perception implies, on this ‘intersection’ view,

mirror neurons in the phonetic/phonological representations of words.

A theory such as the motor theory of speech perception could, if true, solve a problem in language learning. Children are able to imitate the speech sounds they hear; that is, they somehow know how to configure their own vocal tracts so as to produce an auditory impression similar to what they hear, even though the raw information reaching the ear is purely acoustic, and not articulatory. Given the motor theory of speech perception, one can understand how a prelinguistic child can so easily learn to imitate the sound of a word; the acoustic signal is transformed automatically by the ear and brain to a representation at least partly expressed in terms of the articulatory movements required for re-pronouncing the word. This link between mirror neurons and the motor theory of speech perception has been emphasized in the literature discussing the potential of the new discoveries to illuminate language evolution (Gallese et al., 1996; Rizzolatti and Arbib, 1998; Skoyles, 1998). The motor theory has certainly not gained wide acceptance, although neither has it been relegated to the historic waste-bin. Perception always has to start with a discriminatory event that is not motoric in nature, but it is conceivable that automatic motor responses to certain percepts could have evolved.

A theory resembling the Motor Theory of speech perception, though different from it in detail, and possibly overcoming its problems, is the ‘articulatory filter hypothesis’, proposed by Vihman (1993). "On this account,the experience of frequently producing CV [consonant-vowel] syllables sensitizes infants to similar patterns in the input speech stream." (Vihman, 2002) What is common to both theories is the idea that there is some articulatory (i.e. motor) component to children’s representations of speech sounds. Vihman (2002) notes the support given to her hypothesis by the discovery of mirror neurons.

Liberman’s Motor Theory hypothesizes a strong innate component to the perceptuomotor representation of speech sounds, whereas according to Vihman’s articulatory filter hypothesis, the child acquires such representation through experience of its own babbling behaviour. Westermann and Miranda (2002) provide an elegant computer model of the process whereby "mirror neurons responding to both auditory and visual stimuli can develop" (275), based on such feedback from babbling. It is not known whether the mirror neurons in the original experimental monkeys develop in ontogeny influenced by the monkey’s experience of its own grasping gestures, or whether they are epigenetically programmed to develop in any case, regardless of experience. The innate/acquired issue will concern us again in the conclusion of this paper.

1.2 Mirror neurons may aid concept representation

Likewise the ‘sense’, concept, or meaning end of the sign relation is neurally the pattern of activation which constitutes the ‘bringing to mind’ of a particular concept. Here too, in the conceptual domain, there are probably aspects of mirror neuron organization. The central mirror neuron results can most obviously be applied to the mental representations of bodily actions. For instance, if humans are organized in this respect like macaques, the mental representation of the concept GRASP/GRASPING involves some neurons which are involved both in the act of grasping and in the observation of grasping. So thinking of grasping (either by oneself or by someone else) activates these mirror neurons. Similarly, it seems likely that a representation of the concept WALK/WALKING will involve mirror neurons involved both in the observation and the performance of walking. (See the discussion of spontaneous imitative responses in humans in the next section.)

Mirror neurons are, by definition, only involved in the representations of actions, such as grasping and walking. Therefore, adhering to a narrow definition of mirror neuron, we cannot claim that the mental

representations of objects, such as apples and screwdrivers, involve mirror neurons. Apples and screwdrivers are not actions. But it seems likely that representations of objects involve some congruence between motor and sensory neurons, similar to that found in the representations of actions. Attending to or acting on a real apple in an appropriate way, or imagining an apple, involves bringing to mind the concept of an apple. The mental representations of tools involve areas of motor cortex appropriate for handling them, beside sensory information about what the tools look like (Martin et al., 1996). It is hard to dissociate the passive manual feel of an object from active knowledge of what to do with it. Similarly, one's concept of, say, an apple, includes motor information about how to hold it and bite it, as well as sensory information about what it looks/tastes/smells like (Fadiga et al., 2000; Murata et al., 1997). (You don't have to buy this 'mirror' aspect of representations of concepts for the further argument below to go through.)

1.3 Mirror neurons cannot facilitate sign learning

So far, my point has been to agree that the discovery of mirror neurons is a step forward in our understanding of the evolution of human language. I have given a rough characterization in neural terms of the two ends of the Saussurean sign arguments, the sound and the meaning, before even considering the bidirectional relation between them. We might express the Saussurean sign in neural terms as in Figure 2:

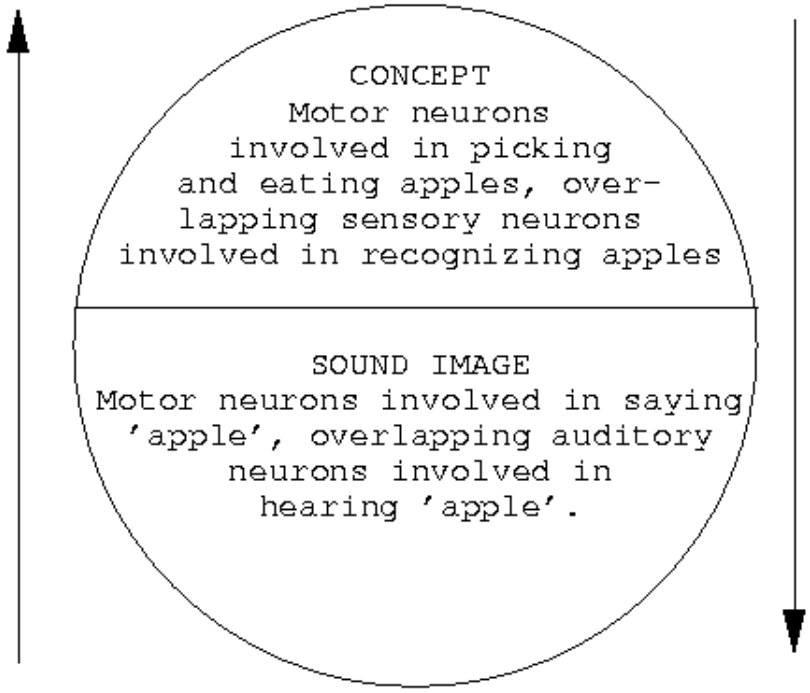

Figure 2: A Saussurean sign represented in neural terms.

But what, in neural terms, might the bidirectional relation between the concept and the sound image be? The well known 'arbitrariness of the sign' implies that in the general case there is no overlap between the neurons involved in the representation of the meaning and those involved in the representation of the sound. The pronunciation of the word 'apple' bears no resemblance whatsoever, in sensory or motor affordances, to apples. The prelinguistic child may well have a fairly solid concept of the category

APPLE, being able to interact in appropriate ways with apples. This concept, of course, involves some neurons. And the child, as argued above, can also represent the sound image of the word ‘apple’, also using neurons, of course. But before the learning of the sound-meaning connection, there is no pre-established overlap between the neurons involved in the concept and the neurons involved in the sound image. This holds generally for all words except for marginal onomatopeic words. So mirror neurons cannot be seen as helping to account for the extreme facility shown by humans in learning the vocabulary of their native language.

There can be no doubt that the extreme facility for learning arbitrary sound-meaning mappings is a specifically human trait. Although trained apes can acquire vocabularies of a few hundred symbols, this is often achieved only with quite laborious training. Even when the learning is somewhat spontaneous, as in Kanzi’s case (Lyn and Savage-Rumbaugh, 2000; Savage-Rumbaugh et al., 1986), the ape’s eventual vocabulary is orders of magnitude smaller than an adult human’s. Adult humans typically have vocabularies in the tens of thousands. The process of a human infant’s acquisition of vocabulary has been labelled ‘fast mapping’ (Carey, 1978; Carey and Bartlett, 1978), reflecting the fact that astonishingly few exposures (sometimes just one!) are needed to learn a word and its meaning. And, as argued above, it cannot be the case that pre-existing mirror neurons facilitate this process.

2. Mirror-neuron-like structures are probably common

This section will argue that a wide range of animal behaviours probably involve arrangements more or less like mirror neurons, depending on how far one is prepared to stretch the term. It will become apparent that a natural definition of ‘mirror neuron’ should be somewhat elastic or fuzzy. It is useful to regard mirror neurons as constituting a fuzzy set, rather than a precisely defined class. There are prototypical, clear central cases of mirror-neuron-like arrangements, and there are cases partially resembling them in relevant ways.

2.1. ‘Automatic’ behaviour possibly reflecting mirror neuron structure

It will be useful to begin by considering behaviours which are involuntary, either automatic, reflex or innate. The focus will be on responses to perceived stimuli which are fast, robust, and hardly subject to suppression or inhibition. Because of the speed of the response to the perceived stimulus and the near-impossibility of inhibition or suppression, these present clear cases where sensory and motor mechanisms are in a tight linkage. An immediate automatic response to a stimulus is, by definition, an action performed when perceiving the stimulus. Action and perception are not absolutely instantaneous; each happens over some brief interval, which we’ll call the perception interval and the motor interval. The onset of the perception interval slightly precedes the onset of the motor interval, but with rapid responses, the two intervals will overlap, and one can look for the possibility of the same neurons being involved in both the perception and the performance of ‘the same action’.

Consider schooling fish and flocking birds. A school of fish appears to act as a single elastic body, with all the member fish swerving uniformly in the same direction, left, right, upwards and downwards. We do not know what specific neurons in the fishes’ brains are involved. Computer modelling of schooling and flocking behaviour (Reynolds, 1987; Toner and Tu, 1998) shows that it is possible to account for it in terms of very simple perception-action responses in the individual animals. Many schooling fish use several sensory modalities to keep a constant spacing between individuals. Both vision and lateral lines down the sides of the fish’s body, sensitive to pressure, are used (Partridge, 1987). Interestingly, the simple basic behavioural principle underlying schooling can be expressed in different ways in English,

one way suggesting that a mirror neuron mechanism is at work, and the other way not suggesting this.

One can say that the schooling fish's basic rule is that sensing a neighbour turn in a given direction automatically triggers the action of turning itself in that same direction. Sensing a left turn triggers a left turn; sensing an upward turn triggers an upward turn, and so forth. Given this constant fast and automatic linkage of perception of action to performance of the same action, it seems almost inescapable that neurons fitting the definition of 'mirror neuron' are involved. Almost certainly, there are neurons involved both in the perception of the neighbours' turn and the immediate turning response.

Alternatively, one can express the facts differently, in a way not suggesting a mirror neuron mechanism. In this version, the basic rule is 'Keep a constant distance from your neighbours'. This rule is implemented by perception of a decrease in the distance to a neighbour triggering a movement away from the neighbour, and by perception of an increase in the distance triggering a move toward the neighbour. It seems wrong to let the issue hinge on the pseudo-question of whether the turning fish acts in response to perception of a turn, or to perception of a change in distance. The two are inseparable; a turn causes a change in distance, and a change in distance implies a turn.

The 'Keep a constant distance' principle is very similar to that used much of the time by a person driving a car in freeway traffic. When the car in front slows down, slow down; when the car in front speeds up, speed up. But would one be tempted to suggest that mirror neurons are involved in this basic aspect of freeway driving? Almost certainly not, for several reasons. The car in front is not an animal of any kind; one may not even be able to see the driver. Although one assumes it is being driven by a creature with a brain, one is not reacting directly to the movements of that other driver, on the gas pedal and the brake pedal. It is not the case that seeing the other driver step on her brake pedal prompts one to step on ones own brake pedal. The response is essentially to the perceived distance to the hard metal shell on wheels in front.

There is, however, a way of seeing this freeway driving behaviour as involving tightly linked perception and performance of 'the same action'. As a thought experiment, consider the car as an outer shell of the driver's body, and thus the whole car+driver ensemble as a single locomotive organism. Its wheels are its limbs; the driver's foot and the pedals are inner working parts of this single organism. The driver's brain is still the brain of the whole car+driver creature. If the driver is well trained, to the point where she regularly brakes involuntarily on perceiving a looming bumper in front, then in terms of the whole car+driver creature, this translates to the rule 'observation of slowing automatically triggers slowing', just as with the schooling fish.

In fact, the imaginative thought experiment is not necessary, if one allows a functional conception of what counts as 'the same action'. If one may count slowing as a recognizable action on the part of a driver, whatever muscles are used to bring it about, and to count observation of a looming bumper in front as observation of slowing, then the statement 'observation of slowing automatically triggers slowing' is a fair description of the driver's behaviour. And to the extent that the slowing behaviour is automatically associated with observation of slowing behaviour, it is likely that some of the neurons involved in the observation are also involved in the action.

Similar to flocking is sudden take-off triggered by observation of sudden take-off by a bird nearby. "A pigeon that signals its intention generally departs without disturbing the others. If a pigeon sees a sign of danger, however, it flies off without giving any intention signals. The other pigeons then immediately take alarm and fly up also." (McFarland, 1987a:13)

In cases where birds and mammals browse the same area in flocks and herds, escape responses in the birds are sometimes triggered by perception of escape behaviour in the mammals, and vice versa. Escape by birds involves taking to the air, escape by the mammals involves running. The events can be described either as 'Perception of escape behaviour triggers escape behaviour' suggesting the possibility of a process which might involve mirror neurons, or as 'Perception of running triggers taking to the air'. Viewed functionally, taking to the air and running off are instances of 'the same action', but viewed purely in terms of what limbs are involved, they are not the same action.

I turn now to another kind of behaviour which prompts questions about whether mirror neurons, or something like them, are involved.

> Some animals, notably the cuttlefish (*Sepia officinalis*) ... are able to alter their coloration to match that of the background. (McFarland, 1987b)
>
> Neural control of the chromatophores enables a cephalopod to change its appearance almost instantaneously, a key feature in some escape behaviours and during agonistic signalling. Equally important, it also enables them to generate the discrete patterns so essential for camouflage or for signalling. ...
>
> ... The chromatophores are controlled by a set of lobes in the brain organized hierarchically. At the highest level, the optic lobes, acting largely on visual information, select specific motor programmes (i.e. body patterns); at the lowest level, motoneurons in the chromatophore lobes execute the programmes, their activity or inactivity producing the patterning seen in the skin. (Messenger, 2001)

Although such colour-changing behaviour is not technically labelled 'imitation' in the animal behaviour literature, it is clearly similar to imitation. Rather than imitating a perceived action, the animal 'imitates' a perceived brightness pattern or texture. Perceiving a stony pattern/texture triggers turning a stony pattern/texture; perceiving a sandy pattern/texture triggers turning a sandy pattern/texture. To the extent that this behaviour is automatic, no doubt some particular neurons are involved both in the pattern perception and in the pattern-changing performance. Should we label these 'mirror neurons'? The question is terminological, not empirical.

" ... caterpillars of some swallowtails (*Papilio spp.*) and cabbage white butterflies (*Pieris brassicae*) change into green pupae when there are many green leaves present, but into brown pupae when the leaves are dead or absent." (McFarland, 1987b) In the case of the caterpillars, it may well not be perception of colour which triggers the colour change. It is just as likely to be something about the comparative smells of green leaves and dead leaves. But if a particular smell is regularly associated with green leaves, it is not unreasonable to say that the caterpillar is, at least indirectly, detecting greenness. If, as seems likely, the neurons responsible for somehow detecting greenness are also involved in the colour change to green, should one call them 'mirror neurons'? Again, the question is terminological, rather than empirical.

The bittern (*Botaurus lentiginosus*) also imitates its surroundings, but in their motion, rather than their colour. A bittern hiding in reeds stretches its neck upwards, lengthening and narrowing its profile, and sways in unison with the reeds as they are moved by the wind (McGowan, 1997; Barrows, 1913)(5). A reasonable description of the bittern's behaviour is 'on seeing leftward swaying (of the reeds), sway left; on seeing rightward swaying, sway right'. The main difference between this and the behavioural rule for

schooling fish is that the bittern is ‘imitating’ not another animal (let alone a conspecific), but a non-sentient organism in its environment.

Another very common form of defense mechanism is freezing, standing absolutely still. Freezing is not necessarily a response to a predator; many animals hold perfectly still for periods between short bursts of activity. An animal freezing acts to match its body to the surroundings by its immobility. A freezing deer in some sense imitates the rocks around it, by not moving. Freezing is a static form of the ‘Keep a constant distance’ behaviour. Despite this ‘imitative’ component to the behaviour, conceptually similar to the colour-changing behaviour of the cuttlefiish, it would be hard to argue that freezing involves mirror neurons, according to the classic definition.

What these examples illustrate is the problematic nature of the phrase ‘the same action’ in the canonical definition of mirror neurons. One problem is with the term ‘same’. It is a matter of judgement whether what the animal observes is ‘the same action’ as what it performs. There are borderline cases, which can be argued either way. If we choose to describe the behaviour of a well-trained freeway driver as ‘Hit the brake pedal when you see the rear bumper of the car in front looming close’, even if perception of the looming bumper and hitting the brake involve some common neurons, these fall outside the definition of ‘mirror neuron’. But if we describe the behaviour as ‘Slow down when you see the car in front slow down’, this seems to be a case involving perception and performance of ‘the same action’, i.e.slowing down.

Another problem is with the term ‘action’, which seems to draw the limits of the class of mirror neurons too narrowly. Intuitively, there is a continuum of related, broadly imitative behaviours stretching from imitation of action (as with fish schooling), through ‘imitation’ of pattern or texture (e.g. by cuttlefish) and ‘imitation of movement of background’ (e.g. by bitterns), to defensive freezing, where the animal ‘imitates the immobility of its surroundings’.

None of these comments are intended to diminish the significance of the experimental work which gave rise to the term ‘mirror neuron’. My suggestion is that mirror neurons occupy one corner of a continuous, extremely diverse, space of possible neuronal arrangements. Neural organization that is mirror-neuron-like to various degrees can be found widely across many species.

Natural selection has shaped schooling by fish and flocking by birds (Partridge, 1987; Krebs, 1987), colour-changing in cuttlefish and some caterpillars and defensive freezing in many species. A fish which turned left when its neighbours (swimming in the same direction) turned right would have become an isolated easy target for a predator (Hamilton, 1971). The selective advantages of camouflaging colour-change and freezing are obvious. Any involuntary behaviour which increases the fitness of an individual is likely to have been naturally selected, with the necessary neurons getting hardwired during the individual’s development. Involuntary imitative behaviours are merely a subcase, and because some of the same neurons are involved in the perception and performance of what can be described as the same action, they can be accorded the special label ‘mirror neuron’.

2.2. Suppressible or learned imitative behaviours

Expressed informally, what happens in the experimental macaques is that on seeing a human grasping a nut, the monkey’s brain takes the first small step toward carrying out a grasping action, but the action is not completed. The action is suppressed, or inhibited. It is the firing of a neuron which normally fires during an action when that action is not being carried out, but merely observed, that attracts so much

scholarly and journalistic attention to mirror neurons. If the action were actually routinely completed, on observation of ‘the same action’, the case would not seem so interesting, and would be classified as a familiar instance of ‘reflex imitative action’, as with schooling fish.

In humans, yawning and laughter are often triggered involuntarily by observation of other people yawning or laughing(6). With some effort of the will, one can resist the temptation to laugh on hearing another person laugh, and one has to be in the right mood for the automatic laugh mechanism to work fully. But there can be no doubt that, in the right circumstances, observing laughter triggers laughter. Entertainment companies boost the perceived funniness of their shows by introducing canned laughter. Yawning on seeing yawning is a weaker, less reliable response, but there is nevertheless an effect. In the cases where the laugh or yawn response is not inhibited, the neurons mediating between stimulus and response conform approximately to the definition of mirror neuron; they fire ‘when’ observing the action and ‘when’ carrying it out. The scare-quotes around ‘when’ here acknowledge the slight delay between the observed laugh or yawn and the evoked laugh or yawn. Any mirror neurons involved presumably fire in the later stages of the observation of the event and in the preparatory stages of the triggered performance. If the response is inhibited, it is an empirical matter whether any such neurons fire. (Here an experimental problem arises, because deliberate (i.e. faked) yawning and laughter are not controlled by the same mechanisms as spontaneous yawning and laughter. But an experiment could be possible in principle, along the following lines. Put a subject alone in a room reading a funny book, and monitor brain activity when the subject laughs spontaneously. The empirical question is how far this activity resembles brain activity on hearing or seeing laughter.)

In fact, a wide variety of actions can trigger spontaneous imitative responses in humans.

> ‘‘Quite often, individuals mirror the behaviors of their conversational partners without ever having conscious intention of doing so (Condon and Ogston, 1967; Kendon, 1970). In an informal group, people may cross their legs at similar angles, hold their arms in similar positions, even simultaneously perform head or hand motions.’’ (Rotondo and Boker, 2002)

Such imitative behaviours are suppressible, and Rotondo and Boker discuss cases of such ‘symmetry breaking’. To the extent that such imitative responses are not suppressed, it is fair bet (though a proposition subject to empirical verification) that mirror neurons are involved. Even when the imitative behaviour is suppressed, there could be some activation of the neurons involved in preparatory stages of such gestures.

Here now is an example involving communication. "[Vervet] monkeys often grunt as they watch another animal, or as they themselves, initiate a group movement across an open plain." (Cheney and Seyfarth, 1990:114). If this description of the vervets’ behaviour is adequate, since this grunt must be initiated by some specific neural activity, this is another clear case of mirror neurons. These grunt-neurons fire both when the animal starts out across open terrain, and when it observes another animal doing so. It can be inferred from the experimenters’ description (using ‘often’) that this behaviour is susceptible to suppression. Other vervet vocalizations are clearly sensitive to differing circumstances, and suppressible.

3. Implications

In this section, I will discuss where we should now take the arguments presented above. Section 3.1 will relate to the earlier Section 1, and to the issue of whether mirror neurons are innate or acquired. Section

3.2, picking up from the earlier Section 2, will ask whether the existence of coordinated perceptual and motor neurons is more significant than the fact that in higher animals many such coordinated schemes are masked by heavy systems of more or less voluntary invocation and inhibition.

3.1 Learned or innate?

The classical experiments revealing mirror neurons were conducted on adult macaques, who had certainly observed and practised grasping many times in their lives. One wants to know whether the same results could be obtained with very young macaques. Is the firing of a mirror neuron on observing grasping an evolved innate response, like the schooling fish's turning responses? Or is it a learned response? An adult macaque could have learned to associate the sight of his own hand grasping a nut with the grasping action, and generalized this association to include sight of a human hand grasping.

M.Arbib (personal communication) reports recent studies by Luciano Fogassi in which some of the same neurons involved in breaking a peanut in half also fire when the monkey hears the sound of a peanut breaking. This suggests a learned response, as the sound of a peanut breaking is very specific and perhaps not likely to have been accurately targetted by natural selection. Again, there is an experiment demanding to be done (suggested to me by M.Arbib). Could a monkey be trained to associate some very different sound, artificially piped into its ears via headphones, with its action of breaking a nut? And would perception of this sound then activate some of the neurons activated in the act of breaking a peanut? (The peanut-breaking case is subtly different from the well-known nut-grasping case. Much of an observed grasping action precedes the actual taking of the nut, whereas the sound of a peanut breaking is simultaneous with, or slightly after, the centre of the action.)

Section 1 above emphasized the arbitrariness of the Saussurean sign, and its consequence that the sound-image and the meaning, or concept, associated with a word, are intrinsically not the same. This lack of sameness is fatal to any straightforward idea that pre-existing mirror neuron structure mediates humans' impressively fast learning of arbitrary symbolic mappings. Some of the imitative behaviours discussed in Section 2 are innate, while others are clearly learned, for example the well-trained freeway driver slowing on observing slowing. 'Unnatural', i.e. non-innate, imitative responses can evidently be instilled by training, just as non-imitative, in fact arbitrary responses can be drummed in. Michael Arbib (personal communication), replying to my challenge about the arbitrary nature of the linguistic sign, made an insightful remark to the effect that the relation between a retinal image caused by observing grasping and a motor neuron firing somewhere else in the cortex is also 'arbitrary'. The macaque's brain is intricately wired up to connect a certain specific pattern of activation of cells of one sort in the retina to an equally specific pattern of activation in an array of cells of a quite different sort in a quite distant part of the brain. The sameness between the observed and the performed action is external to the brain. Consider an analogy. The cities connected by a road are not 'the same cities'. There is nothing intrinsic to each city that in some sense demands that it be connected to the others, as opposed to similar cities across the seas. The connectedness of one city to another by a road is a contingent geographical fact, brought about by the usefulness of a road connection. Neural connections arise, in phylogeny or ontogeny, between intrinsically dissimilar substructures of the brain, for similar functional reasons. Stamenov argues similarly that the 'sameness' of the observed action and the performed action are post-hoc constructions of the experimenter. What happens in the macaque's brain

> ... is due to a resonance-based deictic (here-and-now) attunement of a quite peculiar sort. ...
>
> The appearance of intersubjectivity of MNS, to my mind, is an artefact of the conceptual

> differentiation in its functioning of two separate and different entities --- of ‘observer’ and ‘agent’ --- that are afterwards identified with (or mapped onto) each other. It is their mapping that makes the way MNS functions as if ‘dialogically tuned’ and potentially capable of supporting such high-level cognitive capacities like social learning and intersubjective sharing of experience. (Stamenov, 2002:254)

What humans are amazingly good at is building such arbitrary neural connections. After learning, the association of a word with its meaning is automatic and reflex-like. To an English speaker, utterance of /dog/ brings to mind the concept DOG, and /kæt/ means CAT, and it takes physical injury to disrupt such connections. Facile acquisition of tens of thousands of such arbitrary symbols is not mediated by any pre-existing connection, of a strict mirror-neuron-like sort, between concepts and sound-images.

The firing of the macaque’s grasping mirror neurons may not be innate but acquired by some sort of Hebbian learning, on repeated perception of its own grasping. Then a human child’s acquisition of massive vocabulary could possibly be seen as differing only in degree from this. This over-simple story would attribute a child’s learning of each arbitrary word-meaning mapping to repeated hearing of the word coupled with a clear ostensive indication of its meaning. Very possibly, a few word-meaning mappings are acquired in this way. But it is clear that this cannot be the whole story of vocabulary acquisition. Much vocabulary (e.g. abstract words) cannot be directly tied to any class of percepts.

3.2 Inhibition was the major step

Section 2 above emphasized the widespread occurrence of automatic and semi-voluntary responses to observation of an action by performance of the same action, arguing that mirror-neuron-like organization is common, and often advantageous to the individual. Mirror neuron arrangements are merely a special case, made more interesting by the involvement of the predicate ‘same’, of close neural connections from perception to motor (or premotor) activation. Perception-to-motor linkages are the stuff of animal life. In lower animals, they are largely genetically determined and not subject to inhibition. In higher animals, there is a greater potential for acquired perception-to-motor linkages (e.g. conditioned responses). In the higher animals, too, there is both higher incidence of inhibition of motor responses and greater freedom from immediate stimulus control, as concepts may be ‘brought to mind’ without direct perceptual input.

Using new technology to see something previously hidden is always exciting at first. But reflection may tell us that what we have seen was not, after all, unexpected. The discovery of mirror neurons in fellow primates has been taken to suggest that such linked perceptual and motor mechanisms evolved relatively late and, significantly, probably only in the lineage leading to humans. I have not seen discussion of mirror neurons in non-primate species. Two features mark mirror neurons as discussed in the literature, the perceptuo-motor linkage, and its hiddenness. We may ask, which of these evolved first? We humans clothe our naked bodies; nakedness is the state of nature. The ability to mask the primeval state of nature with clothing came very late in the day.

The data mentioned in Section 2 suggest that imitative perceptuo-motor linkage of all sorts is common in nature. What is probably new in the recent evolutionary lineage of humans is the ability to inhibit or suppress some of the motor aspects of imitative perceptuo-motor linkage. This is consistent with what we know about brain evolution and function. In particular, human brains differ most markedly from chimpanzee and other primate brains in having disproportionately large prefrontal cortex (Deacon, 1997:219). Further, "[t]asks sensitive to prefrontAL damage ... all have to do with using information

about something you've just done or seen against itself, so to speak, to inhibit the tendency to follow up that correlation and instead shift attention and direct action to alternative associations." (Deacon, 1997:263). Most pertinently, Deacon argues that "the ability to overcome the symbol-learning problem can be traced to the expansion of the prefrontal cortical region, and the preeminence of its projections in competition for synapses throughout the brain." (Deacon, 1997:220).

In *Homo sapiens*, the supremely plastic and self-controlling animal, the voracious acquisition of arbitrary symbols certainly involves the creation of neural connections, but these are embedded so deeply in inhibiting systems, and also subject to such complex systems of 'voluntary' evocation, that any behaviourist stimulus-response interpretation of the arbitrary sound-meaning relationship is wholly inappropriate.

4. Last words

I will end as I began, with a quotation from Stamenov, because his conclusions are so parallel to mine, although we arrived at them from quite different directions, and citing different data.

> ... MNS does not perform the work the same way in monkeys and humans (if we assume a causal role of MNS for language origin). In the latter species it can apparently function not only as part of a local brain circuit, but also in an unencapsulated way as a component of the central system supporting the processing of speech and language. If this indeed turns out to be the case after further experimental verification --- that the MNS in humans is a double-action system --- this would entail both good news and bad news.
>
> The bad news would be that one and the same class of neurons functions in different way in two biological species. This means that from studying monkeys' brains we cannot infer for sure how the human brains perform even on the 'low' level of the way classes of neurons function. This is definitely not a good news, as the majority of neurological studies of monkeys and primates are made with an eye that the human brain performs the same way.
>
> The good news would be rather more hypothetical in nature and consequences. It involves the construction of a controversial scenario involving the unencapsulation of the serial component of MNS on an evolutionary scale, and the generalization of its application to the nascent mechanisms of speech and language. (Stamenov,2002:269-270)

There is a long way to go from mirror neurons to language.

References

Arbib, M.A. (2001). The mirror system hypothesis for the language-ready brain. In: *Computational Approaches to the Evolution of Language and Communication* (Cangelosi, A. and Parisi, D., eds.), Berlin: Springer Verlag.

Arbib, M.A. (2002) The mirror system, imitation and the evolution of language. In: *Imitation in Animals and Artifacts* (Nehaniv, C. and Dautenhahn, K. eds.), Cambridge, MA: MIT Press.

Barrows, W.B (1913) Concealing action of the bittern (*Botaurus lentiginosus*). *Ibis*, 30:187-190.

Carey, S. (1978) The child as word-learner. In: *Linguistic theory and psychological reality* (Halle, M., Bresnan, J. and Miller, G.A. eds.), Cambridge, MA: MIT Press.

Carey, S. and Bartlett, E. (1978) Acquiring a single new word. *Papers and Reports on Child Language Development*, 15:17-29.

Cheney, D. and Seyfarth, R. (1990) *How Monkeys See the World: Inside the Mind of Another Species*. Chicago, IL: University of Chicago Press.

Condon, W.S. and Ogston, W.D. (1967) A segmentation of behavior. *Journal of Psychiatric Research*, 5:221-235.

Deacon, T. (1997) *The symbolic species: the co-evolution of language and the human brain*. London: Penguin Press.

Fadiga, L., Fogassi, L., Gallese, V. and Rizzolatti, G. (2000) Visuomotor neurons: ambiguity of the discharge or 'motor' perception? *International Journal of Psychophysiology*, 35(2-3):165-177.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996) Action recognition in the premotor cortex. *Brain*, 119:593-609.

Hamilton, W. (1971) Geometry for the selfish herd. *Journal of Theoretical Biology*, 31(2):295-311.

Kendon, A. (1970) Movement coordination in social interaction: some examples described. *Acta Psychologica*, 32:1-25.

Kohler, E., Keysers, C., Umiltà, M.A., Fogassi, L., Gallese, V. and Rizzolatti, G. (2002) Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297:846-848.

Krebs, J.R. (1987) Flocking in birds. In: *The Oxford Companion to Animal Behaviour* (McFarland, D. ed.), pp 204-208. Oxford: Oxford University Press.

Liberman, A.M. (1957) Some results of research on speech perception. *Journal of the Acoustical Society of America*, 29:117-123.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., and Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychological Review*, 74:431-461.

Liberman, A.M. and Mattingley, I.G. (1985) The motor theory of speech perception revised. *Cognition*, 21:1-36.

Lyn, H. and Savage-Rumbaugh, S. (2000) Observational word learning in two bonobos (*Pan paniscus*): ostensive and non-ostensive contexts. *Language and Communication*, 20(3):255-273.

Martin, A., Wiggs, C.L., Ungerleider,L.G. and Haxby, J.V. (1996) Neural correlates of category-specific knowledge. *Nature*, 379:649-652.

McFarland, D. (1987a) Alarm responses. In: *The Oxford Companion to Animal Behaviour* (McFarland, D. ed.), pp 13-14. Oxford: Oxford University Press.

McFarland, D. (1987b) Camouflage. In: *The Oxford Companion to Animal Behaviour* (McFarland, D. ed.), pp 53-55. Oxford: Oxford University Press.

McGowan, C. (1997) *The Raptor and the Lamb*. London: Penguin Books.

Messenger, J.B. (2001) Cephalopod chromatophores: neurobiology and natural history. *Biological Reviews*, 76(4):473-528.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V. and Rizzolatti, G. (1997) Object representation in the ventral premotor cortex (area f5) of the monkey. *Journal of Neurophysiology*, 78(4):2226-2230.

Partridge, B. (1987) Schooling. In: *The Oxford Companion to Animal Behaviour* (McFarland, D. ed.), pp 490-494. Oxford: Oxford University Press.

Pellegrino, G.D., Fadiga, L., Fogassi, L., Gallese, V. and Rizzolatti, G. (1992) Understanding motor events. *Experimental Brain Research*, 91:176-180.

Reynolds, C.W. (1987) Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, 21(4):25-34.

Rizzolatti, G. and Arbib, M.A. (1998) Language within our grasp. *Trends in Neuroscience*, 21:188-194.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996) Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2):131-141.

Rotondo, J.L. and Boker, S.M. (2002) Behavioral synchronization in human conversational interaction. In: *Mirror neurons and the evolution of brain and language* (Stamenov, M. and Gallese, V. eds.). Amsterdam: John Benjamins. 151-162.

Saussure, F.d. (1916) *Cours de Linguistique Générale*. Paris: Payot.

Saussure, F.d. (1959) *Course in General Linguistics*. New York: The Philosophical Library. Translated by Wade Baskin.

Savage-Rumbaugh, S., McDonald, K., Sevcik, R.A., Hopkins, W.D. and Rubert, E. (1986) Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology: General*, 115 (3):211-235.

Skoyles, J. (1998) Speech phones are a replication code. *Medical Hypotheses*, 50:167--173. Also available via Cogprints at http://cogprints.soton.ac.uk/documents/disk0/00/00/07/82/index.html.

Stamenov, M.L. (2002) Some features that make mirror neurons and human language faculty unique. In: *Mirror Neurons and the Evolution of Brain and Language* (Stamenov, M.L. and Gallese, V. eds.), Amsterdam: John Benjamins. 249-271.

Toner, J. and Tu, Y.H. (1998) Flocks, herds, and schools: A quantitative theory of flocking. *Physical Review*, E 58(4):4828-4858.

Vihman, M.M. (1993). Variable paths to early word production. *Journal of Phonetics*, 21, 61-82.

Vihman, M.M. (2002) The role of mirror neurons in the ontogeny of speech. In: *Mirror neurons and the evolution of brain and language*, (Stamenov, M. and Gallese, V. eds.). Amsterdam: John Benjamins. 305-314.

Westermann, G. and Miranda, E.R. (2002) Integrating perception and production in a naural network model. In: *Connectionist Models of Cognition and Perception* (Bullinaria, J.A. and Lowe, W. eds), London: World Scientific.

Wolf, N.S, Gales, M.E., Shane, E. and Shane, M. (2001) The developmental trajectory from amodal perception to empathy and communication: the role of mirror neurons in this process. *Psychoanalytic Inquiry*, 21/1:94-112.

ENDNOTES

(1) Though he may not be happy for me to mention it, I feel I have gained a better understanding of mirror neurons through discussions with Michael Arbib. Any misunderstandings on my part are not his fault.

(2) In Section 2, we shall see that both the terms ‘same’ and ‘action’ are problematic.

(3) The basic references are Pellegrino et al. (1992), Gallese et al. (1996) and Rizzolatti et al. (1996).

(4) "[L]a langue [...] est de nature homogène: c’est un système de signes où il n’y a d’essentiel que l’union du sens et de l’image acoustique, et où les deux parties du signe sont également psychiques" (Saussure, 1916) What Saussure called the ‘image acoustique’ is rendered in the most popular translation (Saussure, 1959) as the ‘sound image’, a term which I shall use in this paper.

(5) Kenny Smith pointed me to this example.

(6) The relevance of yawning and laughing to mirror neurons is noted by Robert Provine at http://www.edge.org/discourse/mirror_neurons.html .