

driven by selection pressures to sequence, manipulate, and impose hierarchical order on low-cost digital alternations internally as if in a vacuum. But one side of the brain must be anchored in necessity if the other is to experiment with such freedom. One part of the brain must stay alert if the other is to become lost in its own signals. In just the same way, one foot must bear the weight of the dancer's body if the other is to trace fancy patterns in the air, or one hand must grip the slate if the other is to draw marks across its surface. Where the overall context is purely biological, the freely autonomous – normally left-lateralised – activity of imposing structure can certainly still take place. But the resulting movements will not qualify as socially trustworthy signals, being disqualified precisely for appearing so variable and unconstrained.

Even in nature, however, the songs of songbirds and cetaceans show that low-cost autonomous modulations can play a signalling role – on condition that they occur as variables within an otherwise costly, nonarbitrary, and therefore meaningful display. An example will illustrate this point. A weak or frightened animal is likely to be cautious, tentative, and exploratory. It must alternate between action and reaction, coordinating inputs from both hemispheres as it scans the environment for fresh information in advance of each new decision. Normally, for example, it would be risky for a songbird to shut its eyes or block off its ears. Paradoxically, however, for a babbler to “show off” that it can afford to do just that – to sing as if only the song mattered – can be an impressive display of self-confidence. Zahavi and Zahavi (1997) explain this as follows:

Why do babblers use precisely spaced syllables only when they are eager to fight? In order to emit rhythmic, regularly spaced, and clearly defined syllables, one has to concentrate on the act of calling. Any distraction – such as a glance sideways – distorts both the rhythm and the precision of sound; an individual cannot at one and the same time collect information and concentrate on performance. A call composed of precise, rhythmic syllables testifies that the caller is deliberately depriving itself of information, which means either that it is very sure of itself or that it is very motivated to attack, or both. (p. 21)

The Zahavis add that a human being who is in control of a situation likewise tends to issue threats in an ordered, rhythmic sequence, as if celebrating the fact that external reality can be ignored.

To disconnect from reality is to lose touch with the right brain. Less dominant figures cannot afford to do this, which may explain why they tend to rely more heavily on the right hemisphere while speaking (Armstrong & Katz 1983; Ten Houten 1976). Phonological processing is certainly less lateralised in human females than in males (Shaywitz et al. 1995). Lack of dominance makes it vital to stay sensitive to the total environment, drawing on the right hemisphere in order to do so. But autonomous left hemispheric control does not necessarily imply personal dominance. Its fundamental precondition is simply that low-cost signals – whether manual or vocal – need take no account of environmental feedback or resistance. The confident songbird shows off by “deliberately depriving itself of information,” ceding priority to the left hemisphere in the process. When signals need only connect up with one another, free of any requirement to engage with the external environment, it makes sense to encapsulate the computational circuits close together in one cerebral hemisphere while allowing the other to remain in touch with temporarily irrelevant reality.

Following Kobayashi and Kohshima (2001), Corballis notes that humans differ from primates in that human eyes are not inscrutable but enhance cognitive transparency. But this difference is more than an incidental curiosity. Ancestral social networks – even for sexually mature humans – must have been by primate standards anomalously supportive, making it safe to assume that anyone close enough to see the whites of the eyes was likely to be friend, not foe. Direction of gaze is an aspect of ordinary vision. But it may incidentally serve as a signal. A deliberate “wink” can speak volumes at virtually zero cost. Speech may be conceptu-

alised as an extension of the same principle. Where trust is sufficiently high, resistance on the part of listeners disappears, allowing the subtlest of signals to produce effects. Comprehension now involves inserting oneself imaginatively in the signaller's mind (Tomasello 1999). Speech signals do not need to generate their own trust – at the most basic processing level, an assumption of automatic trust is already built in. In fact, on this level it is legitimate to assume a conflict-free – in Chomsky's (1965, p. 3) terms, “completely homogenous” – speech community. So great is the trust, that language works almost as if one component of the brain – or one component of a computing machine – were simply transmitting digital instructions to another (Chomsky 1995; 2002). Quite regardless of whether signs are manual or vocal, it is this bizarre situation which liberates the potential of one hemisphere to arrange complexity independently of the other. We are left with a puzzling intellectual challenge: to elucidate how the necessary levels of trust could ever have been compatible with our selfish genes. Because I believe this to be the key theoretical issue, it will not surprise Corballis that I am critical of his thought-provoking but non-adaptive account, preferring my own more explicitly Darwinian alternative (Knight 1998; 1999; 2000; 2002).

ACKNOWLEDGMENT

Support from the Leverhulme Trust in the associated research is gratefully acknowledged.

Integration of visual and vocal communication: Evidence for Miocene origins

David A. Leavens

Psychology Group, School of Cognitive and Computing Sciences, University of Sussex, Falmer, East Sussex BN1 9QH, United Kingdom.

davidl@central.susx.ac.uk

<http://www.cogs.susx.ac.uk/users/davidl/index.html>

Abstract: Corballis suggests that apes lack voluntary control over their vocal production. However, recent evidence implicates voluntary control of vocalizations in apes, which suggests that intentional control of vocal communication predates the hominid-pongid split. Furthermore, the ease with which apes in captivity manipulate the visual attention of observers implies a common cognitive basis for joint attention in humans and apes.

Corballis suggests that intentionality in communication is exhibited in the visual domain by many primate species (sect. 2.1), but that voluntary control of vocalizations evolved uniquely within our lineage, sometime after the time when gestural language emerged, possibly as late as several hundred thousand years ago. Corballis states that “chimpanzee calls surely have little, if any, of the voluntary control and flexibility of human speech” (sect. 2.1).

Voluntary control over gestural communication by apes is well established (e.g., Leavens 2001; Leavens et al. 1996; Tomasello & Call 1997; Woodruff & Premack 1979), as Corballis notes (sect. 2.1). No researcher can speak to the state of mind of their ape (or human infant) subjects, but operational criteria for intentional communication are relatively standard and uncontroversial in both comparative psychology (e.g., Leavens & Hopkins 1998) and developmental psychology (Bard 1992). Among other criteria, intentional communication requires an audience and is sensitive to changes in the behavioral cues to attention in the audience. With some few exceptions (e.g., Povinelli & Eddy 1996), virtually all experimental and observational studies have confirmed these operational criteria of intentional communication in the gestural production of both free-ranging and captive apes (e.g., Bard 1992; Call & Tomasello 1994; Hostetter et al. 2001; Krause & Fouts 1997; Leavens et al. 1996; Tomasello et al. 1994).

Evidence is growing which is consistent with the interpretation that some voluntary control over vocal production is exhibited by

apes in some circumstances. This evidence derives from regional variations in vocal production, playback experiments in different populations of feral apes, and experimental observations of the co-deployment of gestures and vocalizations by apes in captivity. To briefly elaborate, van Schaik and colleagues (2003) reported regional variations in which wild orangutans (*Pongo pygmaeus*) from Sumatra and Borneo exhibit three vocalizations: kiss-squeak with leaves, kiss-squeak with hands, and "raspberries." Because these vocalizations were exhibited by representatives of only some groups and in fairly constrained contexts, this implies that these particular vocalizations have a learned component. Wilson et al. (2001) reported that the probability of calling by feral male chimpanzees (*Pan troglodytes*) in response to the playback of the pant-hoot calls of an unfamiliar male increased with the number of allied males present, suggesting that chimpanzees can suppress their vocal behavior when it is tactically wise to do so – such as when they may not have a superiority in numbers in the apparent presence of a stranger. Recent studies have also shown that captive chimpanzees deploy their vocalizations seemingly as an attention-getting tactic, vocalizing most when experimenters are less attentive or facing away from the signaler (Hostetter et al. 2001; Leavens et al. 1996; in press). Hence, the data are consistent with the idea that apes can exert voluntary control over their vocal production. Given Corballis's evolutionary assumptions about laterality of function, we might therefore expect to find evidence of functional linkages in the patterns of behavioral asymmetry exhibited by apes. Such evidence has been presented by Hopkins and his associates (cf. Hopkins & Cantero 2003; Hopkins & Leavens 1998; Hopkins & Wesley 2002): Chimpanzees who vocalize while gesturing are more likely to gesture with the right hand than are chimpanzees who do not vocalize while gesturing.

Corballis asserts that "captive chimpanzees can be readily taught by humans to point, and other animals pick up the habit evidently without further human intervention" (sect. 1). We have never consciously trained any of the more than 130 individual chimpanzees we have studied to point or otherwise gesture in the presence of unreachable food. This "spontaneous" development of pointing in captive apes has been noted by others (e.g., Call & Tomasello 1994). That pointing develops so easily in the absence of any explicit training and in populations of apes who have limited interaction with humans is significant insofar as human parents do not consciously train their children to point, yet children begin pointing, typically, by one year of age. We have suggested that in natural habitats, the "problem space," in which one ape is dependent upon another ape to acquire something distant to both interactants, is relatively rare (Leavens et al. 1996). This problem space is encountered on a daily basis not only by apes in captivity, who cannot directly obtain desirable but unreachable food, but also by human children who only slowly develop locomotor independence.

By virtue of the fact that key elements in infants' daily routines involve artifacts that are unreachable by them (e.g., toys, bottles), a problem space exists for year-old human infants in which adult humans must be manipulated to achieve the infants' goals. The relative locomotor autonomy and reduced artifactual dependence of similarly aged apes in the wild (cf. Tomasello 1999) means that they do not encounter, or only very rarely encounter, this problem space. When an object of desire is visible to the cage-bound ape or the relatively immobile human infant, and there is also present an adult human who has delivered similar such objects to the subject, then both ends and means are obviously available. The act of pointing implies that the signaler is aware of the need to draw the visual attention of an observer to the desired entity.

These observations – that some ape vocalizations seem to be either "cultural" (van Schaik et al. 2003) or tactically deployed (e.g., Leavens et al., in press; Wilson et al. 2001), and that apes in captivity spontaneously deploy pointing behavior (Leavens & Hopkins 1998; Leavens et al. 1996) – suggest an earlier evolutionary linkage between vocal and gestural production than that proposed by Corballis. The data are consistent with a claim for continuity

between humans and apes in their problem-solving capacities in these kinds of communicative contexts, which may be fundamental to later acquisition of language in our own lineage (e.g., Baldwin 1995; Butterworth 2001). Parsimony requires that these joint attentional capacities be attributed to the common ancestor of the living great apes and humans, which lived in the middle Miocene, about 12 to 15 million years ago. Because visual and vocal communication seem to be functionally linked in extant apes, language may have been multimodal from its inception.

ACKNOWLEDGMENTS

I thank Kim A. Bard for helpful discussion and William D. Hopkins for his generosity and collegiality. Thanks to Michael Corballis for giving us all something interesting to think about.

**Mouth to hand and back again?
Could language have made those journeys?**

Peter F. MacNeilage

Department of Psychology, University of Texas, Austin, TX 78712-A8000.
macneilage@psy.utexas.edu

Abstract: Corballis argues that language underwent two modality switches – from vocal to manual, then back to vocal. Speech has evolved a frame/content mode of organization whereby consonants and vowels (content) are placed into a syllable structure of frames (MacNeilage 1998). No homologue to this mode is present in sign language, raising doubt as to whether the proposed modality switches could have occurred.

There is an old story about a driver in Maine who was trying to get from one place to another and asked a local for directions. The response was "You can't get there from here." If we reverse the origin and the destination, the Mainiac's problem is my problem with Corballis's assertion that there were two modality switches in the history of language: the first, from vocal to manual language, and the second, back again.

One reason to doubt that either of these transformations occurred at all is that by the time behavior had gone sufficiently up one garden path to be called language, additional selection pressures could not have been strong enough to make us abandon the enterprise in one modality and take it up in the other. We are seriously hampered here in being given virtually no conception of how far up the garden path language had actually gone before we sacrificed one modality for another on each occasion. But I want to take up a more accessible question, the question of *how* these transformations might have been made.

I speak here as one who takes seriously the question of how language transmission modalities actually work. In an earlier paper in this journal, I have argued that modality-specific constraints played a huge role in determining how the mental apparatus underlying modality use in speech (phonology) gets set up in the first place (MacNeilage 1998; see also MacNeilage & Davis 2000a). Corballis for the most part soars above the level of modality constraints. But if, as I suspect, bodily aspects of the transmission modality have a crucial formative role in language phonology, whether spoken or signed, this must have put severe constraints on the freedom to change modalities – in my opinion, *too* severe.

First, let us consider the basic properties of the two transmission modalities, using present-day sign languages as the best available model for the putative early hominid manual language. The manual system consists of two anatomically symmetrical but functionally asymmetrical multijoint limbs arrayed in a signing space centered on the torso and the face. Convention has it that there are four major parameters of sign (Klima & Bellugi 1979): hand shape, hand orientation, location (where in signing space a sign is made), and movement, with some auxiliary functions provided by the face. The vocal system has three subcomponents – respiratory, phonatory, and articulatory – with a directional layering whereby there is res-