



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Language Sciences 26 (2004) 443–466

**Language  
Sciences**[www.elsevier.com/locate/langsci](http://www.elsevier.com/locate/langsci)

# How to do things without words: infants, utterance-activity and distributed cognition

David Spurrett<sup>a,\*</sup>, Stephen J. Cowley<sup>b</sup><sup>a</sup> *Department of Philosophy, University of Natal, Durban 4041, South Africa*<sup>b</sup> *Social Science and Humanities, University of Bradford, Richmond Road, Bradford BD7 1DP, UK*

Accepted 8 September 2003

---

## Abstract

Clark and Chalmers [Analysis 58 (1998) 7] defend the hypothesis of an ‘extended mind’, maintaining that beliefs and other paradigmatic mental states can be implemented outside the central nervous system or body. Aspects of the problem of ‘language acquisition’ are considered in the light of the extended mind hypothesis. Rather than ‘language’ as typically understood, the object of study is something called ‘utterance-activity’, a term of art intended to refer to the full range of kinetic and prosodic features of the on-line behaviour of interacting humans. It is argued that utterance-activity is plausibly regarded as jointly controlled by the embodied activity of interacting people, and that it contributes to the control of their behaviour. By means of specific examples it is suggested that this complex joint control facilitates easier learning of at least some features of language. This in turn suggests a striking form of the extended mind, in which infants’ cognitive powers are augmented by those of the people with whom they interact.

© 2003 Elsevier Ltd. All rights reserved.

**Keywords:** Ape language; Clark; Andy; Deacon; Terrence; Distributed cognition; Language acquisition; Savage-Rumbaugh; Sue; Symbols

---

## 1. Introduction

In ‘The Extended Mind’ Clark and Chalmers (1998) argue for ‘active externalism’—the view that the mind, or what realises it, need not be confined within either

---

\* Corresponding author.

E-mail address: [spurrett@nu.ac.za](mailto:spurrett@nu.ac.za) (D. Spurrett).

the brain, or the body, of the minded individual. We're sympathetic to their position and line of argument. Among the many things outside the brain and body of any particular individual are, of course, other brains and bodies. This paper is a preliminary sketch of what might happen when minds extend into one another. The paper is in two parts—the first establishing some theoretical points of reference, the second being largely descriptive. We note at the outset that what we have written here is speculative and sometimes loose. It is also, hopefully, suggestive of fruitful lines of further reflection and investigation.

Our sub-title refers to 'utterance-activity'. This is a term of art used, here, to refer to the full range of kinetic and prosodic features of the on-line behaviour of interacting humans. Utterance-activity sometimes includes what are usually regarded as words and strings of words, but need not. We regard utterance-activity as at least as good an object of scientific interest in its own right as 'language' traditionally conceived. Further, we regard it as continuous with, and inextricable from, (non-written) language. We combine this continuity thesis with the developmental claim that language, as usually understood, develops out of, or is at least partly an elaboration of aspects of, utterance-activity. This probably sounds at least slightly unorthodox: on a more standard conception, anything deserving the name of (spoken) *language* is a different thing 'in principle' from the rest of behaviour.

One simple argument *for* the standard conception might point out that to do justice to our intuition (if we have one) that written and spoken language are in some fundamental sense the same, we should regard the text-like, or digital, aspects of utterance-activity as language proper, and the remaining twitches, whoops, smiles, wavings and so forth as something else.

Our view, in contrast, is that we get to do things with words (and enable words to do things to us) by means of behaviour in which the wordy and non-wordy are closely integrated, and by going through a developmental period where we do many of the things eventually done with words without them. We maintain that utterance-activity is the arena in which what is standardly regarded as language gets started, and that both the development and ongoing functioning of word-based language are made needlessly mysterious if utterance-activity is sidelined.

We anticipate at least two major objections to our continuity proposal. Briefly, the first points out that powerful and sophisticated models of language treat language as digital, and suggests that the most likely reason these approaches are so powerful is that language is *in fact* digital. If this objection is correct, then what we are doing is urging a retrograde step, where apparently secure results are rendered doubtful. The second objection notes that if utterance-activity includes (as it does) affective display, then it includes signals that are not arbitrary (e.g. Ekman, 1972), whereas we all 'know' that language consists of tokens which *are* conventionally, arbitrarily, connected up to each other and the world. This second objection asserts that we're throwing our net too wide, and running all the risks attendant on ignoring an important partition in the data.

We do not propose to argue directly against either objection, merely suggest how at least one response to each could get started. In the case of the first, note that the power of a theoretical approach is not by itself a compelling argument for the truth

of its assumptions. The success of physical astronomy based on the assumption that planets are point masses does not make it more likely that planets are in fact point masses, or that they truly lack colours or interesting differences in material composition. What it shows is that you can get a lot done by treating them that way.

In the case of the second objection, we note that what counts as arbitrary is a matter of degree, and partly dependent on theoretical perspective.<sup>1</sup> We, now, cannot do much about the association between, say, smiling and feeling good. Plausibly, natural selection could have latched on to some different patterns of facial motion and gone on to build connections between *those* and social and affective states. So smiling could be non-arbitrary to us, but arbitrary from the perspective of one interested in the evolution of patterns of affective signalling in humans. Even supposedly paradigmatic examples of the arbitrary baptism of some referent with a neologism are, of course, constrained by contextual considerations such as what words are already ‘taken’, what phonemes are available to the community in question, what phonetic transitions are easier than others, what the neologism might sound like, etc.

The insistence on viewing language as a formal system of arbitrary elements involves playing up what we call the ‘abstraction-amenable’ aspects of language at the expense of others. One particularly famous instance of this tendency to focus on the abstraction-amenable, or digital, aspects of language is, of course, Turing’s (1950) proposal for an empirical reformulation of the question ‘can machines think?’ Turing regarded it as a virtue of his approach that it had ‘the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man’. We regard it as a competing virtue of our focus on utterance-activity that it demands attending to bodies and environments. By making utterance-activity central, we are not eschewing abstraction and theory.<sup>2</sup> Rather, at least provisionally, we are suspending commitment to the view that there is a theoretically well-motivated gulf separating language ‘proper’ from other aspects of behaviour.

The supposed gulf between language proper and the rest of behaviour finds a suggestive analogue in Clark’s work. Describing *that* gulf will help us get more specific about the kind of extended-mind thesis we are going to sketch.

## 2. A tale of two Clarks

We detect two quite strikingly different registers or moods in Clark (1997). On the one hand there is a line of thinking focused on embodied, and typically mobile,

<sup>1</sup> If some form of determinism is true, then from at least one perspective (i.e. that of the right deterministic theory) *all* relations between signs, other signs and things are no more arbitrary than, for example, the distribution of volcanoes.

<sup>2</sup> In Cowley and Spurrett (2003) we criticise Taylor (in Savage-Rumbaugh et al., 1998) for reacting to what he sees as the failure of traditional linguistics by suggesting that we relax our demands for (scientific) knowledge, partly by means of some Wittgensteinian therapy.

cognition in robots, animals and humans, which emphasises the ways in which traditional expectations concerning the ‘inner’ character of cognition fail to capture the manifest cognitive properties of both living systems and effective engineered ones. On the other hand there are arguments and surveys of evidence centred on the cognitive advantages of language, which also reject the view that cognitive processes are exclusively handled by the brain (a view we call ‘cognitive internalism’) but which focuses on ‘higher level’ functions, paying less attention to embodiment and motion. ‘The extended mind’ is an instance of this line of thinking.

When Clark talks about robots, indeed anything that moves, he emphasises, *inter alia*, the importance of non-neural resources for controlling locomotion and other functions, the greater efficiency and biological plausibility of ‘subsumption architectures’ (Clark, 1997, pp. 13–15; Brooks, 1991) and ‘soft assembly’ (Clark, 1997, p. 42; Thelen and Smith, 1994) as opposed to control systems with fixed hierarchies and/or a central executive. In addition, he combines agnosticism about the necessity of representations with commitment to the view that if there are to be representations they had better pay their way by being directly capable of serving control functions, rather than salvaging outmoded intuitions about the representational nature of thinking (Clark, 1997, pp. 149–153). This is one way of thinking about the ‘extended mind’—an image of brains as parts of embodied coalitions.

When he focuses on language, on the other hand, Clark urges us to relinquish the notion that the primary or only function of language is communication, and instead think of it as an external public and symbolic collection of resources, the exploitation of which grants us a range of cognitive advantages. These cognitive advantages include a capacity for self-stimulation that serves to improve control and performance at tasks (Clark, 1997, p. 202), being able to use symbolic systems to augment memory by using non-neural storage media (Clark, 1997, p. 201), using labels and symbols to simplify our environments and learning processes (Clark, 1997, p. 201; Clark and Thornton, 1997), and simplifying various other types of problem solving. This type of ‘extended mind’ is hooked up to a range of external symbolic resources; language, and language-enabled cognition, is highly *distributed*, but does not seem especially *embodied*.

We are thoroughly sympathetic to both of Clark’s approaches here. We think that he’s on the right track, or two right tracks, and drawing on the right kinds of research. Nonetheless we think that there is an important set of questions which his account of language does not touch on, and which we think need to be part of the type of approach he defends. To see something of what concerns us, consider his discussion of learning with and without labels (Clark, 1993, pp. 69–112; Clark and Thornton, 1997). Whether or not you are surprised that labelling can improve learning efficiency, or open up different types of learning, these results are only possible given a system which operates on labels and data at the same time. With an engineered system which we’ve built ourselves it’s no big deal to add ‘symbolic’ inputs in the form of labels to the inputs already in place for the ‘raw’ data, and adjust the network architecture so that these two streams interact optimally. But with us, with people that is, and some non-humans, there’s a crucial developmental question: “How do we get to be able to make use of ‘symbols’ in the first place?”

Much of the present paper is concerned with this question, which we call the ‘How question’.

This means that for the purposes of what follows, we will for the most part leave Clark’s account of the advantages of language *once you have ‘got’ it*, in place. Another way of saying what interests us, though, is as follows: Clark’s account of language, in common with much linguistic theorising, emphasises the ‘abstraction-amenable’ aspects of language. That is to say that he focuses on labels, signs, symbols and constructions of such elements. But if he is broadly correct about the advantages, then an answer to the question as to how any cognizer can get to count something as a ‘symbol’ at all is needed, and we maintain that part of the answer to that question is to be found by paying closer attention to how talk works *between* people, which is to say drawing on the sorts of ways Clark looks at robots.

### 3. The poverty of the stimulus

A fact indubitably in need of some explanation is that human children typically acquire facility with language within a few years and with little evidence of effort. Debates over the correct explanation are partly organised around a fault line between empiricists defending some version of the view that general learning can account for language acquisition, and nativists insisting that some language-specific innate capacities are essential. Perhaps the most powerful weapon available to the nativists is the poverty of the stimulus argument, which can be glossed as follows:

It is clearly the case that a wide range of sets of organising principles are consistent with the ‘stimulus’ or primary data available to human children, and further that the sub-set of ‘correct’ principles are not preferable by the standards of generic criteria for theory choice, such as simplicity. It consequently seems extraordinarily unlikely that any human child would ever come to behave in ways counted as grammatical for their mother tongue (or tongues) in the event that human children were broadly empiricist learners. Since children do come to be regarded as behaving grammatically with such striking reliability, we can conclude that they are not empiricist learners, but rather that they have language specific innate cognitive endowments.<sup>3</sup>

Debates between empiricists and nativists about language acquisition are not, of course, a series of confrontations between radical ‘tabula rasa’ empiricists and comprehensive nativists who see no role for experience or learning at all. Rather, disagreement concerns, *inter alia*, questions about the real nature of the ‘stimulus’, what mixture of innate and learned capacities are required to explain the phenomena, when particular types of learning start, the extent to which humans and

---

<sup>3</sup> See, e.g. Chomsky (1965, 1967). Laurence and Margolis (2001) is a useful recent and philosophical review of the poverty of the stimulus argument.

particular non-human animals are cognitively alike, and the strengths and limitations of different types of learning.

Although the present paper is not directly concerned with grammar, we may as well stress that we are not Chomskyan nativists. That said, with respect to our ontogenetic concerns we are persuaded that a wide range of innate mechanisms and biases *are* required to explain the available data. Our wariness of Chomsky's brand of nativism is fuelled by two major considerations.

On the one hand, work by such figures as Elman (e.g. Elman, 1991;—see also Clark, 1993) and Christiansen and Chater (in preparation) suggest ways of re-evaluating the properties of the learning involved in coming to behave grammatically. Elman's work seeks to establish what particular connectionist systems are capable of learning, given variations in their architecture, properties of the training data, and the influence of varying general cognitive capacities. An example of this is the role of manipulating the capacity of short-term 'memory' in Elman (1991) which showed that a plausible type of general cognitive maturation could have the same effects as the kinds of 'hyper-benevolent' structuring of training data otherwise required to enable a network to converge on optimal generalisations. Christiansen and Chater, on the other hand, urge a kind of Copernican revolution, in which the vastly greater rate of change of languages as compared to genotypes is a justification for supposing that, to a significant extent, it is languages that are adapted to our cognitive peculiarities and limitations, rather than our cognitive abilities which are specifically and genetically optimised for language.

On the other hand, a range of empirical results concerning the cognitive capacities of non-human animals indicates that many abilities otherwise easily regarded as being language-specific adaptations are found in species without 'language' but with their own versions of utterance-activity. Chinchillas (Kuhl and Miller, 1978) and cotton-top tamarins (Ramus et al., 2000), <sup>4</sup> for example, perform surprisingly well at tasks requiring different (familiar and unfamiliar) language groups to be distinguished from one another—at least as well as human infants of certain ages. <sup>5</sup> To the extent that monkeys can do this, though, it seems reasonable to suppose that the powers of discrimination in question come for 'free' as a consequence of capacities not in any way selected 'for' language.

Equally important, although in different ways, are some of the results from ape language research (ALR), in particular Savage-Rumbaugh's Sherman, Austin and Kanzi (Savage-Rumbaugh, 1986; Savage-Rumbaugh et al., 1998). Kanzi's comprehension is roughly equivalent to that of a two and a half year old human child. His

---

<sup>4</sup> This work tested language discrimination (in this case the ability to distinguish Dutch from Japanese) in both human newborns and cotton-top tamarins. Both types of subject show significant powers of discrimination depending on fairly abstract equivalences rather than simply prosodic features. The authors conclude that 'Since tamarins have not evolved to process speech, we in turn infer that at least some aspects of human speech perception may have built upon pre-existing sensitivities of the primate auditory system'.

<sup>5</sup> The work (see also Nazzi et al., 1998) indicates that rather than distinguishing languages per se, infants distinguish between stress-timed, syllable-timed and mora-timed languages.

production is more difficult to quantify precisely, partly because it is difficult to determine how much it is affected by the physical constraints of the lexigram board system. To be interesting and significant ALR research does not need to produce a non-human ape with levels of fluency comparable to an educated human adult. The point rather is that every increase in performance is a blow against the view that to make *any headway at all* with language requires specifically human biological endowments.<sup>6</sup> For our present purposes what is especially notable about Sherman, Austin and Kanzi is the lexigram board technology used for the research and training, and, in Kanzi's case, an unusual biography and learning history.

First, on lexigram boards, recall that chimpanzees and bonobos have, compared to humans, very limited control over their own vocalisations. Where much other ape language research turned to manual sign-language, Savage-Rumbaugh's team used physical grids of 'lexigram' symbols, both in the form of fixed keyboards which triggered recordings of the relevant spoken term, and as folding boards which could be carried around and used on the move as well as privately by her subjects (who manifestly did engage in self-directed lexigram activity). These external, publicly accessible resources clearly allow some of the memory and other demands of symbolic processing to be handled by non-neural resources, significantly augmenting the cognitive powers of their users (see Cowley and Spurrett, 2003).

Second, and just as importantly, Kanzi's learning biography was unusual. Reared by Matata, a foster mother, he was present during, and apparently uninterested in, her own laborious trials with lexigram boards. Matata managed to show facility with only six different lexigrams, given 30,000 trials over a period of 2 years (Savage-Rumbaugh et al., 1998, p. 17). When she was taken away to be bred at another site, though, Kanzi soon began making use of the lexigram boards to communicate with human laboratory workers, showing, as Savage-Rumbaugh puts it, that he had been 'keeping a secret' (Savage-Rumbaugh et al., 1998, p. 22), concealed by his indifferent progress in prior trials with the boards. On the day before Matata's departure, he used the lexigram board on 21 occasions, asking for 3 different foods. On the following day, he produced 120 lexigram-acts exploiting 12 different symbols (Savage-Rumbaugh et al., 1998, p. 22), twice what Matata had mastered in two years. Savage-Rumbaugh claims that the sudden change suggested that what had changed was not 'his knowledge but [...] his motivation' (Savage-Rumbaugh et al., 1998, p. 22). Consequently ongoing study of Kanzi focussed less on repeated trials, and more on interactions between him and human laboratory workers. An aspect of this shift which we regard as especially important is that in the resulting environment there was a great deal for Kanzi to gain from working out how to manipulate his generally attentive, co-operative, and often indulgent human companions, and to do so with increasing sophistication and precision. Kanzi, then, led a life far closer to that of human infants than most ALR subjects.

---

<sup>6</sup> We note that Savage-Rumbaugh herself accepts the poverty of the stimulus argument and then argues that the genetic similarity between chimpanzees and humans suggests that chimpanzees are likely to have at least some of the same adaptations for language. We prefer the line suggested here, and in Cowley and Spurrett (2003).

Both of the features of Savage-Rumbaugh's research just highlighted (the lexigram boards as part of an extended mind, and Kanzi's own biography) suggest that standard features of debates over the poverty of the stimulus should be re-evaluated. Such debates generally share commitment to the notion that the infant learner is a solitary epistemologist, attempting to make sense of external data on the basis of internal processing, and that it does so with a strikingly scholarly disinterest, or a bare appetite for generalisations. This results in undervaluing or ignoring the ways in which non-neural resources can augment and transform cognitive capacities, and the ways in which social interaction can provide both powerful incentives and mediating structures that support the learning process. If these commitments are conjoined with the tendency, noted above, to focus on the abstraction-amenable aspects of language, the result, we argue, is a grievous misconstrual of the nature of the stimulus and the learning problem, but most strikingly of all, of the nature of the learner.

In the second part of this paper, we present a largely descriptive account of a selection of key episodes—one involving an infant and its mother, one with a child and its father, and one with three interacting adults. We aim, in so doing, to show what it is possible to say about, and identify in, the behaviour of interacting humans when unencumbered either by identification of language with only its abstraction-amenable aspects, or by the view of infants and children as disembodied, or solitary, epistemologists. The re-evaluation of the nature of the learner and of language that this descriptive work suggests, is a further elaboration of the ways in which minds can be extended.

#### 4. The how question

We call the question which we want to put at centre stage the 'how' question: *How can anything come to count as a symbol?*<sup>7</sup> We do not say *be* a symbol because, like e.g. Clark (1993), we are wary of many of the associations carried by the notion of symbols in debates about cognition and language. Any reference to a symbol is too likely, on our view, to suggest some kind of token with fairly precise individuation criteria, determinate intrinsic syntactic properties, and capacities for being more or less literally moved around, operated upon, and combined with other symbols, often in the head. Of course, *whatever* is in (and around) the head, it is undeniable that a great deal of what goes on with people can be *described* in terms of symbols, and structured arrangements of symbols, as well as rules for operating on and with symbols. We want to remain tactically agnostic about what actually goes on under the cognitive hood, so as to try and get a better handle on a particular set of phenomena that we think would be possible without assuming too much about symbols.

Put another way, we do not want to start by buying into a conception of symbols which is too congenial to approaches viewing language largely or com-

<sup>7</sup> A more general form of our question, without the developmental spin of the version in the main text, is: *How do the apparently symbolic aspects of talk relate to wider utterance-activity?*



pletely in terms of its abstraction-amenable aspects. The more one focuses on those aspects, we maintain, the more difficult it is to see how language could possibly get started, or, perhaps, how symbols could be ‘grounded’ (Harnad, 1990).

Recall that utterance-activity embraces both analog (or non-text-like) and non-arbitrary elements. To balance this permissiveness, it is useful to adopt some way of conceptualising how aspects of utterance-activity relate to the how question. For now we’ll use an ‘off the shelf’ solution—the distinctions between iconic, indexical and symbolic reference due to Pierce (1955), especially as appropriated by Deacon (1997). Rather than directly defend the distinctions, we’ll simply take them on board as a taxonomy, leaving aside the empirical question about the extent to which the specified categories are occupied, or the taxonomic analysis is a useful or powerful one.

*Iconic* reference involves some kind of perceived resemblance, perhaps even to the extent of failure to distinguish, between two features of the world. Deacon (1997, p. 75) uses a camouflaged moth as an example, which is only successfully iconic of tree bark to the extent that it is *not* perceptually distinguished from the bark on which it stands. The iconic relationship is, given the range of ways in which two things might be said to resemble one another, a relatively weak one.

*Indexical* reference on the other hand requires some degree of correlation between two re-identifiable types. Again there is a wide range of possible types of correlation, including spatial adjacency and temporal succession. In order for there to be an indexical relationship, a perceiver must be able to identify phenomena as instances of the two types (smoke and fire, say), and note a relationship between them so that, for example, identification of the first can lead to anticipation (or production) of the second.

With *symbolic* reference, the idea is that (to a significant extent conventional) symbols stand in a distributed network of relationships with one another, where the ‘positive’ reference of any symbol is, at least potentially and partly, cashed out in terms of indexically determined equivalence classes. Symbolic reference is, because of the importance of ‘horizontal’ relationships to other symbols, much less hostile to vagaries of correlation than indexical reference, so the boy who cried ‘wolf!’ undermined the indexical value of his utterances, while not changing the symbolic reference of ‘wolf’ (Deacon, 1997, p. 82). Symbolic representation also permits the construction of higher-order types not directly grounded in experience (‘unicorn’) but which do nonetheless partly fix experiential criteria (‘looking like a unicorn’), and others (‘prime number’) which would be impossible, or nearly so, to fix in indexical terms.

Deacon’s view is that symbolic referential relationships are constructed out of indexical ones, which in turn are constructed out of iconic ones, so he envisages a pair of ‘thresholds’ with characteristic cognitive demands and developmental problems in crossing them. For our part we are less confident that the icon, index, symbol taxonomy need be related to cognition and development in such a way, partly because we’re convinced that dispositions to track at least some iconic and indexical relations are ontogenetically innate (see, Cowley et al., in press). That seems to fit

with, for example, the work of Garcia and Koelling (1966) who studied aversion responses to different stimuli in rats. They showed that rats very easily learned to associate (a) a noise and light signal with an electric shock, and (b) a distinctive flavour with (radiation-induced) nausea. In both cases the test populations fairly quickly acquired an avoidance response to the initial signal. Garcia and Koelling also showed that the reversed combinations (light and sound followed by nausea, and distinctive taste followed by a shock) were more difficult for the rats to learn. The innate mechanism suggested here is a bias in favour of connecting nausea with ‘something I ate’ and either no bias at all, or a negative inclination to learn correlations between nausea and flashes and bangs.

According to Deacon (1997, p. 72), the question whether some mark is iconic, indexical, or symbolic, is not about the intrinsic properties of the mark itself, but is a question about the system by which it is actively perceived. So a smile might be a part of some person’s being happy (iconic) or it might be an indicator of happiness (indexical), or even deployed, like Judas’s kiss, as a conventionalised signal (symbolic).<sup>8</sup> While agreeing with Deacon’s general point, we note that the different types of reference each have their own peculiar constraints which, to some extent, make a difference to what can count as a mark. The word ‘hound’ cannot be iconic of dogs, because it cannot be relied upon to be a part of doggy experiences in the same way as hairiness can. Further, wracking sobs are iconic or indexical of misery in ways that conventional labels like “sad” cannot be (Frank, 1988), because we do not generally think anyone can just *decide* to burst into tears, even though we do think that anyone can profess deep sadness.

Note also that on Deacon’s view the distinction between three types of reference implies a distinction between (at least) three degrees of competence (Deacon, 1997, p. 74). A being which could make use of iconic reference to deal with its environment may not be able to manage indexical relations, any more than one that has mastered some indexical relations need be cable of dealing with symbolic ones. The transitions from iconic to indexical, and from indexical to symbolic, are learning problems, with their own distinctive demands. Our primary interest here is in these transitions, and the implied learning problems.

In line with the ‘tale of two Clarks’ above, we note that Clark himself lacks an answer to these questions. This is so even though parts of his work are clearly relevant to these transitions, and highlight aspects of them considered from the perspective of concept formation, and RR learning, that is learning involving ‘representational redescription’ (Clark and Karmiloff-Smith, 1994; Clark, 1993: especially Chapter 4). As we hope to show, though, other parts of his work not specifically concerned with language, but with the demands of robust real-time embodied responsiveness, help us make more headway with approaching the how question.

---

<sup>8</sup> One of us (Cowley, 2002) has critically engaged with aspects of Deacon’s account elsewhere, and accused Deacon of ‘token realism’ about the neural counterparts of apparently symbolic behaviour.

## 5. How to do things without words

Human infants are extraordinarily dependent. They are only able to support their own heads at around three months, cannot reach until around four months, crawl until nine, or walk until thirteen. Unlike other primates, they are unable to cling to their parents in order to be moved around. Almost anything which takes place in accordance with their needs, or, later, their goals, has to be done *for* them. For a being in such a situation there are clearly advantages to be gained from being socially legible—that is from being visibly hungry, distressed, uncomfortable, happy, and so forth, when nourishment, comfort, concerned attention, play, etc., are appropriate. Infants need social relationships in order to survive, and those who take care of infants, typically kin and paradigmatically mothers, need social relationships in order to manage their own energy and resource allocation when caring for the genetic and material investment represented by a child.

The relationships in question are, and have to be, more than simply affiliative. While close mutual interest is undeniably crucial, caregivers have other demands on their attention, especially when an infant has siblings, or is dealing with severe scarcity.<sup>9</sup> And even without siblings, there are times when no matter what a child seems to want, it is more important to make it keep quiet, or wait for some other more urgent goal to be pursued. Infants and caregivers, that is, share an interest in making sense of and to one another, and, although only partly and contingently, share interests in the outcome of their relationship.<sup>10</sup> But they cannot interact in symbolic language, since only one of them is capable of doing so. Symbolic language is an *outcome* of their communication-hungry interaction, rather than a resource available to it from the outset.

Other resources are, though, available. These include facial expressions, direction of gaze, gestures, body-orientation, and prosodic properties of speech, all of which are powerful media of affective signalling. Caregivers are directly affected and motivated by displays of infant affect, especially when the infant is their own offspring (e.g. Wiesenfeld and Klorman, 1978). From birth, or very soon after, infants show interest in faces (e.g. Maurer and Young, 1983), preference for smiling faces (Easterbrook and Barry, 2000)<sup>11</sup> and evidence of facial imitation (e.g. Meltzoff and

<sup>9</sup> There is evidence (see Scheper-Hughes, 1985) that under conditions of severe scarcity a combination of factors relating to the apparent physical health of an infant and its patterns of interaction (including levels of crying) are significant factors in determining levels of care and feeding, possibly determining which offspring will survive. Mann (1992) found that in the absence of serious scarcity, maternal attention tended to focus on the more healthy of two pre-term twins, whether or not the less healthy infant was more responsive, and smiled more.

<sup>10</sup> A parent may have other children to which to allocate resources, or may bet on their chances of success with future offspring, whereas the developing infant has no such options. Haig (1993) documents the ways in which, during pregnancy, the foetus (which has less interest than the mother in her own other and possible future offspring than it does in its own life) can operate more like a parasite than an ally, competing, *inter alia*, over blood supply, and levels of blood sugar. See also Trivers (1974) on some aspects of parent–infant conflict.

<sup>11</sup> This research, with 28 h old infants, showed an appreciable preference for a static and schematic smile over a frown and a bulls-eye figure. The infants showed slightly greater interest in a 6 by 6 checkerboard pattern.

Moore, 1977). By the time of birth they attend to, and prefer, the rhythmic properties of the language they heard most in the muffled world of the womb, and a particular preference for the voice of their mother, which they reliably identify and prefer to other voices following birth (e.g. DeCasper and Fifer, 1980). Some prosodic features of infant-directed utterances have been shown to be indicators of approval, disapproval, etc., in their own way just as universal as facial expressions are indicators of affective state (e.g. Fernald, 1992; Ekman, 1972).<sup>12</sup> Infants across cultures show early preferences for approval vocalisations over ones whose prosodic character is associated with disapproval.

Neither parent nor infant seem, then, to have to *learn* how to get started with affective interaction. In the terms adopted above, we can say that these capacities for affective response make possible a set of innate indexical associations, or serve as the basis for their development. They facilitate the setting up of complex patterns of behavioural co-ordination forming a basis for ongoing development of ever more refined interactive behaviour. By the middle of the second month of life, infants and caregivers begin to engage in interactions often described in terms of mutual ‘delight’, in ways showing evidence of cultural particularity. Trevarthen (1977) refers to such episodes in Britain as manifesting ‘spontaneity, vivacity and delight’, while Bateson (1979) describes interactions in Iran as involving ‘delighted, ritualized courtesy’. We might add that our own data concerning Zulu mothers and infants (see below) includes periods of ‘delighted musical chorusing’. Around the third month interaction between infants and caregivers becomes intensely dialogical, involving the production of protoconversation (Bateson, 1979) and manifesting what Trevarthen (1979, 1998) called intersubjective communication. While caregivers respond to infant behaviour, striking phenomena arise from how they *guide* and *control* the infant’s affectively-based activity. Not only does this involve the development of joint evaluative behaviour but this outcome influences how they motivate and rationalise their own behaviour.

For our purposes an especially important feature of this guiding activity is that it is able to draw on culturally particular expectations concerning appropriate and inappropriate behaviour. What makes this important is that these expectations are, to varying extents, culturally specific, and hence that the particular patterns of expectation have, unlike the responses to smiling, say, to be learned.

It is clear enough that infants occupy what one might call ‘culturally saturated’ environments, in which, for example, the likelihood of an adult allowing an infant’s direction of attention to initiate and fix the focus of interactions, is variable. Other areas of variation include patterns of response to infant distress, where, for example, in some settings attempts to distract the infant by directing its attention to a visible object are more likely, whereas in others attempts to comfort or subdue are common.

<sup>12</sup> Fernald (1992) documents, inter alia, prosodic patterns (found across multiple cultures) indicating approval, prohibition, comforting, and engaging attention. It is important to note one way in which the approach we favour departs from hers. We are interested not only in the ‘internal’ prosodic properties of utterances, but also in *relational* properties discernible in ongoing utterance interactions. Our third example below (‘Oeu!’) is an illustration.

What is not obvious is when infants themselves begin to show evidence of enculturation, that is, of behaviour partly shaped by the patterns of interaction prevalent in their own culturally saturated environment.<sup>13</sup> Our first type of example comes from our own data concerning Zulu infants of between three and four months of age interacting with their mothers, and suggests an answer to this question.

### 5.1. *Thula! (or Shhhhhhh)*

As noted above, there are times when a caregiver will want an infant to fall silent, or in isiZulu to ‘thula’. Zulu children are traditionally expected to be less socially active than contemporary Western children, to initiate fewer interactions, and, crucially, to show a respectful attitude towards adults. An early manifestation of this is in behaviours where a mother attempts to make an infant keep quiet, sometimes saying ‘thula’ (‘quiet’), ‘njega’ (‘no’), while simultaneously gesturing, moving towards or away from the infant, and reacting to details of the infants’ own behaviour (see Cowley et al., in press).

At these times the mother regularly leans forward, so that more of the infant’s visual field is taken up by her face and palms. New vocalisations, and movements or re-orientations of gaze by the infant, are often ‘nipped in the bud’ by dominating vocalisations (sometimes showing prosodic properties indicative of disapproval, comforting, attention and/or arousal towards the mother herself) from the mother, sometimes accompanied by increasingly emphatic hand-waving, and even closer crowding of the infant’s visual field.

While there are distinctive, repeated, elements in many of these episodes, it is important to note that significant portions of the interaction are usually constituted by ‘inter-subjective downtime’ where levels of joint co-ordination are low, and that the interactive ‘game’ being played is characterised by extreme flexibility, manifest in the availability of different routes to a number of acceptable (to the mother) goal-states. There are no simple regularities here where infant distress leads to comforting vocalisations, in turn leading to reduced distress. Rather one sees a rapid alternation of different strategies—comfortings, calls for attention, expressions of disapproval, with, usually, an overall convergence on a parental goal-state in which the infant is quiet. Although it is common to draw on analogies with dancing to describe these interactions, as Stern (1977) noted, boxing also makes an appropriate comparison. Boxers spend a lot of time feinting and otherwise exploring different possible lines of attack, at the same time detecting and closing off their opponent’s explorations. Actual punches thrown, let alone landed, form a small sub-set of a larger number of candidate blows which never make it beyond a slight shifting of weight, or re-orientation of the body.

In spite of this, since our third example below (‘Oeu!’) makes detailed reference to contingent details of interaction on the fly, for the present we focus specifically on the

<sup>13</sup> The contingent patterns need not be cultural: It is well documented that, for example, levels of maternal depression make specific and measurable differences to patterns of affective display and behaviour in infants and children (Lundy et al., 1997).

repeated and strikingly salient aspects of the episodes. With high regularity, and within relatively little time, the particular infant often does ‘thula’, at which point it is generally rewarded with smiling, gentle touching, and other comforting.

At this stage there is no reason to believe that the infant knows what ‘thula’ or ‘njega’ means, or even that it could reliably re-identify the words, let alone produce or contemplate them, so it is extremely unlikely that the word-based aspects of maternal utterance-activity provide labels for the infant. We are considering infants before the stage linguists call ‘babbling’, let alone recognisable speech production. It is not even necessary to suppose that it ‘knows’ that it is supposed to be quiet when behaved at in the ways we have just described. We know that the mother wants the child to be quiet, that this expresses itself in behaviour by the mother, and that the infant comes to be quiet.

If we examine the mother’s behaviour, though, we can make sense of it. She ensures that it is difficult for the infant to attend to anything else by crowding its visual field. She rejects active or new behaviours on its part by cutting off its vocalisations and movements with dominating signals of her own. She largely restricts approval signals, including relaxing the crowding, and reducing the magnitude of her gesturing, as well as expressing comfort through vocalisation, facial signalling and touch, to moments when the infant begins to quieten down. It’s not particularly surprising, then, that it does quieten down.

The mother’s behaviour includes salient, repeated, features which are apt for learning. Her patterns of hand gesturing, for example, could at the outset be iconic of the whole episode including her behaviour and the infant’s becoming quiet, but, when repetition allows the gesture to be individuated and recognised in its own right, go on to become an indexical cue that quietness should follow. The infant’s responses then become indexical *for the mother* of the degree to which the child is co-operative, well-behaved, or, more plainly, ‘good’. Caregiver descriptions of infant behaviour at these times, manifest either in their explicit vocalisations to the child, including references to being ‘good’, or references to possible disciplinary sanctions such as ‘kuza baba manje’ (‘where’s your father now?’) or, in interviews following the video-taping, show that infant behaviour even at this early age is being classified in line with culturally specific expectations of good and bad behaviour. And a crucial part of what makes for a ‘good’ child is responding in ways sensitive to what caregiver behaviour is actually *about*, strikingly in controlling episodes such as the one just described, which make possible the earliest ascriptions of ‘obedience’, ‘co-operativeness’ and so forth.

These ascriptions are over-interpretations. They are, though, necessary over-interpretations, in so far as they motivate caregivers to imbue their own behaviour with regularities manifest regularities in their own behaviour which are then available as structure in the interactional environment for (learning by) the infant. A further episode from our data, in this case concerning a child of around four months, illustrates this point about over-interpretation. In it an infant repeatedly vocalises in ways which to its mother, at least, are suggestive of its saying ‘up’. Each time she says ‘up?’, or ‘you want to go up?’ and after a few repetitions she lifts the child. Prior to the lifting, there is little evidence that the child actually wants to be lifted, or that it

has its attention focussed on anything in particular, except perhaps its own experiments in vocal control. When it is lifted, though, it beams widely. Whatever it did want, if anything, it is now, we suggest, one step closer to figuring out how to behave in ways that lead to its being lifted up.<sup>14</sup>

Still on the subject of lifting, consider the common gesture made around the eighth month by infants who want to be picked up (that is, who subsequently smile or otherwise show approval when they are picked up following such a gesture): a simultaneous raising, or flapping, of both arms (see Lock, 1991). This gesture is not simply copied from common adult behaviours. In the terms we are using here it is partly *iconic*, in virtue of being a common posture of infants while they are in fact being held up, and partly *indexical*, in virtue of being able to stand on its own as an indicator of ‘being up’, as well as being symbolically interpretable as an *invitation* to lift, or a *request* to be lifted. Such gestures are, importantly, serviceable label candidates, in virtue of being amenable to disembedding from behaviour, and eventually coming under deliberate control. An infant need not want to be lifted the first few times it makes such a gesture, it has only to be able to notice that the gesture tends to be followed by liftings.

If and when such learning takes place, it does so in the affectively charged environment we have briefly described. We want to bring discussion of the current example to a close by suggesting a way in which these interactions should be regarded as a further example of how minds can be extended through action. Clark and Chalmers’ suggestion is that paradigmatically mental states and processes can be realised by structures and resources external to the brain. The world beyond the skull of any individual includes, of course, the skulls and brains of others. If active externalism motivates the recognition of a cognitive prosthesis such as a filofax as ‘part’ of what realises a mind, then the embodied brain of another can also play that role. Here, then, is our suggestion: that at times interacting caregiver-infant dyads are neither one individual nor two, but somewhere in between. At the risk of sounding sensational and un-PC at the same time, infant brains can be temporarily colonised by caregivers so as to accelerate learning processes.

If this colonisation does happen, it is made possible by a mixture of affective coupling through interaction, and other mechanisms, such as gaze-following, for co-ordinating attention (see, e.g. Baron-Cohen, 1995 for an attempt to specify the various mechanisms involved). There is ample evidence, some canvassed above, that the affective state of either mother or infant has an immediate impact, especially direct in early life, on the affective state of the other, and that affective state itself generally makes a difference to the ways in which features of the world are observed and remembered (Zajonc, 1980, 1984; Bargh, 1990, 1992),<sup>15</sup> as well as shaping

<sup>14</sup> Papousek (1969) showed that by creating environments in which specific movements by an infant could make things happen in those environments, that the infants smiled when they did ‘work out’ how to exercise control. This suggests that infants are disposed to derive satisfaction from such discoveries.

<sup>15</sup> Zajonc showed that subjects subsequently preferred images which were ‘primed’ with brief (sub-conscious) images of smiles to those primed with frowns. Bargh’s striking research showed, *inter alia*, that subjects exposed to sentences containing words suggestive of age tended to walk more slowly after exposure.

communicative behaviour (e.g. Dimberg et al., 2000; Tartter, 1980).<sup>16</sup> It is not possible directly to ‘install’ some piece of know-how in an infant, but it is possible, some of the time, to direct its attention, modulate its attention and arousal, and regulate various types of reward, to make sure that it is looking in the right direction, at the right time, and in the right way, to pick up on a pattern which is there to be learned. Some of the available patterns are culturally specific indexical relationships which caregivers take as symptomatic of how ‘good’ a particular child is, and which, by structuring caregiver behaviour, open up to the infant a new world of interaction opportunities.

The instances of indexical learning we describe also permit the beginning of a kind of ‘semiotic arms race’ between infants and caregivers. Once an infant has learned, for example, that the arms-up gesture can lead to being lifted, it is possible for ‘requests’ (that is, behaviours *taken as* requests by others, no matter how they are to the infant) to be lifted to be acted on, or to be refused. Prior to the construction and learning of the indexical relationship, this was impossible—a parent would lift a child when the parent wanted to, or thought it would serve some end. Once it has been learned, ‘requests’ can be differentially responded to, depending on their situation in patterns of interaction extending through time. Personal and cultural contingencies about infants and parents will co-determine what patterns are formed, and whether, for example, requested lifting is more likely after relatively quick acquiescence to silencing behaviour, or less likely in the period following failure to attend to objects or events in which a caregiver attempted to arouse interest.

A major shift in the character of this arms race comes with the onset of more deliberate and fine vocal control on the part of the infant, which brings us to our next example.

### 5.2. [nã]/[bø]

Around the tenth month of life a further striking change in infant interaction is noticeable. Where before monadic behaviour gave way to dyadic interaction, the infant now engages the world in a triadic fashion, combining interest in things with joint behaviour with persons. A striking example is given by the linguist Halliday (1975), who describes how at 10½ months his son Nigel came to use his father by means of vocal behaviour.

Nigel produced two distinctive vocal utterances, which Halliday records as [bø] and [nã], and interpreted as, respectively, a request for a favourite toy bird, and a general ‘give me that’ demand. To respond to [nã], in other words, Halliday had to use what was present in the environment to infer what the child was demanding. Indeed, at Nigel’s age, children are likely to be showing early instances of relatively

<sup>16</sup> Dimberg et al. found that observation of, e.g. smiling faces led to neural and muscular activity associated with smiling, even when the images were not consciously perceived. Tartter showed that smiling changes the shape of the human vocal tract, in ways increasing the mean frequency of vocalisations. Vocalisations with high mean frequencies are generally characteristic of approval, making this a fine example of both multiple determination and non-arbitrariness.



fine and ‘deliberate’ vocal control. Even so, as a linguist Halliday may have brought additional (and charitable) interpretive resources to bear on the question whether Nigel, on any two separate occasions, was making the ‘same’ sound again. By doing so he, *perhaps* somewhat more than parents without linguistic training, was lowering the demands on Nigel’s behaviour insofar it could be taken as producing labels which Halliday himself could then go on to take as significant. Although the much younger child taken as ‘asking to be picked up’ in the episode described above undoubtedly had less vocal control than Nigel, Halliday’s criteria for sameness of utterance is similar to that parent’s regarding the successive vocalisations of her child as attempts to say ‘up’. Both cases have in common a movement in the direction of less multi-modal behaviour (one largely gestural, the other largely vocal), and towards producing more effective labels.

In the ‘thula’ case the behaviours we described are likely to be seen as too far from language to count as relevantly related to it. In the present case we need to guard against the opposite tendency, that is to regard Nigel’s various [nā]s and [bø]s as *too much* like mature language. Halliday himself regards the vocalisations as uses of ‘protowords’,<sup>17</sup> and treats them as expressions of relatively well-formed intentions, perhaps even propositional attitudes, to the effect that Nigel wants the bird, or wants some other present object. Thibault (2000) for his part, regards the data as evidence that Nigel has crossed the threshold to indexical reference. We have just seen, though, how infant responses to attempts to quieten them down can be taken by caregivers as indicators of how ‘good’ the child is, and how such ascriptions need not find counterparts in the cognitive world of the infant. Is a similarly deflationary approach possible here?

Clearly it is. Nigel need not initially ‘want’ the bird, any more than the child just described need ‘want’ to be lifted. What is required is that the child be capable of learning the correlation between some aspect of its own behaviour and the regularities produced by attentive adult responses. Nigel could have just gone [bø] at some time when he was shortly after pleased to be presented with the bird toy, and thereafter gone on to learn that [bø]s were reliably followed by bird-givings, and adult utterances of ‘bird’ which partly echoed his own vocalisations. (At the same time Nigel was, of course, acquiring a kind of expertise appropriate to his being in a situation in which 10 month old children get to order parents about at all!) Indexical reference on Nigel’s part can be one *product* of ongoing interaction, scaffolded by Halliday’s production of regularities in the environment, but it need not be the case that Nigel’s initial behaviour be so motivated.

There are, though, important differences between the ‘thula’ case, and that of [nā]/[bø]. Nigel, unlike three month old infants, is capable of behaving in ways which produce highly salient label candidates, not naturally related to affective states in the ways that smiling or crying are, and hence amenable to being conventionally associated with goals, desires and so forth. At his age Nigel also initiates interactions, and,

<sup>17</sup> As is often the case (see Bates and Beghini, 1979), these have imperative uses (e.g. ‘up’, ‘more’). It is of interest that while laboratory trained apes act similarly, even encultured chimpanzees rarely move to ‘declarative’ forms of expression (e.g. ‘dadda’ ‘gone’).

encouraged by caregivers, to engage in active exploration of the world. The regularities in his vocal behaviour, coupled with his greater tendencies to manifest agency, mean that Halliday's (likely) overinterpretations will produce specific opportunities for Nigel, relevant to his level of maturation, and through his exploitation of these opportunities, genuine indexical relationships can come to be established.

### 5.3. *Oeu!*

The discussions of the preceding two examples leave open an interpretation of what we are saying which we wish to dispel. That interpretation would have it that what we are describing is a developmental phase, or perhaps series of phases, during which motor-centric aspects of utterance-activity play an important role because abstraction-amenable ones are relatively underdeveloped, and that once those are properly developed, language 'proper' can get down to business. We maintain, rather, that the full range of aspects of utterance-activity remain in play in all live human interaction.<sup>18</sup> By way of illustration we take a single example from an episode involving several interacting adults.

The episode (for more detail see Cowley, 1998) occurred in Italy, and involved a mother, a father and their adult daughter. In this case, everything begins with Rosa, the mother, evidently seeking sympathy by claiming to Monica, her (adult) daughter, that a 'certain person' had been too lazy to cut some pea-poles she had wanted. This tactic does not succeed in winning Monica's sympathy, and in any event it soon emerges that the husband/father, Aldo, had in fact cut fifteen poles. Rosa changes tack, and instead asserts that the problem is that the pea-poles were unsatisfactory, because they were too long.

Still seeking Monica's sympathy, Rosa now ridicules Aldo by claiming that the pea-poles were 'even longer than this room, if not longer' ('son più lunghe di questa camera se non più'). At this point words fail Aldo, and he uses a response cry (Goffman, 1981) not identifiable with any word, but amenable to being glossed as 'come on, you must be joking', and in the context is clearly legible as an action of gentle mocking. The vocal gesture in this case is a simple vowel ('Oeu') the duration of which can be stretched to that of a short sentence. What is most striking, though, is not the internal prosodic properties of Aldo's 'Oeu' but its relational properties in the context of the interaction, and the shared history of the three people present. To see these features, consider the following figure (Fig. 1).

Notice that Aldo's 'oeu' begins in between Rosa's 'non' and 'più' ('not' and 'longer'), and so follows her assertion that the poles were as long as the room, rather

<sup>18</sup> We would be inclined to argue that this holds, albeit in different ways, in the production and consumption of written texts, even typed ones, as well. Although we do not make this argument here, we draw some inspiration from Dennett's remark: "*Le Penseur's* frown and chin-holding, and the head-scratchings, mutterings, pacings and doodlings that we idiosyncratically favor, could turn out to be not just random by-products of conscious thinking, but functional contributors (or the vestigial traces of earlier, cruder functional contributors) to the laborious disciplining of the brain that has to be accomplished to turn it into a mature mind" (1991: 225).

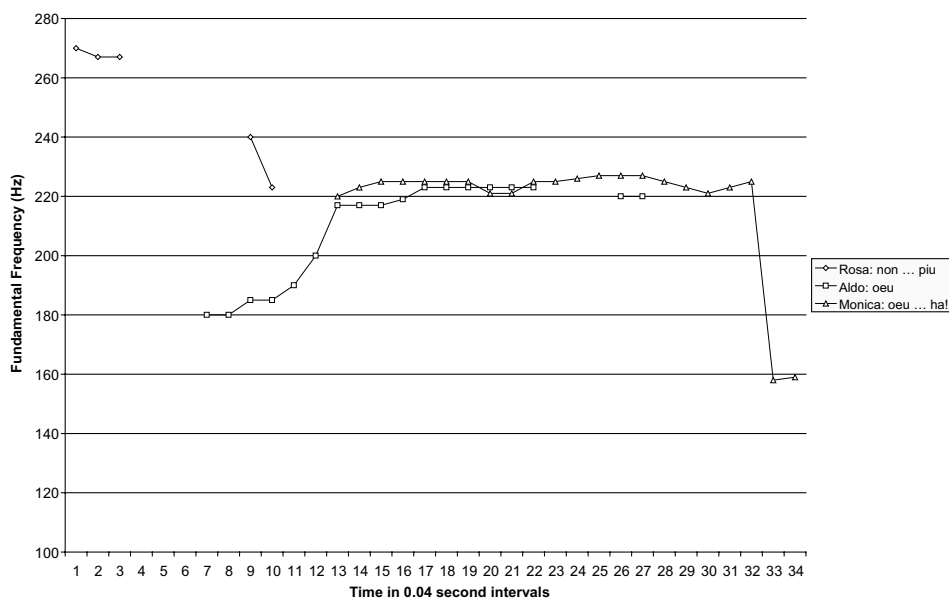


Fig. 1. Oeu!

than waiting for the ‘end’ of her utterance where she adds ‘if not longer’. This violates standard notions of turn-taking while being in keeping with analogies with either dance or boxing. The beginning of Aldo’s vocalisation is at an unusually high pitch for him (about an octave above his usual range), and as he stretches the sound out, he raises his pitch to the same level as the end of Rosa’s ‘più’, indexing her utterance. A little less than half way through Aldo’s ‘oeu’ Monica joins in with an ‘oeu’ of her own, starting with her pitch a little higher than Aldo’s, but joining his in harmony and continuing after he has stopped. Soon after he stops, having run out of breath, Monica drops her pitch to the top of *his* usual range, and gives a short laugh (‘ha!’) at that pitch.

Even without understanding of Italian, the sound recording of this episode makes sense as a brief period during which two people good naturedly mock a third one, and do so *together*. The prosodic details just identified help make sense of why this interpretation is so easy. Aldo and Monica are identifiably ‘together’ because their utterances harmonise, showing a brief allegiance in the same way as bodily orientation shows acceptance or rejection. Their vocalisations are identifiably ‘about’ Rosa’s partly because the pitch on which they converge is indexical of the end of her last utterance, and because Aldo’s unusual starting pitch is also indexical of *her* typical range, rather than his own. Monica’s laugh in turn indexes Aldo, again by being pitched into his normal range. These latter two co-ordinating properties are probably less noticeable to people who do not know the utterers, but are evidence of the ways in which prosodic patterns between people with histories of shared intimacy are modulated by that history, as they can also be by shared cultural experience.

In this case, crucially for our purposes, the gentle mocking which is so clearly accomplished doesn't involve even a single standard 'word'.

Similar forms of indexing can be found by looking beyond pitch, and attending to the ways in which, *inter alia*, accent, timing, and loudness and various kinds of visible movement play out in utterance-activity. Although the 'oeu' example just discussed is very striking, prosodic detail of the same type is all but ubiquitous in utterance-activity at all ages, and occurs in word-based speech as well as in response cries.

## 6. Conclusion

We opened this paper with the assertion that utterance-activity should be regarded as continuous with language, and went on to suggest that approaching our 'how' question from the perspective of distributed cognition would suggest ways of re-evaluating the argument from the poverty of the stimulus. Most of the preceding section is descriptive, rather than argumentative, consisting of an account of how we are inclined to see a number of examples, and in the first two cases, the cognitive and behavioural transitions of which they might be paradigmatic. A question naturally arises, regarding how one sympathetic to our way of describing the episodes might begin to make sense of them.

Here is a somewhat speculative suggestion. In a provocative paper on emotions Ross and Dumouchel (*in press*) argue that emotions should be understood as strategic signals, having the particular effect of encoding preference intensities (which are more difficult to infer than preference orderings) in ways that, unlike standard commitment devices, do not have explicitly to be constructed in advance of strategic interaction. By having preference intensities thus (even if roughly) publicly represented, otherwise intractable strategic problems can be negotiated, and mutually uncongenial prisoners'-dilemma-type situations, sometimes, avoided. Focussing on the first of these possibilities, the idea is that negotiations between agents who are mutually affectively legible involve lower computational demands for each agent's individual strategic decision making. As they say:

On our interpretation of the role of the emotions in bargaining, their status as social conventions enables their expression to be used as early moves in games, ruling out certain outcomes which might otherwise be thought by other parties to be possible equilibria. This can be expected to influence the other party's choice of strategy so long as the structure of the game is such that the other party has a choice at all.

Our suggestion is that a similar function is served by emotional signalling in the epistemic,<sup>19</sup> rather than primarily strategic, interactions between infants and their

---

<sup>19</sup> Evans (2002) is a useful recent attempt to clarify what he calls the search hypothesis of emotion, in which he points out that claims to the effect that emotions solve the 'frame' problem trade on lack of consensus about what that problem actually is, and also notes that we need a positive account of what emotion is, in order to empirically investigate whether emotions really help constrain cognitive searches.

caregivers, and in adult conversation. Our descriptions, unlike many accounts of linguistic and some of strategic phenomena, have not been limited to turn-taking interactions, and instead have emphasised the ways in which roughly simultaneous co-ordination of prosodic and affective display takes place, and how such co-ordinated display can convey significant information about relationships. Such display must convey social information in animals without language, and we contend that it continues to do so in humans. If this speculation isn't obviously wrong, then it suggests two lines of development of the notion of the extended mind.

First, especially considering the 'Oeu!' example, it seems unquestionable that sources of feedback relevant to both Aldo's and Monica's control of their own vocal production, during the period in which they are so strikingly co-ordinated, come from both their own vocal production, and that of the other. More generally, all of the types of affective co-ordination we have described involve integration of inputs from each participant's own behaviour and that of others. This is a striking set of examples of embodied cognition of the sort Clark refers to in the work we have grouped under the 'robots' category. We hope to have shown something of how this type of embodied control could be crucial to the functioning of utterance-activity, and why it merits further empirical investigation.

Second, considering the epistemic pay-offs of the types of embodied co-ordination we have described, it is clear that the model of the solitary infant epistemologist upon which much of the poverty of the stimulus debate is based, is seriously in need of revision. Infants are, in virtue of affective co-ordination, able to function as a kind of cognitive extension of their own caregivers, who focus their attention, regulate their levels of arousal, reinforce and retard patterns in their behaviour, and provide all manner of sources of environmental regularity amenable for infant exploitation. This type of interaction environment permits the construction of socially indexical relationships, and the disembedding of labels and relationships in ways amenable to being recognised as symbolic. The types of embodied co-ordination noted immediately above, that is, permit a particular type of extended mind, in which infants' cognitive powers are augmented by those of the people with whom they interact.

## **Acknowledgements**

Earlier versions of this paper were presented to the mind AND world working group at the University of Natal, Durban in April 2001, and (under the title 'Minded Apes, Talking Infants and the Distribution of Language') at 'The Extended Mind' conference in Hertfordshire, in June 2001. The present version of the paper makes considerably less reference to ape language research than earlier incarnations, although see Cowley and Spurrett (2003). This paper has benefited from comments from and discussions with Andy Clark, Anita Craig, Andrew Dellis, Dan Hutto, Denis McManus, Richard Menary, Mark Rowlands, Fran Saunders, Leslie Stephenson, Susan Stuart, and Michael Wheeler.

## References

- Bargh, J., 1990. Auto-motives: preconscious determinants of social interaction. In: Higgins, T., Sorrentino, R. (Eds.), *Handbook of Motivation and Cognition*. Guilford, New York.
- Bargh, J., 1992. Being unaware of the stimulus vs. unaware of its interpretation: why subliminality per se does matter to social psychology. In: Bornstein, R., Pittman, T. (Eds.), *Perception Without Awareness*. Guilford, New York.
- Baron-Cohen, S., 1995. *Mindblindness*. MIT Press, Cambridge, Mass.
- Bates, E., Begnini, L., 1979. *The Emergence of Symbols: Cognition and Communication in Infancy*. Academic Press, New York.
- Bateson, M.C., 1979. The epigenesis of conversational interaction: a personal account of research development. In: Bullowa, M. (Ed.), *Before Speech: The Beginning of Interpersonal Communication*. Cambridge University Press, Cambridge, pp. 63–77.
- Brooks, R., 1991. Intelligence without reason. In: *Proceedings of the 12th International Joint Conference on Artificial Intelligence*. Morgan Kaufman, Los Altos, CA, pp. 569–595.
- Chomsky, N., 1965. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, Mass.
- Chomsky, N., 1967. Recent contributions to the theory of innate ideas. *Synthese* 17, 2–11.
- Christiansen, M.H., Chater, N. Language as an organism: a connectionist perspective on the acquisition, processing and evolution of language. Oxford University Press, in preparation.
- Clark, A., 1993. *Associative Engines*. MIT Press, Cambridge, Mass.
- Clark, A., 1997. *Being There: Putting Brain, Body and World Together Again*. MIT Press, Cambridge, Massachusetts.
- Clark, A., Chalmers, D., 1998. The extended mind. *Analysis* 58 (1), 7–19.
- Clark, A., Karmiloff-Smith, A., 1994. The cognizer's innards. *Mind and Language* 8 (4), 540–547.
- Clark, A., Thornton, C., 1997. Trading spaces: connectionism and the limits of uninformed learning. *Behavioral and Brain Sciences* 20 (1), 57–67.
- Cowley, S.J., 1998. Of timing, turn-taking and conversations. *Journal of Psycholinguistic Research* 27, 541–571.
- Cowley, S.J., 2002. Why brains matter: an integrational view. *Language Sciences* 24 (1), 73–95.
- Cowley, S.J., Spurrett, D., 2003. Putting apes (body and language) together again. *Language Sciences* 25 (3), 289–318.
- Cowley, S.J., Moodley, S., Fiori-Cowley, A. Grounding signs of culture: primary intersubjectivity in social semiosis. *Mind, Culture and Activity*, in press.
- Deacon, T., 1997. *The Symbolic Species*. Norton, New York.
- DeCasper, A.J., Fifer, W.P., 1980. Of human bonding: newborns prefer their mothers' voices. *Science* 208, 1174–1176.
- Dennett, D., 1991. *Consciousness Explained*. Little, Brown.
- Dimberg, U., Thunberg, M., Elmehed, K., 2000. Unconscious facial reactions to emotional facial expressions. *Psychological Science* 11, 86–89.
- Easterbrook, M.A., Barry, L.A., 2000. Newborns respond differently to smiling and frowning faces. Poster Presentation at the International Society on Infant Studies conference, Brighton, Colorado.
- Ekman, P., 1972. Universals and cultural differences in facial expressions of emotion. In: Cole, J. (Ed.), *Nebraska Symposium on Motivation*. University of Nebraska Press, Lincoln, pp. 207–283.
- Elman, J., 1991. Incremental learning, or the importance of starting small. Technical Report 9101, Center for Research in Language, University of California, San Diego.
- Evans, D., 2002. The search hypothesis of emotion. *British Journal for the Philosophy of Science* 53, 497–509.
- Fernald, A., 1992. Maternal vocalizations to infants as biologically relevant signals: an evolutionary perspective. In: Barkow, J.H., Cosmides, L., Tooby, J. (Eds.), *The Adapted Mind*. Oxford University Press, Oxford, pp. 367–390.
- Frank, R., 1988. *Passions Within Reason*. Norton, New York.
- Garcia, J., Koelling, R.A., 1966. Relation of cue to consequence in avoidance learning. *Psychosomatic Science* 4, 123–124.

- Goffman, E., 1981. *Forms of Talk*. Basil Blackwell, Oxford.
- Haig, D., 1993. Genetic conflicts in human pregnancy. *Quarterly Review of Biology* 68, 495–531.
- Halliday, M.A.K., 1975. *Learning How to Mean: Explorations in the Development of Language*. Elsevier, New York.
- Harnad, S., 1990. The symbol grounding problem. *Physica D* 42, 335–346.
- Kuhl, P.K., Miller, J.D., 1978. Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America* 63, 905–917.
- Laurence, S., Margolis, E., 2001. The poverty of the stimulus argument. *British Journal for the Philosophy of Science* 52, 217–276.
- Lock, A., 1991. The role of social interaction in early language development. In: Krasnegor, N.A., Rumbaugh, D.M., Schiefelbusch, R.L., Studdert-Kennedy, M. (Eds.), *Biological and Behavioral Determinants of Language Development*. Erlbaum, Hillsdale, NJ.
- Lundy, B., Field, T., Pickens, J., 1997. Newborns of mothers with depressive symptoms are less expressive. *Infant Behavior and Development* 19, 419–424.
- Mann, J., 1992. Nurturance or negligence: maternal psychology and behavioral preference among preterm twins. In: Barkow, J.H., Cosmides, L., Tooby, J. (Eds.), *The Adapted Mind*. Oxford University Press, Oxford, pp. 367–390.
- Maurer, D., Young, R., 1983. Newborns' following of natural and distorted arrangements of facial features. *Infant behaviour and development* 6, 127–131.
- Meltzoff, A.N., Moore, M.K., 1977. Imitation of facial and manual gestures by human neonates. *Science* 198, 75–78.
- Nazzi, T., Bertoni, J., Mehler, J., 1998. Language discrimination by new-borns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance* 24, 756–766.
- Papousek, H., 1969. Individual variability in learned responses in human infants. In: Robinson, R.J. (Ed.), *Brain and Early Behavior*. Academic Press, London.
- Pierce, C.S., 1955. *Logic as semiotic: the theory of signs*. In: Buchler, J. (Ed.), *Philosophical Writings of Pierce*. Dover Publications, New York, pp. 98–119.
- Ramus, F., Hauser, M.D., Miller, C., Morris, D., Mehler, J., 2000. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 288, 349–351.
- Ross, D., Dumouchel, P. Emotions as Strategic Signals. *Rationality and Society*, in press.
- Savage-Rumbaugh, S., 1986. *Ape Language*. Columbia University Press, New York.
- Savage-Rumbaugh, S., Shanker, S., Taylor, T.J., 1998. *Apes, Language and the Human mind*. Oxford University Press, Oxford.
- Scheper-Hughes, N., 1985. Culture, scarcity and maternal thinking: maternal detachment and infant survival and a Brazilian Shantytown. *Ethos* 13 (4), 291–317.
- Stern, D., 1977. *The First Relationship*. Fontana, London.
- Tartter, V.C., 1980. Happy talk: perceptual and acoustic effects of smiling on speech. *Perception and Psychophysics* 27, 24–27.
- Thelen, E., Smith, L.B., 1994. *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press, Cambridge, Mass.
- Thibault, P., 2000. The dialogical integration of the brain in social semiosis: Edelman and the case for downward causation. *Mind, Culture, and Activity* 7 (4), 291–311.
- Trevarthen, C., 1977. Descriptive analyses of infant communicative behaviour. In: Schaffer, H.R. (Ed.), *Studies in Mother–Infant Interaction*. Academic Press, London, pp. 227–270.
- Trevarthen, C., 1979. Communication and co-operation in early infancy: a description of primary intersubjectivity. In: Bullowa, M. (Ed.), *Before Speech*. Cambridge University Press, Cambridge, pp. 321–347.
- Trevarthen, C., 1998. The concept and foundations of infant intersubjectivity. In: Bråten, S. (Ed.), *Intersubjective Communication in Early Ontogeny*. Cambridge University Press, Cambridge, pp. 15–46.
- Trivers, R.L., 1974. Parent–offspring conflict. *American Zoologist* 14, 249–264.
- Turing, A.M., 1950. Computing machinery and intelligence. *Mind* 49, 433–460.

- Wiesenfeld, A.R., Klorman, R., 1978. The mother's psychophysiological reactions to contrasting affective expressions by her own and an unfamiliar infant. *Developmental Psychology* 14, 294–304.
- Zajonc, R., 1980. Feeling and thinking: preferences need no inferences. *American Psychologist* 35, 151–175.
- Zajonc, R., 1984. On the primacy of affect. *American Psychologist* 39, 117–123.