


---

13 The Extended Infant: Utterance-Activity and Distributed Cognition

David Spurrett and Stephen Cowley

Introduction

In “The Extended Mind” Clark and Chalmers (1998, reprinted in this volume) argue for “active externalism”—the view that the mind, or what realizes it, need not be confined within either the brain or body of the minded individual. Among the things outside the brain and body of any particular individual are, of course, other bodies and their brains. This paper is a preliminary and self-consciously speculative sketch of what might happen when the systems implementing minds extend into, and get tangled up with, one another. (Such exercises are not unprecedented in the history of philosophy and psychology.) Unlike Clark and Chalmers, our primary focus is not paradigmatically mental states, as understood by philosophers, such as beliefs and desires. Rather, we’re quite generally interested in the control of behavior, whether or not we think of the control systems as operating on beliefs or desires. Our argument here is thus intended as an application of the *parity principle,* Clark and Chalmers’s maxim to the effect (roughly) that any process that we’d have no hesitation in regarding as part of the cognitive process if it went on inside the head, is indeed part of the cognitive process no matter where we end up finding it implemented. Our chapter is in two parts—the first establishes some points of reference regarding language and cognition, the second is both more descriptive and more speculative. Finally, we attempt to connect the issues by discussing the *parity principle* directly.

1 Utterance-Activity and Language

“Utterance-activity” is a term of art (due to Cowley) used here to refer to the full range of kinetic, vocal, and prosodic features of the behavior of
interacting humans. Utterance-activity can include, but is not restricted to, what are usually regarded as words and strings of words. We propose that utterance-activity should be regarded as at least as good an object of scientific interest in its own right as “language” traditionally conceived in mainstream linguistics. Further, we regard it as in crucial ways continuous with, and inextricable from, (nonwritten) language. We combine this continuity thesis with the developmental claim that language as usually understood develops out of, or is to some extent an elaboration of aspects of, utterance-activity. For present purposes these claims can be understood as methodological precepts—claims that useful things might be found out by supposing them—rather than as asserting anything stronger. Even so, this is an unorthodox position because, on a standard conception, anything deserving the name of (spoken) language is different “in principle” from the rest of behavior.

One 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 textlike, or digital, aspects of utterance-activity as language proper, and the remaining twitches, whoops, smiles, wavings, and so forth as something else.

This is not our view. 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 (Cowley, Moodley, and Fiori-Cowley 2004; Cowley in press).

We expect at least two serious objections to our continuity proposal. Briefly, the first points out that powerful and sophisticated models of language treat language as digital and importantly autonomous, and that the most likely reason these approaches are so powerful is that language is in fact digital and autonomous. If this is correct, we are putting apparently secure results needlessly in question. The second objection points out that since we define utterance-activity to include affective display, it includes signals that aren’t arbitrary (see, e.g., Ekman 1972 on universality in facial expression of emotion), whereas we all “know” that language consists of tokens that are conventionally, arbitrarily, connected up to each other and the world. Both objections assert that we are ignoring important partitions in the data.

We won’t offer full arguments against either objection. Both are serious, but we’ll restrict ourselves to offering one suggestion about how to make a start in replying to each.

The first objection derives some of its force from the poverty of the stimulus argument. Some explanation is certainly required for the fact that human children typically acquire facility with verbal language within a few years and with little evidence of effort. Debates over the correct explanation are partly organized 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 we gloss as follows:

It is clearly the case that a wide range of sets of organizing principles are consistent with the “stimulus” or primary data available to human children, and further that the subset 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.

Interesting debates between empiricists and nativists about language acquisition concern, inter alia, questions about the real nature of the “stimulus,” what mixture of innate and learned capacities is required to explain the phenomena, when particular types of learning start, the extent to which humans and particular nonhuman animals are cognitively alike, and the strengths and limitations of different types of learning.

Although the present chapter is not directly concerned with grammar, we note that neither of us is a Chomskian nativist. Our wariness is fueled by two major considerations. On the one hand, work by figures such Elman (e.g., Elman 1991) and Chater and Christiansen (1999) suggest ways of reevaluating 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 manipulating the capacity of short-term “memory” discussed by Elman (1991, discussed in Clark 1993), showing that a plausible type of general cognitive maturation could have some of the same effects as the kinds of “hyper-benevolent” structuring of training data otherwise often required to enable a network to converge on efficient generalizations. (Since inputs in the wild are very unlikely to be
so structured, any approach requiring such benevolence is vulnerable to
the objection that it is unrealistic.) Chater and Christiansen, for their part,
urge a kind of Copernican revolution, in which the vastly greater rate of
change of languages as compared to genotypes is used in claiming that, to
a significant extent, it is languages that are adapted to our cognitive peculiarities and limitations, rather than our cognitive abilities that are specifically and genetically optimized for language.

On the other hand, a range of empirical results concerning the cognitive capacities of nonhuman animals indicates that many abilities we might otherwise regard as language-specific adaptations are found in species without language. Chinchillas (Kuhl and Miller 1978) and cotton-top tamarins (Ramus et al. 2000), 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. 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 specifically selected "for" (digital) language. Ramus et al. (2000, p. 351) 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."

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, Shanker, and Taylor 1998). Kanzi's comprehension is roughly equivalent to that of a two-and-a-half-year-old human child. His skill at production, is more difficult to quantify, partly because it is not clear to what extent it is facilitated or impeded by the constraints of the "lexigram boards." It clearly does not match the level of his comprehension, although Benson et al. (2004) engage in painstaking analysis to suggest that it is more impressive than might have been thought, and that it includes identifiable attempts at vocalization despite inappropriate anatomy. The point we're after here doesn't depend on how much exactly is achieved. It is simply that every increase in nonhuman ape performance is a blow against the view that to make any headway at all with language requires specifically human biological endowments. 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, especially in Kanzi's case, an unusual biography and learning history.

First, regarding the lexigram boards, recall that chimpanzees and bonobos have, compared to humans, very limited control over their own vocalizations. Where much other ape language research used manual sign-languages, Savage-Rumbaugh's team uses physical arrays of "lexigram" symbols, in the form of fixed keyboards that triggered sound recordings of the relevant spoken term, and as folding boards that could be used on the move as well as privately by her subjects (who manifestly did engage in self-directed lexigram activity, and resented others' attempts to get a view of what symbols they were pressing during these episodes). These external, mostly publicly accessible resources allow both memory and other demands of symbolic processing to be handled by nonneural resources, perhaps significantly augmenting the cognitive powers of their users (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 two years (Savage-Rumbaugh, Shanker, and Taylor 1998, p. 17). When she was taken away for a period, though, Kanzi began making use of the lexigram boards to communicate with human laboratory workers, showing abilities 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 three different foods. On the following day, he produced 120 lexigram-acts exploiting 12 different symbols (ibid., p. 22), twice what Matata had mastered in two years. Savage-Rumbaugh et al. claim that the sudden change suggested that what had changed was not "his knowledge but ... his motivation" (ibid.). Consequently, ongoing study of Kanzi focused less on repeated trials, and more on his interactions with human laboratory workers. An aspect of this shift which we regard as especially important is that in the resulting environment Kanzi could gain much from working out how better to manipulate his generally attentive, cooperative, and often downright indulgent human companions. Kanzi, then, lived more like the infants of well-off humans than most ALR subjects.

Both of the features of Savage-Rumbaugh's research just highlighted (the lexigram boards as candidate part of an extended mind, and Kanzi's own biography) suggest that debates over the poverty of the stimulus could stand in need of some reevaluation. Some of these debates often seem to presuppose 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, even a bare appetite for generalizations. This undervalues or ignores both the ways in which non-neural resources can augment and transform cognitive capacities, and those in which social interaction provides both powerful incentives and mediating structures that support the learning process.

In the case of the objection that we needlessly conflate arbitrary and other aspects of activity, we can be briefer. We think that what counts as arbitrary partly depends on where one sets an explanatory frame. We, now, can't do much about the association between smiling and feeling happy and/or friendly. Plausibly, though, natural selection could have latched onto some different patterns of facial motion and gone on to build connections between those and particular social and affective states (see Ross 2007). So the import of smiling could be non-arbitrary to us, but arbitrary from the perspective of one interested in the evolution of patterns of affective signaling in humans.

Insisting on viewing language as a formal system of arbitrary elements involves emphasizing its "abstraction-amenable" aspects at the expense of others. One particularly famous instance of this tendency to focus on the abstraction-amenable aspects of language is, of course, Turing's (1950) proposal for an empirical reformulation of the question "can machines think?" Turing said 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 compelling virtue of a focus on utterance-activity, in the spirit of much contemporary research in cognitive science, that it attends to bodies and their lives in environments. By making utterance-activity central, we are not eschewing abstraction and theory at all. Rather we are suspending, for the purposes of investigation, commitment to the view that there is a theoretically well-motivated gulf separating language "proper" from behavior. How good an idea this is depends on what the research produces.

2 The "How" Question

We call the question we want to put at center stage the "how" question: How can anything come to count as a symbol? We don't say be a symbol because, like Clark (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 on, and combined with other symbols, often in the head. Of course, whatever is in (and around) the head, what goes on with people can usually 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 focus more directly on a particular set of phenomena.

Although utterance-activity embraces both analog (or non-textlike) and non-arbitrary elements, we are interested in its symbol-like aspects. In order to say something about symbols, for present purposes, we use an "off-the-shelf" solution. Accordingly, we use the distinctions between iconic, indexical, and symbolic reference due to Peirce (1955). Rather than directly defend these distinctions, recently appropriated by Deacon (1997), we simply take them on board as a taxonomy. By so doing, we leave aside philosophical questions about whether all of the specified categories are occupied, or whether the taxonomic analysis can be defended independently of how it is used here.

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, where the camouflage is only successfully iconic of tree bark to the extent that it is not perceptually distinguished from the bark on which the moth rests. The iconic relationship is, given the range of ways in which two things might be said to resemble one another and the variety of capacities of different perceiving systems, a relatively weak one.

Indexical reference on the other hand requires some degree of correlation between two reidentifiable types. Again there is a wide range of possible types of correlation, including spatial adjacency and temporal succession (that latter especially important for learning about outcomes of actions). 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 (mostly but not necessarily conventional) symbols stand in a distributed network of relationships with one another, where the "positive" reference of any symbol is partly understood in terms of indexically determined equivalence classes. Symbolic reference is, because of the importance of "horizontal" relationships to other symbols, much less hostage to vagaries of correlation than indexical
reference. For this reason, the boy who cried “wolf!” undermined the utterance’s indexical force without 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 [would]”), and others (“imaginary 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. We are less confident that the “icon, index, symbol” taxonomy need be aligned with a developmental timetable, partly because we’re convinced that dispositions to track at least some iconic and indexical relations are innate.

It is important to bear in mind (Deacon 1997, p. 72) that the question whether some mark is iconic, indexical, or symbolic depends not on the intrinsic properties of the mark, but on the system by which it is consumed. So a kiss might be a part of some person’s being affectionate (iconic) or an indicator of affiliation (indexical), or deployed as a conventionalized signal against the grain of “normal” indexical use (as in Judas’s signal at the Last Supper to the Romans). While agreeing with Deacon’s general point, we note that each type of reference has its own peculiar constraints which, to an extent, make a difference to what can count as a mark. The word “hound” cannot be iconic of dogs as we find them, because it cannot be relied on to be a part of dog-related experiences in the way that hairiness and pointed teeth can. Further, wracking sobs are iconic or indexical of misery in ways that conventional labels like “sad” can’t be (Frank 1988, but see also Ross and Dumouchel 2004), because we don’t 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). An entity that can make use of iconic reference to deal with its environment (e.g., by climbing some kinds of concentration gradients) may not be able to learn from indexical relations around it, any more than one that masters some indexical relations need be capable 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.

We note that Clark himself has relatively little to say about such transitions. This is so even though parts of his work are clearly relevant to them, and highlight aspects of them considered from the perspective of concept formation and “representational redescription” (Clark and Karmiloff-Smith 1994; Clark 1993, esp. chap. 4). As we hope to show, though, other parts of his work concerned not specifically with language, but with the demands of robust real-time embodied responsiveness, help us make more headway with approaching the “how” question.

3 How to Do Things without Words

Human infants are extraordinarily dependent on others. 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 young primates, they are unable to cling to their parents in order to be moved around; even though they have the clinging reflexes, their parents aren’t hairy enough. 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 manifestly hungry, distressed, uncomfortable, happy, and so forth, when nourishment, comfort, concerned attention, play, and so on, are required and possibly available. Equally, they have much to gain if they are able to use caregiver preferences in regulating events. Not only do infants need social relationships; those who take care of them—typically kin and paradigmatically mothers—also need relationships to manage their own energy and resource allocations.

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 under conditions of severe scarcity. Even without competing 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. 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 signaling. 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), and evidence of imitation of facial expression and gesture (e.g., Meltzoff and Moore 1977). By the time of birth they attend to, and prefer, rhythmic properties like those they heard in the muffled world of the womb, and show a 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 are indicators of approval, disapproval, and so on, in their own way just as universal as facial expressions are indicators of affective state (e.g., Ekman 1972). Infants across cultures show early preferences for approval vocalizations over ones whose prosodic character is associated with disapproval. Neither parent nor infant need, then, learn how to get started with affective interaction. In the terms adopted above, we can say that these capacities for affective response form and make possible a set of innate indexical associations. Fernald (1992) documents, inter alia, prosodic patterns (found across multiple cultures) indicating approval, prohibition, comfort-giving, and engaging attention. Our approach departs from hers, as in the final example ("Oeuf"), since we give attention to not only an utterance's "internal" prosodic properties but also to the relational properties that link them (see Cowley 1998, in press).

By the middle of the second month of life, infants begin to engage with their caregivers in interactions in ways showing both some universality, and also evidence, as we shortly explain, 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) include periods of "delighted musical choru­sing" (Cowley 2003). 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) calls intersubjective communication. At this stage culturally or individually specific expectations about appropriate infant behavior begin to play a more serious role in shaping the interactions. This poses a variety of learning problems for both, especially as caregivers attempt to develop ways of guiding and controlling infants' exploiting but transcending the innate indexical relations they start with.

While there are distinctive, repeated elements in many of these episodes, it is important to note that significant portions of the interaction consist in "intersubjective downtime" where levels of joint coordination are low (including for reasons of fatigue), and that the interactive "game" being played is characterized 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 invariably leads to comforting vocalizations, in turn leading to reduced distress. Rather, one often sees a rapid shifting of strategies—for example, offers of feeding, comfortings, calls for attention, expressions of disapproval, with, usually, an overall convergence on a parental goal state in which the infant is content and quiet. As Stern (1977) noted, boxing provides an appropriate comparison for mother-infant interaction. 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, or landed, are rare compared to candidate blows that never develop beyond a slight shifting of weight and gaze, or reorientation of the body.

It is also 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 an interaction is variable. Further areas of variation include patterns of response to infant distress, where in some settings attempts to distract the infant by drawing attention to a salient visible object are likely, but in others attempts to comfort or subdue by holding or rocking occur more often. What is not at all clear to us yet is when infants themselves begin to show evidence of enculturation—of behavior partly shaped by the contingent patterns of interaction prevalent in their own culturally saturated environment. (The contingent patterns need not all be cultural. Levels of maternal depression, for example, make measurable differences to patterns of affective display and behavior in their infants and children [Lundy, Field, and Pickens 1997].) Our example comes from our own observational data concerning Zulu infants of between three and four months of age interacting with their mothers, and suggests that infants may show signs of enculturation (inherited but not genetically encoded and not species-universal patterns of behavior) surprisingly early.

**Thula! (or Shhhhh)**

As noted above, there are times when a caregiver will want an infant to stop being distressed or agitated and fall silent, or in isiZulu to "thula." Zulu children, especially in rural settings, are traditionally expected to be
less socially active than most contemporary European or North American children, to initiate fewer interactions, and, crucially, to show a respectful attitude toward adults. An early manifestation of this is in behaviors where a mother attempts to make an infant keep quiet, sometimes saying “thula” (“quiet”), “njega” (“no”), while simultaneously gesturing, moving toward or away from the infant, and reacting to details of the infants’ own behavior (see Cowley, Moodley, and Fiori-Cowley 2004).

During these episodes the mother regularly leans forward, so that her face and palms take up more of the infant's visual field. When this happens, new vocalizations and movements or reorientations of gaze by the infant are often “nipped in the bud” by dominating vocalizations (sometimes showing prosodic properties indicative of disapproval or comforting) from the mother, sometimes accompanied by increasingly emphatic hand-waving and even closer crowding of the infant’s visual field.

At this age, before the onset of what linguists call “babbling,” let alone recognizable speech production, there is no reason to think that the infant knows what “thula” or “njega” means, or even that it could reliably reidentify the words (let alone produce or contemplate them). It is extremely unlikely, then, that the word-based aspects of utterance-activity provide labels for the infant. It is not even necessary to suppose that the infant “knows” that it is supposed to be quiet when it behaves as we have described. We know that the mother wants the child to be quiet, and not only because she says so, expresses her wish in her behavior, and confirms this in her response to the infant’s falling silent (or failing to do so).

Even without knowing what “thula” means, one can make sense of the mother’s behavior. She makes it difficult for the infant to attend to distractions by moving forward and filling more of its visual field. She rejects repeated or new undesirable behaviors on its part by cutting off its vocalizations and movements with dominating signals that are timed to coincide with infant behaviors. She largely restricts her own approval signals, including relaxing the crowding, and reducing the magnitude of her gesturing, as well as expressing comfort through vocalization, facial signaling (including smiling), and touch, to moments when the infant begins to quiet down. She continues them when it remains quiet, rewarding signs of calm and happiness. It’s not surprising, then, that it often does quiet down.

The mother’s behavior includes salient (sometimes exaggerated), repeated features, including but not limited to innate indexical relations that are apt for reidentification and then learning. The patterns of hand gesturing in some of our examples, could at the outset be iconic of the whole episode, including her behavior and the infant’s becoming quiet. When repetition allows the gesture to be individuated and recognized in its own right, it could go on to become an indexical cue that quietness should follow, that whatever the infant just did was worth doing again because it led to reinforcement. The infant’s responses can then become usefully indexical for the mother of the degree to which the child is cooperative, well behaved, or, more plainly, “good.” Caregiver descriptions of infant behavior at these times, manifest either in their explicit vocalizations to the child, including references to “good” behavior and disciplinary sanctions like “kuza baba manje” (“where’s your father now?”). Further interviews following the videotaping show that infant behavior even at this age is classified in line with culturally specific expectations of good and bad behavior. Part of being “good” here is sometimes doing what you’re supposed to, before learning any words, and early ascriptions of “obedience,” “cooperativeness,” and so forth are often based in episodes of attempted control.

Such parental ascriptions of “goodness” and so forth are almost certainly overinterpretations, if taken to identify explicit infant policies. They are, though, helpful and perhaps necessary overinterpretations, insofar as they motivate caregivers to behave with regularity and thus structure the interaction in ways that become familiar to the infant. A further episode from our data, in this case concerning a child of around four months, illustrates this point about overinterpretation. In it an infant repeatedly vocalizes in ways which to its mother, although not to us, are suggestive of its saying “up.” After several responses along the lines of “up?” or “you want to go up?” the mother lifts the child. Prior to this, to the detached observer, there is little evidence that the child actually wants to be lifted, or that its attention is focused on anything at all. 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.

Still on the subject of lifting, consider the gesture often 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 copied from common adult behaviors (who do not generally flap their arms prior to lifting anything, and for practical purposes are never in a position to raise their arms to enable something considerably larger than they are to get a grip on their torsos). 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.16 Such gestures are, importantly, serviceable label candidates, in virtue of being amenable to disembedding from the behavior they might have been part of, and eventually coming under deliberate control. An infant need not want to be lifted the first few times it makes such a gesture; even though getting into a "being lifted" pose could be part of wanting to be lifted, it has only to be able to notice that the gesture tends to be followed by liftings and learn from this.

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's suggestion is that paradigmatically mental states and processes can be realized by structures and resources external to the brain. The world beyond the skull of any individual includes, as we noted, the skulls and brains of others. If active externalism motivates the recognition of a cognitive prostheses such as a filofax as "part" of what realizes a mind, we don't see why (although see the discussion of the parity principle below) the embodied brain of another couldn't also play that role. Here, then, is our suggestion: at times interacting caregiver-infant dyads are neither one individual nor two, but somewhere in between. At the risk of sounding sensational and un-politically correct at the same time, infant minds can be intermittently colonized by caregivers so as to accelerate learning processes.

If this colonization does happen, it is made possible by a mixture of affective coupling through interaction and mechanisms such as gaze-following that are used in coordinating attention (see, e.g., Baron-Cohen 1995 for an attempt to specify the various mechanisms involved). There is ample evidence, some canvass above, that the affective state of either mother or infant has an immediate impact, and especially direct in very 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), as well as shaping communicative behavior (e.g., Dimberg, Thunberg, and Elmedh 2000; Tartter 1980). It is not possible directly to "install" some piece of know-how in an infant, but it is sometimes possible to direct its attention, to modulate its attention and arousal. Equally, caregivers act to ensure 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 behavior, open the infant to a new world of social opportunities.

The instances of indexical learning we describe also permit the beginning of a kind of signaling arms race (to the extent that their interests sometimes conflict) 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" (i.e., behaviors 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" will be differentially responded to, depending on their situation in patterns of interaction that extend through time. Personal and cultural contingencies about infants and parents will codetermine what patterns are formed, and whether, for example, requested lifting is more likely after relatively quick acquiescence to silencing behavior, 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, at about the same time as infants begin to engage the world in a triadic fashion, combining interest in things with joint behavior with persons. A striking example is given by the linguist Halliday (1975), who describes how at ten months his son Nigel came to exploit his father by means of vocal behavior.

Nigel produced two distinctive vocal utterances, which Halliday records as [be] and [nã], and interpreted as, respectively, a request for a favorite toy bird, and a general "give me that" demand. To respond cooperatively to [nã], then, Halliday had to use pragmatic clues to infer what Nigel wanted. 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. Although the younger child taken as "asking to be picked up" in the episode described above undoubtedly had less vocal control than Nigel, both cases have in common a movement toward more modality-focused behavior (in one case, gestural; in the other, syllabic). This, we think, is crucial in coming to produce behavior that can be interpreted around disembedded labels.

In the "thula" case the behaviors 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 [nəs] and [bəs] as *too much* like mature language. Halliday himself regards the vocalizations as uses of "protowords," and treats them as expressions of relatively well-formed intentions, perhaps even propositional attitudes. Thibault (2000) uses the data as evidence that Nigel has crossed the threshold to indexical reference. On a more deflationary view Nigel need not *initially* "want" the bird, any more than the child described above need "want" to be lifted. A child does need to be capable of learning the correlation between some aspect of its behavior and the regularities produced by attentive adult responses. So Nigel's protowords can be a product of ongoing interaction that is scaffolded by Halliday's production of regularities in the environment. Given their relationship and familiarity with each other's affective signaling, we lack reason to think that what Nigel does is motivated any such "desire."

Oeu!
The discussion of the childhood examples above is consistent with an interpretation of what we are saying that we wish to dispel. On that view what we are describing is a developmental phase during which motor-centric aspects of utterance-activity play an important role because abstraction-amenable ones are relatively underdeveloped. Rather, we maintain that the full range of aspects of utterance-activity remain in play in all live human interaction. By way of illustration we take a single example involving 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 gives a cry not identifiable with any word, but amenable to being glossed as "come on!" in the context clearly legible as an act of good-natured 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 figure 13.1.

Notice that Aldo's "oeu" begins in between Rosa's "non" and "più" ("not" and "longer"), following her assertion that the poles were as long as the room, rather 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 vocalization 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, perhaps having run out of breath, Monica drops her pitch uncharacteristically low, and 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

![Graph](image-url)
help make sense of why this interpretation is so easy. Aldo and Monica are identifiably "together" because their utterances harmonize, showing a brief allegiance in the same way as bodily orientation shows acceptance or rejection. Their vocalizations 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. While these latter two coordinating properties may be below any noticeable threshold, they provide good evidence of how 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 that is accomplished doesn't involve a single standard "word." While the literature on prosodic and visible expression describes many incidents that resemble this one, our focus on relational functions is novel. As with boxing, we stress that inter alia, accent, timing, and loudness and various kinds of visible movement that allow adults to regulate one another's speaking. While the "oeu" example just discussed is striking, prosodic detail of the same type is ubiquitous in utterance-activity at all ages, and occurs in word-based speech as well as cries like Aldo's (see Cowley 1998, in press).

4 The Parity Principle

In the opening paragraph of this chapter we suggested that what we were arguing was an application of Clark and Chalmers's parity principle. Consider the specific criteria that Clark and Chalmers offer for something to count as part of an individual's cognitive system:

1. That the resource be reliably available and typically invoked.
2. That any information thus retrieved be more or less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory.
3. That information contained in the resource should be easily accessible as and when required (Clark, this volume, p. 46).

Scrutiny of the above might suggest that processes we have pinpointed during some kinds of coordinated behavior fail, in some way, on all three. As Clark (this volume) notes with reference to the three criteria, other people "typically would not (but could in rare cases)" meet them. What sort of problem is this for our proposal?

With reference to the second and third criteria, we think that their formulation in terms of "information" that is "retrieved" or "contained" merits revision. Those sorts of images are relatively appropriate where the external resource is taking on cognitive functions related to memory, but less so for other capacities. Our examples have more in common with what Kirsh and Maglio (1994) report of how Tetris is played. Human players of the game are not best thought of as storing and retrieving information. Though there is a sense in which this is what they get the computer to do, they must also trust the machine to perform the rotations and translations they initiate, and not to cheat by modifying the contour at the bottom of the playing area, and so on. Similarly, the various components between keyboard and screen that handle rotations and translations must be "easily accessible as and when required," but, rather than storing information, they perform various manipulations. The "as and when required" part of the third condition and the "reliable availability" requirement of the first must, therefore, be interpreted carefully. It is hardly an objection to the view that experienced Tetris players use external resources to do cognitive work that suitable computers aren't there whenever one might want to play. Similarly, the fact that interaction partners are not always present counts for little if, when they are there, interesting cognitive possibilities arise. (Why, after all, should the boundaries of extended minds be relatively stable, when what the minds do and the resources available are so variable?)

Although a tweaked set of criteria fleshing out the parity principle might allow the sorts of case we're suggesting to count as examples of extended minds, we can imagine other difficulties. Clark (this volume) notes an objection to Clark and Chalmers (1998, this volume) from Sterelny (2004) to the effect that external resources occupy a "common and often contested" space. To his thinking, this suggests an important disanalogy with paradigmatic biological cognitive systems, selected for reliability. If this is correct, then it's bad for our own cases, which often explicitly involve some degree of conflict and attempted control. Far from thinking that this is so, however, the extended infant helps overthrow this case.

First, we are not convinced that the world within an individual's skull is sufficiently harmonious and cooperative to sustain the suggested contrast. Conflict—crucially including various internal battles over how to interpret and remember events and actions—is a regular part of our internal lives. The behavior of a would-be nonsmoker who takes steps now to limit access of her future self (tomorrow morning, say) to cigarettes and thus bring it about that she smoked a few less would be mysterious if we
were agents with consistent sets of preferences (to smoke or not, on balance). The same applies to internal debates over whether this situation is special enough, in a good or bad way, for a cigarette smoked now to be an exceptional moment of celebration or commiseration or, rather, a lapse that trumpets another failed attempt to quit. Such conflicts arise because our default discount function for determining the present value of future rewards is not the exponential curve that alone keeps preference ranking constant over time. It seems, instead, as though our default discount function is hyperbolic, so that our preferences can be unstable, and consequently we can anticipate actions that we now would prefer not to make. (Ainslie 2001 provides an excellent account of this research and its implications.) This is not a specifically human phenomenon: even pigeons can work now to prevent future selves from having opportunities to be impulsive (Ainslie 1974).

As well as not being suitably harmonious ourselves, the contrast comes under pressure from the other direction. Competition isn’t only a source of instability, although Sterelny is right that it can be that. It is also a spur for innovations intended to enhance stability. Further, gadgets that can function as cognitive prostheses are subject to extensive (not always successful) testing for reliability and cognitive fit, not to mention ongoing refinement in the area of security. Those of us who trust records kept behind various walls of encryption and access control over our memories are not obviously making a mistake. And those of us who trust the outputs of the various daemons we set up to filter and otherwise operate on incoming information aren’t always wrong either. (One of us has a mobile telephone that rings with a unique melody when the call is from his wife’s mobile phone. It has never played that melody for calls from any other number, not even [as humans embarrassingly can] confusing it with, say, the number of a previous partner.)

Second, we observe that one of the flagship examples of research into distributed cognition concerns multiple agents in a “common” and at least sometimes “contested” space. This is Hutchins’s (1995) work on navigation in large ships, frequently cited with approval by Clark. Hutchins shows that it is exactly by means of key parts of the navigational process being public, that different agents can monitor and correct aspects of one another’s performance, that is, that joint action in a sometimes contested space can be more effective at handling particular information-processing demands.

These brief remarks don’t entirely dispose of Sterelny’s worry, but they do give reason for thinking that at least some of the time our dealings with others are more like our dealings with ourselves than the worry seems to suppose. (The personal is indeed political.) They also suggest that sometimes external resources might have just the reliability (born partly of the sorts of competition Sterelny takes to distinguish on-board biological systems) that Sterelny takes to be the hallmark of internal ones.

Perhaps, though, we have provided no more than an instance of “causal spread” (in which the direct antecedents of some actions lie partly outside the brain, in our case including systems within the bodies and brains of interaction partners), which is what Wheeler and Clark (1999) regard as a “trivial” example. We don’t think so. Some of these extraneural systems do indeed turn out to be at the root of distinctive target features of the phenomenon of interest, which is the criterion for being nontrivial.

Here is a speculative suggestion as to how this might be. In a provocative paper, Ross and Dumouchel (2004) 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). This is because, unlike standard commitment devices, they do not have to be explicitly constructed before strategic interaction. By having preference intensities thus (even if roughly) publicly represented, otherwise intractable strategic problems can be negotiated, and mutually unengenial prisoners’ dilemma situations, sometimes, avoided. Focusing 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:

Thus most games are embedded in meta-games, and there are few restrictions on the possible complexities in this recursion; embedding relationships may stack infinitely, may loop, and so on. This circular dependency implies uncertainty concerning the objects of analysis for which equilibria should be forecast. Reciprocal affective expression can then be seen as a means of reducing this uncertainty.

Through such things as bodily posture, muscle tone, pitch of voice and facial expression, we negotiate reciprocal intentions into tolerably stable sets of expectations within which our base-level games are well defined. At the meta-game level(s) we do not so much exchange information concerning already formed intentions as dynamically influence and determine each others’ intentions though exchanges of affective expression. (Ross and Dumouchel 2004, p. 271)

Our suggestion is that a related function is served by emotional signaling in both interactions between infants and their caregivers, and adult conversation. Thus [be] may serve Halliday’s infant as a strategic signal in negotiating social events (his father may well give him a toy). Equally,
Aldo may find himself signaling strategically and thus avoid the unfortunate dilemmas that arise in overtly contradicting his wife. Unlike many accounts of linguistic and some of strategic phenomena, playing down turn-taking, we emphasize the roughly simultaneous coordination of prosodic and affective display. We have, moreover, argued that coordinated display itself constitutes significant information about relationships. Affective display functions like this in social animals without language and, we contend, 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 is likely 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 coordinated, include both their own vocal production, and that of the other. More generally, all of the types of affective coordination we have described involve roughly simultaneous integration of inputs from each participant’s own behavior and that of others. Since these regulate behavior they are striking examples of distributed control of precisely the kind Clark refers to in reviewing much of his work on “robots.” We hope to have suggested something of how this type of embodied control contributes to the functioning of utterance-activity, and why it merits further empirical investigation.

Second, considering the epistemic payoffs of the types of embodied coordination we have described, it is clear that the model of the solitary infant epistemologist, on which much of the poverty of the stimulus debate rests, is seriously in need of revision. In virtue of affective coordination, infants are able to function as a kind of cognitive extension of their own caregivers. They can use other people to focus their attention, regulate their levels of arousal, reinforce and retard patterns in behavior, and provide many sources of environmental regularity that invite infant exploitation. This interactional environment permits the construction of socially indexical relationships, and the gradual disembedding of labels from their social contexts. Within close relationships, adults treat infant behavior as symbolic long before such a description is warranted. The types of embodied coordination noted above thus permit a particular type of extended mind, in which infants’ cognitive powers are augmented by those of people with whom they interact. When Clark and Chalmers explicitly consider whether other people might form part of one person’s extended mind, they suggest that, after all, it could happen in an “unusually interdependent couple.” Quite so.

Acknowledgments

Earlier versions of this chapter were presented to the “Mind and World” working group at the University of Natal, Durban, and (under the title “Minded Apes, Talking Infants, and the Distribution of Language”) at “The Extended Mind” conference in Hertfordshire. We are grateful to the audiences on both occasions, and especially to Andy Clark, Anita Craig, Andrew Dellsis, Dan Hutto, Denis McManus, Richard Menary, Mark Rowlands, Fran Saunders, Leslie Stephenson, Susan Stuart, and Michael Wheeler for comments and critical discussion.

Notes


2. One of us (Spurrett) is more sympathetic than the other, and we’re aware of more recent and more naturalist work such as Hauser, Chomsky, and Fitch 2002.

3. Results in this area (see also Nazzi, Bertoncini, and Mehler 1998) indicate that rather than distinguishing individual languages, infants at least initially distinguish between stress-timed, syllable-timed, and mora-timed patterns.

4. 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.

5. 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?

6. Ross (2007) gives reasons, in the context of a discussion of distributed cognition, for thinking better of symbols, or digitally encoded signals, than our remarks here allow.

7. Deacon’s broad proposal regarding the role of the prefrontal cortex (PFC) in constructing symbolic relationships is certainly neurally plausible. The PFC includes areas connected with almost all sensory and motor systems, and has extensive back-projections to “lower” systems. Unlike many other brain areas, PFC subregions are specialized for multimodal convergence, and show remarkable experience-dependent plasticity. See Miller 2000 for a review.

8. Garcia and Koelling (1966) studied aversion responses to different stimuli, showing that rats readily learned to associate (a) a noise and light signal with an
electric shock, and (b) a distinctive flavor with (radiation-induced) nausea. The reversed combinations (light and sound followed by nausea, and distinctive taste followed by a shock) were more difficult to learn. Similarly, we suspect it would be difficult for most humans to learn that smiles predicted hostility or sadness.

9. Falk (2004) considers some questions relating to the evolution of language from motherese with reference to the distinctive challenges faced by apes that walked upright and had unusually dependent infants. See also Spurrett and Dellis 2004.

10. A parent may have other children, or may bet on the chances of success with future offspring, whereas the developing infant has no such options. Haig (1993) documents some of the ways in which, during pregnancy, the fetus (which has less interest than the mother in her own other [including 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.

11. A quantitative analysis of the data from which we select some episodes has yet to be performed.

12. We have collected, for a variety of purposes, video data of mother-infant naturalistic, and sometimes more structured, interaction at various ages, and with subjects from a variety of socioeconomic, geographical and ethnic/cultural backgrounds.

13. We are describing a type of interaction here, but including specific details from one particularly striking example.

14. This infant lived in a non-Zulu sociocultural setting in which mothers tend to go to greater lengths to discern (and satisfy) infants' wishes.

15. Papousek (1969) showed, 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.

16. Kano (1992, p. 164, quoted by Falk 2004) describes a behavior in which chimpanzee mothers walk a few paces and then pause but remain in a walking posture, as a cue for her infant to come and climb onto her back.

17. Zajonc (1980, 1984) showed that subjects subsequently preferred images that were "primed" with brief (subconscious) 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.

18. Dimberg, Thunberg, and Elmehed (2000) 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 (1980) showed that smiling changes the shape of the human vocal tract, in ways that raise the mean frequency of vocalizations. Vocalizations with high mean frequencies are generally characteristic of approval (see Fernald 1992).

19. As is often the case (see Bates and Begnini 1979), these have imperative uses (e.g., "up," "more"). It is of interest that while laboratory-trained apes act similarly, even enculturated chimpanzees relatively rarely move to declarative forms of expression (e.g., "dadda," "gone").

20. 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 don't 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" (Dennett 1991, p. 225).

21. Strictly, it begins during the silent bilabial stop or the [p] of "più." For reasons of emphasis, the sound's duration (perhaps integrated with a gesture) is much extended.

22. It could be interesting to know how players responded to an occasionally "nasty" implementation of Tetris that did such things, and in particular what would happen to the rate at which players engaged in epistemic actions.

References


Utterance-Activity and Distributed Cognition


