How Human Infants Deal with Symbol Grounding

Stephen J. Cowley

School of Psychology, University of Hertfordshire and University of KwaZulu-Natal, South Africa

Abstract

Taking a distributed view of language, this paper naturalizes symbol grounding. Learning to talk is traced to – not categorizing speech sounds – but events that shape the rise of human-style autonomy. On the *extended symbol hypothesis*, this happens as babies integrate micro-activity with slow and deliberate adult action. As they discover social norms, *intrinsic motive formation* enables them to reshape co-action. Because infants link affect to contingencies, dyads develop norm-referenced routines. Over time, infant doings become *analysis amenable*. The caregiver of a nine-month-old may, for example, prompt the baby to fetch objects. Once she concludes that the baby uses 'words' to understand what she says, the infant can use this belief in orienting to more abstract contingencies. New cognitive powers will develop as the baby learns to act in ways that are consistent with a caregiver's false belief that her baby uses 'words.'

Keywords: agency, distributed cognition, distributed language, early child development, epigenetic robotics, intersubjectivity, language acquisition, symbol grounding.

Introduction

Church, Turing and Chomsky emphasised resemblances between human utterances and formal symbol strings. By focusing on syntax, they found striking parallels between language users and physical symbol systems. In the long run, however, recognition of semantic differences led to the demise of purely symbolic models. First, since form-based simulations throw no light on understanding (Searle, 1980), they are of limited use in modelling mind. Second, in spite of Turing's (1950) exhortations, it is difficult to design machines that *learn* linguistic forms. As a consequence, the external grounding of understanding has come to be recognised. To solve the symbol grounding problem (SGP), representations must connect with both classes of referent and the appropriate word-forms (Harnad, 1990). While some action needs no symbols (e.g., Brooks, 1999), agents that interact with an environment can develop symbol-systems that give flexibility to their perception and action. Much can be gained from using action-guided representations (Anderson, 2003a, 2003b).

In their review of attempts to solve the SGP, Taddeo and Floridi (2005) conclude that representational, semi-representational and nonrepresentational models can all make symbols *seem* value-laden. However, this impression relies on human interpretation, because the manipulation of the symbols is hard-coded. This "semantic commitment," they conclude, means that the SGP remains unsolved. Even less progress has been made in mapping symbols to verbal patterns. To ground formal properties of a burst of speech (e.g., *pferd*), agents need to categorise human expression.[1] It is, however, difficult to derive word-forms from natural speech. Further, when approached from a practical perspective (Seabra Lopes & Chaudhun, 2007) words and perception are dealt with together. Simpler solutions, it seems, use human teaching rather than categorisation. Further, it has been found that formal patterns can warp the perception of even artificial agents (e.g., Cangelosi et al., 2002; Steels & Belpaeme, 2005). Today, therefore, there is growing interest in semiotic approaches which, it is hoped, can overcome the limits of formal modelling.

Naturalising Symbol Grounding

Instead of focusing on the logic of symbol grounding, I aim to naturalise the problem. By taking a developmental view, I show how infants ground action-guided representations into language in its broadest sense. While working conceptually, my objective is to encourage engineers to build language-using robots. In reinterpreting empirical work, I trace human symbol grounding to the agent's history of co-ordinating with others. Thus, newborns have become quite different kinds of agents by the time have developed into infants who are able to ground representations into objects. This agency is further transformed by the time that they come to hear utterances as 'words'. As agency changes, moreover, so do the functions of expression. While small babies rely on affect-based dynamics, coordinating prompts them to use human signals in ways that impose norm-based structure on the world.[2] Thus, around their first birthday, infants may begin to use sound-patterns that are heard as, for example, *milk* or *gone.* Using their grasp of situations, they rely on adults to interpret their utterances. Learning to talk thus occurs as infants and caregivers become skilled at acting under dual control (Cowley et al., 2004; Cowley, 2006a). Over time, dependence on innate biases gives way to increased use of body-world co-ordination and, later, to reliance on a nascent self.

Naturalising the process of symbol grounding emphasises how brains and bodies are co-ordinated during joint behaviour. While the importance of social-coordination is often recognised, it is often forgotten that the body enacts language (and other expression). Indeed, infant behaviour is compatible both with Cangelosi's (2006) emphasis on social factors and with Steels and Belpaeme's (2005) emphasis on how a robot can learn to integrate co-ordination with perception (cf. MacDorman, 1999). Further, if social co-ordination anchors robot categories, similar processes may well alter human agency. Below, I suggest that the affective properties of utterances prompt infants to apply new kinds of leverage to caregivers. In emphasising how activity is co-ordinated (see Spurrett & Cowley, 2004), I stress that babies use *culturally* evolved patterns. While initially relying on affect, they come to grasp how values and beliefs play out in a range of situations (e.g., "Be quiet!"). Interaction gives way to co-action that provides a basis for speech-mediated communication and, later, context-free conversation. Pursuing this, I reinterpret old findings against a new perspective. By appealing to the concept of *distributed language*, I show how human agency arises from events under the dual control of mother and caregiver. This allows the division of symbol grounding into phases. While infants initially align behaviour to the sound-patterns of a community, they later develop ways of acting that sustain belief in *language* (and *words*). After summarising this as the extended symbol hypothesis, I sketch later changes and explain how the model applies to machines.

Distributed Language

Following Saussure (1916), linguists often conceptualise language in terms of 'use.' Instead of regarding talk as behaviour, they posit code-like forms (or signs) as constituting 'language.' While this view arose in contesting 19th century work, its logic permeates most 20th century theory (Matthews, 2001). The code view characterises Saussure's (1916) view of language-systems (*langues*) and dominates appeals to habits (Bloomfield, 1933), utterances (Harris, 1951), functions (Halliday & Matthieson, 1999; Martinet, 1960) and form-based systems (Chomsky, 1965; Jackendoff, 2002; Lakoff & Johnston, 1980; Pinker, 1994). In each theory, language is identified with formal units. As in folk psychology, language is reduced to verbal patterns whose forms are independent of history, affect and motor activity (viz. acts of utterance).[3] Code models thus distinguish the patterns that are represented (forms) from the representing agent. In so doing, they posit that cognition is separable from embodied action and perception. This epistemic conception of mind (Cowley & Spurrett, 2003) enables language to be modelled in terms of input/output processes. However, the price paid for this, is high: it implies that how babies act, feel and move has no relevance for learning that exploits (what we call) linguistic forms.

On the distributed view, language is embodied, individual and cultural (Cowley, 2006a). Agents can reflect on the significations of words or, in 19th century terms, draw on historical facts. With time, we learn to reflect on words and, by so doing, to stabilise their value. We thus resemble linguists who, as Paul (1891) observes, rely on how words are "historically derived" (p. xlviii). Language thus permits reflection, unites individual and cultural experience and, as a result, has a constitutive role in behaviour. In recent terms, language links a meshwork of heterogeneous elements (e.g., utterances, texts, pop songs) that emerge as a result of creative drift. As DeLanda (1997, p. 39) says of towns, language is an "interlocking system of complementary... functions" that are spread across space and time. In language, bodies connect a meshwork whose basis is simultaneously social, historical and individual. Like a city, language grows as well-established structures like hierarchies and paradigms come to be articulated in new ways. Most notably, in the last few thousand years, artefacts that include writing instruments, documents, books, and computers have transformed oral practices (Donald, 1991; Clark, 1997). These too interlock cultural patterns while remaining subject to the constraints of human evolution and development.

In a distributed meshwork, no appeal is made to external languages.[4] When learning to talk, infants have to discover ways of integrating their own feeling and moving with that of others. They develop skilled behaviour that extends already grounded symbols. Much depends on how these action-guided representations affect the infant's developing brain. Intrinsic motive formation (Trevarthen, 1998; Trevarthen & Aitken, 2001) exploits how caregivers spread information in getting babies to attend to aspects of the environment. Far from facing a problem of symbol grounding, action-grounded symbols function as cognitive resources. Higher order structures develop out of joint activity that connects an infant's lower-order representations with how adults enact what they are trying to do. In line with Davidson (1986) and the later Wittgenstein (1953), language is an intrinsic part of our cultural life. At the neural level, perception and action connect mammalian representations (symbols) with systems that motivate acting (and thinking). Over time, infants come to hear sounds that can be analysed formally even if, early on, they rely heavily on their affective value. Language is hybrid or, in Love's (in press) terms, both analogue and digital (i.e., categorical). Indeed, this dual nature is crucial in learning both to talk and to acquire many other skills. In reporting conversational events, for example, we can describe how *words* and *tone* contribute to their particular sense. Both features matter because our wordings depend on both the words actually spoken and displays of affect. Language uses cognitive dynamics that alter how others feel, think and act. While verbal patterns constrain language and sense-making, their roots lie in micro-temporal dynamics.[5]

There are many reasons for adopting a distributed view. First, talk is both dynamical and dependent on rate-independent switches. While the ability to repeat what we hear depends on patterns, dynamics are needed to render incomplete information into a new structure. Second, this is compatible with the physics of symbols (Pattee, 2001). Far from needing to implement rules, infants can become rule-followers by drawing on real-time dynamics. Their future action may depend on, not represented wordings, but a history of co-ordination. To engage with language, infants' bodies and brains may attune to contingencies that are normatively constrained. Third, natural history can draw on a co-evolutionary process. Rather than posit a syntactic 'core' to language, weight falls on sensorimotor toil (Cangelosi et al., 2002). Learning to talk is thus grounded in how affect shapes the dyad's use of biomechanics. Fourth, linguistic 'forms' become entities guaranteed, not by brains, but talk about talk (and how this *linguistic reflexivity* is institutionalised in dictionaries, grammars, spell-check programs, etc.). For the modeller, this has advantages. Artificial agents can use programs, spreading activation and extended symbol systems that use a history of co-ordination. New importance falls on building machines that, like infants, align to human expectations. Although the behavioural capacities for meaning-making "*appear* to be symbolic" (Harnad, 1990, p. 337; my italics), their basis may lie in co-ordinated dynamics. Fifth and finally, the distributed perspective fits conversational facts. For example, Cowley (1997) describes an utterance "*marrone*" that is heard as "how can that possibly be chalk?" While based on neither the syntax nor semantics of *marrone* (Italian for 'brown'), an exquisitely timed vocal duet enacts understanding. Just as in using an abacus, bodies exploit both dynamics and relations between virtual resources (words or numbers). To discover verbal patterns, an infant does not rely on hearing words, but on how adults grasp what the infant and adult are doing together.

Co-action and the Emergence of the Person

A physical symbol system changes its powers, first, by altering how it operates on symbols that stand in for aspects of the world. Second, it uses these operations (including perception and action) to re-categorise the perceived environment. This same logic applies in grounding symbols into, say, a simulated mushroom world or displays of colour. Applied to verbal patterns, however, a different logic applies. This is because, the multidimensional and fluctuating complexity of the speech stream relates indirectly to word-forms. Symbols are not simply grounded in the real-world objects, events, and relations they are supposed to represent but also in the continuous vocal fluctuations that contribute to co-action. Indeed, it is gratuitous to assume that babies use speech to develop explicit representations of linguistic forms. While signals (e.g., Morse code) *could* be used to develop something like a notation, speech lacks the necessary salient units.[6] Empirically, it consists neither of words nor phonological units but is, rather, a fluctuating set of multidimensional patterns (Port & Leary, 2005). Segmental structure is an analytical product and, for a baby, less salient than prosodic and voice quality patterns. Why should an infant seek out speech sounds? It is thus more parsimonious to seek the bases of learning to talk in expressive dynamics. Because language spreads across bodies, infants can use caregiver affect together with how, in real time, adults construe the ways in which infants express themselves.

Multimodal expression not only makes certain patterns salient but also imbues them with expressive value. Caregiver expression, moreover, prompts babies to change behaviour and develop their own motivational systems. Given the distributed nature of language, they can derive their powers from co-ordinating with caregivers who integrate verbal, vocal, visible and tactile expression with what is happening. Full-bodied resources serve, initially, in learning to inhibit crying and, later, to track and use features such as caregiver smiles and frowns. Affect thus enables babies to identify the likely results of their response to the caregiver's expression. While this social learning occurs under dual control, it is otherwise compatible with Piaget's view of sensorimotor development. Neural structures can be said to arise from processes of *accommodation* that develop as an infant copes with the environment. Agency changes as the infant exploits physical patterns. When learning under dual control, caregiver reactions to infant appearance and behaviour prompt the baby to use what Piaget would call *assimilation* in developing a "sense of persons" (Legerstee, 2005). Events depend on cycles of orientation and re-orientation where behavioural invariances are less salient than affect and timing. Moments feature vocalisations that are accompanied by smiles and frowns among other things. These come to be associated with contingencies that will both influence current activity and, in the long term, alter infant motive formation.[7] Strange as it may sound, infants exploit what they learn about social situations and, specifically, the "setting." They come to exploit what caregivers believe, for example, as they accede to demands for silence. In human symbol grounding, therefore, much depends on the caregiver's way of presenting situations. This, in turn, affects motive formation, ensuring that social development occurs slowly. For the purposes of this paper, social development is most usefully viewed as falling into two stages:

- infants align behaviour to sound patterns favoured locally
- they develop ways of acting that sustain belief in *language* (or *words*)

By the time children enter school, they treat words as indexing (relatively) stable meanings. Below, however, I focus on the earliest phases.

Learning to talk depends on a distributed linguistic meshwork in which sensorimotor activity becomes sensitised to both affect and shared use of social norms. The infant develops skills by orienting to caregiver behaviour and, later, discovers that others believe in *words*.[8] Early on, brains need only sustain action that can be *described* by the code metaphor. To say, *more* (or *bikkik gon*), the child neither needs phonological nor semantic representations but a capacity to use situated experience in deciding how to vocalise. The baby integrates what can be perceived with expressive action. Later, as a fully fledged person, it will take the circular view that, as a member of Community X, it uses the words of Language X. First, however, the infant must learn to use situations to get what it wants. Since this depend on how adults interpret utterance acts, the child uses the *adult's* language stance in mimicking vocal patterns. To adult ears, of course, it acquires Language X. In fact, however, the baby uses vocalisation to connect contingencies with caregiver expression and how *de facto* interpretation bears on social norms. The process falls into the following stages:

- Living at a body-world boundary (0-6 weeks)
- Early self-implicating behaviour (6-12 weeks)
- Early self-directed behaviour (4-9 months)
- Early self-regulating behaviour (9-12 months)

Using inner motives (action-guided symbols) together with how caregivers strive to control events, the baby constructs its agency. Dyads live a history of interaction where routines emerge from shifting dual control. Perceptions of how adults enact beliefs trigger infant movements based on the dyad's previous experience. By four months, infants begin to use culture in response to caregivers wants (e.g., *hold your tongue*). Capacities for body-world interaction thus develop alongside social skills as infants discover how expression can be used. The growing use of social norms *extends* action-guided symbols.[9] Gradually, co-action enables the baby to motivate the caregiver to re-structuring what they do. The dyad develop games or formats while, in solitary settings, the child learns about the physical world. Towards the end of the first year, the child connects its sense of persons with object knowledge. By linking norm-based patterns with actions on objects, it comes up with new ways of engaging with the caregiver. It develops as a self-regulator who can become skilled in getting others to do things.

Getting Started

Caregivers attempt to stimulate co-action even with newborns. They seek to get babies to respond like 'little persons.' This probably goes some way to explaining Meltzoff and Moore's (1981) finding that newborns respond in kind to a few human expressions. For example, minutes after birth, they can react mimetically to tongue protrusions (Kugiumutzakis, 1999). The function of these imitative capacities depends on the somatic marking of adult response. Affective comeback leads to increased co-ordination where adults seek to re-evoke previous infant response. At times, of course, this will prime the baby for future events. As with early tongue protrusion, this is likely to depend, in part, on parental and cultural views. Given its sensitivity to affect, the infant draws on contingencies. What a baby does in the future depends on *de facto* judgments about events at the body-world boundary.

Babies are active systems. Newborns actively *prefer* speech that has voice quality features correlated with their mothers and the rhythmical patterns of her cultural group (Karmiloff & Karmiloff-Smith, 2001; Legerstee, 2005). They elicit affective expressions that bear a cultural stamp. Although it is not clear when parental beliefs begin to shape behaviour, adult come-back soon begins to encourage, reward and disappoint. By the age of three months, babies and adults exploit what Trevarthen (1979) calls *primary intersubjectivity.* They become self-implicating by, for example, developing *personal* characteristics and, even if blind, co-ordinating movements to adult vocal rhythms (Tønsberg & Hauge, 1996). For example, as discussed below, they may draw on cultural signals to fall silent at an adult's behest. This is, perhaps, the breakthrough: the baby's control systems pick up culturally-specific cues concerning what is wanted. Infant biomechanics gradually sensitise to how caregivers enact local norms. Drawing on operant conditioning, their brains develop higher-level motives that control lower-level mimetic activity (Trevarthen, 1998; Trevarthen & Aitken, 2001). Infants develop a style of autonomy that depends on value-guided co-action.

Talking to Babies: From Co-Action to Alignment

Co-action is the basis for all social development. Unlike in much interaction with the nonhuman world, rewards are associated not only with physical events but also how these relate to culturally-valued contingencies. Thus the neural systems that drive infant action (and inhibition) become sensitised to cultural norms or apparently arbitrary forms of signalling. While depending, in part, on hard-wired systems such as those used in smiling and gaze, the result is situation-sensitive action. Once co-action becomes established, events connect micro-timing, experience and, at times, caregiver interpretation. This ensures that the self-implicating infant begins to act in ways typical of persons. For example, a routine that emerges under dual control allows a three-month-old baby to stop crying (Cowley et al., 2004). What happens arises as behaviour is co-ordinated across (at least) two time domains. While the baby's inhibitory response picks up on micro-features of interaction, the caregiver structures events in the slow domain of deliberate action. Further, this dual temporal structure gives co-action a flexibility that can prompt a baby to sensitise to caregiver attributions. For example, in one family, adults may reward infants who cry, say, in supermarkets. Many babies use contingencies to act in ways that are rewarded in such settings.

Cultural norms (e.g., belief in fun and/or respect) both meet adult expectations (when it suits the baby) and help the baby to develop a personality. When Trevarthen (1979; 1998) and Bråten (1988) first noted the intrinsic motives of babies, they appealed to neural systems that sensitise to the doings of a caregiver-in-the-niche. In contrast, Stern (1977) and Kaye (1982) sought to explain the events in terms of attunement. Regardless of theory, however, the infant's motives form in response to its evaluation of social, cognitive and emotional aspects of co-action. Dyads act according to preferences and motives that reflect what caregivers attribute to infants. Thus, constraints, such as the caregiver's beliefs, that are opaque to the baby can shape the course of co-action. By integrating events across time-scales, social learning sets off spiralling changes in infant reactions to, for example, voice quality and prosody. If we consider twins, they begin to show distinctive personalities when, as infants, they become self-implicating. This is also when cultural differences emerge

(Cowley et al., 2004; Trevarthen, 1988). Caregivers interpret previous experience of co-action through a cultural lens, and it affects their treatment of babies. There is variability both within and across cultures with regard to how caregivers respond to a baby judged to be sick, naughty or tired. Babies thus learn from more than the words that can be heard. Further, while children understand little, parents nurture intelligent response with speech that is also affective expression (*child-directed speech*). The baby's inner symbols are further extended as the functional value of caregiver activity comes to be related to judgements of infant *understanding*. Indeed, what is most extraordinary about human babies, which are hyper-social by the standards of the animal realm, is that they come to exploit signs of culture even before they act in self-directing ways.

In the fourth month, infants typically begin to reach for objects. It is striking that this self-directing behaviour develops in infants who have well-established situation and culture-relevant skills. Self-directed activity develops in a self-implicating infant who has already attuned to cultural settings. Thus while Western caregiver-infants dyads may focus on fun, in Africa weight is also given to activity around, for example, movements resembling song and dance (Cowley, 2003). This is especially marked in the formats that, by the middle of the year, dominate a baby's social life (Bruner, 1978). Strikingly, however, this is also a period in which attention to objects largely eclipses interest in social activity. Where possible, this attention to objects helps babies learn about their tangible and oral properties. While also gaining mastery of vocalisations (Oller, 2000), this may be independent of social motives. Rather it seems related to sensorimotor skills (e.g., sitting and crawling) and preferences for, say, what to suck. It should not be thought, however, that self-directing infants are asocial. Alongside body-world interaction they continue to build formats for co-action.

Caregivers, moreover, link norms with biomechanics in, for example, seeking to prevent *bad habits* (e.g., *eating sand*). Self-directing babies are primarily focused on grounding symbols into the perceived world. At the same time, how caregivers use expression to control what happens influences the ways in which settings are seen.

At the end of the first year, children find new ways of exerting leverage on caregivers. Once again, they exploit manifest attitudes but, in this phase, connect these with the perceived properties of objects. Thus, in *social-referencing* (Campos & Sternberg, 1981; Striano & Rochat, 2000), the caregiver's dynamics may be used, for example, in deciding whether to venture beyond a visual cliff. This self-regulating behaviour, it seems, links object-perception with a history of social experience. Adult responses make the behaviour, in Cowley and Spurrett's (2003) terms, *analysis amenable*. It becomes possible to adduce new mentalistic reasons in accounting for what the baby does (e.g., "He thought that it was dangerous"). In this 'nine-month revolution' (Tomasello, 1999), a baby becomes self-regulating. What Trevarthen and Hubley (1978) call *secondary intersubjectivity* develops or, in our terms, object perception, connects with a sense of persons.

By their first birthday, children regulate activity by using (what adults call) words and gestures. Adult analysis leads to behaviour change that, once again, spurs learning. Further, since thinking involves the slow domain of wordings, children find that inhibiting micro-activity opens up new opportunities. For example, Cowley (in press) describes how Luke, a nine-month old, is induced to fetch a plastic brick. Luke has skills in, for example, holding an object in mind (working memory) and experiencing authorship (a feeling of conscious will). He can thus concentrate on the object while using maternal movements to link sucking, gaze-following, turn-taking, crawling and grasping. After the brick falls from Luke's mouth, nine seconds of focused attention give him time to anticipate that reward will accrue from fetching. As usual, the baby depends on both maternal micro-expression and how she acts to induce understanding. Strikingly, Luke uses the mother's manifest belief that he will be able to grasp what she wants him to do. On the slow time scale, Luke aligns to what she says ("*Do you want to fetch that*?") by, smiling and crawling off to get the brick. The example is of a kind that often contributes to formats and, just as crucially, shows how caregiver movements enact a language stance. If a 'fetching game' develops, over-attributions may even prompt the mother to highlight sound-patterns (viz., "*fetch"*). The child thus becomes an intention reader across repetitions by, literally, using biomechanics to sense what can be done. Analysis-amenable behaviour emerges through the use of the other person's manifest goals.

According to this view, behaviour extends the form of co-action that emerged at three months. What has changed is that the child is now motivated to integrate self-implicating skills with capacities based in self-directed activity. As before, the baby exploits biomechanics that arose during its history of co-action with others. In short, children link their sense of persons with knowledge of what they can do with caregivers. The rise of this *triadic behaviour,* then, further extends grounded symbols. Infants link their doings to objects and real-time expression in ways that align what they do to verbal patterns: When asked to fetch the block, the baby fetches it. The approach proposed here is novel in providing an explanation that depends, not on appeal to wordings, but to the child's having learned to use situations in self-regulating ways.

All agree that such events emerge at this stage and, equally, that they are important in learning to talk. Where the distributed view diverges from others is in presenting this as independent of speech sounds, a lexicon, or words and referents. Nor is there any need for Tomasello's (1999) species-specific intention reading. Rather, a child's self-regulating powers depend on linking person perception with what objects afford. This, of course, is not to deny that sound patterns can mark situations, events and, perhaps, objects. Indeed, in the following weeks, children may supplement (apparent) understanding with acts of utterance and pointing. Given role reversal (Tomasello, 1999, 2003) one-year-olds may make a "*more*" sound to ask for cake (in English-speaking settings) or use "*aah*" to show surprise. Once a baby is self-regulating, systematic use of sound patterns can bring many gains. While still relying on how inner symbols are extended, a caregiver's response to a baby's utterance-acts gives new information about how local norms constrain the situation. As the child gains experience of what to expect, vocalisations can be used more effectively. Eventually, it will learn to use them when referring to utterances made by both caregivers and its own *self* and, at this point, will be close to developing belief in *words*. Having used how caregivers integrate speech and action, aligning to verbal patterns will then reveal other levels of interactional structure. The helpless infant uses such tricks in gradually becoming a 'little person.'

The Extended Symbol Hypothesis

Learning to talk depends on motivated actions which are integrated, in increasingly sophisticated ways, with norm-based expression. Thus, by the end of the first year, a child can align to a few sound patterns used by caregivers. This depends, I have argued, on a history of co-ordination that, by the fourth month, exploits the co-ordinated workings of two brains. According to this view, contingencies that link situations to brain-side events transform the child's agency. Learning to talk depends on how changing patterns of co-action extend the baby's already grounded symbols. This hypothesis can be summarised as follows:

(1) Infants use biomechanics to exploit caregiver bodily movements that represent local norms (values, beliefs and practices).

(2) Biomechanics link affect with caregiver expectations, beliefs and cultural values.

(3) Caregivers attribute sense to the biomechanical output: given affect, these beliefs can shape future behaviour.

(4) Infants integrate a practical grasp of local norms with what caregivers want and thus gain practical knowledge of situations.

(5) Given dual control, just as the caregiver shapes the baby's strategic signals, the baby alters those of the caregiver.

(6) A baby forms a motivation system; a caregiver uses its products to construct a baby's de facto understanding.

According to the extended symbol hypothesis (ESH), social development integrates growth and learning with the social events that forge infant agency. Given openness to culture and changing motivations, learning to talk is surprisingly transparent. The baby depends, above all, on the use of affect to discover cultural norms. Infants thus sensitise to situations in ways that exert leverage on adult doings. Gradually, as adult goals change, infants develop their agency and, after nine months, begin to align their doings to verbal patterns. Learning to talk thus emerges under dual control. It depends on a baby's changing motivations as well as how caregiver attributions are enacted in real-time. What adults hear as first *words* are, thus, based in how affect has shaped a history of co-actions that elicited manifest caregiver beliefs (and thoughts). For from being output from a neural system, these *words* constitute a new form of jointly regulated behaviour.

Beyond the Symbol Grounding Problem

The SGP can be addressed by making invariant features of an environment relevant to action by symbol processing agents. While logical, this conceptualisation throws little light on the grounding of social symbols. Infants extend already-grounded symbols and, in so doing, have no need for speech sounds. Learning to talk, in other words, does not depend on grounding symbols in verbal forms. Rather, it is a consequence of how agency changes as infants become skilled in co-acting. Much thus depends on the caregiver's consistent use of the language stance. Thus while (inner) symbols are physically grounded, early *words* are situationally sensitive actions. Far from representing speech sounds to themselves, infants rely on biomechanics. Given that these shape motive formation the consequence is that infants gradually become self-regulating agents. Avoiding the hard task of grounding verbal forms, babies link experience to how caregivers enact attributions and beliefs. The extended symbol hypothesis thus has striking implications for language-using robots. It implies, above all, that expression can transform agency. Elsewhere, it is argued that attention be given to using multimodal action to simplify cognitive problems (Cowley & MacDorman, 2006). This raises the issue of whether robots can use learning to develop aspects of personhood (MacDorman & Cowley, 2006).

In human symbol grounding, co-action transforms agency. By 12 months, infants exploit situations to utter grounded wordings. Once children become self-regulating, they can use these wordings to develop new wants (or goals). For example, they may discover, as Luke had learned, that caregivers can be used to fetch items.

This, however, depends on acting in self-regulating ways by using sound patterns that, broadly, make sense against another person's expectations. Eventually, using increasingly deliberate forms of action, children come to hear utterances as made up of recurrent patterns of units that can get things done. As the trick gets used, children begin to act as if they were users of a culturally evolved phonological system (Worgan & Damper, 2007). Later, their brains will enable them to develop mastery of constructions (Tomasello, 2003) while, in all likelihood, making use of mimetic schemas (Zlatev, 2002). Even after the first year, adult attitudes and beliefs will nudge children towards use of verbal patterns. Children, however, also learn new skills. Above all, an emergent ability to repeat what they and others say will, eventually, enable them to talk about talk (Taylor, 1997). In learning to hear words an infant acquires routines like saying *"What is that?"* As Vygotsky (1986) noted, a similar process occurs as self-directed speech gradually becomes silent rehearsal. Indeed, it is likely that developed forms of self-regulation use word-infected symbols as the basis for new powers of verbal imagination. This, moreover, fits Viger's (2007) appeal to an *acquired language of thought.* In contrast to the claims of code models, human language is grounded in the use of an emergent grasp of situations to control vocalisations. Symbol grounding allows culturally-evolved norms (and verbal patterns) to be used in forms of action that transform our cognitive powers.

Models and Applications

The claim that babies gradually extend grounded symbols until they are heard to be speaking has many implications. First, it becomes important to separate evolutionary from epigenetic emergence because, implicitly, language and talk have different histories. Conceivably, while verbal patterns derive from cultural selection, these vehicles are of no importance to infants. Accordingly, while it makes sense to model the emergence of verbal (and phonological) patterns in populations, these units may be developmentally marginal. Those concerned with interaction, therefore, might gain much from building machines that use situations to alter their powers. In principle, machines too could use co-action to align to social norms (and, indirectly, words and values). According to this view of human symbol grounding, the roots of utterance acts lie, not in verbal patterns, but in the semantic properties of co-action. While the suggestion is, I think, novel, it is in line with much contemporary thinking. First, following Humphrey (1976), there is a wide consensus that human intelligence has a social basis which, for many, depends on something akin to an extended phenotype (e.g., Dawkins, 1982; Deacon, 1997; Laland, Odling-Smee & Feldman, 2000). Developing such views, Ross (in press) construes human signalling as unique in its capacity to impose structure on the world. This *digitality* draws on compressed cultural information that may appear, for example, in how a mother rebukes her child or a child's attempt to change her attentional focus. Similar lessons seem to be implied by Cangelosi's (in press) simulations which are said to show, for example, that 'verb' is not an abstract category. Rather, the term characterises signals that are "integrated with an agent's past history" in ways that shape behaviour. While emphasis on co-action is novel, one must ask if artificial agents could use cultural contingencies. This would link with the strategic emphasis of game theory and, specifically, the view that affect is the main currency of social interaction (Ross & Dumouchel, 2004). Indeed, in human symbol grounding, affect prompts babies to use cultural constraints. Culture is a supra-individual entity that, among other things, enables biological individuals to become people.

Infants neither possess nor learn quasi-verbal symbols. Rather, human symbol grounding depends on learning to act strategically in the social situation. Language functions, therefore, by enabling babies to develop into persons. Not only is this learning emphasis compatible with Turing (1950) but it fits Dennett's (1991) view of selves. Beyond this, it throws light on, say, the achievements of Alex the parrot (Pepperberg, 1999) and Kanzi the bonobo (Cowley & Spurrett, 2003; Savage-Rumbaugh et al., 1998). Both learn by integrating complex biomechanics with trainer use of the language stance. As a result, both become strikingly person-like. Could artificial agents use co-action to produce similar outcomes? This view of human symbol grounding thus suggests new issues for human-robot interaction and, above all, android science (MacDorman & Ishiguro, 2006a, 2006b). Stressing the importance of experience-prompted action[10], Cowley and MacDorman (2006) propose that human micro-responding could, in principle, enable androids to discover social norms by using human displays of beliefs and values. The distributed view of language thus has striking consequences for robotics. Indeed, a system that mimicked babies would learn by connecting human expression to the world. Integrated intelligence (Seabra Lopes & Connell, 2001) could link material and semantic aspects of utterances. Indeed, by emphasising contingency, models of encounters become relevant to robotics. Such machines might encourage us to co-operate by acting to mimic how we respond to feelings (Cowley & Kanda, 2005).

Conclusion

Symbol grounding is inseparable from Chomsky's insight. Since human agents follow rules that can be formulated (not mindless distributional analysis), language is surely the source of productive, compositional and inference-bearing thought. Tying this to brain-centred models, Harnad (1990) argues that the underlying units are explicitly represented. In his formulation of the symbol grounding problem, therefore, the system maps real-word references onto the symbols on which rules operate. In seeking to naturalise the issue, I have accepted Harnad's view of how agents ground symbols into objects. In relation to learning to talk, however, the evidence suggests that the process proceeds quite differently. In human symbol grounding, I have suggested, infants have no need for explicitly represented linguistic forms (or *words*). Rather they rely on how caregiver construals enable them to develop strategies for acting – and vocalising – in response to situations. The symbols which ground objects are thus extended in ways that differ from those needed to ground 'linguistic forms.' While we learn about the physical world by interacting with objects, human expression and affect mediate our dealings with verbal patterns. Initially at least, babies learn to talk by using strategies that are parasitic on the caregiver's language stance.

Using Saussure's view of forms, Bloomfield, Harris, Chomsky and Harnad all accept the folk view that language is grounded in *words*. This, I have argued, is to confuse an analytic claim with an empirical one. Like other behaviour, language is based in action and perception. Specifically, as babies extend grounded symbols, verbal patterns emerge in co-action where the behaviour of both parties orients to local norms. Thus while formal models capture aspects of what the child masters, the resulting language knowledge differs from what analytical models describe. As Matthews (1967) argues in his review of *Aspects of a Theory of Syntax* (Chomsky, 1965), infants do not learn to talk by representing what formal models make explicit. Developing this view, I suggest that infants learn to talk by elaborating already-grounded symbols with respect to an expected response from a caregiver. Implicitly, the verbal patterns that shape what they do arise, not from a single process (e.g., natural selection, ontogenesis), but events that derive from several time-scales. Far from explicitly representing verbal patterns, infants use expression as a basis for constructing *selves*. When Luke fetches the block, he aligns to what is said and, thus, enacts a semantic interpretation. In his world, he becomes a fetcher.

Far from solving the SGP, appeal to human symbol grounding shows the problem is more complex than previously thought. Indeed, if the extended symbol hypothesis is correct, those who examine language grounding have been looking in the wrong place. Human agents ground symbols into verbal patterns, not by categorising expression, but by prompting caregivers to make consistent use of the language stance. As motivated co-ordination develops, infants slowly develop into agents with the necessary powers. Just as to dance one need to become a dancer, to talk, one needs to become a talking agent. This peculiar capacity depends, on the one hand, on neural systems that shape motives and, on the other, on using affect to become sensitised to expectations. Since human infants are transformed by what situations (objects and people) afford, the grounding agent (a system) cannot be separated from the behaviour that is grounded (language). This claim has important implications for those who aim to develop semiotic models. It suggests that while agents need to become action guided perceivers, it is likely that, for brains, there can be no such thing as a pure symbol.[11] Words are *a posteriori* constructs that exist for a self-regulating agent who acts to draw on the doings and beliefs of others.

For modellers, the perspective poses the issue of how agents learn to exploit cultural constraints (especially nonverbal contingencies). Moreover, it posits that human infants use verbal patterns to shape co-occurring actions. According to the extended symbol hypothesis, however, words need never be explicitly represented.

Rather, co-action serves in discovering verbal patterns that, later, shape both inference-based reasoning and identity. Human subjects, then, are themselves the products of how expressive dynamics connect bodies with the verbal patterns of a cultural heritage. With Ross (in press), signals can function to repartition the world. We are unlike other primates in that we learn to adopt and alter perspectives on what is possible. The distributed view of language thus implies that population models need to explore how the (putatively) co-operative emergence of verbal patterns interacts with game theoretical models of human strategic signalling. The question, then, is whether we can build robots that use human co-action in ways that allow them to construct new levels of agency.

The extended symbol hypothesis has implications for epigenetic models. Given a Piagetian heritage, these often focus on how learning and growth shape interactions with a physically constrained world. Once language is viewed as a distributed meshwork, however, this seems unnecessarily narrow. Indeed, as argued above, much depends on how each party adjusts to the other both biomechanically and in the time domain of deliberate action. Inexperienced agents, I have claimed, unwittingly use this process to exploit culturally-structured events. By means of co-action they integrate aspects of social behaviour across several time scales. For robotics, if the ESH is broadly correct, many aspects of human agency or personhood may emerge as control systems integrate information from diverse sources. In other words to recognise a range of sound patterns as what we can write as *pferd,* we need neither phonological nor semantic categories. Rather the agent must be a self-regulator who has developed in a world where what English speakers call *horses* are salient during co-action.

Much can be gained from rethinking how infants extend the scope of already-developed (physically grounded) symbols. In seeking to develop more plausible views of child development, the two-stage model needs to be evaluated. Ideally, this will be done by robotic modelling. First, it needs to be shown that culturally constrained adult biomechanics can lead to significant changes in agency. In humans, then, one-year olds become agents who, occasionally, align their doings to verbal patterns. Second, we need to explore the view that the capacity develops as children come to act consistently with a language stance. They gain not only from acting as if they knew words but also talking about language (and linguistic knowledge). Humans need not learn natural languages because, oddly, it is enough to imagine that these systems are already known (in a community). A child becomes a person by coming to adopt the language stance. If this is plausible, the view has many consequences for human nature. Among other things, the extended symbol hypothesis offers new scope to models of how we can develop and sustain human kinds of autonomy.

[INSERT NOTES HERE]

References

Anderson, M. L. (2003a). Embodied cognition. *Artificial Intelligence, 149*(1), 91-130.

Anderson, M. L. (2003b). Representations, symbols and embodiment. *Artificial Intelligence, 149*(1), 151-156.

Bloomfield, L. (1933). *Language.* George Allen & Unwin: London.

Bråten, S. (1988). Between dialogical mind and monologic reason: Postulating the virtual other. In M. Campanella (Ed.), *Between Rationality and Cognition*, pp. 205-236. Turin: Meynier.

Bruner, J. S. (1978). From communication to language. A psychological perspective, *Cognition*, 3, 255-287.

Brooks, R. A. (1999). *Cambrian intelligence: The early history of the new AI*. Cambridge, MA: MIT Press.

Campos, J. J. & Sternberg, C. (1981). Perception, appraisal, and emotion: The onset of social referencing. In M. E. Lamb & L. R. Sherrod (Eds.), *Infant social cognition: Empirical and theoretical considerations*, pp. 273-314. Hillsdale, NJ: Lawrence Erlbaum.

Cangelosi, A. (2006). The grounding and sharing of symbols. *Pragmatics and Cognition*, *14*(2), 275-285.

Cangelosi, A. (in press). Adaptive agent modelling of distributed language: Investigations on the effects of cultural variation and internal action. To appear, *Language Sciences*.

Cangelosi, A. Greco, A. & Harnad, S. (2002). Symbol grounding and the symbolic theft hypothesis. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language*, pp. 191-210. Springer: London.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.

Cowley, S. J. (1997). Of representations and language. *Language and Communication*, *17*(4), 279-300.

Cowley, S.J. (2003). Distributed cognition at three months: mother-infant dyads in kwaZulu-Natal. *Alternation*, 10(2), 229-257.

Cowley, S. J. (2004). Simulating others: The basis of human cognition? *Language Sciences*, *26*(3), 273-299.

Cowley, S. J. (2006a). Distributed language: Biomechanics, functions and the origins of talk. In C. Lyon, C. L. Nehaniv & A. Cangelosi (Eds.), *Emergence of Communication and Language*, pp. 105-129. London: Springer.

Cowley, S. J. (2006b). Language and biosemiosis: A necessary unity? *Semiotica,* 162(1/4), 417-444.

Cowley, S. J. (in press). The cradle of language: making sense of bodily connexions. To appear, D. Moyal-Sharrock (Ed.), *Perspicuous presentations: Essays on Wittgenstein's philosophy of psychology*. London: MacMillan Palgrave.

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. (2004). Grounding signs of culture: Primary intersubjectivity in social semiosis. *Mind, Culture and Activity, 11*(2), 109-132.

Cowley, S. J & Kanda, H. (2005). Friendly machines: Interaction-oriented robots today and tomorrow. *Alternation, 12*(1), 79-106.

Cowley, S. J. & MacDorman, K. F. (2006). What baboons, babies and Tetris players tell us about interaction: A biosocial view of norm-based social learning. *Connection Science*, *18*(3).

Davidson, D. (1986). A nice derangement of epitaphs. In E. LePore (Ed.), *Truth and interpretation*, pp. 433-446. Oxford: Blackwell.

Dawkins, R. (1982). *The extended phenotype: The long reach of the gene*. Oxford: Oxford University Press.

Deacon, T. (1997). *The symbolic species: Co-evolution of language and the brain.* London: Norton.

Dennett, D. (1991). The origins of selves. *Cogito*, *3*, 163-173.

Donald, M. (1991). *Origins of the modern mind*. Cambridge, MA: Harvard University Press.

DeLanda, M. (1997). *A thousand years of nonlinear history*. New York: Zone Books.

Eerdmans, S. L., Prevignano, C. L. & Thibault, P. J. (2003). *Language and interaction*. Amsterdam: John Benjamins.

Halliday, M. A. C. & Matthiessen, C. (1999). *Construing experience through meaning: A language-based approach to cognition.* London: Cassell.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335-346.

Harris, Z. (1951). *Structural linguistics*. Chicago: Chicago University Press.

Hauser, M., Chomsky, N. & Fitch, T. (2002). The faculty of language: What is it, who has it and how did it evolve? *Science*, *298*, 1569-1579.

Humphrey, N. (1976). The social function of intellect. In P. P. G. Bateson & R. Hinde (Eds.), *Growing points in ethology*, pp. 303-317. Cambridge: Cambridge University Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Karmiloff K. & Karmiloff-Smith A. (2001). *Pathways to language: From fetus to adolescent*. Cambridge, MA: Harvard University Press.

Kaye, K. (1982). *The mental and social life of babies*: *How parents create persons.* London: Methuen.

Kirsh, D. & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*, 513-549.

Kravchenko, A. (in press). Essential properties of language or why language is not a code. To appear, *Language Sciences*.

Kugiumutzakis, G. (1999). Genesis and development of infant mimesis in early imitation of facial and vocal models. In J. Nadel & B. Butterworth (Eds.), *Imitation in infancy*, pp. 36-59. Cambridge: Cambridge University Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Laland, K. N., Odling-Smee, F. J. & Feldman, M. W. (2000). Niche construction, biological evolution and cultural change. *Behavioral and Brain Sciences, 23*(1), 131-175.

Legerstee, M. (2005). *Infants' sense of people: Precursors to a theory of mind*. Cambridge: Cambridge University Press.

Love, N. (2004). Cognition and the language myth. *Language Sciences*, *26*(6), 525-544.

Love, N. (in press). Are languages digital codes? To appear, *Language Sciences*.

MacDorman, K. F. (1999). Grounding symbols through sensorimotor integration. *Journal of the Robotics Society of Japan, 17*(1), 20-24.

MacDorman, K. F. & Ishiguro, H. (2006a). Toward social mechanisms of android science: A CogSci 2005 workshop. *Interaction Studies, 7*(2), 289–296.

MacDorman, K. F. & Ishiguro, H. (2006b). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies, 7*(3), 297-337.

MacDorman, K. F. & Cowley, S. J. (2006). Long-term relationships as a benchmark for robot personhood. *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication.* September 6-9, 2006. University of Hertfordshire, Hatfield, UK.

Martinet, A. (Ed.). (1960). *Elements of general linguistics*. Chicago: University of Chicago Press.

Matthews, P. H. (1967). Review of Chomsky, N. *Aspects of the theory of syntax*. *Journal of Linguistics 3*(1), 119-152.

Matthews, P. H. (2001). *A short history of structural linguistics.* Cambridge: Cambridge University Press.

Meltzoff, A. N. & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*, pp. 75-78.

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum.

Pattee, H. H. (2001). The physics of symbols: Bridging the epistemic cut. *Biosystems*, *60,* 5-21.

Paul, H. (1891). *Principles of the history of language* (trans. H. A. Strong). London: Longmans Green.

Pepperberg, I. M. (1999). *The Alex studies: Cognitive and communicative abilities of grey parrots*. Cambridge, MA: Harvard University Press.

Pinker, S. (1994). *The language instinct: The new science of language and mind*. London: Penguin.

Port, R. F. & Leary, A. P. (2005). Against formal phonology. *Language*, *81(4)*, 927-964.

Ross, D. & Dumouchel, P. (2004). Emotions as strategic signals. *Rationality and Society, 16*(3)*,* 251-286.

Ross, D. (in press). *H. Sapiens* as ecologically special: What does language contribute? To appear, *Language Sciences*.

Saussure, F. de (1916/1983). *Cours de linguistique générale*. [English translation by R. Harris.] London: Duckworth.

Savage-Rumbaugh, S., Shanker, S., & Taylor, T. J. (1998). *Apes, language and the human mind*. Oxford: Oxford University Press.

Seabra Lopes, L. & Connell, J. H. (2001). Semisentient robots: Routes to integrated intelligence. *IEEE Intelligent Systems*, *16*(5), 10-14.

Seabra Lopes, L. & Chauhun, A. (2007). How many words can my robot learn? An approach and experiments with one-class learning. *Interaction Studies, vol*(issue), page-page.

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*, 417-424.

Spurrett, D. & Cowley, S. J. (2004). How to do things without words: infants, utterance-activity and distributed cognition. *Language Sciences*, *26*(5), 443-466.

Steels, L. & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language. A case study for colour. *Behavioral and Brain Sciences, 28*(4), 469-489.

Stern, D. (1977). *The first relationship: Infant and mother*. London: Fontana.

Striano, T. & Rochat, P. (2000). Emergence of selective social referencing in infancy. *Infancy*, *1*(2), 253-264.

Taddeo, M. & Floridi, L. (2005). Solving the symbol grounding problem: A critical review of fifteen years of research. *Journal of Experimental and Theoretical Artificial Intelligence*. *17*(4), 419-445.

Taylor, T. J. (1997). *Theorizing language.* Oxford: Pergamon Press.

Thibault, P. J. (2003). Contextualization and social meaning-making practices. In S. Eerdmans, C. Prevignano & P. J. Thibault (Eds.), *Discussing John J. Gumperz,* pp. 41-61. Amsterdam: John Benjamins.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition* Cambridge MA: Harvard University Press.

Tønsberg, G H. & Hauge, T. S. (1996). The musical nature of prelinguistic interaction. The temporal structure and organization in co-created interaction with congenital deaf-blinds. *Nordic Journal of Music Therapy*, *5*(2), 63-75.

Trevarthen, C. (1979). Communication and co-operation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before Speech*, pp. 321-347. Cambridge: Cambridge University Press.

Trevarthen, C. (1988). Universal co-operative motives: How infants begin to know the language and culture of their parents. In G. Jahoda & I. M. Lewis (Eds.), *Acquiring culture: Cross-cultural studies in child development,* pp. 37-89. London: Croom Helm.

Trevarthen, C. (1998). The concept and foundations of infant intersubjectivity. In S. Bråten (Ed.), *Intersubjective Communication in Early Ontogeny.* Cambridge: Cambridge University Press, pp. 15-46.

Trevarthen, C. & Aitken, K. J. (2001). Infant intersubjectivity: Research, theory and clinical applications. *Journal of Child Psychology and Psychiatry*, *42*(1), 3-48.

Trevarthen, C. & Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In A. Lock (Ed.), *Action, Gesture, and Symbol,* pp. 183-229. New York: Academic Press.

Turing, A. (1950). Computing machinery and intelligence. *Mind, 49,* 433-460.

Viger, C. (2007). The acquired language of thought hypothesis: A theory of symbol grounding. *Interaction Studies, vol*(issue), page-page.

Vogt, P. The physical symbol grounding problem. *Cognitive Systems Research Journal, 3*(3), 429-457.

Vygotsky, L. (1986). *Thought and Language*. [Trans. and Ed. A. Kozulin]. Cambridge, MA: MIT Press.

Wittgenstein, L. (1953). *Philosophical Investigations*. [Trans G. Anscombe]. Oxford: Blackwell.

Worgan, S. & Damper, R. (2007). Grounding symbols in the physics of speech communication. *Interaction Studies, vol*(issue), page-page.

Zlatev, J. (2002). Mimesis: The missing link between signals and symbols in phylogeny and ontogeny? In A. Pajunen (Ed.), *Mimesis, Sign and Language Evolution*, Publications in General Linguistics 3, University of Turku, Finland.

Author's Address

Dr Stephen Cowley

School of Psychology

University of Hertfordshire

College Lane

Hatfield

Hertfordshire AL10 9AB UK

s.j.cowley@herts.ac.uk

About the author

Stephen Cowley received a BA in English from Cambridge in 1978, a MA in Linguistics from the University of Leeds in 1985, and a PhD in Linguistics from Cambridge in 1993. He is Senior Lecturer in Developmental Psychology at the University of Hertfordshire, United Kingdom and Honorary Research Fellow at the University of KwaZulu-Natal, Durban, South Africa. Previously, he has lectured at the University of KwaZulu-Natal in the Department of Linguistics (1996–1999) and Psychology (2000-2003) and, from 2003 to 2004, in Social and Developmental Psychology at the University of Bradford. His empirical work focuses on cognitive development and interpersonal dynamics as revealed in the study of caregiver-infant interactions.

Notes

* Among the many people who deserve thanks, I would like to give special thanks Karl F. MacDorman who got me hooked on symbol grounding. Since then, I have to thank David Spurrett who kept the interest alive, Angelo Cangelosi who provided a way of linking up with a band of enthusiasts, and Tony Belpaeme who did the tough part of organizing ESG2006. Equally, I am very grateful to the participants who made the event worthwhile and the three anonymous referees who did their best to help me turn a rather unruly manuscript into a focused research paper.

[1] This is the German equivalent of *horse*. I have chosen it to signal interest in grounding symbols into 'speech sounds' (or linguistic form). This may be harder than grounding them into *referents*.

[2] Digital signals impose semantic structure on the world (Ross, in press).

[3] Not only are code models incompatible with empirical facts (Love, 2004, in press), but they violate autopoietic principles (Kravchenko, 2004) and, for Ross (in press), presuppose an individualistic intelligence that is incompatible with natural selection.

[4] Chomsky made a similar move by focusing on a *narrow* language faculty (Hauser, Chomsky and Fitch, 2002). As on the distributed view, this gives biomechanics a major role in the events of talk.

[5] All behaviour is based in neural and bodily microdynamics. Their relevance to language is well illustrated by debate about when, to *some* ears, utterances of "gravy" sounded *unfriendly* (see, Eerdmans, Prevignano and Thibault, 2003). While some appeal to *contextualization cues*, Thibault (2003) treats this as showing the indexicality of language, and Cowley (2006b) claims that voice dynamics shape the feeling of what happens. The word *gravy* is just part of +/- 300ms of speech.

[6] Paul Vogt picks this out as the remaining "non trivial" problem for his models. He puts it that there is currently no way of distinguishing utterances as forms from utterances as physical objects. On the view presented here, the baby depends exclusively on physical (not formal) properties of expression..

[7] Trevarthen and Aitken (2001) give an overview of the neural systems used in intrinsic motive formation.

[8] On a distributed view of language, children represent verbal patterns (off-line) only in talk about what is absent (typically, after age 3). Before then, circumstances prompt them to vocalise. Rather than using linguistic forms, they allow such patterns to constrain expression.

[9] Those interested in modelling have given surprisingly little attention to social norms. Elsewhere Cowley and MacDorman (2006) argue both that the interaction order is crucial to humans and that, to deal with this, natural and social norms need to be brought into a single conceptual framework.

[10] Following Kirsh and Maglio (1994), these "epistemic actions" contrast with actions that bring the agent closer to its goal.

[11] Kravchenko (in press) notes that Peirce shares this view.