Embodiment in Evolution and Culture

Edited by

Gregor Etzelmüller and Christian Tewes

Mohr Siebeck

Gregor Etzelmüller, born 1971; Professor for Systematic Theology at Osnabrück University and Principal Investigator of the Heidelberg Marsilius Project "Embodiment as Paradigm for an Evolutionary Cultural Anthropology".

Christian Tewes, born 1972; adjunct Professor (Privatdozent) for Philosophy at the University of Jena and Principal Investigator of the Heidelberg Marsilius Project "Embodiment as Paradigm for an Evolutionary Cultural Anthropology".

ISBN 978-3-16-154736-2

Die Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliographie; detailed bibliographic data are available on the Internet at *http://dnb.dnb.de*.

The book was typeset by Laupp & Göbel in Gomaringen using Garamond typeface, printed by Laupp & Göbel in Gomaringen on non-aging paper and bound by Buchbinderei Nädele in Nehren.

Printed in Germany.

Table of Contents

Acknowledgements V

Gregor Etzelmüller/Christian Tewes
Introduction 1

1. Philosophical Concepts and Perspectives of Embodiment

Christian Tewes
Introduction 13

Mog Stapleton
Leaky Levels and the Case for Proper Embodiment 17

Christian Tewes
Embodied Habitual Memory Formation: Enacted or Extended? 31

Karim Zahidi/Erik Myin
Radically Enactive Numerical Cognition 57

Christian Spahn
Beyond Dualism? The Implications of Evolutionary Theory for an Anthropological Determination of Human Being 73

2. The Embodied Evolution of Symbolic Competence

Magnus Schlette
Introduction 99

Thomas Fuchs
The Embodied Development of Language 107

Terrence Deacon
On Human (Symbolic) Nature: How the Word Became Flesh 129

Jordan Zlatev
Preconditions in Human Embodiment for the Evolution of Symbolic Communication 151

Matthias Jung
Stages of Embodied Articulation 175

3. Embodiment as a Bridging Concept for Evolutionary and Historical Anthropology

Alexander Massmann
Introduction 193

Gregor Etzelmüller
The Lived Body as the Tipping Point Between an Evolutionary and a Historical Anthropology 205

Eve-Marie Engels
The Roots of Human Morals and Culture in Pre-Human Sympathy. Charles Darwin's Natural and Cultural History of Morals 227

Christoph Wulf
The Creation of Body Knowledge in Mimetic Processes 249

Annette Weissenrieder
"It Proceeded from the Entrance of a Demon into the Man". Epileptic Seizures in Ancient Medical Texts and the New Testament . . . 265

4. The Mutual Intertwinement of Nature and Culture

Miriam Haidle
Introduction 285

Lambros Malafouris
On Human Becoming and Incompleteness: A Material Engagement Approach to the Study of Embodiment in Evolution and Culture 289

Duilio Garofoli
Metaplasticit-*ies*: Material Engagement Meets Mutational Enhancement 307

Shaun Gallagher/Tailer G. Ransom
Artifacting Minds: Material Engagement Theory and Joint Action 337

Wolfgang Welsch
Bodily Changes during the Protocultural Period and Their Ongoing Impact on Culture 353

Contributors 365

Index of Persons 369

Index of Subjects 377

The Embodied Development of Language

Thomas Fuchs

Abstract: The concepts of language prevalent in cultural and cognitive sciences regard it as a complex mental symbol system which is acquired mainly through maturation of suitable cognitive modules. In contrast, from an embodied and enactive point of view there is no fundamental separation between sensorimotor and symbolic interactions of an agent with its environment. The paper first presents arguments for an embodied basis of language production and comprehension, in particular results from cognitive neuroscience which link language processing to motor areas in the brain. The acquisition of language is then conceived as resulting from embodied interactions with others, starting from expressive or interbodily resonance, then proceeding to iconic gestures and finally leading to symbolic modes of communication. This development is essentially based on understanding others as intentional agents, which in turn is enabled by grasping their intentions as embodied in expressive, goal-directed, and pointing gestures in the context of shared practices.

Introduction

Since antiquity man has been primarily distinguished as the being that has *language* – the *zoon logon echon*, as Aristotle defines it, and later as the *animal rationalis.* According to this definition, on the one hand, humans are living beings like animals (*animalia)*, and yet on the other hand are fundamentally different from them due to language and reason. Through these capacities alone, they achieve culture, art, science and technology. They are similar to their animal relations with regard to bodily needs, drives and affections; however, reasoned speech distinguishes them ahead of all other earthly creatures. Thus, *Homo sapiens* is an inherently ambivalent centaur being, a hybrid of animality and rationality, an *animal rationale*.

It may still be attributable to this traditional view of anthropology that for a long time both the cultural as well as the cognitive neurosciences only treated language as a disembodied mental symbol system. Starting with Fodor's "Language of Thought" (1975), words were conceived as producing images or symbols inside the head of the speaker or listener, whose brain would use them to construct a representation of the state of affairs "out there" (Fodor 1998; Pylyshyn 1984). The fact that language originates from speaking with one another, where this primarily represents a bodily movement of *expression* and a joint speech *action*, that is to say in brief – the *bodily performance* of speech was only acknowledged as an accidental attribute, which seemed to have no effects on its structure and the implied contents.

Only recent decades of infant research and evolutionary anthropology have shown the wealth of communication and dialogue that already unfolds in the human individual before learning language (Trevarthen 1979, 2009; Stern 1985; Tomasello 2008). Bodily communication or body language, as we also call it, is mainly conveyed through facial expression and gestures, through the intonation of the voice and ultimately through the body's whole posture. As Darwin ([1872] 1998) already observed, this expressive communication in humans manifests a differentiation and diversity that is unique in the animal kingdom. However, it is also the foundation on which verbal-symbolic forms of communication may initially develop at all during early childhood. For as we shall see later, language acquisition crucially presupposes that children develop an understanding for the intentions of others; and at first these intentions are only accessible to them as embodied, namely as visible, expressive, goal-oriented and pointing movements, whose meaning is exposed in the context of practical bodily interaction.

In what follows, I will proceed from *an embodied and enactive view* on language and its development (Varela, Thompson, and Rosch 1991; Glenberg and Robertson 2000; Ziemke 2002; Zlatev 2007). I will argue for the following theses:

(1) Language is not a representation of the world inside the head, but a form of *embodied intersubjectivity*: The meaning and function of words and sentences is derived from our bodily experience of interacting with the world, which we share in principle with others, and which is evoked both in ourselves and in others by our verbal utterances. This is reflected in recent research on the involvement of sensorimotor brain areas in language processing.
(2) The acquisition of language in infancy is not achieved through an abstract attribution of symbols to references, but through the infant's participation in shared intentional practices of interacting with the world. Only as embedded in an interactive "we-intentionality", can words be learnt and gain their meaning.

In both ways, language thus depends on intercorporeality (*intercorporéité*, Merleau-Ponty 1960), that means, on a sphere of reciprocal bodily understanding and interaction, from which words first draw their references and meanings. Following on from these practical interactions, the infant's brain is also influenced and structured by language: the brain only becomes an organ of the symbolic mind through social interactions (Fuchs 2010, 2011).

In the first part of my paper, I will argue for the embodied nature of language, including the anchoring of language in the brain. In the second part, I will give an account of the embodied development of language in early childhood.

1. Language, Embodiment, and the Brain

The Body as the Medium of Language

In their seminal book "Metaphors we live by", Lakoff and Johnson (1980) have first emphasized the bodily basis of language. They described over 50 systematic schemes of body-related verbal metaphors: basic bodily experiences like those of in and out, up and down, front and back, warm and cold, fast and slow, near and far, etc., cover a wide range of applications in all dimensions of language. They become the basic schemes of conceptual development and imagery, and what we use to call metaphorical or figurative meanings are in fact derived from our bodily experience which is subliminally present and effective even in the seemingly most abstract discourse (see Johnson 1987).

The connection of language and the body has also been examined over the past two decades from the perspective of *embodied and enactive cognition*. This paradigm is based on the assumption that there is no strict separation of "lower" and "higher" cognitive functions, that is, between perception and movement on the one side and thought and language on the other. All forms of cognition are fundamentally considered as a form of interaction between an organism and its environment (Varela, Thompson, and Rosch 1991), which means that there is no abstract level of the mind as a computational symbol system. Instead, motor, sensory, and cognitive functions are always intermodally linked. This has also led to an embodied view of language as involving bodily systems of movement, posture, kinesthesia and proprioception, both in language production and comprehension (e. g. Glenberg and Robertson 2000; Ziemke 2002; Zwaan et al. 2004; Barsalou 2008; Cuffari, Di Paolo, and De Jaegher 2015).

Let us take an example: If we listen to a simple sentence such as "*the book lies on the table*", its meaning is constituted for us by a connection of several components:

(a) the evocation of two objects in our awareness, which does not only include their visual imagination, but also their affordances for our bodily action, for example, as something to grasp, to open and to read (the book), something solid to sit at or to lay things on (the table), etc.;
(b) our operative (motor, postural) bodily intentionality which lets us implicitly grasp the state of "lying", namely as being stretched out flat, wholly supported by the ground;[1]

[1] This involvement of our body in the meaning becomes even more obvious if we think of the difference the German language makes between "lying" and "standing" objects: "Das Buch *liegt* auf dem Tisch" (the book "lies" on the table), but "die Tasse *steht* auf dem Tisch" (the cup "stands" on the table). This usage of the verbs mirrors the different postural imitations that are invoked in our body when looking at a flat versus an upright object.

(c) a spatial relation which we know from your own bodily postures or actions (lying "on" something, being placed "next to", etc.);
(d) a temporal relation of simultaneity to our present experience ("lies");
(e) a syntactical structure which generally combines a subject and a predicate in the same way as we experience *ourselves as doing something* ("the book lies", "the tree stands", "the bell rings", etc.).[2]

So what we implicitly understand when listening to the sentence above would have the unfolded meaning of "the *thing-I-could-take-and-read* is now *lying-like-I-would* on the *thing-I-could-sit-at*", or similar. A sentence thus combines affordance-based terms into patterns of action and relation, or in other words, *the syntax in a sense imitates the operative intentionality of our body*.[3] In its basic grammatical structure, a sentence expresses a subject acting on an object in a way that we could on principle perform ourselves; through this very structure, the sentence *enacts its meaning* and thus enables an embodied understanding, or to use an enactivist term, embodied *sense-making* (Weick 1995; De Jaegher & Di Paolo 2007).

To this, we have to add the person speaking the sentence and her apparent intention in the interactive context, turning the utterance "the book lies on the table" either into an informative answer (there it is!), an implicit request (could you hand it over?), a philosophical example (let's take the following sentence …), or whatsoever. Understanding another thus involves participating in her intentional attitude towards the situation.[4] Moreover, listening to her also involves

[2] The fundamental structure of a sentence (subject – predicate – object) implies an agent performing some kind of operation on an object, which is precisely the basic structure of our embodied relation to the world. Of course, there are many variations – the verb may be intransitive or signify a state rather than an action – but this does not change the fact that a sentence expresses what could on principle be our own experience.

[3] One might object that all these affordances and bodily conditions are far too complex to be present in the immediate understanding of the sentence. As we will see, however, there is now a lot of neurobiological evidence showing that this indeed the case (see below). But apart from that, the question is how one could ever come to understand the meaning of *lying* at all, if not by "what I know from my own lying", even if this embodied knowledge is only activated in the most remote way when hearing the word later on. For otherwise it would be very difficult and circuitous to explain what *lying* actually means, for example, "the spatial relation of an object being in close contact with another object underneath, touching it with its most extended side, whereas its smaller sides remain free and upright." And even then, we would run straightaway into the *symbol grounding problem* (Harnard 1990), for what the symbols "spatial", "contact", "touching", "cover", etc. in that definition mean could only be explained by even more complex definitions, and so on *ad infinitum.* Language cannot be a free-floating system of symbolic references – it must ultimately be grounded in embodied experience. This experience is primarily given as a *knowing how* based on bodily dispositions and habits, not as a *knowing that* represented in a propositional format (Fuchs 2016a).

[4] Usually, this does not require any explicit perspective-taking or mentalizing ("theory of mind"): we do not distinguish between an interlocutor's mental state and his utterances, as if the former would have to qualify the latter, but we understand his words as just what they mean in relation to the shared situation. The intention is inherent in the verbal expression itself. Only in cases of ambiguity or doubt, this unity of intention and utterance may be dissolved, and we apply explicit cognitive procedures of perspective-taking or inference ("what did he mean by that?", "what is he up to?", etc.).

a tendency of subvocalizing her utterances. This becomes obvious for example when listening to a conversational partner who appears to hesitate or to be at loss for the right words, and without hesitation one supplies the missing words, completing the utterance of the speaker. For the speaker in turn, the attentive listener serves as a stimulus for his own speech, as Kleist ([1805] 1951, 43) has famously described in his essay *On the gradual construction of thoughts during speech*: "The other person's face is a curious source of inspiration for a person who speaks. A single glance which indicates that a half-expressed thought is already understood, bestows on us the other half of the formulation."[5] Language production as well as comprehension may thus be described as a special kind of participatory sense-making (De Jaegher and Di Paolo 2007), namely as the *co-enactment* of a sense that is always in the making, through embodied protentions or co-anticipations of both speaker and listener.

If we take all this together, we can assume a *prima facie* evidence that

(a) language is not a free-floating, abstract symbol system, but a network of meanings evoking a certain way of embodied being-towards-the-world (*être-au-monde*, Merleau-Ponty) or acting-towards-the-world;
(b) language production and comprehension are crucially based on embodied and enactive cognition, including the situated verbal interaction itself. That means, "words are patterns available for enacting certain forms of sense-making" (Cuffari, Di Paolo, and De Jaegher 2015), both in speaking and in understanding.

One could now argue that this bodily and operational basis of meaning and grammar does not apply to higher levels of abstraction: there seems to be no enactive account of abstract words like "conclusion", "peace" or "right", etc. However, a closer look reveals that even the meaning of abstract or metaphorical terms is ultimately based on bodily experience (see also Irwin 2015). Let us look at some examples:

– The noun "right" (or German *Recht*) is derived from the Indo-European roots *reg-* ("to move in a straight line, to straighten, to direct") and *regtós* ("straight, upright").[6] Thus, it is related to a bodily operation which implies

[5] It is worthwhile to follow Kleist's description in detail: "Often I sit at my desk, poring over documents and trying to discover the point of view from which some complicated controversy might be judged. … But, lo and behold, if I mention it to my sister, who is sitting behind me and working, I discover facts which whole hours of brooding, perhaps, would not have revealed. … For since I always have some obscure preconception, distantly connected in some way with whatever I am looking for, I have only to begin boldly, and the mind, obliged to find an end for this beginning, transforms my confused concept as I speak into thoughts that are perfectly clear, so that, to my surprise, the end of the sentence coincides with the desired knowledge. … During this process nothing is more helpful to me than a sudden movement on my sister's part, as if she were about to interrupt me; for my mind, already tense, becomes even more excited by this attempt to deprive it of the speech of which it enjoys the possession and, like a great general in an awkward position, reaches an even higher tension and increases in capacity." (Kleist [1805] 1951, 42 ff.)

[6] Cf. also Greek *orektos* (stretched out, upright) or Latin *rectus* (straight, right). See Kluge (1989) and http://www.etymonline.com.

an upright posture or gait and an experience of balance. This refers to the moral sphere as well: being a "righteous", honest or courageous person means an inner or moral attitude which is embodied in a corresponding posture of standing or walking upright. Similarly, the meaning of "justice" or "equity" (German *Gerechtigkeit*) is grounded on the experience of bodily equilibrium achieved in the upright position (as represented also in the balanced scales of Justitia).

- The words "concession" and "concede" are derived from the Latin *cedere* which means to withdraw, to give way. Thus, if I concede a right or a claim to someone, I withdraw, however slightly, from my primary bodily stance which may also be expressed by a conceding gesture of my arm.
- Apart from etymology, embodiment research may also support the bodily basis of metaphorical terms, as for example the connection between *guilt* and *impurity*, or cleansing, respectively. Pilate washed his hands and thus claimed to be innocent of Jesus' death, and Lady Macbeth develops a washing obsession after the murder of King Duncan. Recent research has now shown that cleansing can indeed wash away or alleviate feelings of guilt (Meier et al. 2012, Lee and Schwarz 2011, Zhong and Liljenquist 2006) and have a mildness influence on one's moral judgment (Schnall, Benton, and Harvey 2008).
- When we speak of a "warm welcome", we do so because we actually feel bodily warmth in this situation – the social atmosphere is felt as bodily sensation. Correspondingly, Zhong and Leonardelli (2008) found that test subjects, after having been exposed to a situation of social exclusion or ostracizing, estimated the room temperature to be colder than before. Moreover, Bargh and Shalev (2012) found that persons who experience social loneliness show an increased tendency to take warm baths or showers.

Generalizing such considerations and results, one can describe language as a system of interrelated terms which refer to all kinds of embodied operations and experiences, and which in their syntactical combination imitate our bodily interactions with the world. Even the most abstract terms are ultimately derived from some primary form of operation or interaction: Take "abstraction" as drawing away (from Latin *abs-trahere*), "detection" as pulling away a cover (*de-tegere*), "enlightenment" as sheding a light on something to become visible, or "negation" as an action or resistance against some kind of intrusion (for example, a rejecting gesture of one's hands or a shaking of the head to avoid intake).[7]

[7] Could this thesis even be extended to include abstract systems such as mathematical or logical structures and operations such as $\sqrt[3]{27} = 3$, syllogisms or similar? It seems that from a certain degree of abstraction, such systems can still be comprehended or applied, but do no longer allow for any imagination based on sensorimotor experience. However, it soon becomes clear that even here, the abstract terms and operations are initially derived from experiences of bodily action in the way Piaget ([1936] 1952) has already described it (although he assumed that abstract thought disconnects from the level of primary sensorimotor or preconceptual thinking). Thus, addition, subtraction, multiplication, or division are mental operations which are only acquired initially by performing the concrete operations in an ostensive way (e. g. supported by

Neurobiological Findings

In the last two decades, the embodiment of language has been increasingly confirmed by findings from *neuroscience*, which show that language processing in the brain is functionally connected to sensorimotor systems. Thus, if one listens to words, the same sensorimotor areas are activated as for the practical engagement with the objects that the words refer to, or in other words, language comprehension is crucially based on action-perception circuits in the brain (Gallese 2008; Pulvermüller and Fadiga 2010; Jirak et al. 2010). Let us look at some examples:

- Listening to the words "grasp", "go" or "shout" activates, alongside the receptive language areas, also the motor centers for the corresponding actions (Buccino et al. 2005; Jirak et al. 2010). There is even strong evidence for a *somatotopy* of language, that means a differential activation of motor centers according to the limb or action involved in the sentence one listens to: Pulvermüller (2005) identified specific fMRI-activity patterns in the pre-motor cortex for consonant verbs that refer to mouth, arm or leg movements, such as '*lick*', '*pick*' and '*kick*'. In each case, the premotor cortex is differentially engaged in a topographical bodily pattern.
- When listening to verbs referring to hand movements (give, take, point, etc.) right-handed people show an activation of the left pre-motor cortex, left-handed people an activation of the right (Willems, Hagoort, and Casasanto 2010). This shows that the verbs are processed according to the actual bodily movement that one could perform. Moreover, it strongly suggests that they have already been learnt in this embodied way: "to give" meant originally "handing something over to mom with my right hand" (or left hand, in the other case).
- Words related to odours (for example, "cinnamon") or to sounds (for example, "telephone") cause particular activation in olfactory and auditory brain areas, respectively (Pulvermüller and Fadiga 2010). Thus, listening to the sentence "the alarm sounded and John jumped out of bed" will activate areas both in the auditory and motor cortex related to sounds and movements (Kaschak et al. 2006; Winter and Bergen 2012).
- Moreover, Glenberg and collaborators (2008) and Boulenger, Hauk, and Pulvermüller (2009) found that the *abstract usage* of verbs such as "to give" or "to grasp" (to give a reason, to grasp a notion) activates the motor system no less than the concrete usage. Granted, these results are still open for debate,

---

one's finger or other countable objects). Of course, the habitualization of these operations leads to their formalization which does no longer need (nor afford) operative imagery. However, even though a number such as 1,455,578 cannot be imagined in any sense, we still take it implicitly for granted that it is composed of as many steps of adding 1 + 1 + 1 …, and the same applies for all other kinds of mathematical operations – that is precisely why they are called "operations". The same could be shown for logical operations like conclusions (thus, the famous syllogism "All humans are mortal, Socrates is human, therefore Socrates is mortal" dips into a box in which all objects of a certain type have been put before and picks one out again).

and it may also be possible that the context of words influences the degree to which the motor regions are involved in their comprehension (Jirak et al. 2010).
- Generally, merely listening to speech also activates motor brain regions that are involved in speech production (Wilson et al. 2004, Pulvermüller et al. 2006). This corresponds to the tendency of subvocalization during listening to an interlocutor mentioned above.
- Finally, it emerged that areas which were thought to have purely verbal functions like the Broca and Wernicke area actually combine language and bodily movement with one another, specifically via the *mirror neuron system* (Binkofski and Buccino 2004; Gallese 2008). "Mirror" or sensorimotor neurons, originally found in the premotor cortex of macaque monkeys, generally link one's own motor action to the same action as perceived in conspecifics, enabling a sensorimotor or embodied social perception (e. g., observing someone reaching for a cup activates one's own motor system for the same reaching action, even if only subliminally). In humans, Broca's area has been found to be the core region of the mirror neuron system, and there is increasing evidence showing that this system is at least participating in the connection of verbal sounds and possible action (Aziz-Zadeh et al. 2006; Aziz-Zadeh and Damasio 2008; Jirak et al. 2010).

All these strands of research are still in flux and a final evaluation is not possible yet. Nevertheless, there is at least strong evidence for an enactive concept of language as being crucially based on bodily perception and action. A consequent question is: Does the body also play a constitutive role for the *acquisition of language*, which also means for the establishment of neural action-perception circuits that are necessary to speak and understand language? In the introduction, I have already proposed that language developes as a form of embodied intersubjectivity. I now state some reasons in greater detail, looking at the development from pre-verbal to verbal stages of intersubjectivity in early childhood.

2. The Embodied Development of Language

Primary Intersubjectivity

Infants are attuned from birth to social interactions, in particular by showing a heightened attention to faces and their expressions (Valenza et al. 1996; Turati et al. 2002). Research studies conducted during the last two decades have mostly found that they are also able to imitate adults' gestures like sticking out their tongue, opening their mouth, frowning, and others (Meltzoff and Moore 1977, 1989). This capacity for spontaneous imitation of others' expressions has been considered a crucial basis of early social development (Meltzoff and Brooks 2001, Meltzoff and Prinz 2002). However, recent research with larger samples and a wider range of gestures presented to the infants challenges these results, finding

no significant excess of matching over non-matching reactions (Oostenbroek et al. 2016). But even if it turns out that imitation is not an innate capacity, but develops in the course of mutual exchanges and matching reactions during the first months, it still functions as a major component of what Trevarthen (1979) has termed "primary intersubjectivity".

This stage is characterized by an increasing *emotional resonance* between infant and mother that develops via mutual bodily expressions and reactions. Usually, the mother intuitively answers the baby's signals and initiatives with suitable vocal and gestural reactions that stimulate further resonance. In the first months, mother and infant thus develop dynamic and synrhythmic "proto-conversations" (Trevarthen 2001, 2008), that is, fine-tuned sequences of alternating expressions with imitative utterances, smiles and gestures just like a conversation – the later verbal dialogue is already outlined here. Mothers and fathers intuitively use simplified, prototypical behavioral forms (welcome reaction, eye contact, musical utterances or "motherese", exaggerated facial expressions, among others) that correspond to the child's "musical repertoire" and preference for expressiveness (Papoušek and Papoušek 1987, 1995; Malloch 1999).

This early intensive dialogue is especially influenced by *musical* expressive qualities, by the rhythm and dynamics of facial, vocal, and gestural interaction that express changes of emotion and mood. They may best be described in qualities such as "crescendo", "decrescendo", flowing, frisking, smooth, explosive, etc., which Daniel Stern (1985) termed "vitality contours" or "vitality affects." For example, a sharply rising pitch contour in maternal vocalization alerts the infant, whereas the pitch is low and continuous in comforting or soothing (Fernald 1992, Papoušek 1994). Being the major bridge of emotional exchange, these expressive qualities lead to the mutual "affect attunement" of parent and infant that Stern highlighted. "Even in early weeks, infants learn little rituals of musicality, in vocal games, in simple rhyming songs, sharing with skill and affectionate good humour their recursive events … babies are alert to the pulse and subtle harmonies of a mother's speech, turning to tones of sympathy, or withdrawing from their absence" (Trevarthen 2008, 18, 21).[8] In the course of this preverbal communication, the child increasingly learns to connect the mother's or father's emotional expression with typical recurring situations and thus to distinguish its different meanings. The child also learns that his own reactions motivate the caregiver to specific behavior, and thereby develops interactive expectations. All this conveys to him the basic feeling of living with others in a shared world, of being perceived by them and being connected with them – a central precondition for the steps that now follow.

[8] The baby's particular sensitivity to the lived synchrony of interaction was impressively demonstrated by Murray and Trevarthen (1985) who designed a Double Television set-up that enabled replay of the mother's affectionate and responsive talk with the baby. When a happy minute of the mother's live communication was later replayed to the baby (thus showing the same expressive qualities but lacking synchrony and responsiveness), the baby soon became distressed and turned away.

Secondary Intersubjectivity

(a) Joint Attention and the Pointing Gesture

On the next level of secondary intersubjectivity, the phenomenon of "joint attention", which manifests itself from about the age of 9 months, signifies a key step towards symbolic communication (Trevarthen and Hubley 1978; Tomasello 2002; Bråten and Trevarthen 2007). At this age, babies begin along with adults to turn their attention to objects, in particular by following their *pointing gestures.* Soon the babies also proceed to steer the adults' attention to things through pointing themselves, and in doing so cast each other quick glances to reassure themselves of their attention. In an illuminating experiment by Tomasello and his group, infants aged about 12 months observed how one adult made a hole in a sheet of paper and filed it away in a clip folder. The adult now left the room and another adult entered, took the folder and placed it in a clearly visible cupboard, which he then locked. He left the room, the first adult re-entered and looked around, visibly searching for something, with a sheet of paper in his hand. In most cases, the infants looked attentively at the adults and then pointed to the cupboard (Liszkowski et al. 2006).

How can we interpret this experiment? Obviously, the infants recognized the adult's intention, only due to his previous action and now his questioning expression. Intentions are therefore not only something internal or mental, but they are also perceptible in the goal-oriented bodily actions of others and obtain their meaning from the context of the joint situation. There is no need first for a "Theory of Mind" (ToM) or some kind of inference or mind-reading in order to directly understand others' intentions in a practical context – after all, the usual time of acquiring a sophisticated knowledge of other minds (ToM) is not before the age of 4 years. Considered more closely, what does pointing imply?

Pointing first involves the mutual relation to a third entity that is seen by both partners, being aware that the other is also doing so. Hence, we are no longer concerned with the primary dyadic, but with a *triadic* situation comprising the infant, the adult and the mutually intended object or goal of an action (Tomasello 2002). The joint attention, which is visible in the parallel axes of the child's and adult's gazes, manifests a specifically human form of communication, namely conveying a message about a joint, external reference point. Here lies a fundamental limit to the mental capacities of other primates that cannot develop joint attention (Fuchs 2013). Even though great apes may become capable of so-called imperative pointing ("give me this!") when raised in human environments, there is no declarative or cooperative meaning attached to it (Gómez 2007). In contrast, as we saw in the above study, the infants also attempted to help the adult by pointing to the object being searched for. This communicative and cooperative attitude has been particularly highlighted by Tomasello and his group as a crucial difference from proto-pointing gestures shown by great apes (Tomasello et al. 2005; Tomasello and Carpenter 2007): only through this sharing of intentions,

an actual "we-intentionality" is created ("look at this!", "now we are looking at this object together").

Pointing is a gesture that only makes sense in an intersubjective context: it "indicates" the object with the index finger, instead of grasping it. The other person must understand this meaning, i. e. follow the direction of the finger into empty space until arriving at the object as its goal. The pointing gesture is the origin of mutually shared meanings and thus a precursor of the *sign* – the entity, which stands for something different, and represents it (Fuchs and De Jaegher 2009). Etymology also refers to the genetic link of pointing, sign and later speaking: in German, "*zeigen*" (to point) und "*Zeichen*" (sign) have the same Indo-Germanic root <*deik*>. This root also occurs in the Greek "*deíknymi*" (to point, to show) and "*dáktylos*" (finger), and also in the Latin "*dicere*" (to show, to speak) and "*digitus*" (finger). The same connection becomes manifest in German "*deuten*" (point) and "*bedeuten*" (signify) (see Kluge 1989, 807).

The pointing gesture is a grounding experience in still another way. Infants experience in this instance that other people also have a direction of attention that they can personally influence. Even though we should not be led to assume a mentalistic understanding of others at this level, infants at least begin to understand that the world looks different in their parent's eyes, yet that they can communicate with them about it. They show them an object because they notice that the adult has not seen it yet, but could soon see it, as shown in the experiment of the folder in the cupboard. In other words, infants develop an initial understanding of another perspective with which they identify by a kind of *co-anticipation*, assuming that an object has a meaning for the adult. Such a fundamental new stage of intersubjectivity is manifested here that Tomasello also refers to the "9-month revolution" (Tomasello 2002).[9]

(b) Other Gestures

Apart from the pointing gesture there are also other communicative gestures that develop in the second year of life. In almost all cultures, for example, shaking one's head means "no". The origin of this movement can be observed in babies who move their head to one side to avoid an unpleasant stimulation or to refuse further breast-feeding (Spitz 1957). Presumably this evolved into a ritualization during the course of phylogeny. As the signal must be clear, it was carried out more noticeably, i. e. by more markedly and repeatedly turning the head. On

[9] It should be mentioned here that Tomasello's account of infant pointing goes far into a mentalistic understanding of others even at this stage (see for example Tomasello et al. 2007). As Gómes (2007) has argued, there is also a more parsimonious explanation which emphasizes (as I did above) the embodied intentionality of gestures in the infant's experiential field: "behaviors are directly perceived as intentional, that is, as being directed to things other than themselves … For example, understanding that gaze is directed to an object does not require attributing the mental experience of seeing the object – such directionality is directly attributed to gaze itself" (Gómez 2007, 730). Regarding intentional behaviour as field-related, one can even assume that an infant can "remember and predict the intentional availability of targets for others (e. g., whether they will or not be able to find an object hidden in their absence)" (l. c.).

the other hand, nodding one's head represents "yes" in most cultures. Lowering the head probably meant a sort of gesture of humility signifying: I bow to what you say; I agree (Eibl-Eibesfeld 1972). These gestures are acquired in the course of the 2nd year, with head shaking ("no") before nodding ("yes") (Kettner and Carpendale 2013).

Other gestures, which develop in the course of the second year of life, are of an iconic nature, i. e. they represent pantomime actions or recall something absent in the imagination: raising one's arms means "big", blowing means "too hot", panting represents a "dog", flapping one's arms suggests a "bird" etc. (Tomasello 2009, 159 f.). Thus, the early development of non-verbal communication is characterized by deiktic and iconic gestures which supports an embodied view of language acquisition, although from the 14th month or so the gestures and vocalizations of this 'protolanguage' are already accompanied by the acquisition of verbal speech.

(c) The Development of Language

In the final months of the first year the words adults use to label people, objects or actions attract the infant's attention and invite imitation. Speech acquisition occurs not purely cognitively, however, as though language were just a sign system to be learned abstractly. According to the social pragmatic approach (Bruner 1983; Nelson 1996; Tomasello 2000), language acquisition is scaffolded by situations of intercorporeality, shared attention, joint practice, and ostensive cuing. The conditions for this are:

(1) the child's participation in an interactive framework that is already pre-verbally developed, in other words, verbal interaction presumes intercorporeal exchange;
(2) joint attention to a third entity, and specifically in the practical context that the speech refers to – that is, the triadic situation;
(3) understanding the communicative intentions of others as being based on their goal-directed movements, pointing or expressive gestures.

Hence, social practice represents the reference point and at the same time the scaffolding context within which a symbolic language can be learned. In concrete terms, this means that the first words are connected with already comprehensible gestures, in particular, the pointing gesture. For example, the parents ostensibly look at or point to objects and name them ("Look! A ball!"). The child now must understand that the parent intends for her (the child) to share attention with her to some outside entity, or in other words, the communicative intention (Tomasello 2000). Of course, grasping the word as meaningful does not yet imply higher conceptual capacities, but rather a typification of proto-concepts according to similarities of shape and behavior ("balls" means "such round, rolling things"). In the sequence, this leads to a *reverse imitation*: Now the child uses the first words ("there!", "ball", etc.), often connected with a pointing gesture, to show the adult what she herself finds interesting and wants to share. The adult's

understanding of the verbal gesture then acts as a reinforcement which stabilizes the new gestural meaning.[10]

A crucial question is how cognitively demanding this early communication should be conceived. Tomasello explains it already in terms of Grice's (1989) complex theory of language and meaning: "This is what a linguistic symbol is. It is a noise (or other behavior) that two or more individuals use with one another to direct one another's attention and thereby to share attention – and they both know this is what they are doing" (Tomasello 2000, 405). This is already a high-level account of cognitive intentions, implying some kind of meta-perspective on the communication ("I know that you know what I mean"). It seems highly probable that this rather abstract level is only reached later on, whereas the early language use is based on situated and embodied interaction.

Thus, even if the verbal meanings can increasingly be detached from the concrete situation – at first, all of early speech acquisition is against the backdrop of interactive situations and short episodes: eating, washing, dressing, changing nappies, playing, building a tower out of blocks, feeding ducks, and so on. The child always first learns co-involvement with the relevant practical situation and to form mutual goals, and then he orders the speech, which he has heard, into this context (Bruner 1983). He learns the word "ball" when playing ball, the word "there" in association with the pointing gesture and the word "Ow!" in connection with an expression of pain etc. Children's perception of the environment is synchronized with the corresponding verbal expressions that denote it and with the adult's visible attention and intention. They only adopt a word for a new object when his or her attention is actually directed towards this object. If the adult is looking in another direction or the voice is coming from a tape, the child doesn't connect word and object (Tomasello 2000; Dittmann 2002, 43). The capacity for speech therefore only develops within social scaffolding through an intercorporeal practice that is oriented towards a shared environment.

In fact, the word is a *vocal gesture* and initially only complements the pointing gesture as a first sign. But the voice also separates the sign from the physical movement and transports it into the invisible, no longer localizable medium of sound (Fuchs 2010, 210). Thereby, the possibilities of referencing multiply, and ultimately the sound signs can even be detached from the concrete situation. They are capable of pointing to absent objects, for example to Mummy or Daddy when they are absent; they are even capable to pointing to "something like", that means to similar, general, or abstract objects. The gestural-iconic representation is then increasingly replaced by propositional speech, and the continued gestures accompanying verbal speech serve more visual aspects, for example, to illustrate forms, directions, and structures that are the topic of speech.

[10] Frequently, the interaction also selects wording from spontaneous sound production and the child's babbling, making them into meaningful signals: for example, when the child says "Mummy" or "Daddy", the parents presume her intention is to form these words and reinforce them accordingly. Recognizing the effect of her own sounds then leads the child to learn their "meaning".

Neurobiological Foundations

As we can see from this brief outline of speech acquisition, the body as the medium of all action and interaction plays a fundamental role in the process. How is this reflected in the *neuronal anchoring* of language?

Neuroplasticity is a crucial presupposition for language development; in the course of meaningful interactions with others, the brain also becomes the matrix of language. Two aspects are significant here. Firstly, EEG studies show that up to the 2nd year of life the earlier developing right half of the brain which is the dominant hemisphere for processing music also manifests stronger activation while listening to language than the left half (Patel 2003; McMullen and Saffran 2004). This corresponds to the enhanced role of musical elements, namely, of speech melody, intonation, and rhythm for the perception of the toddler (Trevarthen 1998). The more advanced the development of symbolic speech, the more areas in the left brain take over verbally relevant functions, in particular, the Wernicke and Broca center and other premotor areas as well as the basal ganglia. However, even at a later stage in life, recent results suggest that the neuronal resources for processing speech and music still heavily overlap, in particular, in the Broca region and its counterpart in the right-half of the brain (Koelsch 2005, Koelsch et al. 2005). This suggests that at least in infancy the brain does not process music and speech as separate domains, but rather processes speech as a particular form of music, indeed *that the musical capacities of humans represent a decisive precondition for speech acquisition.*[11]

Both music and language are organized temporally, with the relevant structures unfolding in time, as patterns and sequences of rhythm, emphasis, intonation, phrasing, and contour (McMullen and Saffran 2004).[12] This is in correspondence with the central role of melodious-rhythmic interaction, vitality contours and affective resonance in the early mother-child dyad, which was mentioned above: The musicality of the interaction may be regarded as prefiguring the temporal dynamics in which language may then unfold. The theory of early "Communicative Musicality" is supported by acoustic analyses of the measures of rhythm, quality and dynamics in the vocal interplay between infants and adults (Malloch 1999). Here, an emotional aspect of speech development is involved that is especially manifest in prosody. Accordingly, recent neuroimaging results indicate that responses to human vocal sounds are strongest in the right superior temporal area (Belin, Zatorre, and Ahad 2002), near areas that have been implicated in processing of musical pitch (McMullen and Saffran 2004). This lends

[11] The idea of singing being the ancestral origin of speech was first put forward by Giambattista Vico in his notion of "*Parlare cantando*" (cf. Trabant 1991).

[12] This correspondence of temporal structure has already been noted by Adam Smith in his essay *Of the imitative arts* ([1777] 1982): "Time and measure are to instrumental music what order and method are to discourse; they break it into proper parts and divisions, by which we are enabled both to remember better what has gone before, and frequently to foresee somewhat of what is to come after … the enjoyment of Music arises partly from memory and partly from foresight" (quoted after Trevarthen 2012, 259).

plausibility to accounts of musical and linguistic co-evolution that emphasize emotional communication through prosody as a primary root of both systems.

The second aspect is related to the embedding of speech acquisition in interactive contexts. Specialized systems are required for the neuronal connection of action, perception, and meaning through speech, and there is now plenty of evidence to suggest a crucial role for the sensory-motor system of the *mirror neurons*. The localization of Broca's region in the inferior pre-motor cortex (responsible for speech production, but also for hand and mouth movement) and its coincidence with the main areas of the mirror neuron system suggests that language originally represented an *interpersonal resonance system for action schemes*: via the communication of the mirror neuron system, the voice was able to call up the idea of the intended actions and objects in both speaker and listener.

As mentioned above, the mirror neuron system (MNS) is activated both when observing a conspecific reach for or grasp an object and when imagining oneself reaching or grasping without actually moving one's hand. Thus, the system leads to matching an observed movement to the internally generated enactment of the same movement in the observer.[13] Speculating on a connection to the evolution of language, Rizzolatti and Arbib (1998) have first assumed that the MNS also enables intentional meaning to be assigned to another's vocal gesture. The connection could be spelled out as follows (see in particular Gallese 2008; Jirak et al. 2010):

Mirror neurons also react to suggested *goal-directed* movements, i. e. they are activated when the hand of another individual reaches for an object that was already visible earlier, yet is now out of sight (Umiltá et al. 2001). This clearly corresponds to the pointing gesture which may be directed to a distant or even invisible object. Thus, the MNS would be suitable to support the connection of pointing and the object, by evoking one's own experience of movement and direction of gaze. The discovery of *audiomotor mirror neurons* in the Broca homologous area of monkeys also makes this plausible for *vocal gestures* (Kohler et al. 2002, Keysers et al. 2003). These neurons are activated (1) if the animal observes an action, which generates a sound – for example, knocking on a table or cracking a peanut: (2) if the animal performs the action itself, or also (3) if it only hears the knock or crack without seeing the movement. Transferring this to the voice, this would imply that the heard voice could potentially evoke the same action with an object that the listener could carry out himself.

Hence, in early speech acquisition when pointing and sound gestures are typically linked with each other, a neuronal coupling would be produced between

[13] The question how this matching should be interpreted is still controversial, however. Gallese and Goldman (1998) have originally proposed a simulation theory of mind reading, and Gallese (2008) still defends an embodied simulation of others' expressions on a subpersonal level of the MNS. Such concept have been criticized by phenomenological authors, arguing against the complicated mechanism of an 'as-if'-simulation and backward projection of one's own bodily state onto others (Gallagher 2007, Fuchs and De Jaegher 2009). Instead, one's own bodily resonance may be simply inherent in one's perception of the other, namely as its'proximal' or tacit component (Fuchs 2016b).

(1) the object being pointed to, (2) the related sound, and (3) one's own action with the object. As a result, the originally only accompanying sound becomes capable of evoking the intended object and the object-related action scheme in the listener.[14] At the same time, the gesticulating pointing to objects recedes more and more into the background – as can also be observed in the development of infants.

In the acoustic medium, the word detaches itself from the speaker and is heard by him and the recipient together. The acoustic gesture is thus no longer subject-bound, but for both partners becomes a third entity, an *intersubjective symbol*. Mead (1973) already identified in this reciprocal aspect the decisive attribute of speech: the spoken word as a "significant gesture" becomes a symbol which basically causes the same reaction or idea in the speaker as in the listener. On a neurobiological level, this may be now understood as follows: communication in words is basically grounded in the fact that – in both speaker and listener – via the medium of the MNS the word activates a congruence of neuronal patterns, and thus of ideas or action schemes. The concordant intention in both partners, which manifests itself in the word as an intersubjective symbol, would thus find its match in the resonance which forms between them on the neuronal level. Speech not only produces an intellectual connection among individuals, it additionally involves a biologically anchored *interbodily resonance system*. Thus, it is in virtue of our bodies acquiring, through social interaction, similar neurological structures that we can share the meaning of words and sentences. Although it must be added that the precise functional relevance of the MNS for the evolution and ontogeny of language remains far from being clarified, it already offers strong empirical support for an embodied and enactive view of language.

Summary and Conclusion

My intention in this paper was to show, based on theoretical considerations and empirical evidence, that language cannot be conceived as an abstract, disembodied system of symbols represented in the brains of separated individuals. Instead, language is both produced and understood as a form of embodied interaction, which in speaker and listener evokes the totality of possibilities for action that are mediated by the lived body. Thus, verbal communication is not a transfer of symbolic significances from one mind to another, but a "gesturing with words," co-enacting our actual and possible relations to the world, and scaffolded by our shared practical contexts. Particularly the pointing gesture, through uniting bodily movement and "we-intentionality," may be regarded as the lynchpin that leads from primary intercorporeality to the sharing of meanings through sym-

[14] Apart from the studies on action-related word comprehension which were mentioned at the beginning, this connection is particularly supported by Aziz-Zadeh et al. (2006), who showed that the same cortical regions activated by *action observation* are also activated by the understanding of *action-related sentences*.

bolic interaction. However, as Merleau-Ponty has argued, this transition never loses the gestural, enactive basis from which language first develops:

> "The spoken word is a genuine gesture, and it contains its meaning in the same way as the gesture contains its. This is what makes communication possible. In order that I may understand the words of another person, it is clear that his vocabulary and syntax must be 'already known' to me. But that does not mean that words do their work by arousing in me 'representations' associated with them, and which in aggregate eventually reproduce in me the original 'representation' of the speaker. What I communicate with primarily is not 'representations' or thought, but a speaking subject, with a certain style of being and with the 'world' at which he directs his aim." (Merleau-Ponty [1945] 1962, 213)

In other words, speech is primarily not a symbol system, but transformed gesture, enacted by the body, and evoking possible actions in it. Speaking and understanding are lived acts in which our experiences as embodied agents are always present, both in the content and in the syntactical structure that expresses it.

Speech capacity therefore does not develop merely from a biological *Anlage* or genetic disposition, but like no other human capacity it requires embedding in a sphere of shared meaning structures and communicative practice in order to evolve. Verbal meanings only exist *between* individuals just as pointing with one's finger only attains its meaning from the jointly oriented gaze. Words are carriers of intersubjective meanings, which have formed within a culture and increasingly differentiated into a complex referential system. To learn words, children must primarily be in intercorporeal, emotional and practical contact with others. They must further develop the capacity to focus on the same object and to share this intention with them. Scaffolded by these triadic practical situations the sound gestures may develop whereby we communicate with one another symbolically.

When in the embodied interaction with others the child learns their speech, then his brain functions as an organ of mediation that increasingly matches the heard words with neuronal patterns related to action, interaction and object experiences. This matching only occurs if the child experiences the others as intentional actors who *intend to show him something through their speech* and whose goal is the intended object. In short, the child must experience himself as the *intended* participant of communication. Only then – and not by means of a mechanical-associative connection – can the new words become sedimented as neuronal patterns that are associated with experiences of acting and interacting. The coupling of language perception and motor activity, which is now demonstrated by numerous imaging studies of the brain, shows that the meaning of words always remains connected to the interactive and embodied experiences in which they have been acquired.

The brain as such certainly does not become the location of meanings or the "symbol-processing organ", as it is sometimes referred to. The neuronal patterns, as correlates of speech, are only the *necessary* condition for the child understanding words as meaningful and thus participating in the joint world of the mind conveyed through symbols. Only such participation in the shared symbolic

world is the sufficient condition for speech acquisition. Language is based on *meanings*, and meanings are ultimately based on embodied *relationships*. They are derived from the early childhood experience of joint attention, pointing, from the joint use of speech in practical contexts, and from the intersubjective symbolism of spoken words. Correlates of these meanings are functionally and morphologically inscribed on the brain as neuronal patterns in the course of interaction. In this way, language becomes enmeshed in our organic life: we incorporate into our bodies a linguistic style of being. This is also the reason why "linguistic events have a direct route to even our physiology, why the complex socio-cultural and interpersonal matrix disclosed by an insult or a compliment make our blood rush in quite different ways" (Cuffari, Di Paolo, and De Jaegher 2015, 1116). Language is nothing else than a manifestation of our embodied sociality.

Bibliography

Aziz-Zadeh, L. and A. Damasio. 2008. Embodied Semantics for Actions: Findings From Functional Brain Imaging. *Journal de Physiologie* – Paris 102: 35–39.

Aziz-Zadeh, L., S. M. Wilson, G. Rizzolatti, and M. Iacoboni. 2006. Congruent Embodied Representations for Visually Presented Actions and Linguistic Phrases Describing Actions. *Current Biology* 16: 1818–1823.

Bargh, J. A. and I. Shalev. 2012. The Substitutability of Physical and Social Warmth in Daily Life. *Emotion* 12(1): 154–62.

Barsalou, L. W. 2008. Grounded cognition. *Annual Review of Psychology* 59: 617–645.

Belin, P., R. J. Zatorre, and P. Ahad. 2002. Human Temporal-Lobe Response to Vocal Sounds. *Cognitive Brain Research* 13: 17–26.

Bråten, S. and C. Trevarthen. 2007. Prologue: From Infant Intersubjectivity and Participant Movements to Simulation and Conversation in Cultural Common Sense. In *On Being Moved: From mirror neurons to empathy*, ed. S. Bråten, 21–34. Amsterdam: John Benjamins Publishing.

Bruner, J. 1983. *Child's Talk*. New York: Norton.

Binkofski, F. and G. Buccino. 2004. Motor Functions of Broca's Region. *Brain and Language* 89: 362–369.

Boulenger, V., O. Hauk, and F. Pulvermüller. 2009. Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cerebral cortex* 19: 1905–1914.

Buccino, G., L. Riggio, G. Melli, F. Binkofski, V. Gallese, and G. Rizzolatti. 2005. Listening to Action-Related Sentences Modulates the Activity of the Motor System: A Combined TMS and Behavioral Study. *Cognitive Brain Research* 24: 355–363.

Cuffari, E. C., E. Di Paolo, and H. De Jaegher. 2015. From Participatory Sense-Making to Language: There and Back Again. *Phenomenology and the Cognitive Sciences* 14: 1089–1125.

De Jaegher, H. and E. Di Paolo. 2007. Participatory sense-making: an enactive approach to social cognition. *Phenomenology and the Cognitive Sciences* 6: 485–507.

Darwin, C. (1872) 1998. *The Expression of the Emotions in Man and Animals.* Introduction, Afterword and Commentaries by P. Ekman. London: Harper Collins Publishers.

Dittmann, J. 2002. *Der Spracherwerb des Kindes. Verlauf und Störungen*. München: Beck.

Eibl-Eibesfeldt, I. 1972. Similarities and Differences Between Cultures and Expressive Movements. In *Non-verbal Communication*, ed. R. A. Hinder, 37–48. Cambridge: Cambridge University Press.

Fernald, A. 1992. Human Maternal Vocalizations to Infants as Biologically Relevant Signals. In *Adapted Mind*, ed. J. H. Barkow, L. Cosmides, and J. Toby, 391–426. New York: Oxford University Press.
Fodor, J. 1975. *The Language of Thought*. Cambridge: Harvard University Press
–. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
Fuchs, T. 2010. *Das Gehirn – Ein Beziehungsorgan. Eine Phänomenologisch – Ökologische Konzeption*. 3rd edition, Stuttgart: Kohlhammer.
–. 2011. The Brain – A Mediating Organ. *Journal of Consciousness Studies* 18: 196–221.
–. 2013. The Phenomenology and Development of Social Perspectives. *Phenomenology and the Cognitive Sciences* 12: 655–683.
–. 2016a. Embodied Knowledge – Embodied Memory. In: *Analytic and Continental Philosophy. Methods and Perspectives. Proceedings of the 37th International Wittgenstein Symposium*, ed. S. Rinofner-Kreidl, H. Wiltsche, 215–229. Berlin: De Gruyter.
–. 2016b (in press). Intercorporeality and Interaffectivity. In: Intercorporeality: Emerging Socialities in Interaction, ed. C. Meyer, J. Streeck and S. Jordan. Oxford: Oxford University Press.
Fuchs, T. and H. De Jaegher. 2009. Enactive Intersubjectivity: Participatory Sense – Making and Mutual Incorporation. *Phenomenology and the Cognitive Sciences* 8: 465–486.
Gallagher, S. 2007. Simulation trouble. *Social Neuroscience* 2: 353–365.
Gallese, V. and A. Goldman. 1998. Mirror Neurons and the Simulation Theory of Mind Reading. *Trends in Cognitive Science* 12: 493–501.
Gallese, V. 2008. Mirror Neurons and the Social Nature of Language: The Neural Exploitation Hypothesis. *Social Neuroscience* 3: 317–333.
Glenberg, A. M. and D. A. Robertson. 2000. Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language* 43: 379–401.
Glenberg, A. M., M. Sato, L. Cattaneo, L. Riggio, D. Palumbo, and G. Buccino. 2008. Processing Abstract Language Modulates Motor System Activity. *The Quarterly Journal of Experimental Psychology* 61: 905–919.
Gómez, J.-C. 2007. Pointing Behaviors in Apes and Human Infants: A Balanced Interpretation. *Child Development* 78: 729–734
Grice, P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
Harnad, S. 1990. The Symbol Grounding Problem. *Physica D* 42: 335–346.
Kettner, V. A. and J. I. Carpendale. 2013. Developing Gestures For No and Yes: Head Shaking and Nodding in Infancy. *Gesture* 13: 193–209.
Irwin, B. A. 2015. An Enactivist Account of Abstract Words: Lessons From Merleau-Ponty. *Phenomenology and the Cognitive Sciences* (published online-first).
Jirak, D., M. M. Menz, G. Buccino, A. M. Borghi, and F. Binkofski. 2010. Grasping Language – A Short Story on Embodiment. *Consciousness and Cognition* 19: 711–720.
Johnson, M. 1987. *The Body in the Mind*. Chicago: University of Chicago Press.
Kaschak, M. P., R. A. Zwaan, M. Aveyard, and R. H. Yaxley. 2006. Perception of Auditory Motion Affects Language Processing. *Cognitive Science* 30(4): 733–744.
Keysers, C., E. Kohler, M. A. Umiltà, L. Nanetti, L. Fogassi and V. Gallese. 2003. Audiovisual Mirror Neurons and Action Recognition. Experimental Brain Research 153: 628–636.
Kleist, H. v. (1805) 1951. Über die allmähliche Verfertigung der Gedanken beim Reden. trans. M. Hamburger: On the Gradual Construction of Thoughts During Speech. *German Life and Letters* 5: 42–46.
Kluge, F. 1989. *Etymologisches Wörterbuch der Deutschen Sprache*. 22. Aufl. Berlin: De Gruyter.

Koelsch, S. 2005. Ein neurokognitives Modell der Musikperzeption. *Musiktherapeutische Umschau* 26: 365–381.
Koelsch, S., T. Fritz, K. Schulze, D. Alsop, and G. Schlaug. 2005. Adults and Children Processing Music: An fMRI Study. *Neuroimage* 25: 1068–1076.
Kohler, E., C. Keysers, A. Umiltà, L. Fogassi, V. Gallese and G. Rizzolatti. 2002. Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science* 297: 846–848.
Lakoff, G. and M. Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.
Lee, S. W. S. and N. Schwarz. 2011. Clean Slate Effects: The Psychological Consequences of Physical Cleansing. *Current Directions in Psychological Science* 20: 307–311.
Liszkowski, U., M. Carpenter, T. Striano, and M. Tomasello. 2006. 12- and 18-Month-Olds Point to Provide Information for Others. *Journal of Cognition and Development* 7: 173–187.
Malloch, S. N. 1999. Mother and Infants and Communicative Musicality. *Musicæ Scientiæ* 3(1): 29–57.
McMullen, E. and J. R. Saffran. 2004. Music and Language: A Developmental Comparison. *Music Perception: An Interdisciplinary Journal* 21: 289–311.
Mead, G. H. 1973. *Geist, Identität und Gesellschaft*. Frankfurt: Suhrkamp.
Meier, B. P., S. Schnall, N. Schwarz and J. A. Bargh. 2012. Embodiment in Social Psychology. *Topics in Cognitive Science* 4(4): 705–716.
Meltzoff, A. N. and M. K. Moore. 1977. Imitation of Facial and Manual Gestures by Human Neonates. *Science* 198: 74–78.
–. 1989. Imitation in Newborn Infants: Exploring the Range of Gestures Imitated and the Underlying Mechanisms. *Developmental Psychology* 25: 954–962.
Meltzoff, A. N. and R. Brooks. 2001. 'Like me' as a Building Block for Understanding Other Minds: Bodily Acts, Attention, and Intention. In *Intentions and Intentionality: Foundations of Social Cognition*, ed. B. F. Malle, L. J. Moses, and D. A. Baldwin, 171–191. Cambridge, MA: MIT Press.
Meltzoff, A. N. and W. Prinz. 2002. *The Imitative Mind. Development, Evolution and Brain Bases*. Cambridge, MA: Cambridge University Press.
Merleau-Ponty, M. (1945) 1962. *Phénomenologie de la Perception*. Paris: Gallimard. trans. C. Smith: Phenomenology of Perception. London: Routledge and Kegan Paul.
–. 1960. Le Philosophe et Son Ombre. In *Signes*. Paris: Éditions Gallimard.
Murray, L. and C. Trevarthen. 1985. Emotional Regulation of Interactions Between Two-Montholds and Their Mothers. In *Social perception in infants*, ed. T. M. Field and N. A. Fox, 177–97. Norwood, N. J.: Ablex.
Nelson, K. 1996. *Language in Cognitive Development*. Cambridge University Press, Cambridge.
Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., Clark, S., and Slaughter, V. 2016. Comprehensive Longitudinal Study Challenges the Existence of Neonatal Imitation in Humans. *Current Biology* 26: 1334–1338.
Papoušek, M. 1994. Melodies in Caregivers' Speech: A Species-Specific Guidance Towards Language. *Early Development and Parenting* 3: 5–17.
Papoušek, H. and M. Papoušek. 1987. Intuitive Parenting: A Dialectic Counterpart to the Infant's Integrative Competence. In *Handbook of Infant Development*, ed. H. R. Schaffer, 67–85. London: Academic Press.
–. 1995. Vorsprachliche Kommunikation: Anfänge, Formen, Störungen und psychotherapeutische Ansätze. In *Die Kraft liebevoller Blicke*. Psychotherapie und Babyforschung Bd. II., ed. H. G. Petzold, 123–142. Paderborn: Junfermann.

Patel, A. 2003. Language, Music, Syntax and the Brain. *Nature Neuroscience* 6: 674–681.
Piaget, J. (1936) 1952. *The Origins of Intelligence in Children*. New York: International University Press.
Pulvermüller, F. 2005. Brain Mechanisms Linking Language and Action. *Nature Rev. Neuroscience* 6: 576–582.
Pulvermüller, F., M. Huss, F. Kherif, F. M. del Prado Martin, O. Hauk, and Y. Shtyrov. 2006. Motor Cortex Maps Articulatory Features of Speech Sounds. *Proceedings of the National Academy of Sciences* 103: 7865–7870.
Pulvermüller, F. and L. Fadiga. 2010. Active Perception: Sensorimotor Circuits as a Cortical Basis for Language. *Nature Rev. Neuroscience* 6: 576–582.
Pylyshyn, Z. W. 1984. *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.
Rizzolatti, G. and M. A. Arbib. 1998. Language Within Our Grasp. *Trends in Neurosciences* 21: 188–194.
Schnall, S., J. Benton, and S. Harvey. 2008. With a Clean Conscience: Cleanliness Reduces the Severity of Moral Judgments. *Psychological Science* 19: 1219–1222.
Spitz, R. A. 1957. *No and Yes: On the Genesis of Human Communication*. New York: International Universities Press.
Stern, D. N. 1998. *Die Lebenserfahrungen des Säuglings*. Stuttgart: Klett-Cotta.
Tomasello, M. 2002. *Die kulturelle Entwicklung des menschlichen Denkens. Zur Evolution der Kognition*. Frankfurt: Suhrkamp.
–. 2008. *The Origins of Human Communication*. Cambridge, MA: MIT Press.
Tomasello, M., M. Carpenter, J. Call, T. Behne, and H. Moll. 2005. Understanding and Sharing Intentions: The Origins of Cultural Cognition. *The Behavioral and Brain Sciences* 28: 675–735.
Tomasello, M. and M. Carpenter. 2007. Shared Intentionality. *Developmental Science* 10: 121–125.
Tomasello, M., Carpenter, M., and Liszkowski, U. 2007. A New Look at Infant Pointing. *Child Development* 78: 705–722.
Trabant, J. 1991. Parlare Cantando: Language Singing in Vico and Herder. *New Vico Studies* 9: 1–16.
Trevarthen, C. 1979. Communication and Cooperation in Early Infancy: A Description of Primary Intersubjectivity. In *Before Speech*, ed. M. Bullowa, 321–347. Cambridge, MA: Cambridge University Press.
–. 1998. Language Development: Mechanisms in the Brain. In *Encyclopedia of Neuroscience*, ed. G. Adelman and B. Smith, 1018–1026. Amsterdam: Elsevier.
–. 2001. The Neurobiology of Early Communication: Intersubjective Regulations in Human Brain Development. In *Handbook of Brain and Behaviour in Human Development*, ed. A. F. Kalverboer, A. Gramsberg, 841–881. Dordrecht, Boston, London: Kluwer Academic Publishers.
–. 2008. The Musical Art of Infant Conversation: Narrating in the Time of Sympathetic Experience, Without Rational Interpretation, Before Words. *Musicae Scientiae* 12: 15–46.
–. 2009. The Functions of Emotion in Infancy: The Regulation and Communication of Rhythm, Sympathy, and Meaning in Human Development. In *The Healing Power of Emotion: Affective Neuroscience, Development, and Clinical Practice*, ed. D. Fo-sha, D. J. Siegel, and M. F. Solomon, 55–85. New York: Norton.
–. 2012. Communicative Musicality: The Human Impulse to Create and Share Music. In *Musical Imaginations: Multidisciplinary Perspectives on Creativity, Performance, and Perception*, ed. D. Hargreaves, D. Miell, and R. MacDonald, 259–284. Oxford: Oxford University Press.

Trevarthen, C. and P. Hubley. 1978. Secondary Intersubjectivity: Confidence, Confiding and Acts of Meaning in the First Year. In *Action, Gesture and Symbol: The Emergence of Language*, ed. A. E. Lock, 183–229. London, Oxford: Academic Press.

Turati, C., F. Simion, I. Milani, and C. Umiltà. 2002. Newborns' Preference for Faces: What is Crucial? *Developmental Psychology* 38: 875–882.

Umiltà, M. A., E. Kohler, V. Gallese, L. Fogassi, L. Fadiga, C. Keysers, and G. Rizzolatti. 2001. I Know What You Are Doing: A Neurophysiological Study. *Neuron* 31: 155–165.

Valenza, E., F. Simion, V. M. Cassia, and C. Umiltà. 1996. Face Preference at Birth. *Journal of Experimental Psychology: Human Perception and Performance* 22(4): 892–903.

Varela, F. J., E. Thompson, and E. Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Weick, K. E. 1995. *Sensemaking in organizations*. Thousand Oaks, CA: Sage.

Willems, R. M., P. Hagoort, and D. Casasanto. 2010. Body-Specific Representations of Action Verbs. Neural Evidence From Right- and Left-Handers. *Psychological Science* 21: 67–74.

Wilson, S. M., A. P. Saygin, M. I. Sereno, and M. Iacoboni. 2004. Listening to Speech Activates Motor Areas Involved in Speech Production. *Nature Neuroscience* 7: 701–702.

Winter, B. and B. Bergen. 2012. Language Comprehenders Represent Object Distance Both Visually and Auditorily. *Language and Cognition* 4: 1–16.

Ziemke, T. 2002. Special Issue on Situated and Embodied Cognition. *Cognitive Systems Research* 3(3): 271–274.

Zlatev, J. 2007. Embodiment, Language and Mimesis. In *Body, Language and Mind* vol. 1, ed. T. Ziemke, J. Zlatev, and R. M. Frank, 297–337. De Gruyter Mouton.

Zhong, C. B. and G. J. Leonardelli. 2008. Cold and Lonely: Does Social Exclusion Feel Literally Cold? *Psychological Science* 19: 838–842.

Zhong, C. B. and K. Liljenquist. 2006. Washing Away Your Sins: Threatened Morality and Physical Cleansing. *Science* 313: 1451–1452.

Zwaan, R. A., C. J. Madden, R. H. Yaxley, and M. E. Aveyard. 2004. Moving Words: Dynamic Representations in Language Comprehension. *Cognitive Science* 28: 611–619.