Language-at all times

Action and interaction as contexts for enriching representations

Iris Nomikou,[1,2] Malte Schilling,[1] Vivien Heller[1] and Katharina. J. Rohlfing[2]
[1]Bielefeld University, Germany / [2]Paderborn University, Germany

This article discusses the importance of social interaction for the development of the representations for symbolic communication. We suggest that there is no need to distinguish between different representational systems emerging at different stages of development. Instead, we propose that representations are rich right from the beginning of a child's life, and that they are driven mainly by acting and interacting in the physical and social world. The more variety in a child's interactional experience (i.e., synchrony, sequentiality, and prediction), the more enriched and abstracted the representations become. We review literature providing evidence for the ways in which infants' development toward symbolic communication benefits from repeated social (inter)action and consider some implications for computational approaches.

Keywords: rich representation, language acquisition, interaction, synchrony, contingency, sequentiality

1. Introduction

Child development is commonly described as a progression through specific milestones over time. For language acquisition, this has led to the prominent view that a certain set of competencies is required in order to engage in referential communication; that is, to recognize the meaning of an utterance and to behave accordingly. Traditionally, these competencies are assumed to be acquired one after the other before children are able to participate in verbal interactions (e.g., Piaget, 1937/1954; Werner & Kaplan, 1963). According to Piaget (1937/1954), representational processes begin at approximately 7 to 9 months, are transformed by 16

Interaction Studies 17:1 (2016), **128–153**. DOI 10.1075/is.17.1.06nom
ISSN 1572–0373 / E-ISSN 1572–0381

to 18 months, and become symbolic by the third year of life. In this approach, the infant's own actions play a primary role in constructing the representational capacity that grows out of sensorimotor action (Piaget, 1937/1954). In support of this transformation process, Pruden, Hirsh-Pasek, Golinkoff, and Hennon (2006) showed that 10-month-old infants build an association between a word and an object on the basis of that object's perceptual saliency. In turn, 18-month-olds can make use of signals such as eye gaze to attend to an object of interest and learn new object names. Thus, children make sensorimotor experiences first, and within development, "cognition becomes progressively more abstract and less embodied" (Wellsby & Pexman, 2014, p. 1).

Two main implications can be deduced from this approach: First, when cognition becomes progressively abstract, this implies that embodied experiences become less important. Yet, this contradicts current embodied theories of cognition (see Wellsby & Pexman, 2014). Second, and more relevant for our focus, is the implication that linguistic capabilities are built on top of these cognitive capabilities (e.g., Wellsby & Pexman, 2014; Wojcik, 2013). If this latter idea would prove to be true, it would have serious consequences for the way in which mental representations are used for later language acquisition: Representations connected to early cognitive competencies would have to become mapped onto later emerging linguistic capabilities (see Mandler, 2012; Wellsby & Pexman, 2014; Wojcik, 2013). From our perspective, this view is problematic and not in line with results from current research on the development of growing early linguistic representations in young infants. This suggests that infants as young as 6 months of age understand some language—long before they speak their first words (Bergelson & Swingley, 2012; Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Nomikou et al., 2016).

As an alternative, we argue in favor of a continuous development of rich representations. Thus, instead of a transformation of initially sensorimotor to later symbolic representations, we propose a continuity of representations that is present from early on and shaped throughout development. This process is driven mainly by two factors: acting and interacting in the physical and social world. These two different forces pull each other forward and take turns as to which is pulling which. Yet, representations are not developed only *through* action and interaction. Their primary function is that they are construed *for* actions and interactions. In the following, we review current views on cognitive development, point to some research gaps that we see in the literature on language acquisition, and develop ideas on how to fill them. Our approach aims to sketch rich representation: that is, why representations need to be seen as rich (consisting of [inter]action and language) and how they become rich over the course of histories of interaction, or, in other words, during the process of development.

2. Current views on cognitive development

There is widespread agreement on physical interaction being at the heart of conceptual processes because they constrain how we perceive and conceive the environment (Glenberg, 1997). Conceptualization as a process is fundamental to human cognition because it enables mental representation in, for example, inductive reasoning or language acquisition (Mandler, 2012). On the other hand, the function of conceptualization is to reduce information and to form conceptual spaces (Madole & Oakes, 1999). Yet, despite the agreement on the role of sensorimotor experiences for conceptualization, there is still an ongoing debate on how these sensorimotor experiences serve conceptualization. Wellsby and Pexman (2014, p. 1) identify the core of the debate as being about "whether sensorimotor experiences comprise conceptual knowledge and language or whether accessing this information merely activates sensorimotor areas epiphenomenally." In other words, the debate is over whether sensorimotor experiences are essential for conceptualization or rather are the result of spreading activation.

In his developmental approach, Piaget assumed an abrupt change from the sensorimotor to the symbolic phase (Piaget & Inhelder, 1993). The ability to communicate symbolically plays a key role in this hypothesis. From Piaget's perspective, linguistic symbols can be attached to a mental entity that is already in place as part of a representation. Likewise, an already shaped conceptual representation is often assumed to be a prerequisite for the introduction of a new linguistic form (Bloom, 2000). This process is described fittingly by the "waiting room" metaphor (Johnston & Slobin, 1979). It is only after a concept has been acquired that a child can link a linguistic form to it.

Mandler (1998) has questioned this abrupt change from perceptual to conceptual knowledge. She proposes that as soon as children can parse the world into objects and actions, certain aspects of perceptual information are re-described to serve further cognitive functions (Mandler, 2012). The involvement of own bodily experience and the focus on what objects do and what is done to them is crucial in this approach. Mandler (1998, 2012) introduces two parallel, ongoing forms of conceptualization that develop continuously and take different roles only during development: These are perceptual and conceptual categorization. The former, perceptual categorization, is the process of building representations of visible, tactile, and kinesthetic object features, and it is used for object identification. Whereas this process is more prominent in young infants, it is accompanied continuously by the development of the latter, conceptual categorization, that focuses on the functions and roles of objects and leads to the formation of image schemas and concepts.

Concerning language acquisition, what prior approaches have in common is that they view language skills as developing on top of conceptual ones. One exception is the approach proposed by Göksun, Hirsh-Pasek, and Golinkoff (2010). In their view, there is at least a kind of permeability between concepts and language, allowing the conceptual level to adjust to experiences made during verbal interactions. The authors characterize this permeability as "trading spaces" (Göksun et al., 2010, p. 38). Accordingly, concepts acquired before an infant starts to be verbally productive can become modified —"*traded for*"— according to the categories available within the target language. Other concepts can be formed only when language has become available.

What becomes obvious is that sensorimotor interactions with the environment are hardly ever considered to be important for both conceptual understanding *and* language processing (but see Gibbs, 2006; Wellsby & Pexman, 2014, for exceptions). Instead, in the majority of approaches to language acquisition, conceptualization gives rise to complex mental representations that eventually become semantic and abstracted from immediate and particular experience (Akhtar & Tomasello, 2000). Evidently, this involves not only a qualitative change from sensorimotor experience to conceptual knowledge but also, and most importantly for us, a qualitative change from conceptual to symbolic representations.

In summary, studies on conceptual development propose a qualitative change for the use of language. What remains unclear, however, is how those concepts that are already in place become "translated" for language use.

3. Nature of early representations

To the best of our knowledge, the only clear picture about how children can make use of concepts for language acquisition is presented in Mandler (2012). She suggests that from the beginning of a child's development, concepts (stemming from the child's exploration in space) are formed and then recruited for early words. However, the problem with this view is that it requires universal—and not language-specific—concepts (McDonough, Choi, & Mandler, 2003). Universal concepts are assumed to be acquired before the infant's language use. As they rely on the infant's experience in space, they will be universal in the sense that they are not dependent on categories that may differ from language to language. For example, in Korean, the distinction between putting something into a container tightly (*kkita*) or loosely (*nehta*) is marked linguistically, whereas in English the two events can be described by one spatial preposition (*in*). Choi, McDonough, Bowerman, and Mandler (1999) have suggested that children become sensitive to categories offered in their language early in development. However, another possibility is

that children first acquire finer concepts (such as that of tightness) (McDonough et al., 2003) that can then be used for the categories that are necessary in language (Göksun et al., 2010). At this point, we cannot provide a full picture on the debate about how language influences conceptualization (but see, e.g., Bowerman, 1996; Choi et al., 1999). We would like to point out, however, that, this debate reveals a discrepancy regarding the role of language and interaction: On the one hand, concepts are proposed that rely on infants' spatial experience (Mandler, 2012); on the other hand, it is clear that language is already a resource in interactions with children right from the beginning, and that it affects which words children will learn first (Choi, 2000). This is probably because it accompanies actions to a great degree from early on (Nomikou & Rohlfing, 2011). We shall take up this point in the next section.

Another important aspect that is still unclear in current approaches to language acquisition is how concepts remain up-to-date so that they can be contextualized dynamically for the purpose of the task. In this sense, representations cannot be understood as simply standing for something out there in the world. Behavioral and neuroscientific research has led to a new perspective in which representations have to be dynamic in nature. A growing body of research points out that a core function of representation is to enable meaningful predictions about the ongoing course of actions (Clark, 2013; Gilbert & Wilson, 2007; Glenberg & Kaschak, 2002). Accordingly, mental representations are not just associations relating current knowledge to past similar experiences. Instead, the key function of representation is to unearth the governing invariants of dynamic situations and recognize the structures of situated actions and interactions. This allows us to make *meaningful predictions*, that is, use the known and learned relations for a given context, and project these toward a further point in time. Interdisciplinary research has established prediction to be a key mechanism employed not only in observation, imitation, in understanding others' actions, and in planning ahead (Clark, 2013; Glenberg, 1997; Jeannerod, 2006), but also in communication (Buccino et al., 2005; Feldman & Narayanan, 2004; Fischer & Zwaan, 2008; Pulvermüller, 2005). Because the predictive and dynamic nature of mental representation appears to be at the core of many cognitive competencies, it always has to be kept in mind when considering the characteristics of representations. We shall follow this point and identify predictability as one of the dimensions of rich representations in following sections.

4. What do we miss in language acquisition models?

In this section, we identify a gap between research proposing that prediction is a key mechanism in representations and mainstream developmental approaches. From our perspective, the gap emerges because developmental approaches disregard the influence of the environment (both *physical* and *social*) on learning processes. In particular, we argue that social interaction is the natural habitat for language and that it provides an entry into mental representations and language.

It is generally accepted that interaction plays an important role in the development of social, cognitive, and linguistic skills (Bruner, 1983; de León, 2000). Yet current models addressing the way in which mental representations are established do not account for this fact. Instead, for example, in the trading spaces model proposed by Göksun et al. (2010), infants first explore their physical world. Then, "the more language [children] know, the more attentive they are to" language-specific categories (p. 38). Here, the assumption is that children need to produce language to become sensitive to it. Yet, there is a large body of research suggesting not only that infants orient themselves toward human voices (Colombo & Bundy, 1983; Muir & Field, 1979) or faces (Slater et al., 2010) but also that verbal signals help infants as young as 3 months of age to categorize objects (Ferry, Hespos, & Waxman, 2010). Thus, the assumption that children's sensitivity to language starts later in development seems inconsistent with current research.

The same critique applies to the Emergentist Coalition Model (Golinkoff & Hirsh-Pasek, 2006, 2008; Hirsh-Pasek & Golinkoff, 1996). This model adopts a hybrid approach, departing from the belief that children use multiple sources of information to learn language. According to the model, 10-month-old infants are initially sensitive to attentional cues such as perceptual salience, temporal contiguity, and novelty (Pruden et al., 2006). As they develop, they also become sensitive to social cues such as eye gaze, and, from around 18 months of age, they can use this social information in the form of, for example, eye gaze or pointing gestures, to learn new words.

As in Göksun et al.'s (2010) model, social interaction becomes important only in the later phase of development. It is not before 18 months that children start to pay attention to social cues. However, the transformation from a perception-guided to a socially aware child cannot be explained. Golinkoff and Hirsh-Pasek (2008) posit that the trigger for this transformation might be the recognition of others' intentions. Yet it remains unclear how exactly children come to recognize intentions, and why social information becomes only gradually more important in language acquisition.

In sum, most models attempting to explain the development of representations contrast the preverbal infant with the verbal infant. Her or his own verbal

production is assumed to either induce a different kind of conceptualization (Göksun et al., 2010) or to initiate mapping of already existing concepts onto early words (Mandler, 2012).

Yet, we argue that infants' cognition is influenced by language and interaction long before they produce their first words. A neurophysiological study by Parise and Csibra (2012), in which 9-month-old infants were tested to detect a mismatch between an object appearing and a preceding label, revealed a N400 effect documenting that they understand their mothers' speech. In this vein, Bortfeld, Morgan, Golinkoff, and Rathbun (2005) as well as Bergelson and Swingley (2012) have shown that 6-month-old infants already understand some frequent nouns. How could it be possible for infants to understand language early without this understanding affect their representations? Clearly, we need an approach to the development of representations that accounts for the experience with language that infants make before they start to speak. We call this experience the history of interaction that accumulates over the child's development.

We are convinced that models seeking to explain the development of representations need to focus more on the *history of interaction* on both a microgenetic and an ontogenetic timescale. Concerning the history of interaction, the microgenetic perspective takes into account the *unfolding interaction* in a particular moment within the course of a single interaction. Recently, Liszkowski (2014) has shown that preceding actions are a source by which children can infer others' intentions and recognize the meaning of an utterance. However, rather than focusing on this timescale, we take an ontogenetic perspective and look at the *cumulative history of interaction* actions are recurrent and children also draw from the recurrent context and their expectations when interpreting an ongoing situation (Rohlfing et al., 2016). We thus propose that *situated interaction* is the driving force for symbolic development right from the start (Rohlfing et al., 2016). We further propose that the situated interaction changes in quality as adults and children establish shared activities as *interactional histories* and children gain interactional experience. In the following, we shall elaborate on our view of the co-development of symbolic and non-symbolic skills and how these enhance and scaffold each other. In doing so, we shall attempt to contribute to understanding the nature of representations not only as structures that manage to integrate different sources of information with development (Pauen, Birgit, Hoehl, & Bechtel, 2015) but also as being rich in the sense of involving social information from early on.

5. Toward rich representations: linking embodiment and social interaction

Our argument needs to address two underlying questions: First, what is the force driving the growth of rich representations? And, second, what makes these representations rich?

We consider the key driving factor for the growth and enrichment of representation to be *interaction*, because it elicits meaningful actions from the child. The child has to apply and contextualize her or his accumulated knowledge in the form of a representation. Contextualization brings about the dynamic update of these representations. We propose that it is in the nature of our representations to be applied in interaction, that is, within a specific task involving the physical and/or social world.

The motor ability of self-locomotion allows a child to discover increasingly more affordances of objects and enhances the growth of sensorimotor abilities. For example, Karasik, Tamis-LeMonda, and Adolph (2014) observed that children who move around have free hands not only to handle objects but also to bring them to their caregivers and, thus, to elicit further information about them. This example demonstrates impressively how advancing motor abilities can change the nature of an interaction and enable new interactional experiences such as child-initiated bring-and-take routines. This example also reflects the growth of representations. Furthermore, we presume that representations can also grow through the variety in the forms of interaction that a child experiences: The more the interactional contexts vary, the more the child will come to find similarities or dissimilarities across them. Here, there is a lack of methodologically creative studies seeking ways to compare different contexts and reveal how representations become enriched and abstracted.

Verbal expressions can be applied in early as well as later interactions (Bruner, 1983), and children seem to be particularly sensitive to this kind of behavior, preferring it over other acoustics (Colombo & Bundy, 1983; Muir & Field, 1979). Certainly, the nature of their meaning differs: Whereas concrete actions and functions underscore the meaning of a "ball," more comparisons across different contexts are necessary to account for the word "happy." In this sense, some expressions are a subset of abstract representations (Sadoski & Paivio, 2004) but not all language is abstract.

Now, we turn to the question of what is rich in rich representations. Human representations are rich in that they are *grounded*, *embedded*, and *coupled* to our physical and social interactions. This allows them to be constructed out of our immediate environment and to be adaptive to any kind of context and interaction. Thus, it is not simply skill development that makes representations rich. Instead,

representations become rich when contextualized *in* and *for* an interaction; that is, when the newly developed skills are utilized for a particular task. In this sense, early mother-child interaction may provide a context for stimulating the child's perceptual system. It may well be that early in development, the child does not perceive actions as "caused" by someone else, let alone as intentionally addressed to her or him. It might also be that the child does not perceive these actions as an interaction at all. We propose that what might start off by being only the perception of an immediate effect, a contingency (Cangelosi et al., 2010; Watson, 1985), is then enriched within interaction to become a representation of predictable rules. We shall elaborate on this below. Other existing approaches have proposed that the "initial spatially based conceptual system" receives enrichment via emotions and language (Mandler, 2012, p. 440). We propose that it is not solely due to a child's language that enrichment is initiated, because language is present in early activities (Nomikou & Rohlfing, 2011). Instead, it is an outcome of the child's growing experience in interaction and the awareness of her or his roles when performing joint actions. Thus, we assume that the rich neural structure of representation is constituted of (and cannot be separated from) multiple interactional experiences from the beginning of a child's life (Anderson, 2010).

There is one specific feature of the representational richness that we would like to emphasize here: the bidirectionality of action and language. It is reasonable to assume that language is grounded in rich representations, and that it uses the same resources of (brain) structure that are involved in an agent's actions. In fact, Wagner and Lakusta (2009) postulate equivalence between language and action concepts (cf. Göksun et al., 2010). Yet there is also the other direction, because there have been some analyses of how language can interact with conceptual and motor processes. With the notion of "acoustic packaging," Hirsh-Pasek and Golinkoff (1996) have proposed that even surface properties of language such as prosody act as "carving knives" (p. 161) aiding the child in constructing primary representations of action units. Following this idea, Brand and Tapscott (2007) showed that infants group sequences of actions as belonging together according to the way in which these have been "packaged" by a narration. More recently, Gampe and Daum (2014) showed that for older children, the semantics of language influence action understanding: When primed with an action word, 24-month-old infants will predict the goal of the presented action more quickly. In addition to language facilitating action representations, studies have also demonstrated that accompanying a visual presentation with language enhances infants' conceptualization processes (Balaban & Waxman, 1997; Ferry et al., 2010; Plunkett, Hu, & Cohen, 2008). Research has started to address the issue of co-developing language and action (Cangelosi et al., 2010), but we need to know more about how their mutual influence shapes the representational level from the beginning, and not

just when children start to produce their first words. Furthermore, we propose that we need to know how this mutual influence is manifested in interaction.

Table 1. Characteristics of richness: The table illustrates how different forms of (inter) action give rise to different experiences that, in turn, enrich representations

Form of (Inter-) Action	Experience	Characteristics of Rich Representation	Examples
Intra-personal, intermodal synchrony	Binding of multimodal sources => unity	Embodied, perceptual	Bahrick et al., 2004; Lewkowicz, 2000; Nomikou & Rohlfing, 2011; Zukow-Goldring, 1997
Inter-personal synchrony	Alignment Notion of the other develops.	Coordinated	Jaffe et al., 2001; Stern, 1985
Alternate actions	Contingency, complementarity Organized joint action, including turn-taking (as alternation)	Contingent	Bateson, 1979; Papoušek & Papoušek 1990; Nomikou et al., 2013
Goal directed actions	Purposefulness, joint goal Understanding the joint action and how actions complement each other.	Functional, Purposeful	Rączaszek-Leonardi et al., 2013
Conventionalized goal directed actions	Culturally instilled roles and purposes Role understanding and conventionalized frames are shaped.	Sequential	Bruner, 1983; Bruner & Sherwood, 1976; Ratner & Bruner, 1977; Strähle, 2013; Peters & Boggs, 1986
Pretense	Transcending the here and now Representation (schematic, frames) can be applied in novel contexts– "as if" behavior– allowing a form of meta-cognition	Hypothetical, Fictitious	Harris & Kavanough, 1993; Rakoczy & Tomasello, 2006; Friedman & Leslie, 2007; Sobel, 2007

In the following, we want to pick out a few characteristics to illustrate the process of the development of rich representations. We view these characteristics as contexts that enrich representations and are not bound by a temporal order. However, even though the characteristics do not occur in a developmental sequence, they do depend on each other in the sense that, for example, turn-taking makes a whole routine possible. Table 1 gives a broader overview of the emerging characteristics

and the qualitative change they bring about. Of course, the increase in enrichment implicitly suggests some order. Yet in our approach, this order does not refer to the development of skill or competence in the child, but rather to the complexity of the contexts enabling new qualities of richness in the representations.

5.1 Intermodal synchrony: Embodiment

The infant is born in a vivid world full of sensations that she or he can perceive through a vast array of sensory systems: vision, audition, touch, smell, proprioception, and balance. Gibson (1966) suggests that invariance in the perceptual array can be picked up across modalities. This is called direct perception. For example, what looks like an apple, feels like an apple, and smells and tastes like one will be perceived as belonging together, constituting an apple. This information is already structured in the environment, so the perceiver does not need to integrate it actively in order to perceive unity. For the development of representations from an embodied view, this principle equips naive learners with a mechanism for perceiving and structuring the environment they inhabit.

Accordingly, infants as young as a few weeks old can make use of sensory information originating from different modalities to help them perceive unity in their environment (Bahrick, Lickliter, & Flom, 2004; Lewkowicz, 2000) through intersensory facilitation (Bahrick & Lickliter, 2000; Gibson, 1966). This is biologically plausible, because according to theories of embodied cognition, unity of perception and dynamic categories for objects and actions emerge from the time-locked correlation of multisensory experiences that vary with the infant's activity and actions. Right from the start of their life, infants are, thus, capable of perceiving unitary objects, integrating information from multiple sources, and constructing meaningful units that will in time become linguistic units (Thelen & Smith, 1994). Yet, perception does not develop in a social vacuum but within a certain social context; that is, within daily encounters with social partners who affect the way infants act, interact, and learn language. In a recent study, Cekaite (2016) revealed that perception of touch is used to organize the interaction. By using sustained touch in adult-child interaction, touch was applied to establish a corporeal perceptual field that affords emerging contingent interactivity. In this sense, apart from being a form of action, multimodal synchrony as a type of coordinated behavior can also be a form of interaction.

5.2 Interpersonal synchrony and alignment: Mirroring and contingency

Rączaszek-Leonardi, Nomikou, and Rohlfing (2013) have suggested that synchrony does not just appear within one partner's communicative turn but can

be distributed over interacting parties. Within interaction, we argue, synchrony creates opportunities for the perception of different characteristics of representations. As mentioned above, intermodal synchrony is the perception of similarity in multiple sources of information in the environment. In contrast, interpersonal synchrony is similarity between one's own behavior and the behavior of another person; and these two, we argue, are qualitatively different perceptions. We suggest that because infants are born into a world of social interactions, the perception of intermodal synchrony is embedded in their early social interactions. Thus, we do not view intermodal and interpersonal synchrony as distinct processes, but rather as nested, systemic characteristics.

In interactional synchrony, there is an alignment of the individual contributions of interacting parties that becomes a property of the dyad. Interactional synchrony refers to a type of interaction between a child and caregiver (e.g., mirroring a smile). It is a pattern of interaction that is regulated mutually and is reciprocal (Harrist & Waugh, 2002). As such, it involves dynamic adaptation on the part of both partners. In early mother–infant interactions, this has been confirmed in the temporal alignment of activity level (Beebe et al., 1982; Papoušek & Papoušek, 1989; Stern, 1985) observed in facial expressions (Beebe et al., 1982; Stern, 1974; Trevarthen, 1979) and properties of vocalizations (Leimbrink, 2010; Papoušek & Papoušek, 1989). This behavioral alignment goes beyond pure mimicry to subserve interpersonal coordination (Jaffe, Beebe, Feldstein, Crown, & Jasnow, 2001): To achieve synchronous interactions, the partners align their own behavior to the signals of the other interacting partner, allowing the discovery of regularities between them. In this way, partners "share" subjective states and emotional expressions (Trevarthen, 1979).

Thus, if experiencing intermodal synchrony assists infants in perceiving unity in the environment, experiencing synchrony within interaction is even more powerful because it does not just involve the sensory information originating from one's self. This information also causes a reaction in another person. Harrist and Waugh (2002, p. 564) refer to this as the "experience of effectance." This is the experience of the completion of self-initiated actions.

The experience of effectance can also be seen in dialogical interaction in which a sequence of actions (see Table 1) is construed jointly. Reddy and Uithol (2015) point to the fact that to engage with another dialogically, you need not the same but a different "other," because a dialogue is about complementarity and not mirroring. In dialogical interactions, the behavior of one participant is seen as constraining that of the other (Rączaszek-Leonardi et al., 2013). Consequently, within early social interaction, children might become sensitized to sequential patterns that bear characteristics that are produced collaboratively. As the child learns to participate in sequences of actions such as openings or give-and-take (Nomikou, Rohlfing,

& Szufnarowska, 2013; Papoušek & Papoušek 1990), her or his actions involve simple, locally contingent moves that have a function. As interactional moves become *functional* , they provide a further characteristic of richness. Representations become adaptive to different situations and moldable by experience.

5.3 Establishing rules of exchange: Conventionalized sequential organization

One aspect introducing a further characteristic of richness is interactional history, that is, the experience accumulating through recursive situations (see also Bruner, 1983; Hsu & Fogel, 2003). In the previous section, we suggested that regularities emerge from the interaction itself and constrain the way in which the next action will take place. Here, we further suggest that through repetitions over time, the regularities may enable the emergence of conventionalized action sequences—or routines—that lead to enculturation and the development of language (Rohlfing et al., 2016). We assume that this enables new characteristics of representations that, again, further enrich existing ones.

Routines form a basic constituent of orderly social life (Goffman, 1967; Peters & Boggs, 1986) and have been considered as foundational for early language acquisition (Bruner, 1983; Hirsh-Pasek et al., 2015). Many of the routines established preverbally are designed to involve the child directly. Examples are give-and-take games (Ratner & Bruner, 1978), peek-a-boo (Bruner & Sherwood, 1976), and caretaking routines (Ferrier, 1978; Nomikou, 2015; Tulbert & Goodwin, 2011). What all these activities have in common is that they form *recurrent joint action patterns* or "standard action formats" (Bruner, 1975; Rohlfing et al., 2016). This definition highlights two crucial aspects of routines: The activity is *co-constructed* jointly and it has to occur *repeatedly*. Being co-constructed, these activities are interactive in nature. Thus, they entail not only a particular temporal order of individual actions ("what to do next") but also a particular social organization: Participants assume certain interactive roles and take responsibility for role-related tasks ("who does what"). This enables the participants to anticipate actions and to develop expectations regarding how actions relate not only to each other but also to certain outcomes or effects (Ambrosini et al., 2013; Fantasia, Fasulo, Costall, & López, 2014). Being recurrent, routines are said to enable the child to develop basic cognitive concepts such as that of agent, action, object, or recipient that also underlie linguistic symbols (Bruner, 1975). We suggest that the routines themselves provide a vehicle for developing basic representations underlying verbal communication; they are not linguistic themselves, because a correspondence between action structure and grammatical structure is not necessary.

As demonstrated by Gleason and Weintraub's (1976) study of the trick-or-treat routine, children realize the linguistic constituents of a routine without

understanding (and without being taught) the semantic meaning of the utterances. This suggests that verbal participation in the routine does not necessarily build on the development of linguistic representations. Instead, highly formulaic routines (Gleason, Perlmann, & Greif, 1984) such as trick-or-treat seem to be representations learned as a social *practice* . This also argues against a strict differentiation between nonlinguistic and linguistic representations. Even if children do not employ verbal resources productively in their first months of life, they are involved in verbal interactions from the outset. Given the fact that children are addressed as co-participants right from the beginning (de Léon, 2000), we propose that they do not have to pass through a "prelinguistic stage" (see our argumentation in Section 3 above). Thus, the development of representations is a continuous process. Expanding on social-pragmatic theories (Bruner, 1983; Tomasello, 2003), we propose that by participating in routine activities, children learn to represent *shared purposes* .

The shared purpose of peek-a-boo and vocal-kinetic play routines, for instance, is to experience excitement, togetherness, and joint pleasure (Fantasia et al., 2014). Likewise, when reading a book, co-participants pursue a sharing of the activity with each other. Therefore, we argue that successful participation in routines enables the emergence of an interactive representation of the *shared purpose* of an activity. Most importantly, by assuming interactive representations, we can avoid thinking about representations in terms of "reading" co-participants' intentions, because this automatically presupposes some higher order cognitive ability, the onset of which is then assumed to take place later in development.

With respect to activity-based routines such as peek-a-boo and vocal-kinetic play/nursery rhymes, their purpose can be characterized as conveying immediate bodily and socio-emotional experiences. Sequential analyses have demonstrated that the adult's interactive support is not geared toward making the child understand what the adult has in mind, but rather toward making the child realize "what comes next" (Heller & Rohlfing, 2015; Tarplee, 2010). Rather than forming representations of the co-participant's intentions, the child needs "to recognize, project, and employ unfolding structures of action from the naturally situated behaviors of others" (Lerner, Zimmerman, & Kidwell, 2011, p. 57).

The ability to anticipate a number of sequentially organized actions forms the basis for participating in complex routines that encompass a number of adjacency pairs such as summons–answer, identification–recognition, greetings, and "how-are-you" in greeting routines (cf. Strähle, 2013). By the age of 3 years, children can take the initiative for the transition to the next action. It can be assumed, therefore, that they have developed a representation of the sequential organization of the greeting routine.

Using appropriate conventionalized communicative means, for example, knowing how to respond to a greeting, is, we argue, similar to being able to use

symbols. On the one hand, such symbols are constructed within ontogenetic ritualization: Particular social behaviors emerge from interaction and come to function as intentional communicative signals (e.g., Marentette & Nicoladis, 2012; Spranger & Steels, 2014). For example, a mother–infant dyad may develop a tickling routine, such that some communicative signal may emerge as a convention for initiating the tickling. On the other hand, social behaviors are transmitted within a process of enculturation: By participating in routines, such as greetings, children are instilled with culturally appropriate ways of interacting.

5.4 Predictability as an underlying mechanism

What the above-mentioned characteristics have in common is that they all supply children with a sense of predictability: within their own bodies and for the interaction with the physical or social world. In fact, within the initial cognitive endowment, Bruner (1983, p. 24) postulates that "much of the cognitive processing going on in infancy appears to operate in support of goal-directed activity." An infant seems to be "sensitive to the requirements of prediction" (Bruner, 1983, p. 24), which means that it is important for an infant to establish a knowledge base on which a situation can be interpreted. Papoušek, Papoušek, and Harris (1987) have shown how keen infants are to predict a behavior; they describe the "smile of predictive pleasure" in infants, suggesting that infants feel comfortable and draw pleasure from successfully predicting their environment.

Adding to current approaches postulating that children need to develop a conception of goal-directed activity by recognizing "the relation between the agent and his or her goal" (Gerson & Woodward, 2014, p. 264), we propose that although this ability emerges from sensorimotor experience (Woodward, 1999), it is simultaneously nested within interactional experience (Rączaszek-Leonardi et al., 2013; Reddy & Uithol, 2015). Thus, infants simultaneously develop action control for their own and for collaborative actions. Evidence in favor of our argument is provided by Kaye and Wells (1980) in their work on neonates responding to a contingent action provided by their mothers. In their view, feeding "can appear to be an exchange of turns, in which the infant's pause is answered by the mother's jiggling and the end of jiggling is answered by the next burst" (p. 29). Already in such early interactions, the dynamic and predictive nature of representation appears to be a crucial characteristic. Providing further evidence along these lines, Reddy, Markova, and Wallot (2013) report on 2 month-olds showing anticipatory body adjustments to being picked up. Similarly, Rączaszek-Leonardi et al. (2013) argued that shared intentionality emerges within repeated interactions as the child's actions become incorporated into dialogical events, and that through participation in these shared events, the infant's actions become parts of meaningful wholes in

which these actions have specific causal consequences. Take the example provided by Reddy et al. (2013) of a child anticipating the mother's action of picking up: The infant's representations not only allow her to foresee in the sense of simulating the action; representing involves participating in an action with one's own body. It is likely that growing experience in different interactive activities gives rise to further types of rich representations that can be utilized in more complex predictions such as mental simulations (Gallese & Lakoff, 2005. To apply a mental simulation, it is necessary to differentiate who is doing what for which purpose. Such fully-fledged mental simulations are assumed to underlie language understanding (Fischer & Zwaan, 2008; Pulvermüller, 2005). Our suggestion that language is a part of early interactions right from the beginning may help to reveal a way to investigate the development of the mutual influence of action and language on the representational level.

6. Conclusions

Interaction is crucial for the development of early embodied knowledge and mental representations. Whereas existing approaches focus on children's sensorimotor experiences flowing directly into the structure of cognitive knowledge, we provided evidence strongly suggesting that from early on, children's experiences are shaped by—and benefit from—their participation in social interaction.

We proposed that there is no need to distinguish between different representational systems that come into place at different stages of development. Rather, we argued that representations are rich right from the beginning of a child's life. In time, they are further enriched as they grow during development through the influence of (inter)action. Verbal expressions occur already in early interactions. The nature of their meaning differs depending on which task they are applied in. We therefore proposed that some expressions are a subset of abstract representations (Sadoski & Paivio, 2004).

From our perspective, the role of interaction for enriching representations is twofold: On the one hand, representations are developed *through* action and interaction. Through participation in different interactive activities, the child first experiences intermodal and interpersonal synchrony of actions and emotional states. As children develop increasing motor and cognitive skills and participate in increasingly complex forms of (inter)action, this affords the representation of new dimensions of richness, especially knowledge of purposes and sequentiality of (inter)actions. Thus, on the other hand, representations are also construed *for* actions and interactions: Building up knowledge of the sequential organization of an activity enables the child to initiate activities and act more autonomously.

Thus, early representations may entail a child's action for a particular task. The function of this action is then to be predictive in the sense of initiating and connecting actions to effects in the physical or social world. In its structure, it must embody synchronous and sequential action steps as well as some information about the roles that are to be fulfilled. Representations become rich when newly developed skills are utilized for a particular task.

This approach has implications for both empirical and computational research. We propose that experimental studies should not position infants in a spectatorial stance (Reddy & Uithol, 2015) in which they are treated as observers and analyzers of a situation or an action. Given the importance of engagement in interaction for rich representation, experimental paradigms should embed children into contexts in which they participate in actions and not just observe them. Furthermore, experimental studies should consider the history of interaction as a factor of the settings. Because children bring interactional expectations into any situation, experiments should make use of these expectations by positioning children in an interactional frame in which they can use their expectations to solve a task, make correct inferences, or form categories (e.g., Ferguson & Waxman, 2016).

In observational studies and the analysis of data, coding schemes should incorporate multiple verbal and nonverbal resources (e.g., Nomikou & Rohlfing, 2011; de Barbaro, Johnson, & Deák, 2015). Existing approaches in dynamic systems research have provided insights into the dynamics of, for example, oral–manual coupling throughout infants' development (e.g., Iverson & Thelen, 1999). Furthermore, analyses should consider the dyad as the minimal unit of analysis. There is already a long tradition in qualitative methods and multiple-case analyses of using conversation analysis and microanalysis (e.g., Rossmanith, Costall, Reichelt, López, & Reddy, 2014; Takada, 2005). The generalization from single cases to entire corpora requires dynamic measures that can reveal the nonlinear, emergent quality of the relationship between interacting parties in order to incorporate the situated, embodied aspects of interactions and the process of development within the history of interactions. First studies have started to use nonlinear methods such as recurrence quantification analysis (e.g., Nomikou, Leonardi, Rohlfing, & Rączaszek-Leonardi, 2016; Warlaumont, Richards, Gilkerson, & Oller, 2014).

Further implications involve computational approaches. The view presented here is in line with the fundamental ideas of embodied cognition (Barsalou, Santos, Simmons, & Wilson, 2008). Representations have to be grounded. Importantly, rich representations assume a strong notion of embodiment as proposed by, for example, Lakoff (1987): The meaning of every, even quite high-level (or abstract) concept is rooted in lower-level and embodied (structured) experience. The consequence for modeling approaches is that this requires rich and

connected representations. Most examples in present-day modeling approaches deal only with static features and percepts. But rich representations have to be dynamic in their nature in order to capture temporal progression as well as to allow for predictions. As argued, action and interaction are the key to guiding and driving conceptualization. It is thus necessary for processes themselves to also be capable of being represented. Current work in developmental robotics has started to focus on representation of actions and events (Feldman & Narayanan, 2004; Schilling & Narayanan, 2013; Summer-Stay, Teo, Yang, Fermüller, & Aloimonos, 2013). Importantly, such dynamic rich representations must be able to express more than simple sequences. They should express rich temporal relations allowing for synchronization and co-occurrence, because these are prerequisites for the interaction and cooperation during which multiple co-occurring actions unfold and have to be synchronized (Schilling & Narayanan, 2013).

Central to our view is that these representations are enriched continuously during development. We do not assume multiple representational systems that are bootstrapped one after the other. Instead, conceptual, semantic, and linguistic representations are tightly interwoven and all develop continuously. Over time, they take different roles and differ in their importance. This is in line with Pezzulo et al.'s (2012) view on grounded cognition that argues in favor of embodied computational models and, in particular, of integrated representation. Our approach extends their view, because although their view acknowledges the dynamic nature of representation, it lacks the structure required to organize interaction. Whereas Cangelosi et al. (2010) share the view that conceptualization is driven by action and social interaction, the key differences to the view proposed here are that action and interaction are treated separately and are based on different skills as well as representations. Also, they see language as a functionality built on top of a conceptual space that is already in place. Cangelosi et al. (2010) point out four major current challenges: learning compositional action, lexica, social interaction, and integrating action and language representation. From our point of view, these cannot be separated, and it is this last challenge—the connection between action (as dynamic and structured representations) and language representation (in a broader sense because it should also include interaction)—thatneeds to be addressed first. Only a focus on continuous enrichment will lead toward models that allow us to integrate any interaction experience from early on and to bootstrap language from these experiences. Whereas such a rich representation systems appears difficult be currently realized, Cangelosi et al.'s (2010) suggestion provides a good starting point: Further research on the development and modeling of representations should depart from function (see also Nelson, 1974) as the core of representations.

Acknowledgements

This work was supported by the Cluster of Excellence Cognitive Interaction Technology "CITEC" (EXC 277) at Bielefeld University funded by the German Research Foundation (DFG). We thank the participating families. We also thank Jonathan Harrow for native-speaker advice. Finally, we are much indebted to the two anonymous reviewers for their valuable comments.

References

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff & K. Hirsh-Pasek (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115–135). Oxford, England: Oxford University Press. doi: 10.1093/acprof:oso/9780195130324.003.005

Ambrosini, E., Reddy, V., de Looper, A., Costantini, M., Lopez, B., Sinigaglia, C., & Lappe, M. (2013). Looking ahead: Anticipatory gaze and motor ability in infancy. *PLoS ONE*, 8(7), 1–9. doi: 10.1371/journal.pone.0067916

Anderson, M.L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, 33, 254–313. doi: 10.1017/S0140525X10000853

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, 36(2), 190–201 doi: 10.1037/0012-1649.36.2.190

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, 13(3), 99–102. doi: 10.1111/j.0963-7214.2004.00283.x

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, 64(1), 3–26. doi: 10.1006/jecp.1996.2332

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Symbols, embodiment, and meaning. In M. D. Vega, Glenberg, A. M., & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 245–283). Oxford, England: Oxford University Press. doi: 10.1093/acprof:oso/9780199217274.003.0013

Beebe, B., Gerstman, L., Carson, B., Dolins, M., Zigman, A., Rosensweig, H.,... Korman, M. (1982). Rhythmic communication in the mother–infant dyad. In M. Davis (Ed.), *Interaction rhythms: Periodicity in communicative behavior* (pp. 77–100). New York, NY: Human Sciences Press.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9), 3253–3258. doi: 10.1073/pnas.1113380109

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, 16(4), 298–304. doi: 10.1111/j.0956-7976.2005.01531.x

Bowerman, M. (1996). Learning how to structure space for language: A crosslinguistic perspective. In P. Bloom, Peterson, M., Nadel, L. & M. Garrett (Eds.), *Language and space*. (pp. 385–436). Cambridge MA: MIT Press.

Brand, R. J., & Tapscott, S. (2007). Acoustic packaging of action sequences by infants. *Infancy*, 11(3), 321–332. doi: 10.1111/j.1532-7078.2007.tb00230.x

Bruner, J.S. (1975). The ontogenesis of speech acts. *Journal of Child Language*, 2(1), 1–19. doi: 10.1017/S0305000900000866

Bruner, J.S. (1983). *Child's talk: Learning to use language*. New York, NY: Norton.

Bruner, J. S., & Sherwood, V. (1976). Peekaboo and the learning of rule structures. In J. S. Bruner (Ed.), *Play: Its role in development and evolution* (pp. 277–285. Harmondsworth, England: Penguin.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, 24(3), 355–363. doi: 10.1016/j.cogbrainres.2005.02.020

Cangelosi, A., Metta, G., Sagerer, G., Nolfi, S., Nehaniv, C., Fischer, K., & Zeschel, A. et al. (2010). Integration of action and language knowledge: A roadmap for developmental robotics. *IEEE Transactions on Autonomous Mental Development*, 2(3), 167–195. doi: 10.1109/tamd.2010.2053034

Cekaite, A. (2016). Touch as social control: Haptic organization of attention in adult-child interactions. *Journal of Pragmatics*, 92, 30–42. doi: 10.1016/j.pragma.2015.11.003

Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book reading and toy-play contexts. *Journal of Child Language*, 27, 69–96. doi: 10.1017/s0305000999004018

Choi, S., McDonough, L., Bowerman, M., & Mandler, J. M. (1999). Early sensitivity to language-specific spatial categories in English and Korean. *Cognitive Development*, 14, 241–268. doi: 10.1016/s0885-2014(99)00004-0

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204. doi: 10.1017/S0140525X12000477

Colombo, J., & Bundy, R. S. (1983). Infant response to auditory familiarity and novelty. *Infant Behavior and Development*, 6(2), 305–311. doi: 10.1016/s0163-6383(83)80039-3

de Barbaro, K., Johnson, C. M., Forster, D., & Deák, G. O. (2015). Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother–infant–object interactions. *Child Development*. doi: 10.1111/cdev.12464

de León, L. (2000). The emergent participant: Interactive patterns in the socialization of Tzotzil (Mayan) infants. *Journal of Linguistic Anthropology*, 8, 131–161. doi: 10.1525/jlin.1998.8.2.131

Fantasia, V., Fasulo, A., Costall, A., & López, B. (2014). Changing the game: Exploring infants' participation in early play routines. *Frontiers in Psychology*, 5, 1–9. doi: 10.3389/fpsyg.2014.00522

Feldman, J., & Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language*, 89(2), 385–392. doi: 10.1016/s0093-934x(03)00355-9

Ferguson, B., & Waxman, S. R. (2016). What the [beep]? Six-month-olds link novel communicative signals to meaning. *Cognition*, 146, 185–189. doi: 10.1016/j.cognition.2015.09.020

Ferrier, L. (1978). Some observations of error in context. In N. Waterson & C. Snow (Eds.), *The development of communication* (pp. 301–309). London, England: Wiley.

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3-and 4-month-old infants: An advantage of words over tones. *Child Development*, 81(2), 472–479. doi: 10.1111/j.1467-8624.2009.01408.x

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology*, 61(6), 825–850. doi: 10.1080/17470210701623605

Friedman, O., & Leslie, A. M. (2007). The conceptual underpinnings of pretense: Pretending is not "behaving-as-if." *Cognition*, 105(1), 103–124. doi: 10.1016/j.cognition.2006.09.007

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in reason and language. *Cognitive Neuropsychology*, 22, 455–479. doi: 10.1080/02643290442000310

Gampe, A., & Daum, M. M. (2014). Productive verbs facilitate action prediction in toddlers. *Infancy*, 19(3), 301–325. doi: 10.1111/infa.12047

Gerson, S. A., & Woodward, A. L. (2014). Learning from their own actions: The unique effect of producing actions on infants' action understanding. *Child Development*, 85(1), 264–277. doi: 10.1111/cdev.12115

Gibbs, R.W. (2006). *Embodiment and cognitive science*. New York, NY: Cambridge University Press.

Gibson, J.J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton Mifflin.

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science*, 317, 1351–1354. doi: 10.1126/science.1144161

Gleason, J. B., Perlmann, R. Y., & Greif, E. B. (1984). What's the magic word: Learning language through politeness routines. *Discourse Processes*, 7(4), 493–502. doi: 10.1080/01638538409544603

Gleason, J. B., & Weintraub, S. (1976). The acquisition of routines in child language. *Language in Society*, 5(2), 129–136. doi: 10.1017/S0047404500006977

Glenberg, A.M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20(1), 1–55.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558–565. doi: 10.3758/bf03196313

Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2010). Trading spaces carving up events for learning language. *Perspectives on Psychological Science*, 5(1), 33–42. doi: 10.1177/1745691609356783

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. Garden City, NY: Anchor.

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby wordsmith from associationist to social sophisticate. *Current Directions in Psychological Science*, 15(1), 30–33 doi: 10.1111/j.0963-7214.2006.00401.x

Golinkoff, R. M., & Hirsh-Pasek, K. (2008). How toddlers begin to learn verbs. *Trends in Cognitive Science*, 10, 397–403. doi: 10.1016/j.tics.2008.07.003

Harris, P. L., & Kavanough, R. D. (1993). Young children's understanding of pretense. *Monographs of the Society for Research in Child Development*, 58(1), 1–91. doi: 10.2307/1166074

Harrist, A. W., & Waugh, R. M. (2002). Dyadic synchrony: Its structure and function in children's development. *Developmental Review*, 22(4), 555–592. doi: 10.1016/s0273-2297(02)00500-2

Heller, V., & Rohlfing, K. J. (2015). From establishing reference to representing events independent from the here and now: A longitudinal study of depictive practices in early childhood. *GESPIN Gesture and Speech in Interaction proceedings*, Nantes 2–4 September 2015, 143–148.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A.,... Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, 26(7), 1071–1083. doi: 10.1177/0956797615581493

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA: MIT Press.

Hsu, H. C., & Fogel, A. (2003). Stability and transitions in mother-infant face-to-face communication during the first 6 months: A microhistorical approach. *Developmental Psychology*, 39(6), 1061–1082. doi: 10.1037/0012-1649.39.6.1061

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 6(11–12), 19–40.

Jaffe, F., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. D. (2001). Rhythms of dialogue in infancy. *Monographs of the Society for Research in Child Development*, 66(2, Serial No. 265).

Jeannerod, M. (2006). *Motor cognition – what action tells the self*. Oxford, England: Oxford University Press. doi: 10.1093/acprof:oso/9780198569657.001.0001

Johnston, J. R., & Slobin, D. I. (1979). The development of locative expressions in English, Italian, Serbo-Croatian and Turkish. *Journal of Child Language*, 6, 529–545. doi: 10.1017/S030500090000252X

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K.E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental science*, 17(3), 388–395. doi: 10.1111/desc.12129

Kaye, K., & Wells, A. J. (1980). Mothers' jiggling and the burst-pause pattern in neonatal feeding. *Infant Behavior and Development*, 3, 29–46. doi: 10.1016/s0163-6383(80)80005-1

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226471013.001.0001

Leimbrink, K. (2010). *Kommunikation von Anfang an: Die Entwicklung von Sprache in den ersten Lebensmonaten* [Communication right from the start: Language development in the first months of life]. Tübingen, Germany: Stauffenburg.

Lerner, G., Zimmerman, D. H., & Kidwell, M. (2011). Formal structures of practical tasks: A resource for action in the social life of very young children. In J. Streeck, Goodwin, C., & C. D. LeBaron (Eds.), *Embodied interaction. Language and body in the material world* (pp. 44–58). Cambridge, England: Cambridge University Press.

Lewkowicz, D. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, 126(2), 281. doi: 10.1037/0033-2909.126.2.281

Liszkowski, U. (2014). Two sources of meaning in infant communication: Preceding action contexts and act-accompanying characteristics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369 (1651), 20130294. doi: 10.1098/rstb.2013.0294

Madole, K. L., & Oakes, L. M. (1999). Making sense of infant categorization: Stable processes and changing representations. *Developmental Review*, 19, 263–296. doi: 10.1006/drev.1998.0481

Mandler, J.M. (1998). Representation. In D. Kuhn & R. Siegler (Eds.), *Cognition, perception, and language* (pp. 255–308). Volume 2 of W. Damon (Series Ed.), Handbook of child psychology. New York, NY: Wiley.

Mandler, J.M. (2012). On the spatial foundations of the conceptual system and its enrichment. *Cognitive Science*, 36(3), 421–451. doi: 10.1111/j.1551-6709.2012.01241.x

Marentette, P, Nicoladis, E. (2012) Does ontogenetic ritualization explain early communicative gestures in human infants? In S. Pika & Liebal, K (Eds.), *Developments in primate gesture research* (pp. 33–54). Amsterdam, The Netherlands: John Benjamins.

McDonough, L., Choi, S., & Mandler, J. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology*, 46, 229–259. doi: 10.1016/s0010-0285(02)00514-5

Muir, D., & Field, J. (1979). Newborn infants orient to sounds. *Child Development*, 50, 431–436. doi: doi: 10.2307/1129419

Nomikou, I. (2015). *The collaborative construction of early multimodal input and its significance for language development* (Doctoral thesis, Bielefeld University, Germany).

Nomikou, I., & Rohlfing, K. J. (2011). Language does something: Body action and language in maternal input to three-month-olds. *IEEE Transactions on Autonomous Mental Development*, 3(2), 113–128. doi: 10.1109/TAMD.2011.2140113

Nomikou, I., Rohlfing, K. J., Mandler, J. M. & Cimiano, P. (submitted). Evidence for early comprehension of action verbs.

Nelson, K. (1974). Concept, word, and sentence: interrelations in acquisition and development. *Psychol. Rev.* 81, 267–285. doi: 10.1037/h0036592

Nomikou, I., Rohlfing, K. J., & Szufnarowska, J. (2013). Educating attention: Recruiting, maintaining, and framing eye contact in early natural mother–infant interactions. *Interaction Studies*, 14(2), 240–267. doi: 10.1075/is.14.2.05nom

Nomikou, I., Leonardi, G., Rohlfing, K. J. & Rączaszek-Leonardi, J. (2016). Constructing interaction: the development of gaze dynamics. *Infant and Child Development*, 25(3), (277–295).

Papoušek, M., & Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language*, 9(6), 137–157. doi: 10.1177/014272378900900603

Papoušek, M., & Papoušek, H. (1990). Excessive infant crying and intuitive parental care: Buffering support and its failures in parent-infant interaction. *Early Child Development and Care*, 65(1), 117–126. doi: 10.1080/0300443900650114

Papoušek, M., Papoušek, H., & Harris, B. J. (1987). The emergence of play in parent-infant interactions. In D. Görlitz & J. F. Wohlwill (Eds.) *Curiosity, imagination, and play: On the development of spontaneous cognitive and motivational processes* (pp. 214–246). Hillsdale, NJ: Erlbaum.

Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, 23, 128–733. doi: 10.1177/0956797612438734

Pauen, S., Birgit, T., Hoehl, S., & Bechtel, S. (2015). Show me the world: Object categorization and socially guided object learning in infancy. *Child Development Perspectives*, 9(2), 111–116. doi: 10.1111/cdep.12119

Peters, A. M. & Boggs, S. T. (1986). Interactional routines as cultural influences upon language acquisition. In B. B. Schieffelin & E. Ochs (Eds.), *Language socialization across cultures* (pp. 80–96). Cambridge, England: Cambridge University Press.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2012). Computational grounded cognition: A new alliance between grounded cognition and computational modeling. *Frontiers in Psychology*, 3

Piaget, J. (1954). *The construction of reality in the child* (M. Cook, Trans.). New York, NY: Basic Books. (Original work published 1937) doi: 10.1037/11168-000

Piaget, J., & Inhelder, B. (1993). *Die Entwicklung des räumlichen Denkens beim Kinde* [The child's conception of space]. Stuttgart, Germany: Klett-Cotta.

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, 106(2), 665–681. doi: 10.1016/j.cognition.2007.04.003

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development*, 77(2), 266–280. doi: 10.1111/j.1467-8624.2006.00869.x

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576–582. doi: 10.1038/nrn1706

Rączaszek-Leonardi, J., Nomikou, I., & Rohlfing, K. J. (2013). Young children's dialogical actions: The beginnings of purposeful intersubjectivity. *IEEE Transactions on Autonomous Mental Development*, 5(3), 210–221. doi: 10.1109/TAMD.2013.2273258

Rakoczy, H., & Tomasello, M. (2006). Two-year-olds grasp the intentional structure of pretense acts. *Developmental Science*, 9(6), 557–564. doi: 10.1111/j.1467-7687.2006.00533.x

Ratner, N., & Bruner, J. S. (1978). Games, social exchange and the acquisition of language. *Journal of Child Language*, 5, 391–401. doi: 10.1017/S0305000900002063

Reddy, V., Markova, G., & Wallot, S. (2013). Anticipatory adjustments to being picked up in infancy. *Plo S one*, 8(6), e65289. doi: 10.1371/journal.pone.0065289

Reddy, V., & Uithol, S. (2015). Engagement: looking beyond the mirror to understand action understanding. *British Journal of Developmental Psychology*, 34(1), 101–114.

Rohlfing, K. J., Wrede, B., Vollmer, A.-L., Oudeyer, P.-Y. (2016). *An alternative to mapping a word onto a concept in language acquisition: Pragmatic frames. Frontiers in Psychology*, 7.

Rossmanith, N., Costall, A., Reichelt, A. F., López, B., & Reddy, V. (2014). Jointly structuring triadic spaces of meaning and action: book sharing from 3 months on. *Frontiers in Psychology*, 5. doi: 10.3389/fpsyg.2014.01390

Sadoski, M. & Paivio, A. (2004). A dual coding theoretical model of reading. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 1329–1362). Newark, DE: International Reading Association.

Schilling, M., & Narayanan, S. (2013), Communicating with executable action representations. In *Proceedings of AAAI Spring Symposium Series 2013*, Stanford, CA. http://www.aaai.org/ocs/index.php/SSS/SSS13/paper/view/5772/5944

Slater, A., Quinn, P., Kelly, D., Lee, K., Longmore, C., McDonald, P., & Pascalis, O. (2010). The shaping of the face space in early infancy: Becoming a native face processor. *Child Development Perspectives*, 4, 205–211. doi: 10.1111/j.1750-8606.2010.00147.x

Sobel, D.M. (2007). Children's knowledge of the relation between intentional action and pretending. *Cognitive Development*, 22(1), 130–141. doi: 10.1016/j.cogdev.2006.06.002

Spranger, M., & Steels, L. (2014). *Discovering communication through ontogenetic ritualization*. In Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on (pp. 14–19). IEEE.

Stern, D.N. (1974). Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors. In M. Lewis &. L. A. Rosenblum (Eds.), *The effect of the infant on its caregiver* (pp. 187–213). Oxford, England: Wiley-Interscience.

Stern, D.N. (1985). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology*. New York, NY: Basic Books.

Strähle, P. (2013). *Emergenz globaler Diskursfähigkeiten im Rahmen von Begrüßungsroutinen: Rekonstruktion interaktiver Erwerbsprozesse* [Emergence of global discourse abilities in greeting routines: Reconstruction of interactive acquisition processes] (Doctoral thesis, Technische Universität Dortmund, Germany). Retrieved from https://eldorado.tu-dortmund.de/bitstream/2003/31172/1/Dissertation.pdf

Takada, A. (2005). Early vocal communication and social institution: Appellation and infant verse addressing among the Central Kalahari San. *Crossroads of Language, Interaction, and Culture*, 6, 80–108.

Tarplee, C. (2010). Next turn and intersubjectivity in children's language acquisition. In H. Gardner & M. A. Forrester (Eds.), *Analysing interactions in childhood. Insights from conversation analysis* (pp. 3–22). Chichester, England: Wiley-Blackwell.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

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

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech. The beginning of interpersonal communication* (pp. 321–347). Cambridge, England: Cambridge University Press.

Tulbert, E., & Goodwin, M. H. (2011). Choreographies of attention: Multimodality in a routine family activity. In J. Streeck, Goodwin, C., & C. D. LeBaron (Eds.), *Embodied interaction: Language and body in the material world* (pp. 79–92). Cambridge, England: Cambridge University Press.

Wagner, L., & Lakusta, L. (2009). Using language to navigate the infant mind. *Perspectives on Psychological Science*, 4(2), 177–184. doi: 10.1111/j.1745-6924.2009.01117.x

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological science*, 25(7). 1314–1324. doi: 10.1177/0956797614531023

Watson, J. (1985). Contingency perception in early social development. In T. Field & N. A. Fox (Eds.), *Social perception in infants* (pp. 157–176). Norwood, NJ: Ablex.

Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology*, 5, 506. doi: 10.3389/fpsyg.2014.00506

Werner, H., & Kaplan, B. (1963). *Symbol formation*. New York, NY: Wiley.

Wojcik, E.H. (2013). Remembering new words: Integrating early memory development into word learning. *Frontiers in Psychology*, 4, 151. doi: 10.3389/fpsyg.2013.00151

Woodward, A. (1999). Infants' ability to distinguish between purposeful and nonpurposeful behaviors. *Infant Behavior and Development* 22, 145–160. doi: 10.1016/S0163-6383(99)00007-7

Authors' addresses

Iris Nomikou
CITEC
Bielefeld University
Inspiration 1
33619 Bielefeld
Germany

iris.nomikou@uni-paderborn.de

Malte Schilling
CITEC
Bielefeld University
Inspiration 1
33619 Bielefeld
Germany

mschilli@techfak.uni-bielefeld.de

Vivien Heller
CITEC
Bielefeld University
Inspiration 1
33619 Bielefeld
Germany

vheller@uni-wuppertal.de

Katharina. J. Rohlfing
Paderborn University
Warburger Str. 100
33098 Paderborn
Germany

katharina.rohlfing@uni-paderborn.de

Biographical notes

Iris Nomikou received her Dipl. degree in Translation Studies from the Ionian University, Corfu, Greece in 2006 and her M.A. degree in Linguistics from Bielefeld University at Bielefeld Germany in 2010. She then obtained her PhD within the project "Symbiosis of Language and Action," which was supported by the Volkswagen Foundation. She then worked as a Postdoctoral researcher at Center of Excellence for 'Cognitive Interaction Technology' (CITEC) of Bielefeld University. She is currently a Postdoctoral researcher at Paderborn University. Her interests include language development and particularly the interactive foundations of language learning. Her research focuses on the interplay between language and bodily experience for the development of meaning.

Malte Schilling is a Responsible Investigator at the Center of Excellence for 'Cognitive Interaction Technology' in Bielefeld. His work concentrates on mental representation their grounding in behavior and application in higher-level cognitive function like planning ahead or communication. Before he was a PostDoc at the ICSI in Berkeley and did research on the connection of linguistic to sensorimotor representation. He received his PhD in Biology from Bielefeld University in 2010 working on biologically-inspired minimal cognitive systems. He has studied Computer Science at Bielefeld University and finished 2003 the Diploma with his thesis on knowledge-based systems for virtual environments.

Vivien Heller received a grant from the international Research School Education and Capabilities and obtained her doctorate in linguistics from the University of Technology Dortmund. She was a post doctoral researcher in the CRC 673 'Alignment in Communication' at the 'Center of Excellence for Cognitive Interaction Technology' (CITEC). Since 2016, Vivien is professor of German linguistics and language education at the University of Wuppertal where her research focuses on the acquisition of discourse competence in everyday and educational settings, multimodal interaction, and the interrelation between discourse practices and subject matter learning. She uses insights from Conversation Analysis to inform the modeling and analysis of acquisitional mechanisms and processes.

Katharina J. Rohlfing received the Master's degree in linguistics, philosophy and media studies from Paderborn University in 1997. As a member of the Graduate Program Task Oriented Communication, she received the Ph.D. degree in Linguistics from Bielefeld University in 2002. In 2006, with her interdisciplinary project on the Symbiosis of Language and Action, she became a Dilthey-Fellow (VolkswagenStiftung). From 2008 to 2014, she was Head of the Emergentist Semantics Group within the Center of Excellence Cognitive Interaction Technology. Since 2015, she is professor of Psycholinguistics at Paderborn University.