Young children’s dialogical actions: The beginnings of purposeful intersubjectivity

Joanna Rączaszek-Leonardi, Iris Nomikou, and Katharina J. Rohlfing

Abstract—Are higher-level cognitive processes the only way that purposefulness can be introduced into the human interaction? In this paper, we provide a micro-analysis of early mother-child interactions and argue that the beginnings of joint intentionality can be traced to the practice of embedding the child’s actions into culturally shaped episodes. As action becomes co-action, an infant’s perception becomes tuned to interaction affordances.

Index Terms—intentions, intersubjectivity, social interaction, cognitive development

I. INTRODUCTION

An increasing number of studies confirm the existence of low-level automatic mechanisms that allow humans to imitate and synchronize their activities [1], [2]. The importance of these mechanisms is demonstrated both in simple physical interactions and in complex cooperative and communicative events in which participants align both their movements and verbal actions [3], [4]. However, for purposeful, intentional coordination, higher-level cognitive mechanisms are usually evoked, such as “understanding” the actions of others and “representing” them in terms of others’ intentions [5], [6] or predicting the other’s behavior on the basis of “internal models”. The reason for such a theoretical strategy is probably the fact that it is indeed difficult to explain the emergence of this type of coordination just from low-level imitation and alignment. However, before resorting to a dichotomous solution, one should ask whether higher-level cognitive processes are the only way that purposefulness can be introduced into human interactions. In this paper, we present a micro-analysis of early mother-child dialogues, in which we show how the child’s actions are being incorporated into dialogical events. We argue that, through participation in these shared events, the infant’s actions become parts of meaningful wholes, in which these actions have specific causal consequences. This, in turn, educates the infant’s perception toward picking up cooperation opportunities.

The problem of the purposefulness of an interaction conceptually resembles another problem that has been recognized for decades in cognitive psychology: that of the motor control of a purposeful, individual action. Whereas coordinated perception and movement in an environment is ensured by low-level perceptual integration and effector synchronization, the goal-directedness of a movement – within the information processing approach – is provided by a ‘motor program,’ (i.e., a high-level cognitive representation) [7]. Thus, the representational approach proposes that individual actions are driven by top-down processes that account for the functionality and intentionality of movements and deeds, and collective actions are similarly treated as based on mind-reading capacities, simulations, and the construction of predictions of the others’ behavior.

However, in the case of explaining individual actions, valuable alternatives to the top-down, information processing approaches have been introduced. The ecological psychology approach [8] – [10], [11] suggests the inseparability of action from perception and proposes that organisms perceive the environment in terms of affordances. Affordance is a relational concept that involves specifying both the perceiver/actor and the environment in which s/he acts [10]. Perception is calibrated and educated to pick up from the environment “what it offers the animal, what it provides or furnishes, either for good or ill” [10, p. 127]. Thus, the objects’ involvement in the process of tuning purposeful activity links them meaningfully to actions. As many authors, most notably Heft, note, the notion of affordance actually introduces intentionality into the objects’ perception [12]. What guides behavior toward an object is thus specified both by the capacities and history of an organism acting in the environment and by the information directly picked up from the world.

Drawing a parallel between the problem of control of individual and collective actions, we will follow the above line of reasoning on the goal-directedness of individual actions in the environment and apply it to the mechanisms that give rise to the shared intentionality in interactions. “Shared intentionality” refers to human participation in collaborative
activities involving shared goals and socially coordinated action plans [6]. In developmental research, it is still unclear how children become aware of others’ action plans and goals, however, most researchers hold a view ascribed to the “Dogma of Autonomy” [13, p. 227], which implies that intentions are a form of mental objects that ‘reside’ in an individual and can be ‘shared’ during an interaction. In contrast, in our alternative approach, we view shared intentionality not as objects, but as acts, cf. [13] [14]. Accordingly, the early forms of a dialogue are co-actions that generate dialogic states rather than individual representations. In this sense, we propose that early forms of intentionality arise from initial, perhaps automatic, ‘moving with others’, which is sculpted in multiple social episodes to become ‘acting with others.’ Within this view, actions acquire both individual and systemic (dyadic) meaning, cf. [13]. Thus, although mere synchrony and imitation is clearly not sufficient for interindividual coordination to become purposeful, the evoking of abstract higher-level cognitive representational processes may not be necessary. Instead, the basis for purposeful intersubjectivity may lie in the shaping of individual behaviors into meaningful events through repetitive interactions. As our examples from early mother-child interactions show, the main mechanism for this shaping is the education of action-perception to enable the child to pick up and create interactive affordances. This process instills modes of engagement with others in purposeful acting (as a system) within environmental contexts.

II. BASIC INTERACTIVITY AND ITS LIMITATION

A. Basic adaptation for social life

A fundamental difference between acting toward inanimate objects and acting with conspecifics is the reciprocity of action. Recent years have brought substantial confirmation to earlier claims that low-level automatic mechanisms allow humans to focus on, mimic, and synchronize their activity [15] – [21], [1] – [3].

These imitative and synchronous properties of human behavior are present in newborns, which points to the biological preparation for at least some of the basic interactive mechanisms [22], [15], reflecting a specific adaptation for social interaction [23]. It has been suggested that both the function of imitation [24], and more obviously, the function of temporal synchronization, for example [25], go far beyond simple practice of novel behavioral forms into the realm of interpersonal regulation [26]. The question arises as to what these functions might be.

In the traditional cognitive psychology literature, these propensities, particularly the imitative one, are said to aid ‘understanding each other’; by simulating and recreating in our own body the movement of the other person, we can better perceive the movement itself and link it to possible intentions [27], [28]. However, in early development research, these skills are also cast in terms of forming a less cognitive and more emotional link between an infant and a caregiver (see [22] for a review). Thus, automatic imitation and synchronization are mechanisms for ‘locking’ onto the other, which allow individuals to ‘enter’ the dyadic system. They serve as a flexible ‘glue’ for establishing and sustaining contact. The state of synchronizing may be rewarding in itself, which is indirectly supported by the visible distress of very young infants in the case of a delayed [20] or missing [29] maternal response.

Thus, we accept that low-level coupling mechanisms exist that help individuals enter interactions and maintain or stabilize these interactions. However, we also claim that these mechanisms are not sufficient to make the interaction ‘meaningful’.

B. Beyond basic interactivity: Making the interaction meaningful

The mere ‘coupling in time and activation’ is simply too unspecific to fit the coupled system to environmental demands. In other words, for a child-caregiver dyad to realize a meaningful interaction (both meaningful for each other and meaningful in the sense of realizing a common goal), a simple synchronized mirroring will not suffice (for the same argument concerning adult communication through language, see [30]). This basic coupling must be ‘sculpted’ by the demands of the situation or task at hand.

Some researchers who acknowledge the child’s input into structuring the interpersonal interaction from the earliest moments of life claim that the functionality originates from innate mechanisms that provide much more than general imitation and synchronization. For example, the innate ‘motives’ of [22] appear to be such functional units of cooperation: already narratively shaped for structured interaction. However, it also appears plausible that the structure of this interactivity ‘sneaks in’ through very early encounters between the baby and her/his con-specifics, which flexibly distribute roles in the interactive episodes.

In explaining how the joint actions of an infant-caregiver dyad may acquire functionality through relevant structuring, we refer to the notion of affordances. First we remind how in the case of individual actions toward objects this notion has been used to explain (i) the interdependence of action and perception [9], [10] and (ii) the intentional character of perception [12]. We then show that in the case of joint actions, this notion is crucial for understanding how purposefulness emerges in development at the systemic level and how it is created and upheld by individual behaviors.

III. PHYSICAL, CULTURAL, AND SOCIAL AFFORDANCES: FROM ‘BEHAVING TOWARD’ TO ‘ACTING WITH’

A. Affordances and intentionality

As mentioned previously, the concept of affordance was introduced by Gibson [31], [10] in an attempt to express the inseparability of the organism from the environment in the explanation of perceptual processes. Due to the evolutionary tuning in of the perceptual systems, these are able to directly perceive the possibilities for actions, i.e., perceive objective environmental features in relation to a particular body acting in that environment. This concept is important because it deals with the problem of the meaning of environmental structures.
and events without making the mind of the perceiver solely responsible for this meaning.

Gibson’s approach is attractive because it does not rely on the controversial concept of mental representation; however, it was criticized for its suitability to account merely for the immediate automatic reactions to the objects in the natural environment [32]. More recently, Heft [12] dealt with the question of how this framework could explain culture-based action possibilities (affordances) and how to account for the possibility of choosing an action toward a single object according to a current goal.

Heft [12] claimed that it is misleading to think of affordances as causing or eliciting behavior. Instead, these constrain actions that may be produced. Thus, affordances are not defined just with respect to a body shape or its dimensions. Heft reminds, quoting Merleau-Ponty: “having a body is [...] to identify oneself with certain projects and to be continually committed to them” [33, p. 82]; quoted in [12]. This intentional approach to the body allows us to understand affordances in relation to intentional acts: they are those environmental features that are implicated in goal-directed behaviors. The meaning of an object for an organism is defined both by the environmental possibilities and by its placement in an intentional act. Because it is possible to shape intentional acts (i.e., to add to the intentional repertoire), culturally meaningful objects can also become affordances. Accordingly, the time-scale of ‘in-forming’ (shaping) the perceptual systems can be shortened from an evolutionary one to the cultural and/or experiential. The state of an organism and the history of interaction with a given object modify the shaping processes occur. The enculturation process involves broadening the “intentional repertoire” of acts in a culture-specific way [12, pp. 18-19].

Although it was the physical reality that was of primary focus for Gibson and Heft, one might risk a statement that the shaping of the intentional repertoire occurs first not toward objects, but toward, or rather – as we will shortly see – with people.

B. Affordances of living organisms — ‘behaving with’

In this section, after introducing the concept of affordances for intentional actions, we return to our central question: How can purposeful joint intentionality arise in an interaction among people. There is a crucial difference between acting that involves objects and acting that involves other members of the species. As mentioned previously, actions specify parts of the environment: “the act of grasping is only comprehendible in relation to a thing that can be grasped” [12, p. 6]. Similarly, one may state that the act of smiling is comprehendible in relation to a thing that can be smiled at. However, the ‘thing’ is not immobile: unlike inanimate objects, ‘it’ can smile back. Once it does this, ‘it’ is smiled with. Thus, the key difference is that the participant in an interaction is both a source of affordances for the other and a user of affordances made available by the other.

This interactional perspective opens a way to extend the ecological framework to encompass purposeful co-actions. We point to the possibility that, just as intentional individual action (purposeful subjectivity) arises as a result of tuning a perceiving/acting body to the physical constraints of the environment (objects and forces), acting with others requires shaping sensitivities to the behaviors of others as parts of one’s intentional act. This prepares a body for ‘acting a part’ in a coordinated system.

It now has to be demonstrated how the repertoire of intentional acts of a child is broadened in the social realm and how this shapes social affordances. In the case of affordances of the inanimate world, Heft suggested that this happens mainly through observing others and instruction. However, in the case of interactions with other agents, this statement should be qualified in two important aspects.

First, because intention refers to possibilities instantiated in an actual situation, forming social affordances requires actual behaving with a child in such a manner that the future behaviors of others will be perceived as possibilities for action [25]. Participation (not just observation or instruction) is crucial: the actual performance of acts to be molded into intentional ones by being embedded in an interactive structure.

Second, another important specificity of the interactive situation is that interaction affordances are also created and made available for the other. Thus, the repertoire of acts is shaped in interactions both with respect to affordances of the other and as affordances that are displayed to the other. In this way, we learn the meaning of our own acts. Once they are shaped, they can be adjusted on-line, which results in dynamical engagement. As mentioned above in section II, the ‘glue’ of automatic imitation and synchronization helps the establishment and maintenance of this engagement; however, it is the perception of other’s behavior in relation to own intentional acts that can make this engagement functional and specific to a goal. The functionality of a dyad’s action with respect to external goals is preserved if one remembers that the individual’s behavior in an interaction does not cease to be constrained by the environmental possibilities for actions. Yet due to the developmental shaping in interaction as described above, the presence of others and co-acting with them is able to change the perception of what the world affords here and now [34].

This ‘togetherness’, i.e., the dynamical engagement in an interaction, is not often noted, even within the ecological perspective. For example, in an excellent ecological analysis of the development of purposeful behavior from a dynamical systems perspective, Fogel and Thelen [35] treated affordances in social interactions as constraints on an individual’s expressive behavior toward others and not as possibilities for engagement that enable the emergence of a new systemic unit. Here, we argue that social behavior is not behavior toward, but behavior with. Intersubjectivity is scaffolded by imitation and synchronization, but just like in the case of individual actions constrained by the physical world, the meaning is given by tuning to the social context. However, in the case of other people, the external context is...
co-created: a child does not learn to behave ‘toward’ a caregiver but rather ‘with’ a caregiver, i.e., the child learns to be a part of a relational whole (see also [36] and [37] for an enactive view, which similarly underscores actual interaction).

It is important to underscore the cultural origin and cultural specificity of the acquired repertoire of intentional acts. In this way, culture shapes the engagement from day one. Already 14-week-old children use “culture-specific biomechanics” [38]. Ways to culturally co-regulate interaction develop as the infant’s behavior “meshes with caregiver use of cultural dispositions to (over)interpret events” [39, p. 114].

IV. FROM INTERACTION TO LANGUAGE

A. Indexicality and repetitive interactions.

In the previous section we proposed that joint intentionality emerges by participating in interactions in which social affordances are shaped. The next step is to consider how this interactive view can help construe the beginnings of language use. Although we have argued that a co-action is crucial for meaning to emerge during an interaction, language consists of symbols that direct perception and action toward events and objects in the world. For this, the interaction needs to shift from reciprocal to indexical. The function of indexicality is to relate to the person-space-time-structure of a produced communicative act [40]. In early interactions, participants share space as well as time, thus they follow each other’s attention to work on a joint goal. In this framework, each other’s acts are interpreted in relation to the joint goal. Liszkowski [41] argued that the communicative goals of affecting a recipient’s body (such as a child’s early pointing) are very important steps on the road to intentional communication aimed at affecting the recipient’s mind. Thus, indexicality is considered to be a necessary step from a more situated to a more abstract meaning. Tomasello [42] as well as Marcos [43] suggested that ritualization (i.e., repetitive interaction) helps children learn the forms of indexicality (pointing, eye-gaze, etc.) as gestures of mutually shared attention. Bruner [44] also suggested that such repetitive “predictable formats of interaction” (p. 31) enable the emergence of rules and the development of conventions (see also [43]). As was shown by Nomikou et al. [45], early joint actions can develop into routines that are then applied as means for communicative purpose, and young infants learn to process those means more rapidly. Bruner [44, p. 39] suggested that this is the “vehicle” toward language: As the formats become increasingly conventionalized, they develop in ways that are less idiosyncratic and indexical and more recognizable to a broader cultural community.

In the examples below, we attempt to show that training to do things together begins very early, through learning what the adult’s behaviors afford in the interaction and, perhaps more importantly, through experiencing in interactive episodes that one’s own (the infant’s) behavior can be an affordance for the other.

In the following study, we ‘slow down’ real interactions to analyze them in a search for how mothers embed infants’ behavior within meaningful events. We hypothesize that in this embedding, the mothers select the babies’ behaviors to become affordances for their (mothers) actions. This sensitizes the infants’ perception of their own actions as social acts. Additionally, mothers create slots for children’s appropriate response to their (mothers) behavior. This sensitizes the babies’ perception to the mothers’ behaviors as social affordances. Through repetitive interactions, these two educative processes instill flexible interactional frames with a specific, dialogic time structure that become purposeful acts and indexical patterns that drive the verbal exchange.

Before presenting the microanalysis, a word of warning is due about the methodology of this study. While this research is observational in nature, we do not aim to stop at this illustrative level. However, the shift in the theory that is proposed here calls for new observables that we do not believe are firmly established in the field of study of early interactions. This work is thus considered a step on the way to the development of new categories that would aid the classification of behaviors for future quantitative analyses.

V. MICROANALYSIS OF EARLY INTERACTION

We followed a methodological approach that allowed us to analyze behavior as it is constructed within the interactive process [46]. We zoomed into the microstructure of interactional sequences and observed behavior as it occurs within a concrete situation or “contextual configuration” [47, p. 1490] and thus collected single instances of the phenomena expected on the basis of the above theoretical framework, which we made visible through detailed qualitative micro-analysis.

A. Method

In the following, we present selected case examples originating from a corpus of video-recordings of mother-infant interactions that were collected outside laboratory conditions.

B. Participants

The participants were recruited through a number of different visits to mother-infant classes in Bielefeld (North Rhine-Westphalia, Germany). The corpus includes seventeen mothers and their infants. The examples presented herein were recorded during two visits to the families: one when the infants were between 14 and 16 weeks of age and a second visit 12 weeks later, when the infants were between 26 and 29 weeks of age.

C. Procedure

The mother and infant dyads were filmed at their homes during the common and familiar activity of changing the infant’s diaper (for further information on these data, see [48], [45]).

D. Analysis and Interpretation

Example 1: CUBE

In an example of a mother and her 3-month old infant, at the beginning of the sequence, it is evident how the mother
includes the infant’s own behavior in the activity. She does it by commenting on the infant’s action (see transcript line 2) as it becomes the focus of the interaction. Although we cannot suppose (and do not suggest) that the infant understands what the mother is talking about, her utterances are evidence of her other-orientation, which informs her behavior in the situation.

1 I: <vocalization>

2 M: immer alles in den Mund, ne?\(^a\)
   always everything in the mouth = huh?
3 (0.9 s)
4 M: so (.) mach immer alles rein da
   so (.) put always everything in= there
5 (1.2 s)

Until this point, the infant has been gazing up toward the ceiling (image a). In line 6 of the transcript, the infant turns his head to the side. The mother, who is monitoring the infant’s behavior, follows the infant’s gaze and looks in the same direction (image b). This shift in the infant’s gaze is sufficient to open a new topic of the interaction. In her work, Filipi [49] suggested that gaze is the first resource available to infants for communicating their interest and attention. The mother reacts to the gaze and repeatedly asks the infant what he is looking at (lines 6–10).

6 M: \(^b\)ja was ist denn da? (0.4 s) was= ist denn da?
   Yes what is there? (0.4 s) what is= there?
7 (0.3 s)
8 M: Mäuschen (1 s) hm?
   little mouse (1 s) huh?

9 (0.4 s)
10 M: was ist denn da? (..) ist das ein= Würfel?
   What is there? (..) is that a cube?

Interestingly, after the pause in line 10, she continues interpreting the direction of the infant’s gaze. She asks the infant if what he is looking at is a stuffed toy in the form of a cube, thus providing the label for the object of the infant’s attention. Following this, the mother repeats the labeling of the object (line 11), whereupon she simultaneously actually grasps, shakes (image c), and repositions the object at a higher location (image d).

This example is interesting for the following reasons: It shows how the interaction is structured in a way to incorporate any behavior of the infant. The mother is not only monitoring the behavior of the infant and is responsive to his eye-gaze, but she is also interpreting the infant’s gaze as the infant’s intention to communicate about an object. By being responsive to the infant’s gazing action, she enables the infant to actively affect the development of the interaction sequence. In this way, the child’s ‘random’ behavior becomes meaningful in the dialogical event: it becomes causal in this situation because it affects the change in the direction of the mother’s gaze and even her behavior toward the object. Thus, the mother is acting as though the infant’s gaze were an intentionally produced affordance for her action. By doing so, she not only ascribes intentionality to the infant, but this intentionality also constrains her own actions.
In the next example, the tactile behavior of the infant is treated as intentional. In this case, the infant is given a functional role in the completion of a joint task.

**Example 2: ASSISTANCE**

This is an example of a mother with her six-month-old infant. The sequence begins with the infant being undressed on the changing table. He is looking to the side and moving his arms and legs (image a).

![Image a](image-a)

During these movements, he touches the diaper with one hand (image b). The mother responds to this action by placing it in the infant’s hand (image c). She responds to the infant touching the diaper as if he had expressed a wish or intention of holding it (line 3) as she starts her utterance by acknowledging the preceding action with a “yes”.

![Image b](image-b)

3 M: ja du kannst ja schonmal die Windel halten
Yes you can start by holding the diaper

The infant then grasps the diaper (image d) and starts moving it toward his face. The mother again embeds this action of the infant by treating it as a part of their joint action of changing the infant’s diaper. In line 6, she provides an explanation for the infant’s supposed intentional action, namely that he can now actively participate in the task as her assistant.

![Image c](image-c)

![Image d](image-d)

4 M: hier halt mal die Windel (..)= genau
here hold the diaper (..) right
5 (0.4)
6 M: dann kannst du mir gleich assistieren
and then you can assist me

In this example, the produced behavior of the infant is treated as having a function; the infant is considered to have an active role in the unfolding event. Here, the ascription of intentionality constrains not only the action of the mother but also that of the infant, as he is given an object to manipulate.

In both this and the earlier example, the infants are experiencing that their own behavior is being contingently responded to. The role of contingent responsiveness has been considered central for the development of early protocconversation [50] because it is said to foster the temporal
coordination of the mother and infant [51] and support the active interactional participation of the infant, which leads to reciprocal communication [52]. Moreover, contingent responsiveness is about learning to exchange different roles in an interaction [53] and the alternation of being active and inactive. However, we, among others, suggest that it is not merely about temporal synchronization and alternation but about complementarity [54]. The movements of participants in an interaction are complementary if they match each other as parts of a bigger whole. This also points to the functionality of the interaction sequences: a response can only be complementary to another with respect to a systemic function [30]. Accordingly, in the first example, the infant looks in a certain direction, and something happens in that region; in the second example, the infant touches something, and this object is then given to him to hold. We suggest that this goes beyond simple contingency because the action produced by the infant constrains what the response to this action should be, i.e., what this action affords as a next action. We propose that in these very short interaction sequences, the behavior of the infant ‘makes sense’, becomes meaningful, as mother and infant are acting together as parts of a coordinated system. We propose that this coordination develops because the child is treated as an active participant of the interaction. Not only is he or she given an opportunity to act, but also the action produced is ‘contextualized’ and responded to as intentional and meaningful [14], [55].

For infant vocalizations, Snow [56] suggested that it may help the infant realize that vocal behavior is effective, which would thus pave the way to the acquisition of language. We propose that already for young infants, participating in a goal-directed activity – in which the infant is treated as an intentional agent who is able to influence the outcome of the activity – may be very powerful in creating shared meaning, as it is illustrated in the following example.

Example 3: SHOE
In this sequence with a three-month-old baby, the mother is putting on the infant’s shoes. She first announces her action (line 1).

1 M: jetzt ziehe ich dir noch die = Schuhe an
   now I will put your shoes on too

2 (0.5 s)

She holds the shoe (image a) and then pauses her movement and looks at the infant’s feet. The infant kicks with his feet, during which action he stretches one foot a little higher than the other (image b).

3 M: so

The mother selects this foot as the one to put the first shoe on (images c and d). In doing so, she is treating the infant’s movement as his decision on how they should carry out the specific task. This is also made evident by the fact that she then asks a clarification question, as if she wants to make sure that she has understood his decision or proposal correctly (line 5).
The sequence continues with the mother tying a bow with the shoe’s laces as she verbalizes her action. Then, some seconds later, it is the second shoe’s turn. The mother holds the shoe and waits as the infant shakes his legs. She again pauses her movement and looks up at the infant (image c).

She takes the other foot (image f) to put on the other shoe (image g). In this sequence, the structuring role of the mother becomes evident. She repeats the exact sequence of actions as before: she stops and waits for the infant, looking up at his face, as if expecting him to do something, thereby creating a ‘slot’ for his action within this frame. This can also be observed in her accompanying utterance (line 12) in which she tells the infant how this was supposed to work, namely that he is supposed to assist her by stretching the other leg. Only then does she put on the second shoe.
This sequence exemplifies two interesting aspects: First, the infant’s behavior is treated as originating from his own agency relevant to the situation. The infant’s movements are interpreted as proposals or decisions that influence the outcome of the interaction. This is consistent with the view of communicative agency, which suggests that young infants learn to identify goal-directed actions and intentional agents because they witness agency within communicative actions, e.g., observe presentations of actions addressed to them in infant-directed ways [57]. We take this a step further and propose that apart from the observation of demonstrations, the active participation in interactions with caregivers may also serve as an additional ostensive cue that educates the infant in becoming an intentional agent. The infant’s stretched leg is selected as the goal of the mother’s subsequent action, and the infant is also provided with a time window, i.e., a ‘slot’ inviting his reaction as the mother stops her activity and looks at him.

The second aspect is the emerging structure of the activity, which is made evident by the mother’s repetition of the same action sequence when she wants to put the second shoe on. Having established a ‘frame’ in the sense of a recurrent interactive structure [58, p. 171], the mother uses it in her next action. Regularities emerge from the interaction itself and constrain the way in which the next action is to take place. From this repetition, expectations may emerge: not of random behaviors (e.g., the infant kicking his legs) but rather specific ones (i.e., after stretching one leg, one must stretch the other). Over time, this may enable the emergence of conventionalized behaviors, which would lead to interpersonal coordination, enculturation, and even language.

However, this process is gradual and involves the shaping and structuring of joint actions into experiences of shared intentionality. This is achieved by constraining interactions in specific ways, as elaborated in the following example.

**Example 4: FERTIG**

This sequence is the last in the particular interaction. The mother has finished diapering and marks the end by saying ‘finished’, elongating the vowels in the utterance (line 1) and making a large gesture with her hands while repeating the same utterance (image b and line 2).

What is interesting in this interaction is that the mother is apparently expecting a specific reaction from the infant, which she is not getting, and this becomes evident in her subsequent pursuit. The pursuit of a particular response has been suggested to be an educational tool used by parents in conversations with children learning how to speak [59]. Here, we can observe how this process of learning to participate in conventionalized interactions may have its outsets.

Throughout the entire sequence, the mother is looking at the infant’s face, but it appears that she cannot get him to look at her. Thus, in the beginning of the sequence, when she says ‘finished’, the infant is looking at her hand (image a). She then repeats her utterance adding the hand gesture, but that still does not attract the infant’s attention to her face: he follows the movement of her hand (image b). Having no success, she attempts again, this time calling the infant’s name (line 4); but, as marked in image c with a grey arrow, the infant is now looking away.
In her next pursuit, she taps the infant on the belly while calling him with his full name (line 5 and image d). At this moment, she first succeeds in establishing eye-contact (image d); therefore, she returns to her original utterance (line 6), but she has already lost the infant’s gaze again (image e).

Having lost the infant’s attention, she moves her upper body into his visual field (image f) and continues with a greeting (line 8). Having recruited his attention, she repeats her original phrase, concluding with a smile (image g). This conclusion is further marked by a modified intonation of the repeated phrase. The infant smiles too (image g) and only then does she end the interactional sequence and reaches to lift the infant from the changing table (image h).

This sequence illustrates the way in which mutuality is dynamically negotiated. This exemplifies the process of co-shaping mutual events as emerging both during the course of a particular interaction sequence and over the interactional history of mother and infant [60]. Here, we can see that this dyad has established a repertoire of interactional formats such that not any response of the infant is treated as relevant, but a specific action sequence is expected for the interactional episode to be successful. In this case, the mother and infant have to mutually acknowledge the termination of the interaction by looking and smiling at each other. We propose...
that the mutual acknowledgment could explain how, through interaction, the infant may move from shared-intentional to conventionalized communication, as is also illustrated in the example below.

**Example 5: HELLO**

This sequence with a six-month-old baby begins with the mother spreading her fingers to touch the infant’s extended arm and fingers (image a).

1 M: kleiner oder großer Räuber?  
small or big thief?

In line 2, the mother produces a prolonged interjection ‘oh’ while grasping the infant’s open palm and pulling the arm up (image b).

2 M: o:u:

She then freezes the movement and pauses (line 4). She repeats the same greeting in a more exaggerated manner, prolonging the duration of the first syllable (line 5). At the same time, she grasps the other arm of the infant and moves it in a waving movement (image d).

3 M: c) h a:lo

hello

5 M: ha:a) llo

hello

In this sequence, we can see how a coordinated event, such as the touching of the mother’s and infant’s hands, is interpreted and turned not only into a greeting sequence, but more importantly, a sequence of paired actions that is conventionalized, as the mother is acknowledging the infant’s hand movement as a conventionalized gesture.

**VI. SUMMARY OF THE FINDINGS**

At the beginning of our paper, we proposed that acting with others requires shaping sensitivities to the behaviors of others as parts of one’s intentional act. This prepares an organism for ‘acting a part’ in a coordinated system. We then asked how the ‘shaping of social sensitivities’ occurs. We hypothesized that infants’ perception becomes shaped socially by the babies’ own movements being treated as responses to and at the same time affordances for mothers’ behavior (by mothers enacting...
the consequences of a child’s movement.

In this respect, with the above-presented case examples, we have shown how early interactions are structured in a way that incorporates the behaviors of the infant. These interactions are not structured by the mother’s conscious effort to educate a baby by changing its internal representations. It is the continuous re-enacting of parts that carries the “play” through time. No script is written here, and none is represented in anyone’s head. The child’s observable behavior becomes meaningful for the dialogical event as the mother (i) treats it as intentional and considers it causal in a particular situation, as it effectuates the change in the direction of the mother’s gaze and the manipulation of an object (as in example 1); (ii) embeds it into a goal-directed activity and assigns the infant a functional role for the completion of the joint activity (as in example 2); (iii) interprets it as a gesture within a conventionalized social frame (such as a greeting in example 5). In this sense, the mother is acting as though the infant’s behavior was an affordance for her action. The crucial point is not only that the mother provides contingent behavior but also that the infant himself constrains what the mother’s response to this action should be. The responses are then generated from the intentional repertoire of the mother’s action. Complementarily, the mother’s response (or better, her interpretation) provides the infant new action constraints (e.g., having the diaper put in his hand in example 2 or the touch of the mother’s hand in example 5). Thus, the emergent ‘togetherness’ does have a power to change perception. This may be the developmental mechanism through which “(...) the presence of another person extends the action possibilities” [61, p. 326]. Being together modifies what is perceived as affordance [34].

Furthermore, with the case examples presented above, we have shown how the process of educating an infant to purposeful intentionality may take place. In examples 3 and 4, we see that within the course of the interaction, the dyad establishes a repertoire of interactional ‘frames.’ Within these repeating structures, the infant is given a ‘slot’ in which he or she is expected to act or respond. As a result of the history of the interaction, both within one interaction episode and within a developmental time, not just any response of the infant is treated as relevant. The correct response needs to be negotiated (example 4). Thus, through recurrent interactions, the set of acceptable responses is narrowed.

Finally, based on examples 1, 2, and 5, we speculate that over time, the mother-infant system moves from shared intentionality to indexicality and from there to conventionalized forms of communication. In these examples, we can see how in an early interaction the gaze of the infant, the touch of the infant as well as the movement of his hand are treated not only as intentional but also as indexical means, such as pointing or referring to something that he or she is interested in or wishes to communicate about. We assume that although in the early interaction, many actions performed by the infant will be interpreted as indexical or intentional, with the infant’s development and experience in interactions with others, only specific means will become established (see also example 4). We see such selections in e.g., giving priority to the perception of the face in 6 month-olds [45]: whereas 3 month-olds focus on many bodily behaviors that occur during an interaction, 6-month-olds seem to perceive the face more rapidly. In addition, mothers significantly increase the use of the face and decrease other modalities. We propose that through recurrent interaction and co-constructed repetitive interactional formats (examples 3, 4, and 5), conventionalized means of communication, such as language, may emerge (cf. [45]).

VII. CONCLUSIONS

The microanalysis above showed that it is possible to achieve the purposeful coordination of an interacting dyad via an ecological psychology path and that one can go a long way toward explaining shared intentionality without either referring to higher-order cognitive representations or proposing elaborate innate motives.

We think that there are advantages of such a perspective over both the nativist and the representationalist approach: educating sensitivity to others’ behavior in terms of interactive affordances leads to a more flexible coordination than having a specific innate ‘motive’ or a definite cognitive representation (script) of a possible encounter. Within the view presented here, the interaction is structured online both because attention is educated (in the sense of Gibson [9]) to perceive the affordances for action and co-action online and because participants adjust in real time the affordances for others. This and the simultaneous coupling with the environment leave more details to be specified by an unfolding situation, which confers plasticity to the system.

We also suggest (and will follow up in future work) that the presented perspective on purposeful intersubjectivity may contain valuable explanatory elements for the theory of intentionality in language. We underscore that there is no such being as a prelinguistic child because interactions with infants almost always involve speech as one of multiple modalities [35], [48]. This also means that spoken language is immersed in physical interactions between a child and a caregiver. The infants’ vocalizations also appear to be very important for the mothers. They are almost always repeated or responded to [62]. Often the vocalizations are also interpreted as being relevant to the present activity and thus embedded into it and structured as turn-taking sequences unfolding in parallel to the activity [55]. The idea of grounding expressive and communicative actions in a systemic view of a developing organism has previously been acknowledged [35], but we claim that it is crucial to see ‘communicative actions’ not as actions of an individual ‘toward’ another individual but as possible affordances for a dialogical interaction [63].

Finally, we wish to underscore that this work is a first step in advancing a more quantitative approach for the analysis of early interactions. Our case examples can serve as instances of a behavioral category that encompasses different ways of how purposeful intersubjectivity is established. In the next step, the turns in a whole interaction can be classified according to this category. The distribution of instances within an interaction
and over a developmental time can then be assessed, and the conclusions can be generalized. Such quantitative treatment should lead to a better understanding of the progress of tuning to social affordances and allow for observations of the individual differences in the quality and effectiveness of this process. In future work, we envision that our qualitative results can lead to studies that quantitatively investigate how frames (i.e., recurrent interaction protocols) are formed as a basis for symbolic communications and how collaborative tasks are carried out in interactions with young infants.

**REFERENCES**


Joanna Rączaszek-Leonardi received her MA at the University of Warsaw and her PhD from the Center of Complex Systems and Brain Sciences at Florida Atlantic University, USA. She is a professor at the Faculty of Psychology, University of Warsaw and a co-founder of the interdisciplinary research group Cognitive Systems Warwick. Her interests involve basic human interactivity and the nature and role of symbols as the means through which this interactivity is controlled.

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 is currently finishing her Ph.D. within the project “Symbiosis of Language and Action,” which is supported by the Volkswagen Foundation. She is a member of the Emergentist Semantics Group and the Graduate School of the Center of Excellence Cognitive Interaction Technology (CITEC) of Bielefeld University. Her interests include language development, particularly the interactive foundations of language learning. Her research focuses on the interplay between language and bodily experience for the development of meaning.

Katharina J. Rohlfing received her Master's degree in linguistics, philosophy, and media studies from the University of Paderborn in 1997. As a member of the Graduate Program Task-Oriented Communication, she received her 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). Since 2008, she has been Head of the Emergentist Semantics Group within the Center of Excellence Cognitive Interaction Technology (CITEC) of Bielefeld University. Her habituation in 2009 on early semantics attests to her interest in the interface between cognitive development and the early stages of language acquisition.