Consciousness and Cognition 20 (2011) 643–657

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Extended cognition and the space of social interaction

Joel Krueger *

Center for Subjectivity Research, University of Copenhagen, Njalsgade 140-142, 2300 Copenhagen S, Denmark

ARTICLE INFO

Article history:
Received 8 April 2010
Available online 20 October 2010

Keywords:
Extended mind
Social cognition
Theory Theory
Simulation Theory
Moebius Syndrome
Gesture
Coordination
Intersubjectivity
Phenomenology

ABSTRACT

The extended mind thesis (EM) asserts that some cognitive processes are (partially) composed of actions consisting of the manipulation and exploitation of environmental structures. Might some processes at the root of social cognition have a similarly extended structure? In this paper, I argue that social cognition is fundamentally an interactive form of space management—the negotiation and management of "we-space"—and that some of the expressive actions involved in the negotiation and management of we-space (gesture, touch, facial and whole-body expressions) drive basic processes of interpersonal understanding and thus do genuine social-cognitive work. Social interaction is a kind of extended social cognition, driven and at least partially constituted by environmental (non-neural) scaffolding. Challenging the Theory of Mind paradigm, I draw upon research from gesture studies, developmental psychology, and work on Moebius Syndrome to support this thesis.

1. Introduction

The extended mind thesis (EM) asserts that some cognitive processes are (partially) composed of actions consisting of the manipulation and exploitation of environmental structures (Clark, 2008; Clark & Chalmers, 1998; Hurley, 1998; Hutchins, 1995; Menary, 2007; Rowlands, 2003). Some cognitive processes are thus partially driven by environmental (non-neural) scaffolding. Might some of the processes at the root of social cognition have a similarly extended structure? Despite intense recent interest in both EM and social cognition in philosophy and cognitive science, this question has yet to receive sustained consideration.

In this paper, I argue that social cognition is a kind of extended cognition. Specifically, I argue that social cognition is fundamentally an interactive form of space management—the negotiation and management of "we-space"—and that some of the expressive actions involved in the negotiation and management of we-space (gesture, touch, facial and whole-body expressions, etc.) drive basic processes of interpersonal understanding and thus do genuine social-cognitive work. Some social cognitive processes are therefore partially driven by and composed of non-neural scaffolding; and social cognition is in this way not reducible to individual, intracranial mechanisms but instead emerges from within the dynamics of the interactive process itself. Put otherwise, social *interaction* is a form of social *cognition*—the self-structuring negotiation of what I call "we-space". Challenging the dominant Theory of Mind paradigm in social cognition literature, I draw upon empirical research from gesture studies, developmental psychology, and work on Moebius Syndrome to support this thesis.

2. Embodiment, interaction, and "we-space

The basic mode of human sociality is face-to-face interaction with others who are physically co-present (Zhao, 2003). Co-presence occurs within shared contexts of interaction in which others become physically accessible and subject to one

* Fax: +45 35 32 86 81.
E-mail address: joelk@hum.ku.dk

1053-8100/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.concog.2010.09.022

another (Goffman, 1963, p. 22). Accordingly, the social characteristics of co-presence are anchored in the spatiality of the body (Giddens, 1984, p. 64). The notion of "we-space" is offered to elucidate the social significance of another's bodily co-presence.

In one sense, "we-space" is simply body-centric *action*-space. Recent neuropsychological research has explored distinctions between different forms of sensorimotor, action-space: *personal space* (inhabited by the subject's body, comprised of proprioceptive and tactile information); *peripersonal space* (immediately surrounding the subject's body, structured by the multimodal integration of visual, auditory, and tactile information); and *extra-personal space* (just outside of the subject's immediate reach, structured by visual and auditory information) (Cardinali, Brozzoli, & Farnè, 2009; Legrand, Brozzoli, Rossetti, & Farnè, 2007). Most of this research focuses on how multisensory representations of body-centric space are integrated and what sort of role these representations play in action execution.

Yet little attention has been paid to what sort of role these shared action-spaces *themselves* might play in driving various social cognitive processes. While retaining the practical character of previous discussions of body-centric space, the notion of "we-space" is offered to foreground the manner in which practical space is reconfigured as social space: the face-to-face locus of joint attention and mutually-coordinated social interactions.[1] Whereas neuropsychological accounts of body-centric space have primarily emphasized its practical character—it is structured relative to different action-possibilities in the agent's local environment—we-space, while practical, is additionally an emotion-rich coordinative space dynamically structured via the ongoing engagement of social agents. Moreover, whereas the single subject remains the primary locus of agency in neuropsychological accounts of body-centric space—reflecting the one-sided emphasis on practical or goal-directed action—the notion of a shared we-space instead emphasizes its co-regulated nature. Within we-space, agency does not emerge atomistically from a single source (the individual acting agent) but is instead distributed across the temporally-extended dynamics of co-regulated interaction. For instance, the coordinative ensemble of socially-salient behaviors (body posture, expressive gestures, gaze patterns, vocalizations, etc.) that emerge naturally in face-to-face interaction comprise "a self-organized, softly assembled (i.e., temporary) set of components that behave as a single functional unit" (Shockley, Richardson, & Dale, 2009; see also Bernstein, 1967). The bodily negotiation of we-space is in this way framed not as an explicitly cognitive-inferential process but rather as an ongoing form of embodied coping involving the continuous mutual adjustment of actions and intentions (Fogel & Garvey, 2007).

Inspiration for this understanding of we-space comes from theoretical and empirical sources. The most prominent theoretical source is the phenomenological tradition of western philosophy. Edmund Husserl, Max Scheler, Maurice Merleau-Ponty, and Emmanuel Levinas each in their own way argue for the importance of shared space in constituting the unique character of face-to-face interaction. Merleau-Ponty, for instance, argues that "our body is not primarily *in* space: it is of it" (Merleau-Ponty, 1945/1962, p. 171). He insists further that when it comes to social cognition, the task is to understand how it is that, "[b]etween my consciousness and my body as I experience it, between this phenomenal body of mine and that of another as I see it from the outside, there exists an internal relation which causes the other to appear as the completion of a system" (Merleau-Ponty, 1945/1962, p. 410). Further motivation for this characterization of we-space comes from developmental psychology. Affirming and often utilizing descriptive claims offered by phenomenologists, psychologists such as Peter Hobson (2002, 2008), Vasudevi Reddy (2008; Reddy & Morris, 2004), Colwyn Trevarthen (1979), Trevarthen (1992) and Phillipe Rochat (2009) among others have emphasized how emotional expression, turn taking, gaze coordination, and affect attunement serve as the developmental platform for antecedently-developing "higher" forms of cognitive social understanding.

Gesture and bodily expressivity are crucial for understanding the negotiation of we-space. I use "gesture" and "expressivity" (or alternatively, "expression") interchangeably to refer to actions central to the dynamic construction of shared meanings and sympathetic attunement. These terms encompass a range of bodily and body-related traits such as posture, movement, facial, hand and whole-body expressions. I thus argue for a differentiation between gesture and other forms of non-expressive bodily action (reflex, locomotive, and instrumental). For what distinguish gestures from other non-expressive actions is precisely how they exploit local space in the service of driving basic forms of interpersonal understanding. In other words, gestures scaffold the *shared narrative space* of communicative exchanges (Cole, Gallagher, & McNeill, 2002, p. 61). This is an activity of establishing we-space, an instance of gestures doing social-cognitive work.

2.1. We-space as focused interaction

This thesis clearly requires extensive unpacking. However, I want to first make a few more preliminary remarks about the relation between we-space and expressive scaffolding. In what follows, we-space refers to face-to-face interaction. Within face-to-face interactions—which are primarily dyadic in structure but which can, at times, involve more than two participants—bodily expressions convey extensive information about one individual to another, such as their moods, emotional states (both type and intensity), intentions, as well as their social status and behavioral competence. Expressive behavior is rich with social information. Much of this information is potentially available to all of the individuals within an agent's social context. For example, someone sitting across the room from me in a large lecture hall will likely be able to discern my mood (concentration) or intention (to comprehend the speaker's utterances) simply by observing my posture and facial

[1] Vittorio Gallese has spoken suggestively of "we-centric space" and "we-ness" (Gallese, 2009). However, Gallese's concern is primarily with the subpersonal functional mechanisms (e.g., mirror neurons) responsible for social cognition, whereas I am concerned with the social dynamics of we-space as they manifest at the personal, or experiential, level.

expression (unwavering gaze, leaning forward expectantly, brow furrowed in concentration). However, this is an instance of unfocused interaction since the signals my body conveys are communicated merely in virtue of my observer's co-presence within the common space of the lecture hall (Goffman, 1963). Other examples of unfocused interaction include individuals walking down a street, sitting in a waiting room, or sharing a bus. In these contexts, there is no genuine interpersonal engagement, only bodily co-presence. Additionally, there is no *co-regulation* in the sense that the expressive activity of one individual plays no direct role in shaping the expressive activity of another (my observer in the lecture hall might chuckle at my furrowed brow and attentive posture but it has no bearing on my behavior).

We-space captures a richer, and more structurally complex, form of social interaction that rests on bodily co-presence—it is rooted in the perceptual and communicative modalities of the body—but which additionally incorporates several other important elements. It involves focused interaction: "the kind of interaction that occurs when persons gather close together and openly cooperate to sustain a single focus of attention" (Goffman, 1963, p. 24). This interactive process of cooperation and attention is necessary for the creation of we-space. It occurs when two or more individuals openly coordinate their expressive activities though a continued intersection of bodily and vocal gestures (Giddens, 1984, p. 72). This coordination is, however, simultaneously a process of *co-regulation*. The expressive activity of one individual shapes that of the other, and vice-versa—a back-and-forth process of continuous reciprocal causation[2] (Clark, 1997) which establishes the narrative character of that particular interaction (e.g., coordinated sequences of bright-eyed facial expressions, friendly gestures and enthusiastic utterances establish an interaction's mutually-affirming positive valence). Co-regulation thus occurs when both partners are responsive to mutual influence, resulting in the emergence of new information not previously available to participants prior to their joint engagement (Fogel, 1993). Additionally—and this is the key point—face-to-face focused interaction, rooted in the body's expressive dynamics and ongoing processes of co-regulation, erects a "conventional engagement closure" (Goffman, 1963, p. 156). It marks off those interacting from others who are bodily co-present but not directly engaged (think of lovers nuzzling in a crowded movie theater, or a mother interacting with her infant in the middle of a bustling sidewalk). This "marking off" characteristic gives we-space its particular ethos of intimacy, its particular shared meaning. The establishment of we space is, in essence, the bodily establishment of a conventional engagement enclosure. Construed thusly, the scaffolding role of the expressive body in this process is reaffirmed. This is because the mechanisms responsible for establishing and negotiating we-space are (1) the *control of the body* (the regulation of its behavioral and expressive capacities), and (2) the adherence to various *body-related interactive norms or conventions* (expressive coordination, turn-taking rhythms, spatial proximity, gaze patterns, etc.) (Giddens, 1984, p. 73). The focused encounters that create we-space are the foundation of social interaction and the basis of interpersonal understanding.

3. Cognitive scaffolding, cognitive niche

The twin notions of "cognitive scaffolding" and "cognitive niche" are important parts of the EM story. Niche construction consists of "the activities, choices, and metabolic processes of organisms, through which they define, choose, modify, and partly create their own niches" (Laland, Odling-Smee, & Feldman, 2000, p. 132). Organisms act on physical structures in their local niche in ways that introduce novel selection pressures and enhance adaptive fitness. For example, beavers enact ecological changes (dam modifications, nests, burrows, paths, etc.) in response to seasonal and predatory challenges that carry over to subsequent generations by altering selection pressure (Kirsh & Maglio, 1994). An important point is that the environmental niche is not a fixed constant—a boundary offering up a fixed set of constraints on action—but is, rather, a dynamic source of ecological information and adaptive opportunity. A related and equally important point is that by altering physical structures in the environment, niche construction generates mind-expanding feedback cycles (Clark, 2008, p. 62). The environmental modifications involved in niche construction are processes that filter and transform the action of the environment as it redounds back onto the agents *doing* the niche-constructing, opportunistically transforming the informational load placed on the agent and thus opening up new avenues for efficient thought and action (Lungarella & Sporns, 2005). Environmental interaction is in some cases *cognitive performance* (Kirsh, 1995). This is particularly evident with human cognition.

One of the most effective ways that we transform the informational load on our cognition is via the construction of location-specific material cultures. In addition to language and norm-governed rituals, humans employ various epistemic artifacts (tools, weapons, clothes, forms of shelter, means of transportation, etc.) designed to scaffold intelligent thought and action (Sterelny, 2004). These epistemic tools are tools for managing cognitive space. For example, we routinely exploit various forms of environmental storage (sticky notes on the side of a computer monitor; street signs, arrows pointing to the toilet; a box of to-be-recycled items positioned strategically by the apartment door; cultural songs and narratives) to ease the burden of holding information in biological memory and prompt rapid recall. Via informational offloading, we allow bits of the world to store information for us and access these bits only on a need-to-know basis. Similarly, we exploit the informational space of our workspaces to simplify *choice* (laying out cooking ingredients in the order needed), *perception* (arranging jigsaw pieces into shape or color piles), and *internal computation* (physically manipulating Tetris pieces or continually reordering Scrabble tiles to prompt new word associations) (Kirsh, 1995; Scribner, 1986; see also Sterelny, 2004). Space is thus a cognitive resource, and manipulating space a cognitive performance.

[2] Within a social context, this is the idea that my behavioral responses to events in my social milieu elicit and shape the very events they are responsive to; responsive action and eliciting event form a functional unity or "new whole" (Merleau-Ponty, 1942/1963, p. 13).

How does this relate to social cognition? I suggest that similar principles are at work even in our earliest interpersonal engagements. In other words, the interactive character of our social engagements alters the structure of that particular cognitive niche (we-space), generating new feedback cycles and process of shared feeling and sympathetic understanding uniquely specified to that exchange. Co-regulated, focused interactions are thus a form of interpersonal niche construction. As a preliminary example, consider Reddy's claim that infants as young as 2 months, prior to evidence of joint attention (12 months) or Theory of Mind capacities (3–4 years), show both awareness of others as attending beings as well as awareness of themselves as objects of others' attention (Reddy, 2003). Within instances of focused interaction, very young infants exhibit a variety of emotional reactions to others' attention (increased smiling at eye contact, less when adults turn away; elaborate tempo-sensitive expressivity in response to attention; distress at still-face or non-contingent gazes; coy reactions to renewals of attention); and they also employ a range of embodied-expressive strategies to direct others' attention to specific aspects of *themselves* (social actions such as clowning, showing-off, clever displays, or teasing) (Reddy, 2003). They skillfully manage we-space to motivate social engagements.

According to Reddy, these interactions are developmentally primitive forms of interpersonal understanding rooted in self-other conscious affects (Reddy, 2003, p. 400). They are active negotiations of we-space: the infant's actions are continuously and self-consciously influencing and responding to caregivers, and vice-versa. And while these interactions precede the emergence of the infant's ability to formulate conceptual representations of self and other, they nevertheless indicate an early, pre-joint-attentional, perceptual and affective grasp of others as intentional agents (Reddy, 2008). Important for present concerns is the fact that these focused interactions are driven by two forms of non-neural scaffolding: the material scaffolding of the expressive body as well as the scaffolding provided by the local structure of we-space itself. Both of these features qualify as cognitive scaffolding in that they are an active constitutive element of the process of perceiving another as an intentional agent (the basis of interpersonal understanding); to remove one kind of scaffolding, either through physiological impairment or within an experimental set-up, results in a diminishment of the infant's social-ecological competence, as we'll see below. I turn to a more focused consideration of these ideas now. First, I discuss empirical research supporting the idea that the body's expressive dynamics serves as the material scaffolding for some cognitive and affective processes, pushing these processes (at least partially) beyond the head. In the following section, I say more about how expressive dynamics actively structure we-space and constitute part of social cognitive processes at the root of interpersonal understanding.

4. Gesture as material scaffolding

4.1. Gesture and cognition

Despite having no experience of their face, newborn infants are capable of perceiving and imitating the facial expressions of caregivers (Kugiumutzakis, 1999; Meltzoff & Moore, 1977, 1997). This gestural mirroring is an important component of focused interaction. And it doesn't stop in infancy: adults, too, mimic the nonverbal behaviors and gestures displayed by conversational partners, establishing rapport and motivating social interactions (Chartrand & Bargh, 1999; Kendon, 1970). Gestures are thus material scaffolding available to both speaker and listener. They have a public presence; their semantic quality is externalized. In virtue of this publically accessible character, various lines of evidence suggests that they have a systematic cognitive-affective impact on both (1) the subject who initiates them as well as (2) the subject(s) who perceives them (Clark, 2008, p. 127; Goldin-Meadow, 2003).

In some instances spontaneous gestures[3] don't merely express fully-formed thoughts or intentions. Rather, they appear to constitute part of the material process of thought-in-action (Goldin-Meadow, 2003; Laird, 2007; McNeill, 2005). Let me consider this idea by bringing together several streams of empirical research. Consider first that gestures occur in a surprising array of contexts: when talking on the phone, to ourselves, or in the dark; gestures co-vary with task difficulty, and increase when speakers must choose between increased options; we gesture more when reasoning about a problem as opposed to simply describing the problem or recounting a known solution (Goldin-Meadow, 2003, pp. 136–149). Moreover, everyday observation reveals how seamlessly gestures integrate with our speaking. Gestures are not mere automatic reflexes—but on the other hand, we're not normally aware of them in an explicit self-monitoring sense. Under normal conditions, gestures appear to be ubiquitous feature of the active *performance* of speech and thought.[4]

But how might gestures play an active role in driving the thinking process itself?[5] Goldin-Meadow and colleagues (Goldin-Meadow, 1999, 2000, 2003, 2009; Goldin-Meadow & Wagner, 2005) have argued that, in some contexts, gestures lighten the speaker's cognitive load by freeing up additional resources for memory and recall. In one experiment, children and adults were asked to explain how they solved a math problem while simultaneously remembering a list of words or letters (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001). Those told to refrain from gesturing during the intervening mathematical test did

[3] Spontaneous gestures are those gestures that speakers produce along with (though not necessarily tied to) speech, sometime also referred to as "illustrators". "Emblems" are conventionally-specified gestures with particular meanings, such as *thumbs-up*, *okay*, *shush*, the infamous "middle finger", etc. (Goldwin-Meadow & Wagner, 2005, p. 234; see also Ekman & Friesen, 1972).

[4] Merleau-Ponty argues similarly that language emerges from movement, that is, it is generated when "the body converts a certain motor essence into vocal form" (Merleau-Ponty, 1945/1962, p. 181).

[5] Clark (2008, pp. 123–131) asks this same question. My discussion here is very much indebted to his analysis.

significantly poorer on the memory test than the gesturing group. Gesture may improve performance because it facilitates conservation of cognitive resources on the explanation task—some of that burden is off-loaded onto the gestures; they let their hands "do the talking"—freeing up more resources for recall (Goldin-Meadow & Wagner, 2005, p. 237). Other studies suggest a causally-reciprocal link between gesture and learning. For instance, children working on mathematical equivalence problems are more likely to learn successful problem-solving strategies when they mimic an instructor's gestures representing a correct problem-solving strategy (Cook & Goldin-Meadow, 2006). Gesturing during the learning of a new mathematical concept, as opposed to just speaking about it, assists concept retention (Cook, Mitchell, & Goldin-Meadow, 2008). And early (prior to 14 months) and prodigious gesturing, such as pointing, seems to play a crucial role in later vocabulary development (Rowe, Ozcaliskan, & Goldin-Meadow, 2008).

Additional support comes from a study on the cognitive benefits of doodling, which may at times actually aid concentration and recall (Andrade, 2010). The physicality of doodling may provide a kind of self-stimulating, concentration-enhancing loop diminishing one's tendency to daydream by maintaining an optimal level of arousal during an otherwise boring task. Alternatively, doodling may also perform an off-loading function, sharpening concentration on task-specific information by helping the doodler organize relevant information for later recall. For example, doodling each time a bit of salient task-specific information is processed might structure a visuo-spatial representation of the relevant information, making it easier to remember later on. These interpretations are not inconsistent; perhaps they're both correct. The point is that the material scaffolding of the body seems to play an active cognitive role in the doodler's ability to process, parse, and recall information.

Consider next the case of Ian Waterman, who lacks both a sense of touch and proprioceptive feeling from the neck down (Cole, Merton, Barrett, Katifi, & Treede, 1995). In order to maintain motor control, Ian continually focuses on his movements and tracks his limbs in space by keeping them in his visual field at all times. Despite his proprioceptive deficit, Ian has, with much effort, regained his ability to gesture, including spontaneous gesturing while speaking (Cole et al., 2002; Gallagher, Cole, & McNeill, 2001). As with other movements, he consciously initiates gesture while talking. But Ian gestures naturally even *without explicit visual feedback and attentional control,* which he requires for other motor movements (Gallagher, 2005, p. 117). Ian's gestures thus appear to inhabit a different category than instrumental or locomotive action. Rather than inhabiting a pre-established narrative space, they are expressive actions that instead help to create the narrative space within communicative contexts (Gallagher, 2005, p. 117). Ian Waterman's case affirms that "the mutual co-occurrence of speech and gesture reflects a deep association between the two modes that transcends the intentions of the speaker to communicate" (Iverson & Thelen, 1999, p. 19).[6] These different strands of evidence do not, of course, decisively show that gesture is quite literally the embodiment of thought. But they do strongly suggest that gesture is more than an auxiliary support system standing behind thought and language (e.g., Butterworth & Beattie, 1978; Hadar, 1989; Levelt, Richardson, & La Heij, 1985), and that it is, rather, deeply interwoven into the material processes of cognition and speech production.[7]

4.2. Gesture and affect

Other research indicates that the phenomenal character of emotion and affect may crucially depend on the material scaffolding of bodily expression. For instance, a recurrent theme in the narratives of those with Moebius Syndrome—a congenital form of bilateral facial paralysis—is a persistent sense of diminished affect somehow tied to their condition (Cole, 2010). Some Moebius subjects report the feeling of assuming a spectatorial as opposed to a participatory stance with respect to their emotional experiences and social interactions. James, a priest in his fifties, says that

> I have a notion which has stayed with me over much of my life—that it is possible to live in your head, entirely in your head. . .I do think I get trapped in my mind or my head. I sort of think happy or I think sad, not really saying, or recognizing, actually feeling happy or feeling sad. . .I've often thought of myself as a spectator rather than a participant (Cole, 1999, p. 308).

Cole also quotes Oliver, a university student in his early twenties, who contracted Bell's Palsy and over the course of six months gradually succumbed to total facial paralysis before later recovering fully. Oliver was able to track the progressive diminishment of his emotional life commensurate with his loss of facial expression: "I suppose I didn't feel constantly happy, but then I didn't feel sad. . .I felt almost as if in a limbo between feelings—just non-emotional. . .I didn't think that I felt it as much because I was not actually smiling" (Cole, 1999, p. 310). Oliver adopted a surrogate scaffolding (writing) which allowed him to recapture some of the phenomenal character of his diminished emotional experience. The physical act of writing it-

[6] Some individuals with congenital absence of limbs report the experience of gesturing (Brugger et al, 2000; Ramachandran & Blakeslee, 1998). The phantoms are not reported to be active in contexts calling for instrumental action but only in communicative contexts that involve gesture. Additionally, congenitally blind individuals gesture while speaking, despite never having seen their own hands or having seen another person gesture (Iverson & Goldin-Meadow, 1998; Goldin-Meadow and Wagner, 2005). Like sighted gesturers, they do so in a variety of situations (Iverson, 1999; Iverson, Tencer, Lany, & Goldin-Meadow, 2000), produce the same range of gesture forms as sighted speakers (Iverson & Goldin-Meadow, 2001), and even gesture when speaking to others they know to also be blind (Iverson & Goldin-Meadow, 1998, 2001).

[7] Multiple converging lines of research from neurophysiology and neuropsychology indicate links between language and movement at the neural level. For discussion, see Iverson and Thelen (1999) and Gibbs (2006, pp. 165–174). There is now widespread agreement that speech and gesture originated from the same neural system (Corballis, 1994).

self—beyond merely expressing the significance of a particular episode—partially recalibrated the emotional experience's normal phenomenal intensity. Other Moebius subjects report adopting scaffolding with similar results: the physicality of artistic expression, for instance—painting, dancing, or playing the piano—becomes a vehicle for enacting the phenomenal character of emotions and moods (Cole & Spalding, 2009). Another Moebius subject began to mimic gestures she observed while on holiday in Spain, claiming that, "I do not think I had emotion as a child but now I have it. How did I get it? It was in Spain. . .The body language I had learnt and used at university could be exaggerated in Spain, using the whole body to express one's feelings" (Cole & Spalding, 2009, p. 154). Moebius subjects are thus open to utilizing alternative channels of embodied expression such as prosody, gestures, and verbalization to communicate and experience emotion (Rives Bogart & Matsumoto, 2010).

Apart from Moebius narratives, the idea of continuous reciprocal causation between bodily expression and the phenomenal character of emotion receives support from many studies.[8] The largest and most consistent body of evidence in support of this theory concerns facial expressions (Laird, 2007, p. 23): an overwhelming majority of existent studies seem to indicate that manipulation of expressive behavior produces corresponding changes in feelings. For example, multiple studies have found that when subjects are induced to adopt a particular emotion-specific facial expression (grimacing, frowning, etc.) or posture, they report experiencing the corresponding emotion (disgust, anger, etc.) (Duclos et al, 1989; Duclos & Laird, 2001; Edelman, 1984; Flack, Laird, & Cavallaro, 1999; for extensive review, see Laird & Bresler, 1992; Kellerman & Laird, 1982; Niedenthal & Maringer, 2009; see also Niedenthal, 2007). Paula Niedenthal surveyed other research indicating that (1) adopting emotion-specific facial expressions and postures influences preferences and attitudes, and (2) inhibition of bodily expression (motor movements) leads to diminished emotional experience (reduction in the experience's phenomenal intensity), as well as interference in processing emotional information (Niedenthal, 2007; see also Niedenthal, Barsalou, Ric, & Krauth-Gruber, 2005). This research, coupled with the narratives of Moebius subjects, strongly suggests that the embodied expression of emotional states—along with their social sharing—may be necessary for their being experienced (Cole, 2010, p. 667).

5. Expression, interaction and we-space

Based upon the studies and narratives discussed above, there are reasons to think that some cognitive and affective processes have a spatially extended structure (i.e., they are partially constituted by the material vehicles of gesture and bodily expressivity). However, this fact alone does not show that gesture and bodily expression play a role in establishing the narrative structure of we-space—that is, by embodying and externalizing intentions, and by soliciting and co-regulating the communicative expressions of others—and, in doing so, perform genuine social-cognitive work. Yet this is the claim at the heart of the paper. Having put the necessary pieces in place, I now want to develop this argument in more detail. First, some background.

5.1. Social cognition and Theory of Mind

For a number of years, the Theory of Mind (ToM) paradigm has framed debates in social cognition. Within ToM, "mindreading" or "mentalizing" (Whiten & Perner, 1991)—the ability to impute mental states to others, and in doing so attribute intentions and interpret patterns of behavior—is said to be the basis of social understanding. A related supposition is that the mind (and *inter alia*, intentions) is localized inside the head, directly accessible only to the introspecting individual (Proust, 2003).[9] Given the interpsychic distance between subjects, social cognition is fundamentally a project of developing the requisite mindreading mechanisms to overcome, or at least lessen, this epistemic distance; these mechanisms are what allow one subject to represent what is happening in the mind of another.

One generally finds two proposed mechanisms: the theories advocated by Theory Theory (TT) and the simulations advocated by Simulation Theory (ST). According to TT, interpreting and predicting behavior is the product of innate or acquired theories about how minds work, how mental states interrelate, and how mental states causally motivate behavior (Baron-Cohen & Frith, 1985; Carruthers, 1996; Premack & Woodruff, 1978). They allow us to make inferences about another's mental life and to anticipate and interpret their behavior based on these inferences. ST, on the other hand, urges that this sort of inferential theory-making is unnecessary in virtue of the immediate access we have to our own cognitive and emotional resources. We instead exploit the rich inner resources of our own mental life to imaginatively model the mental states of others *as if* we were in their situation, yielding a practical understanding of another's intentions (Dokic & Proust, 2002; Goldman, 2006; Goldman & Sripada, 2005; Gordon, 1996; Heal, 1986). The particulars of the ToM debate are more intricate than this caricature suggests.[10] But what this gloss conveys is that the ToM framework (as well as TT and ST) rests on the "Myth of the Hidden" (Torrance, 2009), assuming that we can only ever attain indirect knowledge of other's mental life by first (1) perceiving external behavior and then (2) inferring to hidden inner states and intentions via theorization, simulation, or some

[8] See Laird (2007) for a survey of hundreds of such studies.

[9] Alan Leslie, for example, writes that, "[o]ne of the most important powers of the human mind is to conceive of and think about itself and other minds. Because the mental states of others (and indeed ourselves) are completely hidden from the senses, they can only ever be inferred" (Leslie, 2004, p. 164).

[10] Some simulation theorists argue that simulating is not an explicit process of imaginative projection but rather an offline or subpersonal process (Gallese, 2001, 2009). Other theorists have begun endorsing hybrid theories combining aspects of both TT and ST (Currie & Ravenscroft, 2002; Goldman, 2006; Perner & Kuhberger, 2005).

combination of the two, since (1) is merely a secondary *expression*, and not a material or *constitutive* aspect of, (2). This mentalistic and individualistic emphasis, as well as the tendency to favor inferential and/or simulative models, has led to a failure to recognize the centrality of both bodily expression and interactive engagement in facilitating interpersonal understanding (Hobson, 2002; Reddy, 2008). Based on this neglect, ToM has recently been subjected to a number of phenomenologically-informed criticisms.[11] These criticisms, too, differ in their details. Yet they are united in shifting the focus away from the intrapsychic and onto the coordinated interaction of face-to-face engagement.

Coordination is the realization of synchronized patterns of relations between entities (Baron, 2007, p. 180) on both a microscopic level (configurations of tensile states or patterns of cellular and vascular activities), as well as on a macroscopic level (patterns of body and limb motions relative to objects and events in the environment) (Turvey, 1990). In the context of social cognition, coordination refers to the temporally-configured alignment of movements, gestures, and vocalizations that allow participants to share intentions and construct felt contexts of sympathetic attunement. However, this coordination is not merely a physical process. It essentially involves embodied subjects, and thus within these couplings there is also a coordination of *meaning*—meanings which emerge, align, transform and evolve via interaction (Fuchs & De Jaegher, 2009, p. 471; Trevarthen, 1979). And once more, within the focused interaction of we-space, interpersonal coordination entails co-regulation: patterns of mutually-governed action and response that ground sympathetic attunement and the co-creation of shared meanings.[12]

This is a process that begins at birth. For example, consider newborn infants who are capable of perceiving and imitating the facial expressions of caregivers despite never having seen their own face (Kugiumutzakis, 1999; Meltzoff & Moore, 1977, 1997). Though they lack the cognitive capacities for inferential theorization or projective simulation (mindreading), neonates nevertheless seem to pick up on these gestures as interpersonally significant—imitation only occurs within the affective ethos of actual face-to-face engagements with other humans (Reddy, 2008, p. 79)—and they attempt to respond with meaningful gestures of their own. This mimetic behavior is part of a broader repertoire of engagement strategies that includes tracking intonation of adult frequencies (Lieberman, 1967) and co-vocalizing with caregivers, even at three days, significantly above chance (Rosenthal, 1982). However, within the space of face-to-face imitation, the intense focus on the face of the other, as well as the complementarity of the partner's actions, establishes an acute form of affective intimacy that serves as the basis of interpersonal communication (Zeedyk, 2006, p. 323). Within this imitative space, the infant also begins to bodily negotiate self-other boundaries; imitation orients attention away from self and toward the other (Zeedyk 2006, p. 332). Portraying imitation as a dynamic process of negotiating self-other boundaries thus highlights the fact that imitating infants are not merely passive observers but rather active participants in composing the gesture-based "communicative musicality" of early engagements (Malloch, 1999; Malloch & Trevarthen, 2009). For within weeks, imitation develops into reciprocal, context-sensitive *interaction* (Reddy, 2008, p. 59). Even newborns therefore appear at some level to recognize bodily expressivity as a vehicle for negotiating we-space—and they actively exploit it to do just this.[13]

How does gesture, even at this early stage, facilitate we-space management and perform rudimentary social-cognitive work? Recall that, in Kirsh's treatment, the primary function of the intelligent use of space is to reduce the descriptive complexity of the local environment (Kirsh, 1995, p. 66). Gestures, I suggest, potentially perform a similar space-management function. Not only do they scaffold (and thus externalize) some sender-side cognitive and affective processes, as the gesture studies and Moebius narratives indicate. Additionally, gestures and bodily expressivity exert a social-cognitive impact on the receiver: namely, they lighten their cognitive load within focused interactions by reducing the descriptive complexity of that interpersonal niche, sculpting the back-and-forth character of the interaction and thus motivating shared understanding. This "load lightening" is accomplished in at least two interrelated ways. First, gestures ease the epistemic burden on the receiver by scaffolding spatial arrangements that simplify *choice* and *perception*. Second, they ease the epistemic burden on the receiver by scaffolding spatial dynamics that simplify *internal computation*. This reduction of descriptive complexity is an active structuring and manipulation of we-space—a jointly-constituted interaction that serves as a mechanism for driving interpersonal understanding (Cowley, 2003).[14]

[11] See, e.g., De Jaegher and Di Paolo (2007), Di Paolo (2009), Gallagher (2004, 2008a, 2008b), Gallagher and Zahavi (2008), Hobson (2002), Hutto (2008), Overgaard (2006), and Zahavi (2008).

[12] De Jaegher and Di Paolo (2007) term this process as "participatory sense-making".

[13] A deflationary criticism might question interpreting early imitative episodes as constituting a genuine kind of pre-linguistic communication. Admittedly, this remains an open question, well beyond the scope of this discussion. However, in response it can first be observed that infants exhibit a strong affective motivation to engage in bouts of imitation (which, defined expansively, include not just discrete, consecutive acts of turn-taking but also bodily gestures and emotional expressions that are overlapping or simultaneous) (Zeedyk, 2006, p. 334). They actively seek out opportunities for engagement—and while engaging, they appear to enjoy it. Increased smiles are observed before, during, and after imitative exchanges (Kugiumutzakis, Vitalaki, Kokkinaki, & Makrodimitraki, 2005). Infants exhibit increased heart rate (suggesting heightened attentiveness and anticipation) when *provoking* interactions instead of simply *responding* (Nagy & Molnar, 2004). Moreover, neonatal imitation demonstrates two characteristics basic for social development and communicative competence: selectivity of human stimuli, and ontological detection of identity (Nadel, Guerini, Peze, & Rivet 1999, p. 211). Finally, even the earliest bouts of infant-caregiver imitation seem to share a number of features that establish emotional and communicative contact (Trevarthen 1979, 1993). For example, infants and caregivers share 3-D space, and companion space; similar tendencies to interact in turns, and to establish shared temporal patterns; the ability for self-other discrimination, as well as the ability for recognition of face and voice isomorophism; a shared code of communication, and the ability to read and respond to emotional expressions (Kugiumutzakis, 1999, p. 55). Therefore, it seems plausible to characterize these early engagements as having a legitimately social and communicative function (Reddy, 2008, p. 61; Uzgiris, 1981). I'm grateful to an anonymous reviewer for pressing this point.

[14] Despite my using divisional terms like "sender-side" and "receiver", it is important to be mindful that, within focused interaction, gestures and patterns of bodily expressivity are in fact co-regulated or *responsive*—that is, they are part of an ongoing interactive process shaped by gestures and expressivity on both sides of the dyad, and therefore are part of a dynamic, self-organizing process rightfully conceived of as a single functional unit.

5.2. Gestures simplify choice

Gestures actively structure we-space by simplifying choice. In particular, gestures and other kinds of bodily expressiveness draw attention to social affordances within we-space that both constrain as well as cue trajectories of available interaction. By "social affordances", I am referring to the joint actions within focused interaction—comprised of tactile, auditory, and visual contact with others—that, unlike contact with other physical objects, are dynamic and interactively constituted. The interactive, jointly-constituted (co-regulated) character of social affordances ensures that the bodily co-presence of another person is different in kind, experientially speaking, from a piece of equipment or other physical object. This is because "[t]he latter, as affordances, are ignored or used up according to present interests, whereas the presence of another opens up, whether you like it or not, a world of constraints and possibilities that cannot be ignored in the same way" (Still & Good, 1998, p. 56).

For example, consider the rich array of social affordances manifest in facial expressions and whole-body gestures. A slight pulling back of the head and upper trunk, a brief distracted glance away from the face of one's partner mid-conversation, a well-timed raised eyebrow, or a delayed response out-of-sync with the rhythm of previous vocal and gestural exchanges can indicate a desire to socially disengage. These expressive movements occlude receiver-side[15] affordances that would otherwise open up possibilities for continued engagement and further intimacy. In other words, they constrict we-space such that the receiver (assuming she is sensitively attuned to them) feels her range of social options suddenly less than those available to her a moment ago. The fact that infants, for example, become highly agitated and distressed during still face experiments—social interactions in which opportunities for further engagement are abruptly closed off due to the mother's sudden lack of expressivity—indicate that a sensitivity to constraints established by social affordances is at work very early, developmentally speaking (Murray & Trevarthen, 1985; Tronick, Als, Adamson, Wise, & Brazelton, 1978).[16] Similarly, lack of gestural coordination and postural mirroring can signal—and even exacerbate—discord within cooperative contexts (LaFrance, 1985).

But gestures don't merely hide social affordances. They also highlight various opportunistic affordances. For example, gestures can index moments of cognitive instability in young learners by reflecting thoughts not yet capable of being given verbal articulation (e.g., a child's gestures indicate an understanding of how to solve a mathematical equivalence task but the child cannot yet verbally articulate their successful strategy) (Goldin-Meadow, 2003, p. 56). Gestures can signal both what a student knows and doesn't know about a particular task. Sensitive teachers may therefore perceive certain kinds of gesture as affording opportunities to intervene and help the learner integrate different information by providing the conceptual framework in which to do this. These gesture-speech "mismatches"—moments when gesture conveys different information from speech—are material carriers of the learner's working through a task that publically signal crucial transition points within the learning process (Church & Goldin-Meadow, 1986; Goldin-Meadow & Wagner, 2005; see also Crowder & Newman, 1993). A sensitive teacher, rather than having to guess or inferentially work out the inner content of her student's mind, can instead directly perceive the learning process dynamically play out within the spatial arrangements of student's gestures.

Something similar is often at work in our experience of others' emotions. Nonverbal bodily-expressive behavior is widely recognized to play an essential role in managing interpersonal encounters. It expresses emotion, articulates interpersonal attitudes, presents one's personality, and is crucial in negotiating dynamical aspects such as turn-taking, feedback, and attention (Argyle, 1975). Congruent postures convey mutuality of topic and interest, and build rapport between partners (Goodwin, 1981). Building off of the previous discussion, however, there are many everyday cases where a gesture, facial expression, or whole body expression can articulate information at odds with an individual's verbal report, thus highlighting an opportunistic affordance. For example, head and bodily cues (facial expression versus posture, hand gesture, movement, etc.) often carry different affective information (Ekman, 1965). Whereas the former carries information about the specific emotion or affect being experienced but little about the intensity or level of arousal, the latter carries information about the intensity or level of arousal but little about the specific emotion or affect.

Consider a situation where I encounter a co-worker in the hall. Something about her bodily comportment strikes me as "off": a subtle slowness and lack of dynamism in her gait; slightly hunched shoulders; a heavy, leaden posture, in contrast to her normal more buoyant way of carrying herself. Convinced that something is wrong, I approach her and ask if everything is ok. She smiles, brow slightly furrowed, and says, "No, I'm fine, just a bit tired and down from the winter weather, I suppose". However, I remain convinced that her body is articulating something more affectively intense and I press her gently: "Are you sure that's it?" My colleague looks at me suggestively for a moment and pauses, uncrossing her arms and leaning toward me slightly. I perceive this leaning-in as an invitation to continue, so I ask again: "I'm here if you want to talk about any-

[15] By "receiver-side" affordances, I mean affordances perceived but not initiated by a particular partner within a social interaction (e.g., I perceive (receive) my partner's smile and wink as affording further intimacy within our present exchange or, conversely, their frown and downcast gaze as closing off such a possibility).

[16] An important question—though somewhat orthogonal to this discussion—is how best to characterize, within a given context, the experiential relation between (1) *actualized* affordances (affordances used to guide further action) and (2) *unactualized* affordances (affordances potentially available but not used to guide further action). When we act on (1), it's not clear how far our awareness of (2) extends; for we're surely not always aware of *every* available affordance when choosing to act on specific ones. Nevertheless, it does seem correct to say that picking out and acting on specific affordances brings with it an attenuated awareness of at least some of the unactualized affordances "left behind", so to speak. Thinking of the horizon of experience as having a focus-fringe structure (James, 1890/1950) might help us begin to get a grip on this issue. See Rietveld (2008) for a fuller discussion of these and other related matters.

thing". My colleague then proceeds to recount a recent event that has wounded her deeply. In this case, I correctly perceived the mismatch between her verbal report and bodily expression (her posture, gait, bodily comportment, etc.), and I used this mismatch as an opportunistic affordance to step in, establish we-space inviting her honest participation, and in doing so console my colleague.

These sorts of encounters are not aberrant. To the contrary, they make up the marrow of our everyday social lives.[17] Gestures and bodily expressions thus reveal salient information about an individual's cognitive status (their mood or confusion over a particular task)—both by very literally articulating material aspects the state itself as well as potentially signaling a mismatch of one sort or another—that receivers may then utilize to alter their input and, in doing so, transform the sender's cognitive and affective status. This gesture-based back-and-forth dynamic, the rudiments of which are already enacted from birth, is a process of bodily negotiating we-space that drives interpersonal understanding.

5.2.1. An objection

But is it really an instance of genuine social cognition? An objection is that this picture conflates understanding the *expression* of, e.g., desire (the desire to socially disengage or invite further intimacy) with an understanding of the desire *itself*.[18] To genuinely understand another's desire, the objection goes, is to understand the other as instantiating a representation with a world-to-mind direction of fit—that is, as instantiating a desire that I no longer socially engage with them—and that, therefore, I may perceptually pick up on a social affordance (a distracted glance or delayed response), respond appropriately (end the conversation and leave), but ultimately fail to understand anything about the other's motivational desire. More simply, I can be perceptually responsive to expressions without understanding anything about the intentional states behind them; purely seeing a *movement* is not the same thing as seeing it as *intentional*. The latter requires additional cognitive resources.

This is an important point; I'll say more about it in Section 5.4. We can first note, however, that it's not clear this objection passes the circularity smell test in that it seems to presuppose that genuine social cognition inheres not in *interaction* but rather only in the process of deploying an observational, other-directed Theory of Mind enabling me to *infer* what another is thinking—which is, precisely, one of the ideas here being contested. Moreover, it similarly presupposes a tidy division between an (inner) mental state and its (outer) behavioral expression which, as we've already seen above, may not adequately reflect the integrated character of some cognitive and affective processes.

With respect to the first point, many developmental studies (some of which were touched on earlier) indicate that, prior to developing the capacity to explicitly attribute mental states to others (i.e., a Theory of Mind), infants interact with caregivers in ways that suggest not only a perceptual awareness of social affordances but, additionally, a basic understanding of their *interpersonal significance* (Hobson, 2002). First, even very young infants are relatively sophisticated "smart" perceivers (Gallagher, 2008b), socially speaking, surprisingly attuned both to the timing and emotional quality of gestures and bodily expressions (Nadel et al., 1999) as well as to the emotive values carried by the harmonic and melodic parameters of the human voice (Trevarthen, 2002). And their perceptual skills are more robust than previously thought: newborn infants not only pick out faces for imitation but can reliably discriminate their mother's face and voice (Field, Cohen, Garcia, & Greenberg, 1984; Mehler, Bertoncini, & Barriere, 1978), discriminate accents (Mampe, Friederici, Christophe, & Wermke, 2009), and track frequencies of adult utterances by changing their intonation with different speakers (e.g., higher with mother, lower with father) (Lieberman, 1967); very young infants can pick out and attend to fine-grained auditory properties of music such as pitch, melody, tempo, and musical phrase structure (Schellenberg & Trehub, 1996; Trehub & Schellenberg, 1995; Trehub, Schellenberg, & Kamenetsky, 1999; Trehub & Trainor, 1993); at 2 months, infants can remember short melodies and later discriminate a remembered melody from other melodies (Plantinga & Trainor, 2009); at 3–6 months they can vocalize a matched pitch to sung tones (Wendrich, 1981) and learn to turn toward a loudspeaker whenever they perceive a change in background melody (Trehub, Thorpe, & Morrongiello, 1987).

These diverse perceptual skills enable sensitivity to social affordances. By 5–7 months, infants can detect correspondences between acoustic and optic information that specifies an affective expression (Walker, 1982). By 9 months, they are sensitive to the referential information of eye gaze and can perceive body movements as meaningful, goal-directed actions (Senju, Johnson, & Csibra, 2006); a few months later, they can parse continuous actions according to intentional boundaries (Baldwin & Baird, 2001; Baldwin, Baird, Saylor, & Clark, 2001). By 18 months, they can complete an observed subject's incomplete action, suggesting an understanding of the subject's unfulfilled intention (Herrmann, Call, Hernandez-Lloreda, Hare, & Tomasello, 2007; Meltzoff, 1995). During this same period they begin to participate in cooperative actions—both problem-solving activities and social games—indicating the presence of shared intentionality which emerges through dyadic collaboration, rooted in responsive attentiveness to others (see also Gibbs, 2001; Warneken, Chen, & Tomasello, 2006). These and other studies suggest, therefore, that rudimentary forms of interpersonal understanding—an attunement and responsiveness to social affordances, including other's intentions—inhere *in the interactive process itself*, a basic form of social understanding constituted not by inferential theorizing but rather within the expressive perception-action loops of focused interaction that shape our context-specific responsiveness to others (Gallagher, 2008b, p. 540). Even young infants seem to perceive bodily movements and expressive actions as goal-directed and intentional without a conceptual capacity coming into play. They directly perceive dynamic interactions between behavior and environment as involving specific body parts,

[17] Kendon (1970) observed synchronous movements and gestural mirroring among interactants in a public drinking house, speculating that their bodily synchrony both heightened rapport as well as differentiated them from others (i.e., it erected a conventional engagement closure).

[18] I'm grateful to an anonymous reviewer for pressing this point.

oriented toward specific portions of space, and governed by specific temporal parameters (Proust, 2003, p. 300). More on this later.

5.3. Gestures simplify perception

In addition to scaffolding spatial arrangements that simplify choice by constraining or cueing social affordances, gestures also scaffold spatial arrangements that sculpt the attention of receiver—and thus ease their epistemic burden—by simplifying perception. In other words, gestures and bodily expression can manage we-space by eliminating, or at least reducing, perceptual interference and drawing the receiver's attention to salient features of a particular interpersonal context that convey intentions and afford social responses.

Consider breastfeeding, the newborn's most immediate and complex form of social interaction (Kaye, 1982, p. 37). Kenneth Kaye has argued that breast feeding—which consists of affective cycles of touch and movement—may play a crucial role in the infant's social-cognitive development. Human infants are the only mammalian infants who breastfeed in short bursts. When human infants pause in their feeding all mothers—including new mothers who've never held a baby—instinctively jiggle the infant, or exhibit similar tactile behavior, as a prompt to resume feeding. And it seems to work: infants are more likely to resume feeding in the pause just after jiggling than they are during the jiggling or if they hadn't been jiggled at all (Kaye, 1982, p. 38). This behavioral interaction is significant in that it is arguably one of the earliest instances of interpersonal understanding: the mother's tactile behavior communicates a nonverbal intention, and the infant perceives and responds to this intention. And by organizing this primitive and repetitive engagement, the mother's actions provide the temporal prototype for the infant's future turn-taking interactions (Wexler, 2008, p. 111).

But are these interactions genuinely instances of shared *meanings*? Potentially so—as long as we accept, once again, that basic forms of interpersonal understanding can develop *interactively*, *sans* the explicit ascription of mental content (Hobson, 1993).[19] Granted, it's implausible that the infant grasps the mother's intention in an explicit or propositional sense. Nevertheless, the infant exhibits responsive tactile behavior of her own that affects the mother's state—the infant also sculpts her mother's attention to a degree—establishing this interaction as a mutually-governed, dynamically developing social system (Hopkins, 1983, p. 131). For instance, from the very first feeding onward, mothers adapt their tactile behavior to the bout-pause behavior of the babies' sucking (Alberts, Kalverboer, & Hopkins, 1983, p. 157); their behavior is modulated by the babies' sucking (Lavelli & Poli, 1998; Widström et al, 1990). But babies are equally as adaptable: they are reactive to the mother, consistently postponing their resumption of sucking until the mother ends her tactile behavior (Alberts et al., 1983, p. 157). Their mutual responsiveness thus establishes a tightly coupled system in which both partners play an equal role in regulating the exchange; its meaning emerges from the collaborative process of interaction.[20] Additionally, during breastfeeding another level of adaptive synchrony emerges via co-vocalization, in which even three day old infant's vocalizations are responsive to the presence and absence of maternal vocalization (Rosenthal, 1982).

Again, however, it starts with the mother bodily exploiting we-space to sculpt the infant's attention. First, the tactile behavior of breastfeeding structures a region of felt intimacy (conventional engagement closure) between infant and mother. The beginning of feeding is one of the few situations in which most infants keep their eyes open (Paul & Dittrichová, 1989); the relatively fixed distance between infant and mother provides optimal conditions for visual contact. And with the mother's tactile behavior (jiggling), the bodily structure of this we-space immediately orients the infant towards the mother's intention. It makes this intention explicit (resume feeding) as well as the salient local affordances (the nipple as affording feeding) the mother intends for the infant to make use of. The tactile behavior also reduces competing perceptual interference (ambient noise, mother's shirt as an object of interest, etc.), constraining the range of local affordances. Again, this is an instance of the mother bodily exploiting space to manage the infant's attention—which, in the long term, scaffolds the infant's developmental trajectory as it learns to care for itself by picking up and responding to salient affordances and by responding to others' intentions.

A similar attention–sculpting example can be found in the way that Zulu mothers, working in a culture where children are traditionally expected to be less socially active than contemporary European or North American children, cause their children to "thula" (fall silent) (Spurret & Cowley, 2010). During episodes of infant fussiness, the mother will sometimes say "thula" (quiet) or "njega" (no) while leaning forward

> so that her face and palms take up more of the infant's visual field. When this happens, new vocalizations and movements or reorientations of gaze by the infant are often "nipped in the bud" by dominating vocalizations. . .from the mother, sometimes accompanied by increasingly emphatic hand-waving and even closer crowding of the infant's visual field (Spurret & Cowley, 2010, p. 306).

[19] Kaye himself remains unconvinced that these early exchanges are genuinely social since they are "a neurological phenomenon" (rooting reflex) guided solely by the mother's intervention (Kaye, 1982, p. 40). I disagree with this teleological assessment for the reasons I give here.

[20] Mothers' social behavior is also modulated by the baby in other ways. For instance, mothers of infants who touch or lick the areola and nipple within 30 minutes of birth leave their infants in the nursery for shorter periods and speak more them during breastfeeding periods than do controls (Widström et al., 1990). Another example stems from the fact that, by the third month, the coordination of the head and eye is nearly mature (Trevarthen, 1974). The muscles governing head control are quite developed relative to other muscles at birth (Schloon, O'Brien, Scholten, & Prechtl, 1976), providing the young infant with a precocious approach-avoidance mechanism (Hopkins, 1983). Though lacking the ability to point, even two-week olds initiate processes of "co-orientation" in which they sculpt their mothers' attention by "pointing" with their head and eyes (Hopkins, 1983).

This bodily negotiation of we-space makes it difficult for the infant to attend to competing distractions. And the mother's expressive coordination with the infant's behavior (restricting approval signals such as diminished gesturing, increased smiles, comforting vocalizations and touch for moments when the child begins to quiet) co-orients the dyadic unit toward a shared attentional focus (Hopkins, 1983); the mother skillfully exploits we-space to manipulate the child into silence.

5.4. Gestures simplify internal computation

Finally, if the analysis of the previous sections is on the right track, the visuo-spatial dynamics of gesture and expression provide direct, real-time access to cognition- and emotion-in-action. Accordingly, the problem space of interpersonal understanding is shifted from individual heads into the interactive space of embodied engagements. This is because gestures scaffold and partially externalize intentions—intentions become perceptually accessible in gesture and action, instead of residing exclusively inside the head—which reduces the need for inferential theorizing or projective simulating (the core mechanisms within the ToM paradigm) (Gallagher, 2008b). An epistemically demanding cognitive process is transformed into a less demanding process of direct perception and interactional engagement, easing the computational burden on the receiver by making vital social information accessible in a perceptually immediate way. And by easing the receiver's informational burden, the gesturer plays a crucial role in shaping their responses and thus actively manipulates the local structure of we-space. Merleau-Ponty characterizes the phenomenological aspect of this engagement thusly: "The communication or comprehension of gestures comes about through the reciprocity of my intentions and the gestures of others, of my gestures and the intentions discernible in the conduct of other people. It is as if the other person's intentions inhabited my body and mine his (Merleau-Ponty, 1945/1962, p. 215).

How does this computational simplification occur? Again, the idea is not that *all* of cognition is externalized and publically accessible. Rather, *some* of it is, enough such that the dynamics and structure of the embodied engagement itself is sufficient to secure a relatively rich kind of interpersonal understanding. Additionally, by focusing on the way that interpersonal understanding is driven by the mutual manipulation of we-space, *action*—and not simply detached theorization or simulating—becomes prioritized as a valid source of attaining knowledge of others (Scribner, 1986). For it is within the real-time dynamical flow of embodied engagement that we have perceptual access to fine-grained social information not so obvious from a detached theoretical standpoint (Bavelas, Black, Lemery, & Mullett, 1987). Crucially, this includes sensitivity to various social contingencies (affordances), some of which have already been mentioned: subtleties of intention, of "responsiveness, of emotional attentiveness, of responsive or emotion-filled pauses, of the coordination of the other's expressions—widening of the eyes, partial opening of the mouth, sudden stilling of the limbs, the quality of the attention directed to us—in invitation or response to us" (Reddy & Morris, 2004, pp. 657–658).

The high-speed and dynamic character of our engagements indicates that developing perceptual sensitivity to the timing of social contingencies is a critical social skill and the bedrock of social interaction (Crown, Feldstein, Jasnow, Beebe, & Jaffe, 2002). Gestural and expressive strategies for managing we-space highlight the nature of this contingency-detecting skill. In Kirsh's treatment, our space-structuring actions reduce the descriptive complexity of our environment by making it easier to track the state of that environment (Kirsh, 1995, p. 65). By scaffolding intentions, bodily expressivity affords a similar state-tracking function. Gestures make it easier to detect and respond to social contingencies (and their emotional significance) than if intentions were localized exclusively in the head, requiring the adoption of a more demanding theoretical or simulative stance.

This idea is empirically supported by research indicating that a perceptual contingency-detecting skill is at the root of our earliest social engagements. Based upon microanalysis of face-to-face infant-mother interaction, it seems that caregivers are, not surprisingly, the main organizers and initiators of these early social exchanges (Stern, 1977). Caregivers "scaffold their infants within particular play frames characterized by exaggerated contours, marked changes of tempo, and systematic repetitions" (Rochat, Querido, & Striano, 1999, p. 951). They offer up expressive "packages" of facial-visual-tactual-auditory information that infants detect and exploit to enact an infant-caregiver "interactional synchrony" (Beebe & Gerstman, 1984).[21] As previously noted, this engagement takes the form of imitation (Kugiumutzakis, 1999; Meltzoff & Moore, 1977, 1997). Infants almost immediately begin monitoring caretakers' eyes to establish eye contact (Bruner, 1999; Reddy, 2003) and smile more when adults look at them and less when they look away (Muir & Hains, 1999; Reddy, 2003). Within weeks of birth, however, this interactional synchrony becomes a genuine dyadic exchange established by the precise temporal coordination (synchronized to within .10 seconds) of lip and tongue movements, expressive head movements, eye movements, hand gestures, finger movements, and pointing—bodily-expressive "proto-conversations" (Trevarthen, 1974, 1977). Face-monitoring quickly develops into look-regulation (Fogel, 1993; Stern, 1985; Tronick, 2003); imitative actions quickly become reciprocal engagements involving shared arousal and affectivity (Reddy, 2008, p. 59). And the key component of these engagements is the rhythmic coupling of patterns of expressive movement. Within these embodied couplings, infants can more readily discern socially salient information since, crucially, the sensorimotor skills we have for dealing with the external world—especially at a very young age—go beyond those we have for dealing with the internal world (Kirsh, 1995, p. 64).

This idea is also affirmed by research on sensitivity to teasing, which necessarily involves sensitivity to others' intentions (without this sensitivity teasing would not be perceived as such). By 7–9 months, infants perceive certain actions as playful

[21] The Zulu mothers in the previous section are one instance of such a "package".

intentions (ambiguous acts like offering and withdrawing objects) with different goals and outcomes than when the same intentions are interpreted literally (Legerstee, 2005, p. 124; Reddy, 1991, 2008). But this perceptual sensitivity is present even earlier. For example, 5.5 month-old infants are similarly sensitive to teasing intentions within natural play conditions; they distinguish between caregivers' mischievous versus neutral-faced expressions when a ball is offered and then taken away—they spend more time inspecting the first kind of look than the second—and produce more person-specific than object-specific looks (Legerstee, 2005). Infants know where to look to perceive intentions. Other evidence—much of which has already been surveyed—similarly suggests that infants are, from a very young age, innately disposed to directly (i.e., nonconceptually) see certain actions and bodily movements as goal-directed and intentional. As mentioned several times now, newborns as young as ten minutes old perceive facial movements not as mere sensorimotor patterns but both as socially salient as well as in an action-specific format that can be mapped onto their own motor possibilities (Meltzoff & Moore, 1977). Three month-old infants can discriminate biological motion from non-biological movements, even when the visual cues are limited to a few light points on the moving joints themselves (Johansson, 1977). Perceptual access to others' intentions is possible because the structure and dynamics of expressive actions embody intentional content. Crucially, even young infants perceive certain patterns of gestural behavior as associated with some target event—whether that target event is communication, teasing, sharing of affect or attention, or some other process intended to influence their motivational states in some way (Proust 2003, p. 302). It seems, therefore, that "perception offers all the evidence needed, in many cases, for judging—without inference—not only what an agent does, but what she is up to" (Proust, 2003, p. 303).[22] Within focused interactions, gestures can exhibit an epistemic function by offering up socially salient information (e.g., intentions)—gesture becomes a knowledge-oriented *action*—that simplifies the computational burden for the interactant and, in so doing, manipulates we-space and furthers even rudimentary processes of interpersonal understanding.

6. Concluding thoughts

I have argued that aspects of EM offer insight into understanding the interactive nature of basic levels of social understanding. Specially, I have drawn upon several lines of empirical research to defend the idea that gesture and bodily expressivity may in fact be a kind of material scaffolding essential to some basic social-cognitive and affective processes. I explored this idea by looking at the cognitive and affective benefits of gesture from both the perspective of the sender as well as the receiver, and attempted to highlight the role they potentially play in managing interpersonal we-space.

To be clear, the discussion above is not intended to offer a comprehensive theory of social cognition. Obviously we do not always perceive intentions, motives, or emotions in others in this immediate perceptual and nonconceptual way. Nor does this view preclude an important place for theories or simulation. There are occasions when we might employ theorizing, such as when trying to see through another's suspicious utterances or behavior in order to discern their true motives and intentions. Likewise, we might employ simulation by imaginatively placing ourselves in another's position when they exhibit behavior we find ambiguous, perplexing or strange. Both theorizing and simulating have a heuristic value that makes them important tools in the mature adult's social toolkit. Additionally, many of our social engagements simply fall outside the scope of immediate face-to-face encounter. For example, when speaking to someone on the telephone or via an instant message client on our computer, many, if not all, of the expressive contingencies discussed above are absent from that communicative context. Yet we are by and large able to understand one another within these disembodied encounters. Once more, this is because we have other strategies to employ beyond always having to rely upon the active bodily negotiation of we-space. But the point, rather, is that these are derived cases of social understanding which rest on more basic embodied skills—skills cultivated and refined within our history of face-to-face engagements. And it is here that EM might be summoned to do some helpful explanatory work, both in directly challenging the "Myth of the Hidden" informing dominant approaches within the TOM paradigm as well as by lending conceptual resources for clarifying how body and space are exploited in the process of engaging with and understanding others.

Acknowledgments

I am grateful to audiences in Copenhagen, Lyon, and Dublin, as well as three anonymous reviewers, for helpful comments on earlier versions of this paper. Special thanks also to Nivedita Gangopadhyay and Julian Kiverstein for their exceedingly thorough comments.

References

Alberts, E., Kalverboer, A. F., & Hopkins, B. (1983). Mother-infant dialogue in the first days of life: An observational study during breast-feeding. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 24*(1), 145–161.
Andrade, J. (2010). What does doodling do? *Applied Cognitive Psychology, 24*(1), 100–106.
Argyle, M. (1975). *Bodily communication*. New York: International Universities Press.
Baldwin, D., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences, 5*(4), 171–178.
Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72*(3), 708–717.
Baron, R. M. (2007). Situating coordination and cooperation between ecological and social psychology. *Ecological Psychology, 19*(2), 179–199.

[22] See Proust (2003) for an extended defense of this idea.

Baron-Cohen, S., & Frith, U. (1985). Does the autistic child have a theory of mind? *Cognition, 21*(1), 37–46.
Bavelas, J. B., Black, A., Lemery, C. R., & Mullett, J. (1987). Motor mimicry as primitive empathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its Development* (pp. 317–338). Cambridge: Cambridge University Press.
Beebe, B., & Gerstman, L. (1984). A method of defining 'packages' of maternal stimulation and their functional significance for the infant with mother and stranger. *International Journal of Behavioral Development, 7*(4), 423–440.
Bernstein, N. (1967). *Coordination and regulation of movements*. New York: Pergamon Press.
Brugger, P., Kollias, S. S., Müri, R. M., Crelier, G., Hepp-Reymond, M., & Regard, M. (2000). Beyond re-membering: Phantom sensations of congenitally absent limbs. *Proceedings of the National Academy of Sciences of the United States of America, 97*(11), 6167–6172.
Bruner, J. (1999). The intentionality of referring. In P. Zelazo & J. Astington (Eds.), *Developing theories of intention: Social understanding and self-control* (pp. 329–339). Manwah, New Jersey: Earlbaum.
Butterworth, B., & Beattie, G. (1978). Gestures and silence as indicators of planning in speech. In R. Campbell & P. Smith (Eds.), *Recent advances in the psychology of language: Formal and experimental approaches* (pp. 347–360). London: Plenum.
Cardinali, L., Brozzoli, C., & Farnè, A. (2009). Peripersonal space and body schema: Two labels for the same concept? *Brain Topography, 21*(3), 252–260.
Carruthers, P. (1996). Simulation and self-knowledge: A defence of theory–theory. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 22–38). Cambridge: Cambridge University Press.
Chartrand, T., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893–910.
Church, R., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23*, 43–71.
Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge: MIT Press.
Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford: Oxford University Press.
Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19.
Cole, J. (1999). On 'Being Faceless': Selfhood and facial embodiment. In S. Gallagher & J. Shear (Eds.), *Models of the self* (pp. 301–318). Charlottesville: Imprint Academic.
Cole, J. D., Merton, W. L., Barrett, G., Katifi, H. A., & Treede, R. D. (1995). Evoked potentials in a subject with a large-fibre sensory neuropathy below the neck. *Canadian Journal of Physiology and Pharmacology, 73*(2), 234–245.
Cole, J., Gallagher, S., & McNeill, D. (2002). Gesture following differentiation: A phenomenologically informed experimental study. *Phenomenology and the Cognitive Sciences, 1*(1), 49–67.
Cole, J., & Spalding, H. (2009). *The invisible smile: Living without facial expression*. Oxford: Oxford University Press.
Cole, J. (2010). Agency with impairments of movement. In D. Schmicking & S. Gallagher (Eds.), *Handbook of phenomenology and cognitive science* (pp. 655–670). Dordrecht: Springer.
Cook, S. W., & Goldin-Meadow, S. (2006). The role of gesture in learning: Do children use their hands to change their minds? *Journal of Cognition and Development, 7(2)*, 211.
Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106*(2), 1047–1058.
Corballis, M. (1994). Neuropsychology of perceptual functions. In D. Zaidel (Ed.), *Neuropsychology handbook of perception and cognition* (pp. 83–104). San Diego: Academic Press.
Cowley, S. J. (2003). Distributed cognition at three months: Mother–infant dyads in kwaZulu Natal. *Alternation, 10*(2), 229–257.
Crowder, E., & Newman, D. (1993). Telling what they know: The role of gesture and language in children's science explanations. *Pragmatics and Cognition, 1*, 173–208.
Crown, C. L., Feldstein, S., Jasnow, M. D., Beebe, B., & Jaffe, J. (2002). The cross-modal coordination of interpersonal timing: Six-week-olds infants' gaze with adults' vocal behavior. *Journal of Psycholinguistic Research, 31*(1), 1–23.
Currie, G., & Ravenscroft, I. (2002). *Recreative minds: Imagination in philosophy and psychology*. Oxford: Oxford University Press.
De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences, 6*(4), 485–507.
Di Paolo, E. (2009). Editorial: The social and enactive mind. *Phenomenology and the Cognitive Sciences, 8*(4), 409–415.
Dokic, J., & Proust, J. (2002). *Simulation and knowledge of action (Vol. 45)*. Amsterdam: John Benjamins Publishing Company.
Duclos, S. E., Laird, J. D., Schneider, E., Sexter, M., Stern, L., & Van Lighten, O. (1989). Emotion-specific effects of facial expressions and postures on emotional experience. *Journal of Personality and Social Psychology, 57*(1), 100–108.
Duclos, S. E., & Laird, J. D. (2001). The deliberate control of emotional experience through control of expressions. *Cognition & Emotion, 15*(1), 27–56.
Edelman, B. (1984). A multiple-factor of body weight control. *Journal of General Psychology, 110*, 99–114.
Ekman, P. (1965). Differential communication of affect by head and body cues. *Journal of Personality and Social Psychology, 2*(5), 726–735.
Ekman, P., & Friesen, W. V. (1972). Hand movements. *The Journal of Communication, 22*(4), 353–374.
Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother-stranger face discrimination by the newborn. *Infant Behavior and Development, 7*(1), 19–25.
Flack, W., Laird, J. D., & Cavallaro, L. A. (1999). Separate and combined effects of facial expressions and bodily postures on emotional feelings. *European Journal of Social Psychology, 29*(2–3), 203–217.
Fogel, A. (1993). *Developing through relationships: Origins of communication, self and culture*. Chicago: Chicago University Press.
Fogel, A., & Garvey, A. (2007). Alive communication. *Infant Behavior and Development, 30*(2), 251–257.
Fuchs, T., & De Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences, 8*(4), 465–486.
Gallagher, S. (2004). Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, & Psychology, 11*(3), 199–217.
Gallagher, S. (2005). *How the body shapes the mind*. Oxford: Oxford University Press.
Gallagher, S. (2008a). Inference or interaction: Social cognition without precursors. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action, 11*(3), 163–174.
Gallagher, S. (2008b). Direct perception in the intersubjective context. *Consciousness and Cognition, 17*(2), 535–543.
Gallagher, S., & Zahavi, D. (2008). *The phenomenological mind: An introduction to philosophy of mind and cognitive science*. New York: Routledge.
Gallagher, S., Cole, J., & McNeill, D. (2001). The language-thought-hand system. In C. Cave, I. Guaitella, & S. Santi (Eds.), *Oralite et gestualite: Interactions et comportments multimodaux dans la communication* (pp. 420–424). Paris: L'Harmattan.
Gallese, V. (2001). The 'shared manifold' hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies, 8*(5–7), 33–50.
Gallese, V. (2009). Mirror neurons, embodied simulation, and the neural basis of social identification. *Psychoanalytic Dialogues: The International Journal of Relational Perspective, 19*(5), 519–536.
Gibbs, R. W. (2001). Intentions as emergent products of social interactions. In B. F. Malle, L. Moses, & D. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 105–122). Cambridge: MIT Press.
Gibbs, R. W. (2006). *Embodiment and Cognitive Science*. Cambridge: Cambridge University Press.
Giddens, A. (1984). *The constitution of society*. Cambridge: Polity Press.
Goffman, E. (1963). *Behavior in public places*. New York: The Free Press.
Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences, 3*(11), 419–429.
Goldin-Meadow, S. (2000). Beyond words: The importance of gesture to researchers and learners. *Child Development, 71*(1), 231–239.
Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge: Belknapp Press.
Goldin-Meadow, S. (2009). How gesture promotes learning throughout childhood. *Child Development Perspectives, 3*(2), 106–111.
Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*(6), 516–522.

Goldin-Meadow, S., & Wagner, S. (2005). How our hands help us learn. *Trends in Cognitive Sciences, 9*(5), 234–241.
Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford: Oxford University Press.
Goldman, A. I., & Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition, 94*(3), 193–213. doi:10.1016/j.cognition.2004.01.005.
Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.
Gordon, R. M. (1996). 'Radical' simulationism. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind*. Cambridge: Cambridge University Press.
Hadar, U. (1989). Two types of gesture and their role in speech production. *Journal of Language and Social Psychology, 8*(3–4), 221–228.
Heal, J. (1986). Replication and functionalism. In J. Butterfield (Ed.), *Language, mind, and logic* (pp. 135–150). Cambridge: Cambridge University Press.
Herrmann, E., Call, J., Hernandez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317*(5843), 1360–1366.
Hobson, P. (2002). *The cradle of thought: Exploring the origins of thinking*. London: MacMillan.
Hobson, R. P. (1993). The emotional origins of social understanding. *Philosophical Psychology, 6*(3), 227–249.
Hobson, P. (2008). Interpersonally situated cognition. *International Journal of Philosophical Studies, 16*(3), 377.
Hopkins, B. (1983). The development of early non-verbal communication: An evaluation of its meaning. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 24*(1), 131–144.
Hurley, S. L. (1998). *Consciousness in action*. Cambridge: Harvard University Press.
Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.
Hutto, D. (2008). *Folk psychological narratives: The sociocultural basis of understanding reasons*. Cambridge: MIT Press.
Iverson, J. M. (1999). How to get to the cafeteria: Gesture and speech in blind and sighted children's spatial descriptions. *Developmental Psychology, 35*(4), 1132–1142.
Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature, 396*(6708), 228.
Iverson, J. M., Tencer, H. L., Lany, J., & Goldin-Meadow, S. (2000). The relation between gesture and speech in congenitally blind and sighted language-learners. *Journal of Nonverbal Behavior, 24*(2), 105–130.
Iverson, J. M., & Goldin-Meadow, S. (2001). The resilience of gesture in talk: Gesture in blind speakers and listeners. *Developmental Science, 4*(4), 416–422.
Iverson, J., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies, 6*(11–12), 19–40.
James, W. (1890). *The principles of psychology (Vols. 1 and 2)*. New York: Dover.
Johansson, G. (1977). Studies on visual perception of locomotion. *Perception, 6*(4), 365–376.
Kaye, K. (1982). *The mental and social life of babies: How parents create persons*. Chicago: Chicago University Press.
Kellerman, J. M., & Laird, J. D. (1982). The effect of appearance on self-perceptions. *Journal of Personality, 50*(3), 296–351.
Kendon, A. (1970). Movement coordination in social interaction: Some examples described. *Acta Psychologica, 32*, 101–125.
Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence, 73*(1–2), 31–68.
Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science, 18*(4), 513–549.
Kugiumutzakis, G. (1999). Genesis and development of early infant mimesis to facial and vocal models. In J. Nadel & G. Butterworth (Eds.), *Imitation in infancy* (pp. 127–185). Cambridge: Cambridge University Press.
Kugiumutzakis, G., Vitalaki, E., Kokkinaki, T., & Makrodimitraki, M. (2005). Emotions in early mimesis. In J. L. Nadel & D. Muir (Eds.), *Emotional development* (pp. 161–182). Oxford: Oxford University Press.
Laird, J. D. (2007). *Feelings: The perception of self*. Oxford: Oxford University Press.
Laird, J. D., & Bresler, C. (1992). The process of emotional experience. A self-perception theory. In M. Clark (Ed.), *Emotion: Review of personality and social psychology (Vol. 13)* (pp. 213–234). Newbury Park, CA: Sage.
LaFrance, M. (1985). Postural mirroring and intergroup relations. *Personality and Social Psychology Bulletin, 11*(2), 207–217.
Laland, K. N., Odling-Smee, J., & Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences, 23*(01), 131–146.
Lavelli, M., & Poli, M. (1998). Early mother–infant interaction during breast- and bottle-feeding. *Infant Behavior and Development, 21*(4), 667–683.
Legerstee, M. (2005). *Infants' sense of people: Precursors to a theory of mind*. Cambridge: Cambridge University Press.
Legrand, D., Brozzoli, C., Rossetti, Y., & Farnè, A. (2007). Close to me: Multisensory space representations for action and pre-reflexive consciousness of oneself-in-the-world. *Consciousness and Cognition, 16*(3), 687–699.
Leslie, A. (2004). Children's understanding of the mental world. In R. L. Gregory (Ed.), *The Oxford companion to the mind* (pp. 167–169). Oxford: Oxford University Press.
Levelt, W. J., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language, 24*, 133–164.
Lieberman, P. (1967). *Intonation, perception, and language*. Cambridge: MIT Press.
Lungarella, M., & Sporns, O. (2005). Information self-structuring: Key principle for learning and development. In *Proceedings of the 4th international conference on development and learning* (Vol. 5, pp. 25–30).
Malloch, S. (1999). Mothers and infants and communicative musicality. *Musicae scientiae, special issue on rhythms, musical narrative, and the origins of human communication* (pp. 29–57).
Malloch, S., & Trevarthen, C. (2009). Musicality: Communicating the vitality and interests of life. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 1–11). Oxford: Oxford University Press.
Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*(23), 1994–1997.
McNeill, D. (2005). *Gesture and thought*. Chicago: Chicago University Press.
Mehler, J., Bertoncini, J., & Barriere, M. (1978). Infant recognition of mother's voice. *Perception, 7*(5), 491–497.
Meltzoff, A. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*(5), 838–850.
Meltzoff, A., & Moore, K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*, 75–78.
Meltzoff, A., & Moore, K. (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting, 6*, 179–192.
Menary, R. (2007). *Cognitive integration: Mind and cognition unbounded*. Basingstoke: Palgrave Macmillan.
Merleau-Ponty, M. (1942). *The structure of behavior (A. L. Fisher, Tran.)*. Pittsburgh: Dusquene University Press.
Merleau-Ponty, M. (1945). *Phenomenology of perception (C. Smith, Tran.)*. London: Routledge, Kegan Paul.
Muir, D. W., & Hains, S. (1999). Young infants' perception of adult intentionality: Adult contingency and eye direction. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life* (pp. 155–187). Manwah, New Jersey: Earlbaum.
Murray, L., & Trevarthen, C. (1985). Emotional regulation of interactions between two month-olds and their mothers. In T. Field & N. Fox (Eds.), *Social perception in infants*. Norwood, NJ: Ablex.
Nadel, J., Guerini, C., Peze, A., & Rivet, C. (1999). The evolving nature of imitation as a format for communication. In J. Nadel & G. Butterworth (Eds.), *Imitation in infancy* (pp. 209–234). Cambridge: Cambridge University Press.
Nagy, E., & Molnar, P. (2004). Homo imitans or homo provocans? Human imprinting model of neonatal imitation. *Infant Behavior and Development, 27*(1), 54–63.
Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*, 1002–1005.
Niedenthal, P., Barsalou, L. W., Ric, F., & Krauth-Gruber, S. (2005). Embodiment in the acquisition and use of emotion knowledge. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 21–50). New York: The Guilford Press.
Niedenthal, P. M., & Maringer, M. (2009). Embodied emotion considered. *Emotion Review, 1*(2), 122–128.

Overgaard, S. (2006). The problem of other minds: Wittgenstein's Phenomenological perspective. *Phenomenology and the Cognitive Sciences, 5*(1), 53–57.
Paul, K., & Dittrichová, J. (1989). Feeding behaviour in infants. *Czechoslovak Medicine, 12*(4), 224–240.
Perner, J., & Kuhberger, A. (2005). Mental simulation: Royal road to other minds? In B. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 158–173). New York: Guilford Press.
Plantinga, J., & Trainor, L. J. (2009). Melody recognition by two-month-old infants. *The Journal of the Acoustical Society of America, 125*(2), 58–62.
Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences, 1*, 515–526.
Proust, J. (2003). Perceiving intentions. In J. Roessler & N. Eilan (Eds.), *Agency and self-awareness: Issues in philosophy and psychology* (pp. 296–320). Oxford: Oxford University Press.
Ramachandran, V., & Blakeslee, S. (1998). *Phantom limbs in the brain: Probing the mysteries of the human mind*. New York: William Morrow.
Reddy, V. (1991). Playing with others' expectations: Teasing and mucking about in the first year. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation in everyday mindreading* (pp. 143–158). Oxford: Blackwell.
Reddy, V. (2003). On being the object of attention: Implications for self-other consciousness. *Trends in Cognitive Sciences, 7*(9), 397–402.
Reddy, V. (2008). *How infants know minds*. Cambridge: Harvard University Press.
Reddy, V., & Morris, P. (2004). Participants don't need theories: Knowing minds in engagement. *Theory & Psychology, 14*(5), 647–665.
Rietveld, E. (2008). The skillful body as a concernful system of possible actions. *Theory & Psychology, 18*(3), 341–363.
Rives Bogart, K., & Matsumoto, D. (2010). Living with Moebius syndrome: Adjustment, social competence, and satisfaction with life. *The Cleft Palate-Craniofacial Journal, 47*(2), 134–142.
Rochat, P. (2009). *Others in mind: Social origins of self-consciousness*. Cambridge: Cambridge University Press.
Rochat, P., Querido, J. G., & Striano, T. (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental Psychology, 35*(4), 950–957.
Rosenthal, M. (1982). Vocal dialogues in the neonatal period. *Developmental Psychology, 18*(1), 17–21.
Rowe, M. L., Ozcaliskan, S., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language, 28*(2), 182–199.
Rowlands, M. (2003). Externalism: Putting mind and world back together again. McGill-Queen's University Press.
Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals. *Psychological Science, 7*(5), 272–277.
Schloon, H., O'Brien, M. J., Scholten, C. A., & Prechtl, H. F. (1976). Muscle activity and postural behaviour in newborn infants. A polymyographic study. *Neuropädiatrie, 7*(4), 384–415.
Scribner, S. (1986). Thinking in action: Some characteristics of practical thought. In R. Sternberg & R. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 13–30). Cambridge: Cambridge University Press.
Senju, A., Johnson, M. H., & Csibra, G. (2006). The development and neural basis of referential gaze perception. *Social Neuroscience, 1*(3–4), 220–234.
Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science, 1*(2), 305–319.
Spurret, D., & Cowley, S. (2010). The extended infant: Utterance-activity and distributed cognition. In R. Menary (Ed.), *The extended mind* (pp. 295–323). Cambridge: MIT Press.
Sterelny, K. (2004). Externalism, epistemic artefacts and the extended mind. In R. Schantz (Ed.), *The externalist challenge* (pp. 239–254). Berlin: Walter de Gruyter.
Stern, D. (1977). *The first relationship: Infant and mother*. Cambridge: Harvard University Press.
Stern, D. (1985). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology*. New York: Basic books.
Still, A., & Good, J. (1998). The ontology of mutualism. *Ecological Psychology, 10*(1), 39–63.
Torrance, S. (2009). Contesting the concept of consciousness. *Journal of Consciousness Studies, 16*, 111–126.
Trehub, S. E., & Schellenberg, E. G. (1995). Music: Its relevance to infants. *Annals of Child Development, 11*, 1–24.
Trehub, S. E., Schellenberg, E. G., & Kamenetsky, S. (1999). Infants' and adults' perception of scale structure. *Journal of Experimental Psychology, 25*(4), 965.
Trehub, S. E., Thorpe, L. A., & Morrongiello, B. A. (1987). Organizational processes in infants' perception of auditory patterns. *Child Development, 58(3)*, 741.
Trehub, S. E., & Trainor, L. (1993). Listening strategies in infancy: The roots of music and language development. In S. Mcadams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition*. Oxford: Oxford University Press.
Trevarthen, C. (1974). The psychobiology of speech development. *Neurosciences Research Program Bulletin, 12*, 570–585.
Trevarthen, C. (1977). Descriptive analyses of infant communicative behavior. In H. Schaffer (Ed.), *Studies in mother–infant interaction* (pp. 227–270). London: Academic 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: Cambridge University Press.
Trevarthen, C. (1992). The self born in intersubjectivity: The psychology of an infant communicating. In U. Neisser (Ed.), *The perceived self: Ecological and interpersonal sources of self-knowledge*. Cambridge: Cambridge University Press.
Trevarthen, C. (1993). The functions of emotion in early infant communication and development. In J. Nadel & L. Camaioni (Eds.), *New perspectives in early communicative development* (pp. 279–308). London: Routledge.
Trevarthen, C. (2002). Origins of musical identity: Evidence from infancy for musical social awareness. In R. Macdonald, D. Hargreaves, & D. Miell (Eds.), *Musical identities* (pp. 21–38). Oxford: Oxford University Press.
Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child Psychiatry, 17*(1), 1–13.
Tronick, E. Z. (2003). Things still to be done on the still-face effect. *Infancy, 4*(4), 475–482.
Turvey, M. T. (1990). Coordination. *American Psychologist, 45*(8), 938–953.
Uzgiris, I. C. (1981). Two functions of imitation during infancy. *International Journal of Behavioral Development, 4*(1), 1–12.
Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology, 33*(3), 514–535.
Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development, 77*(3), 640–663.
Wendrich, K. (1981). Pitch imitation in infancy and early childhood: Observations and implications. Unpublished doctoral dissertation, University of Connecticut, Storrs.
Wexler, B. (2008). *Brain and culture: Neurobiology, ideology, and social change*. Cambridge: MIT Press.
Whiten, A., & Perner, J. (1991). *Fundamental issues in the multidisciplinary study of mindreading*. Oxford: Blackwell.
Widström, A., Wahlberg, V., Matthiesen, A., Eneroth, P., Uvnäs-Moberg, K., Werner, S., et al (1990). Short-term effects of early suckling and touch of the nipple on maternal behaviour. *Early Human Development, 21*(3), 153–163.
Zahavi, D. (2008). Simulation, projection and empathy. *Consciousness and Cognition, 17*(2), 514–522.
Zeedyk, M. S. (2006). From intersubjectivity to subjectivity: The transformative roles of emotional intimacy and imitation. *Infant and Child Development, 15*(3), 321–344.
Zhao, S. (2003). Toward a taxonomy of copresence. *Presence: Teleoperators and Virtual Environments, 12*(5), 445–455.