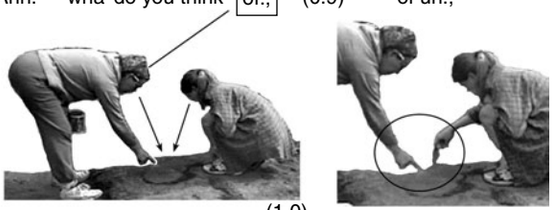
1 Embodied Interaction in the Material World: An Introduction

Jürgen Streeck, Charles Goodwin, and Curtis LeBaron

HUMAN ACTION

The chapters in this volume take as their focus the organization of action in human interaction. The question immediately arises as to where and how the structure of human action might be investigated. Different disciplines have taken very different kinds of phenomena, ranging from the mental intentions of individual actors to large, historically shaped social structures, as the proper locus for such a study. Here we take as our point of departure events in which multiple parties are carrying out endogenous courses of action in concert with each other within face-to-face human interaction. A concrete example can make clearer what we mean by this. In Transcript 1.1 Ann, a senior archeologist and director of the field school where the current excavation is taking place, is working with Sue, a new graduate student, as Sue works to outline the shape of an archeological feature faintly visible in the color patterning of the dirt they are examining (this sequence is examined in more detail, from a slightly different perspective in Goodwin (2007a).

1 Ann: wha' do you think of:, (0.9) of uh:,



2 (1.0)

3 Sue: Does it kinda go aro:und=

4 =┌en come ┌over around there

5 Ann: └*hhh └I think it's-

Transcript 1.1. Embodied Interaction.

The actions occurring in Transcript 1.1 are not organized within a single medium, such as talk, but are instead constructed through the simultaneous use of multiple semiotic resources with quite different properties. Thus in line 1, Ann says, "Wha'do you think of:,". In English, of begins a prepositional phrase that requires a noun for its grammatical completion. However, no appropriate noun occurs in Transcript 1.1. A similar argument can be made about "aro:und" in line 3, where the entity being gone around is never specified in the talk. If one focuses only on the talk occurring here, and the linguistic structure emerging within that talk, what is said here does not conform to the requirements of English grammar. However the participants do not in any way treat this talk as defective. Instead the "it" in Sue's line 3 "Does it kinda go aro:und" explicitly ties back to what Ann indicated, and thus not only treats what Ann was talking about as unproblematically understood, but incorporates that recognition into the structure of the utterance responding to Ann's talk.

There is of course no mystery in how Sue was able to appropriately understand what Ann was telling her. As Ann said "of:," in line 1, she used her right arm and index finger to point toward a particular patch of color patterning in the dirt they were working on together. The slot for the noun in the prepositional phrase in the talk was thus filled by the combination of a pointing gesture and the visible structure in the environment it indicated. Ann was showing Sue something in the dirt that should now become the focus of their joint scrutiny and work. Well before she produces "it" in line 3, Sue displays precisely this embodied, work-relevant understanding of the complex structure of Ann's action by moving her own hand and trowel to just the spot in the dirt that Ann indicated. She then uses that positioning as the point of departure for the gesture with the trowel tracing structure in the dirt that accompanies "kinda go around" in line 3.

The interaction between Ann, Sue, and the world that is the focus of their work is organized through the structured exchange of different kinds of signs. These include not only language but also a variety of diverse

signs constituted through the visible organization of the participants' bodies. Ann uses her pointing finger in line 1 to indicate to Sue a specific place in the dirt. Sue's movement of the trowel in line 3 is used to show Ann the path in the dirt that is described in the talk as "kinda go around" and thus constitutes a sign for that path. Each party builds action by producing signs for the other. Thus, to build relevant action in Transcript 1.1, the participants simultaneously make use of a number of quite different kinds of semiotic resources that have different properties and are instantiated in different kinds of semiotic materials (linguistic structure in the stream of speech, signs such as pointing displayed through the visible body, the patterning of phenomena in the environment that is the focus of their work, etc.). The recognizable and consequential actions they are building for each other cannot be found in any single semiotic medium. As noted earlier, by itself the talk is incomplete both grammatically and, more crucially, with respect to the specification of what the addressee of the action is to attend to in order to accomplish a relevant next action. Similarly the embodied pointing movements require the co-occurring talk to explicate the nature and relevance of what is being indicated. Indeed the mutual organization of talk and gesture has long been a central theme in gesture studies (Kendon, 2004; McNeill, 1992). By itself each individual set of semiotic resources is partial and incomplete (Agha, 2007; Goodwin, 2007a). However, when joined together in local contextures of action, diverse semiotic resources mutually elaborate each other to create a whole that is both greater than, and different from, any of its constituent parts (Goodwin, 2000a). Describing how action is built here thus requires an analytic framework that recognizes the diversity of semiotic resources used by participants in interaction, and takes into account how these resources interact with each other to build locally relevant action.

Having the ability to build action by combining resources with diverse properties has clear advantages and greatly expands the repertoire of possible action available to participants. To note one very simple example: In line 3, Sue is tracing with her trowel a complex, irregular shape in the dirt. Describing the precise shape of the phenomena they uncover in the dirt being excavated is crucial to the work of archeology. Suppose the resources available for doing this were restricted to a single semiotic field, such as language. If each different shape encountered in an excavation had to be categorized semantically, the vocabulary of archeology would quickly become unmanageably large – indeed, useless. However if a limited set of semantic categories ("feature," "postmold," "disturbance," etc.) can be supplemented by analogic signs capable of continuous variation (gestures over a shape such as line 3, drawings on maps, etc.), precision and flexibility become not only possible, but quite literally ready at hand as working hands and trowels articulate for others relevant structure in the world they are acting upon together.

To try and demonstrate as clearly as possible how action is built by combining resources with diverse properties that mutually elaborate each other, the discussion has so far been restricted to how talk, gesture, and structure in the world mutually elaborate each other. This might be glossed as the referential domain that the participants are focusing on: what they are talking about and formulating as particular kinds of structure in the dirt they are excavating. However this does not in any way exhaust the different kinds of semiotic resources that are implicated in the organization of their action.

For example, how can Ann unproblematically assume that Sue will take her gesture into account, something an addressee must do in order to properly understand what Ann is telling her and thus build an appropriate next action? Note that Ann places her gesture right in front of Sue's eyes, over the dirt she is already looking at. Ann treats Sue's gaze as a sign for where she is attending and what she is attending to. More generally, the mutual orientation of the participants' bodies creates what Goffman (1964: 64) called an "ecological huddle," which publicly demonstrates through visible embodied practice that the participants are mutually oriented toward each other and frequently toward particular places, objects, and events in the surrounding environment (Heath, Luff, vom Lehn, Hindmarsh and Cloeverly, 2002). Such embodied participation frameworks (Goodwin, 2000a) or F-formations (Kendon, 1990) are central to the organization of action in face-to-face interaction. Like gestures, these displays of mutual orientation are constructed through embodied signs. However, they differ from gesture in a number of important respects. First, they are not "about" the substance of what the participants are talking about (e.g., relevant structure in the dirt these parties are working on), but instead have as their subject matter the orientation of the participants toward each other, and the world that is the focus of their activity. Second, they have a quite different temporal organization. Unlike particular elements of talk, or specific gestures, which disappear and are replaced by other words or gestures almost as soon as they occur, embodied participation frameworks can be sustained over extended stretches of talk and action. Third, even not being about the substance of what is being talked about, they contribute to the organization of that talk in other important ways. For example, the shared orientational frameworks they make publicly visible deictically ground many of the indexical expressions that occur within that talk (including "you," "it," and "there" in Transcript 1.1) while making possible other indexical, context sensitive uses of language, such as the "incomplete" prepositional phrase noted earlier. These embodied orientational frameworks create local environments where participants can treat each other as attending to, and working together within, a shared world of perception and action, something crucial to the way in which Ann and Sue are building action together by attending to

how each other is interpreting and operating on the dirt that is the focus of their work. In essence, the signs used to create and continuously sustain, modify or dismantle participation frameworks (Goodwin, 1981, 2007b; Kendon, 1985) create a public semiotic environment within which other kinds of sign exchange processes, such as talk and gesture, can flourish.

Events of the type found in Transcript 1.1, in which multiple parties are carrying out a course of action together through the use of talk and other embodied action while attending to each other and frequently to the phenomena in the world that are the focus of their scrutiny and activities, provide a perspicuous environment for the systematic investigation of a range of phenomena that are central to the organization of human language, social organization, culture, and cognition. First, insofar as a common course of action is being accomplished through the joint, collaborative work of multiple parties, such events provide pervasive examples of elementary human social organization, a place where one can investigate in detail the actual practices used to build endogenous social order. Simmel (1950: 21–22) argued that "if society is conceived as interaction among individuals, the description of the forms of this interaction is the task of the science of society in its strictest and most essential sense." Such sites, in which action is organized with reference to the properties of embodied co-presence, render clearly visible many of the central features of human interaction noted by Goffman (1963), including mutual monitoring and the reflexivity of embodied interaction. Second, as has long been noted by conversation analysts (Sacks, 1992; Schegloff, 2006), face-to-face interaction is a central place where language emerges in the natural world. Third, if participants are to carry out courses of collaborative action together, they must in some relevant sense understand what each other is doing, and the nature and detailed structure of the events they are engaged in together. Such sites thus permit investigation of the practices of sense making noted by Garfinkel (1967) and of cognition as public practice more generally. They are also central to contemporary work in Europe, such as Linell (2009), which is attempting to rethink language, the mind, and the world dialogically. Fourth, though organized through general practices, the particulars of what participants must see and understand in order to build action together, such as how color patterns in a patch of dirt can be interpreted as archeological features, are lodged within specific communities. Situations such as these are places where the content and organization of culture as practice, as well as the ways in which such knowledge, skills and practices are appropriated by newcomers just entering its distinctive phenomenal world of a community, can be examined in fine detail (Sue is a beginning archeologist at her first excavation). Fourth, in such events, it is possible to investigate both the part played by the individual body in the organization of cognition and action, including how such bodies gain the skills required for relevant action within specific communities (Ingold, 2000), and how participants see and understand each other's bodies so that they can anticipate what each other is about to do and joint action can be successfully accomplished.

It is not being argued that such events are the only place where human action occurs, or that they are in some sense primordial. Many actions, such as the words now being written, are created by solitary individuals, though ones using culturally structured resources such as language. An individual can come to know the world and its distinctive properties through exploration and work with her own hands (Streeck, 2009), and much phenomenal knowledge is lodged within the experience of an individual embedded within a consequential world. The interactive organization of multi-party action does, however, provide a fruitful arena for investigating from an integrated perspective a host of crucial phenomena that are central to the organization of human action, cognition, and social life.

In brief, by looking at events such as that found in Transcript 1.1, it is possible to systematically examine some of the practices used by human beings to build action in concert with each other. As has long been strongly demonstrated by conversational analysts (Jefferson, 1988; Sacks, Schegloff, and Jefferson, 1974; Schegloff, 2007), sequential organization is central to both the structure of action and the way in which it is understood by the participants themselves (Sue's talk is built in response to what Ann has just said and done, and, as noted earlier, a number of constructional features of her utterance explicitly display this, including the "it," which ties to what Ann has just said and done). One phenomenon that quickly emerges from records that preserve not only the talk but also the bodies of actors, is that action is built through the mutual elaboration of diverse semiotic resources with quite different properties, each of which, including language, can make only a partial, incomplete contribution to the action in progress. The participants themselves attend to both this diversity and to the unique, distinctive contributions made by different kinds of semiotic resources. Thus Sue builds a new action to Ann both with talk and with relevant actions of her body – for example, by moving her own trowel to the place in the dirt indicated by Ann and then using that trowel to outline what she has been asked to see, and on another level, by aligning her body toward both Ann and the patch of dirt they are examining together.

A unifying thread running through all of the papers in this volume, though one developed in very different ways, is the systematic investigation of how multiple participants build action together in the midst of situated interaction, typically by using different kinds of semiotic resources that mutually elaborate each other. One aspect of this process that the current volume is not able to adequately address is prosody. However, this is the focus of rich and important work, much of it in

Europe, by several linked groups of scholars including Elizabeth Couper-Kuhlen, Margaret Selting, Dagmar Beth-Weingarten, Elisabeth Reberm, John Local and his collaborators in York, and many others (see, for example, Couper-Kuhlen and Selting, 1996b). This volume's focus on the organization of action within interaction differentiates it from some other approaches to what is sometimes glossed as multimodality. It is, however, consistent with a growing body of work, in Europe, Japan, and the United States, that has begun to engage in intensive analysis of how action is built through the inter-elaboration of talk, the body, encompassing activities and features of the setting (see, for example, Heath and Luff, 2000; Mondada, 2008a, 2008b; Nishizaka, 2007), and reflexive analysis of the transcription practices that can make such phenomena visible and amenable to analysis (Lindwall and Lymer, 2008; Murphy, 2005; Mondada, 2006).

The approach taken in this volume, with its focus on systematic investigation of the different kinds of semiotic resources and meaning-making practices that participants themselves attend to, and treat as relevant, as they build action within interaction together, seems to us not only fruitful, but straightforward and uncontroversial. The simultaneous use of diverse semiotic resources – currently discussed under the heading *multimodality* (see the fourth section of this chapter) – is pervasive in the organization of endogenous human action. The issue therefore arises as to why the relevance of adopting a perspective that takes this into account must be clearly argued. Briefly, much existing research has avoided the crucial issues posed by the heterogeneous semiosis that sits at the center of actual human action by focusing on the analysis of individual semiotic systems as self-contained wholes. For example, Saussure (1959: 16) envisioned a general science focused on "the life of signs within society." Such a goal is entirely compatible with the work in this volume. However, Saussure then argued that linguistics should confine its study to just one part of this larger field by investigating language as an isolated self-contained whole:

> A science that studies the life of signs within society is conceivable; it would be a part of social psychology and consequently of general psychology; I shall call it "semiology" (from Greek *semefon*, "sign"). Semiology would show what constitutes signs, what laws govern them. Since the science does not yet exist, no one can say what it would be; but it has a right to existence, a place staked out in advance. Linguistics is only a part of the general science of semiology; the laws discovered by semiology will be applicable to linguistics, and the latter will circumscribe a well-defined area within the mass of anthropological facts. To determine the exact place of semiology is the task of the psychologist! The task of the linguist is to find out what makes language a special system within the mass of semiological data.

Language is thus demarcated as a "special system" that not only can be, but should be investigated without reference to other semiotic processes with which it characteristically co-occurs. Delimiting the scope of inquiry in this way, and thus defining the phenomenal and analytic field within which all subsequent inquiry will occur, has had enormous consequences. Such limits defined the scope of formal linguistics, and were carried over as unquestioned assumptions when new fields, such as cognitive science, emerged. Thus it took much creative innovation for cognitive science to reshape itself so that phenomena such as embodiment (Clark, 1997; Gibbs, 2005) and the distribution of cognitive processes beyond the individual brain to encompass the situated practices of communities (Hutchins, 1995; Suchman, 1987) were recognized as essential to the analysis of human cognition.

From a slightly different perspective, human language possesses rich, intricate, and varied structure combined with extraordinarily powerful representational capacities. Moreover, for thousands of years it has been possible to use writing to capture much of this richness in another, more permanent medium. Writing does, however, have the effect of rendering invisible the embodied frameworks within which language in face-to-face interaction is embedded, including the crucial part played by co-present hearers. Rather than simply being constraints, the restrictions and distinctive properties of writing, as a semiotic medium in its own right, make possible new and important ways of using language and preserving some of its detailed structure not only across encounters, but also across generations. In part because of the powerful resources provided by writing, many fields, including some that strongly oppose the formal and monologic assumptions of Saussure and argue persuasively for the crucial importance of dialog (Bakhtin, 1981; Volosinov, 1973), have nonetheless restricted the scope of their inquiry to phenomena that fall within a broad conception of language. While offering a powerful and most important arena for study, such logocentricism – what Linell (2005) calls the written language bias in linguistics – nonetheless renders invisible many of the crucial forms of semiosis that shape human action in actual interaction (for example many of the embodied phenomena found in Transcript 1.1, as well as the crucial role of structure in the world that is a focus of the participants' talk and action).

Not all interaction occurs within the fully embodied frameworks of mutual orientation found in Transcript 1.1. Indeed this is a systematic consequence of the very semiotic structure of such events. Because action is being built through the co-articulation of different semiotic fields, it is possible to remove some of these fields while adapting the structure of others so that the accomplishment of relevant action remains. Throughout human history, from hunter gatherers talking across campfires in the dark to contemporary talk over telephones, human beings have been able to build rich interaction with each other through talk alone. Situations with such restricted semiotic structure do, however, eliminate for participants as well as analysts many of the crucial resources implicated in the organization of action in face-to-face

interaction. Thus, in fully embodied situations, utterances are not constituted exclusively within the stream of speech by the actions of the speaker. Instead the visible actions of hearers, including both orientation toward the speaker and operations on the specifics of the talk as it is being spoken, can systematically lead the speaker to change the structure of a sentence in progress (Goodwin, 1981; M. H. Goodwin, 1980). Many of the consequential actions of the hearer are performed through visible displays of the body rather than with talk. Within such frameworks, both the utterance and the turn-at-talk within which it emerges are not only intrinsically multiparty activities, but also ones built through the interplay of structurally different kinds of semiotic processes, including the talk of the speaker and the visual displays of hearer (the speaker also makes consequential visual displays, for example using gaze to indicate address). Noting this is not to deny the powerful analysis that has been developed from audio recordings of interaction, but it does demonstrate the relevance of analysis that takes into account the distinctive semiotic structure of fully embodied co-presence.

THE INTERACTIONIST PERSPECTIVE

The study of embodied interaction as it is presented in this book takes inspiration from a variety of sources, most of which are familiar names: Mead, Vygotsky, Bakhtin, Bateson, Goffman, Sacks, Schegloff and Jefferson, Kendon. Some would want to include Wittgenstein in the list, others Merleau-Ponty and Heidegger, or Bourdieu, de Certeau, and Marx. Even though there may be minor disagreements about the list, on the whole our field is not lacking in intellectual cohesion. We cannot account for these influences in detail, but want to remind the reader of some especially pertinent intellectual forces that continue to shape the ways in which interactionist researchers think about their subject matter and the proper ways to analyze it.

Of particular importance for work on embodied interaction has been G. H. Mead's critique of methodological individualism (Mead, 1909, 1934), that is, of those accounts of social life and symbolic interaction that posit the self as given and treat meaning, mind, and intersubjectivity as epiphenomena or products of individual minds. Mead (1934: 222–223) maintained that a theory which

> … assumes individual selves as the presuppositions, logically and biologically, of the social process or order within which they interact…., cannot explain the existence of minds and selves…. [In contrast, a theory which] assumes a social process or social order as the logical and biological precondition of the appearance of the selves of the individuals involved in that process or belonging to that social order, … can explain that which it takes as logically prior, namely the existence of the social process of behavior, in terms of such fundamental biological relations and interactions as reproduction.

Mead conceived interaction as a conversation of gestures. Gestures in Mead's conception are not hand gestures as they are studied today, but more broadly early parts of acts, components that can become separated as free-standing units with organic and motivated, yet conventional, relationships to the social acts in which they have emerged. Nevertheless, Mead's conception is quite compatible with interactionist accounts of hand gestures. He observed that

> … throughout the entire process of an interaction, we analyze the incipient actions of others by our own instinctive reactions to changes in their postures and other signs of developing social actions (Mead, 1909: 219).

Thus, making gestures that come from and designate acts, we creatively hatch courses of joint action. Through gestures in Mead's sense, we rapidly and incessantly indicate to one another – and thus prepare – what is to come next (McDermott and Roth, 1978). Mead draws our attention to the forward-design of human action. The foreshadowing of imminent actions is made possible not least by the multimodal structure of the human body – its ability to move some of its parts independently from one another and thus create multiple, heterogeneous signs at the same time.

As the self is mediated by interaction, it is also inextricably embedded in a community and draws on this community's historically evolved sense-making tools, in the first place a language and the typified categories of experience that it offers. Vygostky, a near-contemporary of Mead, called such "mediational means" (Wertsch, 1991) psychological tools (Vygotky, 1978). Individual minds are produced through cultural apprenticeship. Bakhtin (1986) proposed an analogous view of language: Speaking means to rent words from a community, to fashion oneself (and one's utterance) by using communal means. In every act of speaking, individual and society are intertwined.

From Gregory Bateson we have learned to think of speech and "nonverbal communication" not as a combination of signs, but as a relation between act and context. Contexts *frame* or *type* behavior. Context can be a metamessage, for example, "[T]his is play" (Bateson, 1956), which instructs us not to take anything that is contextualized by it at face value. But the relation is mutual: The context is also created by the act, a relationship that Gumperz (1992) expresses in his notion of "contextualization cues." The act is "part of the ecological subsystem called context and not … the product or effect of what remains of the context once the piece which we want to explain has been cut out from it" (Bateson, 1972: 338).

In Goffman's dramatistic view of interaction, characteristic especially of his earlier work (1959, but see 1976), the entire setting insofar as it is under the actor's control can be manipulated to display the committing of acts, to embody the working consensus, or to represent something as something else. He wrote: "[T]he representation

of an activity will vary in some degree from the activity itself and therefore misrepresent it" (Goffman, 1959: 45). He also noted that we cannot separate bodily signs from the settings in which the bodies that make them operate:

> The individual gestures with the immediate environment, not only with his body, and so we must introduce this environment in some systematic way … while the substratum of a gesture derives from the maker's body, the form of the gesture can be intimately determined by the microecological orbit in which the speaker finds himself. To describe the gesture, let alone uncover its meaning, we … have to introduce the human and material setting in which the gesture is made (Goffman, 1964: 164).

This rarely cited dictum could serve as a motto for this book; it presages the common ground of much contemporary research on embodied and multimodal interaction.

Goffman's term *footing* (Goffman, 1981) also reveals his interest in embodiment, in the question of how aspects of the interaction order are given corporeal form. The term "footing" designates the differing forms of alignment and presence in an utterance that can be taken up by the range of structurally differentiated participants who are implicated in the organization of a strip of talk. For example the current speaker, or animator, may be voicing words authored by either herself or others, and while quoting the words of others can display varying stances toward the talk and action being reported (see also Bakhtin, 1981; Goodwin, 2007b; Hanks, 1996; Levinson, 1988; Volosinov, 1973). Non-speaking participants can have a range of quite different kinds of alignment toward the current utterance, both in terms of typology of different kinds of hearers Goffman offered in footing, and with respect to local operations on the structure of emerging utterances (M. H. Goodwin, 1980). When we observe conversations among people who are standing, we can indeed often read off changes in footing from the reshuffling of the participants' feet, as they reconfigure their spatial arrangement: It was this type of modality-crossing representations of the interaction order that Goffman was especially interested in.

What inspires all contributions to this volume is a view of speakers and listeners as profoundly and inextricably "intervolved" (Dreyfus, 1991) with the material context that they operate in – with the world at hand (Schütz, 1967). When we imagine a speaker, we typically envision her with pen and wrench in hand, or preparing a blood vessel for surgery, or with feet firmly planted in a hopscotch grid. This analytic orientation – to picture speaker and listener at work, doing things *with things* (Streeck, 1996a) – resonates with a certain conception in philosophical anthropology, dating back to the Enlightenment, of humanity as *homo faber*, as makers of artifacts, caught up in the never-ending project of sustaining the world and surviving in it by making and remaking it over and over and over. A phenomenological perspective shapes the work of an increasing number of linguists, anthropologists, cognitive scientists, and other researchers of communicative practice (Gehlen, 1988; Hanks, 1996). Herder (1772), and Plessner (1965, 1980), among many others, have conceived of human-made material culture and language as an *Ersatz* for a missing biosphere. The human species suffers from its "excentric positionality" (Plessner, 1975) in the world: It is not biologically adapted to a biosphere, but must create its own artifactual work and adapt itself to it, each group to its own, in order to survive. The evolution of the human mind is part of this adaptation. Our ability to adopt a reflective attitude toward our own words and gestures – to regard and scrutinize them as our own objectivations – must have evolved from our primary ability to manufacture – and then behold, probe, and modify – meaningful things. Just like artifacts, words and gestures are external objects brought into existence by human action (Donald, 1991).

Our capacity for manufacture is grounded in specific abilities of hand-eye coordination and certain kinds of precision grip, that is, the ability to closely inspect, rotate, and modify objects while firmly holding on to them (Napier, 1980). The grounding of manufacture and reflexivity in hand-eye coordination, central already to the work of Gehlen (1988) and Plessner (1965) and, much later, Bruner's theory of language acquisition and grammatical relations (Bruner, 1969), is central to any kind of craft (McCullough, 1996; Sennett, 2008; Streeck, 2009). The conception of interaction as multimodal, as it is presented in this book, is consistent with this philosophical-anthropological notion of the excentric positionality of the human species: We have survived by means of our multiple and hetereogeneous objectivations, which include language and artifacts such as tools, skilled practices, rituals, and institutions. These objectivations can only be understood and explained in relation to one another. Such a view contrasts sharply with approaches that seek to abstract language from this nexus and attribute to an innate faculty or claim the centrality of texts to human social life and reproduction. Phenomenological philosophers have given us a notion of the body as a vehicle for being in the world (Merleau-Ponty, 1962) and a primarily haptic – rather than visual – epistemology. Manipulations are our primary understandings of the world (Heidegger, 1962). "Understanding is not in our minds but in our skillful ways of comporting ourselves" (Dreyfus, 1991: 75). It is the body thus conceived – in its concrete, unique, pre-verbal, skilled, and practical coupling with a world – that occupies center stage in the studies of embodied interaction that are collected here.

In another theoretical context, the French social anthropologist Marcel Mauss, nephew and co-worker of Émile Durkheim, proposed the study of *techniques corporelles* (1973), of movement and action skills that people acquire by living in some social milieu. Bourdieu elaborated this focus on the body as practice in the concept *habitus* (Bourdieu, 1977), which designates the

socially contexted bodily dispositions, sensibilities, and skills that permeate our sensory cognition and action skills. Previously, Bateson and Mead had worked from a similar concept when they described the Balinese by focusing on "the way in which they, as living persons, moving, standing, eating, sleeping, dancing, and going into trance, embody that abstraction which (after we have abstracted it) we technically call culture" (Bateson and Mead, 1942: xii).

Anthropologists have produced many textual and visual accounts of embodied culture. As examples for many others, Keller and Keller (1996) have analyzed the sensory cognition of blacksmiths, and Harper (1987) the working knowledge of a car mechanic (see also Csordas, 1994; Ingold, 2000; Jackson, 1989; Strathern, 1996). French anthropologists have developed film-based methods for the praxeological study of cultural transmission (Comolli, 2003; de France, 1983), as exemplified by the *gestes de savoir* (Comolli, 1991) of housewives and violinists.

The phenomenological conception of the body as situated in and "intervolved" (Dreyfus, 1991) with a material-practical world is in many ways a forerunner (sometimes acknowledged, sometimes not) of the currently popular cognitive science program known as *embodied cognition*. Its agenda is neatly summed up in the subtitle of A. Clark's book: "putting brain, body and world together again" (Clark, 1997), whose title, *Being There*, is a direct translation of Heidegger's term *Dasein* (Heidegger, 1962). Cognitive scientists who conceive cognition as embodied widely agree on the following points:

(a) the computational view of the mind, according to which the mind-brain operates by manipulating abstract (amodal) symbols, is rejected;
(b) experience (memory) is modally stored, in the form of "perceptual symbol systems" (Barsalou, 1999); the sensory, perceptual dimensions of experience are retained in the formation of concepts;
(c) the brain is multimodal: it allows us to recode experience, to structure it in terms of schemata from other domains (Deacon, 1997);
(d) the original function of any brain is to control motion – only mobile organisms have brains; other functions of the brain must have evolved from this primary ability (Llinàs, 2001);
(e) cognition and emotion are inseparable; emotion is a form of (embodied and social) cognition (Damasio, 1994, 1999);
(f) perception and motor control are not separate in the brain; perceiving another human being's action means producing an internal (i.e., inhibited, simulated) version of that action (this is known as common coding of motor-control and perception).

Many cognitive scientists interested in embodied cognition, while granting that the body must be conceived as a body in action, even in joint action (Knoblich and Sebanz, 2006), are reluctant to situate it fully within the material, external, human-made world. Psychologists, keen to maintain the separate integrity of the psychological system(s), have a hard time accepting the idea of distributed cognitive systems as agents of cognitive activity, as proposed, for example, by Hutchins (1995) and contributors to this volume. Thus, Wilson (2002: 126) grants that "cognition is situated, … takes place in the context of a real-world environment, and … must be understood in terms of how it functions under the pressures of real-time interaction with the environment, [and] we off-load cognitive work onto the environment." She rejects, however, the notion that, because "the environment is part of the cognitive system, … the mind alone is not a meaningful unit of analysis" (loc.cit.). For the researchers represented in this volume, an understanding of cognition as socially shared and distributed across mind, communication media, and other artifacts is essential.

EMBODIED INTERACTION

This volume contributes to a stream of research that has gradually emerged and matured during the past four decades. In this section, we seek to account for the convergence of several strands of research and delineate the place of our own attempts in this development.

In the 1970s, scholars from various disciplines began to lament the artificial separation and isolation of so-called "verbal" and "nonverbal" behavior. For instance, Kendon (1972) observed that "it makes no sense to speak of 'verbal communication' and 'nonverbal communication'" (443); he argued that theories of language derived from a study of only speech should be thought of as special language theories, whereas general language theories would show how vocal and visible behaviors function together (Kendon, 1977). In a similar spirit, Margaret Mead (1975) rejected nonverbal research as a "discipline-centric" neglect of vocal phenomena: She argued against Ekman's (1973) theory that facial expressions have universal meanings, suggesting that members of cultures derive meaning from facial expressions by relating them to the context in which they occur, which includes vocal behavior. Such laments in the 1970s were coincident with the mass marketing of a new technology called "videotape," which set the stage for more programmatic explorations of face-to-face interaction.

In the 1980s, a handful of seminal studies clearly and empirically established how talk and embodied behavior co-occur as interdependent phenomena, not separable modes of communication and action. Researchers in the tradition of conversation analysis explored the relationship between talk and eye gaze. Goodwin (1979) examined a videotaped dinner conversation and focused on a single spoken sentence that was shaped and reformed in the process of its utterance as the speaker shifted his gaze among recipients who had different knowledge states – which called into question the linguistic notion of

a sentence as something whose organization was lodged within the mental life of a single individual, the speaker. In an other work, C. Goodwin (1980) analyzed a collection of videotaped instances to show subtle forms of coordination between utterance-initial restarts and shifts in participants' eye gaze (hence attention) toward the speaker. Atkinson (1984) dissected recordings of political speeches to show how politicians elicit applause from audiences, not merely through vocal devices such as "contrastive pairs" and "three-part lists," but also through their rhythmic coordination of talk and gaze shifts toward their audience. Heath (1986) studied the organization of talk and gaze during medical consultations, whereby patients may direct their doctor's attention toward parts of their bodies that need medical attention. Although some prior research had explored the relationship between talk and gaze (e.g., Kendon, 1967), these studies in the 1980s were seminal because they emphasized the sequential unfolding of human activity within specific situations: Rather than code the phenomena and count the frequencies of occurrences, these scholars transcribed and carefully analyzed particular strips of situated interaction.

Researchers who conducted sequence-analytic studies of videotaped interaction also turned their attention to hand gesture (e.g., Kendon, 1983, 1988; Goodwin and Goodwin, 1986), which has become an especially fruitful branch of naturalistic inquiry. When people gesture, they usually talk at the same time, coordinating their behaviors to be understood as an ensemble (e.g., Goodwin, 1986; Goodwin and Goodwin, 1986). Schegloff (1984) considered the connection between gestures and their "lexical affiliates" as evidence for the "projection space" during which an element of talk is in play, without having been uttered, allowing co-interactants anticipatory adaptations. Streeck (1993) showed how gestures may be "exposed" (i.e., made an object of attention during moments of interaction) through their coordination with indexical forms of speech (e.g., words such as "this") and eye gaze (which may perform "pointing" functions). Hands move within three-dimensional spaces that include objects and artifacts, and gestures may be largely recognized and understood through their relationship to the material world within reach (e.g., Goodwin, 1997, 2000b; Heath and Hindmarsh, 2000; LeBaron and Streeck, 2000). Furthermore, gesture may be embedded within extended processes or activities, such that any particular gesture is understood through its relationship to the whole activity (e.g., Koschmann, LeBaron, Goodwin and Feltovich, 2006). During this time, David McNeil (1992) and colleagues at the University of Chicago, including Susan Duncan (2002) and Susan Goldin-Meadow (2003), developed important frameworks for the analysis of gesture that were consistent with their orientation in psychology.

Meanwhile, interaction-focused researchers of gesture demonstrated that communicative acts are always "environmentally coupled" (Goodwin, 2007a), but can also structure the perception of the environment. Working as an anthropologist in Chiapas, Mexico, Haviland (2000) documented the directional precision of a farmer's pointing gestures, suggesting that his gestures made his "mental map" interactively available, even interactively constructed. Gestures have been explicated as a locus of shared knowledge and emergent understanding (e.g., Enfield, 2008; Koschmann and LeBaron, 2002; LeBaron and Koschmann, 2003), organizing social interaction on the one hand and shaping individual cognition on the other (LeBaron and Streeck, 2000). Such studies of gesture have been more anthropological than psychological (e.g., Sidnell, 2005), emphasizing the public nature of "individual" cognition (Streeck, 2002), treating the human mind as something that extends beyond the skin to include social and material worlds. This research offers an alternative to views that are more psychologically oriented, such as McNeill's, who suggested that "gestures are the person's memories and thoughts rendered visible … [belonging] not to the outside world, but to the inside one of memory, thought, and mental images" (1985: 12).

All the chapters in Schmitt (2007) focus on the delicate *coordination* of modalities, both intrapersonal and interpersonal, that bring about ordered and intelligible sequences of interaction. Deppermann and Schmitt (2007), who have done much to establish the study of multimodal interaction as a recognized field within European linguistics, conceive the study of multimodality as a study of *coordination*, on the one hand of different strands of bodily action within the single participant (self-organization), and on the other the coordination between co-interactants (interactional organization). The structuring of actions in one modality – for example, gaze – is clearly constrained by, or interacts with, those in another modality – for example, postural configurations or "F-formations" (Tiittula, 2007; cf. Kendon, 1976). As Mondada (2007b) has shown, self-organization is of particular importance where people participate in multiple activities at the same time ("multi-activities" such as conducting a conversation while driving a car or performing surgery while explaining the process to a remote audience). Lindström and Mondada (2009), building on work by Goodwin and Goodwin (1987, 1992), exemplify the multimodal nature of human interaction in a single language game, assessments of which are often performed through careful orchestration of talk, gaze, and facial displays (Ruusuvuori and Peräkylä, 2009), among other modalities. Krafft and Dausendschön-Gay (2007) introduce a useful distinction between "direct coordination" (coordination through the spatial organization of the bodies of the interactants) and coordination via objects, which occurs when participants use gestural and verbal acts of deixis to achieve a shared orientation to the setting of the interaction. Deppermann and Schmitt (2007) point out that research on multimodality complements the analysis of sequencing that is at the core of conversation analysis by an additional focus on *simultaneity*, that is, close attention to which behaviors are produced

at the same time and how such synchronous productions are possible. Simultaneity is a constitutive feature of any interaction, which implies the importance of spatial relations: how participants are positioned in relation to one another or where they look at any point in time is as important as the temporal relations between their talk and movements. This, in turn, points up the relevance of the *materiality* of communication modalities, for example the affordance of gesture to be perceived and processed simultaneously with speech as well as to attract and direct the addressee's visual attention (Heath, 1986; Streeck and Hartge, 1992).

That even speech alone comprises several modalities that must be explicated both in relation to one another and to their shaping, and functions in real-time interaction is the theme of a new paradigm within linguistics known as *interactional linguistics* (Selting and Couper-Kuhlen, 2001). One focus of this conversation-analysis-based field of studies have been the roles of rhythm and prosody in conversational interaction (Auer, Couper-Kuhlen, and Müller, 1999; Couper-Kuhlen and Selting, 1996a, 1996b; Uhmann 1992, 1996); another the emergence and operation of syntactic constructions in interactional contexts (Auer, 2009; Deppermann, Fiehler and Spranz-Fogasy, 2006; Günthner and Imo, 2006; Streeck, 1996b. See also Ford, Fox, and Thompson, 1998; Ochs, Schegloff, and Thompson, 1996.) Although we cannot cover this field here, it is important to note that it is guided by the same view of interaction as multimodal and of structural forms (constructions) as in part interactionally motivated.

Several of the contributors to this volume are linguistic anthropologists. Linguistic anthropology has given us several distinct analytic traditions; it is centrally concerned with the symbolic structuring of behavior. We have learned from linguistic anthropologists to attend to the social-symbolic significance of minimal differences in interactively produced forms (e.g., phonetic choices or prosodic contours; see Gumperz, 1982a, 1982b), but also to investigate such dimensions of embodiment in the context of culturally defined, regulated, and recognized events (Agar, 1975). In *Linguistic Anthropology*, Duranti (1997) presents the study of embodied interaction as one of the standard methodologies in contemporary linguistic anthropology. His own work is a good example of the inevitably "multimodal" nature of anthropological research into linguistic practice: studying the Samoan honorific system (which is expressed in verb morphology), Duranti (1992, 1994) discovered that the system is inextricably bound up with ways in which Samoans position themselves in relation to the place they are in and to one another.

MULTIMODALITY: EMBODIED INTERACTION IN THE MATERIAL WORLD

In a recent review of Tomasello's (2008) *Origins of Human Communication*, Kendon has emphasized, without employing the term, the inherently multimodal nature of human communication:

> [T]he transition into referential or language-like expressions involved hands and body, face and voice and mouth, all together, as an integrated ensemble. What so many writers on this topic – "gesture firsters" and "speech firsters" both – pay little attention to is the fact that modern humans, when they speak together in face-to-face situations … always mobilise face and hands and voice together in complex orchestrations… Every single utterance using speech employs, in a completely integrated fashion, patterns of voicing and intonation, pausings and rhythmicities, which are manifested not only audibly, but kinesically as well, and always, as a part of this, there are movements of the eyes, the eyelids, the eyebrows, the brows, as well as the mouth, … patterns of action by the head, and … from time to time variously conspicuous hand and forearm actions or "gestures" (Kendon, 2009: 363).

In the same vein, Stivers and Sidnell write that "face-to-face interaction is, by definition, multimodal interaction in which participants encounter a steady stream of meaningful facial expressions, gestures, body postures, head movements, words, grammatical constructions, and prosodic contours" (Stivers and Sidnell, 2005: 1).

Following Enfield (2005), they distinguish between "vocal/aural" and "visuospatial modalities." In contrast, we regard the abstraction of the interacting body from the material world as an abstraction with problematic consequences and – although we acknowledge the usefulness of terminological distinctions between different kinds or groups of modalities of communication – nevertheless insist that embodied interaction *in the material world*, which includes material objects and environments in the process of meaning making and action formation, is primary. Many of the contributions to this book therefore go beyond the study of the ways in which several bodily "channels" are coordinated in social interaction to show how environmental sources of meaning are drawn into the production of inter-subjective understanding and how interaction, in turn, structures its own semiotic and material environment.

Long before the term "multimodal(ity)" entered the field of interaction studies, it was established as a technical term in two entirely different fields, logistics and therapy. In the logistics industry, "multimodal" refers to the coordinated transportation of goods by air, land, and water; in medicine and psychotherapy, to the combination of multiple therapeutic practices, for example music therapy and talking cure or surgery and radiation. More recently, the term has taken center stage in computer science, where it describes human-computer interfaces that allow for multiple simultaneous input (e.g., by voice and gesture) and heterogeneous representations. Not very different from this usage is the term "multimodal corpora" applied to linguistic research, that is, the production of data representations that combine auditory and visual with textual representations (Kipp, Martin, Paggio, & Heylen, 2002). The term "multimodal communication"

is also used by various groups of researchers who seek to expand the semiotic analysis of texts so as to accommodate text-image combinations, but also other artifacts including films, buildings, and objects of daily use (Kress and van Leeuwen, 2001; Norris, 2004). Some of these researchers draw on Halliday's systemic-functional perspective (O'Halloran, 2004), others develop their own versions of discourse analysis (Levine and Scollon, 2004), but none approach human interaction in the material world with the rigorous microanalytic focus on the formation of action sequences that is characteristic of the contributions to this book.

When exactly the term "multimodal(ity)" entered the microanalysis of interaction is not entirely clear – certainly long before the appearance of Stivers and Sidnell (2005). What is equally certain is that the reconceptualization of embodied interaction as multimodal and the subsequent recognition of the importance of material contexts and artifacts drew a great deal of inspiration from – and partly overlapped with – two new, interdisciplinary research programs: *studies of work* (or *workplace studies*) and *science and technology studies* (see, among many others, Lynch and Woolgar, 1988; see also Heath, Luff, and Knoblauch, 2004). Inspired by these studies, sociologists became interested in the contingent, local production of practical, normatively accountable actions in the context of labor rather than conversational interaction. One of the hallmarks of this research program was recognition of the paramount importance of physical objects – things – in the conduct of work-related activities. Explaining the new research program, Garfinkel wrote that

> it was evident from the availability of empirical specifics that there exists a locally produced order of work's things; that they make up as massive domain of organizational phenomena; that classical studies of work, without remedy or alternative, depend upon the existence of these phenomena, make use of the domain, and ignore it (Garfinkel, 1986: vi).

In an early, seminal study, Suchman (1987) demonstrated that normative rules of use are unable to guide (or explain) the operation of technological objects (in her case: copy machines), but that usage of such objects – and the normative accountability of such usage – represents ongoing, situated, contingent, and interpretive accomplishments. Understanding technology-supported action, as well as designing "user-friendly" technologies, thus requires the precise, moment-by-moment study of people's physical actions and the practical reasoning displayed by them, rather than reliance on decontextualized models of cognitive "plans" in the vein of Miller, Gallanter, and Pribram (1960). In another study, Suchman (1996) investigated how competent actors construct shared workspaces and arrange resources and tools to assemble readily interpretable surfaces that facilitate collaborative action. Suchman's work contributed to a growing trend among microanalysts of interaction to investigate talk and embodied communication not apart from, but within complex material environments that they simultaneously make intelligible and coherent (Button, 1993; Engeström and Middleton, 1996).

A wealth of new research into hitherto unexplored domains of human action and interaction thus emerged. Heath and Luff (2000), in their wide-ranging research in contexts such as control rooms of the London Underground, computer-assisted architectural design, video-conferencing, and software development, focused attention on the difficulties of adapting new technologies to established orders of mundane reasoning and interaction. Rather than simulating face-to-face interaction, communication technologies such as video-conferencing demand that participants reconfigure participation frameworks and practices of turn-taking and speaker-listener coordination. With this widening of scope, compared to the initial focus on conversation over the telephone, ethnomethodological and interactionist researchers began to seriously implement Wittgenstein's vision that the study of a language must encompass the entirety of the community's language games and explicate them as forms of life. As McHoul (2008: 825) writes, "what we are ultimately interested in is taking pretty much any bit of ordinary everyday interaction as a means of understanding forms of life (Lebensformen) as such and not simply for its own sake as a technical object.... Conversation may be our favourite 'game', but it is not the only one in town."

A type of workplace that attracted particular attention were science laboratories, in which the study of work took on the form of studying the practices, instruments, and representations by which scientific findings are assembled and ratified as facts by the relevant community of scientific practice (Knorr-Cetina and Mulkay, 1983; Latour and Woolgar, 1986; Lynch, Potter, and Garfinkel, 1983; Lynch and Woolgar, 1988). Whereas scientific work – especially laboratory work – is inherently multimodal (it is the normatively guided coordination of practices of perceiving, experimenting, measuring, and representing that constitutes legitimate scientific practice), particular attention was paid to the production and interpretation of visual representations. In Latour's (2005) influential conception, dubbed "actor-network theory," agency is seen as being distributed across human actors and material things. Even though this view may not be universally shared by researchers of science, technology, and interaction, Latour's work has undoubtedly contributed to a scientific climate in which it is much easier to find acceptance for the notion that interaction, cognition, and work are inherently multimodal affairs that cannot be studied on the basis of what goes on in a single "channel" alone or by relying on textual representations abstracted from the rich contexts of the phenomena represented by them. The domain of things had rarely entered the picture in studies of conversational interaction, and never in studies of telephone conversation (but see Mondada, 2008b; Whalen, 1995). What is sometimes referred to as the "logocentric" bias in conversation analysis (e.g., Erickson, 2010) certainly has