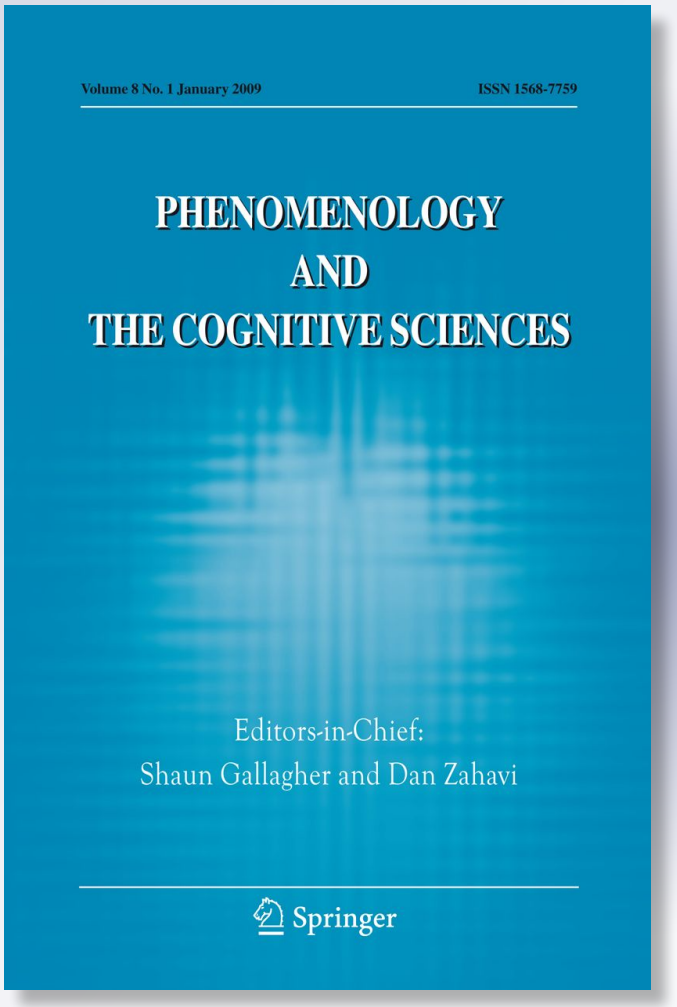
Gestural sense-making: hand gestures as intersubjective linguistic enactments

Elena Cuffari

Phenomenology and the Cognitive Sciences

ISSN 1568-7759

Phenom Cogn Sci
DOI 10.1007/s11097-011-9244-9



Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media B.V.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Author's personal copy

Phenom Cogn Sci
DOI 10.1007/s11097-011-9244-9

Gestural sense-making: hand gestures as intersubjective linguistic enactments

Elena Cuffari

© Springer Science+Business Media B.V. 2011

Abstract The ubiquitous human practice of spontaneously gesturing while speaking demonstrates the embodiment, embeddedness, and sociality of cognition. The present essay takes gestural practice to be a paradigmatic example of a more general claim: human cognition is social insofar as our embedded, intelligent, and interacting bodies select and construct meaning in a way that is intersubjectively constrained and defeasible. Spontaneous co-speech gesture is markedly interesting because it at once confirms *embodied* aspects of linguistic meaning-making that formalist and linguistic turn-type philosophical approaches fail to appreciate, and it also forefronts *intersubjectivity* as an inherent and inherently normative dimension of communicative action. Co-speech hand gestures, as linguistically meaningful speech acts, demonstrate both *sedimentation* and *spontaneity* (in the sense of Maurice Merleau-Ponty's dialectic of linguistic expression (2002)), or features of *convention* and *nonconvention* in a Gricean sense (1989). Yet neither pragmatic nor classic phenomenological approaches to communication can accommodate the practice of co-speech hand gesturing without some rehabilitation and reorientation. Pragmatic criteria of intersubjectivity, normativity, and rationality need to confront the non-propositional and nonverbal meaning-making of embodied encounters. Phenomenological treatments of expression and intersubjectivity must consider the normative nature of high-order social practices like language use. Reciprocally critical exchanges between these traditions and gesture studies yield an improved philosophy that treats language as a multi-modal medium for collaborative meaning achievement. The proper paradigm for these discussions is found in enactive approaches to social cognition. Co-speech hand gestures are first and foremost emergent elements of social interaction, not the external whirring of an isolated internal consciousness. In contrast to current literature that frequently presents gestures as uncontrollable bodily upsurge or infallible imagistic phenomenon that drives and dances with verbal or "linguistic" convention (McNeill 1992, 2005), I

E. Cuffari (✉)
Department of Philosophy, University of Oregon, 1415 Kincaid St. PLC 17, Eugene, OR 97403, USA
e-mail: ecuffari@uoregon.edu

Published online: 30 November 2011

Author's personal copy

suggest that we study gestures as dynamic, embodied, and shared tools for collaborative sense-making.

Keywords Gesture · Language · Intersubjectivity · Normativity · Convention · Enaction

Introduction: social cognition as embodied and intersubjectively normative

The ubiquitous human practice of spontaneously gesturing while speaking demonstrates the embodiment, embeddedness, and sociality of cognition.[1] The present essay takes gestural practice to be a paradigmatic and illuminating example of a more general claim: human cognition is social insofar as our embedded, intelligent, and interacting bodies select and construct meaning in a way that is intersubjectively constrained and defeasible. Spontaneous co-speech gesturing is markedly interesting because it at once confirms *embodied* aspects of linguistic meaning-making that both formalist and linguistic turn-type philosophical approaches fail to appreciate or accommodate, and it also forefronts *intersubjectivity* as an inherent and inherently normative dimension of communicative action.

Linguistic turn thinking has made philosophers concerned with meaning appropriately savvy to the dangers and limits of foundationalism or reductive naturalism (Rorty 1979, 1989; Brandom 1994); yet it raises these flags at the double cost of eschewing any bodily activity of meaning-making and crowning propositional verbal language use as the paradigm of rational achievement practices (as pointed out by Lakoff and Johnson (1999, 440); Johnson (2007, 198–199, 205–206); and Koopman (2011)). Keeping these fears as well as these costs in mind, I argue that acknowledging the pivotal role of the body in our conversational practices does not at all have to mean ceding our claims to rationality (reasoning) and normativity (correctness or incorrectness). Gestures are not an "anything goes" phenomena; they do not signify in a purely natural way; they are not irrational upsurges of bodily excess. Just as we have pragmatic criteria for the giving and asking of reasons and semantic and syntactic criteria for the production of grammatical sentences, our gestural behavior finds its intelligibility by conforming to certain collectively held standards on multiple levels.

In what follows, I seek to make a contribution to recent and growing cross-disciplinary efforts to bring social cognition properly into focus.[2] Language use is both a social and a cognitive act; reflecting on actual linguistic practices ought to assist in illuminating certain questions and target areas for new treatments of social cognition. Furthermore, situating the study of linguistic activity within a research context that explicitly prioritizes what is social about meaning-making may cast fresh light on overlooked phenomena. I take the ubiquitous yet under-theorized phenomena of hand gesturing while speaking as a paradigm case of people making

[1] The author would like to thank Dr. Emma R. Jones, George Fourlas, and an anonymous reviewer for their great help and suggestions in revising this article.

[2] For example, Strawson (2004), Gallese (2005), Stawarska (2006), De Jaegher and Di Paolo (2007), Fuchs and de Jaegher (2009), and Steiner and Stewart (2009) extend aspects of the 4E paradigm (Menary 2010, particularly enactive treatments such as Hutto (2005), Thompson (2005), and Simpson (2010)) to encompass sociality in explaining embodied cognition.

Author's personal copy

sense together, hence a prime avenue for inquiry into social cognition. Along this route, I find that improvements can be made to canonical philosophical treatments of language and meaning, which consistently ignore gestural phenomena. Language use operates by submission to various normative authorities. Twentieth century philosophy and linguistic science presumes these authorities to be grammaticality and objectivity, i.e., competence-based rule-following and appropriate mappings to the way the world is. However, pairing insights from phenomenology and pragmatics yields an alternative account of normativity, in which what matters to meaning achievement in language is the difference made to a lived situation and practical, purposive projects of understanding and cooperation. In other words, real communication practices such as speaking and co-speech gesturing are normatively constrained first and foremost at the level of pragmatic action that is thoroughly situated in a (or several) shared horizon(s) of meaning possibility.

I argue that in order for theorists to properly get the *social* as well as *embodied*, and *nonverbal* as well as *verbal* aspects of linguistic phenomena into view—and thus be equipped to appreciate co-speech gesture as linguistic enactments of social cognition—we need to draw at once on a phenomenological understanding of language use as interpretative and intentional world-relation; a pragmatic focus on communication as intersubjective action; and an enactive (non-representational) approach to cognition. Throughout the essay, I argue that each of these fields holds resources for understanding social cognition, yet at the same time does not tell a complete story. In particular, while gesture scholars borrow selectively from the domains of pragmatics and phenomenology, a full, reciprocal cross-disciplinary encounter that might yield a gesture-inclusive philosophy of language is overdue. Philosophical treatment can play a role in negotiating contestations between emerging paradigms within the field of gesture studies, as I point out at the end of the article.

After offering some introductory remarks on the meaningfulness of co-speech hand gestures and the significance this meaningfulness has for a recognition of nonverbal linguistic activity ("Convention, nonconvention, and normativity in co-speech gesturing"), I facilitate two reciprocally critical cross-disciplinary encounters. In "Gricean and neo-Gricean pragmatics", this encounter is between Gricean and neo-Gricean pragmatics on one hand, and on the other, gesture-inclusive descriptions of communicative action in which participants collaboratively negotiate conventional and nonconventional usages in multiple modalities. The goal here is to recover from Grice's reflections on communicative intention an account of local and constrained conversational reasoning that is not necessarily propositional, verbal, nor private (in a coding–decoding, mental processing sense). "Towards a phenomenology of gestural enacting" continues the demonstration of the intersubjective basis of linguistic normativity in an encounter between a psychological theory of speech–gesture interaction and certain phenomenological treatments of embodied experiences of communicative achievement. I caution gesture theorists to avoid adopting an overly rigid dialectical view of gesture production, which misunderstands the constituent ambiguity of the phenomenological account. At the same time, current phenomenologically-rooted dialogues regarding intersubjectivity are invited to attend to the vicissitudes of the linguistic level of interaction in complement to studies of "non-linguistic"

Author's personal copy

phenomena such as perception and the communicative behavior of pre-linguistic infants.

In addition to making a case for rehabilitating both pragmatics and phenomenology via confrontation with co-speech gesture practice, which ought to be included in any analysis of meaning achievement in language, in "Philosophical preference for a paradigm of intersubjective gestural sense-making" I recapitulate the consequences of these encounters for the emerging interdisciplinary field of gesture studies. I submit that co-speech hand gestures are first and foremost emergent elements of social interaction, not the external whirring of an isolated internal consciousness. Rather than treat them as wayward missives from the subconscious that defy convention and demand psychological decoding, I suggest that we study gestures as dynamic, embodied, and shared tools for collaborative sense-making.[3]

Convention, nonconvention, and normativity in co-speech gesturing

In recent decades, scholars across disciplines have shown that hand gestures significantly contribute to the meaning of the utterances in which they occur. I take this as a central premise in the following arguments, along with the strong claim that no proper analysis of a speech act that contains co-speech gestures is possible without taking these accompanying hand motions into account. Hand gestures that spontaneously accompany speech offer iconic depictions (McNeill 1992; Streeck 2008), specify verb manner (McNeill and Duncan 2000), enact metaphorical reasoning (Müller 2007; Cienki and Müller 2008), and regulate speaker–hearer interaction (Sweetser and Sizemore 2008), to name just some of the myriad identified functions. Consider two general examples of the rich meaningfulness of co-speech gesturing.

Cornelia Müller offers the following instance from her research: in a therapist's session, a young woman charts the course of a relationship gone wrong by riding the "ups and downs" with the palm of her right hand (Müller 2007, 114). Müller explains, "What is in the focus of attention of the speaker is foregrounded in a verbal–gestural utterance" (114). This example not only shows gestures accompanying speech; gestures aid in carrying the metaphor ("up and down") the speaker is using to understand her experiences as she converses with her therapist.

[3] Throughout this work I argue yet more strongly that co-speech hand gestures are linguistic. This is done in deliberate effort to broaden our understanding of what languaging activities actually involve, as one anonymous reviewer helpfully summarizes my point. The goal here is not to lose important distinctions by broadening a term; rather, the goal is to correct a shocking oversight in the received way scholars carve up phenomena. My view is perhaps usefully located in relation to contemporary pragmatists such as Mark Johnson and Colin Koopman, who point out that there are meaningful human practices that are normatively structured—subject to criticism and defense, success and failure—which are non-linguistic (Johnson 2007, 34, 207–209; Koopman 2011, 75). Agreeing with Johnson and Koopman, my goal is to reserve a place for language use as a sub-region of human practices of meaning achievement and locate co-speech gesture as belonging to this place. Within the hierarchy of meaning-making practices and communicative acts that humans engage in, co-speech gestures occur *with* speech and achieve their meaning in close interaction with speech. I can communicate by throwing things, and this is probably not well-classified as a linguistic act. But surely pointing to a book is closer to the act of verbally bringing someone's attention to that book than it is to throwing the book at them.

Author's personal copy

Another ready example comes from Adam Kendon's work. A Neopolitan speaker discusses the geographical boundaries of the city of Naples. She concludes her account with the words "'*il proprio' o core e Napulë è hchesto ccà'* ('the true heart of Naples is this here'). As she says '*il proprio' o corë'* she extends her arm forwards and lowers her hand vertically, drawing the fingers into the *grappolo* as she does so" (Kendon 2004, 237). The *grappolo* is the recognizable "purse hand" or "finger bunch" gesture form in which the fingers are vertically extended but flexed at the knuckles and drawn together so that fingertips are all in contact (Kendon 2004, 229). For the community under discussion, Kendon identifies this hand shape as "involved in gestures which appear to mark the topic of a speaker's discourse but also in gestures which are employed when the speaker is asking certain kinds of questions or demanding an explanation or justification for something" (2004, 228). In the "true heart of Naples" example, the *grappolo* is positioned outward, in correspondence to a particular central region of an area the speaker has constructed via a virtual gestural map, thus both deictically and pragmatically highlighting this intention of this speech act as locating the city center and placing beyond question the characterization of this region as the place best identified as "Naples."

Claiming that linguistic normativity is found in gestural practices might be seen as a radical move insofar as nonverbal behavior has for decades been deemed categorically distinct from verbal (i.e., "linguistic") behavior precisely for its willful shrugging off of convention, for its non-linear, holistic, non-segmental nature, and interpretive subjectiveness, as Kendon points out in his history of gesture studies (Kendon 2004). In counterpoint to this received view, a growing literature demonstrates both semantic and pragmatic systematicity and conventionality of co-speech gestures.[4] Geneviève Calbris offers a book-length study of the semiotics of French gestures (1990). Sotaro Kita reviews four factors governing cross-cultural variation of co-speech gestures, finding culture-specific conventions for (1) form-meaning associations, (2) spatial cognition, (3) verbal means of expressing spatial information, and (4) pragmatics, or "the principles under which gesture is used in communication" (2009). Meanwhile, gesture researcher and historian Adam Kendon has identified four culturally specific gestural forms (of which the *grappolo* is one), which are identifiable functions of particular gestures that are widely used in a particular communication community (2004, 226ff). Kendon's method is to generalize over a range of related pragmatic usages to get at a more general meaning that can then be associated with the gestural form and movement pattern (a semantic theme). He writes that "this theme, being introduced as it is in different ways in different contexts, through the way it interacts with the (usually verbal) meaning of the spoken component of the utterance, contributes to the creation of a highly specific local meaning" (2004, 226).

Each of these scholars' projects makes evident that, while complex, it is quite possible to recognize and explain multiple processes of gestural sense-making. Furthermore, such work points the way towards improved, nuanced, and non-foundationalist understandings of linguistic operations like representation and

[4] Thanks to anonymous reviewer for bringing my attention to several of these references.

Author's personal copy

reference generally.[5] As I argue below, gestures are meaningful by virtue of some of the same conditions and constraints that render verbal language use meaningful. At times gestures serve referential functions, generating and specifying propositional content in interaction with speech (Kendon 2004, 158ff), but only because they are enacted *for* an audience prepared to take them as doing so. The sense we gather from speech and gesture always has to do with what the speaker is *doing* in and through these modalities and what *difference* this doing makes (or fails to make). Better or worse conceptualizations and communicative actions via gesturing hands are possible. This is evidenced by "word-search" phenomena, when participants help supply a speaker with a missing word as he gestures to retrieve it (Goodwin & Goodwin 1986; Streeck 1994); repairs made to both gesture and speech mis-fires (Kendon 2004, 133–134; Goodwin 2006); and studies of the effect of audience attention and inattention on the structure, placement, shape, amplitude, and frequency of gestures (Streeck 1994). While a fuller theoretical articulation of this kind of normativity is desired, empirical observation indeed points to hand gesturing as an intersubjectively monitored and shaped practice of meaning-making.

Gricean and neo-Gricean pragmatics

Ordinary language philosopher H.P. Grice demonstrates that the successful use of language to communicate requires a certain knowledge of and cooperation with norms and conventions. Like his predecessor J.L. Austin (1961), Grice takes nonverbal human behavior to be social and meaningful on the same grounds and by the same operations of verbal activity.[6] Indeed, without the nonverbal (at least not-necessarily verbal) background conditions of cooperation, intent, environment, custom, knowledge, action, and action-interpretation (what today gets called "mind-reading"), utterances

[5] See, in particular, Jürgen Streeck (2008). Complicating the traditional yet unsatisfactory understanding of iconicity as resting on perceived similarity, Jürgen Streeck identifies and analyzes a variety of practices by which gesturers achieve an interpretation in depiction (2008). Streeck names 12 methods by which hand gestures construe something *as* something for their receivers (2008, 292–295). These gestures make sense to participants immediately as the hand motions transparently give way to the selected schemata or features they enact. For Streeck, the "pictorial language" by which gestures construe consists "of schematized acts of making, handling, drawing, and so on: whatever is depicted—things, inanimate processes, actions—is depicted and at the same time *analyzed* in terms of *manual* acts. Knowledge of these acts… is not in the first place knowledge of the gesture methods (or gestures' meanings), but of ways of acting in the material world" (2008, 298–299). Common practices and familiar action sequences in a shared world, rather than formal resemblance or mirroring, thus enable our understanding of depictive gestures.

[6] Grice (1989, 28) gives the examples of fixing a car or baking as activities in which expectations for others' rational and helpful contributions holds. Austin saw all speaking as an *act*. Austin writes, "There are a great many devices that can be used for making clear, even at the primitive level, what act it is we are performing when we say something—the tone of voice, cadence, gesture—and above all we can rely upon the nature of the circumstances, the context in which the utterance is issued" (1961, 231). For Austin, the conventions guiding how we perform utterances or "do things with words" are importantly relative to a society of language users. Austin explains "The social habits of the society may considerably affect the question of which performative verbs are evolved and which, sometimes for rather irrelevant reasons, are not" (1961, 232). Clearly Austin's sense of "convention" is never simply or solely located in grammar or literal sentence meaning.

Author's personal copy

would not be intelligible. Yet Grice (like Austin) was followed by thinkers eager to formalize his discussions of "literal" phenomena, effectively ghettoizing the indispensible conditions of performance and pragmatics central to the original ordinary language vision.[7] I prefer to take from Grice the idea of communication as an achievement activity, something that can be done better or worse, something that is done for a reason, something that knows of success or failure. I contend that if we sustain an effort to see language use as a communicative practice, and meaning as an enactment or achievement of communication that tends to use language as a preferred medium, then we will lose the imperative to maintain a privileged place for verbal symbols and verbal utterances. Rather, it begins to make more sense to see "language" as including all those semiotic modalities that humans deploy together to achieve meaning in communicative action.[8]

Grice's nonconventional conversational implicatures

Grice's theory demonstrates that it is our exploitation of standards that make our meaningful exchanges both interesting and ultimately indicative of intersubjective rationality. *Conversational implicatures* presume the existence of (and participants' tacit adherence to) a principle that governs conversations as intentional communicative activities. Grice offers the Cooperative Principle, which states: "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (1989, 26). Grice then analyzes the Cooperative Principle into four more specific categories, each with attendant maxims (1989, 26–27): quantity, quality, relation, and manner, which generally have to do with what sort of information the speaker shares, how much, how true that information is, the manner it is conveyed, etc. The purpose of talk presupposed by these conversational maxims is "a maximally effective exchange of information," and talking is on Grice's view "a special case or variety of purposive, indeed rational, behavior," (1989, 28). These maxims are thus normatively enforced by a community of speakers; a speaker who utterly failed to meet these rational expectations would most likely be "subject to rebuke" of varying degrees (1989, 27).

The precise connection between the Cooperative Principle and its related maxims and the nonconventional phenomenon of conversational implicature is failure to fulfill the maxims (Grice 1989, 30). Failure to adhere to these codified conventions by conversational participants who are still presumably rational and still presumably adhering to the overarching Cooperative Principle triggers a need for nonconventional interpretation, that is, interpretation that is not fully determined or specified in advance, but which relies on context and inference in various ways. Out of the kinds

[7] For example, consider Sperber and Wilson's neo-Gricean Relevance Theory (1986), discussed below. Much like Searle's famous elaboration of Austin, performance and deep social context is left out in favor of an individualized, interiorized, competence-based representational account of cognition.

[8] Consider Adam Kendon's point: In contrast to explanation of speech–gesture co-expressiveness according to a dialectic processing plan, Kendon suggests "that the conjunction of the stroke with the informational center of the spoken phrase is something that the speaker *achieves*. In creating an utterance that uses both modes of expression, the speaker creates an *ensemble* in which gesture and speech are employed together as *partners* in a single rhetorical enterprise" (2004, 127).

Author's personal copy

of failure possible, the type that Grice calls "flouting" most generally leads to conversational implicature via *exploitation* of the maxims.

A classic example to demonstrate both this sort of implicature and the reasoning process Grice attributes to the hearer is the case in which a philosophy professor writes a pointedly brief letter of recommendation praising a student's grammar (Grice 1989, 33). Though he appears to flout maxims of quantity and relevance, the professor is not opting out, since he is writing the letter. He knows more about the student than his command of grammar, and he knows that more is requested in a recommendation letter. He must then be reluctant to say anything else, and this is understandable if what he would say would be negative. Thus, the professor thinks his student is no good at philosophy: he has demonstrated this by saying all the good things he can say about the student, and these do not include philosophical prowess. In the end, then, although at the level of what is *said* there is maxim violation, at the level of what is *implicated*, the maxims or minimally the Cooperative Principle is satisfied (Grice 1989, 33).[9] In other words, when a person communicates, they perform speech acts at multiple levels, not all of them verbal or vocal; their interlocutors are expected to track the emerging meaning across all of them (Grice 1989, 35). This is quite successful most of the time, and when it is not, metalinguistic clarification is always an option.

The kind of reasoning process that Grice's hearers go through to make sense of conversational implicatures is almost egregiously propositional (taken in the sense of conventionally configured representational structures of verbal signs). Grice gives many examples of this and offers a generic formula for the interpretation or "working out" of such an implicature:

> "He has said that *p*; there is no reason to suppose that he is not observing the maxims, or at least the Cooperative Principle; he could not be doing this unless he thought that *q*; he knows (and knows that I know that he knows) that I can see that the supposition that he thinks that *q* is required; he has done nothing to stop me thinking that *q*; he intends me to think, or is at least willing to allow me to think, that *q*; and so he has implicated that *q*." (Grice 1989, 31)

While he does not expect that the average hearer will deploy such technical vocabulary, Grice insists that "the presence of a conversational implicature must be capable of being worked out; for even if it can in fact be intuitively grasped, unless the intuition is replaceable by an argument, the implicature…will be a conventional implicature" (1989, 31). Why does failure to be worked out in rational, propositional form make an implicature conventional instead of nonconventional? Precisely because it is the additional *effort* required to make sense of an odd, unexpected, not conventionally predetermined utterance that characterizes it as nonconventional. In these cases, what is literally said or directly conventionally implicated is not enough to render the speech act rational and cooperative. In conventional implicatures, the rationality or saliency of the utterance is automatically apparent, no extra thinking or

[9] In his later "Retrospective Epilogue," Grice notes, "What I have been calling conversational implicature is just those assumptions which have to be attributed to a speaker to justify him in regarding a given sequence of lower-order speech-acts as being rationalized by their relation to a conventionally indexed higher-order speech act" (1989, 370).

Author's personal copy

attributing of various intentions and shared background knowledge required. With cases of nonconventional implicatures such as conversational implicatures, the hearer must reason across standard forms and conventions as well as concrete, particular situations and knowledge, and presumed rational communicative intentions. Yet as Grice's pattern clearly demonstrates, all of this effort takes place via forms that are "linguistic" in the traditional sense. These nonconventional implicatures that are expressed silently in what is not said, or in the excess of what is said, contribute significantly to the meaning-making processes of conversations. The conversational implicatures are not spoken, but they are linguistic. However, it appears from the foregoing that they are linguistic in virtue of being able to be put into propositional form in a hearer's inner dialogue with himself, in his reasoning processes.

Despite Grice's basic insistence that we see language as just another form of rational activity (1989, 341), and overlooking a few suggestive remarks about nonverbal and non-propositional aspects of speech such as prosody making nonconventional contributions to meaning (1989, 50–53), subsequent scholars, particularly in psychology, follow the implicit requirement of propositionality outlined above. This interpretation of Grice has generated several decades of pragmatic research that presupposes verbal propositions as the medium of inference-based conversational negotiations and posits module-based monological reasoning as the process by which those inferences are carried out. Leading neo-Griceans Dan Sperber and Deirdre Wilson defend "the broadly Gricean view that pragmatic interpretation is ultimately an exercise in mind-reading, involving the inferential attribution of intentions," and they explain that this mind-reading is accomplished by a "dedicated" "comprehension module" (2002).[10] On their view, when a speaker communicates to a listener, she gets his attention and gives information. This signals to the listener that the speaker thinks her message is relevant to him. Having access to his own conscious and unconscious mental states, the listener selects out of a wide range of possible meanings (given that literal sentence meaning always under-determines speaker meaning) the one that is most easily processed and most salient to him. "Our claim is that all human beings automatically aim at the most efficient information processing possible," whether consciously done or not (Sperber and Wilson 1986, 49). Note that a coding–decoding model of cognition and a presumed direct access to one's own consciousness (but not another's) facilitates communication on this view.

Neo-Gricean, "cognitive" pragmatist theories struggle with gesture findings insofar as they adhere to the dominant, operative philosophical view of meaning as

[10] Compare a recent study by Ivan Enrici, Mauro Adenzato, Stefano Cappa, Bruno G. Bara, and Marco Tettamanti, which argues that "Human communicative competence is based on the ability to process a specific class of mental states, namely, communicative intention," and recruits fMRI results to argue for a neural network that is used in processing communicative intention and that is modality-neutral yet accessed differently through different modalities (2011). This work seeks the neural correlates for intention processing in communication, which Sperber and Wilson presume but have not studied in a cognitive neuroscience paradigm. Yet the type of gestures they seek to add as a modality of communicative interaction are primarily emblematic, and the study measures reactions to pictures and sentences, which seems a bit thin for a "full communicative exchange" situation (2011, 2429). They conclude that communicative context transcends modality, but define communicative context by the presence of intentions to communicate, such that intentions as mental states remain a presupposed, unexplained explainer.

Author's personal copy

mentalistic, individual, and thus over-determining regarding the outcome of inquiry into normativity and convention in gesture in favor of maintaining a strong verbal/nonverbal or "linguistic/extra-linguistic" divide (see, e.g., Bara 2010; Enrici et al. 2011). For example, Tim Wharton's neo-Gricean continuum of *showing* and *non-natural meaning* places conventionalized emblematic gestures (such as a thumbs-up sign) in the middle, as a perfect example of the mixed (natural and non-natural) signs he is attempting to analyze. Yet Wharton insists that spontaneous gestures are categorically instances of natural meaning (*natural meaning*), as clouds "mean" that it will rain (2009, 149). This approach fails to appreciate the fully integrated roles spontaneous gestures play in cognition and expression, which are normative and social activities.[11] Fortunately, other routes are now emerging from Grice's lead, such that it is plausible to understand communicative activity and intersubjective sense-making as an embodied and multi-modal practice of rationality that is not always propositionally structured or processed.

Embodied and enactionist interpretations of Gricean pragmatics

Consider psycholinguistic researcher N.J. Enfield, who writes, "There is meaning in language for the same reason that there is meaning elsewhere in our social lives: because we take signs to be evidence of cognitive processes (Peirce 1955), evidence of others' communicative intentions (Grice 1957, 1975)" (2009, 2). Enfield follows Grice in claiming that language users interpret across both conventional and nonconventional signs and usages (Enfield 2009, 12). Guided by his detailed investigations into the composite speech-and-gesture utterances of Lao speakers, Enfield draws out far more fully Grice's implicit insight that *all contributing communicative behavior* is subject to *both conventional and nonconventional* usages and interpretations.

> [Nonconventional signs] become signs only when taken as signs in context. This is the key to understanding the asymmetries we observe in composite utterances like speech-and-gesture ensembles. A hand gesture may be a convention sign (e.g. as "emblem"). Or it may be non-conventional, only becoming a sign because of how it is used in that context (e.g. as "iconic" or "metaphoric"). …Hand gestures are not at all unique in this regard: the linguistic component of an utterance may, similarly, be conventional (e.g. words, grammar) or non-conventional (e.g. voice quality, sound stretches), or symbolic indexical (e.g. demonstratives like *yay* or *this*). (2009, 13)

For Enfield, *both* verbal and gestural elements in a composite utterance put burdens of *both* recognition and interpretation on the conversation participant (14). "Composite utterances are interpreted through the recognition and bringing together of these multiple signs under a pragmatic unity heuristic or co-relevance principle, i.e. an interpreter's steadfast presumption of pragmatic unity despite semantic complexity" (15). In other words, attention to intention and the effort to charitably and reasonably

[11] In a recent review of Wharton's book, Kensy Cooperrider notes in criticism, "It seems that gesture—perhaps preeminently among the nonverbal behaviors Wharton discusses—challenges the ease of disentangling nature and convention, biology and culture" (Cooperrider 2011, 87).

Author's personal copy

interpret another's communicative acts in context are necessary no matter what modality the communicative signs are in.[12] The multiple signs are easily taken together and understood at once when this sort of interpretive effort is in play. "…the mere fact of language being used triggers a process of interpretation, and the gestures which accompany speech are straightforwardly taken to be associated with what a speaker is saying" (16). Enfield's work highlights "the collaborative, public, socially strategic nature of the process of constructing composite utterances. These communicative moves are not merely designed but designed for, and with, anticipated interpreters. They are not merely indices of cognitive processes, they constitute cognitive processes" (20–21). Here, then, is a possible way out of the propositionality requirement of theories of communicative action that rest on norms of rationality.[13] Enfield demonstrates that Grice's theory can be applied to co-speech hand gestures if we take them as linguistic *enactments* of social cognition.

As I turn to next, further empirical work on the multi-modal and co-participatory nature of meaning achievement in communication demonstrates that enactive rather than representational or linguistic coding–decoding models can serve as appropriate analytic tools for intersubjective communicative action. As it turns out, a variety of possibilities exist for understanding how humans coordinate actions and intricately communicate with one another without rationally reconstructing the inner monologues of our interlocutors or fellow participants in propositionally structured monologues of our own.

Charles Goodwin's observations and writings on his father, Chil, a severe aphasic who literally communicates *with* others, helps get at the enactive, normative, and social event of making-meaning without propositions. Following a stroke, Chil can only say "yes," "no," and "and." He can make tonal noises, and he can point and gesture. Goodwin observes

> Despite his almost complete lack of productive language, [Chil] nonetheless acts as a powerful speaker in conversation. He accomplishes this by using a range of meaning-making practices beyond language itself to bring phenomena to the attention of his interlocutors who attribute relevant communicative intentions to his actions and who work hard to figure out what he wants to tell them. …The way in which Chil uses systematic practices to get others to produce the language he needs again demonstrates the relevance of focusing on the public organization of collaborative action within interaction. (Goodwin 2006, 98)

[12] It is important to note that this sort of normativity (rational interpretation) is not the only source of convention that conditions gesture meaning. We also recognize certain kinds of gestures *as* gestures, rather than other types of actions (Kendon 1980, 208). As Streeck writes in describing what are typically seen as "iconic" gestural representations, "Depiction is always a matter of convention… Whether I recognize a cluster of paint particles or a sequence of motions in the limbs as a likeness of an object or not is a matter of the methods by which these images have been made, and whether these methods are part of my cultural repertoire" (2009, 120). Thanks to an anonymous reviewer for pointing out the slippage in Enfield's use of 'sign,' quoted above.

[13] While not included in exegetical discussion here, the communicative action theories found in the work of Jürgen Habermas (1981), as well as Robert Brandom's inferential pragmatist account of semantics (1994), require at least the potential for explicit propositionality in order for an act to be rational (and therefore communicative, on this view).

Author's personal copy

Goodwin thus details a fully collaborative way in which meaning is made, in language, yet across modalities and individual bodies and consciousnesses. Chil's acts of meaning construction are co-productions of several participants who immediately recognize his communicative intent and begin working to give back to him the sense that he wants to make. Importantly, the work of the participants in helping Chil achieve his meaning is public and out in the open, offered immediately for Chil's acceptance, rejection, or modification. There is no need to wonder about the participants' inner processes. The proof is in the pudding, as it were; they are thinking with Chil, and that thinking takes place *in* their responses and gestures.[14]

Goodwin notes that "Chuck," Chil's addressee, must first recognize "that Chil's pointing finger embodies the intention to indicate something to Chuck" (2006, 106). This is not enough; Chuck must also "construe [the gesture] in a way that is relevant to the activities in progress at the moment, and to use the pointing gesture as the point of departure for a relevant next move" (106). Goodwin offers a thorough analysis of an episode in which Chil, via prosody, "nonsense" syllables, and pointing, told his son that he liked the bagel he was eating and that he noticed his son's new haircut. In this case, Chuck rightly takes a pointing gesture as Chil's indication of his enjoyment of the bagel. Goodwin explains this process by arguing that "Chil's gesture does not stand alone as an isolated pointing hand, but is instead elaborated by a number of other co-occurring signs, including a range of quite different kinds of embodied displays" (2006, 106). While especially explicit in Chil's case, Goodwin argues that this multimodality is "quite general in the organization of human gesture and action" (ibid).

The salient difference that sets Chil's communicative action apart from standard cases of speech–gesture composite utterance communication is that others offer the explicatory or elaborating speech that completes the meaning of the gesture (Goodwin 2006, 106–108). (This is not particularly rare; consider cases of "word searches" in which conversation participants overtly interpret a speaker's gestures to help her supply a temporarily missing word (see Streeck 1994)). Even as participation from interlocutors becomes putatively more active in Chil's case, highlighting the multi-modal and participatory nature of all human communication, there is no doubt that Chil is the speaker when he is the speaker. His agency and authorship are not lost; in fact, the deliberate process of judging the emerging meanings as they are offered may result in more control than the typical speaker has over the meaning of what she says. As Goodwin analyzes the event:

> …Chil and his interlocutor animate different elements of the complex carrier (gesture+talk) used to construct Chil's action, although Chil alone is the principal who commits himself to what is being asserted. His genuine agency arises from the way in which he is implicated in different stages of this process and visibly responsible for the proposition voiced by his interlocutor. (Goodwin 2006, 109)

[14] A similar conclusion is reached in Hanne De Jaegher and Ezequiel Di Paolo's analysis of an example from Currie in which a woman nonverbally draws her partner's attention and orients him generally to fully appreciate a moment she is experiencing. De Jaegher and Di Paolo write, "…it is through a process of coordination and modulation of sense-making activities that the orientee is directly affected by the orienter's intentions and sense-making and therefore he does not need to figure out what these intentions are in order to respond accordingly. A coordinated response already embodies a practical understanding" (2007, 499).

Author's personal copy

With little remaining command of the conventional verbal forms that most take to be the exhaustive constituents of language, Chil is able to communicate quite effectively, and to take on the appropriate role and commitments of a language user. This example ought to trouble the frequent conflation of communicative action with manipulation of verbal signs, driving a conceptual wedge between the two and opening up some space for broader possibilities. It may be the case that what makes human communication rational has very little to do with the modality in which it occurs; at the very least, it is clear that more than the verbal modality is crucially in play when meaning is achieved in language.

Another alternative approach to mainstream neo-Gricean accounts can be recovered from American philosopher and psychologist Eugene Gendlin, who centrally features relevance as one of the guiding features of the way humans make meaning linguistically, yet who characterizes that relevance in terms of experience and a felt sense, rather than in terms of propositional coding and decoding. In describing the progress of a conversation, Gendlin writes,

> The felt meaning (relevance) of what has gone before enables one to understand what comes next. Often one has a fairly specific sense of what will be said next, but often one is wrong. Something quite different is said next; something quite different was being lead up to. Yet, when the listener hears this rather surprising thing, he can still understand it from out of the same felt meaning that—he guessed—would lead to something else. …Both what the listener expected, and what was actually said next, were understandable from out of the relevant felt meaning. (Gendlin 1962, 129)

The meaning of the conversation unfolds, and both participants are at work in that unfolding. For Gendlin, meaning is a functional relationship between symbols and experiencing (1962). When describing bodily contributions to meaning, Gendlin uses the term "moreness" (or ". . . . .") to capture the felt sense or feeling or experiencing beyond verbal symbols that interacts with verbal symbols in the achievement of meaning. The moreness is not *preverbal*—it is part of language in the sense of intention to communicate and the informing material of communication. Gendlin explains,

> The felt meanings that function in experienced creation of meanings are always just *these* (directly referred-to) felt meanings, having whatever meaning they have. They are not indeterminate, they are merely capable of further symbolization. …If this felt meaning functions, the results will be different than if some other felt meaning functioned instead. (1962, 148)

I take this idea of the possibility of further symbolization as an important clue to the puzzle of rationality, propositionality, and the realm of the linguistic. Whatever the modality or modalities one is communicating in to others, further elaboration is always possible and is frequently required. Communication is cooperation and coordination via meanings built up for that process, in that process—it is not an exhaustive activity, or perhaps it is more accurate to say, it is not an activity that is ever finished. Communication is an achievement activity, and in this sense language can be seen as a set of the symbolic modalities dynamically deployed to accomplish such interaction.

Author's personal copy

Following Enfield and Goodwin, I suggest we recast the Gricean ideas of intentionality and mind-reading by finding them in embodied enactments. Via Enfield's notion of pragmatic unity, it becomes clear that conversation participants may treat each other's gestures as in Austin's and Grice's speech act paradigms, where interlocutors have expectations about, make assumptions regarding, and generally navigate the intentions and implications of their interlocutors' verbal performances in both their conventional and nonconventional aspects. Yet these expectations and assumptions need not exist prior to or apart from an interlocutor's response. Speech act theory can be improved by including nonverbal performances in its pragmatic analyses and by re-orienting inquiry into participant interaction (involving shared experience and cooperative symbolic practices) instead of presuming that this interaction takes the form of mutual deciphering. Following Gendlin, we can maintain the significance of sign and symbol use for the collaborative event of meaning-making without insisting that the materials used are the whole story. In other words, we can maintain a pragmatic order of explanation, yet pick up Gendlin's offer of a guiding felt meaning that is linguistic not because it is implicitly propositional but because it interacts co-dependently with linguistic forms for the purposes of co-construal in communication.

Towards a phenomenology of gestural enacting

Note that the foregoing analyses and arguments for the intersubjectively normative nature of co-speech gesturing already implicitly draw on phenomenological notions of worldhood and embeddedness in a situation as well as on embodied notions of intelligent, environmentally coupled enactions of meaning.[15] In seeking to maintain a pragmatic focus on communicative action and of "non-natural" possibilities of rightness or wrongness for specific local communities of speaker-gesturers, I am critical nonetheless of the Anglo-American tendency to restrict rational action to propositional language use, and to restrict this use for the purposes of problem solving and justification (as seen in Rorty 1989, or Brandom 1994). This tendency also colors cognitive science treatments of interaction, as Hanne De Jaegher and Ezequiel Di Paolo have noted (2007).[16] Our uses of language regularly go beyond giving and asking for reasons; linguistic activity (verbal and nonverbal) is also a medium of intentional world-relation, to use phenomenological parlance.[17] In

[15] As Dan Zahavi blends these perspectives: "The very possibility of intersubjectivity is rooted in the bodily constitution of subjectivity" (Zahavi 2005, 163).

[16] "Over-emphasis on the latter skills [explanation and prediction] has led most of contemporary social cognitive science to paint a picture of individuals who have to work out each other's minds much like they do with scientific problems. In this view, what counts as 'social' differs from non-social problem solving merely as a matter of degree" (De Jaegher and Di Paolo 2007, 486).

[17] Even Jürgen Habermas, who offers quite a disembodied and formalist account of communicative action, specifies three kinds of world relation achieved in communication action. He offers a broad, "phenomenological" model, which includes three ways (objective, social, and subjective) in which we thematize an always already given, pre-thematic, shared lifeworld (1981, 83). On this phenomenological model, "…rational expressions have the character of meaningful actions, intelligible in their context, through which the actor relates to something in the objective world. The conditions of validity of symbolic expressions refer to a background knowledge intersubjectively shared by the communication community" (1981, 13).

Author's personal copy

particular, in order to see how closely spontaneous gestures interact with verbal language use in collaborative meaning-making and reasoning, it is necessary to shift out of the paradigm of coding and decoding, of inner rehearsals of projections of others' interior monologues, and begin instead to see meaning as taking place in interactions between intelligent and dynamically expressive embodied persons. The meaning and mindedness that thus emerges is not that of two isolated consciousnesses meeting, but rather takes place exactly in the movements and the speech coordinated and orchestrated by participants in conversation.[18]

Yet this emergent, co-constructed meaning is not a free-for-all; it does not lack normativity. Indeed, as demonstrated above (Goodwin 2006), composite speech-and-gesture conversational exchanges are under perpetual monitoring and modification by all involved parties.[19] Normativity operates at various levels, ranging from the co-enacted content of what is expressed to how the interaction is going for those involved. Semantic and pragmatic criteria are jointly in play. As Dan Zahavi points out, "Our experience and understanding of others is fallible" (2005, 155). The criteria for success and failure, for better or worse takings, also comes from shared embodied structures of experience and shared practices of expression, which constrain possibilities without over-determining the sense that is made.

One must tread carefully here. On the basis of the demonstrations of co-speech gesture meaningfulness, I argue that in face-to-face conversation, meaning is a dynamic, emergent, and continually revised co-construction that is multi-modal in nature and constantly demands interpretive takings-as, which are dependent upon yet also transcendent of convention. This line of reasoning, if accepted, may only serve to prompt further questioning of *how* it is that such embodied social cognition is enacted. While in the above discussions I have taken a pragmatic route to avoid getting stuck inside intracranial phenomena, robust debate persists as to proper explanations of interactive, embodied intersubjectivity.[20]

[18] Following Scheler's critique of empathy and the problem of other minds, Zahavi writes, "we should avoid construing the mind as something visible to only one person and invisible to everyone else. The mind is not something exclusively inner, something cut off from the body and the surrounding world, as if psychological phenomena would remain precisely the same even without bodily and linguistic expressions" (2005, 152).

[19] A wonderful, detailed example of how "the audience's orientation to or away from the speaker determines the fate of the gestures that the speaker makes" is found in Streeck (1994, 257ff). Streeck describes a case in which an artist attempts to explain her exhibit to an audience of politicians visiting her art opening. At first, the politicians stare fixedly at the exhibit and miss her gestures, which grow increasingly small and half-hearted. When she notices the gaze of a member of an utterance is directed at her, the artist begins gesturing again, this time much more boldly and symbolically (259–265). Streeck's description of this event gives us some fledgling criteria for gesture failure ("…they neither merge into one another nor combine into complex constructions. They remain isolated, bounded simple events" (259)) and success (the gestures "becoming structures in space that are set up in a sequence of preparation and stroke…" (262)).

[20] See for example the conversation between Hanne De Jaegher and Shaun Gallagher on how best to frame what goes on in social interaction (De Jaegher 2009a, b). As Ezequiel Di Paolo suggests in a recent interview, "We're still facing the messy beginnings in this [enactive] approach, still exploring our precursors, finding affinities with research lines in different disciplines that are enactive in all but name and clarifying open issues, getting at better foundations" (Interview with Ezequiel Di Paolo. *New APPS: Art, Politics, Philosophy, Science*. 1 Jun. 2011. Web.)

Author's personal copy

Contemporary phenomenological (or phenomenologically based) inquiry into this field tends to focus on encounters and experiences that are considered "pre" or "non" linguistic, such as facial expressions, emotional displays, and just being there with—or for—the other in a shared world.[21] I suggest that as another, complementary line of inquiry, we need a phenomenological account of the intersubjectivity of co-speech gesturing.[22] As Dan Zahavi demands, "any convincing theory of social cognition should be able to account for our face-to-face encounters with others" (2008, 515). The face-to-face encounters of enlanguaged humans frequently involve linguistic behavior. An account of social cognition will involve a treatment of language, which in turn will properly involve a treatment of co-speech gestures.

In the final discussions that follow, I examine just one encounter between gesture scholarship (David McNeill's *imagery–language dialectic*) and phenomenology (Maurice Merleau-Ponty's *sedimentation–spontaneity dialectic*). In appraising a gesture scholar's uptake of select phenomenological insights, I hope to open up a broader, reciprocally critical exchange between disciplines regarding intersubjectivity at the level of embodied linguistic practice. I maintain that to better understand the event of bodily expression that gesture enacts as an intersubjectively normative achievement, we must be sensitive to ambiguity and encounter, and wary of an overly strict dialectical processing approach. In particular, I contend that McNeill (2005) does not adequately appreciate the *sedimented* aspects of Merleau-Ponty's picture, and hence the socio-culturally constituted nature of the possibilities of meaning construction that his phenomenological view offers.

Merleau-Ponty on gesture: sedimentation–spontaneity dialectic

In *Phenomenology of Perception*, Maurice Merleau-Ponty presents the dialectic interplay of sedimentation and spontaneity as an explanation for how speech is meaningful and how thoughts come into being via bodily accomplishment (2002). Gesture, for Merleau-Ponty, is the spurring force, the bodily act of a speaker using conventional language to say something original. In this model, *sedimentation* describes spoken speech, which provides the material of thoughts and the verbalizations. On the other hand, *spontaneity* marks true speech, original speaking, that is, the birth of a new relationship between myself, the world, and others that brings thought into existence through an appropriation of constituted language (2002, 213). These previous acts of expression are not merely a static assemblage of dictionary definitions or an inanimate corpus; Merleau-Ponty describes the spoken word as an embodied habit that enables in the first place any response to our "momentary desires" to make meaning out of "the primordial silence" (213). For Merleau-Ponty, spoken words are gestures, because they are embodied appropriations (217).

[21] "Without ever denying the eminently intersubjective character of *language*, phenomenologists have often endeavored to unearth pre- or extra-linguistic forms of intersubjectivity, be it in simple perception or in tool-use, in emotions, drives, or body-awareness" (Zahavi 2005, 176). Zahavi takes this to be "decisively different" in approach from Habermas.

[22] Notably, Beata Stawarska develops a dialogical phenomenology (2009a) that engages with empirical work on infant–caregiver interactions. See also Stawarska 2009b. While gesturing as Stawarska discusses is not paradigm co-speech gesturing, as pre-linguistic infants are involved, this work assists in defeating the theory-theory of mind and paving a way for the sort of inquiry being called for here.

Author's personal copy

A gesture for Merleau-Ponty is the way that meaning *inhabits* a body and a body inhabits acquired ways of expressing, which is to say, the way that a particular existing, thinking, and communicating body-subject *lives*—and creates—a particular meaning. A gesture is a meaningful bodily act, the way a human body always transcends itself towards some significance. Gesture, then, is precisely the simultaneous constitution of thought and expression. In distinction from the Cartesian theater paradigm that dominated philosophy of mind since modernity, language for Merleau-Ponty does not represent some interior item awaiting expression. Merleau-Ponty tells us that "thought is no 'internal' thing and does not exist independently of the world and of words" (213). Thought is *achieved* or *completed* in bodily expression; once expressed via the *gesture* of taking up constituted speech it may recur in what appears, but only appears, to be an inner monologue. Expression is the outcome of the dialectic of sedimentation and spontaneity, in that expression is the body's appropriation of acquired form in a new act of meaning-giving. The body is always the medium of expression. Speech is already gesture: the use of words is an instance of body movement and expression. Gesture is the happening, or enactment, of thought.

Sedimented language is stretched, bent, and inhabited in order that new thought come into being, into form and presence. Acquired (conventional) words and thoughts are taken up through a stylistic inhabiting, which Merleau-Ponty calls gesture. The spontaneity–sedimentation dialectic unites thought and language completely. For Merleau-Ponty, gesture is this *stroke* through which a speaking subject incorporates the past into the present, establishing continuity with a previous context of thought and meaning in the same moment that she gives embodied expression to a new and original idea that is at once communicable and recognizable as such. We can define "gesture" for Merleau-Ponty as the stylistic inhabiting of acquired words and thoughts to make new meaning, to think new thoughts or speak originally.

McNeill's imagery–language dialectic

Psychologist David McNeill's treatments of co-speech gesture (e.g., 1992, 2000, 2005) are influential in the field of gesture studies, and McNeill offers one of the most theoretically robust accounts of how the gestural modality operates in both thought and expression. On McNeill's view, gesturing and gestural thinking aids the speaker in cognitive preparation for making an utterance, rhythmically guides the execution of the utterances, and carries certain expressive content in the utterance that is meaningful to other conversation participants (1992, 2005). While Merleau-Ponty's mentions of gesture and his claim that "the spoken word is a genuine gesture" (2002, 213) are not necessarily intended to explain the same phenomena that McNeill researches, the *imagery–language dialectic* that McNeill introduces to account for the cognitive processing/expression producing dual role of gestures is inspired by Merleau-Ponty's sedimentation–spontaneity dialectic for speech significance (McNeill 2005).

McNeill's *imagery–language dialectic* runs on the tension of two unlike cognitive modes juxtaposed in time and in concept as the two sides of an underlying idea unit. Speech and gesture respectively embody two unlike modes of thinking, according to McNeill: the speech mode is linear, analytic, categorical, constrained, and

Author's personal copy

conventional, while the gesture mode is characterized as holistic, imagistic, less constrained, and idiosyncratic. The conflict of these modes gets resolved in a well-formed utterance that usually includes gesture. In the framework that McNeill is establishing, then, "to make a gesture… is to bring thought into existence on a concrete plane… The greater the felt departure of the thought from the immediate context, the more likely is its materialization in a gesture, because of this contribution to being" (2005, 99).[23]

McNeill explains the occurrence and absence of co-speech gestures in terms of *communicative dynamism*, arguing that gestures are more likely to occur when the utterance content emerges as a salient point or contrast to the unfolding communication scenario. Thus, "the higher the newsworthy content, the more elaborate the image" and the more likely the gesture (1992, 57). The gesture increases in likelihood and complexity in relation to the idea unit's differentiation from current discourse. McNeill's more recent "H-model," named after Heidegger and meant to capture fleeting disclosures of being (one's particular being, for McNeill), presents communicative dynamism in existential terms: "The H-model is in this way an extension of Merleau-Ponty's "existential content of speech" (and gesture). It gives existential content an interpretation on the level of cognitive being" (2005, 99). An *absence* of gesture indicates little or no contrast in a speaker's ongoing expressing. Thought is an unpacked idea unit that arises as a break from a given context; therefore, an absence of gesture—or the cessation of the dialectic and the remainder of "pure verbalism"—indicates an absence of thought. McNeill surmises, "All this implies that the dialectic itself varies proportionately with communicative dynamism and memory, and when these conditions are absent speaking is no longer merging with thinking" (2005, 103). McNeill boldly concludes: speech without gesture is "speech without thought" (ibid).

Dueling dialectics?

McNeill's idea of a not-yet-articulated but fully intentional idea unit is analogous to Merelau-Ponty's spontaneous flash or upsurge of "new intention" or "pure thought"—a new moment in the body's ongoing meaningful engagement with the world, one that initiates its own inhabiting of sedimented forms to new expressive ends. It follows that McNeill's process of unpacking, or the playing out of the imagery–speech dialectic, is comparable to the *act* of appropriation and inhabitation of constituted speech, that is, the process of Merleau-Ponty's spontaneity–sedimentation dialectic. The suggestion

[23] In an often referenced example, a participant in McNeill's lab re-tells the narrative of a Sylvester and Tweety cartoon wherein Tweety drops a bowling ball into the drainpipe of which Sylvester is concurrently climbing up the interior. The speaker's sentence expresses that Tweety takes the bowling "ball and drops it down the drainpipe." She makes a symmetrical two-handed gesture with palms loosely curved and facing down. The downward stroke of the gesture is synchronous with 'down'. Importantly, the gesture stroke does not coincide with the verb 'drops', but is withheld to co-occur with 'down'. The gesture is withheld because the core concept to be accomplished in this instance, according to McNeill, is what the *bowling ball* was doing and how it pushed Sylvester down a drainpipe. Though Tweety was still the agent in the utterance according to the verbal information, the gesture aided in transitioning to an understanding of the bowling ball as the real agentive force and 'it down' as the true "anchor" of the sentence (McNeill 2005, 122). In this way, the gesture manifests what the speaker takes as the important point requiring emphasis in that moment of the dialogue.

Author's personal copy

then is that the act of appropriation (in Merleau-Ponty's terms) could also be explained by the interaction of gesture (imagistic thinking) and speech (linear analytic formal thinking), with gesture in McNeill's sense being a way of talking about active, true, original, or "speaking" speech in Merleau-Ponty's sense (and McNeill's speech being tantamount to Merleau-Ponty's spoken, constituted speech). Then gesture is the root (as that imagistic "half" of thinking) and the manifestation (as the physical expressive action that accompanies speech) of Merleau-Ponty's "new sense-giving intention."

Rather than attempt to make these dialectics line up (which requires a bit of gymnastics to get around Merleau-Ponty's broader use of "gesture"), the more exciting result that emerges when we put these dialectics side by side is not that they are mutually translatable, but that they are reciprocally critical. Merleau-Ponty's notion of gesture as stylistic, intentional inhabiting can act as a corrective to McNeill's strict modal dichotomy. For Merleau-Ponty, the thinking–expressing gesture is an embodied and embedded act that carries, creates, and delivers its meaning holistically. The dialectic enabling the appropriative act of gesture marks a distinction between spoken speech and speaking speech, a difference between speech that says something new as opposed to inauthentic or mindlessly recycled speech. The tension is between the creative versus the conventional quality of this gesture: To what degree does it stretch, elaborate, push, or reinterpret the forms it takes up? To what extent is this taking up of given forms able to say something that has not been said before? What is the effect of this appropriative inhabiting? How does this gesture make a difference to meaning at this moment? While McNeill walks a similar path with the notion of H-model and a speaker's shifting "cognitive being," he does not need to take on the weight of authenticity claims. Rather than winding up with the conclusion that a speech act that lacks gesture is "inauthentic" or mindless ("without thought"), it is interesting, plausible, and productive to question how our verbal linguistic behavior can be spontaneous and how our gestural linguistic behavior demonstrates sedimentation and "rule-following."[24] Put another way, for Merleau-Ponty, "spontaneous" action is never blind, automatic, or cleanly separable from convention.[25]

McNeill's contribution here consists in pointing out the ubiquity and semantic richness of hand gestures, which Merleau-Ponty does not seem to notice; therefore we can now say that Merleau-Ponty gives us an impoverished picture of linguistic expression as being only verbal—and it is this verbal behavior that Merleau-Ponty describes as gestural. Yet in turn, McNeill seems to miss exactly that insight—that our use of verbal language is idiosyncratic, contextual, stylistic, intentional, and particularly meaningful—whenever he characterizes the "linguistic" or verbal side of his dialectic as strictly linear, conventional, and so on. While philosophers of language ought to attend

[24] Furthermore, it is not clear that Merleau-Ponty would conscience a clean dialectical separation of the conventional (non-natural) and nonconventional (natural) within us. As he writes in *The Structure of Behavior*, "Man is not a rational animal. The appearances of reason and mind do not leave intact a sphere of self-enclosed instincts in man" (2006, 181). For Merleau-Ponty, the acting body is always discriminating and *taking as*, whether in speech, perception, or gesture.

[25] "Action is the action of subjects; it is the action of minded individuals" (Zahavi 2005, 161). As Etienne Bimbenet describes the defining ambiguity of perception in Merleau-Ponty's account, "It turns out, and this is the ultimate point, that there is in perception as much passivity as there is spontaneity, or that perception is a feeling at the same time that it is a thought" (Bimbenet 2009, 73).

Author's personal copy

to McNeill's general claim that language has been construed too narrowly, he himself should avoid falling into the same trap, merely supplementing language (or psycholinguistic processing models of thinking-for-speaking) with a gestural channel of cognition and expression, rather than rethinking linguistic activity as such.

Philosophical preference for a paradigm of intersubjective gestural sense-making

It is worthwhile to note how McNeill's interpretation of Merleau-Ponty is in line with a general eschewing of the social, normative, and conventional aspects of co-speech gestural practice. Given the importance of the normativity in language meaning, there are good philosophical reasons to be cautious of accounts of gesture that want to see them as natural, romantically expressive, or incapable of failure, critique, or conscious control. While he rightly points out that gestures are part of language (1992, 2) and that they "accomplish thought" in a Merleau-Pontian sense, McNeill walks a precarious path whenever he insists on categorizing spontaneous speech-accompanying hand gestures as unconventional and whenever he speaks of them as especially revelatory of a speaker's inner thought processes. Regarding the impact that gestures have on thought, McNeill writes that "gesture supplies the idiosyncratic, the personal, and the context-specific aspects of thought, to be combined with the socially regulated aspects that come from the conventions of language" (1992, 2). This perspective easily renders gesture incapable of failure and seemingly immune to audience reception.[26]

In fact, a Merleau-Pontian interpretation should point in the other direction. Precisely since "expressions are not merely exterior manifestations of something that was already internally present" but instead "what is expressed is fully realized only in the expression," (Zahavi 2005, 152–153), word choice matters, and gestures can go wrong. Indeed, following this line of thought, we can start to realize all of the many ways that gestures may turn an expression in an unintended direction, for example, or elicit shades of discomfort from an interlocutor.[27] Why take as given that co-speech gestures are always true, helpful, or readily and successfully interpreted? We do better to at the very least include a "meaning-building" inquiry into gestural phenomena alongside this well-established "meaning-leaking" research paradigm.

On my reading, a meaning-leaking paradigm retains traditional representationalist and individualist characterizations of cognition by holding gestures to be uncontrollable, unconscious windows to speakers' thought patterns and intentions.[28]

[26] A narrow sense of linguistic convention can yield vexing results in empirical gesture scholarship, for instance a recent study that investigates whether a hand gesture is conventional *or* motivated by a conceptual metaphor, as if these are mutually exclusive options (Parill 2008). An overly rigid, overly dichotomous view of imagistic and spontaneous thinking as being *opposed* to "linguistic" thinking wrongly presents some aspects of cognition as sheltered, personal, and pure of culturally shared habituation and sedimentation.

[27] Recall Merleau-Ponty's discussion of the angry Japanese person who smiles (2002, 219).

[28] For example, gestures unwittingly give away our lies (Franklin 2007), while speech–gesture mismatches may "…point to a state of overload, in which the speaker's emotional content exceeds the means of expression, and marks the search of this means of expression by the speaker" (Waisman 2010, 173). Such explanations logically fall out of a theory of speech–gesture interaction that ascribes to gesture all of the personal, idiosyncratic, and nonconventional aspects of cognition and communication. See McNeill and Duncan (2000) for a clear statement of the "window" view.

Author's personal copy

This can be contrasted with a meaning-building paradigm that portrays gestures as external objects accessible for both speakers and listeners to monitor and interact with. Those interested in demonstrating how embodied cognition is socially structured have reason to improve upon a line of inquiry that sees certain embodied communicative practices as accidental or irrational. As we have seen from the foregoing speech-act analysis, it oversimplifies matters to see gestures as *only* and *ever* "spontaneous" (unconventional, or non-normative) and speech as only and ever "sedimented" (conventional, or normatively regulated). Furthermore, a focus on individual cognitive processes tends to overlook the irreducible sociality of linguistic activity, and at its worst locates intersubjective meaning achievement in mind-reading modules rather than in the shared tangible space and actions of participants in dialogue.

On a meaning-building model, by contrast, both speaker and listener "co-author" the emerging dialogue *and the gestures* that constitute it (Streeck 1994, 248). Jürgen Streeck's research on the interaction of gaze and co-speech gestures identifies a broad cross-cultural tendency:

> … as speakers…initiate gestural events that are integral parts of the message, they shift their gaze to the gesture, and then, as they produce the word that is most intimately tied to it, they look back at the recipient. (1993, 288)

In this activity, "the gesture is thus made part of what is reasonably "visible" in the interaction (I know that you know that I know that you have seen it)" (Streeck 1993, 289). This is a directly embodied performance of the kind of mutual interpretation of intentions that Grice highlights as integral to communication and rationality. Since both the speaker and the listener varyingly and jointly attend to—look at—a gesture, speaker and listener are able to relate it to the meaning that is unfolding, which in turn allows them to collaboratively direct this emerging meaning. "Since the speaker can see and does see what she is doing, she could *suppress* her manual behavior any time if it appeared to her as symbolically or otherwise undesirable" (Streeck 1993, 289). Rather than taking gestures as uncontrollable leakages of one's thought processes, by attending to their visibility and tool-like object-ness, theorists may follow participants in taking their conversation partner's gesturing while speaking *as* part of the enactment of his meaning.[29]

One consequence of the linguistic turn can be seen in decisions that various traditions make to self-consciously avoid foundationalism and offer explanations that are normatively grounded not by brute objective reality to which we have

[29] See Streeck's more recent writings (2009, 2010) for a careful identification of six gestural ecologies, which on his account are kinds of world-relation accomplished through hand gestures made in conversation and coupled with local environments in different ways. Streeck writes, "…this heuristic enables us to take note of the fact that hand gestures not only embody meaning and mediate communication in heterogeneous ways, but also bring the communicating body in contact with the world in a variety of distinct modes" (2010, 226). While beyond the scope of the current essay, I find Streeck's recent formulations promising for dialogue with someone like Habermas. Following this lead, we might say that gesture ecologies' various ways of relating to the world in communicative action are so many enactments of rationality. Streeck's work, both earlier and later, is particularly attuned to the recipient-designed nature of gestures, demonstrating that gestures require intersubjective recognition—one of Habermas's criteria for rational attempts. Streeck's example of the artist ceasing and then redesigning her gestures for her audience show that gesture attempts can fail (another Habermasian criterion).

Author's personal copy

unmediated access, but rather by intersubjective and post-conventional (self-critical) standards that communities build and rebuild together. I have argued here that this insight is correct but does not go far enough; it leaves us stranded on the shores of lingualism, when our communicative practices and our ways of reasoning and knowing together are richly embodied and regularly involve nonverbal and non-propositional, or to put it positively, visible, kinesic, bodily practices. Furthermore, it is not clear that our received understanding of rightness or wrongness has been well-conceived for verbal language; too often it is measured in terms of competence (with forgiveness more readily available for slips in performance), when what matters is better or worse takings relative to a shared purpose and a community. Yet as this turn evolves and is corrected in various ways, whether by analytic returns to consciousness studies (e.g., McDowell 1994), by a pragmatic turn to practices of achievement (see Egginton and Sandbothe 2004; Allen 2008), or by continued investigations in an enaction research paradigm (Stewart et al. 2010), we do well not to revert to strict dichotomies and dialectics that merely supplement linguistic activity with a vast, undifferentiated field of "meaningful" yet "nonlinguistic" remainders.

References

Allen, B. (2008). *Artifice and Design: Art and Technology in Human Experience*. Ithaca: Cornell University Press.

Austin, J. L. (1961). *Philosophical Papers*. Oxford: Clarendon.

Bara, B. G., & Douthwaite, J. (2010). *Cognitive Pragmatics: The Mental Processes of Communication*. Cambridge, Massachusetts: The MIT Press.

Beata Stawarska 2009a. *Between you and I: dialogical phenomenology.* Ohio University Press.

Beata Stawarska & Meeting. 2006.*: International Association for Phenomenology and the Cognitive Sciences; Intersubjectivity and embodiment* Springer; 2006.

Bimbenet, E. (2009). Merleau-Ponty and the quarrel over the conceptual contents of perception. *Trans. Robin Muller. Graduate Faculty Philosophy Journal, 30*(1), 59–78.

Brandom, R. (1994). *Making It Explicit: Reasoning, Representing and Discursive Commitment*. Cambridge: Harvard University Press.

Calbris, G. (1990). *The Semiotics of French Gestures*. Bloomington: Indiana University Press.

Cienki, A., & Müller, C. (2008). Metaphor, gesture, and thought. In R. W. Gibbs Jr. (Ed.), *The Cambridge Handbook of Metaphor and Thought*. Cambridge: Cambridge University Press.

Cooperrider, K. (2011). *Review* of *Pragmatics and Nonverbal Communication*, *by Tim Wharton (2009)*. *Gesture, 11*(1), 81–88.

De Jaegher, H. (2009a). Social understanding through direct perception? Yes, by interacting. *Consciousness and Cognition, 18*(2), 535–542.

De Jaegher, H. (2009b). What made me want the cheese? A reply to Shaun Gallagher and Dan Hutto. *Consciousness and Cognition, 18*(2), 549–550.

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

Egginton, W., & Sandbothe, M. (2004). *The Pragmatic Turn in Philosophy: Contemporary Engagements Between Analytic and Continental Thought*. Albany: State University of New York Press.

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge: Cambridge University Press.

Enrici, I., Adenzato, M., Cappa, S., Bara, B. G., & Tettamanti, M. (2011). Intention processing in communication: a common brain network for language and gestures. *Journal of Cognitive Neuroscience, 23*(9), 2415–2431.

Ezequiel Di Paolo. *New APPS: Art, Politics, Philosophy, Science*. 1 Jun. 2011. Web.

Franklin, A. (2007). Blending in deception: tracing output back to its source. In E. T. Levy, S. D. Duncan, & J. Cassell (Eds.), *Gesture and the Dynamic Dimension of Language: Essays in Honor of David McNeill* (pp. 99–109). Amsterdam: John Benjamins.

Author's personal copy

Fuchs, T., & de Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences, 4*(8), 465–486.
Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences, 1*(4), 23–48.
Gendlin, E. (1962). *Experiencing and the creation of meaning: A philosophical and psychological approach to the subjective*. Evanston: Northwest University Press.
Goodwin, C. (2006). Human sociality as mutual orientation in a rich interactive environment. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of Human Sociality: Culture, Cognition, and Interaction*. Oxford: Berg.
Goodwin, C., & Goodwin, M. H. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica, 62*(1–2), 51–76.
Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge: Harvard University Press.
Habermas, J. (1981). *A Theory of Communicative Action* (Vol. 1). Boston: Beacon.
Hutto, D. (2005). Knowing What? Radical Versus Conservative Enactivism. *Phenomenology and the Cognitive Sciences, 4*(4), 389–405.
Johnson, M. (2007). *The Meaning of the Body: Aesthetics of Human Understanding*. Chicago: University of Chicago Press.
Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *Nonverbal Communication and Language* (pp. 207–227). The Hague: Mouton.
Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes, 24*(2), 145–167.
Koopman, C. (2011). Rorty's linguistic turn: Why (more than) language matters to philosophy. *Contemporary Pragmatism, 8*(1), 61–84.
Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.
McDowell, J. H. (1994). *Mind and World*. Cambridge: Harvard University Press.
McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago: The University of Chicago Press.
McNeill, D. (2005). *Gesture & Thought*. Chicago: University of Chicago Press.
McNeill, D., & Duncan, S. D. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and Gesture*. Cambridge: Cambridge University Press.
Menary, R. (2010). Introduction to the special issue on 4E cognition. *Phenomenology and the Cognitive Sciences, 4*(9), 459–463.
Merleau-Ponty, M. (2002). *Phenomenology of Perception*. London: Routledge Classics.
Merleau-Ponty, M. (2006). *The structure of behavior. Trans. Alden L. Fisher*. Pittsburgh: Duquesne University Press.
Müller, C. (2007). A dynamic view of metaphor, gesture and thought. In E. T. Levy, S. D. Duncan, & J. Cassell (Eds.), *Gesture and the Dynamic Dimension of Language: Essays in Honor of David McNeill*. Amsterdam: John Benjamins.
Parrill, F. (2008). Form, meaning, and convention: A comparison of metaphoric gesture with an emblem. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture*. Amsterdam: John Benjamins Publishing Company.
Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
Rorty, R. (1989). *Contingency, Irony, Solidarity*. Cambridge: Cambridge University Press.
Simpson, D. (2010). Language and know-how. *Phenomenology and the Cognitive Sciences, 4*(9), 629–643.
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Cambridge: Harvard University Press.
Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind and Language, 17*(1), 3–23.
Stawarska, B. (2009). Merleau-Ponty and Sartre in response to cognitive studies of facial imitation. *Philosophy Compass, 4*(2), 312–328.
Steiner, P., & Stewart, J. (2009). From autonomy to heteronomy (and back): The enaction of social life. *Phenomenology and the Cognitive Sciences, 4*(8), 527–550.
Stewart, J. R., Gapenne, O., & Di Paolo, E. (2010). *Enaction: Toward a New Paradigm for Cognitive Science*. Cambridge: MIT.
Strawson, G. (2004). Real intentionality. *Phenomenology and the Cognitive Sciences, 3*(3), 287–313.
Streeck, J. (1993). Gesture as communication I: Its coordination with gaze and speech. *Communication Monographs, 60*(4), 275–299.

Author's personal copy

Streeck, J. (1994). Gesture as communication II: The audience as co-author. *Research on Language and Social Interaction, 27*(3), 239–267.
Streeck, J. (2008). Depicting by gesture. *Gesture, 8*(3), 285–301.
Streeck, J. (2009). *Gesturecraft*. Amsterdam: John Benjamins Publishing Co.
Streeck, J. (2010). *New Adventures in Language and Interaction*. Amsterdam: John Benjamins Publishing Co.
Sweetser, E., & Sizemore, M. (2008). Personal and interpersonal gesture spaces: Functional contrasts in language and gesture. In A. Tyler, Y. Kim, & M. Takada (Eds.), *Language in the Context of Use: Discourse and Cognitive Approaches to Language*. New York: Mouton de Gruyter.
Thompson, E. (2005). Sensorimotor subjectivity and the enactive approach to experience. *Phenomenology and Cognitive Sciences, 4*(4), 407–427.
Waisman, O. S. (2010). *Body, Language and Meaning in Conflict Situations: A Semiotic Analysis of Gesture-Word Mismatches in Israeli-Jewish and Arab Discourse*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Wharton, T. (2009). *Pragmatics and Nonverbal Communication*. Cambridge: Cambridge University Press.
Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First Person Perspective*. Cambridge: MIT.
Zahavi, D. (2008). Simulation, perception, empathy. *Consciousness and cognition, 17*(2).