Language Sciences 97 (2023) 101524

Contents lists available at ScienceDirect

Language Sciences

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

Multimodal coordination and pragmatic modes in conversation

Camila Alviar [a,c,*], Christopher T. Kello [a], Rick Dale [b]

[a] University of California, Merced, USA
[b] University of California, Los Angeles, USA
[c] Vanderbilt University Medical Center, Nashville, TN, USA

A R T I C L E I N F O

Article history:
Received 4 September 2020
Received in revised form 1 October 2022
Accepted 22 November 2022
Available online 3 February 2023

Keywords:
Pragmatics
Coordinative structures
Ecological psychology
Dynamical systems
Gesture and speech

A B S T R A C T

Language is intrinsically multimodal. Speakers use gestures, prosody, gaze, and facial expressions as cues that complement and expand the meaning expressed in their words. These varied signals operate in remarkably flexible coordination, constantly adapting to the conversational partners and topics as they change over time. We argue that an ecological approach to multimodal behavior offers a promising account of natural conversation as it takes place both in experimental contexts, and in natural ones outside the lab. After reviewing major historical themes in the study of language and communication, we describe how this ecological perspective situates future work, especially work that seeks to quantify these processes. We describe a quantitative hypothesis that multimodal signals are projected on manifolds of lower dimension that can be described in terms of dynamical systems. We refer to these lower dimensional patterns as “pragmatic modes,” and compare this idea to a number of prior theoretical proposals. We describe how the notion of pragmatic mode frames a quantitative basis to supplement and extend prior research with explicitly quantitative goals. The paper concludes with an outline to link quantitative descriptions of multimodality with more abstract, qualitative theories of the past few decades, and describe how future research might explore pragmatic modes, how they change over the course of conversation, and relate to our understanding of human communication.

1. Introduction

Language is intrinsically multimodal. Across cultures and in very many contexts, language remains a predominantly face-to-face endeavor, in which at least two people are involved. Not only words are exchanged, but also looks, tones, gestures, and objects are weaved together smoothly to form meanings. Language use may also predominantly involve a joint purpose, because it helps to coordinate interlocutors to perform tasks. Tasks can include story-telling, fighting, flirting, or even “phatic communion” in search of a conversation topic; tasks can also include elaborate spatial directions to a stranger visiting a city, or technical descriptions of a scientific domain to a whole group of strangers. In all such instances of interaction, one sees a process of coordination among interlocutors. This coordination is a complex multiscale

* Corresponding author. C. Alviar was affiliated with UC Merced when the first version of the manuscript was conceived. She is now affiliated with Vanderbilt University Medical Center, Music Cognition Lab, Department of Otolaryngology, 1408 17th Ave. S., Nashville, TN, 37212.
E-mail address: maria.c.alviar-guzman@vumc.org (C. Alviar).

https://doi.org/10.1016/j.langsci.2022.101524

performance. Just intuitively introspecting the scales of this performance presents an imposing scientific landscape: Phonemes constitute syllables, syllables words, words utterances, and utterances conversational turns and topics; gesture and prosody also respond to this structure, mirroring it to help group information, aid turn taking, and accomplish pragmatic goals. In mere moments of interaction, human groups do all of this simultaneously and relatively smoothly. A long tradition across a number of realms of the language and communication sciences supports these observations (among many: Bavelas et al., 1986; Clark, 1996, 2003; Dale et al., 2013; Di Paolo et al., 2018; Enfield, 2013; Fusaroli and Tylén, 2012; Garrod and Pickering, 2004; Giles, 2016; Goodwin, 1979, 2018; Grosz and Hirschberg, 1992; Grosz & Sidner, 1986; Mondada, 2016; Newtson, 1993; Rasenberg et al., 2020; Rączaszek-Leonardi, 2010; Sacks, 1992; Sacks et al., 1978; Schegloff, 1990, 1991, 1996; Watzlawick et al., 1967/2011).

In the past few decades, research has begun to quantify these patterns of coordination using novel analysis and measurement techniques. For example, such projects have measured how people follow each other's eye gaze (Jermann and Nüssli, 2012; Nomikou et al., 2016; Richardson and Dale, 2005), subtly mimic each other's movements and tones (Grammer et al., 1998; Levitan and Hirschberg, 2011), coordinate words over time (Brennan and Clark, 1996; Fusaroli, Bahrami et al., 2012), continuously exchange the floor with each other with little delay (Stivers et al., 2009), adapt their words and constructions to each other's needs (Branigan et al., 2000; Healey et al., 2014), and recover quickly from miscommunication (Paxton, Roche et al., 2021; Dingemanse et al., 2015). These programs of research have yielded rich quantitative confirmation of a highly complex process—people shape multiple aspects of a conversation into a smooth flow and without much conscious effort (cf. Garrod and Pickering, 2004). This confirms long-standing work in conversation analysis and related domains that made early observations of these characteristics of interaction (e.g., Goodwin, 1979), especially ones that articulate complex relationships across the verbal and nonverbal actions of interlocutors (e.g., Schegloff, 1996).

This coordination is astonishing, even puzzling, given all the modalities and behaviors available to convey any single message (cf. Buder et al., 2010; Oller et al., 2013). Indeed, the many types of verbal and nonverbal behavior afford a great many *potential* combinations during an episode of interaction. We can consider these potential combinations as the degrees of freedom available to interlocutors, who must select and constrain them on the fly while communicating, all while abiding the importance of timing, rhythm, and social norms in language use (Fusaroli et al., 2014; Heras-Escribano, 2019; Shockley et al., 2009).

Despite these observations, many domains of the language sciences still focus on single or small numbers of behaviors, void of context, in individual tasks involving the "linguistic modality," such as the cognitive processes that allow individuals to produce and understand words (see discussion in: Garrod and Pickering, 2004; Linell, 2004; Perniss, 2018; Rączaszek-Leonardi, 2014; Tanenhaus and Brown-Schmidt, 2008). This focus is useful and justifiable when seeking to manage the complexity of language to yield controlled, experimental findings. It has produced a vast and valuable amount of knowledge on the mechanisms of some aspects of language like grammar and phonology, but it can only offer a more complete view by integrating them broadly. Over the past several decades, prominent alternative research traditions, ones maintaining focus on multimodality in natural interaction, have gained traction in the language sciences. These traditions have consistently shown the importance of the other modalities and joint actions for both discourse regulation and meaning making, especially in natural language use, amidst multiple parties to a conversation (e.g.: Clark, 1996; Enfield, 2013; Schegloff, 1996; Stivers and Sidnell, 2005; Streeck and Jordan, 2009). Now, even approaches like constructional grammar, traditionally focused exclusively on studying the verbal components of language daily use, are actively debating whether, and under which cases, the nonverbal information should also be considered as a fundamental and constitutive part of the constructional sign (see Zima and Bergs, 2017 and the papers in that special issue).

An important next step towards understanding the intrinsic multimodality of language is to integrate it into a broader theoretical framework, especially one that helps bridge these basic facts about language to quantitative approaches in the language sciences (Dale et al., 2013; Rasenberg et al., 2020; Shockley et al., 2009). If it is true that multimodality is intrinsic to natural language, then a critical new understanding of language will derive from understanding how natural interaction *integrates* these many "signals." Again drawing from the concept of degrees of freedom: This perspective would solve the fundamental puzzle of how interlocutors can operate so smoothly among such a wide array of potential behavioral choices. There was a time when asking such a question may have seemed overwrought—how does one measure many dimensions of language and communication and then proceed to quantify their combinations? Indeed, core studies in conversation analysis, where this theoretical tradition is strong (Goodwin, 1979; Goodwin and Heritage, 1990), often remain focused on qualitative analysis of a specific behavior despite that its ultimate theoretical direction is a multimodal one (for an early such lament, see Schegloff, 1996). But what are the dynamic processes by which two or more people come together and bring about these multimodal performances? What are the characteristic stabilities and dynamic patterns that drive them? Modern language sciences are now in a position to take on these large-scale problems of multimodality directly, through automated and semi-automated measurements and statistical methods.

But these methods are only informative insofar as they are combined effectively with a theoretical perspective to frame them. In this paper, we argue that a theoretical framework rooted in ecological psychology has great promise for understanding the kind of coherence amidst complexity we see in interaction. We define ecological perspectives in general terms here, though there are important variants of these theories, with distinct relationships to other theoretical perspectives. For scope, we avoid these detailed discussions here, but the reader is encouraged to consider Szokolszky et al. (2019) which includes excellent detail and subtlety (see also Chemero, 2009). Indeed, Szokolszky et al. (2019) offer a clear frame for our brief discussion here:

> The theoretical starting point for Ecological Psychology has always been the recognition that perceiving is an activity of organisms, and it is functionally and logically coupled to action … The foundational idea of the reciprocal relationship between perceiving and acting, between organism and environment, indicates a profound departure from the traditional view and goes against the prevailing mechanistic and computational approach to living systems and life in general. (Szokolszky et al., p. 365)

We take this to be a statement of the ecological approach "writ large." This ecological tradition, rooted to a great extent in the inextricable connection between perception and action (Bernstein, 1967; Gibson, 1966, 1979), is now influential in various realms of cognitive science. Perspectives inspired by ecological psychology quickly influenced developmental psychology, various domains of language and communication, social cognition generally and much more (for some review see Chemero, 2009; Dent-Read and Zukow-Goldring, 1997; Hodges and Fowler, 2015; Iverson and Thelen, 1999; Marsh et al., 2006; Szokolszky et al., 2019; Reed, 1996; Thelen and Smith, 1996).

We draw from ecological psychology for a specific purpose in the present manuscript. The ecological approach, and research it inspired, has introduced a robust tradition of operationalizing and quantifying complex, interacting systems. For example, there is a strong tie between this approach to psychology and an approach rooted in dynamical systems (Chemero, 2009). Under this approach, a variety of aspects of perception and action have been explored using explicit quantitative models. Especially prominent here are processes of coordination (Turvey, 1990).

Consider the body's network of effectors interacting with its world. When a body interacts with objects, like shaking a stick or picking up a ball, it induces dynamic change across this network. In other words, behavioral elements making up our performance (body motion, eye movements, etc.) are *not* independent of each other. They coordinate. So the degrees of freedom get constrained, we move in particular ways, and we can describe that behavior in simpler terms. This coordination can be described using fewer dimensions, ones that are compressed from our possible degrees of freedom. This is sometimes termed the *effective* degrees of freedom—the *combinations* of these behaviors we engage in rather than each separately (Turvey, 2007). This lower dimensional characterization is sometimes referred to as a "manifold." In ecological terms, such manifolds may be a source of information that we directly perceive (Turvey and Carello, 2011). This ecological perspective has also inspired models of coordination involving multiple people. These quantitative models show how oscillators capture perception-action dynamics coupling two people, such as in periodic movements (similar to walking together) or even quasi-periodic processes like conversational nonverbal behavior (Schmidt et al., 1990, 2014) and multi-person task coordination (Riley et al., 2011).

This tradition frames an explicit quantitative hypothesis that approaches the "degrees of freedom" problem noted above (cf. Bernstein, 1967). How do we manage all the complexity of interaction so systematically and smoothly? Ecological psychology suggests that the structure that inheres in various couplings—especially in perceiving and acting with another person—also generates a "manifold" over which communicative goals can be perceived and achieved. Below, we illustrate this idea by looking to nonverbal behavior and how it is coupled dynamically to traditionally "high-level" aspects of language and communication, like pragmatics and discourse. We argue below that the relevant fields traditionally studying these aspects of language and communication are poised to be further integrated. As noted above, new capacities for experimental measurements across many dimensions of communication, and algorithms to sort and analyze these measurements, create opportunities to find this "mutuality" across many aspects of language and communication. We will refer to this hypothesis as hunting for particular statistical patterns, which we'll call "pragmatic modes," a phrase we motivate in further discussion below.

Before we explicate our approach to multimodal coordination and discourse, we briefly sketch a history of primary competing models for language use. This review will serve as a contrast to the proposal here, and an illustration of key ingredients required for a theoretical approach to be rich enough to capture multimodal coordination.

2. Two theoretical themes in the language sciences

We begin by briefly contrasting two major theoretical traditions in the study of language, in particular "big picture" theories that frame the fundamental question "what is language?" The first of these could be called the "classical" approach, as it is often termed in cognitive science (Markman and Dietrich, 2000). Under this very influential approach, language as a multimodal object of study has not been the dominant agenda. For most of the early decades of modern cognitive science, approaches inspired by classical symbolic computing dominated the scientific study of language (Chomsky, 1957; Fodor, 1983). Researchers under this tradition focused their efforts on understanding the generativity of language and its importance for thought, while leaving aside its communicative function (Bechtel, 2001). Syntax, in particular, often occupies a special place in these and current studies of language (Hauser et al., 2002) alongside other "mental representations" such as properties of sounds and words and more (Pinker and Jackendoff, 2005). Internal representational structure is a powerful conceptual tool for describing the organization of thought and meaning comprehension.

From this point of view the many nonverbal signals that accompany language, though they may be critical for communication, are not central and are rarely brought to bear on linguistic theory (see discussion in Perniss, 2018; Rączaszek-

Leonardi, 2014). Under this tradition, language tends to be approached scientifically as an *individual* endeavor, with a primary function of organizing and making public the private representations inside one's mind through words (see discussion in Hauser et al., 2002; Chomsky, 2012; Everaert et al., 2017; cf. Rączaszek-Leonardi, 2014; Spivey and Richardson, 2008; Trueswell & Tananhaus, 2005).

A quite different understanding of language emerged after these early models. These alternative theories emphasize the importance of processes taking place not just within individual minds but also *between* them. Many of these theories preserve the representational commitments of the classical approach, but shift focus to their interactive nature. This approach can be illustrated through the influential work of Clark and collaborators, who reinterpreted several findings in the 1980s and early 1990s (e.g., Clark, 1996; 2003; Clark and Brennan, 1991). From this perspective, language participates in *joint* action, and its main purpose is to facilitate coordination between individuals to do things in the world together. It is not about the expression of internal representations, but about the development of *shared* representations that facilitate understanding and joint action (for review see Brennan et al., 2010). As a mainly social activity, language needs to be studied both at the level of the individual action and at the level of the joint product the individual actions make possible. At the level of the individual actions, interlocutors have a diverse set of communicative signals at their disposal: gestures, words, prosody, gaze, body orientation, interactions with objects in the shared environment, and even fillers play a role in communication. But these theories see such actions as *coordinated* in language use, at the level of a joint product.

The influential approach of Clark and others is just one of many that have taken quite a different turn from classical approaches. For example, the prominent domain of conversation analysis had a similar historical time course. The work of Sacks and Schegloff and others ushered in a closely related theoretical approach, under which language and communication are fundamentally embedded in the interactive context and sociocultural sphere (for an early review see Schegloff, 1991). Here too, much of the structure of actual language use depends on the goals and the evolving sequential patterns of interactions (Schegloff, 1990). For example, the domain of communicative repair studies how we recover from mishaps during interaction (such as misunderstandings), and has unveiled a variety of systematic strategies that interlocutors use to adapt to each other. To understand the words "um" or "uh," for example, it is not enough to simply describe the grammatical representations into which they figure, but also the social and cultural functional role of these seemingly simple segments (Clark and Fox Tree, 2002; and cross-culturally: Dingemanse et al., 2013).

Multimodality from this theoretical perspective is at the service of successfully communicating, and as such the goal has been to understand how people use diverse communicative signals to coordinate for joint action (e.g., Goodwin, 2018). Many in this general research tradition still use concepts of internal representations to describe language and communication—such as the intentions behind signals to accomplish specific communicative goals (Bratman, 1993; cf. Tollefsen and Dale, 2012). The level of commitment on such explanations varies. For example, conversation analysis may only imply a subtle representational commitment. This commitment comes from simply implying that there are internalized goals or meanings that are used during interaction. Despite this implicit commitment, conversation analysts see communication as much more dynamic and locally managed. In general, these perspectives in conversation analysis, psycholinguistics, dialogic and "dynamic" syntax (e.g., Du Bois, 2014; Kempson et al., 2016) move beyond the classical approach to language. They instead frame language as fundamentally interactive, while only implicitly preserving some representational qualities of the classical theory. Indeed, some in this tradition have begun to adopt an explicitly anti-representational agenda, seeing language and communication as rooted in coordinated *action* instead (e.g., Gregoromichelaki et al., 2020 and "actionism").

3. An ecological approach rooted in dynamics

So far, we have sketched a theoretical transition from classical approaches that emphasize grammatical processes exclusively in individuals, to alternative approaches that highlight the critical importance of social coordination. The classic approach is tightly tied to assumptions of an intentional and representational system that maps meanings onto behaviors; some in the more recent coordinated approach have begun to move away from such commitments.

We now return to the ecological approach that began this paper. This theoretical perspective for understanding language claims that the sort of coordination patterns that we observe in conversation are not simply the result of intentional choices of the speakers. Instead, these patterns can emerge from low-level interactions and are modulated by multiple situational constraints operating at many scales (Shockley et al., 2009; Wallot and Van Orden, 2011). These constraints include the limitations and capabilities of the individual brain and body (Dale et al., 2016; Shockley et al., 2009), what is possible under a given situation and goals of interaction (Chemero, 2009; Gibbs and Van Orden, 2012), the local culture and conventions of language use and the affordances they highlight (Heras-Escribano, 2019; Rączaszek-Leonardi and Kelso, 2008), the emergent behavioral patterns at the level of the dyad (Shockley et al., 2009), and more.

These are, in a sense, the observations from conversation analysis and discourse psycholinguistics and so on, described above: the critical importance of interactivity and multimodality in natural language use. But under the ecological lens, the goal is to uncover the couplings between interlocutors—to understand how interlocutors achieve this coherent meshwork. Notions of computation and representation might underperform here—we do not always "consciously compute" each gesture, each pitch modulation, each glance away, nor even each word that is either produced or comprehended (e.g., Galantucci et al., 2018). Indeed, considering the complexity of the multimodal character described above, it would seem intractable to do so. Under the ecological lens, coordination of this sort, such as complex multimodal performance of language, is seen as a self-organized phenomenon in which these constraints operate dynamically, and mostly without "awareness," to shape human interaction. These constraints convey the cognitive system from its very high number of potential degrees of behavioral freedom, to the linguistic and pragmatic choices in conversation, and back again, iteratively (Gibbs and Van Orden, 2012; Newtson, 1993; Wallot and Van Orden, 2011).

In this approach, linguistic elements such as words or grammatical frames are treated as another source of constraint, albeit more abstract than bodily and environmental constraints. Linguistic units act as constraints by virtue of their repeated pattern of use through learning, and by sustaining predictable effects over the dynamics of a system (cf. Elman, 2009). In that way, they are brought into the interaction to help regulate and constrain the dynamics of the other variables of the system during conversation in specific ways (Rączaszek-Leonardi and Kelso, 2008). Multimodality thus has to do with the configuration of language as an integrative process of brains, inside bodies that move, embedded within contexts, and that pursue particular goals in coordination with other brains and bodies around them.

While such a perspective may seem overly broad, one could argue that this wide relevance of language, from brains to bodies and external reference, is a hallmark feature demanding closer theoretical attention (Anderson, 2014; Dale et al., 2016). The question of multimodal coordination here has to do with successful action in such complex situations. Specifically, it has to do with the regulation of degrees of freedom by interlocutors. Consider the questions that could underlie one cognitive agent communicating with another: "When should I take the floor? What's the best way to describe this? Should I gesture this way or that? How long should I look at their eyes, how long away? Was that the right tone?" The stultifying effect of all such conscious assessments seems rather unpromising. And somehow, and rather fluidly in most contexts, interlocutors simply carry on without them. In "solving" this sort of problem, a classical approach would look quite different from one founded on ecological principles. Borrowing a visualization from Turvey (1990), we illustrate this in Fig. 1. On the left in Fig. 1, the notion that we have a set of centralized, abstract symbol manipulators governing these complex processes raises curious challenges. The contrast would be Fig. 1, right, in which we see various verbal and nonverbal behaviors constrain each other via local, incremental influence (as, indeed, recommended by some historical conversation analytic theory: Schegloff and Sacks, 1973; and recently: Kendrick et al., 2020).

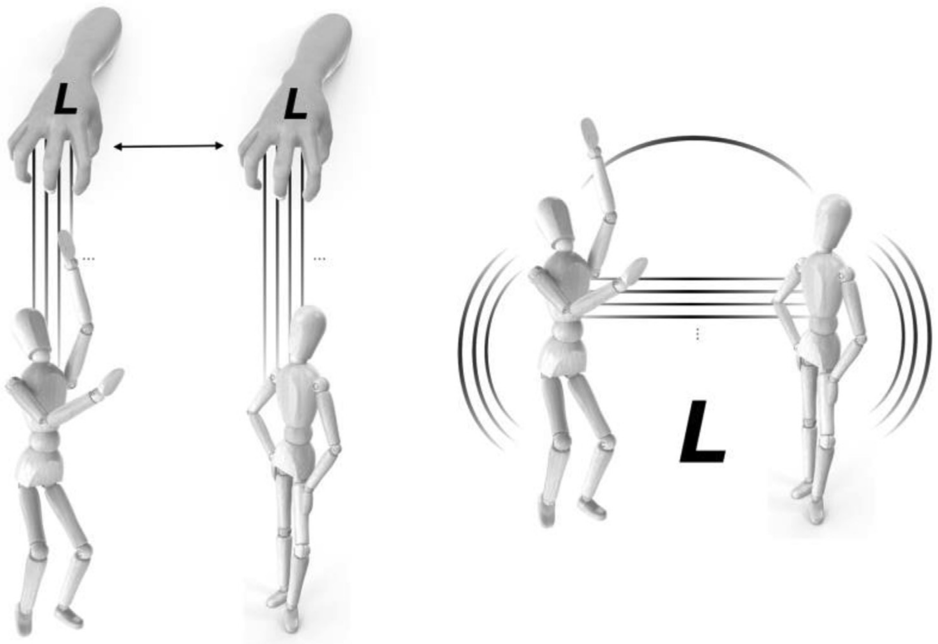

Fig. 1. Left: The classical approach to language sees each mind as harboring a set of abstract processes we refer to as "language"; these abstract processes, by some accounts centralized and encapsulated, independently control each effector and communication is achieved via reconstructing representations. Right: Coordination across modalities, verbal and nonverbal, underlie language. The ecological approach frames language as a coordinated activity in which there is a mutuality—coordination within each interlocutor's effectors, a network of constraints, but also extending across interlocutors (figure inspired by Turvey, 1990; cf. Shockley et al., 2009). Language is taken to be best defined as the outcome of this multi-person coordination.

An analogy may be fruitfully drawn with theories of motor control, as noted above (cf. Fusaroli et al., 2014). Most movements are highly complex, even seemingly simple movements. Multiple muscles are marshaled in the service of movement, sometimes with complex timings (d'Avella et al., 2003). A dominant approach in motor control is that specific muscles are not controlled *separately* but rather in *combination*. These combinations have been termed "synergies" (Bernstein, 1967; Turvey, 2007) or "coordinative structures" (Tuller and Turvey, 1982; Shockley et al., 2009). Multimodal coordination has been proposed to self-organize too into goal-dependent coordinative structures, both at the level of the individual and at the level of the dyad or group (Dale et al., 2013; Fusaroli and Tylén, 2012; Fusaroli et al., 2014). There is now methodological capacity to dynamically measure multimodal behaviors, and quantify these "structures." This is not the same as simply noting the importance of multimodality. It is instead quantifying the very multimodal patterns themselves. Again in the realm of conversation analysis, it may be fruitful to see such measurements as quantifying the theoretical concept of "unified composites" described by Enfield (2013, p. 65).

This ecologically inspired approach helps to integrate a few clear characteristics of language, bringing them to the forefront of its scientific study. These characteristics are the complex and dynamic nature of language. By recognizing the very complexity of language in context, and that its use offers a moving target among interlocutors, it alters our sense of what must underlie it. It highlights how language can be emergent and not necessarily representationally or intentionally driven (at least, not all the time). Importantly, the perspective explains how we can sustain smooth dynamics amidst such multimodal complexity: Each interlocutor in an interaction is shaped by varied constraints not just "inside" them, but also from the environment, which includes other interlocutors in these interactions (cf. Heras-Escribano, 2019; Dale et al., 2013).

4. Quantitative hypothesis: "Pragmatic Modes"

The work of ecologically inspired psychologists described above recommends a particular quantitative hypothesis. We remarked on the notion of multimodal "composites" or "combinations." This fundamental idea refers to a loosely assembled configuration of various states of individuals and groups acting together under a common a goal, and reassembling into new configurations under new constraints (cf. Kelso, 2009). This can be given explicit quantitative definition. Consider the work of d'Avella et al. (2003) on how frogs control limbs to carry out particular actions. By measuring muscle activity of the frog's limb structures, the researchers isolated particular combinations. These combinations are characteristic sequences of muscle actions that can then be combined under particular tasks or environmental conditions. The level of analysis can now become the combination rather than the individual muscles. We refer to these combinations as "modes," though they could also be called synergies (d'Avella et al., 2003) or coordinative structures (Tuller and Turvey, 1982). We will describe these as "modes" because it distinguishes this specific statistical hypothesis from the broader theories we have considered, for which these terms already have specialized uses.

In d'Avella et al. (2003) these modes are the basis for particular kicking actions, and appear to be "functionally assembled" for jumping, swimming, etc. The research to identify such modes has two parts. First, it uses dynamic, higher dimensional measurement, capturing how the physical parts of the performance are carried out individually. Second, it then isolates lower dimensional groupings by statistical techniques (in particular, dimensionality reduction). This strategy could be used in human language. After measuring a variety of behaviors simultaneously, the statistical patterns emerging from such an analysis may be at the "functional" level too, and so the "pragmatic mode" hypothesis is that the lower-order statistical character of multimodal performance would reveal these sorts of "functionally assembled" combinations. With such statistical methods and multimodal measurements, we may find the overall mesh of "mutuality" or "coupling" that the ecological approach generally predicts: a kind of coupling between interlocutors, their perception, action, and so on into functional combinations.

This idea is illustrated schematically in Fig. 2, again borrowing from Turvey's (1990) depiction of coordination. Measuring a variety of behaviors permits extraction of lower-dimensional circuits of constraints—characteristic patterns of behaviors that tend to travel together (we give examples from speech and gesture below). These "modes" are not necessarily discrete—in fact, as in d'Avella et al. (2003), they would likely be *statistical regularities*, inferred from the various behaviors measured (cf. Enfield and Sidnell, 2017). When these regularities unfold in conversation, they blend into each other (illustrated in B, at the bottom). We offer specific examples of this below with specific designs and measures.

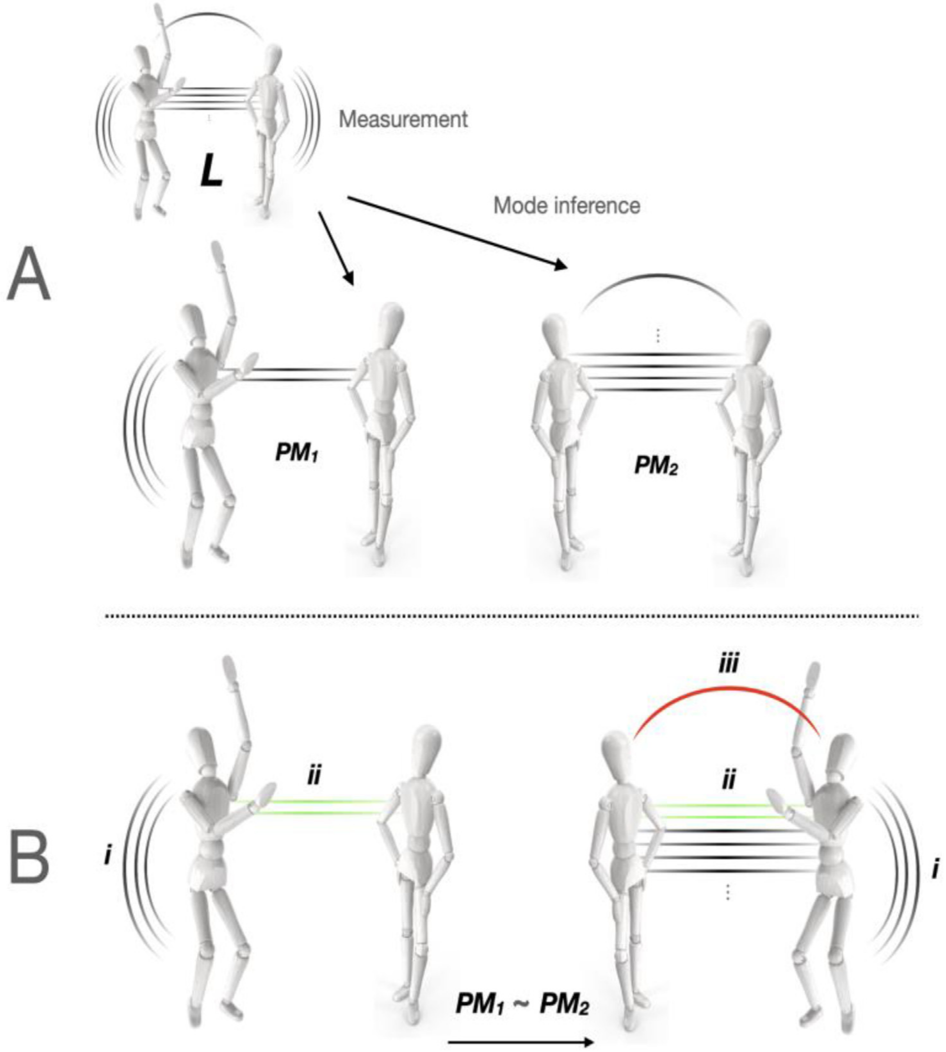

Fig. 2. A: Novel experimental and measurement techniques permit extraction of a high dimensional set of measurements. Using machine learning and other techniques, the pragmatic mode hypothesis is that natural linguistic performances are resolvable into dynamic "circuit"—temporary constraints that characterize different combinations of modalities (cf. d'Avella et al., 2003). B: Once obtained pragmatic modes (PM) may be explored for how they can model the dynamics of interaction, exploring the potential mutualities within and across nonverbal and verbal behaviors, such as speech-gesture synchrony (i), adaptiveness (ii) and reorganization around topics or other goals that emerge from the interaction (iii). These correspond to the subsections on speech-gesture below.

The precise timescale of these modes is, of course, an empirical question that should be studied both at the level of the dyad as a relational system embedded in a shared context, and at the level of the individual as embedded in an interaction with another. Perhaps the modes will primarily be at the level of what is called the adjacency pair—characteristic combinations of local moments in interaction (e.g., question-answer; Schegloff and Sacks, 1973). Alternatively, they may pertain to coarse, general pragmatics, such as giving an apology vs. sharing directions vs. giving a scientific talk. For instance, a first exploration of this last mode—delivering a talk—suggests some consistent multimodal patterns between the speaker's body movements, their prosodic inflections, and their use of PowerPoint slides that emerge as participation in this pragmatic context unfolds (Alviar et al., 2019). Nevertheless, the idea here is that "pragmatic modes" emerge from a collection of constraints at various levels, including the goals of conversation partners, and the constraints imposed by perception-action coupling. Pragmatic goals, for example, may organize speakers into temporary configurations like telling jokes, giving directions, showing compassion, and so on, just as phonological targets organize the jaw muscles in particular resilient configurations (c.f., Kelso et al., 1984). The pragmatic mode hypothesis supplements prior theories by seeing these as solutions to the degrees of freedom problem: Multimodal signals have rapid (and perhaps directly coupled) properties; we can rapidly adapt to each other during conversation because we transition between these combinations by forming and dissolving loosely coupled structures. We do not have to "compute" the positions of every interactive behavior independently because the behaviors are embedded in these configurations or synergies.

Adaptivity is common and seemingly rapid in interaction. These trends hint at the underlying synergies in various situations. Indeed, there is much evidence in multimodal research that interlocutors can be rapidly responsive to one another. Listeners laugh and join in when presented with the exaggerated gesture and prosody of humorous self-mockery, but rarely do it when speakers engage in the self-deprecating version (Yu, 2013). Listeners stop all other activities and attentively look to speakers when they engage in reenactments, playing along when they are used as characters of the reenacted situation (Sidnell, 2006). They politely wait and listen while speakers perform a thinking face and search for the next word, participating in the search only when directly or indirectly asked to do so (Bavelas and Chovil, 2018). When allowed to interact normally, interlocutors subtly coordinate their movements with each other during knock–knock jokes at all the time scales important for the interaction (Schmidt et al., 2014). They also coordinate their facial gestures, their manual gestures, their language use, and even their laughs while describing a route to a listener, and they may become even more synchronized the more difficult the task gets (Louwerse et al., 2012). These behavioral couplings may have neural implications. Interlocutors show synchronization in their neural oscillations while engaged in conversation (Perez et al., 2017; cf. Hasson et al., 2012).

This proposal that there may be pragmatic modes extractable statistically from multimodal signals has other empirical implications. Because these are probabilistic, dynamic processes–temporary combinations transitioning from one to the next—they imply that multimodal coordination is labile under different situations. Again, conversation analysis research offers illustration of this. In one example, Jensen (2018) offers a qualitative analysis of spontaneous humor in which he analyzes the coordinated verbal and non-verbal exchanges between interlocutors and shows how humor is jointly constructed by multimodal signaling that communicates the existence of humorous affordances and values in the situation. It is possible that this new mode of interacting emerges as people shift to playfulness, and humor looks different at the level of the multi-person system than the mode of interaction they were pursuing before. Perhaps now, prosody, gesture and gaze get less coupled intrapersonally and more coupled interpersonally, as to embody the shared understanding of the contradictions of humor. It is possible that this change is not obvious at the individual level (e.g., prosody: Flamson et al., 2011), but may instead be observable in the *dyadic* system.

5. A quantitative toolkit for pragmatic modes

Pursuing measures and analyses of the dyad as the basic unit is a step towards a quantitative approach to pragmatics that follows the intuitions of complex dynamic systems and ecological theories of language (cf. Favela, 2020; Fusaroli et al., 2014; Marsh et al., 2006; Schmidt et al., 1990; M. J. Richardson et al., 2015; Schilbach et al., 2013). There are existing quantitative proposals for analyzing interactive dynamics, and they range across a number of traditions and approaches (for reviews see Favela, 2020; M. J. Richardson et al., 2014; Paxton et al., 2016; Riley and Van Orden, 2005). We organize these possibilities into a kind of taxonomy for assessing "modes." Our use of the term "modes" is inspired by the very first branch of this taxonomy.

(i) **"*Dimensional reduction.*"** The use of this term "modes" is motivated partly by the growing literature in machine learning that seeks to recover the underlying properties of dynamical systems from rich measurements of their behavior (for some review see Dale and Bhat, 2018). For example, a relatively recent method is called "mode decompositions." These techniques are able to reconstruct a complex dynamic system from the raw data of its behavior (Rehman and Mandic, 2010; Tu et al., 2014). In Fig. 3, we refer to this process as "dimensional reduction," because the method takes a very large set of measurements—such as thousands or hundreds of thousands of pixels in a video—and recovers a lower-dimensional mathematical model for that exact system (Schmid, 2022). What is recovered in this reduction approach? In our case, it may involve gathering a large number of measurements from interaction (Fig. 3, top), then the dimensionality reduction methods look for shared dynamics across these components—which go together, and which lead into another (e.g., gesture, nodding, speech, etc.). These linkages can be thought of as a network or cluster, and we can give names to them (for convenience, we use "PM" for "pragmatic mode" in Fig. 3). Some methods suited to this kind of extraction are illustrated in Fig. 3 (e.g., Dale, 2015; Saltzman & Holt, 2014; Schmid, 2022; Wallot et al., 2016). These modes are now explicitly lower-dimensional characterizations of higher-dimensional data. In recent interpersonal research, it can be trivial to obtain rich dynamic data for these methods (see Cornejo et al., 2017 for some review). Much as complex, dynamical physical systems can have "modes" of behavior (particular flow patterns), so might human behavioral data yield similar insights.

(ii) ***"Dimensional expansion."*** At the exact opposite end of the spectrum is a family of methods that could be called "dimensionality expansion." These methods build up a more complex model of interaction by using statistics and machine learning on a relatively smaller amount of data, sometimes just a *single* time series. These methods including phase-space reconstruction, (cross-)recurrence quantification, complexity matching and other fractal methods and more (see Abney et al., 2018; Kello et al., 2010; M. J. Richardson et al., 2014; Paxton et al., 2016). These methods take densely sampled time series and then characterize the underlying composition of the "system" producing the data. An example is in Abney et al., (2021), who show that similarities in temporal complexity (the "fluctuations") in body motion can be compared across persons in an interaction, and can index the likelihood of success in their task. This is possible because the single time series contains variation that can be exploited by various methods, some of which we list in Fig. 3, bottom.

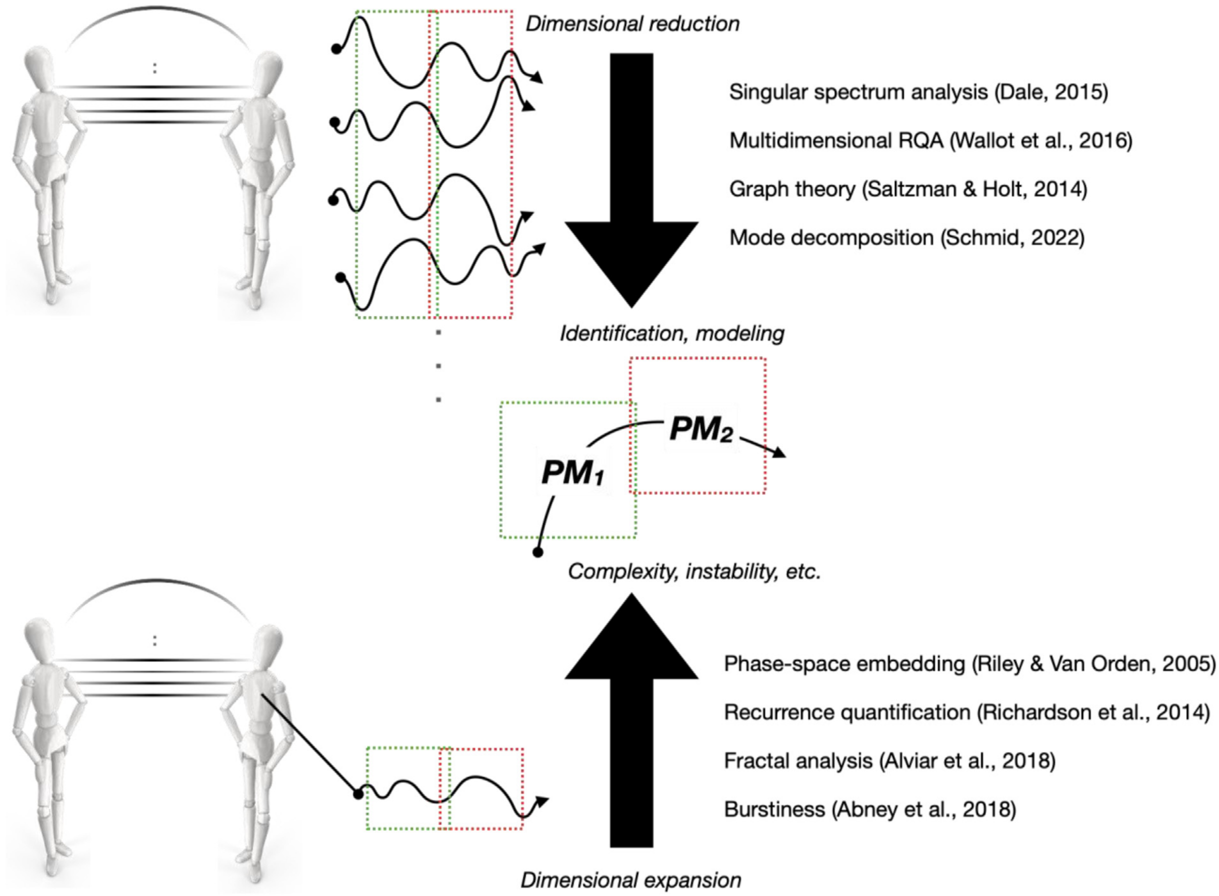

Fig. 3. Top: Methods designed to take large numbers of measurements (such as multimodal signals, video, etc.) and reduce them to an identified model; hypothetically illustrated here is the idea of taking many signals and finding underlying "modes" that characterize the sequencing among conversants (Kendrick et al., 2020). Bottom: It is also possible to describe the dynamic characteristics of interactions by recovering their properties through a single time series. These measures describe how the transitions from mode to mode may be more complex or ordered.

It is outside the scope of the present discussion, but both these types of method go beyond the data and "reconstruct" underlying regularities from them. In dimensional reduction methods, dimensionality is reduced into a more compressed form, identifying the underlying modes that are coordinating behaviors. In dimensional expansion methods, researchers can describe these modes, such as their complexity and how they change. It is possible to link these types of analysis in single studies, too (cf. Kello, 2013).

Above in Fig. 2, we describe these "modes" as dynamic, emerging gradually and fusing with others into their pragmatic contexts. So an important innovation in all these methods is to apply them in a sliding-window fashion (depicted at the center of Fig. 3). This means you conduct the same kind of analysis incrementally across a conversation (cf. Alviar et al., 2019). Thus it may be possible to assess how the modes *change* over interactive time (for simple examples, see Trulla et al., 1996 in a model system and Alviar et al., 2018 with human data). Together, these quantitative techniques may offer ways of identifying novel descriptions of the underlying constraints that we are calling "pragmatic modes." The hypothesis is that these statistical modes will be richly covarying with the goals in conversation, and like machine-learning techniques in other domains (e.g., Tu et al., 2014), may yield new quantitative descriptions of interaction we are referring to as "modes." Moreover, the statistical insights could relate to classic notions in the domains we have reviewed above, such as conversation analysis. Consider the theory of "social action" (Schegloff and Sacks, 1973), according to which one contribution to an interaction sets the stage for the next in incremental, sequential fashion. These structural ingredients have sequential and incremental properties that may be central to human interaction (Kendrick et al., 2020). Statistical analysis of diverse behavioral signals may yield those transitory modes that partners are negotiating, thus offering a new sort of quantitative picture of social action and sequence organization. Indeed, tongue-in-cheek monikers for the two main branches of this toolkit described above might be: *pragmatic mode decomposition* (dimensionality reduction) vs. *pragmatic mode recomposition* (dimensionality expansion).

6. Gesture and speech

So far, we've described the coordination of behavior during complex, natural language use. This coordination could be described quantitatively as a set of combinations or "modes" that are functionally assembled. However the evidence we review above offers just piecemeal support for the hypothesis. Though we have sketched a survey of relevant quantitative methods and some examples, are there domains in the study of communication that best support the idea that behaviors become coordinated into adaptive units under particular pragmatic goals or contexts? In those domains, do segments of behavior act like "modes," coupling together temporarily and adapting under pragmatic conditions?

In this section, we review perhaps the best example of this sort of mutuality between behavioral signals: how gesture and speech are richly integrated, constrained by each other and by the individuals and contexts in an interaction (cf. Rasenberg et al., 2020). This is a large literature, and so naturally we cannot review it in its entirety. We focus on work that illustrates quantitative and experimental approaches. Several facts emerge from this literature, consistent with the hypothesis we describe: (i) the importance of coupling between gesture and speech of one person (synchrony), (ii) the functional (semantic) value of gestures, and (iii) the coupling between one's gestures and those of one's communication partners (interpersonal coupling). This review suggests that the gesture-speech relationship serves as a robust illustration of the interdependence and adaptivity among behavioral signals that our proposal implies. In other words, these basic facts suggest gesture and speech (and more) are operating as a kind of unit that adapts under different communicative pressures.

(i) ***Synchrony of Gesture and Speech.*** The first robust fact about the relationship between speech and gesture has to do with their remarkable synchronization. At a very basic level, the movements of the vocal apparatus that produce speech, like the area of the opening of the mouth, directly relate to the amplitude of the speech signal (Chandrasekaran et al., 2009). At a more general level, iconic gestures tend to occur at the times that are relevant for the content of the uttered sentence (Özyürek et al., 2005), and beat gestures tend to align with the peak amplitudes of the speech signal (Pouw et al., 2019; see also Pouw et al., 2020). Importantly, such synchronization seems to be necessary for the successful integration of the two signals (Habets et al., 2011; Obermeier et al., 2012). Iverson and Thelen have colorfully referred to speech and gesture as "two fins" of a fish, coordinating carefully in order to progress (Iverson and Thelen, 1999). Desynchronizing speech and gesture has consequences for fluency. Numerous studies have shown this in various ways, primarily through disrupting the timing between speech and gesture experimentally (e.g., Chu and Hagoort, 2014; Pouw and Dixon, 2019). Intriguingly, research using neuroscience methods like event-related potentials offers further evidence for this tight timing between speech and gesture (Habets et al., 2011; Obermeier et al., 2012). Though this research derives from outside ecological psychology, it is nevertheless consistent with what we described above regarding the "combinations" of various modalities of communication. The collection of papers in De Wit and Withagen (2019) offers discussion of ecological approaches to neuroscience, in which "Gibsonian" perspectives on neuroscience are developed, leading to differing positions contrasted in their introduction. For example neural processes may index perception-action couplings between persons, and thus may serve as a strong source of information about these interpersonal processes.

(ii) ***The Semantic Role of Gestures.*** A second fact that has become evident about the gesture-speech relationship is their complementarity and mutual influence with the semantic interpretation of a message. Listeners expect gesture to inform the content of speech and be congruent with it. In fact, when presented with non-matching speech-gesture pairs (e.g., hearing "chop" and seeing "twist") participants become slower and less accurate to recognize the pair as matching an action prime (Kelly et al., 2010). Like speech-gesture synchrony, EEG and brain-imaging studies have also produce findings consistent with this congruency expectation (Habets et al., 2011). Indeed some results suggest that there is overlap in brain areas participating in speech and gesture together (Özyürek, 2014; Willems et al., 2006). This area of research also shows that gesture can facilitate interpretation of difficult language (Holle et al., 2012), perhaps even to the point of directly modulating the interpretation of words as they are being heard (Biau et al., 2015).

(iii) ***Sensitivity to Functional Constraints.*** A third fact of the gesture-speech relationship is its sensitivity to diverse functional constraints that modulate the gestures produced. Consider, for example, the language being spoken. The preferred syntactic encoding of a language seems to be one such functional constraint. Consider some cross-linguistic comparison of Turkish speakers, whose grammar requires the expression of path and manner of movement in separate clauses (e.g., "the ball descended the hill while rolling"), vs. English speakers, whose grammar allows the joint expression of the two characteristics in the same clause (e.g., "the ball rolled down the hill"). When asked to describe the movement of cartoon figures in an animation, speakers of these two languages showed differences in their gesturing patterns that were in line with their language's preferred way of linguistically encoding the movement. English speakers tended to produce compound gestures showing both the path and manner of movement at the same time, whereas Turkish speakers tended to produce gestures that depicted just one of the two characteristics but not both (Özyürek et al., 2005).

Another functional factor of importance is the intention to communicate. Research suggests that gestures help the speaker to assemble their message, but also that gestures are modulated in the presence of an addressee to improve communication (see Alibali et al., 2001; Galati and Brennan, 2014; Holler et al., 2013; Krauss et al., 2000; Trujillo et al., 2018; Wagner and Bryhadyr, 2017). Similarly, the dynamic structure in the movement amplitude and the sound amplitude of a performance is more tightly correlated in situations in which the movements are more informative for the message (e.g., spontaneous speech) than in situations in which the movements are responding more strongly to stylistic (e.g., poetry) or non-communicative constraints (e.g., classical music; Alviar et al., 2020). And factors that threaten communication, such as auditory noise, result in changes to the kinematics of facial and manual gestures (Trujillo et al., 2021).

7. Summary, conclusion and future directions

The responsiveness of gesture-speech combinations to such a diversity of factors like the syntactic structure of language, the location and knowledge of the addressees, the ambiguities in the information, and the communicative nature of a situation, is a feature that is hard to explain in “classical” models. All this information would have to be specified in advance in the form of conditions that mandate the selection of specific collections of rules inside a contextual module that would later serve as input for the operation of the other modules to make their outputs sensitive to the context (cf. Onnis and Spivey, 2012). This may be *possible* cognitively, but it would not be a very parsimonious model (Casasanto and Lupyan, 2015). On the other hand, the dynamical proposal includes at its core the idea of constraints, both internal and external, individual and collective, that give rise to different emergent coordinative structures and shape systems differently depending on their specific combinations over time. The speed with which we decide and coordinate so quickly can be explained more easily when we do not take these processes as computationally modular, but rather ecologically integrated in real time (Wallot and Van Orden, 2012). From this theoretical point of view, speaker and addressee could also be understood as another emergent system with coordinative structures of its own (Dale and Spivey, 2019; Schmidt et al., 1990; Sebanz et al., 2006; Shockley et al., 2009). This system encompasses the *two* individuals, and its reorganization happens not for the speaker or for the listener uniquely, but for the two-person system’s functional stability and coordinated response to changing constraints.

We began this paper by putting the multimodal complexity of natural language performance into stark relief with prior theoretical traditions. We argued that an ecological approach that integrates an individual language user with her conversation partners and their mutual environment is critical to understanding *how* such multimodal complexity is possible. Using speech-gesture coordination as an illustration, we showed that these modalities subtly constrain each other, involving constraints from both listeners and speakers, and can be influenced by the context and goals of an interaction. The case of speech and gesture offers a specific instance of the proposed “pragmatic modes” hypothesis as a statistical feature of language use that emerges as a result of multimodal coordination.

We also offered a quantitative taxonomy of methods that may help to identify these statistical patterns (cf. Rasenberg et al., 2020). It may be possible to leverage stable statistical characteristics to gain insight into general trends that occur in interactions, or ones that may vary across individuals or cultures (cf. Dingemanse et al., 2013). Will there be modes characteristic of “irony” or “conflict” and so on that could inform these domains of discourse? Conversely, these statistical trends may be highly variable so much that the linkage between qualitative and quantitative remains crucial. In the absence of completely lawful statistical behavior measured with these methods, qualitative methods can help to contextualize and abstractly characterize the approximate output of such a quantitative exercise. It echoes a classic notion influential in the philosophy of language, originating perhaps in the later Wittgenstein (1953). Pragmatic modes may exhibit a “family resemblance” across their combinations, not to be strictly defined by necessary and sufficient conditions (conditions which, incidentally, are characteristic of classic cognitive approaches too and perhaps Wittgenstein’s earlier *Tractatus*). A later Wittgensteinian frame might encourage us to fuse the qualitative and quantitative to understand the more complex interconnections among types of discourse as “language games” in our “form of life.”

The pragmatic mode hypothesis is inspired by the kind of perception-action and organism-environment coupling that ecological psychology, writ large, urges we should make central in our theories (Chemero, 2009; Szokolszky et al., 2019). The influence of multimodal structure may relate in some ways to the kind of specification that the optic array supplies to organisms that engage their environment. The organism in the midst of this visual array picks up affordances for engagement with the world. Beyond *just* the optic array, here we also have multimodal “arrays” that are in flux, constraining each person in the dialogue and only temporarily establishing conditions for dialogic structure (some prominent recent discussion on language has called this the “now or never bottleneck”: Christiansen and Chater, 2016). Consider some recognizable examples: looking down at a watch, turning up one side of one’s mouth, furrowing the brow, a long pause while looking at the ground, etc. An intriguing hypothesis is that members of successful conversational systems may *directly perceive* the significance of these moments *as* members of these conversational systems (see Valenti and Gold, 1991 for early conceptualizations of this idea, and Gallagher, 2020 for contemporary accounts). The “mesh” of multimodal structure may be a kind of visual (haptic, auditory, etc.) array that forms the Gibsonian-like structure of conversational synergies (see again Fig. 1).

There are obvious limitations to our proposal here. More work is needed to develop the concept of a pragmatic mode to the point of generating testable, informative hypotheses. There are some explicit designs in place for something similar. We offered a toolkit to think about this issue in the above section, but we also offered some existing examples. As noted earlier, research in multisegment motor control and machine learning illustrates the promise of studying the “statistical composition” of complex action (d’Avella et al., 2003). In addition, some explicit formal proposals are on hand already to help articulate this network of multimodal influences, such as in the work of Saltzman and colleagues (Saltzman and Munhall, 1992; Saltzman and Holt, 2014). In the framework of Saltzman and Munhall (1992) for example (updated in Saltzman and Holt, 2014), complex human behaviors are likened to interconnected networks or graphs. For our purposes, these graphs would be connected "nodes" that represent behavior, such as interconnected patterns of hand movements, arm movements, eye movements and so on. These graphs can also be dynamic because the connectivity on them can change depending on the "mode." So researchers can articulate several layers of analysis of this behavior, from the "activation" of nodes themselves (e.g., activation of some gesture) to the relationships among the nodes—the connectivities can change (e.g., hand, arm and head movement during "giving directions"). Although Saltzman, Munhall and others formulate their ideas in very general motoric terms, their formalism and other graph-theoretic approaches hold promise in this respect (see also Dale, 2015; Dale and Bhat, 2018; section 6 in Dale et al., 2013).

There are rich illustrations of multimodal measurements and analysis in human interaction itself to showcase these methods (Fusaroli et al., 2016; Cornejo et al., 2017; Dale, 2015; Duran et al., 2019). Nevertheless, we would argue that the examples of gesture and speech serve as very strong illustrations of these concepts, even though "pragmatic mode" is introduced as a new frame for this quantitative strategy. A proposal of this kind would be successful to the extent that it recommends new empirical research. Here we conclude by offering one such example.

Consider a potential test of these ideas in the real-time dynamics of conversations and their disruption. One prediction that follows from the idea of pragmatic modes has to do with overhearers or interlocutors that join a conversation at a later point, after its "coordinative structures" have already been established by other conversation partners. As external to the system, the new interlocutors should show a different dynamic regime than the other two, and they should show variability while they reorganize themselves to participate in the existing system. The original dyadic system should also show signatures of reorganization as members adapt to the new interlocutor and combine modalities in a way that supports the new constraints that come with it. Articulating these dynamics across multiple modalities under these contexts would serve two exciting theoretical purposes. First, measured dynamics would specify the manner in which varied modalities and behavioral patterns are assembled uniquely, and how they flexibly adapt. And second, they would shed more light on the nature of the underlying control processes of language and communication themselves. They would also again serve as a quantitative bridge between the work of conversation analysis who have studied "schisming" (Egbert, 1997), conversational disruptions and general disruptions or interference in dialogue (e.g., Dingemanse et al., 2015) or even pragmatic shifts in conversation, such as telling a joke (Stivers, 2021). Such measures may also contribute dynamic quantification of what Stivers and Rossano (2010) have called the "mobilizing response"—how interlocutors induce particular responses from their partners under particular conversational conditions. The hypothesis of "pragmatic modes," though, would link these observations from conversation analysis to the quantitative principles of ecological psychology, and associated methods in the study of complex dynamical systems (Chemero, 2009; Richardson et al., 2014).

Human language is a multimodal phenomenon that has at its core the coordination of dyadic and group actions in complex environments. In this paper, we have argued that there is much evidence that the relationship among signals like gestures, prosody, words, and gaze is not a simple and rigid one as would be predicted from the "classical" models of multimodal language. Speakers adapt their multimodal behaviors to the needs of interlocutors, the changing constraints of the environment, the goal and constraints of the tasks, and reorganize their behavior to keep the system stable when faced with perturbations that jeopardize the successful communication of the message (cf. inter alia: Brennan et al., 2010; Fusaroli et al., 2014; Newtson, 1993; Sidnell, 2006). Such adaptability of behavior seems to be more in line with the ideas of ecological psychology and associated methods in complex dynamical systems: interactions are processes of self-organization, where the various temporary "pragmatic modes" that are assembled bring about successful communication despite the moving targets of conversation partners and contexts. As illustrated in the review of speech-gesture dynamics, speakers also use modalities differently depending on the pragmatic goals pursued and discursive moves used in the conversation. We propose that it may be time to leverage rich data and quantification of this sort to show statistical structure more generally—from perception and action to pragmatic goals across as broad an array of conversational phenomena as our measurements take us.

Declaration of interest

The authors declare no conflict of interests.

Aknowledgements

An earlier version of this paper was written as part of the first author's dissertation and was supported by funding from the Cognitive and Information Sciences Department at UC Merced.

References

Abney, D., Dale, R., Louwerse, M., Kello, C., 2018. The bursts and lulls of multimodal interaction: Temporal distributions of behavior reveal differences between verbal and non-verbal communication. Cogn. Sci. 42 (4), 1297–1316.
Abney, D., Paxton, A., Dale, R., Kello, C., 2021. Cooperation in sound and motion: Complexity matching in collaborative interaction. J. Exp. Psychol. Gen.
Alibali, M.W., Heath, D.C., Myers, H.J., 2001. Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen. J. Mem. Lang. 44 (2), 169–188.

Alviar, C., Dale, R., Dewitt, A., Kello, C.T., 2020. Multimodal coordination of sound and movement in solo music and speech. Discourse Process 57 (8), 682–702. https://doi.org/10.1080/0163853X.2020.1768500.
Alviar, C., Dale, R., Kello, C.T., 2018. The fractal structure of extended communicative performance. In: Rogers, T.T., Rau, M., Zhu, X., Kalish, C.W. (Eds.), Proceedings of the 40th Annual Meeting of the Cognitive Science Society. Cognitive Science Society, Austin, TX, pp. 1292–1297.
Alviar, C., Dale, R., Galati, A., 2019. Complex communication dynamics: exploring the structure of an academic talk. Cognit. Sci. 43 (3), e12718.
Anderson, M.L., 2014. After Phrenology: Neural Reuse and the Interactive Brain. MIT Press.
Bavelas, J., Chovil, N., 2018. Some pragmatic functions of conversational facial gestures. Gesture 17 (1), 98–127.
Bavelas, J.B., Black, A., Lemery, C.R., Mullett, J., 1986. I show how you feel": motor mimicry as a communicative act. J. Pers. Soc. Psychol. 50 (2), 322–329.
Bechtel, W., 2001. Linking Cognition and Brain: The Cognitive Neuroscience of Language. *Philosophy and the Neurosciences: A Reader*. Basil Blackwell, Oxford.
Bernstein, N.A., 1967. The Coordination and Regulation of Movements. Pergamon Press, Oxford, UK.
Biau, E., Torralba, M., Fuentemilla, L., de Diego Balaguer, R., Soto-Faraco, S., 2015. Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. Cortex 68, 76–85.
Branigan, H.P., Pickering, M.J., Cleland, A.A., 2000. Syntactic co-ordination in dialogue. Cognition 75 (2), B13–B25.
Bratman, M.E., 1993. Shared intention. Ethics 104, 97–113.
Brennan, S.E., Clark, H.H., 1996. Conceptual pacts and lexical choice in conversation. J. Exp. Psychol. Learn. Mem. Cognit. 22 (6), 1482.
Brennan, S.E., Galati, A., Kuhlen, A.K., 2010. Two minds, one dialog: coordinating speaking and understanding. In: Psychology of Learning and Motivation, 53. Academic Press, pp. 301–344.
Buder, E.H., Warlaumont, A.S., Oller, D.K., Chorna, L.B., 2010. Dynamic indicators of mother-infant prosodic and illocutionary coordination. In: Speech Prosody 2010-Fifth International Conference.
Casasanto, D., Lupyan, G., 2015. All Concepts are Ad Hoc Concepts. In: The Conceptual Mind: New Directions in the Study of Concepts, pp. 543–566.
Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., Ghazanfar, A.A., 2009. The natural statistics of audiovisual speech. PLoS Comput. Biol. 5 (7), e1000436.
Chemero, A., 2009. Radical Embodied Cognitive Science. MIT press.
Chomsky, N., 1957. Syntactic Structures. Mouton & Co.
Chomsky, N., 2012. On the nature, use and acquisition of language. In: Language and Meaning in Cognitive Science. Routledge, pp. 13–32.
Christiansen, M.H., Chater, N., 2016. The now-or-never bottleneck: A fundamental constraint on language. Behav. Brain Sci. 39.
Chu, M., Hagoort, P., 2014. Synchronization of speech and gesture: evidence for interaction in action. J. Exp. Psychol. Gen. 143 (4), 1726.
Clark, H.H., 1996. Using Language. Cambridge University Press, Cambridge.
Clark, H.H., 2003. Pointing and placing. In: Pointing. Psychology Press, pp. 251–276.
Clark, H.H., Brennan, S.E., 1991. Grounding in Communication.
Clark, H.H., Fox Tree, J.E., 2002. Using uh and um in spontaneous speaking. Cognition 84 (1), 73–111.
Cornejo, C., Cuadros, Z., Morales, R., Paredes, J., 2017. Interpersonal coordination: methods, achievements, and challenges. Front. Psychol. 8, 1685.
Dale, R., 2015. An integrative research strategy for exploring synergies in natural language performance. Ecol. Psychol. 27 (3), 190–201.
Dale, R., Bhat, H.S., 2018. Equations of mind: data science for inferring nonlinear dynamics of socio-cognitive systems. Cognit. Syst. Res. 52, 275–290.
Dale, R., Fusaroli, R., Duran, N.D., Richardson, D.C., 2013. The self-organization of human interaction. In: Ross, B. (Ed.), Psychology of Learning and Motivation, 59. Academic Press, pp. 43–95.
Dale, R., Kello, C.T., Schoenemann, P.T., 2016. Seeking synthesis: the integrative problem in understanding language and its evolution. Top. Cognit. Sci. 8 (2), 371–381.
Dale, R., Spivey, M.J., 2019. Weaving oneself into others. In: Eye-tracking in Interaction: Studies on the role of eye gaze in dialogue, vol. 10, p. 67.
d'Avella, A., Saltiel, P., Bizzi, E., 2003. Combinations of muscle synergies in the construction of a natural motor behavior. Nat. Neurosci. 6 (3), 300–308.
Dent-Read, C.E., Zukow-Goldring, P.E. (Eds.), 1997. Evolving Explanations of Development: Ecological Approaches to Organism–Environment Systems. American Psychological Association, Washington DC.
de Wit, M.M., Withagen, R., 2019. What should a "Gibsonian neuroscience" look like? Introduction to the special issue. Ecol. Psychol. 31 (3), 147–151.
Di Paolo, Cuffari, De Jaegher, 2018. Linguistic Bodies: The Continuity between Life and Language. MIT Press.
Dingemanse, M., Roberts, S.G., Baranova, J., Blythe, J., Drew, P., Floyd, S., et al., 2015. Universal principles in the repair of communication problems. PLoS One 10 (9), e0136100.
Dingemanse, M., Torreira, F., Enfield, N.J., 2013. Is "Huh?" a universal word? Conversational infrastructure and the convergent evolution of linguistic items. PLoS One 8 (11), e78273.
Du Bois, J.W., 2014. Towards a dialogic syntax. Cognit. Ling. 25 (3), 359–410. https://doi.org/10.1515/cog-2014-0024.
Duran, N.D., Paxton, A., Fusaroli, R., 2019. ALIGN: analyzing linguistic interactions with generalizable techNiques—a Python library. Psychol. Methods 24 (4), 419.
Egbert, M., 1997. Schisming: the collaborative transformation from a single conversation to multiple conversations. Res. Lang. Soc. Interact. 30 (1), 1–51.
Elman, J.L., 2009. On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. Cognit. Sci. 33 (4), 547–582.
Enfield, N.J., 2013. Relationship Thinking: Agency, Enchrony, and Human Sociality. Oxford University Press.
Enfield, N.J., Sidnell, J., 2017. On the concept of action in the study of interaction. Discourse Stud. 19 (5), 515–535.
Everaert, M.B., Huybregts, M.A., Berwick, R.C., Chomsky, N., Tattersall, I., Moro, A., Bolhuis, J.J., 2017. What is language and how could it have evolved? Trends Cognit. Sci. 21 (8), 569–571.
Favela, L.H., 2020. Cognitive science as complexity science. Wiley Interdiscipl. Rev. Cognit. Sci. 11 (4), e1525.
Flamson, T., Bryant, G.A., Barrett, H.C., 2011. Prosody in spontaneous humor: evidence for encryption. Pragmat. Cognit. 19 (2), 248–267.
Fodor, J.A., 1983. The Modularity of Mind. MIT press.
Fusaroli, R., Rączaszek-Leonardi, J., Tylén, K., 2014. Dialog as interpersonal synergy. New Ideas Psychol. 32, 147–157.
Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., Tylén, K., 2012. Coming to terms: quantifying the benefits of linguistic coordination. Psychol. Sci. 23 (8), 931–939.
Fusaroli, Riccardo, Tylén, Kristian, 2012. Carving language for social coordination: a dynamical approach. Interact. Stud. 13 (1), 103–124. https://doi.org/10.1075/is.13.1.07fus.
Fusaroli, R., Bjørndahl, J.S., Roepstorff, A., Tylén, K., 2016. A heart for interaction: shared physiological dynamics and behavioral coordination in a collective, creative construction task. J. Exp. Psychol. Hum. Percept. Perform. 42 (9), 1297.
Galantucci, B., Roberts, G., Langstein, B., 2018. Content deafness: when coherent talk just doesn't matter. Lang. Commun. 61, 29–34.
Galati, A., Brennan, S.E., 2014. Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture. Lang. Cognit. Neurosci. 29 (4), 435–451.
Gallagher, S., 2020. Action and Interaction. Oxford University Press.
Garrod, S., Pickering, M.J., 2004. Why is conversation so easy? Trends Cognit. Sci. 8 (1), 8–11.
Gibbs, R.W., Van Orden, G., 2012. Pragmatic choice in conversation. Top. Cognit. Sci. 4 (1), 7–20.
Gibson, J.J., 1966. The Senses Considered as Perceptual Systems. Houghton Mifflin, Boston, MA.
Gibson, J.J., 1979. The Ecological Approach to Visual Perception. Houghton Mifflin, Boston, MA.
Giles, H. (Ed.), 2016. Communication Accommodation Theory: Negotiating Personal Relationships and Social Identities across Contexts. Cambridge University Press.
Goodwin, C., 1979. The interactive construction of a sentence in natural conversation. Everyday language: Studies in ethnomethodology. Irvington Publishers, New York, pp. 97–121.

Goodwin, C., 2018. Co-operative Action. Cambridge University Press.
Goodwin, C., Heritage, J., 1990. Conversation analysis. Annu. Rev. Anthropol. 19 (1), 283–307.
Grammer, K., Kruck, K.B., Magnusson, M.S., 1998. The courtship dance: patterns of nonverbal synchronization in opposite-sex encounters. J. Nonverbal Behav. 22 (1), 3–29.
Gregoromichelaki, E., Kempson, R., Howes, C., 2020. Actionism in syntax and semantics. Dialogue and perception-extended papers from DaP2018. CLASP Pap. Computat. Linguist. 2, 12–27.
Grosz, B., Hirschberg, J., 1992. Some intonational characteristics of discourse structure. In: Ohala, J. (Ed.), Proceedings of the International Conference on Spoken Language Processing. Banff, Alberta, Canada, pp. 429–432, 1992.
Grosz, B., Sidner, C.L., 1986. Attention, intentions, and the structure of discourse. Comput. Ling. 12 (3), 175–204, 1986.
Habets, B., Kita, S., Shao, Z., Özyürek, A., Hagoort, P., 2011. The role of synchrony and ambiguity in speech–gesture integration during comprehension. J. Cognit. Neurosci. 23 (8), 1845–1854.
Hasson, U., Ghazanfar, A.A., Galantucci, B., Garrod, S., Keysers, C., 2012. Brain-to-brain coupling: a mechanism for creating and sharing a social world. Trends Cognit. Sci. 16 (2), 114–121.
Hauser, M.D., Chomsky, N., Fitch, W.T., 2002. The faculty of language: what is it, who has it, and how did it evolve? Science 298 (5598), 1569–1579.
Healey, P.G., Purver, M., Howes, C., 2014. Divergence in dialogue. PLoS One 9 (6), e98598.
Heras-Escribano, M., 2019. The Philosophy of Affordances. Palgrave Macmillan, Cham.
Hodges, B.H., Fowler, C.A., 2015. Fields, waves, and particles: finding common ground in understanding language as a public activity. Ecol. Psychol. 27 (3), 175–189.
Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A.D., Ward, J., Gunter, T.C., 2012. Gesture facilitates the syntactic analysis of speech. Front. Psychol. 3, 74.
Holler, J., Turner, K., Varcianna, T., 2013. It's on the tip of my fingers: Co-speech gestures during lexical retrieval in different social contexts. Lang. Cognit. Process. 28 (10), 1509–1518.
Iverson, J.M., Thelen, E., 1999. Hand, mouth and brain. The dynamic emergence of speech and gesture. J. Conscious. Stud. 6 (11–12), 19–40.
Jensen, T.W., 2018. Humor as interactional affordances: an ecological perspective on humor in social interaction. Psychol. Lang. Commun. 22, 238–259. https://doi.org/10.2478/plc-2018-0010.
Jermann, P., Nüssli, M.A., 2012. Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, pp. 1125–1134.
Kello, C.T., Brown, G.D., Ferrer-i-Cancho, R., Holden, J.G., Linkenkaer-Hansen, K., Rhodes, T., Van Orden, G.C., 2010. Scaling laws in cognitive sciences. Trends Cognit. Sci. 14 (5), 223–232.
Kello, C.T., 2013. Critical branching neural networks. Psychol. Rev. 120 (1), 230.
Kelly, S.D., Özyürek, A., Maris, E., 2010. Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. Psychol. Sci. 21 (2), 260–267.
Kelso, J.S., 2009. Coordination dynamics. In: Meyers, R.A. (Ed.), Encyclopedia of Complexity and Systems Science. Springer, New York, pp. 1–41.
Kelso, J.S., Tuller, B., Vatikiotis-Bateson, E., Fowler, C.A., 1984. Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. J. Exp. Psychol. Hum. Percept. Perform. 10 (6), 812.
Kempson, R., Cann, R., Gregoromichelaki, E., Chatzikyriakidis, S., 2016. language as mechanisms for interaction. Theor. Ling. 42 (3–4), 203–276. https://doi.org/10.1515/tl-2016-0011. October 2016.
Kendrick, K.H., Brown, P., Dingemanse, M., Floyd, S., Gipper, S., Hayano, K., et al., 2020. Sequence organization: a universal infrastructure for social action. J. Pragmat. 168, 119–138.
Krauss, R., Chen, Y., Gottesman, R., 2000. Lexical gestures and lexical access: a process model. In: McNeill, D. (Ed.), Language and Gesture. Cambridge University Press, Cambridge, pp. 261–283. https://doi.org/10.1017/CBO9780511620850.017.
Levitan, R., Hirschberg, J., 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In: Twelfth Annual Conference of the International Speech Communication Association.
Linell, P., 2004. The Written Language Bias in Linguistics: Its Nature, Origins and Transformations. Routledge.
Louwerse, M.M., Dale, R., Bard, E.G., Jeuniaux, P., 2012. Behavior matching in multimodal communication is synchronized. Cognit. Sci. 36 (8), 1404–1426.
Markman, A.B., Dietrich, E., 2000. Extending the classical view of representation. Trends Cognit. Sci. 4 (12), 470–475.
Marsh, K.L., Richardson, M.J., Baron, R.M., Schmidt, R.C., 2006. Contrasting approaches to perceiving and acting with others. Ecol. Psychol. 18 (1), 1–38.
Mondada, L., 2016. Challenges of multimodality: language and the body in social interaction. J. SocioLinguistics 20, 336–366. https://doi.org/10.1111/josl.1_12177.
Newtson, D., 1993. The dynamics of action and interaction. In: Smith, L.B., Thelen, E. (Eds.), A Dynamic Systems Approach to Development: Applications. The MIT Press, Cambridge, MA, US, pp. 241–264.
Nomikou, I., Leonardi, G., Rohlfing, K.J., Rączaszek-Leonardi, J., 2016. Constructing interaction: the development of gaze dynamics. Infant Child Dev. 25 (3), 277–295.
Obermeier, C., Dolk, T., Gunter, T.C., 2012. The benefit of gestures during communication: evidence from hearing and hearing-impaired individuals. Cortex 48 (7), 857–870.
Oller, D.K., Buder, E.H., Ramsdell, H.L., Warlaumont, A.S., Chorna, L., Bakeman, R., 2013. Functional flexibility of infant vocalization and the emergence of language. Proc. Natl. Acad. Sci. USA 110 (16), 6318–6323.
Onnis, L., Spivey, M.J., 2012. Toward a new scientific visualization for the language sciences. Information 3 (1), 124–150.
Özyürek, A., 2014. Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. Philos. Trans. R. Soc. B 369 (1651), 20130296.
Özyürek, A., Kita, S., Allen, S., Furman, R., Brown, A., 2005. How does linguistic framing of events influence co-speech gestures?: insights from cross-linguistic variations and similarities. Gesture 5 (1), 219–240.
Paxton, A., Dale, R., Richardson, D., 2016. Social coordination of verbal and non-verbal behaviours. Interpersonal Coordination and Performance in Social Systems. Routledge, pp. 277–292.
Paxton, A., Roche, J.M., Ibarra, A., Tanenhaus, M.K., 2021. Predictions of miscommunication in verbal communication during collaborative joint action. J. Speech Lang. Hear. Res. 64 (2), 613–627.
Perez, A., Carreiras, M., Duñabeitia, J.A., 2017. Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. Sci. Rep. 7 (1), 4190.
Perniss, P., 2018. Why we should study multimodal language. Front. Psychol. 9, 1109.
Pinker, S., Jackendoff, R., 2005. The faculty of language: what's special about it? Cognition 95 (2), 201–236.
Pouw, W., Dixon, J.A., 2019. Entrainment and modulation of gesture–speech synchrony under delayed auditory feedback. Cognit. Sci. 43 (3), e12721.
Pouw, W., Harrison, S.J., Dixon, J.A., 2019. Gesture–speech physics: the biomechanical basis for the emergence of gesture–speech synchrony. J. Exp. Psychol. Gen.. https://doi.org/10.1037/xge0000646.
Pouw, W., Paxton, A., Harrison, S.J., Dixon, J.A., 2020. Acoustic information about upper limb movement in voicing. Proc. Natl. Acad. Sci. USA 117 (21), 11364–11367.
Rączaszek-Leonardi, J., 2010. Multiple time-scales of language dynamics: an example from psycholinguistics. Ecol. Psychol. 22 (4), 269–285.
Rączaszek-Leonardi, J., 2014. Multiple systems and multiple time scales of language dynamics: coping with complexity. Cybern. Hum. Knowing 21 (1–2), 37–52.
Rączaszek-Leonardi, J., Kelso, J.S., 2008. Reconciling symbolic and dynamic aspects of language: toward a dynamic psycholinguistics. New Ideas Psychol. 26 (2), 193–207.

Rasenberg, Marlou, Özyürek, Asli, Dingemanse, Mark, 2020. Alignment in multimodal interaction: an integrative framework. Cognit. Sci. 44 (11). https://doi.org/10.1111/cogs.12911.
Reed, E., 1996. Encountering the World: Toward an Ecological Psychology. Oxford University Press.
Rehman, N., Mandic, D.P., 2010. Multivariate empirical mode decomposition. Proc. Math. Phys. Eng. Sci. 466 (2117), 1291–1302.
Richardson, D.C., Dale, R., 2005. Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. Cognit. Sci. 29 (6), 1045–1060.
Richardson, M.J., Dale, R., Marsh, K.L., 2014. Complex Dynamical Systems in Social and Personality Psychology: Theory, Modeling, and Analysis.
Richardson, M.J., Harrison, S.J., Kallen, R.W., Walton, A., Eiler, B.A., Saltzman, E., Schmidt, R.C., 2015. Self-organized complementary joint action: behavioral dynamics of an interpersonal collision-avoidance task. J. Exp. Psychol. Hum. Percept. Perform. 41 (3), 665.
Riley, M.A., Richardson, M., Shockley, K., Ramenzoni, V.C., 2011. Interpersonal synergies. Front. Psychol. 2, 38.
Riley, M.A., Van Orden, G.C., 2005. Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences. Retrieved. http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp. (Accessed 8 April 2005).
Sacks, H., 1992. Lectures on Conversation, 2 vols. Blackwell, London.
Sacks, H., Schegloff, E.A., Jefferson, G., 1978. A simplest systematics for the organization of turn taking for conversation. In: Studies in the Organization of Conversational Interaction. Academic Press, pp. 7–55.
Saltzman, E., Holt, K., 2014. Movement forms: a graph-dynamic perspective. Ecol. Psychol. 26 (1–2), 60–68.
Saltzman, E.L., Munhall, K.G., 1992. Skill acquisition and development: the roles of state-, parameter-, and graph-dynamics. J. Mot. Behav. 24 (1), 49–57.
Schegloff, E., Sacks, H., 1973. Opening up closings. Semiotica 7 (4), 289–327.
Schegloff, E.A., 1990. On the organization of sequences as a source of "coherence" in talk-in-interaction. Conversat. Org. Dev. 38, 51–77.
Schegloff, E., 1991. Reflections on talk and social structure. Talk and Social Structure: Studies in Ethnomethodology and Conversation Analysis. Polity Press, Cambridge.
Schegloff, E.A., 1996. Issues of relevance for discourse analysis: contingency in action, interaction and co-participant context. In: Computational and Conversational Discourse. Springer, Berlin, Heidelberg, pp. 3–35.
Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al., 2013. Toward a second-person neuroscience. Behav. Brain Sci. 36, 393–414. https://doi.org/10.1017/S0140525X12000660.
Schmid, P.J., 2022. Dynamic mode decomposition and its variants. Annu. Rev. Fluid Mech. 54, 225–254.
Schmidt, R.C., Carello, C., Turvey, M.T., 1990. Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. J. Exp. Psychol. Hum. Percept. Perform. 16 (2), 227.
Schmidt, R.C., Nie, L., Franco, A., Richardson, M.J., 2014. Bodily synchronization underlying joke telling. Front. Hum. Neurosci. 8, 633.
Sebanz, N., Bekkering, H., Knoblich, G., 2006. Joint action: bodies and minds moving together. Trends Cognit. Sci. 10 (2), 70–76.
Shockley, K., Richardson, D.C., Dale, R., 2009. Conversation and coordinative structures. Top. Cognit. Sci. 1 (2), 305–319.
Sidnell, J., 2006. Coordinating gesture, talk, and gaze in reenactments. Res. Lang. Soc. Interact. 39 (4), 377–409.
Szokolszky, A., Read, C., Palatinus, Z., Palatinus, K., 2019. Ecological approaches to perceptual learning: learning to perceive and perceiving as learning. Adapt. Behav. 27 (6), 363–388.
Spivey, M.J., Richardson, D.C., 2008. Language embedded in the environment. In: The Cambridge Handbook of Situated Cognition. Cambridge University Press, Cambridge, UK, pp. 383–400.
Stivers, T., 2021. Is conversation built for two? The partitioning of social interaction. Res. Lang. Soc. Interact. 54 (1), 1–19.
Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al., 2009. Universals and cultural variation in turn-taking in conversation. Proc. Natl. Acad. Sci. USA 106 (26), 10587–10592.
Stivers, T., Rossano, F., 2010. Mobilizing response. Res. Lang. Soc. Interact. 43 (1), 3–31.
Stivers, T., Sidnell, J., 2005. Introduction: multimodal interaction. Semiotica 2005 (156), 1–20.
Streeck, J., Jordan, J.S., 2009. Communication as a dynamical self-sustaining system: the importance of time-scales and nested context. Commun. Theor. 19 (4), 445–464.
Tanenhaus, M.K., Brown-Schmidt, S., 2008. Language processing in the natural world. Phil. Trans. Biol. Sci. 363 (1493), 1105–1122.
Thelen, E., Smith, L.B., 1996. A Dynamic Systems Approach to the Development of Cognition and Action. MIT press.
Tollefsen, D., Dale, R., 2012. Naturalizing joint action: a process-based approach. Phil. Psychol. 25 (3), 385–407.
Trueswell, J.C., Tanenhaus, M.K. (Eds.), 2005. Approaches to Studying World-Situated Language Use: Bridging the Language-As-Product and Language-As-Action Traditions. MIT Press.
Trujillo, J., Özyürek, A., Holler, J., Drijvers, L., 2021. Speakers exhibit a multimodal Lombard effect in noise. Sci. Rep. 11 (1), 1–12.
Trujillo, J.P., Simanova, I., Bekkering, H., Özyürek, A., 2018. Communicative intent modulates production and comprehension of actions and gestures: a Kinect study. Cognition 180, 38–51.
Trulla, L.L., Giuliani, A., Zbilut, J.P., Webber, C.L., 1996. Recurrence quantification analysis of the logistic equation with transients. Phys. Lett. A 223 (4), 255–260.
Tu, J.H., Rowley, C.W., Luchtenburg, D.M., Brunton, S.L., Kutz, J.N., 2014. On dynamic mode decomposition: theory and applications. J. Comput. Dynam. 1 (2), 391–421.
Tuller, B., Turvey, M., 1982. The Bernstein perspective: ll. the concept of muscle linkage or coordinative structure. In: Kelso, J.A.S. (Ed.), Human Motor Behavior: An Introduction. Taylor & Francis, New York, pp. 253–270.
Turvey, M.T., 1990. Coordination. Am. Psychol. 45 (8), 938.
Turvey, M.T., 2007. Action and perception at the level of synergies. Hum. Mov. Sci. 26 (4), 657–697.
Turvey, M.T., Carello, C., 2011. Obtaining information by dynamic (effortful) touching. Phil. Trans. Biol. Sci. 366 (1581), 3123–3132.
Valenti, S.S., Gold, J.M., 1991. Social affordances and interaction I: Introduction. Ecol. Psychol. 3 (2), 77–98.
Wagner, P., Bryhadyr, N., 2017. Mutual visibility and information structure enhance synchrony between speech and co-speech movements. J. Multimodal Commun. Stud. 4 (1–2).
Wallot, S., Roepstorff, A., Mønster, D., 2016. Multidimensional Recurrence Quantification Analysis (MdRQA) for the analysis of multidimensional time-series: a software implementation in MATLAB and its application to group-level data in joint action. Front. Psychol. 1835.
Wallot, S., Van Orden, G., 2011. Grounding language performance in the anticipatory dynamics of the body. Ecol. Psychol. 23 (3), 157–184.
Wallot, S., Van Orden, G., 2012. Ultrafast cognition. J. Conscious. Stud. 19 (5–6), 141–160.
Watzlawick, P., Bavelas, J.B., Jackson, D.D. (Eds.), 1967/2011. Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies and Paradoxes. WW Norton & Company.
Willems, R.M., Özyürek, A., Hagoort, P., 2006. When language meets action: the neural integration of gesture and speech. Cerebr. Cortex 17 (10), 2322–2333.
Wittgenstein, L., 1953/2001. Philosophical Investigations. Blackwell Publishing.
Yu, C., 2013. Two interactional functions of self-mockery in everyday English conversations: a multimodal analysis. J. Pragmat. 50 (1), 1–22.
Zima, E., Bergs, A., 2017. Multimodality and construction grammar. Linguistics Vanguard 3 (s1).