

Chapter citation in APA format: Cooney, G., Wheatley, T. (2025). Conversation. In D. T. Gilbert, S. T. Fiske, E. J. Finkel, & W. B. Mendes (Eds.), *The handbook of social psychology* (6th ed). Situational Press. <https://doi.org/10.70400/ZKHH6259>

Conversation

Gus Cooney, University of Pennsylvania

Thalia Wheatley, Dartmouth College

“The great pleasure of conversation, and indeed of society, arises from a certain correspondence of sentiments and opinions, from a certain harmony of minds, which like so many musical instruments coincide and keep time with one another.”

Adam Smith, *The Theory of Moral Sentiments*

Conversation is an ancient and enduring human technology. Throughout history, people have conversed to persuade, cooperate, exchange knowledge, express emotion, gossip, laugh, and fall in love. However, the pervasive and seemingly effortless nature of conversation conceals its complexity as a highly coordinated system that facilitates social interaction. Despite conversation’s status as the centerpiece of sociality, its scientific study has a relatively brief history. This history is also difficult to appreciate because it is scattered across disciplines, as its complexity demands, from linguistics to sociology to psychology to computer science. This chapter synthesizes these varied literatures to explore a set of core features that characterize the conversation system.

Three interconnected themes guide the discussion. First, conversation is built around a highly coordinated system of turn-taking. Second, this turn-level coordination serves as the foundation for mental coordination, a state of “intersubjectivity,” or shared minds, that fuels the conversation machine. Third, with this extensive degree of turn-level and cognitive coordination, conversation partners coordinate two primary goals: the informational (i.e., the transmission of information) and the relational (i.e., the formation and maintenance of social bonds). Such coordination gives rise to the distinct phenomenology of conversation, the norms that govern it, and even common conversational errors. This chapter seeks to foster a deeper appreciation of the exquisite coordination that lies at the heart of the conversation system—of turns, minds, and goals—with the hope of inspiring new research and encouraging scholars from diverse disciplines to embrace conversation as a primary object of inquiry.

THE TURN-TAKING SYSTEM

A hallmark of conversation is that it is built around a system of turn-taking. One person goes, the other person goes, and the first person goes again. This familiar rhythm, in which the "right" to speak flips back and forth, is the primary mechanism by which conversation unfolds. This system is highly efficient, but it is also flexible. After all, if the first person finishes their turn, sometimes instead of the second person taking a turn, there is a prolonged pause, and the first person resumes speaking. Or the first person may take such a long turn that the second person decides to end the conversation on the spot. Turn order, turn length, the relative distribution of turns, and the overall length of the conversation are all subject to considerable variation, which must be negotiated on the fly (Sacks et al., 1974; see also Levinson, 2016).

Thinking about how all this variability is coordinated in real time naturally gives rise to several key concepts and areas of research: *turn exchange*, or how people manage to pass the floor back and forth in an orderly and efficient manner; *turn duration*, or how long speakers talk before they turn over the floor; and *listener feedback*, or what listeners are doing to facilitate coordination while speakers are talking. Importantly, these aspects of conversation not only contribute to efficient coordination but are also imbued with considerable social meaning and importance in their own right.

Turn-Taking: The Exchange Of The Floor

Although turn-taking in conversation may seem effortless, doing so in a timely and efficient manner is a surprisingly complex cognitive operation. In the early days of research on turn-taking, two approaches emerged from very different perspectives. One was that of Duncan, a psychologist at the University of Chicago, who was an early pioneer of face-to-face interaction research (Duncan & Fiske, 1977/2015). Duncan drew inspiration from Mead and the broader pragmatist tradition, recognizing the importance of "starting with the social whole of complex group activity" but also eager to "analyze (as elements) the behavior of each separate individual composing it" (Mead, 1934; p.7). The other approach was founded by Sacks, who was inspired by Garfinkel's ethnomethodology and Goffman's detailed inquiries into spoken interaction (Garfinkel & Sacks, 1970/1986; see also Clayman et al., 2022; Heritage, 1984/2013; Schegloff, 1989). Sacks died young, but not before his influential papers and lectures helped birth a new subfield in the 1960s called conversation analysis (CA), an offshoot of sociology (Sacks et al., 1974; see also Heritage & Clayman, 2010; Sidnell & Stivers, 2012). For Sacks, behavior and meaning in conversation were profoundly contextualized, and therefore elements of the current turn could only be understood in relation to what happened in the prior turn as well as what will happen in the next turn. The differences between these approaches can clearly be observed in their respective ideas about "turn-yielding cues."

Turn-Yielding Cues

How do people know when it is their turn to talk? Duncan and colleagues believed that with sufficiently fine-grained and reliable data, it should be possible to extract the turn-yielding cues that listeners use to determine when a speaker's turn is ending. It is worth noting that the simplest cue might be silence: We wait one second after our partner stops talking before we begin. Using such a standardized, relatively long gap as a strict rule, however, would come across as overly robotic or artificial for natural conversation. Duncan and colleagues painstakingly hand-coded interactions in search of other turn-yielding cues and came up with a list of possibilities, such as the completion of a grammatical clause, a drop in pitch, and direct eye contact (Duncan, 1972, 1974; Duncan & Fiske, 1977/2015; Duncan & Niederehe, 1974).

Although Sacks and Duncan were not in direct dialogue, Sacks and colleagues made it clear that the idea of turn-yielding cues was insufficient, likening them to an "over and out" cue on a radio (Sacks et al., 1974). For one thing, cues occurring late in a speaker's turn would not enable timely speaker changes. Moreover, if listeners were simply waiting for distinct cues at the end of a speaker's utterances, then speech overlaps would necessarily reflect mistakes in the cueing system, which fails to jibe with the frequency of overlaps in naturally occurring speech. And finally, unlike "over and out," which always means the same thing, the meaning of cues depends on the interactional context (e.g., rising intonation in English can mean termination when it appears in a question but continuation when it appears in a statement; Levinson & Torreira, 2015). What Sacks offered in place of turn-yielding cues was the suggestion that turns happen at certain "transition relevant places," which are interactionally determined by a variety of inputs across a broader window of time, allowing people to "project" (i.e., predict) the end of the turns. Ultimately, Sacks' account of how this process worked was incomplete, but his critique and several predictions of the account had a large influence on subsequent research, which attempted to fill in the details.

Turn Preparation And Prediction

Contemporary research in psychology and linguistics has latched onto one particular finding that has helped constrain the discussion of how turn-exchange works: In everyday conversation across cultures and languages, there is evidence that the mean and median turn transition is a short gap lasting approximately 100-200 milliseconds, roughly the duration of an eyeblink (Heldner & Edlund, 2010; Stivers et al., 2009). Of course, typical conversations will include many longer gaps and overlaps, but the vast majority of transitions occur within a +/- one-second window surrounding a seamless no-gap, no-overlap transition, with the most common transitions being very short gaps (Heldner & Edlund, 2010; Stivers et al., 2009). These short gaps between turns are noteworthy because of just how short they are—so short, in fact, that there is not enough time for listeners to produce an utterance starting from scratch. Going from mind to mouth takes time. For example, in extrapolating from studies using picture-naming tasks, it becomes clear that articulating a single-word response requires a minimum of around 400-600 milliseconds of preparation and even longer for a short clause (Indefrey, 2011; Indefrey &

Levelt, 2004; see also Castellucci et al., 2022). In light of the language production system being relatively slow, the speed of listener responses in conversation is remarkable.

What this means is that listeners do not simply wait for speakers to stop vibrating their vocal cords before they start composing their responses, nor do they rely solely on cues that occur at the end of a speaker's turn. Instead, listeners must be preparing their response before a speaker's turn is complete (Bögels, Magyari, & Levinson, 2015; Corps, Gambi, & Pickering, 2018; Levinson & Torreira, 2015; Lindsay et al., 2019; see also Barthel et al., 2016; Bögels, 2020; Sjerps & Meyer, 2015). While preparing their response, listeners also anticipate the end of the speaker's turn, using incoming semantic and syntactic, and perhaps prosodic, information to predict the structure and maybe even the remaining words of the turn (Brehm & Meyer, 2021; De Ruiter et al., 2006; cf. Bögels & Torreira, 2015; Garrod & Pickering, 2015; Magyari & De Ruiter, 2012; Magyari et al., 2014). Then, because listeners can project the remaining content of the turn (i.e., "how it will end"), they may be able to use this information to estimate the timing of the turn (i.e., "when it will end"; Magyari, 2022; cf. Corps, Crossley, Gambi, & Pickering, 2018). Listeners may even decide when to respond by aligning to the speech rates of the speaker and using the expected speech rate to know when to jump in (Corps et al., 2020; see also Garrod & Pickering, 2015).

In sum, one can think of the turn-transition process as a relay race in which the runner from the next leg gets a head start to ensure the baton's smooth passing. The new runner speeds up but also monitors how far the current runner is from the line. Listeners are doing the same, actively *preparing* their turn while *predicting* the end of a speaker's turn, matching their tempo and stride to assume the floor as close to the transition point as possible. Meanwhile, as the turn is unfolding, listeners monitor the speaker for "turn-holding" cues that are meant to prevent interruption—such as an averted gaze or filled pauses such as "uh" or "um" (Ball, 1975; Bavelas et al., 2002; Cassell et al., 1999; Clark & Fox Tree, 2002). Finally, there is evidence for Duncan's original suggestion of turn-yielding cues, now often referred to as *turn-final cues*, which may occur near the end of the turn, such as pitch drop, the re-establishment of gaze, or cues designed especially to mobilize a response from listeners (e.g., tags such as "right?" or "you know?"; e.g., Barthel et al., 2017; Bavelas et al., 2002; Bögels & Torreira, 2015, 2021; Local & Walker, 2012; Rossano, 2012; Stivers & Rossano, 2010; see Rühlemann & Gries, 2020, p. 2, for a list of cues). These may act as a final trigger or go signal, which cues the listener to say what they had planned.

Turn Models

Clearly, turn-taking involves many separate processes: (1) processing and comprehension of the current turn, (2) aligning to the speaker's speech rate, (3) predicting the not-yet-encountered part of the turn, (4) preparing a response, (5) estimating the timing of the end, (6) monitoring for turn-holding cues that halt turn exchange, and (7) waiting for turn-final cues that launch articulation of the response. Current research and models of turn-taking are working out whether all these processes are necessary, what their mechanisms are, and how everything is organized and allocated in the speech comprehension and

production systems (e.g., Garrod & Pickering, 2015; Levinson & Torreira, 2015; Magyari, 2022). There is also a growing interest in incorporating gesture and body movements into turn models (e.g., Holler et al., 2018) and renewed efforts to synthesize research that has explored the developmental trajectory of turn-taking (Nguyen et al., 2022). Indeed, what makes this research doubly impressive is the need to harness expertise across diverse disciplinary subfields to respect the inherent complexity of natural conversation.

When delving into this body of literature, the care and experimental precision devoted to cataloging these various inputs into the turn-taking system is evident. One puzzle, however, is that if one simply reads a conversation transcript that has been segmented into turns by a computer, there is often a large number of "turns" that consist of only a couple of words. Researchers have thus asked how listeners manage to take over the floor in a timely manner even when the previous "turn" is seemingly too short to allow the planning of a response (Corps et al., 2022; see also Meyer et al., 2018). One answer is that what appears in a transcript as multiple turns is not always experienced that way by the people talking.

In everyday conversation, people often engage in surface-level back-and-forth to keep the dialogue going, while simultaneously formulating more substantial responses in their heads. These substantial responses are held until exactly the right moment to contribute. This is an example of a pattern of turn-taking that is difficult to examine because automated segmentation of large conversation corpora usually defines a "turn" as any stretch of speech before a speaker switch. This definition is based on timing without consideration of the content of the turn, and thus can result in speech that is too aggressively segmented into short verbal segments that do not always correspond to what people *in the conversation* psychologically perceive as natural turns. For example, a speaker's story might artificially be broken up into multiple turns in a transcript due to a series of short interjections by a listener—"mhm"... "okay"... yeah". As such, the empirical conclusions of this literature depend critically on the thorny question of what counts as a turn, coupled with the equally thorny question of how to apply such definitions computationally and at scale (Cooney & Reece, 2024). Large naturalistic datasets with more bespoke segmentation algorithms might work in concert with controlled experiments to produce a deeper and more satisfying account of how listeners simultaneously execute the comprehension, prediction, response planning, alignment, and timing processes that ultimately produce the seamless turn-taking that we all take for granted.

Turn-taking is not only an efficient solution to the problem of coordinating communication among multiple minds; it is *the* solution used by humans everywhere. Turn-taking shows consistency across cultures and languages, including speakers of signed languages (Stivers et al., 2009; de Vos et al., 2015). Moreover, turn-taking emerges early in development, appearing in "protoconversations"—the rhythmic back-and-forth vocalizations of caregivers and preverbal infants (Bruner, 1983; Gratier et al., 2015; Hilbrink et al., 2015; Nguyen et al., 2023; Trevarthen & Aitken, 2001). There are also intriguing parallels in the communication systems of other animals, such as the vocal alternation or "duetting" observed in various primate lineages (Levinson & Holler, 2014;

Pika et al., 2018). These collective findings imply a deep and shared cognitive infrastructure that has yet to be fully examined (Levinson, 2016).

Social Implications

Turn-taking is not just a cognitive optimization problem (i.e., how to swap the floor while minimizing gaps and overlaps). Gaps and overlaps are more than just lost time and verbal logjams; they can be intentional and often serve specific social functions—they *mean* something to people in conversation. For example, a well-researched finding is that when a listener's response is delayed by a certain amount of time, it can signal a "dispreferred" turn (i.e., saying something that conflicts with the expectation of the previous turn; Bögels, Kendrick, & Levinson, 2015; Bögels, 2020; Brennan & Williams, 1995; Enfield, 2017; Kendrick & Torreira, 2015; Levinson, 1983; Pomerantz, 1984; Schegloff, 1968; Smith & Clark, 1993; for a computational account, see Gates et al., 2021). For example, when rejecting an invitation, people will often hesitate before delivering their regrets. This is also why it is offensive (or funny) if someone asks a friend whether they are coming to their party, and the friend answers "no" without skipping a beat, offering no further account. This would violate the norm of delivering dispreferred responses with a delay, a deferral, and/or an account. In general, however, the various social functions and implications of gaps and pauses remain relatively understudied, particularly in experimental settings.

At this point, it may be helpful to distinguish between "cooperative" and "competitive" overlaps in conversation. Cooperative overlaps are speech acts that contribute to turn stability (e.g., words such as "yeah," "mm-hmm," etc.; see the section on listener feedback) as well as many affiliative phenomena such as "co-production" or "choral speech," in which one person anticipates what their partner is going to say and joins them to finish the utterance in unison (Lerner, 2004; Pfänder & Couper-Kuhlen, 2019; Tannen, 1989/2007; see also Cummins, 2009). On the competitive side, listeners may infringe upon the rights of the current speaker for various reasons, ranging from the genuinely apologetic to the blithely unaware to the explicitly hostile. Similarly, there are many reasons why speakers may try to reassert a claim to the floor once an overlap has been initiated: They may be motivated by "survivorship" (i.e., they may want to finish their current thought in full), by the "thrust of their turn" (i.e., they may want to bring their turn to the point that the upshot is recognized), or "sequential priority" (i.e., they may wish to ensure that subsequent speakers in a group conversation respond to *their* turn rather than to the interrupter's turn) (Schegloff, 2000). All of these factors suggest that there are many types of overlaps, each with varying degrees of social impact.

Although this discussion is nowhere near exhaustive, it highlights that turn-taking, gaps, and overlaps are part of the social fabric of conversation. Of course, in most cases, gaps go relatively unnoticed as listeners respond in time, and most overlaps are resolved instantaneously or at least swiftly upon the first insistence by one of the involved parties. But gaps sometimes lengthen into pregnant pauses or awkward silences, and overlaps into interruptions or a "fight for the floor." These more dramatic displays are socially impactful and relatively hard to study because of their contextual complexity and nuanced

meaning. Notably, much of this research on the social implications of turn-taking comes from "conversation analysis" (CA), the field introduced at the beginning of this section. Unfortunately, CA remains underappreciated across the broader social sciences, due in part to its inductive methods based on a case-by-case analysis of naturally occurring conversation. However, just as CA influenced the psycholinguistic work discussed earlier, analogously, the CA work described here can serve as a valuable foundation for additional experimental work on the social implications of turn-taking.

Research is starting to examine these questions in increasing detail. For example, studies have shown that shorter average gaps between turns predict how much the conversation partners felt they had "clicked" and enjoyed the conversation (Templeton et al., 2022; see also Koudenburg et al., 2011). One reason for this association may be that short gaps telegraph an eagerness to respond. Another is that short gaps provide an "honest" signal that the listener was paying close attention because responding so quickly cannot be consciously controlled. As gaps lengthen, feelings of connection tend to diminish, at least for strangers. Notably, long gaps between friends do not suffer the same fate. Friends frequently experience long gaps in their conversations and, in fact, experience *heightened* connection when they occur. Rather than preceding a topic change, these long gaps provide space for friends to reflect on what was just said or to savor a shared experience (Templeton et al., 2023)—a case of silence being golden rather than awkward. Future research should explore how the dynamics of turn exchange and enjoyment are moderated by the definition of conversational turn that is used, the type of turn interval (i.e., gap versus overlap), the relationship between the conversation partners, and the activity/social action that precedes the turn interval. Overall, the existing work only scratches the surface, and there remains a significant lack of research, particularly experimental research, that links the turn-by-turn mechanics of conversation to people's beliefs, attitudes, and feelings that arise during and after conversation.

Summary

Conversations would sound very different if listeners waited for speakers to fall silent at the end of their turns before composing their responses. Instead, listeners are constantly planning their own turns, predicting the timing of the speakers' turns, and monitoring for cues to either halt or launch the articulation of their response. This nuanced cognition remains in the service of timely and efficient turn-taking despite the cognitive costs associated with doing so, particularly on the speech production system that must manage many of these processes simultaneously. Additionally, conversations would sound very different if turn-taking were simply a matter of efficiency designed to minimize gaps and overlaps. Instead, people strategically manage the time between turns for social effect, pausing, letting a silence grow, or even cutting their partners off if the situation calls for it.

These characteristics of conversation give rise to many unanswered questions about the dynamics of turn exchange, how it varies across contexts and cultures, how it relates to people's perceptions of their conversation partners, how it affects the overall quality of

conversations, and how to build a machine capable of passing the ultimate Turing test of real human conversation replete with exquisitely timed turns.

Turn Duration: Patterns Of Talk And Silence

Another basic feature of the turn-taking system is the duration of turns, which once again is not fixed or specified in advance but rather varies and needs to be coordinated on the fly. Such coordination is not a bloodless activity. As with all social systems, whether a board meeting, a poker game, or a family picnic, people come ready with assumptions and preferences about the timing of activities. So too in conversation, where we can intuitively feel that people's timing preferences and their different temporal rhythms of speaking and listening strongly influence the "vibe" we get when talking to them. However, scientifically examining temporal patterns of talk has been an enduring challenge.

It was Chapple (1939, 1940), an anthropologist, who first embarked on the study of what he called "interaction chronography." Initially, he was put off by anthropology because he wanted to draw conclusions more general than the culturally specific descriptions common at the time, and he turned to psychology instead. What he found in psychology, however, were personality tests such as the Rorschach test, which he thought required too many assumptions to draw reliable conclusions about people's stable traits (Chapple et al., 1954). Long before the Big Five personality traits were established, Chapple's goals were similar: to find an indicator of personality that could be reliably measured and had predictive power. To do so, he focused on people's sequences of talk and silence, hoping that he might extract from this basic fabric of conversation something revealing about human nature.

Over the next decades, a handful of researchers joined Chapple in his quest, each with a unique perspective. For example, beginning in the 1960s, a group of researchers was interested in psychopathology and clinician-patient interactions (e.g., Jaffe & Feldstein, 1970), and later, psychologists and communication scholars became interested in temporal patterns as indicators of interpersonal responsiveness (e.g., Cappella, 1981). Despite their different perspectives, these researchers shared a commitment to innovative data collection methods and new forms of statistical analyses. Even by today's standards, extracting and analyzing sequences of talk and silence is onerous, let alone considering the technological limitations of the time.

What transpired was a series of sometimes crude, but always clever, devices designed to capture patterns of talk over time. Chapple's original device was called the "interaction chronograph," a recorder that drove a tape at a uniform speed with the help of observers who pressed a key every time a speaker started talking and then released the key when that speaker stopped talking. This recorder captured various measures of duration, such as vocalizations, silences, pauses, and so forth. In the 1960s, researchers working at William Alanson White Institute in collaboration with colleagues at New York University built a new system called the AVTA (Automatic Vocal Transaction Analyzer), which removed the

need for human coders (Cassotta et al., 1964; see also Dabbs & Ruback, 1987; Martz & Welkowitz, 1977). More flexible and sophisticated systems followed, with better acronyms, such as the FIASSCO (Fundamental Interpersonal Arrangement of Sounds and Silences in Conversational Occurrences; Cappella & Streibel, 1979). Alongside the development of these data collection methods, statistical advances such as time series modeling emerged, which together led to a surge of research related to on-off temporal dynamics in face-to-face conversation.

Historically, these developments were significant because although the promise of understanding conversation as a dyadic system was evident, it was not clear from early work in social psychology or sociology how to execute such research (e.g., Cottrell, 1942; Lewin, 1951; Mead, 1934). Addressing this challenge empirically, and with methodical and statistical rigor, became a primary objective for this early group of researchers interested in turn duration.

The earliest research on turn duration can be more fully appreciated with an understanding of the contextual influences of the time. First, this research occurred as a debate raged over personality and whether stability or variability was more important in explaining human behavior (Allport, 1937; MacKinnon, 1944). Second, there were new dyadic models of communication that, coupled with methodological and statistical advances, enabled the modeling of the simultaneous, bidirectional exchange during face-to-face interaction. Third, the researchers interested in these temporal dynamics were also focused on applied problems, such as improving patient-therapist relationships or determining the best hiring practices within organizations.

Against this backdrop, three natural questions arose regarding turn duration. First, can the duration of people's turns be understood as a trait that is characterized by consistency and stability over time? Second, to the extent that people's patterns of turn duration are variable, how do these patterns change and adapt during and across conversations? Third, how is the stability and variability of people's turn duration patterns related to important psychological and applied outcomes such as the breakdown of effective dialogue, perceptions of leadership, and the cultivation of empathy, responsiveness, and social connection?

Stability

Do people exhibit distinctive and stable patterns of talk time? The evidence suggests that they do. Although early studies did not specifically focus on unguided naturalistic conversation, the findings based on the speakers' average durations indicated that people were reliably differentiated by their typical duration patterns. This was true across repeated conversations with the same or different conversation partners and even across different conversation formats, including stressful conversations, biographical interviews, and therapy sessions (Cassotta et al., 1967; Chapple, 1940, 1949; Feldstein, 1972; Jaffe & Feldstein, 1970; Matarazzo, Saslow, & Guze, 1956; Matarazzo, Saslow, & Matarazzo, 1956; Saslow et al., 1955).

The idea behind demonstrating the stability of these temporal patterns was to identify an objectively observable trait, part of a "conversation personality," that could be used to predict important outcomes. Chapple himself, in an early article published in the *Harvard Business Review*, emphasized the importance of turn duration as a predictive measure for hiring and personnel decisions (Chapple & Donald, 1946). This applied perspective captured the attention of other researchers, particularly in organizational behavior, who were intrigued by the notion that hierarchies within groups could be based on individuals' verbal participation, with speaking time, like any limited resource, serving as a basis for power and status (Bales, 1950; Bales et al., 1951; Bass, 1949; Borgatta & Bales, 1956; Kirscht et al., 1959; Mullen et al., 1989). For example, in small groups that were assigned problem-solving tasks, turn duration—defined as the total time spent speaking or the ratio of talk time between conversation partners—proved highly predictive of who would be perceived as a group leader by the end of the conversation. Over time, this finding became known as the "babble hypothesis" (Bass & Stogdill, 1990).

The babble hypothesis remains one of the most well-studied ideas about talk time (MacLaren et al., 2020). It also provides something of a template for what additional work on speaking duration could look like. For example, research on the babble hypothesis has often pitted "quantity" and "quality" of speech against each other to see which is most important. In the case of perceptions of leadership, both show consistent effects, with quantity being surprisingly important in many situations (Jones & Kelly, 2007; Sorrentino & Boutillier, 1975). Having this quantity-quality distinction at the ready may help advance new lines of research on talk time, such as the relationship between talk time and gender (Schmid Mast, 2001, 2002; Brescoll, 2011) or between talk time and negotiation outcomes (Curhan & Pentland, 2007). Moreover, studies have also compared the impact of talk time on leadership to its impact on interpersonal liking—another interesting comparison that might be deployed in additional research. Early research suggested an intriguing inverted U-shaped relationship between talk time and liking (Hayes & Meltzer, 1972; Stang, 1973; see also Borgatta & Bales, 1956; Bales, 1970): Although people who dominate the conversation may be perceived as leaders, they tend not to be the most liked, which requires striking a balance between speech and silence.

Despite the lasting momentum surrounding the babble hypothesis, research on people's stable patterns of talk duration appears to have stalled, leaving many interesting questions unanswered, such as how to compare the quantity and content of speech and whether and how talk time can predict other psychological traits and people's perceptions of their conversation partners (e.g., Hirschi et al., 2022). As Hayes and Meltzer (1972) concluded, this early work could have easily justified "an intensive interest [...] in the quantitative aspects (including frequency, amount, timing and patterning) of [...] talk. But the strange fact is that most social psychologists have been curiously uninterested [...]" (p. 540). We echo their observation and believe these ideas remain ripe for a reboot, especially given recent advances that enable efficient and sensitive measurement of conversational dynamics, which may shed light on the correlates and predictive power of people's stable individual tendencies in conversation.

If these ideas were to be revived, it is worth considering the contextual variability that affects how stable traits are expressed, as well as their impact on important conversational outcomes. Perhaps the greatest contextual modulator of how one person's trait is expressed is the other person with whom they are paired. For example, research has shown that extraversion has a positive effect on certain conversation outcomes through positive emotional displays, but this effect depends heavily on whether one's partner is extraverted as well (Wilson et al., 2016). Analogously, understanding the impacts of being a "long-talker" may require knowing whether one's conversation partner is a fellow long-talker, a short-talker, or somewhere in between.

Moreover, individual differences often interact strongly with the conversational context. For example, certain types of speech may be highly dependent on duration, such as how a short answer to a question may seem underwhelming, a medium answer smart, and a long answer pedantic. Conversely, other speech acts, such as praise, might be equally effective regardless of duration, whether expressed briefly, moderately, or at length. These are just two examples of how the idea of duration as a stable trait might be revitalized while managing the complexities and contextual variability involved.

Variability

Alongside the idea that duration patterns can be understood as stable traits, researchers also recognized that variability must be part of the story. Talkativeness, or verbosity, is influenced by many factors—including one's conversation partner and the topic being discussed—but amidst all possible questions about variability in duration, early research focused extensively on how two conversation partners mold and adapt their turn durations to each other during an ongoing interaction, or how their durations "converge."

A dyadic conversation yields two simultaneously unfolding patterns of talk and silence. A natural question then is how these patterns relate as a dynamic system. Studies using clinical assessment interviews revealed not a simple one-to-one relationship between the length of an interviewer's question and the length of a respondent's direct answer, but rather, an interspeaker effect, one of "mutual synchrony," "entrainment," or "convergence" over time. For example, when a clinician begins by asking short questions, and then, at an experimentally predetermined point in the interview, doubles the duration of their questions, and then subsequently halves them back to baseline, the average length of the respondents' answers proportionally mirrors these changes (Matarazzo et al., 1963; see also Matarazzo & Wiens, 1972). Nonexperimental work from the conversation analysis literature further bolsters this finding—pithily referred to with the maxim, "little questions get little answers" (Heritage & Raymond, 2012).

This work rocketed to prominence with two further papers, the first being a natural experiment involving NASA astronauts communicating with ground control operators (Matarazzo et al., 1964), and the second focusing on President Kennedy's political interviews (Ray & Webb, 1966). Across space flights and presidential interviews, the

duration of the astronauts' and the president's speech appeared to adapt in response to the length of their conversation partners' utterances.

Although these results were eye-catching, they were limited by small sample sizes and interviews with predetermined structures and predefined roles. This raised questions about generalizability—specifically about what happens in naturalistic conversations—which produced early conflicting results. In a review of the literature, Cappella (1981) concluded that the evidence for mutual influence was stronger for variables such as pauses and silences, and weaker for speaking duration. This pattern was true across the two typical indicators of interspeaker influence: (a) changes in average response duration following some manipulation; and (b) intraclass correlations between people's mean response durations across multiple dyads. Cappella and Planalp (1981) followed up on these studies using more sophisticated time series modeling techniques to capture the unfolding convergence. The results supported a significant convergence effect of duration in naturalistic conversation. Nevertheless, the data also revealed many cases of "compensation" where durations showed anticorrelation rather than convergence. Just as reciprocity has a certain face validity—two excited people syncing up with a rapid-fire back and forth—so, too, does the idea of compensation whereby a reticent partner yields more floor time to a verbose partner over the course of the conversation.

The interpretation of duration convergence remains controversial. In the literature, researchers occupied various positions (see Crown, 1991), ranging from the idea that high coordination signals rapport (Chapple, 1970) to the idea that high coordination indicates distress (Gottman, 1979), to a U-shaped account suggesting that moderate coordination is optimal (Warner et al., 1987). All of these hypotheses carry some intuitive appeal. If durations converge, it may feel as though people are similar and on the same page—for example, a couple giving each other equal space to express their perspectives during an argument, with their turn lengths evening out. On the other hand, if duration patterns are too coordinated, it might suggest that people have converged into a highly predictable, rigid, and potentially dysfunctional conversation—such as a couple locked in an unhealthy cycle of accusations and defensiveness, with their turn lengths becoming extremely regular. Finally, perhaps there is some golden mean of convergence whereby people adapt to each other while retaining enough independence to allow their personalities to shine through or to adapt to the interactional needs of the conversation (e.g., Erica had something important happen to her at work today, so she is doing most of the talking). These accounts appear never to have been empirically adjudicated.

Although research specifically focused on convergence of talk durations has declined, the broader field of convergence research, often referred to as "interpersonal coordination" (i.e., "mutual influence," "synchrony," "accommodation," "alignment," etc.), has thrived. Accordingly, researchers have widened the lens to investigate how people in conversation converge across nearly every communicative channel, including speech rate, syntax, vocal intensity, gesture, accent, posture, gaze, and so forth. (Chartrand & Bargh, 1999; Fusaroli et al., 2012, Fusaroli et al., 2016; Giles et al., 1991; Ireland et al., 2011; Ostrand & Chodroff, 2021; Rasenberg et al., 2020; Richardson et al., 2008; Wynn & Borrie, 2022). Although interesting and relevant to conversation (see Section 2 on intersubjectivity), a

full review of the research on convergence would take us too far afield of this chapter's purpose. Overall, early empirical evidence demonstrated that people's patterns of duration exhibited mutual contingency, but there is not enough work, especially experimental work, to fully understand how durations coordinate and what this means in naturalistic conversation.

Rhythm

Examining the stability and variability of people's turn durations at a turn-by-turn level is revealing, but it is worth noting that duration patterns can also be aggregated and examined at a higher level. After all, when people are experiencing conversation and judging their conversation partners, they are not simply paying attention to the length of individual turns but also evaluating longer stretches of dialogue: in the two-to-five-minute range or even longer. Dabbs (1983) referred to these larger blocks of vocal activity as "megaturns" that can give rise to their own temporal dynamics and significantly impact the conversational experience. Indeed, talk times may ebb and flow in conversation as people delve into a topic for which there are practical or emotional reasons to let one person have more of the floor. In return, the activity may flow back toward the other person later on. For example, when reminiscing about a shared experience, one person may take an extended "megaturn" to walk the other through the story's details, only to have the roles reverse later as their partner recounts their own perspective on those same events. Average turn durations and the convergence of these durations do not capture these larger waves of activity that are often characteristic of a meaningful and engaging conversation (Cummins, 2012).

One notable effort to explore these ideas was undertaken by Warner, who wrote a series of papers extending Chapple's research to account for these conversational macro rhythms (e.g., Warner, 1979, 1988, 1992a, 1992b, 2002; Warner et al., 1987; see also Chapple, 1970). Early findings suggested that conversations do have stable rhythms, with speakers exchanging periods of high vocal activity approximately every two to five minutes. In terms of where these stable rhythmic patterns come from, it may again be the interplay of people's conversation personalities, or the acting out of certain established behavioral scripts (e.g., introductory conversations, conversations between professor and student, etc.), or intriguingly, even physiological factors tied to respiration—only the most committed and largest-lunged can monopolize the floor for long periods.

Seeing patterns of duration through this macro lens has many interesting implications. For example, any cycle can be understood not only by its periodicity but also by other cyclic parameters such as changes in amplitude: When speakers are exchanging periods of high vocal activity, does the swap between speakers happen quickly or gradually? Second, to the extent that these rhythms are stable, the tendency is not for a person's vocal activity to return to a uniform average level, but rather, to the average value of wherever that person is on the rhythmic cycle—just as the human body returns its temperature to a different stable level depending on its circadian rhythm. Presumably, the variability and convergence of these rhythms are the result of people's dominance and responsiveness in

relation to their conversation partners and ultimately part of what makes some people overbearing and others meek, or some easy and others awkward.

The complexity involved is clear, especially when trying to draw general conclusions in the face of considerable contextual variability, such as the unpredictability of topic flow within conversation, the differing relationships between conversation partners, and so forth. Perhaps for these reasons, after early progress, the topic appears to have been abandoned. However, with recent methodological advances that enable the collection of larger and more diverse data, automated feature extraction, and multi-timescale analysis, these ideas are ready for reinvention.

Challenges

Before rushing to use the latest methods to collect and analyze data, there is again a fundamental question that the literature has not adequately addressed: What counts as a turn? As previously discussed, this was an issue for research on turn exchange as well. In a similar vein, duration research (and most conversation research performed at scale) often defines turns based on the raw audio signal, counting a turn as finished as soon as any audio from another speaker appears. But this creates many splintered turns where one-word interjections from a listener (e.g., "yeah") get counted as turns. Although this is not always problematic, it can be, depending on the research question. In the case of trying to predict important psychological outcomes from duration—and perhaps even more so here compared to the previously discussed research on turn exchange—turns will likely need to be defined in a way that closely corresponds to what people *in the conversation* perceive as meaningful turns. Providing this definition to a computer, however, is far from trivial, which complicates large-scale data analysis (Cooney & Reece, 2024; Reece et al., 2023). Questions of duration are also highly context dependent. Factors such as the physical location of the conversation, the topic being discussed, the relationship between conversation partners, and their communicative goals (e.g., efficient information exchange versus relationship building) heavily influence duration patterns. But ultimately, this contextual richness is a feature rather than a bug—assuming that scholars manage to wrangle the data.

The duration of our turns and how long we speak contain a remarkable amount of information about us. Although duration patterns may not always be front of mind during conversation—unless we are talking to someone who talks *far* too much or too little—subtle differences in duration can still have a profound impact on our experience: an answer to a question can go from glib to perfect to excruciating simply due to the response's duration; conversations can falter when one person wants to tell lengthy anecdotes while their partner is in the mood for a more rapid-fire back and forth; and conversations can soar when that perfect rhythm is achieved—a volley of perfect durations in which the floor is passed at exactly the right moment, and then, when one's response has welled up again, the floor is deftly handed back.

Summary

Ella Fitzgerald famously sang, "It don't mean a thing if it ain't got that swing." As with jazz, the art of conversation depends on a good rhythm. One important component of this rhythm is the duration of people's vocal activity—their patterns of talk and silence—which are the result of both standard riffs and also improvisation when two people adapt to one another's conversational styles. These temporal patterns are constantly being evaluated during conversation; time can act as a scarce resource reflective of people's power and status, as well as a fundamental currency of affiliation. Many mysteries remain, both on a turn-by-turn level and across the larger ebbs and flows of vocal activity, about how exactly temporal patterns of talk determine or reflect who hits it off and who leaves frustrated. This idea that we might extract from the basic on-off rhythm of human conversation something important about who we are and how we relate to others remains as intriguing today as it was when Chapple first suggested it all those many years ago.

Listener Feedback: Yeah, Yeah, Yeah

Listeners in a conversation are not just speakers-in-waiting. Rather, they play an active role in keeping conversation coordinated and running smoothly. One of listeners' main jobs is to plan their response while simultaneously predicting the end of the speakers' turns so that they can take over the floor with split-second precision. But in addition to managing the content and timing of their own turns, listeners' other main job is to help speakers with their turns by providing them with short snippets of verbal and nonverbal feedback while speakers are talking (e.g., "mm-hmm," "yeah," "exactly!"). This interactive feedback helps speakers feel understood and acknowledged, and it encourages their ongoing participation in the conversation.

Early studies of the feedback that listeners supplied were once again the work of a motley crew of researchers. There was Yngve, originally trained as a cosmic ray physicist, who in the process of launching high-altitude measurement balloons, became interested in the circuits related to early computers. With this knowledge, he pursued pioneering work in machine translation, taking a broader view than was popular in linguistics at the time. He was unapologetically "concerned with much more than just language. I am also concerned with people, and [...] the major use that people make of their language" (Yngve, 1970, p. 567). One major use of language that fascinated him was the idea that listener feedback constituted a separate "channel" of information, which he referred to as the "back channel" to distinguish it from what speakers were doing in the main channel. Other researchers had similarly varied backgrounds and similarly circuitous paths to the subject. Kendon (1967), for example, did a stint with Chapple during his graduate studies and was later invited to join the Institute of Experimental Psychology at Oxford. Although initially focused on gaze, Kendon noticed that listeners' head movements and nods were tightly coupled with their verbal feedback, together forming what he called

"accompaniment signals," which built upon even earlier work by Fries (1952) on how listeners manage to signal their continued attention in conversation.

These early researchers lacked similar backgrounds, shared methods, or even a common set of terms, but one can appreciate their commitment to the raw data of interaction. For example, Fries conducted his research using recordings from *his own* telephone conversations. Dittmann, who produced the most direct work on the subject (Dittmann, 1972; Dittmann & Llewellyn, 1967, 1968), wrote of his initial research: "When we found the very low responsiveness in the first child we recorded, we went to a nearby private school to find out if he was peculiar in this respect, or if the absence of Listener Response was generally true of children his age. We were relieved to find that he was not strangely different from his age-mates since he was [my colleague's] seven-year-old son" (Dittmann, 1977, p. 134). This moment of interest in social interaction was part of a larger push in the 1960s to decode the secrets of nonverbal behavior. In the process, this peculiar class of conversational behavior kept cropping up. It sits somewhere "between speaking and listening" and acts as a kind of collaborative glue that coordinates talk (Gardner, 1998).

In reading this older work, there is a palpable sense of excitement at using new technology to record and analyze large amounts of detailed conversational behavior. Researchers were eager to apply these tools to understand everything from the disordered speech of psychiatric patients to the developmental trajectories of children as they become adult members of their linguistic communities. Over time, however, it appears that this early enthusiasm was tempered by the inherent difficulties in conducting and analyzing studies that capture the complexity of interaction data, which is characterized by enormous contextual and interpersonal variability, without sacrificing external validity. In response to these challenges, research seems to have gone in two directions at once. One direction was deeper into disciplinary silos, exemplified by conversation analysis, which sought to maximize external validity by adhering to a specific set of philosophical and methodological principles. Notable figures such as Jefferson made seminal contributions to this field, pushing the boundaries of our understanding of conversation (e.g., Jefferson, 1984). The other direction involved arduous experimental research that few scholars attempted, such as the work of Bavelas, who, after spending much of the 1960s working at the interdisciplinary Mental Research Institute in Palo Alto, California, went on to spend her career nominally as a social psychologist. However, given her strong commitment to the close examination of face-to-face interaction, she perhaps never felt quite at home within the confines of traditional social psychology (e.g., Bavelas, 2022).

In the end, Yngve's (1970) term appears to have won the day, and this distinctive "listener feedback" is most commonly, but not exclusively, referred to as "backchanneling," and the specific things listeners say as "backchannels" (although "listener responses" is also frequently used). Backchannels have many functions in conversation, but the three most important are that they act as "continuers" (i.e., they facilitate the steady uptake of information), help further coordinate turns, and serve key emotional, affiliative, and relational goals.

Continuers

By design, backchannels often blend into the background hum of conversation. In fact, there is not even a colloquial term in English for backchannels, although some languages have an approximation, such as "aizuchi" in Japanese (which refers to the brief interjections a listener makes to show that they are paying attention and understanding what the speaker is saying). But if one intentionally shifts their focus to backchannels during a conversation, they become distractingly obvious, sometimes occurring more than a thousand times in a single hour of talk (Gardner, 2001; Reece et al., 2023).

Although not limited to storytelling, consider a paradigmatic case of the most basic form of backchanneling: "One person is explaining something or telling a story, the other person is paying attention and understanding, and produces a typical word or sound to indicate this, and also to indicate that he wishes the story-teller to continue. The story teller, without showing any awareness of this response, continues with his story, perhaps slightly encouraged to know that his listener is still interested" (Ward & Tsukahara, 2000, p. 1178).

From this everyday example, it is easy to pick out some important features of backchannels. They are short, spoken quickly, and often unconsciously; they tend to overlap with the speaker's speech without interrupting the flow of their turn; they do not require explicit acknowledgment by the speaker; and their presence functions to convey the listener's continued acknowledgment, attention, and understanding of what is being said. In their most basic form, backchannels have been referred to as "continuers" (e.g., "mhm," "uh-huh," etc.). In essence, continuers confirm that "[t]here is nothing in what you have said to this point that requires me to ask for repetition, clarification, or correction," or even more fundamentally, they answer the question that is implicit in every conversational turn: "Do you understand me and accept what I am proposing up to this point?" (Enfield, 2017, pp. 109–110, summarizing Schegloff, 1982; Clark, 1996).

Continuers function as a green light. At various intersections of conversation, they give listeners an unobtrusive way to signal that they are ready for the speech to continue down the road. If necessary, however, listeners can also change the traffic signal to yellow or even red. For example, if listeners withhold continuers, speakers may slow down or ask if everything is okay, which is especially salient in contexts such as a telephone call in which visual cues are absent and the possibility of talking into the void is very real (e.g., "Are you still there?"). More subtly, even repeating the same continuer over time (e.g., "mhm...mhm...mhm") runs the risk of being interpreted as inattentive, often prompting speakers to slow down and check in. Furthermore, backchannels can function as a request for more nuanced types of slowdowns, such as when listeners need some time to adjust to a new state of the world (e.g., "Oh"; Heritage, 1985). Finally, there are also full-on red lights (e.g., "Huh?" "What?"), which function as the opposite of continuers, usually initiating a process of "repair," which we will explore later (see Section 3 of this chapter on the main goals of conversation).

Overall, it is quite remarkable how effectively and fluidly backchannels can manage the flow of conversational traffic (Enfield, 2017; see also Bergey & DeDeo, 2024). They help speakers and listeners proceed at the right pace, assuming neither too much nor too little of one another and ultimately allowing for the construction of "intersubjectivity" or mutual understanding, which is one of the core goals of conversation (Dingemanse, 2024). Note, however, that so far, backchannels have functioned as signals from listeners to speakers about the pacing and progression of the speakers' turns. But what if listeners do not want the speaker to continue, slow down, or clarify—what if listeners want to take a turn of their own?

Turn Coordination

If backchannels were limited to mere continuers, "uh-huh," or "mm-hmm," it would present a dilemma: Either keep backchanneling—which signals to speakers that they should keep talking—or stop backchanneling and run the risk of signaling to speakers that there is confusion that requires clarification. So, how can a listener both provide feedback that they are not confused but also nudge the speaker toward completion of a turn? Conversation has a convenient and efficient solution: Use different types of backchannels—one type to signal receipt of what a speaker is saying and another type to signal the desire to take a turn. In essence, backchannels not only function as continuers but also signal the listener's "orientation toward the floor."

In the language of the literature, it has been proposed that backchannels range from signaling "passive reciprocity" to "incipient speakership" (Jefferson, 1984, 1983/1993; see also Drummond & Hopper, 1993a; cf. Zimmerman, 1993)—with listeners tending to use "mm-hmm" or "uh-huh" for receptivity and "yeah" or "okay" when they are ready to shift to speakership. Consider a scenario where a long turn is being delivered and how listeners' backchannels might evolve: "mm-hmm...uh-huh...yeah...okay...got it." The specific words being used will vary across linguistic communities and over time, with prosody and embodied behavior also playing a role, but the important point is that specific words appear to be slotted into different functional roles related to the listener's desire to assume the floor.

Building on Jefferson's early work, Drummond and Hopper (1993a, 1993b, 1993c) provided further support for the basic idea that backchannels occur in nonrandom sequences within a turn, discovering in the process even more nuanced functions, such as responses that appear to signal turn closure with maximum strength (e.g., "that's beautiful," or "lovely"). For example, imagine a speaker is telling their friend a story about their recent vacation, and to nudge them toward wrapping up, the friend chimes in with "That's beautiful." While delving into the functional analysis of specific words is valuable, it is important to note that these studies typically rely on close analysis of a limited number of transcripts, and therefore, the literature would benefit from larger-scale empirical analysis and, where feasible, careful experimental work on backchannel function (for an overview of backchannel functions, see Gardner, 2001).

Close attention to the placement of backchannels reveals that they are not randomly sprinkled across turns. They can be found playing the role of continuers with a certain reassuring regularity, they cluster at key moments during narratives, and they are even deployed strategically toward the end of turns to encourage speakership change. Moreover, backchannels often go beyond passive encouragement; they actively mirror a speaker's emotions (e.g., interjecting a timely "wow" while a speaker is telling a story that is meant to be perceived as astonishing). Finally, if one observes conversation closely, it appears that the speakers themselves may have a role to play in eliciting backchannels (e.g., a speaker increasing eye contact when they arrive at a key detail in an explanation). These additional features of backchannels hint at how they might be used in pursuit of relational and affiliative ends.

Affiliation And The Co-Construction Of Dialogue

In order to understand the emotional and affective functions of backchannels, it is helpful to distinguish between "generic" and "specific" backchannels. Continuers (e.g., "mhm," "uh-huh") are examples of generic backchannels because they are context generic, displaying attention and understanding but not depending for meaning on the content of the speaker's turn. In contrast, specific backchannels (e.g., "Exactly!" "Wow!"), sometimes referred to as "assessments," encompass a broader range of responses, commenting on the speech that came before and typically displaying a degree of affective and affiliative alignment with what the speaker is saying (Stivers, 2008).

A key finding is that specific backchannels are important not only for establishing rapport but also for shaping conversation. Across a series of careful experiments, specific backchannels were examined using a paradigm that had speakers tell a "close call" story (e.g., "My friend and I went camping...we made some hotdogs over the campfire...night fell...and then we woke up to a bear!"; Bavelas et al., 2000). Listeners were instructed either to listen normally or to count the number of "t's" in the speaker's story. Listeners in the counting condition produced fewer specific backchannels. Consequently, without these context-appropriate, affiliative displays, the speakers told worse stories—especially stories that had less impactful climactic endings. Consider the bear story's climax: If the listener does not react to or mirror the speaker's fear, the speaker's narrative might veer off path, producing all sorts of awkward compensatory behaviors, such as trying to "justify" the close call, circling back around to the punch line to salvage the story, or simply bailing on the story with a rushed ending. This narrative fumbling underscores the degree to which listeners, through their use of specific backchannels, act as "co-tellers" during conversation.

Additional research has expanded upon these ideas (Tolins & Fox Tree, 2014; see also Bangerter & Clark, 2003; Beukeboom, 2009; Kuhlen & Brennan, 2010). One noteworthy finding is that generic backchannels often lead speakers to continue with the story's content, whereas specific backchannels encourage speakers to elaborate on the content that elicited the backchannel. For example, if a listener backchanneled, "Oh my gosh!" at a specific moment in the speaker's story about the bear, the speaker might double down at

that moment, "Seriously, I don't think I'll ever go camping again." This doubling down does not function to clarify or alleviate confusion, but rather to elaborate a focal piece of the story, bringing speakers and listeners into closer social and emotional alignment. Narratives are truly co-constructed by speakers and listeners in conversation, and backchannels are a key mechanism by which this happens.

Summary

Despite the importance of backchannels for the narrative and affective dynamics of conversation, there are still many gaps in our understanding. Fundamental questions, such as the link between backchannels and rapport, their relevance for turn exchange, and their role in narrative development, require additional experimental support. Cross-cultural work is also needed to answer questions of universality and variability (Beach & Lindstrom, 1992; Clancy et al., 1996; Deng, 2008; Ward & Tsukahara, 2000), and research on children and adolescents—despite the initial interest (e.g., Dittmann, 1972)—is limited (although, see Liu et al., 2022; Park et al., 2017). Research is hampered, in part, by computational challenges such as how to detect and label backchannels at scale (e.g., Edlund et al., 2010), a task made particularly challenging because the lexical properties of backchannels alone do not reveal their function without also understanding the prosody, the frequent visual-gestural component, and the wider conversational context (e.g., Heldner et al., 2013; Schegloff, 1982; Tomlinson & Fox Tree, 2011). Finally, it would be intriguing to situate backchanneling within a broader theoretical framework of what makes a good listener during conversation (e.g., Collins, 2022), perhaps drawing on constructs from the literature on close relationships, such as "responsiveness" (Itzchakov et al., 2022; Itzchakov et al., 2023).

Although it is tempting to locate the flow and fluidity of a perfect conversation in the speakers themselves, research from various fields suggests that a speaker's success often relies on being paired with a skillful listener. Brilliance in the main channel is often due to brilliance in the backchannel. At first, it may be hard to appreciate how repeatedly saying the word "yeah"—the most common backchannel in American English in 2022 (Reece et al., 2023)—could be described as anything close to brilliant, but listeners are known to be considerably resourceful within the constraints they are given. Bounded by time and the need for maximal impact with minimal footprint, listeners occupy the backchannel with great creativity, skill, and purpose, a key part of the collaborative dance that we call conversation.

Turn-Taking Summary

A remarkable feature of conversation's system of turn-taking is that there is no single agreed-upon signal for *turn-exchange*. Speakers do not end their turns with "period" or "over and out." Rather, speakers and listeners come together to negotiate turns on the fly. A second and equally remarkable feature is that there is no single agreed-upon way to manage *turn duration*, no shot clock or agreement to tug one's ear twice to end a turn. And

yet, with some salient exceptions, people behave in durationally sensible ways. A third remarkable feature of the system is the existence of a *backchannel*, which listeners can occupy in order to provide feedback to facilitate the steady uptake of information without interrupting the main channel. Finally, coordinating turn exchange, duration, and backchannels is not simply a matter of efficiency; it is also managed by speakers and listeners for social purposes, adjusting gaps between turns for impact, synchronizing rhythms of talk and silence, and using backchannels to shape narratives and pursue affiliative goals.

Despite its under-the-hood complexity, the dominant feeling of having a conversation is one of seamless coordination. Of course, breakdowns happen—two people wait expectantly for one another to speak only to start, awkwardly, at the same time; the dreaded long-talker prattles on; or uneasiness grows when a conversation partner does not backchannel enough. However, such breakdowns actually underscore the remarkable success of the implicit rules and norms that govern the turn-taking system. After all, without an underlying system of coordination, these departures from smooth turn-taking would not register as noticeable violations. For the most part, the conversation machine just *works*, perhaps the most complicated thing that most people do well. But what justifies this complex, highly coordinated system of turn-taking?

INTERSUBJECTIVITY: A SHARED STATE OF MIND

Why invest in a highly complex, mechanically precise, and socially nuanced system of turn-taking? The answer is that turn-level coordination affords *mental* coordination—a state of shared minds or “intersubjectivity” that lies at the heart of the conversation system. This section explores first how conversation cultivates intersubjectivity in a uniquely effective way (“Conversation as a Unique Form of Communication”). Second, it examines the diverse social and cognitive tools that are necessary to build a shared space of intersubjectivity, such as shared intentionality, mentalizing, and perspective-taking (“Intersubjectivity”). Third, it explores various proposals that have attempted to capture how intersubjectivity unfolds in real time (“Intersubjectivity in Action”). These range from radical perspectives that question the value of cataloging mental processes, to a variety of historical and contemporary approaches that focus on how people build common ground, align their speech, and synchronize their minds.

Conversation As A Unique Form Of Communication

The social psychologist Ragnar Rommetveit was born on Stord, a rural island on the southwest coast of Norway (Hagtvet et al., 2020). Influenced by his upbringing as a speaker of a Norwegian dialect, he often felt like an outsider. He was also famously dissatisfied with his ability to communicate in writing, often preferring instead to lecture and engage with people in dialogue (Hagtvet & Wold, 2003; Wertsch, 2003). Rommetveit regarded conversation as humanity’s greatest invention for bridging sociocultural divides

and thought it should be approached with wonder and awe that it managed to work at all. He understood that everyday conversation is much more than transmitting information from speaker to listener, but involves “a process of building a fragile, temporary, and partially shared intersubjectivity in a deeply pluralistic world.” (Wertsch, 2003, p. 184). But how exactly does conversation construct this important state of “intersubjectivity”? Grappling with this very question inspired pioneering research across multiple disciplines. While the process of constructing intersubjectivity through conversation is nuanced and specialized, researchers have sought to first understand the fundamental mechanisms of how communication itself functions. At its core, communication involves the exchange of representations—“a process involving two information-processing devices. One device modifies the physical environment of the other. As a result, the second device constructs representations similar to [or at least overlapping to some degree with] representations already stored in the first device” (Sperber & Wilson, 1986/1995, p. 1). This broad definition captures communication as it is studied in fields ranging from cell biology to computer science, but it leaves open the question of precisely how the representations stored in one device come to be constructed by the second device. The history of human communication research can be understood, in part, as different ways of modeling this process (Krauss & Fussell 1996; see also Krauss & Chiu, 1998).

The most basic model of communication can be likened to an encoder/decoder framework in which a source encodes information into a message, which is transmitted over a channel and then decoded by a receiver. In humans, the source and the receiver might be the brains of speakers and listeners, respectively, and the message might be encoded in words. This general model is often linked to Shannon, despite the fact that it was not his intention to describe human communication, and he remained suspicious of attempts to do so using his ideas (Shannon, 1948a, 1948b; Shannon & Weaver, 1949; see Ritchie, 1986). Weaver, aiming to adapt Shannon’s theorems to the social sciences, posited that their broader application should extend beyond the technical issue of “How accurately can the symbols of communication be transmitted?” to the semantic issue of “How precisely do the transmitted symbols convey the desired meaning?” (Weaver, 1949, p. 2).

There is, of course, some truth to the idea that meaning is transmitted by being mapped one-to-one with a message consisting of words (e.g., “Will you marry me?”). However, any cursory observation of human communication reveals many instances of talk that defy such a simple model. Consider statements such as, “Can you shut the door?”, “Is it drafty in here?”, and “Were you raised in a barn?”—all of which essentially convey the same request. However, to extract their full meaning, listeners must go beyond the literal words to understand the speaker’s *intentions* (Holtgraves, 2002). This basic fact has inspired many models of communication, especially in the “pragmatic” tradition. For example, Grice (1975; see also Levinson, 2000) examined how deviating from certain conversational norms is interpreted as intentional and conveys meaning beyond the actual words (e.g., describing a singer’s performance as “producing a series of notes” leaves much to be inferred beyond the literal meaning). Likewise, in the “speech act” tradition, emphasis was placed on the difference between direct speech acts (in which the literal and intended meaning of an utterance are the same) versus indirect speech acts (in which this is not the

case; e.g., Austin, 1962/1975; Searle, 1969). Although representations are encoded as words in the transmission of messages, listeners often need to go beyond these words to infer the speaker's intentions if they are to fully grasp the message's meaning.

One can go a step further and note that the same words *and* the same intentions can still elicit different representations in different recipients. This is because people have different backgrounds, assumptions, and degrees of shared knowledge. For example, telling someone "I'll pick you up at 8 p.m." will likely evoke different representations based on contextual assumptions—some may expect a ride in a car, while others envision meeting on foot if we both love walking. As a result, speakers must consider their listeners' unique perspectives and structure their communication accordingly (e.g., if I usually pick someone up walking, but want to drive today, I should give them a heads up in advance). Models of communication that emphasize this aspect of meaning transmission became popular starting in the 1980s, going by various names such as perspective taking, accommodation, audience tuning, recipient design, and so forth (e.g., Bell, 1984; Clark & Carlson, 1982; Clark & Marshall, 1981; Giles et al., 1991; Higgins, 1981, 1992; Krauss & Fussell, 1991b; Fussell & Krauss, 1992).

It is worth noting that while language is certainly the primary way people build intersubjectivity, communication also involves nonverbal behaviors such as gesture, eye gaze, face, posture, and body movement which supplement, reinforce, and can even alter linguistic meanings (e.g., Bavelas et al., 1986). In fact, nonverbal behavior can sometimes convey shared understanding more efficiently and powerfully than words can (e.g., two colleagues locking eyes and smirking whenever Karen speaks at a meeting). Despite the widespread study of nonverbal behavior, its direct contribution to intersubjectivity has received less research attention, partly due to methodological challenges that are now becoming increasingly surmountable (e.g., Bavelas, 2022; see also Holler, 2022; Wohltjen & Wheatley, 2021).

The purpose of this section is not to chart a detailed history of communication models, but rather to highlight the central question they seek to address: How are representations transmitted? This question often hinges on where meaning is located—whether it resides in the words themselves, in the speakers' intentions, or in listeners' idiosyncratic perspectives. Conversation, as a form of communication, can seem like a chaotic and haphazard process, requiring people to fill in the blanks, furrow their brows for clarification, and rephrase things that have already been said to ensure understanding. Meaning appears to emerge in fits and starts, with representations partially transmitted, clarified, crystalized, and even created anew in real time (e.g., Clark, 1996; Gernsbacher & Givon, 1995).

As such, conversation could easily seem like a degenerate version of some ideal form of communication, in which words are carefully chosen, intentions entirely gleaned, and perspectives fully considered. But what seems a mess is, in fact, an intricate and finely tuned system. Conversation involves not only the transmission of semantic meaning, inferring intentions, and taking another's perspective, but it also deploys these processes within a dynamic back-and-forth that enables moment-by-moment mutual adaptation,

allowing people to co-build new representations that go beyond what each individual had before. The goal of any form of communication is to faithfully transmit representations from one person to another, but conversation's real-time co-construction makes it uniquely effective in achieving a state of shared minds or "intersubjectivity"—arguably conversation's fundamental purpose.

Intersubjectivity

In conversation, the state of intersubjectivity is coordinated in a distinctive way, by transmitting, creating, and clarifying meaning in real time using mutual influence and continuous feedback. A storyteller needs a listener; a sad person, a sympathetic ear; a gossip, a co-conspirator. And critically, listeners, sympathetic ears, and co-conspirators are not just receptacles for stories, sadness, and gossip. Rather, speakers and listeners actively clarify and co-construct meaning through mutual feedback as well as through the unique interaction of their backgrounds, personalities, and relational dynamics.

In conversation, people's thoughts and feelings only take on their full shape through ongoing interaction. Rather than independent information processing nodes, people in conversation are akin to co-drivers in a car, with two driver's seats, two gas pedals, and two steering wheels—conversation becomes a shared mental road trip that people collectively steer.

Many scholars have pointed out the importance of this state of intersubjectivity, the common mental world that people create and inhabit in conversation. However, since its famous discussion in Husserl's *Cartesian Meditations*, the term *intersubjectivity* has been used to mean many things, and, consequently, some scholars have opted to abandon the term. This abandonment has yielded the opposite problem, with various disciplines now using different words to describe roughly the same phenomena (e.g., "communion of mutual engagement," "neural synchrony," "psycholinguistic alignment," and "shared reality"; for a related discussion, see Gillespie & Cornish, 2010).

Regardless of the terminology, intersubjectivity lies at the heart of conversation, and many important scientific inquiries remain: What are the basic processes required to achieve this state of shared minds? Are these processes unique to our species? How does the capacity for intersubjectivity develop in children? How might it vary in clinical populations? How exactly is intersubjectivity maintained in a live interaction and what causes it to break down? And what does intersubjectivity feel like to those on the inside—how is it related to "clicking" or "vibing" or thinking of someone as a good conversation partner? Answering such questions will contribute to a deeper understanding of the intricacies of intersubjectivity and its significance in human communication and connection.

Matter Over Mind

The most radical and pioneering ideas about the construction of intersubjectivity—at least from the perspective of an experimentally trained social scientist—come from the field of conversation analysis. Intersubjectivity, Schegloff famously wrote (1992, p. 1299), “is woven into the very warp and weft of ordinary conversation.” This is more than a metaphor for many scholars of conversation analysis; it is a rallying cry. How is intersubjectivity constructed? The conversation analyst might say, “Look, it’s right there in front of you,” laid bare in the conversation itself. Consider this recent exchange with a colleague:

A: What days are you teaching next semester?

B: Tuesdays and Thursdays.

A: Okay.

That’s it. That is intersubjectivity being constructed, mundanely but beautifully woven into even the sparest exchange. But how do we achieve so much from so little?

Even in the most casual conversation, talk is assumed to be “orderly,” governed by a system of principles that are in one sense obvious, but also far from the forefront of our minds while actually engaged in a conversation (Sacks, 1985; Sacks et al., 1974). For example, A’s utterance, “When are you teaching next semester?” is immediately followed by B’s response, “Tuesday and Thursday.” This question-answer couplet is an example of a so-called “adjacency pair” (Schegloff & Sacks, 1973), which is a unit of “sequence construction” that organizes much of conversation (Schegloff, 2007; Stivers, 2012). Question-answer pairs may seem straightforward, but this line of thinking generalizes to many other similarly structured sequences, such as greeting-greeting, inform-acknowledge, or even blame-denial. This is sometimes referred to as a “next turn proof procedure,” the idea that “speakers display in their ‘next’ turns an understanding of what the ‘prior’ turn was about” (Hutchby & Wooffitt, 1998/2008). For example, “It’s six o’clock” can be responded to with “Thanks,” or “Yes, I’m nearly ready,” or “Oh no, we’ve missed it”—each of which can be appropriate and accepted or rejected as a correct interpretation by the original speaker, showcasing the continuous ongoing coordination that is a hallmark of conversation. The general point is that intersubjectivity is woven so deeply into conversation that the fundamental unit of analysis can be thought of not as individual utterances but rather as pairs of utterances that directly demonstrate how conversation partners coordinate their shared understanding. Without the use of Likert scales or measuring BOLD signal, a person’s next turn is taken as evidence of their orientation to the prior turn, and this is CA’s primary gateway into people’s understanding—and especially the construction of intersubjectivity.

Moreover, as might be expected from the locus of intersubjectivity, adjacency pairs are not just what people do in conversation; they are what people *reinforce*. People hold each other accountable for fulfilling the expected actions within an adjacency pair, considering any deviation a breakdown in communication (Corti & Gillespie, 2016; Enfield et al., 2010; Stivers & Rossano, 2010). Consider a slightly more complicated scenario: A speaker

initiates a new adjacency pair by asking, "Do you want to come to my birthday party?" What happens if a listener wants to decline this invitation? As previously articulated, when a response in an adjacency pair is not a preferred response, these "dispreferred seconds" are often delivered with a timing delay and various linguistic markers such as apologies, disfluencies, and hedges (e.g., "...oh...ah, I'm terribly sorry, but..."; Clayman, 2002; Pomerantz, 1984; Pomerantz & Heritage, 2013; cf. Kendrick & Torreira, 2015). This response pattern is exactly what one would expect if people perceived violations of adjacency pairs (e.g., failing to respond to the action of inviting someone to a party in a preferred way) as threats to intersubjectivity, requiring that they signal their awareness of the momentary disruption and their continued commitment to building shared understanding. Further, if the decline of the invitation involved a potential mix-up (e.g., "I'm sorry, I can't make it *this weekend*."), the original speaker might offer a relevant correction to restore intersubjectivity (e.g., "Oh, the party is actually *next weekend*..."). Thus, in terms of intersubjectivity, a vital complement to the adjacency pair is the process of "repair," conversation's mechanism to fix misunderstandings (see section 3 of this chapter, on conversational goals).

In addition to adjacency pairs and the process of repair, close attention to the raw data of interaction reveals many strategies that people use to maintain intersubjectivity. Consider several strategies that a speaker might use in the space of a single turn. First, a speaker might begin their turn by briefly monitoring intersubjectivity using a "pre-sequence" or prefatory clause flagging necessary background knowledge (e.g., "Do you remember Mark?"; Terasaki, 2004). Subsequently, a speaker might add "installment" phrases that sequentially provide clarifying information until the listener signals understanding (e.g., "Mark... we met him at that party... the guy who loves wine..."; Clark & Wilkes-Gibbs, 1986). Finally, a speaker might include a "try-marker," such as a rising intonation, to indicate something as potentially incorrect and thus a threat to mutual understanding (e.g., "I think he lives... *in DC*..."; Sacks & Schegloff, 1979/2007). Meanwhile, as the turns unfold, listeners are continuously signaling their degree of engagement using backchannels. People in conversation use every available communicative channel to manage expectations, monitor perturbations, and reaffirm their commitment to building intersubjectivity in the face of the many obstacles that arise in the bump and grind of conversation.

Intersubjectivity is constantly in flux, and close attention to line-by-line dialogue reveals just how much interactional work speakers and listeners are doing to manage this joint construction project. Attempting to review the large and sometimes unwieldy body of work on this topic by scholars of conversation analysis is beyond the scope of this discussion (e.g., Sidnell & Stivers, 2012; see also Horton, 2017). However, as highlighted in the previous section on turn-taking, there is considerable untapped synergy between ideas born from qualitative analysis and experimental methods (De Ruiter & Albert, 2017). Overall, conversation analysis contributes significantly to the study of intersubjectivity by reminding us to appreciate the intricacies within each conversational moment and to recognize how deeply intersubjectivity is woven into every moment of talk.

Although much can be gleaned by detailed analysis of talk, there is still much progress to be made by incorporating the perspectives of psychology, neuroscience, psycholinguistics, computer science, and other related disciplines. Such an interdisciplinary approach holds considerable potential for advancing our understanding of intersubjectivity and the mechanisms that underlie this fundamental aspect of human interaction.

Shared Attention, Intentions, And Goals

Building intersubjectivity requires a variety of interconnected abilities, including the capacity for shared attention, shared intentions, and shared goals. Shared attention refers to the human ability to attend to something while simultaneously knowing that another person is attending to the same thing. Alongside shared attention, humans also possess the capacity for shared intentions. Intentionality is simply the power of mental states to be about, represent, or stand for something. So, shared intentionality is the ability for one person's mental state to be about something while knowing that another person's mental state is also about that same something, such as a goal that both individuals share.

Shared attention, intentions, and goals may seem obvious in one sense, but they are a defining feature of our species that emerges early in our development, and humans are exceptional in their attunement and motivation to build shared intentional goals in shared attentional spaces (e.g., Moore et al., 1995/2014; Tomasello, 2008, 2022). For example, if an adult points to one of two buckets, a toddler will inspect the bucket that was pointed at, assuming that something interesting is inside; this illustrates how creating a shared frame of attention allows for the attribution of shared intentions and goals, such as jointly searching for toys. Shared attention, intentions, and goals are critical because they enable people to co-inhabit a shared psychological space, which is a prerequisite for engaging in various human-style collaborative activities, the act of conversation being a prime example.

Close attention to the raw data of interaction reveals how, at every stage, people exhibit a strong commitment to managing conversation as a shared activity (sometimes referred to as a "joint" activity as well; Clark, 2006/2020; Levinson, 2006/2020; Tomasello et al., 2005; see also Genty et al., 2020; Goffman, 1981). Even before the conversation officially begins, people may engage in "pre-beginnings," involving glances, mutual gaze, and perhaps large bodily or gestural movement, which serve to open a frame of reference and propose and establish shared attention (Mondada, 2009; see also Kendon, 2004; Schegloff, 1998). Initial greetings can then play a crucial role, for example, by involving comment on the location and timing of the conversation and thereby setting expectations about one's level of commitment to the joint goal (e.g., "I'm running off to the gym but..."; De Stefani & Mondada, 2018). As the conversation unfolds, people repeatedly reaffirm their commitment to the joint goal, justify any disruption or suspension, and even signal changes of focus to avoid confusion (e.g., "I think we've done enough work talk, let's just chill now"; Bangerter et al., 2010; Chevalley & Bangerter, 2010). Finally, there is a preclosing and closing phase which can involve reminiscing about the interaction, reinforcing the strength of the relationship, checking in that there is nothing else

substantive to add (e.g., "Okay?" "Okay"), and finally dissolving the joint action by exchanging well wishes and walking away (Albert & Kessler, 1976; Bangerter et al., 2004; Broth & Mondada, 2013; Clark & French, 1981; Schegloff & Sacks, 1973). These various stages all reinforce the extent to which people "bear witness to the importance of constructing the psychological state of 'togetherness' inherent to human joint action" (Genty et al., 2020, p. 360).

Throughout the process of proposing, beginning, maintaining, and ending the joint activity of conversation, people are exquisitely attuned to slackers, fakers, and free riders—anything that might threaten intersubjectivity, such as signs of distraction, self-centeredness, excessive self-consciousness about the interaction itself rather than being fully present in it, or even being too focused on the topic at hand without attending to their partner's emotional state, where deeper meaning often resides (Goffman, 1957; see also Rumsey & Sidnell, 2011). In essence, the state of intersubjectivity depends on the fundamental capacity of both people to inhabit a shared attentional space, to pursue shared intentions and goals, and their precise attunement to whether others are maintaining an optimal level of engagement in this joint endeavor. Achieving this degree of intersubjectivity, however, also requires unparalleled access to other minds.

Mentalizing

Humans are experts at mind reading, or forming beliefs about what others perceive, believe, and intend. This ability is rooted in basic social perception, such as the direct perception of a partner's smile, laughter, or lack of eye contact. But intersubjectivity also requires going beyond what is simply given by perception to represent, conceptualize, and reason about the mental states of others. For example, when Mark smiles, we can infer that he is likely in a positive mood. But why and what will he do next because of it? Understanding the complex web of someone's mental states, and their relation to past and future behavior, is commonly referred to as "theory of mind" or "mentalizing," or sometimes in communication research as "perspective taking" (Frith & Frith, 2006; Premack & Woodruff, 1978; see also Clark & Marshall, 1981; Krauss & Fussell, 1991b).

Interestingly, there also appears to be untapped potential in viewing theory of mind not only as a means to predict other minds but also as a tool to influence and change other minds. This perspective links theory of mind more directly to its application in conversation, expanding its scope beyond inferring intentions to consider its role in nuanced social tasks such as impression management and persuasion (Ho et al., 2022). It is also worth noting that emotions and feelings fall within the realm of mental states, although they are usually discussed in terms of "empathy" or "empathic accuracy"—and there has been little exploration of how emotional intersubjectivity is achieved in conversation.

In summary, mentalizing also plays a key role in how people extract and co-create meaning during conversation. The state of intersubjectivity, however, is not simply the result of two people trying to understand each other's minds and the ability to share

intentions and goals. Intersubjectivity is characterized by a specific way of pursuing these ends, facilitated by the orderly and coordinated system of turn-taking that this chapter previously examined.

Intersubjectivity And Turn-Taking

It is worth remembering that in conversation, mental coordination is supported by coordination at the turn level (e.g., Levinson, 2006/2020; Sacks et al., 1974; Schegloff, 1992). Shared attention requires alert participation, which is reinforced by conversation's rapid and continuous exchange of turns. Shared intentions and goals require ongoing maintenance and constant attunement to others' perspectives, which backchannel feedback can provide. Moreover, mentalizing is far from perfect, and so people take turns confirming, updating, and correcting their shared understanding. Finally, as previously noted, turn-taking is more than a matter of efficiency. Gaps, overlaps, and pauses all have considerable social meaning, which can provide a real-time window into other minds and the progress of one's shared goals—like a trail of breadcrumbs guiding the moment-by-moment adjustments that facilitate intersubjectivity.

This exquisite turn-taking choreography allows two minds to engage, in real time, with a high degree of continuous attention, clarification, and feedback. When its system of alternating turns and collective steering is functioning well, conversation may be the most sophisticated method for two human beings to achieve and maintain a true "meeting of minds" or intersubjective state.

Intersubjectivity In Action

The conversation machine requires a distinctive set of sociocognitive abilities to achieve intersubjectivity (see Enfield & Levinson, 2006/2020; Enfield & Sidnell, 2022; Levinson, 2019; Levinson, 2006/2020). The previous section briefly outlined some of the machine's possible gears, but what does it sound like when it's running? In other words, how do the various psychological processes just discussed come together to achieve intersubjectivity in real time? Answering this specific question is core to conversation research—even more so than examining the underlying processes that act as prerequisites, which are broadly relevant to many subfields in psychology. The following sections compare and contrast some of the main proposals, ranging from psycholinguistics to neuroscience, social psychology, and computational modeling. However, it should be noted that the field of conversation research is still far from having a detailed account of "intersubjectivity in action."

Building Common Ground

There are many explanations for how intersubjectivity is constructed and maintained on a turn-by-turn basis, but one of the best known is the process of "grounding," explored in

detail by Clark and colleagues (Clark & Wilkes-Gibbs, 1986; see also Fussell & Krauss, 1989a). A classic paradigm used to investigate grounding is a referential communication task that asks people to communicate about abstract geometric figures (e.g., "Tangram figures"; Krauss & Weinheimer, 1964, 1966). For example, imagine how a circle (i.e., a face) and two triangles (i.e., ears) could form a cat. A key feature of the task is that varying shapes are used to produce figures that are somewhat recognizable but more difficult to describe than a cat (e.g., a dancer holding an arabesque pose). In a typical version of the experiment, these tangram figures are printed on a set of cards. One person is assigned to be the Director and the other the Matcher. Each has the same set of cards, but they are separated by a partition so neither can see the other (cf. De Ruiter et al., 2012; Holler & Wilkin, 2011). The Director's cards are laid out in a specific order, while the Matcher's cards are in a random order. Through conversation, the Director must instruct the Matcher on how to order their cards. Without clear names for the figures, the Director and Matcher must coordinate on how to talk about the ambiguous figures. Simply, but beautifully, this creates a divergence in perspectives, and a front-row seat for watching how people assess gaps in their perspectives and try to repair them. This assessment process is referred to as "grounding," and it is ultimately responsible for creating "mutual knowledge" or "common ground" (Brennan et al., 2010; Clark & Brennan, 1991). In general terms, grounding is one process by which people achieve a state of intersubjectivity.

People engaged in a referential communication task start with varying degrees of overlapping perspectives. In some cases, their perspectives align relatively well, facilitating the construction of mutual understanding (e.g., "The first card is a dancer holding an arabesque pose." "Gotcha."). However, there are also many instances in which perspectives may diverge, and people need to work harder to build a shared understanding (e.g., "It looks like a dancer holding an arabesque pose." "Um, what is an arabesque?" "Oh, sorry, I was a dancer. I didn't realize that wasn't a common term—it looks like a person balanced with one leg in the air." "Okay, kind of like a dancer blowing in the wind?" "Exactly."). In many cases, people exhibit similar patterns in their utterances, such as proceeding cautiously in their assumptions and double-checking their understanding. As the conversation progresses, people then converge on abbreviated names for the figures (e.g., "the dancer"), reflecting ad hoc convention formation that underpins the construction of intersubjectivity (Hawkins et al., 2020).

Additional work builds on these ideas, exploring details such as how speakers monitor, modify, and design their speech to increase understanding (Clark & Krych; 2004; see also Horton & Gerrig, 2005), as well as how the listener's backchannels play a role in coordinating common ground (Tolins et al., 2017). Interestingly, studies even examine what happens when a third person overhears the conversation and is tasked with ordering the pictures based solely on what they overheard. In this case, the Overhearers make more mistakes ordering the cards than the Matchers do, presumably because Overhearers are not in direct dialogue with the Director, and have no opportunity for questions, feedback, and so forth. Because of these limitations, Overhearers cannot achieve the same degree of intersubjectivity with the Director that the intended Matcher can (Schober & Clark, 1989; see also Fox Tree, 1999; Fox Tree & Mayer, 2008). These results

highlight the collaborative nature of conversation and how meaning is transmitted, clarified, and co-created—how intersubjectivity is achieved on a turn-by-turn basis.

It is worth noting that a comprehensive theory of language-use in dialogue requires not just an account of how speakers and listeners establish common ground, but also how common ground and beliefs about shared knowledge impact language *processing*, the real-time production and comprehension of language. Many different models of how this occurs have been proposed, and although they contain interesting insights, a thorough exploration takes us too far from a focus on the core features of conversation (e.g., Barr & Keysar, 2007).

Alignment

Pickering and Garrod have laid out a detailed account of how intersubjectivity is constructed based on what they call “interactive alignment,” which serves as an interesting complement to the process of grounding (Pickering & Garrod, 2004, 2006, 2013, 2021; Garrod & Pickering, 2009; see also Wachsmuth et al., 2013). Their model also starts with the idea of conversation as a joint action with the goal of aligning mental representations, and at the highest level, the state of alignment appears to be equivalent to the state of intersubjectivity, in which mental representations are fully shared. However, Pickering and Garrod depart from the previous account of common ground in terms of how the state of intersubjectivity is achieved. In their view, the bread and butter of intersubjectivity does not require further intentional action by which people model one another’s mental states and engage in conscious strategies to collaborate and take each other’s perspectives (cf. Brennan & Hanna, 2009). Rather, the core of intersubjectivity is achieved through a process of automatic imitation that happens across many levels of linguistic representation and that flows directly from the production and comprehension of speech.

To begin, when a word or grammatical construction is activated during comprehension, it enhances its activation during production. That is, when a speaker says a certain word or describes something in a certain way, this primes the listener to imitate it on their subsequent turns. For example, if a speaker uses a rare but vivid word like “serendipitous” to describe an event, it may prime a listener to mirror that same distinctive word choice later in the conversation. Further, it is proposed that alignment happens at many levels of language use simultaneously—phonetic, lexical, syntactic, semantic, prosodic, conceptual – and evidence for this claim comes from research that goes by many names, such as synchrony, mirroring, accommodation, entrainment, and convergence (Branigan et al., 2000; Brennan & Clark, 1996; Garrod & Doherty, 1994; Kim et al., 2011; Levitan & Hirschberg, 2011; Pardo, 2006).

Another key proposition is that alignment occurs not only within each level but also percolates between levels (Cleland & Pickering, 2003; cf. Hawkins et al., 2023). For example, alignment at the level of words might lead to the alignment of higher-level “situation models” (i.e., people’s understanding of the general state of affairs). People who use the

same words to describe things are also primed to have a shared perspective and approach to a given situation.

Beyond linguistic alignment, varied research traditions have provided evidence for the alignment of many other conversational features such as gaze, facial expression, bodily movement, gesture, turn-taking, and emotion (e.g., Bavelas, 2007; Bernieri & Rosenthal, 1991; Cappella & Planap, 1981; Chartrand & Bargh, 1999; Hatfield et al., 1993; Niedenthal et al., 2010; Richardson et al., 2007). All this coordination presumably plays some role in the construction of intersubjectivity—not to mention the rapport that develops between conversation partners (e.g., Tickle-Degnen, 2006). However, it has also been suggested that this coordination can reflect emergent coordinative structures that do not necessarily involve high-level cognitive representations (e.g., Shockley et al., 2009). It is worth noting how linguistic and nonlinguistic alignment might interact in conversation. For example, aligning moods may lead to similar linguistic expressions, such as when two people who are both feeling excited start to mirror each other's use of high-arousal, positive words like "amazing." In contrast, there is also evidence that linguistic and nonlinguistic features can interact in complementary ways, such as when two people synchronize their movements but allow their linguistic contributions to diverge in order to creatively solve a task (Trujillo et al., 2013; see also embodied cognition perspectives to consider how they may relate to alignment in conversation, e.g., Winkielman et al., 2015).

Finally, as the automaticity of all these alignment processes suggests, the basic idea is that the construction of intersubjectivity is not primarily conscious or intentional. Of course, conversation partners may have the overarching intentional goal of arriving at a mutual understanding. And surely, achieving intersubjectivity can, at times, benefit from conscious and effortful strategies to coordinate perspectives. Indeed, even linguistic mirroring, for example, can be conceptualized as a more explicit, conscious process, in which individuals recycle their partners' expressions to enhance mutual understanding (e.g., "conceptual pacts," Brennan & Clark, 1996; Fusaroli et al., 2012). Nevertheless, Pickering and Garrod would likely emphasize—as indicated by the title of an early paper, "Why is conversation so easy?" (Garrod & Pickering, 2004)—that we should not confuse what is salient for what is fundamental. Although certain salient moments in conversation require maximum conscious effort to achieve intersubjectivity, the process of alignment—and therefore intersubjectivity—is less a matter of deploying intentional strategies and more about the automatic and effortless priming of representations, which naturally occurs during the mutual comprehension and production of speech.

The Many Faces Of Intersubjectivity

The construction of intersubjectivity in human conversation is not a one-size-fits-all process, as it heavily depends on the type of conversation being examined. For example, researchers commonly study conversation using structured communication tasks primarily concerned with the transfer of information, such as one person directing another to solve a maze or to chart a route on a map. This type of conversation is worlds apart from something like a negotiation, in which two people are strategically trying to

get the best deal while simultaneously seeming friendly on the surface, or a confrontation between friends in which one friend has recently seemed aloof but does not recognize or acknowledge it. As the competitiveness, ambiguity, and social complexity of an interaction increase, automatic mechanisms of alignment seem less likely to reliably produce intersubjectivity. Moreover, in certain circumstances, people may not even want to align at lower levels of speech. For example, people may choose to reframe a partner's words using their own words. Or, there may be situations in which someone is concealing something, and thus their speech does not align with their representation of the situation (Duran & Fusaroli, 2017).

Inspired by this contextual variation, a third account of intersubjectivity has been proposed by Fusaroli and colleagues (e.g., Dideriksen et al., 2023; Fusaroli et al., 2014; Fusaroli et al., 2016). Rather than focusing on the automaticity of interactive alignment or the conscious strategies of grounding, this account explores how "interpersonal synergy" can function to reduce the complexity of conversation. After all, intersubjectivity must be constructed amidst the whirlwind of planning one's turn while predicting the end of another's turn, listening while also backchanneling, and monitoring our nonverbal behavior and that of others. The cognitive challenge of intersubjectivity is, in part, to reduce all this complexity, which can be accomplished through various synergistic processes that constrain each person's degrees of freedom and reduce the amount of control needed. Linguistic entrainment, for example, is a clear case of such synergy, in which conversation partners linguistically mirror each other. If one person starts using a particular phrase or grammatical structure, like "I reckon" or the passive voice, their partner may unconsciously begin to use those same linguistic patterns, reducing the cognitive load of having to generate novel constructions. Try it out on your spouse to build intersubjectivity while also subtly deflecting blame: "Mistakes were made, I reckon." Other key conversational features, such as backchanneling and adjacency pairs, can also be seen as synergistic processes, in which one person's behavior triggers a complementary behavior in their partner. Critically, it is proposed that *regardless* of whether these synergies are automatic, conscious, or normative, the result is an increased interdependence between people, bringing the seemingly infinite complexity of intersubjectivity within reach of finite minds.

This approach stands apart from previous accounts by emphasizing the need to reduce complexity and make conversation more manageable. Consequently, it does not necessarily expect the percolation effect between modalities proposed by Pickering and Garrod. Instead, differential alignment between modalities can be observed in conversation, whereby alignment at one level can potentially free up resources for greater complementarity at other levels (e.g., Ostrand & Chodroff, 2021). For example, imagine two friends discussing a movie they recently watched together. As they talk, they may unconsciously align their speech rates, speaking at a similar pace. This alignment could reduce the cognitive load of temporal coordination, freeing up resources for complementary processes at another level, such as the divergence in viewpoints required to thoroughly evaluate the film. Another advantage of this approach is its recognition and incorporation of the previously described contextual variations in the construction of intersubjectivity. It emphasizes how automatic priming may reliably produce

intersubjectivity in certain conversational contexts, highlights how alignment can predict positive outcomes in affiliative conversations but negative outcomes in task-based conversations, and explores how key conversational behaviors—such as listeners' inclination to repair mistakes versus let speakers keep talking—may vary based on whether the interlocutors' overall goals are more informational or affiliative (e.g., Dideriksen et al., 2023; Duran & Fusaroli, 2017; Tylén et al., 2023).

Overall, achieving intersubjectivity across the full range of conversational contexts is likely to involve many of the historical, current, and even yet-to-be-theorized models of intersubjectivity working in concert.

New Approaches To Intersubjectivity

Intersubjectivity has many distinctive features: its inherent fragility, its dependence on a toolkit of uniquely human sociocognitive abilities, its relationship with the continuous feedback provided by the turn-taking system, its wide-ranging conscious strategies that people deploy to establish common ground, its extensive automatic alignment that flows directly from the production and comprehension of speech, and its remarkable variability depending on the conversational context and goals. A comprehensive understanding of intersubjectivity will require concerted interdisciplinary effort across areas of expertise. The upcoming sections explore advances in three fields—neuroscience, computational modeling, and natural language processing—that offer promising new insights into the intricate dynamics of intersubjectivity.

Neural Synchrony And Beyond

Work in cognitive neuroscience has moved beyond the "isolated brain" to study the brains of people interacting (e.g., Redcay & Schilbach, 2019; Wheatley et al., 2019). The earliest work related to conversation involved noninteractive paradigms in which one group of individuals recorded an offline, one-way exchange that was later replayed to another group. For example, an early fMRI study related the brain activity of speakers telling an unrehearsed real-life story to the brain activity of a separate group of listeners (Stephens et al., 2010; for a similar result using EEG, see Kuhlen et al., 2012; for early nonverbal results, see Schippers et al., 2010; and for emotional alignment, see Anders et al., 2011). A model of neural coupling based on intersubjective correlation analysis (ISC) indicated that speakers' and listeners' brains exhibited temporally coupled response patterns, and the extent of this coupling was related to the listener's reported understanding of the story. Furthermore, when people's beliefs about an ambiguous story written by J.D. Salinger were deliberately manipulated, people given the same perspective displayed more synchronous brain activity while listening to the story than people given divergent perspectives (Yeshurun et al., 2017; see also Lahnakoski et al., 2014; Nguyen et al., 2019; Nummenmaa et al., 2012). The broad idea that emerged from this work was that once brains are coupled, the coupled systems can generate complex cooperative behaviors, such as conversation (Hasson et al., 2012). Being "in sync" may be a

precondition of joint action. Still, is increased neural synchrony a *cause* of intersubjectivity or simply epiphenomenal with it (Holroyd, 2022)?

Newly developed tools are raising the possibility of answering this question. In particular, "hyperscanning," which involves scanning the neural activity of two or more people simultaneously while they engage in a task, may provide insight into the underlying mechanisms of intersubjectivity. For example, conversation likely requires more than time-locked synchrony in which two people's brains exhibit the same activity at the same time (Hasson & Frith, 2016; Wheatley et al., 2019). After all, the nature of conversation, with its turn-taking structure, is characterized by complementary actions in which speakers build on each other's contributions to weave a shared narrative (Fusaroli et al., 2014; Marková et al., 1995; Mills, 2014). Furthermore, there is evidence that moving *out of* synchrony may play a crucial role in social interaction, particularly when individual exploration and creativity are required (Mayo & Gordon, 2020).

One can also imagine more complex types of coupling that can occur during conversation, such as those related to someone issuing a threat and responding to a threat, providing comfort and receiving comfort, or making a joke and laughing at a joke. These joint actions are the lifeblood of conversation, and although it is expected that two conversation partners' neural patterns should have some lawful relationship in these cases, new experimental and statistical methods are necessary to examine them. It is also becoming clear that brains in interaction have complex temporal dynamics that are not well captured by time-locked interbrain synchrony. For example, studies have found that much of the listeners' coupled activity is on the order of a few seconds delayed from the speakers' activity; there are also times in which the listener's brain activity *precedes* activity in the speakers' brains, which has been interpreted as listeners actively predicting the speaker's upcoming utterances (e.g., Stephens et al., 2010).

Overall, key features of the experiments on neural coupling—including the nature of the tasks used, the activated regions, and the temporal dynamics—suggest that what is being indexed may be closely related to the interactive alignment account proposed by Pickering and Garrod (Hasson et al., 2012). Neural coupling appears to be capturing the information flow between brains, which kicks off the cascade of mutual priming during the production and comprehension of speech. Such neural resonance between speaker and listener may thus reflect and/or facilitate a state of shared minds and shared representations throughout a conversation.

This is certainly one effective way that people build intersubjectivity, but it is not the only way. As previously discussed, other accounts of how people build intersubjectivity emphasize the co-construction of meaning beyond the explicit communicative signal itself (e.g., the words in an utterance). If two romantic partners are cooking together and one asks, "Where's that whatchamacallit slicer thing?" the other might infer the communicative intent and hand over the mandolin. Another key input into building intersubjectivity is a speaker's tendency to add recipient-designed information to their utterances, such as saying, "Where's that thing *you always use to slice stuff*?" This practice demonstrates that speakers do not simply trust the signal to convey meaning without

additional tailor-made information based on who their listeners are. Building intersubjectivity is more than a cascade of priming based on the production and comprehension of speech; rather, it involves perceiving intentions, mentalizing, perspective-taking, and various other cognitive processes (for a review of these abilities from a neuroscientific perspective, see Redcay & Schilbach, 2019). Both speakers and listeners continuously engage in these efforts, contributing to the development of a shared conceptual space that evolves dynamically during dialogue. People then draw on this shared frame of reference to negotiate meaning on the fly.

What are the neural signatures of this shared conceptual space? Interactive paradigms have emerged as a promising approach (Stolk et al., 2013; Stolk et al., 2014; Stolk et al., 2023). Consider a task in which a Communicator must inform an Addressee how to configure a token on a grid in real time. To succeed, the Communicator must take the Addressee's perspective, and the Addressee must infer the Communicator's intentions so that they can align their understanding on a trial-by-trial basis. One study using magnetoencephalography (MEG) showed that people who approached this novel communicative problem displayed upregulated neural activity in regions involved in processing conceptual knowledge and mental models of other agents (i.e., the right ventrolateral temporal lobe and ventromedial prefrontal cortex). Precise temporal information available from MEG revealed that the upregulation was also *not* time-locked to the processing of the communicative signal but rather was present well *before* the occurrence of a specific communicative problem. This suggests the presence of a shared conceptual space that people could draw on to negotiate meaning during conversation. Additional fMRI work using the same task found neural coherence between conversation partners in neighboring portions of a brain region (i.e., right superior temporal gyrus) that is related to interpreting motion based on mental states as well as interpreting pragmatic features of utterances. Once again, cerebral activity seemed to closely track the dynamics of mutual understanding over the course of the interaction.

These findings provide evidence for the specific brain regions and cerebral dynamics involved in how people establish shared meaning. And clearly, this meaning goes beyond being a property of the signal itself or solely residing in the individual minds interpreting that signal. Meaning is best understood within a conceptual frame of reference or context that people develop *together* in dialogue. The capacity and commitment of individuals to maintain and update this shared conceptual space are fundamental to the achievement of intersubjectivity and the richness of conversation itself.

It is important to emphasize that the real debate is not about whether intersubjectivity requires coordinating meaning that lies apart from the signal itself (e.g., de Ruiter et al., 2010; Levinson, 2000). Context plays a role in how to get from signal to meaning, whether that communicative signal is a word, an emotion, or a gesture. Even accounts of intersubjectivity based on automatic priming acknowledge that dialogue is characterized by high-level conceptual alignment. The real question is *how* this shared conceptual space is constructed—whether something like automatic priming can account for the formation and maintenance of this shared space during conversation, or whether more intentional, conscious strategies involving mentalizing, theory of mind, perspective-taking,

grounding, or some other not yet characterized processes, are necessary, or all of the above.

In short, neuroscientific results are not needed to prove that communicating meaning involves shared context. This much is known. What neuroscience does promise, however, is to help clarify the cognitive mechanisms by which people manage to align everything from their low-level pronunciation of words to their emotional expressions to shared higher-level conceptual spaces—and how these processes come together in real time to create intersubjectivity across the diverse array of conversations that shape our everyday lives.

Computational Approaches

One difficulty in discussions of intersubjectivity is the lack of theoretical clarity, which is the result of a long history of ideas from multiple disciplines across multiple levels of analysis. Although this section has done its best to cut a path through this conceptual jungle, another useful machete is formalization and computational modeling. There is still a considerable distance to cover before arriving at comprehensive computational models of how intersubjectivity is constructed on a turn-by-turn basis in naturalistic conversation, but recent findings are encouraging.

This section has discussed several aspects of constructing intersubjectivity in conversation. Broadly speaking, aside from automatic priming, there are two key categories of psychological processes involved: (1) those related to the theory of mind and mentalizing, which are basic capacities necessary to build intersubjectivity; and (2) those related to “grounding,” the back-and-forth process people engage in during conversation to arrive at a shared understanding. Each of these processes is integral to intersubjectivity: (1) theory of mind enables people to go beyond the literal meaning of a communicative signal to convey and understand pragmatic meaning; and (2) because mentalizing abilities are imperfect and speakers and listeners do not always take each other’s perspectives, grounding facilitates the creation, clarification, and coordination of novel meaning when required. In short, two key features of constructing intersubjectivity are the ability to infer pragmatic meaning beyond literal meaning and coordinating shared meaning using feedback. How might these abilities be formalized?

A core feature of theory of mind—and by extension the inference of pragmatic meaning—is a specific kind of recursive thinking (e.g., “If she thinks I think this about XYZ, then she will understand if I refer to XYZ in the following way.”). Formalizing and applying recursive thinking is at the heart of the Rational Speech Act framework (Frank & Goodman, 2012; Goodman & Frank, 2016). In its simplest form, the framework starts with a literal listener, an agent who reasons about the states of words based simply on their literal meaning. The model then adds a pragmatic speaker who reasons *not* about semantics but about what the literal listener knows. The model then continues with a pragmatic listener who reasons about the pragmatic speaker, and so on and so forth. With

very few free parameters, this simple recursive probabilistic model of thought can yield powerfully accurate predictions about how people infer pragmatic meaning.

This basic idea can be applied to people's behavior during specific types of referential communication tasks (Glucksberg & Krauss, 1967; Lewis, 1969). For example, imagine two people are both presented with the same set of three items, [P, E, p]. The speaker then needs to determine how to tell the listener to select the middle item from the set, "E," but the speaker's message can only consist of *one word* selected from a given list: vowel, consonant, capital, or lowercase. This example is easy because "vowel" uniquely differentiates "E" from "P" and "p." But what about another example: [P, E, e]? In this case, there is no single descriptor that uniquely identifies "E" from the other options. If forced, however, a speaker might say "vowel," reasoning that if they were trying to talk about "e," they could have said "lowercase." Assuming the listener is following this same logic, they should then realize that by picking "vowel" (and avoiding "lowercase"), the speaker was trying to refer to "E." When people are given this task, the RSA framework explains the resulting pattern of data remarkably well. It sounds complicated when written out (e.g., "If she had meant to refer to 'e,' I think she would have said 'lowercase,' therefore by 'vowel,' I think she means 'E'..."), but this is exactly the sort of recursive social reasoning that people engage in all the time, and it is fundamental for inferring pragmatic meaning.

The preceding task, although valuable for showcasing the role of recursive thinking in pragmatic language use, did not involve actual back-and-forth communication. Although mentalizing and recursive thinking are essential inputs into the construction of intersubjectivity, allowing people to "talk" to their conversation partners without actually talking *with* them, this does not fully capture the aspect of intersubjectivity that is unique to conversation: the collaborative process of coordinating and co-creating new meaning together on a turn-by-turn basis. This dynamic process of meaning construction is central to the concept of grounding as previously described (e.g., Clark, 1996; Clark & Wilkes-Gibbs, 1986; see also van Arkel et al., 2020).

Additional research has explored these ideas by attempting to formalize the cognitive mechanisms involved (Hawkins et al., 2019; Hawkins et al., 2020; Hawkins et al., 2023). These efforts aim to go beyond understanding how speakers make educated guesses about the minds of others and instead focus on how people use feedback from their partners to adapt their shared understanding. To develop and evaluate these models, researchers have used another referential communication task: a variant of the tangrams task, in which a speaker has to tell a listener how to select a certain tangram from a set when there is no perfect way to do so because the abstract geometric shapes are not entirely recognizable (e.g., "It looks sort of like a dancer, maybe with her arms up..."). Crucially, this task provides valuable insights into how people collaborate and construct shared meaning through repeated turns in conversation—and modeling this process provides an account of the inferential mechanisms that speakers may use to incorporate feedback from their partners over time. Initially, basic models treated listener feedback as a simple indicator of whether they understood or not, which allowed speakers to update their understanding using purely Bayesian principles. However, more advanced models have integrated deep

neural networks to enhance computational efficiency, allowing for the incorporation of a wider range of linguistic feedback that occurs in real-life conversations.

Formalizing the cognitive mechanisms associated with inferring pragmatic meaning (e.g., recursive thinking) and collaborating to create new shared meaning (e.g., the grounding process across repeated rounds of listener feedback) brings several advantages to our understanding of intersubjectivity. First, formalization allows researchers to develop unified frameworks that can be applied to diverse linguistic puzzles. For example, the Rational Speech Act framework has been successfully used to explain a range of linguistic phenomena, such as scalar implicature (Goodman & Stuhlmüller, 2013), adjectival vagueness (Lassiter & Goodman, 2017), overinformativeness (Degen et al., 2020), indirect questions (Hawkins et al., 2015), and other nonliteral language use (Kao et al., 2014; for a review, see Degen, 2023). Second, once formalized, mechanisms can easily be built upon, such as how ideas from the RSA framework can be combined with neural nets to account for richer and more complex communicative contexts (e.g., Monroe et al., 2017). Third, a formalized understanding of cognitive mechanisms can contribute to the development of artificial intelligence with more human-like conversational abilities. Lastly, formalization works to dispel the notion of intersubjectivity as a mysterious and impenetrable shared consciousness. The true marvel resides in comprehending the cognitive mechanisms that enable this remarkable feat, ultimately producing a deeper understanding of how intersubjectivity is achieved across the range of conversation types that characterize people's lived experiences.

"Perceived" Intersubjectivity

Conversation is a marvelous device for constructing intersubjectivity, a state of shared minds. So far, this section has discussed the basic psychological prerequisites needed to build intersubjectivity, their possible implementation in the brain, computational models of these processes, various theories of how intersubjectivity is constructed in real time, and even the microstrategies people use during turn-by-turn coordination. But intersubjectivity also *feels* like something to those "in" the conversation. Unfortunately, research on this subjective aspect of intersubjectivity is scattered across disciplines that use inconsistent terminology, but it seems most logical to consider the possibility that alongside *actual* intersubjectivity—a state of shared minds that causes, and is caused by, the accurate exchange of representations—there also exists *perceived* intersubjectivity, or the subjective experience of sharing (or thinking one shares) inner states and representations of the world with one's conversation partner.

Perceived intersubjectivity is most closely captured by research on "shared reality" (Echterhoff, 2012; Echterhoff et al., 2009; Hardin & Higgins, 1996) and related work on "I-sharing" (e.g., Pinel, 2018). For possible neurobiological mechanisms underlying these phenomena, see Lieberman (2022). Shared reality is defined as the subjective experience (accurate or not) of having inner states in common with another person (e.g., shared attitudes, feelings, and beliefs). It has developed into a wide-ranging theory (see Echterhoff & Higgins, 2018), but historically, it focused on a process known as "audience

tuning," whereby people adapt their communication to their partners (Higgins, 1992; Higgins & Rholes, 1978; see also Zajonc, 1960).

A typical paradigm asks people to read an ambiguous description of a target: "Once Michael makes up his mind to do something, it is as good as done, no matter how long it might take or how difficult the going might be. Only rarely does he change his mind even when it might well be better if he did" (Higgins et al., 2021, p. 104). In this case, Michael can be described as either "persistent" or "stubborn." Research has shown that people will tune to their audience by describing Michael as persistent to someone who likes him and stubborn to someone who dislikes him. This phenomenon may seem equivalent to perspective-taking or to certain aspects of "grounding" that have already been discussed. But whereas this section has previously focused on the processes and strategies that people use to build actual intersubjectivity, this line of research emphasizes the subjective experience that arises and its impacts on people's perceptions of the world. For instance, in the Michael example, tuning to their audience distorts people's memory of Michael, such that they later recall Michael in a way that matches their partner's attitude (Echterhoff et al., 2005; Echterhoff et al., 2008; Echterhoff et al., 2009; Higgins et al., 2021). Tuning to a communication partner can have lasting effects on our perceptions of reality.

More relevant to the current discussion is additional work on "generalized" shared reality, which has started to explore how moments of perceived intersubjectivity can add up to a more general perception that two people see the world in the same way, whether across a single conversation or multiple conversations (Rossignac-Milon et al., 2021; see also Higgins et al., 2021). Instead of studying one-shot messages, research on generalized shared reality measures how much conversation partners felt like they were thinking the same thoughts at the same time and anticipating what they were going to say. Such measures cohere, and for both conversations with strangers and those with close friends or family, generalized shared reality predicts whether two conversation partners feel as though they "clicked" and "vibed" together. Conversation partners who experienced a greater sense of generalized shared reality also felt more certain of their interpretations of the ambiguous stimuli they discussed. Lastly, the construct of generalized shared reality appears distinct from other important measures of interpersonal liking, such as similarity and partner responsiveness.

Future research could explore the relationship between people's subjective experience of building intersubjectivity and the various processes discussed previously, such as grounding, turn-taking, and so forth (e.g., Echterhoff & Schmalbach, 2018; Koudenburg, 2018). This is particularly important because an account of the relationship between actual intersubjectivity and perceived intersubjectivity is surprisingly absent. Although one would expect that the construction of actual intersubjectivity during conversation gives rise to the perception of it, there may be instances where perception and reality diverge, or where perception itself shapes reality. Moreover, people's feelings of intersubjectivity may also arise from specific salient events in conversation, rather than being a cumulative aggregate of all intersubjectivity constructed throughout the interaction. For example, instances of good listening behavior may contribute significantly to the state of perceived intersubjectivity, an interesting and underexplored

area (Collins, 2022). Finally, research has demonstrated that sharing experiences with another person, even in the absence of verbal communication—such as looking at a piece of artwork or tasting chocolate together—can amplify those experiences (Boothby et al., 2014; Boothby et al., 2016). This finding highlights the fundamental non-linguistic processes that might underlie and support the interactional conversational processes that build perceived intersubjectivity. These topics present rich opportunities for future research and exploration.

Another intriguing finding is that an overall measure of semantic similarity, derived from latent semantic analysis (LSA), consistently tracks people's experience of shared reality during conversation (Rossignac-Milon et al., 2021). Moreover, there is a proposed connection between semantic similarity and the construction of actual intersubjectivity (Babcock et al., 2014; Ta et al., 2017). In addition to further exploring the conceptual relationship between intersubjectivity and semantic similarity, research should examine whether these results using LSA (Landauer et al., 1998) can be generalized to larger datasets using newer techniques for creating numeric representations (in vector space) of the meaning and conceptual content of words and utterances, such as Word2vec (Mikolov et al., 2013), GloVe (Pennington et al., 2014), ELMo (Peters et al., 2018), BERT (Devlin et al., 2018), USE (Cer et al., 2018), and GPT (e.g., Radford et al., 2018).

Overall, work on perceived intersubjectivity reminds us of a basic truth: Alongside all the psychological processes devoted to building intersubjectivity in real time, people also have another voice in their heads that reacts to how this process feels. Over time, this inner voice can grow into a chorus, and at key moments—such as at the end of an initial conversation with a new acquaintance, a conversation with a friend after a long absence, or a conversation with a spouse after an argument—people update their perceptions of whether their conversation partners see the world as they do. These feelings about intersubjectivity certainly relate to that universal but elusive sense of being “in sync” or “clicking” with others, and carry profound implications for people's relationships and general well-being.

Intersubjectivity Summary

Ragnar Rommetveit often said that some of the most humbling episodes of his life were when he came home from his worldly travels to the fjords of Norway. After all his reflection on intersubjectivity and his encounters with the thinking of Wittgenstein, Vygotsky, and Husserl, the farmers and fishermen of the island of Stord eagerly wanted to know what he had learned. Rommetveit had to laugh at himself when he realized how much he struggled to give them a good answer. If placed in Rommetveit's shoes, how might one rise to the same challenge?

Intersubjectivity relies on creating a shared attentional space with others, within which people can pursue the shared goal of coordinating their minds. This coordination process is facilitated by a range of uniquely human mentalizing abilities and people's keen sensitivity to their partners' level of engagement in the shared task, reinforced by strong

conversational norms. Various frameworks have been proposed to understand how these psychological abilities seamlessly integrate in real-time to build intersubjectivity. The process of grounding, for instance, emphasizes the conscious deployment of perspective-taking strategies to create common ground, while interactive alignment highlights how intersubjectivity is the direct result of priming during the production and comprehension of speech. As intersubjectivity takes shape, the turn-taking system reinforces key processes by compensating for limitations in perspective-taking, enhancing priming by encouraging listeners to closely attend to speakers' turns, and simplifying tasks like knowing when to speak, which frees up cognitive resources for more complex aspects of coordination. Neuroscience and computational modeling approaches have begun to uncover the neural mechanisms and cognitive processes responsible for key perspective-taking abilities, the priming cascade, and the shared conceptual space that enables people to coordinate complex goals and derive higher-order meaning beyond the words themselves. Finally, close analysis of naturalistic conversation transcripts has revealed how people's motivation and talent for building intersubjectivity are deeply woven into the micromechanics of turn-by-turn interaction. Collectively, these interdisciplinary efforts contribute to our evolving understanding of how intersubjectivity is cultivated and sustained in diverse social contexts.

If the Norwegian fishermen's eyes were glazed over at this point, it would mean that the preceding discussion of intersubjectivity had, ironically, decreased it. This realization may serve as motivation to reestablish the phenomenon, perhaps by changing the topic or asking the fishermen a good question to reel them back in. After all, the state of intersubjectivity *feels* like something to those "in" the conversation. It is profoundly frustrating to talk with someone who is not committed to building intersubjectivity, unsettling when we try but cannot create it, and incredibly rewarding to find a conversation partner who makes the whole process feel effortless. Ultimately, as Rommetveit might say, no matter how much one learns about intersubjectivity, there remains an irreducible sense of wonder in the fact that we can achieve it at all—engaged as we are in the uniquely human ritual of putting our heads together to create a shared mental world.

Conversation is a marvelous vehicle, one that we must mentally drive together, with a coordinated system of turn-taking that keeps us on the road. However, the final big question is: *Where exactly are we headed?*

THE GOALS OF CONVERSATION

There are almost as many reasons people converse as there are topics of conversation. Yet, all of these reasons can—with some simplicity but also some wisdom—be understood in terms of two non-exclusive goals: the informational and the relational. This informational-relational distinction is a simple yet powerful lens through which to view the structure of conversation, its phenomenology, the norms that govern it, and even the

systematic errors that people make when they try to coordinate these goals during conversation.

Two Primary Conversational Goals

Dorothy Parker was a poet, writer, and critic best known for her famously trenchant wit. Her reputation was so legendary that many witticisms have been attributed to her; some perhaps apocryphally, such as, "The first thing I do in the morning is brush my teeth and sharpen my tongue" (Oscar Levant is a likely source). Throughout her career, her writings contained keen social observations, often touching on core issues in communication and especially featuring characters who try valiantly (but fail miserably) to execute the main goals of conversation.

One of Parker's short stories, *Here We Are*, is about two newlyweds traveling by Pullman coach to New York City for the first night of their honeymoon (Parker, 1931/1942). The story begins:

HUSBAND: "Well!"

WIFE: "Well!"

HUSBAND: "Well, here we are."

WIFE: "Here we are. Aren't we?"

HUSBAND: "Eeyop. Here we are."

This all-too-human exchange consists of talk that simply exists to prolong talk—speech not designed to transmit novel information about the world but rather to keep the conversation going for social reasons. The newlywed wife continues:

WIFE: "I mean, goodness, we've only been married about three hours, haven't we?"

After studying his wristwatch, the husband responds,

HUSBAND: "We have been married exactly two hours and twenty-six minutes."

In his response, the husband shifts from social talk to talk that is purely based on factual information transmission—the precise timing of how long they have been married. In response to this mechanical precision regarding their marital timeline, the wife remarks:

WIFE: "My, it seems like longer."

Although this is an extreme example of comedic effect, it is hard to avoid the nagging suspicion that, too often, we have difficulty striking and maintaining an optimal balance between the informational and relational goals in a conversation.

This chapter is not the first to suggest a core tension between two primary communicative goals (see Pinker, 2007). Similar dualities have been posited, such as: transactional/interpersonal (Brown & Yule, 1983); referential/emotive (Richards, 1924/2001); epistemic/relational (Higgins et al., 2021); content/relationship (Watzlawick et al., 1967); phatic/nonphatic (Malinowski, 1923/1946); cognitive/social (Schiffrin, 1987; see also Maschler & Schiffrin, 2015); and others. With these different terms come different shades of emphasis associated with the historical context and discipline in which they were coined. In the past, the informational side of the coin was often privileged because of its association with reason and logic. For example, Malinowski used the term "phatic" communication to refer to the pursuit of relational goals, describing it as "a flow of language, purposeless expressions of preference or aversion, accounts of irrelevant happenings, comments on what is perfectly obvious" (1923/1946, p. 314). Over time, the importance of the relational side was better understood, including the "meta-message," the social subtext of an utterance (Tannen, 1993); as well as discussions of "gossip," which recognize the social functions of conversation as a key "grooming" behavior—*homo sapiens'* verbal version of what other primates manage to accomplish with their fingers (Dunbar, 1996).

In this section, the terms "informational" and "relational" are used to represent the intuitive difference between a person at a bus stop asking another, "Do you know when the bus leaves?" and that same person saying to the other, "Nice weather today, huh?" Informational goals are those related to the efficient transmission of information about the world, and relational goals are those related to forging, maintaining, and adjusting social relationships. Few conversations are solely one or the other but rather a blend of both, replete with moment-by-moment changes in emphasis (Pinker, 2007; Yeomans et al., 2022).

Information Transfer And Relationship Building

Language has been referred to as "the first and greatest human device for stepping up the observational intake" because it allows people to engage in "enlarged or vicarious observation" (Quine & Ullian, 1970/1978, p. 51). Some of what people know comes from firsthand experience, but much more is learned by talking to others. As discussed previously, this massive flow of information is facilitated by turn-taking and intersubjectivity, but it is also coordinated by strong norms.

Norms Of Information Transfer

Historically, the study of norms of information transfer is most closely associated with Grice, who, in "Logic and conversation" (1975), based on his notable 1967 William James lectures, argued that conversation has an inherent rationality. He immortalized this rationality in his conversational "maxims" or norms that people in conversation seem to honor: quantity (be as informative as necessary), quality (be truthful), manner (be clear), and relevance (be relevant). These maxims are the original information transmission

commandments. However, as one might expect, they have also been the subject of debate. Grice himself considered the work incomplete, referring to it as his “tottering steps” toward a theory of communication and meaning (Grice, 1989, p. 4). Critiques have ranged from too much overlap between the maxims to the maxims being too underspecified or the list incomplete. In response, various neo-Gricean proposals have been advanced (e.g., Huang, 2017). For our purposes, what matters is the spirit of these maxims. Speakers and listeners are subject to normative forces when they enter into the contract of conversation. And when conversation partners have the mutual goal of transmitting information, the essence of these governing norms is informativeness, transparency, clarity—or, in a single word, *directness*.

One way to understand the norms governing information transfer is that they exist to preserve the “forward drive” of conversation, the continued unspooling of mutually understood information. And yet, what about all the instances in conversation in which directness is in short supply: when people appear to say more or less than is necessary, when meaning is not stated plainly, and when people bend and blur the truth? Conversations do not unfold like a pilot communicating with air traffic control; rather, they involve friends, business partners, and families whose utterances strike a delicate balance between their informational and relational concerns. It turns out that *indirectness* has its utility—not in the efficient transmission of factual information but rather in the service of relational goals.

Norms Of Relationship Building

Perhaps the greatest contribution of Grice’s maxims lies not in their content but in the accompanying observations about what happens when the maxims are breached. Consider someone describing their recent restaurant meal as “a sequence of flavors.” It is doubtful that this description would lead to a rush of new restaurant reservations. When an utterance deviates from the directness that is emphasized by informational norms, meaning is inferred, which, in this case, is a lackluster dining experience. Grice used the term “conversational implicature” to refer to this act of implicating, rather than simply stating, what one means (see also Levinson, 2000). Metaphors, euphemisms, hedges, and simple expressions of uncertainty are all ways to provide just enough ambiguity to maneuver. An off-color metaphor that lands poorly can be plausibly denied with a swift pivot. Or a hedged position can be backpedaled. For instance, if it turns out that one’s conversation partner actually enjoyed the maligned restaurant, a savvy speaker might lean into the sarcasm as a benign (if misguided) attempt at humor. Although being direct optimizes the clear transmission of information, being indirect greases the social wheels.

Among the various forms of indirectness in conversation, politeness is the most well-studied, owing to the seminal work by Brown and Levinson (1987). Their research built on Goffman’s (1959; 1967) idea that people have a “face,” which is a social construct that represents an individual’s self-image and their need to be respected and valued by others. Politeness theorists distinguish between “positive face,” which pertains to the desire to be liked, accepted, and approved by others, and “negative face,” which involves the desire to

be free from imposition and to be able to act independently. Notably, this distinction bears strong resemblance to the well-known duality of social perception (i.e., the "Big Two") that has been constantly invented and reinvented across many disciplines (Martin & Slepian, 2021). These dualities include warmth/competence (e.g., Fiske et al., 2007; Yzerbyt, 2018), communion/agency (Abele & Wojciszke, 2018; Bakan, 1966); communal sharing/authority ranking (Fiske, 1992); beta/alpha (Digman, 1997); love/power (Leary, 1957; Wiggins, 1979), and so on.

Initially, work on politeness analyzed linguistic behavior across cultures to show that when people formulate requests, they abide by norms of politeness. The specific degree of politeness is dictated by various features of the situation, such as the magnitude of the burden, the closeness of the relationship, and the relative status of the interaction partners. For example, imagine someone who forgot their wallet and needs to borrow money from a friend. The least polite request would be something like, "Gimme some cash," which would undoubtedly threaten all forms of face for nearly all recipients. In the middle, there are various shades of politeness, such as speech that privileges positive politeness (e.g., "Hey, old friend, old buddy, old pal...could I borrow...") or speech that prioritizes negative face (e.g., "I'm terribly sorry to ask this of you, but could I..."). Even more polite, albeit in a riskily indirect way, would be to simply look broke, maybe by opening one's wallet and sighing deeply, hoping that one's conversation partner will intuit the implied request.

Overall, experimental results consistently show that these indirect requests are indeed perceived as more polite than their direct counterparts (e.g., Clark & Schunk, 1980; Francik & Clark, 1985; Fraser, 1990; Holtgraves & Joong-nam, 1990; for summaries, see Holtgraves, 2002, 2010). Moreover, positive and negative politeness strategies often operate together in conversation, with speakers employing different combinations depending on the social context, the competitiveness of the interaction, and their relationship with the listener (Pinker, 2007; Pinker et al., 2008; Yoon et al., 2020).

In summary, although the term "politeness" colloquially invokes asking someone to please pass the salt at the dinner table, this does not do justice to the extent to which social motives are embedded within every fiber of conversation. Politeness is, at its core, meant to capture the basic but important idea that when people talk, they must consider the relational impact on others of what they say. Even a listener's simple request for clarification implies that the speaker was not clear enough to begin with, and the mere act of initiating and sustaining a conversation itself is a demand on someone's time and attention. Thus, maintaining a conversation and continually clarifying the flow of talk—the two most basic prerequisites of having a conversation—are both possible affronts to the two most basic social motives of agency and affiliation. Conversation, no matter how purely informational it may seem, invokes fundamental relational goals. As a result, the balance between informational and relational goals must be continuously and mutually coordinated, which requires both informational and relational norms.

Informational And Relational Norms In Action

A good example of the interconnectedness of informational and relational norms can be observed in the fundamental process of "repair." During a typical conversation, there are frequent informational gaps, such as unknown names and ambiguous references, and *repair* is the process by which these potential misunderstandings, misspeakings, and mishearings are corrected (e.g., Sacks et al., 1974; Schegloff et al., 1977; see also Dingemanse & Enfield, 2024). In their review, Albert and de Ruiter (2018) examined the main typologies of repair, its relevance for understanding intersubjectivity, and the considerable potential for further cross-disciplinary research. For our purposes, what matters is the specific way that repair is executed and what this reveals about the main goals and norms of conversation.

It may be helpful to think of conversation as having a sort of stoplight system (Enfield, 2017). In this analogy, listener-backchannels might function as a green light, signaling that the conversation is flowing smoothly and encouraging speakers to continue. However, there are also many instances in conversation that require something in the flow of speech to be repaired. The most straightforward solution would be to have a signal in conversation that functions as a red light that can be used to halt the conversation to sort out any confusion before proceeding (e.g., "Hold on, who's Jenny?"). Such clarification does happen, but conversation's process of repair is also considerably more nuanced, reflecting both its informational and relational norms. Ideally, conversation would find a way to repair ongoing talk without publicizing the fact that the speaker was unclear in the first place and without interrupting the flow and forward drive. This is exactly what happens.

In conversation, a red light that fully interrupts the speaker's turn is often a last resort. Instead, listeners will sit tight to see if speakers will recognize and repair any misunderstandings or misspeakings on their own. This should sound familiar to anyone who has let a speaker keep talking in the hope that things will become clear in a few more sentences. Speakers typically oblige by closely monitoring their speech for potential problems, such as small lexical or phonological errors (e.g., "Can you pass the sal—uh—I mean pepper?"), or larger misunderstandings when mutual understanding is not being established (e.g., "Jenny called—oh sorry, um—Jenny was my college roommate..."). Speakers may even use "filled pauses," words such as "uh" and "um," to signal that they mean to initiate a repair but require a short delay to do so (e.g., Clark & Fox Tree, 2002; Levelt, 1983, 1993; Gauvin & Hartsuiker, 2020). By engaging in this process of "self-initiated self-repair," speakers preserve the flow of conversation and avoid placing listeners in the position of publicly highlighting the speaker's lack of clarity (Schegloff et al., 1977). This approach beautifully blends informational and relational goals and serves as the initial line of defense against misunderstandings.

Furthermore, while listeners wait patiently for speakers to correct themselves, they may betray nothing. Or, if needed, they may give off various yellow light signals that indicate the need for clarification, such as staring blankly, narrowing their eyes, or quickly furrowing their brow. Finally, if speakers do not catch the drift, listeners must sometimes bring the conversation to a brief halt and ask for clarification, a full red light. These "other-initiated repairs" are also a normal part of talk (Dingemanse & Enfield, 2015; Dingemanse

et al., 2015; Dingemanse et al., 2016), but are notably deployed with considerable care and nuance that reflects the delicate balance between informational and relational goals.

In short, conversation is driven by both informational and relational goals, each often guided by norms that favor directness or indirectness, respectively. The dynamic interplay between these goals and norms pervades conversation, explored here by examining the basic process of repair (for another similar example related to how people negotiate conversational endings, see Mastroianni et al., 2021). As people engage in conversation, they balance the continued uptake of information with their fundamental social goals of agency and affiliation, and repair serves as a captivating example of this delicate equilibrium. However, repair is just one instance that showcases how the fusion of informational and relational norms—and the uneasy marriage of directness and indirectness—gives rise to intriguing aspects of the social cognition of conversation.

Tradeoffs, Errors, And Miscommunication

Achieving a harmonious blend of informational and relational goals in conversation is no easy task. A listener might jump in prematurely to request clarification when a speaker is about to initiate a repair on their own, or they may opt for silence, hoping a speaker's murky narrative will eventually clarify—only to panic when it becomes obvious a few sentences later that they are missing a key detail. These everyday examples are instances of the many errors, biases, and miscommunications that can result from the nuanced dance between informational and relational norms. This section proposes that there are two key biases in conversation—*informational optimism* and *relational pessimism*—each of which appears to act as a default tendency during conversation and flows directly from the dominant norms of conversation.

Informational Optimism

To transmit information clearly, people need to consider the perspectives of others. For example, if someone asks for directions to the nearest coffee shop in a local accent versus a nonlocal accent, the respondent will provide less detail to a local and more detail to an out-of-towner (Krauss & Fussell, 1991b). Similarly, when asked to generate names for ambiguous shapes, people use shorter descriptions if they need to jog only their own memories rather than someone else's; and messages designed for a specific friend are more easily interpreted by that friend than by others (Fussell & Krauss, 1989b, 1989b, 1991; Krauss & Fussell, 1991a; see also Hawkins et al., 2021). Although people tend to formulate messages with others' perspectives in mind, this ability is far from flawless.

One key strategy for perspective-taking is initially to assume that other minds are similar to one's own. In many ways, this assumption of a shared perspective makes sense. Speakers of the same language often have overlapping representations of word meanings as well as shared cultural references. Even people from extremely varied backgrounds have had relatable experiences simply by virtue of being fellow humans with similar

biological constraints, environmental concerns, life stages, and so forth. However, this "egocentric" assumption of shared perspective—although a useful starting point—is often overapplied and undercorrected, leading to systematic errors. For example, people who know the meaning of an uncommon word, such as "sidereal," tend to overestimate how many other people will also know the meaning (Hayes & Bajzek, 2008; Nickerson et al., 1987). People who are exposed to a puzzle's solution think that the puzzle is easier to solve than people who have never seen the solution (Kelley & Jacoby, 1996). And already having procedural knowledge, such as how to swing a croquet mallet, causes people to underestimate how long it will take a novice to learn the task from scratch (Hinds, 1999). These "curse of knowledge" effects—sometimes thought of as special cases of the "false consensus effect" (Ross et al., 1977)—involve people falsely attributing their knowledge to others, and these biases persist despite considerable financial incentives to the contrary (Camerer et al., 1989). Finally, people also display an "illusion of transparency," assuming that their inner states, such as anxiety or nervousness in social interaction, are readily apparent to others (Gilovich et al., 1998; Gilovich et al., 2000; Savitsky & Gilovich, 2003; cf. Boothby et al., 2017).

It is easy to imagine how this egocentric assumption and the resulting errors that occur might bedevil communication (Keysar et al., 2003). For example, researchers asked people to read ambiguous sentences aloud (e.g., "Angela shot the man with the gun") while viewing pictures that clarified their meaning (e.g., the man was holding a briefcase, not a gun, when Angela shot him). People then overestimated how well listeners would understand the correct meaning of the sentences solely from hearing the sentences without visual cues (Keysar & Henly, 2002). In another study, one group of people was informed that the archaic idiom, "the goose hangs high," meant that the future looked bright, and another group was told that "the goose hangs high" meant that the future looked bleak. Again, people mistakenly thought naive listeners would be more likely to interpret the idiom in the manner that they perceived as correct (for the curious reader, the real meaning of the idiom is that things will be pleasant, an allusion to the fact that geese tend to fly higher during pleasant weather; Keysar & Bly, 1995). Once people have knowledge themselves, they find it hard to separate their own perspective from others, which hinders their ability to communicate clearly.

This egocentric tendency to assume that "what's in my head is also in yours" extends to natural conversation and can lead to similar errors. For example, research has shown that people expect extraordinary experiences, such as drinking rare wines or going on exotic vacations, to be delightful. However, a less obvious consequence of such extraordinary experiences is that they are unlikely to be shared by others, which can impact conversations. In a series of studies, people predicted that they would be better off after having a great experience (e.g., watching an amazing video) compared to a mediocre experience (e.g., watching an ordinary video) *even* when they knew that they would later converse with a group of people who all had the ordinary experience. These predictions turned out to be wrong. Although people who had a great experience did indeed enjoy it, they subsequently felt excluded by their conversation partners, and the unanticipated pain of exclusion outweighed the pleasure of having a great experience (Cooney et al., 2014). Moreover, speakers also tend to miscalibrate their ability to communicate novel

experiences during conversation. Although much of conversation revolves around speakers sharing novel experiences with their listeners—cities they have never visited, books they have never read, or foods they have not tasted—experiments have demonstrated that people underestimate the difficulty of effectively conveying these novel experiences and therefore overestimate how much their conversation partners will enjoy hearing about them (Cooney et al., 2017). Extraordinary and novel experiences offer clear and foreseeable benefits for the experiencer, but the subsequent conversational costs are more difficult to anticipate.

What animates the preceding research is the idea that people are informational optimists—quick to assume that they can clearly transmit information to others, while underestimating how difficult it is to properly account for all the “informational gaps” inherent in speech. Theoretically, this work extends well-known research on egocentrism (e.g., Epley et al., 2004; Keysar, 2007; Nickerson, 1999) to the domain of face-to-face interaction, shedding light on how egocentric tendencies can influence topic choice, storytelling, and information transmission during conversation.

Although speaker egocentrism may offer the most straightforward explanation for excessive informational optimism, listeners may also play a role. After all, it takes (at least) two to dance the conversational tango, and listeners have at their disposal many mechanisms to resolve communication errors. So why do listeners not simply ask for clarification? Why is informational optimism a lingering threat to the success of conversation rather than a mere blip in understanding that is immediately corrected?

The answer once again lies in the delicate balance that listeners must strike between their informational and relational goals. This is *not* to say that speakers and listeners are unable to function as an effective team. There are many instances in which people are not “mind blind,” and indeed, behave in “resource rational” ways, ramping up or down their perspective-taking efforts depending on how actively listeners are providing feedback because the achievement of intersubjectivity depends on it (e.g., Hawkins et al., 2021). Nevertheless, listeners have a difficult job. Confronted with egocentric speakers who overestimate their communicative abilities, listeners face the challenge of correcting unclarity while preserving social harmony. Possible manifestations of this delicate balancing act have been discussed previously—namely, the process of repair and the norms of politeness. Both of these dictate that listeners err on the side of letting speakers correct themselves. By elevating indirect norms of politeness above direct norms of clarity, misunderstandings can linger and deepen.

Interestingly, research has provided even more dramatic demonstrations of listeners’ insensitivity to potential communication problems (e.g., Galantucci & Roberts, 2014; Galantucci et al., 2018; Galantucci et al., 2020; Roberts et al., 2016). In a series of studies, researchers introduced a range of incoherencies into conversation: surreptitiously swapping in a random partner for 30 seconds during an instant-messaging exchange, inserting an incoherent reference to a celebrity in an unrelated discussion, and even having a confederate—apropos of nothing—interject the nonsensical phrase “colorless green ideas sleep furiously” into face-to-face conversations. Remarkably, listeners not

only failed to request clarification, but they often reported not even noticing the nonsense utterances, even when the conversation ended shortly thereafter.

Importantly, these effects appear to go beyond run-of-the-mill politeness, which is often a fully conscious decision to ignore incoherence. These phenomena have been collectively referred to as "content deafness," drawing parallels to the well-known phenomena of "change blindness" or "inattention blindness" in visual attention, in which people fail to notice unexpected stimuli, such as a gorilla that appears on the screen during an attention task (Rensink et al., 1997; Simons & Levin, 1998; Simons & Chabris, 1999; see also Simons & Rensink, 2005). Just as visual gorillas go unnoticed, verbal gorillas also vanish into the conversational mist. Moreover, in the same way that change blindness is important for understanding perception and attention, content deafness may be important for understanding the role of motivation in maintaining a conversational environment that nurtures social ties—including psychological processes that go beyond the consciously executed strategies typically associated with politeness.

It is worth noting that despite this apparent insensitivity, listeners can also be quite sensitive to incoherence. Listeners initiate repairs at exactly the right places, their degree of confidence tracks their requests for clarification, and so forth (Micklos et al., 2020; see also Fay et al., 2018). To take a concrete example, people's vigilance seems to know no bounds when the subject is office politics: If a speaker accidentally confuses two individuals—such as saying "John" instead of "Mark"—a listener might quickly interject, seeking immediate clarification (e.g., "Wait! Hold on, John said that?"). Listeners often keep close tabs on who is present in a story, who knew what, and who did what to whom. This stands in stark contrast to instances in which listeners gloss over quite dramatic cases of incoherency. What explains listeners' behaviors, ranging from unconsciously wiping incoherence from their minds to attentively hanging onto every word to ensure pinpoint accuracy?

The explanation for such varied listener behavior—whether lack of clarity is ignored or pounced upon—lies once again in the tradeoff between maximizing clarity (informational goal) and maintaining social harmony and conversational flow (relational goal). This is why it can be so jarring to get direct feedback on a presentation or lecture: Once listeners are freed from relational concerns, every last instance of egocentric unclarity is highlighted. But, in the relational mode, egocentric unclarity can go unchecked. For the sake of the relationship, listeners may let speakers believe they are better understood than they are. Assuming that silence implies comprehension, speakers may fail to appreciate the many instances in which listeners choose not to flag misunderstandings. Most of the time, these overlooked incoherencies never come to light or are inconsequential. For example, missing a detail about a conversation partner's trip to Rome is not the end of the world, so why ruin the flow of talk? On the other hand, there are instances where miscommunication does surface later on (e.g., "Didn't I just tell you that?", "Why didn't you ask if you didn't understand!"). Such cases are likely the exceptions that prove the rule: In conversation, speakers tend to believe that if something is unclear, listeners will ask for clarification, while listeners may err on the side of just letting it go by.

This tacit coordination may be a default equilibrium that generally works—until it doesn't.

Given the nuance and complexity of social interaction, it is important to acknowledge the limitations of any overarching theory that attempts to explain the systemic errors observed in conversation. For example, consider an error that stands in contrast to informational optimism, such as the tendency for speakers to *overconfirm* understanding when it is not necessary. This behavior can be readily observed in everyday conversations (e.g., "Did you get that...does that make sense...are you sure...know what I mean?"). Although there is a lack of direct empirical evidence, it is plausible that this behavior is relatively uncommon due to the potential threat it poses to the listener's negative face (e.g., "Yes, I heard you the *first* time..."). Relatedly, it is conceivable that listeners could also excessively seek clarification during a speaker's narrative, leading to a pedantic exchange. This certainly happens, too, but it does not seem to be a stable equilibrium in conversation, again likely due to swift and pointed feedback (e.g., "Just let her finish the story!"). Informational optimism is not the only error that arises from the coordination of informational and relational goals, but current empirical evidence suggests that it is a dominant one.

Overall, this section has argued that informational optimism results from speakers' naturally egocentric default together with listeners' reluctance to address unclarity due to their relational and affiliative goals. However, speakers' own relational goals can also contribute to the issue. For example, when making requests, speakers may choose to express themselves indirectly to minimize face-threat (e.g., "Getting to the airport on the train is just so confusing these days."), but in doing so, they may overestimate the listener's understanding of their intentions (Holtgraves, 2021; see also Chaudhry & Lowenstein, 2019). Sometimes, this sort of miscommunication is resolved in the moment by speakers being more direct, but other times, the miscommunication persists, only to be revealed later (e.g., "Why didn't you just tell me you wanted a ride to the airport!"). Here, the root cause of speakers' overoptimism, and the resulting miscommunication, is not solely the listener's failure to address the unclarity, but rather the speaker's overly optimistic assumption that their veiled intentions were understood in the first place.

In summary, effective communication requires overcoming egocentrism—transcending one's own perspective to take that of another person. As discussed in previous sections, turn-taking and the construction of intersubjectivity are both fully dedicated to the cause. However, the challenge of overcoming egocentrism is compounded by the need to coordinate informational and relational goals. Sometimes people are informational logicians, striving for perfect accuracy. At other times, people are social magicians, nodding their heads, smoothing over mistakes, and creating the illusion of seamless communication. Perfection would require that speakers and listeners fully agree on when and how to adopt these roles, but alas, conversation is marked by unspoken preferences and an impossible balance of informational and relational concerns that shift depending on the topic of conversation and the evolving nature of people's relationships. This appears to result in certain stable patterns of miscommunication, such as the bias of

informational optimism, which may persist as an imperfect solution to yet another coordination problem that lies at the heart of conversation.

Relational Pessimism

In contrast to the optimism that characterizes people's pursuit of informational goals, research has also shown that people are often pessimistic about their ability to execute the relational goals of conversation.

People need social connection almost as much as they need food, shelter, and sleep. Despite the widespread benefits of social interaction for health and happiness, people do not always capitalize on the social opportunities afforded to them. Dunn and Lok (2022) have likened this situation to a metaphorical island of abundance—a land of milk and honey—where the inhabitants are nonetheless starving by failing to take advantage of the sustenance. In some ways, modern society resembles this island, but instead of food, it is the opportunity for social connection that is being ignored (e.g., Epley et al., 2022). Conversation, in particular, serves as social nourishment, but a growing body of work demonstrates that people are remarkably pessimistic about their ability to deploy conversation to form, maintain, and adjust their social relationships.

One basic manifestation of relational pessimism involves underestimating the value of conversations with strangers. For example, people in a field study in Chicago underestimated how rewarding it would be to strike up a conversation with a stranger during their commute home on the train (Epley & Schroeder, 2014; see also Schroeder et al., 2022). Even brief moments of social connection with strangers, such as a short conversation with a barista, have been shown to increase well-being (Sandstrom & Dunn, 2014a, 2014b; see also Gunaydin et al., 2021). Beyond the hedonic benefits, people also appear to incorrectly predict the depth of topics that they can discuss without feeling awkward (Kardas, Kumar, & Epley, 2022) and how long they can talk before running out of things to say (Kardas, Schroeder, & O'Brien, 2022).

People's pessimism about talking to strangers springs from a variety of sources, such as fear of rejection, doubting that they will enjoy the conversation, and a perceived lack of conversational ability (Sandstrom & Boothby, 2021; see also Welker et al., 2023). Unfortunately, these fears run deep and are resistant to change—even after a positive interaction, people will assume that their next conversation will be worse—although evidence suggests that a week of practice approaching and talking to strangers can alleviate some of these anxieties and better align people's beliefs with reality (Sandstrom et al., 2022).

Not only do people underestimate the value of striking up a conversation with someone new, but this relational pessimism extends to people's perceptions of the impressions they leave on others. For example, people exhibit a "liking gap," systemically underestimating how much their conversation partners actually liked them after an initial conversation (Boothby et al., 2018). This pessimism about first impressions is evident in children as

young as five years old, occurs in both Western and non-Western cultures, applies to coworkers forming new teams, and extends to other positive traits such as intelligence and conscientiousness (Elsaadawy & Carlson, 2022; Li et al., 2020; Mastroianni et al., 2021; Wolf et al., 2021).

Relational pessimism is not just a feature of conversations with strangers; it persists even when people interact with close others and as they engage in a variety of speech acts that build social connection. For example, people tend to underestimate the positive impact of expressing gratitude to others (Kumar, 2022; Kumar & Epley, 2018), giving compliments (Boothby & Bohns, 2021; Zhao & Epley, 2021), asking for help (Zhao & Epley, 2022), and providing social support (Dungan et al., 2022; for related reviews, see Abel et al., 2022; Epley et al., 2022). People, of course, know that prosocial speech can positively affect others, but they nonetheless consistently underestimate the magnitude of this positivity and often drastically overestimate the possibility that these acts will be delivered poorly or received awkwardly.

Pursuing relational goals in conversation is not all compliments and kindness. People must also be prepared to have difficult conversations, such as resolving conflict, providing constructive feedback, broaching sensitive topics, and engaging with diverse views. Once again, people consistently underestimate how well these conversations will go (e.g., Levine & Cohen, 2018; Sanchez et al., 2022; see also Abi-Esber et al., 2022; Collins et al., 2022; Hart et al., 2021; for a review of related ideas, see Levine et al., 2020). Moreover, after such conversations, people appear to display a "thought gap," erroneously believing that they thought more about their conversation partners than their conversation partners thought about them in the minutes, hours, and days following an interaction. This feeling of being alone with one's thoughts is heightened in situations of high emotional salience, such as after an argument with a romantic partner (Cooney et al., 2022; see also Boothby et al., 2017).

In summary, people appear overly pessimistic about their ability to execute the fundamental relational goals of conversation. They avoid conversations that would improve their well-being, doubt their ability to start, maintain, and end such interactions, are overly self-critical during and after conversations, and fail to express prosocial sentiments such as showing sympathy, gratitude, and giving compliments due to a misplaced sense of awkwardness. Why do we harbor such gloomy expectations about our conversations and social abilities?

One factor contributing to this relational pessimism may be the same mechanism that underlies informational optimism: egocentrism. People often project their own insecurities and doubts onto their conversation partners. For example, if someone has concerns about their conversational competence, they may assume their partner values competence over warmth.

They might also project their own self-critical and socially anxious thoughts onto others, even when their conversation partners do not share these same thoughts (for a related discussion, see Epley et al., 2022).

Although egocentrism explains why people project their self-perceptions onto others, it does not fully account for the negative skew of those projections. This negativity may be particularly related to the indirectness that characterizes the relational mode. Consider the contrast between pursuing informational goals and relational goals in conversation. Although *informational* goals benefit from direct listener feedback, which allows for a clear distinction between successful and failed communication (e.g., "Hold on, is the restaurant on the corner of 43rd or 44th street?"), *relational* goals are fundamentally different. They are often governed by indirect norms, which makes it challenging to receive unambiguous feedback. When a stranger agrees to talk, do they genuinely want to be there, or are just being polite? Is a potential romantic partner really enamored, or are they feeling obliged? Colleagues may thank each other for feedback, even if they find it unhelpful, blurring the lines between true appreciation and social niceties. Unlike the directness of information transfer, which encourages feedback to confirm mutual understanding, relational goals are characterized by indirectness, which makes it hard to know where one stands. Of course, people *do* get feedback when they pursue relational goals—words of encouragement, prosodic engagement, eye gaze—but such cues are often ambiguous: A smile after a compliment does not distinguish a genuine compliment from a merely perfunctory one. As a result, people find themselves guessing how their social overtures are being received, and assuming the worst may seem safer than assuming the best.

The relationship between indirectness, uncertainty, and pessimism could be a promising avenue for future research, serving to connect three extensive and growing literatures: (1) the large positive impact of social interaction on people's health and happiness; (2) the pessimistic errors that arise when people reason about the value of social interactions and their ability to navigate them; and (3) the intriguing social cognition that emerges from an ongoing need to coordinate the informational and relational goals of conversation.

Goals Summary

Exploring the goals of conversation could easily become an unwieldy discussion of all the things for which people use conversation. As in the other sections, this section's focus was to identify a core feature of the conversation system, here, its two primary goals: transmitting information and managing relationships. These informational and relational goals are each governed by specific norms characterized by directness and indirectness, respectively.

Like turn-taking and intersubjectivity, conversation's goals and norms require considerable coordination, both within and across individuals. This delicate balance affects everything from how we repair misunderstandings to our most common conversational errors, such as people's tendency to overestimate their ability to transmit information while underestimating their ability to use conversation to forge and maintain social bonds. Moving forward, researchers might continue to explore the underlying logic of informational and relational goals to produce a deeper understanding of the existing tradeoffs. This exploration will ultimately help unravel what sustains the beautiful yet

delicate equilibrium of directness and indirectness that we perpetually negotiate when we talk.

CONCLUDING THOUGHTS

Conversation rides on an ancient system of turn-taking, which can be observed early in infancy and even has intriguing parallels in nonhuman primates and other animals (Bateson, 1975; Levinson & Holler, 2014; Pika et al., 2018; Trevarthen & Aitken, 2001). This turn-taking mechanism, coupled with the exchange of linguistic meaning and the coordination of goals, enables conversation—a uniquely human act of cooperation and a platform for minds to meet. As such, conversation requires an astounding level of coordination across many levels of cognition (e.g., Enfield, 2017; Levinson, 2020; Pickering & Garrod, 2021). Once established, the utility of conversation is hard to overstate. It has been primary technology by which humans throughout history have exchanged information (Grice, 1989), shaped culture (Henrich, 2015), established norms (Hawkins et al., 2019), developed relationships (Dunbar, 1996, 2004), negotiated conflict (Boothby et al., 2023), conducted economic activity (Bartkus & Davis, 2009), and learned about the world (Quine & Ullian, 1970/1978). Finally, the frequency and success of people's conversations have a profound impact on their mental and physical health (e.g., Diener & Seligman, 2002; Forgas et al., 2022; Hawkley & Cacioppo, 2010; Holt-Lunstad et al., 2015, 2017; House et al., 1988; Sun et al., 2020).

Despite the ubiquity and importance of conversation, the science of conversation remains surprisingly underdeveloped. The goal of this chapter was to synthesize disparate literatures, and in doing so, identify the core features of the conversation system, often described using words such as "coordinate," "cooperate," and "collaborate." To converse, people must ordinate, operate, and laborate *together*. This togetherness transcends mere presence, as is captured in the most common analogies used to describe conversation, such as a duet or a waltz (e.g., Clark, 1996). However, even these deeply collaborative activities undersell the depth of the conversation's togetherness, which is the most intricate, multimodal, and synchronized dance that humans perform.

This complex togetherness requires exquisite coordination at multiple levels, which give rise to the unique structure and experience of conversation: the highly organized yet flexible back-and-forth dance of turn-taking; the unconsciously primed yet consciously managed construction of intersubjectivity; and, finally, the integration of informational and relational goals that speakers and listeners must interweave during every moment of talk.

The Past, The Future, And Technology

Beyond understanding the core features of the conversation system, the science of conversation must also plunge deeper into the complex dynamics of moment-to-moment

behavior. This endeavor will rely heavily on new and still developing technologies, many of which go beyond the scope of any single review. Nevertheless, since the earliest days of conversation research, scholars without a shared background, method, or set of terms have displayed a fascination with the raw data of interaction. In reading this earlier work, one senses the excitement at the possibility of using new technology to record and analyze vast amounts of detailed conversational behavior. Today, that same excitement can be felt, but with the availability of advanced computational techniques and emerging developments increasingly at hand.

First, consider features closest to the raw signal of the audio, video, and text of a conversation (e.g., the words uttered, the timing of speech and silence, pitch, eye gaze, contractions of facial muscles, etc.). These features vary on a continuous timescale, and although they often require some degree of inference to generate, such as the linguistic inferences made by an automated transcription service, it can be argued that they constitute something like the "objective" properties of conversation. Due to the proliferation of video-mediated conversation and growing interest in developing conversational AI, researchers are now processing, extracting, and organizing these features at scale—a boon for conversation research (Cooney & Reece, 2024; Reece et al., 2023).

Second, moving beyond lower-level features, the study of conversation requires the extraction of what might be considered "mid-level" features that vary at an intermediate timescale and are meant to capture meaningful intra-conversational psychology: a rising vocal intensity, an expression of sadness, a timely change of subject. Small-scale research could, with some success, use coders to label such features, but increasingly, this level of analysis is being examined with the help of algorithmic tools that are trained to attend to specific aspects of speech, sound, and movement to infer meaning. Some examples are language and signal-processing algorithms such as Word2vec, GloVe, ELMo, BERT, USE, and GPT (e.g., Cer et al., 2018; Devlin et al., 2018; Mikolov et al., 2013; Pennington et al., 2014; Peters et al., 2018; Radford et al., 2018), along with purpose-built "feature detectors" across text, audio, and visual modalities, such as those capturing politeness (Yeomans et al., 2018), emotion expression (Mollahosseini et al., 2019), vocal intensity (Livingstone & Russo, 2018), and vocal tone (Knox & Lucas, 2021). These feature detectors are usually trained on labeled datasets and then applied at scale. Although there are still many useful feature detectors missing from the literature, such as laughter and self-disclosure, these gaps are rapidly closing, which will accelerate research into the moment-to-moment flow of talk.

Third, and finally, consider what might be termed "high-level" features of conversation, which encompass individuals' subjective judgments about their conversations, are formed on a coarse timescale, and are usually reflected in post-conversation survey responses. These subjective assessments include factors such as enjoyment, conversational flow, and perceptions of warmth and intelligence. As scholars from across the social sciences show increasing interest in conversation, we will surely see a greater emphasis on methods and large-scale data collection, especially those that permit sophisticated links between post-conversation outcomes and intra-conversational dynamics (Reece et al., 2023). For

example, research has linked patterns of question-asking behavior in conversation to people's overall perceptions of liking (Huang et al., 2017; Yeomans et al., 2019); and other studies have linked the speed of turn-taking to overall conversational enjoyment (Templeton & Wheatley, 2023). This research links the dynamics of "lower-level" turn-by-turn conversation with "higher-level" outcomes of interest, mirroring the structure of pioneering research on how listeners' backchannels influence the success of speakers' narratives (Bavelas et al., 2000). Overall, this "vertically integrated" research, which is genuinely committed to bridging levels of analysis and analyzing real face-to-face conversation, has been all too rare across the social sciences, particularly in social psychology. However, it will likely become increasingly popular and important in the coming years (Yeomans et al., 2023).

At this point in a chapter on conversation, it has become something of a trope to highlight the complexity of conversation by presenting a simple conversational task that Artificial Intelligence (AI) cannot yet perform. This chapter refrains from doing so, and for the AI reading this, we apologize for the limited imagination of our species. Nevertheless, it is clear that real human-like conversation continues to pose a challenge. For AI to engage in successful conversation, it will need to exhibit considerable skill across multiple communicative channels and levels of abstraction. This entails not only choosing words that respond to the previous turn but also exchanging turns with precision, grace, and social savvy. Further, such an AI must be able to build intersubjectivity and execute social goals even without, presumably, *feeling* the same phenomenology or motivation as humans. This complexity is the reason why conversational AI—AI capable of seamlessly taking turns, building shared conceptual spaces, and juggling goals in real-time across various linguistic and paralinguistic channels—remains an unsolved challenge (Dinan et al., 2020; Ram et al., 2018; Roller et al., 2021). Undoubtedly, successful conversational AI will also be a game-changer for the experimental science of conversation. Imagine the ability to use a perfectly trained conversation AI as a confederate, allowing researchers to exert precise control, manipulation, and measurement over every aspect of conversational behavior, thus enabling elusive experimental work on many big unanswered questions.

It is worth noting that perhaps the biggest unanswered question that this chapter did not address is the difference between dyadic and group conversation. Although the archetype of a conversation may be between two people, a considerable portion of our lives is spent talking in groups. And as any person who has talked in groups can attest, everything that this chapter has explored—be it turn-taking, intersubjectivity, or the pursuit of goals—has the potential to change when more voices, minds, and preferences are added to the conversation. Unfortunately, these differences, which one can feel intuitively, have been scarcely examined (Cooney et al., 2020; Moreland, 2010; Stivers, 2021). May some intrepid researchers take up the heavy mantle of investigating how group dynamics shape the intricate coordination that defines conversation and develop the technological innovations necessary to do so.

Finally, progress in conversation science has historically gone hand in hand with the availability of large public datasets. Examples include the Map Task corpus (Anderson et

al., 1991), the Switchboard corpus (Godfrey et al., 1992), large text corpora like the Fisher corpus (Cieri et al., 2004), or newer multimodal datasets, such as the MELD (Poria et al., 2019; see also Chen et al., 2018) and OMG-Empathy datasets (Barros et al., 2019). These vast resources will continue to become more accessible, as exemplified by the CANDOR corpus, which consists of thousands of unscripted conversations totaling over 7 million words and 850+ hours of audio and video, along with hundreds of extracted features and detailed post-conversation measures (Reece et al., 2023; for a review of relevant datasets, see Serban et al., 2015).

Any attempt to fully review the computational developments related to conversation will be incomplete, and likely obsolete by the time this chapter appears. Nonetheless, amidst the complexity of low-level, moment-by-moment behavior, people's mid-level turn-by-turn perceptions of talk emerge, ultimately leading to their high-level conversation-by-conversation perceptions of the interaction overall. This integration of all levels of analysis across text, audio, and visual modalities will be a central focus of future conversation research. Such an endeavor will be aided by developments in data science, machine learning, and artificial intelligence (e.g., Reece et al., 2023; Wheatley et al., 2024). Indeed, one aim of this chapter has been to offer a conceptual, theoretical, and historical framework to accompany these exciting technological developments, the pace of which can sometimes overshadow the larger narrative of conversation research.

The history of conversation research is spread across multiple disciplines, and it is the main contention of this chapter that planting some of the core features here in one place will help the science grow. At present, the literature is constrained by a lack of experimental studies, research that remains within disciplinary silos, and the complexity of the subject matter—all of which resist overarching theorizing and comprehensive conceptual models. We hope that this gathering together of the research and the resulting cross-pollination of ideas will lead to a bloom of foundational findings and new theories, as an increasing number of scholars take conversation as their primary object of inquiry.

Coda

To borrow an observation from Wilbur Schramm, conversation is “an academic crossroad where many have passed, but few have tarried” (1963, p. 2). Yet, in recent times, a growing number of researchers appear to be setting up camp there, drawn by the allure of this unique crossroads, and wondering, like pioneers in a new land, who their fellow settlers are and how this encampment will eventually evolve into a thriving scientific community.

In trying to advance the science of conversation, researchers might simply follow their noses, but it is far wiser to develop a master plan. This chapter does not claim to provide such a plan, but it does offer a starting point by highlighting the importance for conversation researchers to continually situate themselves within the broader narrative of conversation scholarship, which is currently scattered across disciplines. A robust science of conversation requires collaboration, and perhaps as researchers linger and

work together, they may even tarry together long enough that the insights from various disciplines will start to intermingle in each person's mind, leading to a fuller appreciation of the vast landscape that still awaits populating with new research.

We return then to where this chapter began, with Adam Smith's epigram: "The great pleasure of conversation, and indeed of society, arises from a certain correspondence of sentiments and opinions, from a certain harmony of minds, which like so many musical instruments coincide and keep time with one another." This statement was prescient in every way: Sentiments and opinions perfectly presage people's relational and informational goals; a harmony of minds is precisely what intersubjectivity strives for, and the keeping of time is the purview of the turn-taking system. These are the forms of coordinated togetherness that lie at the very heart of the conversation system.

The word "conversation" derives from the Latin "con" and "versare," meaning "to turn together." In conversation, we indeed turn together, literally by taking turns; in our minds as we create intersubjectivity; and, finally, in pursuit of our basic goals. All of this turning requires significant coordination, about which this chapter said much while also acknowledging that there is still so much more to say. Fortunately, there is no shortage of talk to examine; for as long as the earth keeps turning, people will turn together, billions of times a day. Around and around, we go—taking turns, sharing minds, and coordinating goals—engaged in that ancient human ritual, so precious to our species, known as conversation.

AUTHOR NOTE

We are indebted to all those who provided feedback on previous drafts: Saul Albert, Ruth Corps, Mark Dingemans, Nick Epley, Riccardo Fusaroli, Rod Gardner, Dan Gilbert, Robert Hawkins, Sid Horton, Lilla Magyari, Martin Pickering, Maya Rossignac-Milon, Tanya Stivers, Arjen Stolk, and Becky Warner—thank you for contributing to this conversation. We also extend our thanks to Pedro Gallino, Sarah Gorbach, Evelyn Huang, George James, and Amelia Scharff for their assistance.

REFERENCES

Abel, J. E., Vani, P., Abi-Esber, N., Blunden, H., & Schroeder, J. (2022). Kindness in short supply: Evidence for inadequate prosocial input. *Current Opinion in Psychology*, 48, 101458. <https://doi.org/10.1016/j.copsyc.2022.101458>

Abele, A. E., & Wojciszke, B. (Eds.). (2018). *The agency-communion framework*. Routledge. <https://doi.org/10.4324/9780203703663>

Abi-Esber, N., Abel, J. E., Schroeder, J., & Gino, F. (2022). "Just letting you know ..."
Underestimating others' desire for constructive feedback. *Journal of Personality and*

- Social Psychology*, 123(6), 1362–1385. <https://doi.org/10.1037/pspi0000393>
- Albert, S., & de Ruiter, J. P. (2018). Repair: The interface between interaction and cognition. *Topics in Cognitive Science*, 10(2), 279–313. <https://doi.org/10.1111/tops.12339>
- Albert, S., & Kessler, S. (1976). Processes for ending social encounters: The conceptual archaeology of a temporal place. *Journal for the Theory of Social Behaviour*, 6(2), 147–170. <https://doi.org/10.1111/j.1468-5914.1976.tb00363.x>
- Allport, G. W. (1937). *Personality: A psychological interpretation*. Holt.
- Anders, S., Heinzle, J., Weiskopf, N., Ethofer, T., & Haynes, J. D. (2011). Flow of affective information between communicating brains. *Neuroimage*, 54(1), 439–446. <https://doi.org/10.1016/j.neuroimage.2010.07.004>
- Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The HCRC Map Task corpus. *Language and Speech*, 34(4), 351–366. <https://doi.org/10.1177/002383099103400404>
- Austin, J. L. (1975). *How to do things with words*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198245537.001.0001> (Original work published 1962)
- Babcock, M. J., Ta, V. P., & Ickes, W. (2014). Latent semantic similarity and language style matching in initial dyadic interactions. *Journal of Language and Social Psychology*, 33(1), 78–88. <https://doi.org/10.1177/0261927x13499331>
- Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. Rand McNally.
- Bales, R. F. (1950). *Interaction process analysis; a method for the study of small groups*. Addison-Wesley.
- Bales, R. F. (1970). *Personality and interpersonal behavior*. Holt, Rinehart and Winston.
- Bales, R. F., Strodtbeck, F. L., Mills, T. M., & Roseborough, M. E. (1951). Channels of communication in small groups. *American Sociological Review*, 16(4), 461–468. <https://doi.org/10.2307/2088276>
- Ball, P. (1975). Listeners' responses to filled pauses in relation to floor apportionment. *British Journal of Social and Clinical Psychology*, 14(4), 423–424. <https://doi.org/10.1111/j.2044-8260.1975.tb00198.x>

- Bangerter, A., Chevalley, E., & Derouwau, S. (2010). Managing third-party interruptions in conversations: Effects of duration and conversational role. *Journal of Language and Social Psychology*, 29(2), 235–244. <https://doi.org/10.1177/0261927x09359591>
- Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, 27(2), 195–225. https://doi.org/10.1207/s15516709cog2702_3
- Bangerter, A., Clark, H. H., & Katz, A. R. (2004). Navigating joint projects in telephone conversations. *Discourse Processes*, 37(1), 1–23. https://doi.org/10.1207/s15326950dp3701_1
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54(7), 462–479. <https://doi.org/10.1037/0003-066x.54.7.462>
- Barr, D. J., & Keysar, B. (2007). Perspective taking and the coordination of meaning in language use. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 901–938). Academic Press. <https://doi.org/10.1016/b978-012369374-7/50024-9>
- Barros, P., Churamani, N., Lim, A., & Wermter, S. (2019). The OMG-Empathy dataset: Evaluating the impact of affective behavior in storytelling. In *2019 8th international conference on affective computing and intelligent interaction (ACII)* (pp. 1–7). IEEE. <https://doi.org/10.1109/acii.2019.8925530>
- Barthel, M., Meyer, A. S., & Levinson, S. C. (2017). Next speakers plan their turn early and speak after turn-final “go-signals.” *Frontiers in Psychology*, 8, 393. <https://doi.org/10.3389/fpsyg.2017.00393>
- Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology*, 7, 1858. <https://doi.org/10.3389/fpsyg.2016.01858>
- Bartkus, V. O., & Davis, J. H. (Eds.). (2009). *Social capital: Reaching out, reaching in*. Edward Elgar Publishing. <https://doi.org/10.4337/9781848445963>
- Bass, B. M. (1949). An analysis of the leaderless group discussion. *Journal of Applied Psychology*, 33(6), 527–533. <https://doi.org/10.1037/h0058164>
- Bass, B. M., & Stogdill, R. M. (1990). *Bass & Stogdill's handbook of leadership: Theory, research, and managerial applications*. Simon and Schuster.
- Bateson, M. C. (1975). Mother-infant exchanges: The epigenesis of conversational interaction. *Annals of the New York Academy of Sciences*, 263(1), 101–113. <https://doi.org/10.1111/j.1749-6632.1975.tb41575.x>

- Bavelas, J. B. (2007). Face-to-face dialogue as a micro-social context. In S. Duncan, E. Levy, & J. Cassell (Eds.), *Language in mind, body, and context* (pp. 127–146). John Benjamins Publishing Company. <https://doi.org/10.1075/gs.1.15bav>
- Bavelas, J. B. (2022). *Face-to-face dialogue: Theory, research, and applications*. Oxford University Press. <https://doi.org/10.1093/oso/9780190913366.001.0001>
- Bavelas, J. B., Black, A., Lemery, C. R., & Mullett, J. (1986). "I show how you feel": Motor mimicry as a communicative act. *Journal of Personality and Social Psychology*, 50(2), 322–329. <https://doi.org/10.1037/0022-3514.50.2.322>
- Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, 79(6), 941–952. <https://doi.org/10.1037/0022-3514.79.6.941>
- Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, 52(3), 566–580. <https://doi.org/10.1111/j.1460-2466.2002.tb02562.x>
- Beach, W. A., & Lindstrom, A. K. (1992). Conversational universals and comparative theory: Turning to Swedish and American acknowledgment tokens in interaction. *Communication Theory*, 2(1), 24–49. <https://doi.org/10.1111/j.1468-2885.1992.tb00027.x>
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2), 145–204. <https://doi.org/10.1017/s004740450001037x>
- Bergey, C.A., & DeDeo, S. (2024). From "um" to "yeah": Producing, predicting, and regulating information flow in human conversation. arXiv:2403.08890 [cs.CL]. <https://doi.org/10.48550/arXiv.2403.08890>
- Bernieri, F., & Rosenthal, R. (1991). Interpersonal coordination: Behavioral matching interactional synchrony. In R. S. Feldman & B. Rime (Eds.), *Foundations of nonverbal behavior* (pp. 401–432). Cambridge University Press.
- Beukeboom, C. J. (2009). When words feel right: How affective expressions of listeners change a speaker's language use. *European Journal of Social Psychology*, 39(5), 747–756. <https://doi.org/10.1002/ejsp.572>
- Bögels, S. (2020). Neural correlates of turn-taking in the wild: Response planning starts early in free interviews. *Cognition*, 203, 104347. <https://doi.org/10.1016/j.cognition.2020.104347>
- Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, 52, 46–57.

<https://doi.org/10.1016/j.wocn.2015.04.004>

Bögels, S., & Torreira, F. (2021). Turn-end estimation in conversational turn-taking: The roles of context and prosody. *Discourse Processes*, 58(10), 903–924.

<https://doi.org/10.1080/0163853x.2021.1986664>

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never say no... How the brain interprets the pregnant pause in conversation. *PLOS One*, 10(12), e0145474.

<https://doi.org/10.1371/journal.pone.0145474>

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2020). Conversational expectations get revised as response latencies unfold. *Language, Cognition and Neuroscience*, 35(6), 766–779.

<https://doi.org/10.1080/23273798.2019.1590609>

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, 5(1), 12881.

<https://doi.org/10.1038/srep12881>

Boothby, E. J., & Bohns, V. K. (2021). Why a simple act of kindness is not as simple as it seems: Underestimating the positive impact of our compliments on others. *Personality and Social Psychology Bulletin*, 47(5), 826–840.

<https://doi.org/10.1177/0146167220949003>

Boothby, E. J., Clark, M. S., & Bargh, J. A. (2014). Shared Experiences Are Amplified.

Psychological Science, 25(12), 2209–2216. <https://doi.org/10.1177/0956797614551162>

Boothby, E. J., Clark, M. S., & Bargh, J. A. (2017). The invisibility cloak illusion: People (incorrectly) believe they observe others more than others observe them. *Journal of Personality and Social Psychology*, 112(4), 589–606.

<https://doi.org/10.1037/pspi0000082>

Boothby, E. J., Cooney, G., Sandstrom, G. M., & Clark, M. S. (2018). The liking gap in conversations: Do people like us more than we think?. *Psychological Science*, 29(11), 1742–1756.

<https://doi.org/10.1177/0956797618783714>

Boothby, E. J., Cooney, G., & Schweitzer, M. E. (2023). Embracing complexity: A review of negotiation research. *Annual Review of Psychology*, 74, 299–332.

<https://doi.org/10.1146/annurev-psych-033020-014116>

Boothby, E. J., Smith, L. K., Clark, M. S., & Bargh, J. A. (2016). Psychological Distance Moderates the Amplification of Shared Experience. *Personality and Social Psychology Bulletin*, 42(10), 1431–1444.

<https://doi.org/10.1177/0146167216662869>

Borgatta, E. F., & Bales, R. F. (1956). Sociometric status patterns and characteristics of interaction. *The Journal of Social Psychology*, 43(2), 289–297.

<https://doi.org/10.1080/00224545.1956.9919225>

- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2), B13–B25. [https://doi.org/10.1016/s0010-0277\(99\)00081-5](https://doi.org/10.1016/s0010-0277(99)00081-5)
- Brehm, L., & Meyer, A. S. (2021). Planning when to say: Dissociating cue use in utterance initiation using cross-validation. *Journal of Experimental Psychology: General*, 150(9), 1772–1799. <https://doi.org/10.1037/xge0001012>
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482. <https://doi.org/10.1037/0278-7393.22.6.1482>
- Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, 1(2), 274–291. <https://doi.org/10.1111/j.1756-8765.2009.01019.x>
- Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, 34(3), 383–398. <https://doi.org/10.1006/jmla.1995.1017>
- Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. *Psychology of Learning and Motivation*, 53, 301–344. [https://doi.org/10.1016/S0079-7421\(10\)53008-1](https://doi.org/10.1016/S0079-7421(10)53008-1)
- Brescoll, V. L. (2011). Who takes the floor and why: Gender, power, and volubility in organizations. *Administrative Science Quarterly*, 56(4), 622–641. <https://doi.org/10.1177/0001839212439994>
- Broth, M., & Mondada, L. (2013). Walking away: The embodied achievement of activity closings in mobile interaction. *Journal of Pragmatics*, 47(1), 41–58. <https://doi.org/10.1016/j.pragma.2012.11.016>
- Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511805226>
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511813085>
- Bruner, J. (1983). Play, thought, and language. *Peabody Journal of Education*, 60(3), 60–69. <https://doi.org/10.1080/01619568309538407>
- Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, 97(5), 1232–1254. <https://doi.org/10.1086/261651>

- Cappella, J. N. (1981). Mutual influence in expressive behavior: Adult-adult and infant-adult dyadic interaction. *Psychological Bulletin*, 89(1), 101–132. <https://doi.org/10.1037/0033-2909.89.1.101>
- Cappella, J. N., & Planalp, S. (1981). Talk and silence sequences in informal conversations III: Interspeaker influence. *Human Communication Research*, 7(2), 117–132. <https://doi.org/10.1111/j.1468-2958.1981.tb00564.x>
- Cappella, J. N., & Streibel, M. J. (1979). Computer analysis of talk-silence sequences: The FIASSCO system. *Behavior Research Methods & Instrumentation*, 11(3), 384–392. <https://doi.org/10.3758/bf03205681>
- Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsón, H., & Yan, H. (1999). Embodiment in conversational interfaces: Rea. *CHI'99 conference proceedings* (pp. 520–527). Association for Computing Machinery. <https://doi.org/10.1145/302979.303150>
- Cassotta, L., Feldstein, S., & Jaffe, J. (1964). AVTA: A device for automatic vocal transaction analysis. *Journal of the Experimental Analysis of Behavior*, 7(1), 99–104. <https://doi.org/10.1901/jeab.1964.7-99>
- Cassotta, L., Feldstein, S., and Jaffe, J. (1967). The stability and modifiability of individual vocal characteristics in stress and nonstress interviews. *Research Bulletin No. 2*. William Alanson White Institute.
- Castellucci, G. A., Kovach, C. K., Howard, M. A., III, Greenlee, J. D., & Long, M. A. (2022). A speech planning network for interactive language use. *Nature*, 602(7895), 117–122. <https://doi.org/10.1038/s41586-021-04270-z>
- Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y. H., & Kurzweil, R. (2018). Universal sentence encoder. arXiv:1803.11175 [cs.CL]. <https://doi.org/10.48550/arXiv.1803.11175>
- Chapple, E. D. (1939). Quantitative analysis of the interaction of individuals. *Proceedings of the National Academy of Sciences*, 25(2), 58–67. <https://doi.org/10.1073/pnas.25.2.58>
- Chapple, E. D. (1940). "Personality" differences as described by invariant properties of individuals in interaction. *Proceedings of the National Academy of Sciences*, 26(1), 10–16. <https://doi.org/10.1073/pnas.26.1.10>
- Chapple, E. D. (1949). The interaction chronograph: Its evolution and present application. *Personnel*, 25, 295–307.
- Chapple, E. D. (1970). *Culture and biological man: Explorations in behavioural anthropology*. Holt, Rinehart and Winston.

- Chapple, E. D., & Donald, G., Jr. (1946). A method for evaluating supervisory personnel. *Harvard Business Review*, 24(2), 197–214.
- Chapple, E., Chapple, M., & Repp, J. (1954). Behavioral definitions of personality and temperament characteristics. *Human Organization*, 13(4), 34–39.
<https://doi.org/10.17730/humo.13.4.e5t531u645200253>
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893–910.
<https://doi.org/10.1037/0022-3514.76.6.893>
- Chaudhry, S. J., & Loewenstein, G. (2019). Thanking, apologizing, bragging, and blaming: Responsibility exchange theory and the currency of communication. *Psychological Review*, 126(3), 313–344. <https://doi.org/10.1037/rev0000139>
- Chen, S. Y., Hsu, C. C., Kuo, C. C., & Ku, L. W. (2018). Emotionlines: An emotion corpus of multi-party conversations. arXiv:1802.08379 [cs.CL].
<https://doi.org/10.48550/arXiv.1802.08379>
- Chevalley, E., & Bangerter, A. (2010). Suspending and reinstating joint activities with dialogue. *Discourse Processes*, 47(4), 263–291.
<https://doi.org/10.1080/01638530902959935>
- Cieri, C., Miller, D., & Walker, K. (2004). The Fisher corpus: A resource for the next generations of speech-to-text. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of the fourth international conference on language resources and evaluation (LREC'04)* (pp. 69–71). European Language Resources Association (ELRA).
- Clancy, P. M., Thompson, S. A., Suzuki, R., & Tao, H. (1996). The conversational use of reactive tokens in English, Japanese, and Mandarin. *Journal of Pragmatics*, 26(3), 355–387. [https://doi.org/10.1016/0378-2166\(95\)00036-4](https://doi.org/10.1016/0378-2166(95)00036-4)
- Clark, H. H. (1996). *Using language*. Cambridge University Press.
- Clark, H. H. (2020). Social actions, social commitments. In N. J. Enfield, & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and human interaction* (pp. 126–150). Routledge. <https://doi.org/10.4324/9781003135517-6> (Original work published in 2006)
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association. <https://doi.org/10.1037/10096-006>

- Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, 58(2) 332–373. <https://doi.org/10.1353/lan.1982.0042>
- Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73–111. [https://doi.org/10.1016/s0010-0277\(02\)00017-3](https://doi.org/10.1016/s0010-0277(02)00017-3)
- Clark, H. H., & French, J. W. (1981). Telephone goodbyes. *Language in Society*, 10(1), 1–19. <https://doi.org/10.1017/s0047404500008393>
- Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1), 62–81. <https://doi.org/10.1016/j.jml.2003.08.004>
- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge University Press.
- Clark, H. H., & Schunk, D. H. (1980). Polite responses to polite requests. *Cognition*, 8(2), 111–143. [https://doi.org/10.1016/0010-0277\(80\)90009-8](https://doi.org/10.1016/0010-0277(80)90009-8)
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1–39. [https://doi.org/10.1016/0010-0277\(86\)90010-7](https://doi.org/10.1016/0010-0277(86)90010-7)
- Clayman, S. E. (2002). Sequence and solidarity. In S. R. Thye & E. J. Lawler (Eds.), *Group cohesion, trust and solidarity* (pp. 229–253). Elsevier Science. [https://doi.org/10.1016/s0882-6145\(02\)19009-6](https://doi.org/10.1016/s0882-6145(02)19009-6)
- Clayman, S. E., Heritage, J., & Maynard, D. W. (2022). The ethnomethodological lineage of conversation analysis. In D. W. Maynard & J. Heritage (Eds.), *The ethnomethodology program: Legacies and prospects* (pp. 252–286). Oxford Academic. <https://doi.org/10.1093/oso/9780190854409.003.0010>
- Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, 49(2), 214–230. [https://doi.org/10.1016/s0749-596x\(03\)00060-3](https://doi.org/10.1016/s0749-596x(03)00060-3)
- Collins, H. K. (2022). When listening is spoken. *Current Opinion in Psychology*, 47, 101402. <https://doi.org/10.1016/j.copsyc.2022.101402>
- Collins, H. K., Dorison, C. A., Gino, F., & Minson, J. A. (2022). Underestimating counterparts' learning goals impairs conflictual conversations. *Psychological Science*, 33(10), 1732–1752. <https://doi.org/10.1177/09567976221085494>

- Cooney, G., Boothby, E. J., & Lee, M. (2022). The thought gap after conversation: Underestimating the frequency of others' thoughts about us. *Journal of Experimental Psychology: General*, *151*(5), 1069–1088. <https://doi.org/10.1037/xge0001134>
- Cooney, G., Gilbert, D. T., & Wilson, T. D. (2014). The unforeseen costs of extraordinary experience. *Psychological Science*, *25*(12), 2259–2265. <https://doi.org/10.1177/0956797614551372>
- Cooney, G., Gilbert, D. T., & Wilson, T. D. (2017). The novelty penalty: Why do people like talking about new experiences but hearing about old ones?. *Psychological Science*, *28*(3), 380–394. <https://doi.org/10.1177/0956797616685870>
- Cooney, G., Mastroianni, A. M., Abi-Esber, N., & Brooks, A. W. (2020). The many minds problem: Disclosure in dyadic versus group conversation. *Current Opinion in Psychology*, *31*, 22–27. <https://doi.org/10.1016/j.copsyc.2019.06.032>
- Cooney, G., & Reece, A. G. (2024). NaturalTurn: A Method to Segment Transcripts into Naturalistic Conversational Turns. arXiv:2403.15615 [cs.CL]. <https://doi.org/10.48550/arXiv.2403.15615>
- Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition*, *175*, 77–95. <https://doi.org/10.1016/j.cognition.2018.01.015>
- Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes*, *55*(2), 230–240. <https://doi.org/10.1080/0163853x.2017.1330031>
- Corps, R. E., Gambi, C., & Pickering, M. J. (2020). How do listeners time response articulation when answering questions? The role of speech rate. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(4), 781–802. <https://doi.org/10.1037/xlm0000759>
- Corps, R. E., Knudsen, B., & Meyer, A. S. (2022). Overrated gaps: Inter-speaker gaps provide limited information about the timing of turns in conversation. *Cognition*, *223*, 105037. <https://doi.org/10.1016/j.cognition.2022.105037>
- Corti, K., & Gillespie, A. (2016). Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. *Computers in Human Behavior*, *58*, 431–442. <https://doi.org/10.1016/j.chb.2015.12.039>
- Cottrell, L. S. (1942). The analysis of situational fields in social psychology. *American Sociological Review*, *7*(3), 370–382. <https://doi.org/10.2307/2085366>

- Crown, C. L. (1991). Coordinated interpersonal timing of vision and voice as a function of interpersonal attraction. *Journal of Language and Social Psychology*, 10(1), 29–46. <https://doi.org/10.1177/0261927x91101002>
- Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 37(1), 16–28. <https://doi.org/10.1016/j.wocn.2008.08.003>
- Cummins, F. (2012). Looking for rhythm in speech. *Empirical Musicology Review*, 7(1-2), 28–35. <https://doi.org/10.18061/1811/52976>
- Curhan, J. R., & Pentland, A. (2007). Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology*, 92(3), 802–811. <https://doi.org/10.1037/0021-9010.92.3.802>
- Dabbs, J. M., Jr. (1983). Fourier analysis and the rhythm of conversation. Atlanta: Georgia State University. (ERIC Document Reproduction Service No. ED 222 959).
- Dabbs, J. M., Jr., & Ruback, R. B. (1987). Dimensions of group process: Amount and structure of vocal interaction. In *Advances in experimental social psychology* (Vol. 20, pp. 123–169). Academic Press.
- De Ruiter, J. P., & Albert, S. (2017). An appeal for a methodological fusion of conversation analysis and experimental psychology. *Research on Language and Social Interaction*, 50(1), 90–107. <https://doi.org/10.1080/08351813.2017.1262050>
- De Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science*, 4(2), 232–248. <https://doi.org/10.1111/j.1756-8765.2012.01183.x>
- De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3), 515–535. <https://doi.org/10.1353/lan.2006.0130>
- De Ruiter, J. P., Noordzij, M. L., Newman-Norlund, S., Newman-Norlund, R., Hagoort, P., Levinson, S. C., & Toni, I. (2010). Exploring the cognitive infrastructure of communication. *Interaction Studies*, 11(1), 51–77. <https://doi.org/10.1075/is.11.1.05rui>
- De Stefani, E., & Mondada, L. (2018). Encounters in public space: How acquainted versus unacquainted persons establish social and spatial arrangements. *Research on Language and Social Interaction*, 51(3), 248–270. <https://doi.org/10.1080/08351813.2018.1485230>
- De Vos, C., Torreira, F., & Levinson, S. C. (2015). Turn-timing in signed conversations: Coordinating stroke-to-stroke turn boundaries. *Frontiers in Psychology*, 6, 268.

<https://doi.org/10.3389/fpsyg.2015.00268>

- Degen, J. (2023). The rational speech act framework. *Annual Review of Linguistics*, 9, 519–540. <https://doi.org/10.1146/annurev-linguistics-031220-010811>
- Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A Bayesian approach to “overinformative” referring expressions. *Psychological Review*, 127(4), 591–621. <https://doi.org/10.1037/rev0000186>
- Deng, X. (2008). The use of listener responses in Mandarin Chinese and Australian English conversations. *Pragmatics*, 18(2), 303–328. <https://doi.org/10.1075/prag.18.2.06xud>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805 [cs.CL]. <https://doi.org/10.48550/arXiv.1810.04805>
- Dideriksen, C., Christiansen, M. H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2023). Quantifying the interplay of conversational devices in building mutual understanding. *Journal of Experimental Psychology: General*, 152(3), 864–889. <https://doi.org/10.1037/xge0001301>
- Diener, E., & Seligman, M. E. (2002). Very happy people. *Psychological Science*, 13(1), 81–84. <https://doi.org/10.1111/1467-9280.00415>
- Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, 73(6), 1246–1256. <https://doi.org/10.1037/0022-3514.73.6.1246>
- Dinan, E., Logacheva, V., Malykh, V., Miller, A., Shuster, K., Urbanek, J., Kiela, D., Szlam, A., Serban, I., Lowe, R., Prabhumoye, S., Black, A. W., Rudnicky, A., Williams, J., Pineau, J., Burtsev, M., & Weston, J. (2020). The second conversational intelligence challenge (convai2). In S. Escalera & R. Herbrich (Eds.), *The NeurIPS'18 competition: From machine learning to intelligent conversations* (pp. 187–208). Springer, Cham. https://doi.org/10.1007/978-3-030-29135-8_7
- Dingemanse, M. (2024). Interjections at the Heart of Language. *Annual Review of Linguistics*, 10, 257–277. <https://doi.org/10.1146/annurev-linguistics-031422-124743>
- Dingemanse, M., & Enfield, N. J. (2015). Other-initiated repair across languages: Towards a typology of conversational structures. *Open Linguistics*, 1(1), 96–118. <https://doi.org/10.2478/opli-2014-0007>
- Dingemanse, M., & Enfield, N. J. (2024). Interactive repair and the foundations of language. *Trends in Cognitive Sciences*, 28(1), 30–42. <https://doi.org/10.1016/j.tics.2023.09.003>

- Dingemanse, M., Kendrick, K. H., & Enfield, N. J. (2016). A coding scheme for other-initiated repair across languages. *Open Linguistics*, 2(1), 35–46. <https://doi.org/10.1515/opli-2016-0002>
- Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladdottir, R. S., Kendrick, K. H., Levinson, S.C., Manrique, E., Rossi, G., & Enfield, N. J. (2015). Universal principles in the repair of communication problems. *PloS One*, 10(9), e0136100. <https://doi.org/10.1371/journal.pone.0136100>
- Dittmann, A. T. (1972). Developmental factors in conversational behavior. *Journal of Communication*, 22(4), 404–423. <https://doi.org/10.1111/j.1460-2466.1972.tb00165.x>
- Dittmann, A. T. (1977). The development of conversational behavior. In N. Freedman & S. Grand (Eds.), *Communicative structures and psychic structures: A psychoanalytic interpretation of communication* (pp. 133-147). Plenum Press. <https://doi.org/10.1007/978-1-4757-0492-1>
- Dittmann, A. T., & Llewellyn, L. G. (1967). The phonemic clause as a unit of speech decoding. *Journal of Personality and Social Psychology*, 6(3), 341–349. <https://doi.org/10.1037/h0024739>
- Dittmann, A. T., & Llewellyn, L. G. (1968). Relationship between vocalizations and head nods as listener responses. *Journal of Personality and Social Psychology*, 9(1), 79–84. <https://doi.org/10.1037/h0025722>
- Drummond, K., & Hopper, R. (1993a). Back channels revisited: Acknowledgment tokens and speakership incipency. *Research on Language and Social Interaction*, 26(2), 157–177. https://doi.org/10.1207/s15327973rlsi2602_3
- Drummond, K., & Hopper, R. (1993b). Acknowledgment tokens in series. *Communication Reports*, 6(1), 47–53. <https://doi.org/10.1080/08934219309367561>
- Drummond, K., & Hopper, R. (1993c). Some uses of yeah. *Research on Language and Social Interaction*, 26(2), 203–212. https://doi.org/10.1207/s15327973rlsi2602_6
- Dunbar, R. I. M. (1996). *Grooming, gossip, and the evolution of language*. Harvard University Press.
- Dunbar, R. I. (2004). Gossip in evolutionary perspective. *Review of General Psychology*, 8(2), 100–110. <https://doi.org/10.1037/1089-2680.8.2.100>
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23(2), 283–292. <https://doi.org/10.1037/h0033031>

- Duncan, S. (1974). On the structure of speaker–auditor interaction during speaking turns. *Language in Society*, 3(2), 161–180. <https://doi.org/10.1017/s0047404500004322>
- Duncan, S., & Fiske, D. W. (2015). *Face-to-face interaction. Research, methods, and theory*. Routledge. <https://doi.org/10.4324/9781315660998> (Original work published 1977)
- Duncan, S., Jr., & Niederehe, G. (1974). On signaling that it's your turn to speak. *Journal of Experimental Social Psychology*, 10(3), 234–247. [https://doi.org/10.1016/0022-1031\(74\)90070-5](https://doi.org/10.1016/0022-1031(74)90070-5)
- Dungan, J. A., Munguia Gomez, D. M., & Epley, N. (2022). Too reluctant to reach out: Receiving social support is more positive than expressers expect. *Psychological Science*, 33(8), 1300–1312. <https://doi.org/10.1177/09567976221082942>
- Dunn, E. W., & Lok, I. (2022). Can sociability be increased?. In J. P. Forgas, W. Crano, & K. Fiedler (Eds.), *The psychology of sociability* (pp. 98–115). Routledge. <https://doi.org/10.4324/9781003258582-8>
- Duran, N. D., & Fusaroli, R. (2017). Conversing with a devil's advocate: Interpersonal coordination in deception and disagreement. *PloS One*, 12(6), e0178140. <https://doi.org/10.1371/journal.pone.0178140>
- Echterhoff, G. (2012). Shared-reality theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology: Volume 2* (Vol. 2, pp. 180–199). SAGE Publications Ltd.
- Echterhoff, G., & Higgins, E. T. (2018). Shared reality: Construct and mechanisms. *Current Opinion in Psychology*, 23, iv–vii. <https://doi.org/10.1016/j.copsyc.2018.09.003>
- Echterhoff, G., Higgins, E. T., & Groll, S. (2005). Audience-tuning effects on memory: The role of shared reality. *Journal of Personality and Social Psychology*, 89(3), 257–276. <https://doi.org/10.1037/0022-3514.89.3.257>
- Echterhoff, G., Higgins, E. T., & Levine, J. M. (2009). Shared reality: Experiencing commonality with others' inner states about the world. *Perspectives on Psychological Science*, 4(5), 496–521. <https://doi.org/10.1111/j.1745-6924.2009.01161.x>
- Echterhoff, G., Higgins, E. T., Kopietz, R., & Groll, S. (2008). How communication goals determine when audience tuning biases memory. *Journal of Experimental Psychology: General*, 137(1), 3–21. <https://doi.org/10.1037/0096-3445.137.1.3>
- Echterhoff, G., & Schmalbach, B. (2018). How shared reality is created in interpersonal communication. *Current Opinion in Psychology*, 23, 57–61.

Edlund, J., Heldner, M., Al Moubayed, S., Gravano, A., Hirschberg, J. (2010). Very short utterances in conversation. In *Proceedings from Fonetik 2010, Lund, June 2–4, 2010*. (pp. 11–16). Lund University.

Elsaadawy, N., & Carlson, E. N. (2022). Do you make a better or worse impression than you think? *Journal of Personality and Social Psychology*, 123(6), 1407–1420.
<https://doi.org/10.1037/pspp0000434>

Enfield, N. J. (2017). *How we talk: The inner workings of conversation*. Basic Books.

Enfield, N. J., & Levinson, S. C. (Eds.). (2020). *Roots of human sociality: Culture, cognition and interaction*. Routledge. <https://doi.org/10.4324/9781003135517> (Original work published 2006)

Enfield, N. J., & Sidnell, J. (2022). *Consequences of language: From primary to enhanced intersubjectivity*. MIT Press. <https://doi.org/10.7551/mitpress/14795.001.0001>

Enfield, N. J., Stivers, T., & Levinson, S.C. (2010). Question-response sequences in conversation across ten languages: An introduction. *Journal of Pragmatics*, 42(10), 2615–2619. <https://doi.org/10.1016/j.pragma.2010.04.001>

Epley, N., Kardas, M., Zhao, X., Atir, S., & Schroeder, J. (2022). Undersociality: Miscalibrated social cognition can inhibit social connection. *Trends in Cognitive Sciences*, 26(5), 406–418. <https://doi.org/10.1016/j.tics.2022.02.007>

Epley, N., Kumar, A., Dungan, J., & Echelbarger, M. (2023). A prosociality paradox: How miscalibrated social cognition creates a misplaced barrier to prosocial action. *Current Directions in Psychological Science*, 32(1), 33–41.
<https://doi.org/10.1177/09637214221128016>

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, 40(6), 760–768. <https://doi.org/10.1016/j.jesp.2004.02.002>

Epley, N., & Schroeder, J. (2014). Mistakenly seeking solitude. *Journal of Experimental Psychology: General*, 143(5), 1980–1999. <https://doi.org/10.1037/a0037323>

Fay, N., Walker, B., Swoboda, N., Umata, I., Fukaya, T., Katagiri, Y., & Garrod, S. (2018). Universal principles of human communication: Preliminary evidence from a cross-cultural communication game. *Cognitive Science*, 42(7), 2397–2413.
<https://doi.org/10.1111/cogs.12664>

Feldstein, S. (1972). Temporal patterns of dialogue. In A. W. Siegman & B. Pope (Eds.) *Studies in dyadic communication* (pp. 91–113). Pergamon General Psychology Series.
<https://doi.org/10.1016/b978-0-08-015867-9.50009-5>

- Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, 99(4), 689–723.
<https://doi.org/10.1037/0033-295X.99.4.689>
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83.
<https://doi.org/10.1016/j.tics.2006.11.005>
- Forgas, J. P., Crano, W., & Fiedler, K. (Eds.). (2022). *The psychology of sociability: Understanding human attachment*. Routledge.
<https://doi.org/10.4324/9781003258582>
- Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes*, 27(1), 35–53. <https://doi.org/10.1080/01638539909545049>
- Fox Tree, J. E., & Mayer, S. A. (2008). Overhearing single and multiple perspectives. *Discourse Processes*, 45(2), 160–179.
<https://doi.org/10.1080/01638530701792867>
- Francik, E. P., & Clark, H. H. (1985). How to make requests that overcome obstacles to compliance. *Journal of Memory and Language*, 24(5), 560–568.
[https://doi.org/10.1016/0749-596x\(85\)90046-4](https://doi.org/10.1016/0749-596x(85)90046-4)
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084), 998–998. <https://doi.org/10.1126/science.1218633>
- Fraser, B. (1990). Perspectives on politeness. *Journal of pragmatics*, 14(2), 219–236.
[https://doi.org/10.1016/0378-2166\(90\)90081-n](https://doi.org/10.1016/0378-2166(90)90081-n)
- Fries, C. C. (1952). *The structure of English. An introduction to the construction of English sentences*. Harcourt, Brace.
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4), 531–534.
<https://doi.org/10.1016/j.neuron.2006.05.001>
- Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science*, 23(8), 931–939. <https://doi.org/10.1177/0956797612436816>
- 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. *Journal of Experimental Psychology: Human Perception and Performance*, 42(9), 1297–1310. <https://doi.org/10.1037/xhp0000207>

- Fusaroli, R., Raćaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32, 147–157.
<https://doi.org/10.1016/j.newideapsych.2013.03.005>
- Fussell, S. R., & Krauss, R. M. (1989a). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, 25(3), 203–219.
[https://doi.org/10.1016/0022-1031\(89\)90019-x](https://doi.org/10.1016/0022-1031(89)90019-x)
- Fussell, S. R., & Krauss, R. M. (1989b). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, 19(6), 509–525. <https://doi.org/10.1002/ejsp.2420190603>
- Fussell, S. R., & Krauss, R. M. (1991). Accuracy and bias in estimates of others' knowledge. *European Journal of Social Psychology*, 21(5), 445–454.
<https://doi.org/10.1002/ejsp.2420210507>
- Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, 62(3), 378–391. <https://doi.org/10.1037/0022-3514.62.3.378>
- Galantucci, B., Langstein, B., Spivack, E., & Paley, N. (2020). Repair avoidance: When faithful informational exchanges don't matter that much. *Cognitive Science*, 44(10), e12882. <https://doi.org/10.1111/cogs.12882>
- Galantucci, B., & Roberts, G. (2014). Do we notice when communication goes awry? An investigation of people's sensitivity to coherence in spontaneous conversation. *PloS One*, 9(7), e103182. <https://doi.org/10.1371/journal.pone.0103182>
- Galantucci, B., Roberts, G., & Langstein, B. (2018). Content deafness: When coherent talk just doesn't matter. *Language & Communication*, 61, 29–34.
<https://doi.org/10.1016/j.langcom.2018.01.001>
- Gardner, R. (1998). Between speaking and listening: The vocalisation of understandings. *Applied Linguistics*, 19(2), 204–224. <https://doi.org/10.1093/applin/19.2.204>
- Gardner, R. (2001). *When listeners talk: Response tokens and listener stance*. John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.92>
- Garfinkel, H., & Sacks, H. (1986). On formal structures of practical actions. In H. Garfinkel (Ed.), *Ethnomethodological studies of work* (pp. 157–189). Routledge.
<https://doi.org/10.4324/9780203996867> (Reprinted from *Theoretical sociology*, pp. 337–366, by J. D. McKinney, & E. A. Tiryakian, Eds., 1970, Appleton-Century Crofts)

- Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3), 181–215. [https://doi.org/10.1016/0010-0277\(94\)90048-5](https://doi.org/10.1016/0010-0277(94)90048-5)
- Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy?. *Trends in Cognitive Sciences*, 8(1), 8–11. <https://doi.org/10.1016/j.tics.2003.10.016>
- Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2), 292–304. <https://doi.org/10.1111/j.1756-8765.2009.01020.x>
- Garrod, S., & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, 6(751), 1–12. <https://doi.org/10.3389/fpsyg.2015.00751>
- Gates, V., Callaway, F., Ho, M. K., & Griffiths, T. L. (2021). A rational model of people's inferences about others' preferences based on response times. *Cognition*, 217, 104885. <https://doi.org/10.1016/j.cognition.2021.104885>
- Gauvin, H. S., & Hartsuiker, R. J. (2020). Towards a new model of verbal monitoring. *Journal of Cognition*, 3(1), 17. <https://doi.org/10.5334/joc.81>
- Genty, E., Heesen, R., Guéry, J. P., Rossano, F., Zuberbühler, K., & Bangerter, A. (2020). How apes get into and out of joint actions: Shared intentionality as an interactional achievement. *Interaction Studies*, 21(3), 353–386. <https://doi.org/10.1075/is.18048.gen>
- Gernsbacher, M. A., & Givon, T. (1995). *Coherence in spontaneous text*. John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.31>
- Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, N. Coupland, & J. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge University Press. <https://doi.org/10.1017/cbo9780511663673.001>
- Gillespie, A., & Cornish, F. (2010). Intersubjectivity: Towards a dialogical analysis. *Journal for the Theory of Social Behaviour*, 40(1), 19–46. <https://doi.org/10.1111/j.1468-5914.2009.00419.x>
- Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology*, 78(2), 211–222. <https://doi.org/10.1037/0022-3514.78.2.211>
- Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology*, 75(2), 332–346. <https://doi.org/10.1037/0022-3514.75.2.332>

- Glucksberg, S., & Krauss, R. M. (1967). What do people say after they have learned how to talk? Studies of the development of referential communication. *Merrill-Palmer Quarterly of Behavior and Development*, 13(4), 309–316.
- Godfrey, J., Holliman, E., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of the IEEE International Conference on Acoustics, Speech, & Signal Processing (IEEE ICASSP-92)* (pp. 517–520). IEEE.
<https://doi.org/10.1109/icassp.1992.225858>
- Goffman, E. (1957). Alienation from interaction. *Human Relations*, 10(1), 47–60.
<https://doi.org/10.1177/001872675701000103>
- Goffman, E. (1959). *The presentation of self in everyday life*. Doubleday.
- Goffman, E. (1967). *Interaction ritual: Essays on face-to-face interaction*. Aldine.
- Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.
<https://doi.org/10.1017/s0047404500009623>
- Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11), 818–829.
<https://doi.org/10.1016/j.tics.2016.08.005>
- Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, 5(1), 173–184.
<https://doi.org/10.1111/tops.12007>
- Gottman, J. M. (1979). Detecting cyclicity in social interaction. *Psychological Bulletin*, 86(2), 338–348. <https://doi.org/10.1037/0033-2909.86.2.338>
- Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, 6(1167), 236–245.
<https://doi.org/10.3389/fpsyg.2015.01167>
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts* (pp. 41–58). Brill. https://doi.org/10.1163/9789004368811_003
- Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press.
<https://doi.org/10.1017/s0031819100064330>
- Gunaydin, G., Oztekin, H., Karabulut, D. H., & Salman-Engin, S. (2021). Minimal social interactions with strangers predict greater subjective well-being. *Journal of Happiness Studies*, 22, 1839–1853. <https://doi.org/10.1007/s10902-020-00298-6>

- Hagtvet, B. E., & Wold, A. H. (2003). On the dialogical basis of meaning: Inquiries into Ragnar Rommetveit's writings on language, thought, and communication. *Mind, Culture, and Activity*, 10(3), 186–204. https://doi.org/10.1207/s15327884mca1003_2
- Hagtvet, B. E., Linell, P., Wertsch, J. V., & Wold, A. H. (2020). Ragnar Rommetveit: A full life. *Culture & Psychology*, 26(3), 528–539. <https://doi.org/10.1177/1354067x19898679>
- Hardin, C. D., & Higgins, E. T. (1996). Shared reality: How social verification makes the subjective objective. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (pp. 28–84). Guilford Press.
- Hart, E., VanEpps, E. M., & Schweitzer, M. E. (2021). The (better than expected) consequences of asking sensitive questions. *Organizational Behavior and Human Decision Processes*, 162, 136–154. <https://doi.org/10.1016/j.obhdp.2020.10.014>
- Hasson, U., & Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1693), 20150366. <https://doi.org/10.1098/rstb.2015.0366>
- 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 in Cognitive Sciences*, 16(2), 114–121. <https://doi.org/10.1016/j.tics.2011.12.007>
- Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science*, 2(3), 96–100. <https://doi.org/10.1111/1467-8721.ep10770953>
- Hawkins, R. D., Frank, M. C., & Goodman, N. D. (2020). Characterizing the dynamics of learning in repeated reference games. *Cognitive Science*, 44(6), e12845. <https://doi.org/10.1111/cogs.12845>
- Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., Griffiths, T. L., & Goodman, N. D. (2023). From partners to populations: A hierarchical Bayesian account of coordination and convention. *Psychological Review*, 130(4), 977–1016. <https://doi.org/10.1037/rev0000348>
- Hawkins, R. X. D., Goodman, N. D., & Goldstone, R. L. (2019). The emergence of social norms and conventions. *Trends in Cognitive Sciences*, 23(2), 158–169. <https://doi.org/10.1016/j.tics.2018.11.003>
- Hawkins, R. D., Gweon, H., & Goodman, N. D. (2021). The division of labor in communication: Speakers help listeners account for asymmetries in visual perspective. *Cognitive Science*, 45(3), e12926. <https://doi.org/10.1111/cogs.12926>

- Hawkins, R. X. D., Stuhlmüller, A., Degen, J., & Goodman, N. D. (2015). Why do you ask? Good questions provoke informative answers. In R. D. Noelle, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 878–883). Cognitive Science Society.
- Hawkley, L. C., & Cacioppo, J. T. (2010). Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of Behavioral Medicine*, *40*(2), 218–227. <https://doi.org/10.1007/s12160-010-9210-8>
- Hayes, D. P., & Meltzer, L. (1972). Interpersonal judgments based on talkativeness: I fact or artifact? *Sociometry*, *35*(4), 538–561. <https://doi.org/10.2307/2786532>
- Hayes, J. R., & Bajzek, D. (2008). Understanding and reducing the knowledge effect: Implications for writers. *Written Communication*, *25*(1), 104–118. <https://doi.org/10.1177/0741088307311209>
- Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555–568. <https://doi.org/10.1016/j.wocn.2010.08.002>
- Heldner, M., Hjalmarsson, A., & Edlund, J. (2013). Backchannel relevance spaces. In E. L. Asu & P. Lippus (Eds.), *Nordic Prosody: Proceedings of the XIth Conference, Tartu 2012* (pp. 137–146). Peter Lang, Germany.
- Henrich, J. (2015). Culture and social behavior. *Current Opinion in Behavioral Sciences*, *3*, 84–89. <https://doi.org/10.1016/j.cobeha.2015.02.001>
- Heritage, J. (1985). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 299–345). Cambridge University Press. <https://doi.org/10.1017/cbo9780511665868.020>
- Heritage, J. (2013). *Garfinkel and ethnomethodology*. John Wiley & Sons. (Original work published 1984)
- Heritage, J., & Clayman, S. (2010). *Talk in action: Interactions, identities, and institutions*. Wiley-Blackwell. <https://doi.org/10.1002/9781444318135>
- Heritage, J., & Raymond, G. (2012). Navigating epistemic landscapes: Acquiescence, agency and resistance in responses to polar questions. In J. P. de Ruiter (Ed.), *Questions: Formal, functional and interactional perspectives* (pp. 179–192). Cambridge University Press. <https://doi.org/10.1017/CBO9781139045414.013>
- Hess, L. J., & Johnston, J. R. (1988). Acquisition of back channel listener responses to adequate messages. *Discourse Processes*, *11*(3), 319–335.

<https://doi.org/10.1080/01638538809544706>

- Higgins, E. T. (1981). The "communication game": Implications for social cognition and persuasion. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium* (Vol. 1, pp. 343–392). Erlbaum.
- Higgins, E. T. (1992). Achieving "shared reality" in the communication game: A social action that creates meaning. *Journal of Language and Social Psychology*, 11(3), 107–131. <https://doi.org/10.1177/0261927X92113001>
- Higgins, E. T., & Rholes, W. S. (1978). "Saying is believing": Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology*, 14(4), 363–378. [https://doi.org/10.1016/0022-1031\(78\)90032-x](https://doi.org/10.1016/0022-1031(78)90032-x)
- Higgins, E. T., Rossignac-Milon, M., & Echterhoff, G. (2021). Shared reality: From sharing-is-believing to merging minds. *Current Directions in Psychological Science*, 30(2), 103–110. <https://doi.org/10.1177/0963721421992027>
- Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: A longitudinal study of mother–infant interaction. *Frontiers in Psychology*, 6, 1492. <https://doi.org/10.3389/fpsyg.2015.01492>
- Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied*, 5(2), 205–221. <https://doi.org/10.1037/1076-898X.5.2.205>
- Hirschi, Q., Wilson, T. D., & Gilbert, D. T. (2022). Speak up! Mistaken beliefs about how much to talk in conversations. *Personality and Social Psychology Bulletin*, 49(10), 1454–1465. <https://doi.org/10.1177/01461672221104927>
- Ho, M. K., Saxe, R., & Cushman, F. (2022). Planning with theory of mind. *Trends in Cognitive Sciences*, 26(11), 959–971. <https://doi.org/10.1016/j.tics.2022.08.003>
- Holler, J. (2022). Visual bodily signals as core devices for coordinating minds in interaction. *Philosophical Transactions of the Royal Society B*, 377(1859), 20210094. <https://doi.org/10.1098/rstb.2021.0094>
- Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin & Review*, 25(5), 1900–1908. <https://doi.org/10.3758/s13423-017-1363-z>
- Holler, J., & Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior*, 35(2), 133–153. <https://doi.org/10.1007/s10919-011-0105-6>

- Holroyd, C. B. (2022). Interbrain synchrony: On wavy ground. *Trends in Neurosciences*, 45(5), 346–357. <https://doi.org/10.1016/j.tins.2022.02.002>
- Holtgraves, T. M. (2001). *Language as social action: Social psychology and language use*. Psychology Press. <https://doi.org/10.4324/9781410601773>
- Holtgraves, T. M. (2002). *Language as social action: Social psychology and language use*. Erlbaum.
- Holtgraves, T. (2010). Social psychology and language: Words, utterances, and conversations. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 1386–1422). John Wiley & Sons. <https://doi.org/10.1002/9780470561119.socpsy002036>
- Holtgraves, T. (2021). Understanding miscommunication: speech act recognition in digital contexts. *Cognitive Science*, 45(10), e13023. <https://doi.org/10.1111/cogs.13023>
- Holtgraves, T., & Joong-nam, Y. (1990). Politeness as universal: Cross-cultural perceptions of request strategies and inferences based on their use. *Journal of Personality and Social Psychology*, 59(4), 719–729. <https://doi.org/10.1037/0022-3514.59.4.719>
- Holt-Lunstad, J., Robles, T. F., & Sbarra, D. A. (2017). Advancing social connection as a public health priority in the United States. *American Psychologist*, 72(6), 517–530. <https://doi.org/10.1037/amp0000103>
- Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science*, 10(2), 227–237. <https://doi.org/10.1177/1745691614568352>
- Horton, W. (2017). Theories and approaches to the study of conversation and interactive discourse. In M. F. Schober, D. N. Rapp, & A. M. Britt (Eds.), *The Routledge handbook of discourse processes* (pp. 22–68). Routledge Press. <https://doi.org/10.4324/9781315687384-3>
- Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, 96(2), 127–142. <https://doi.org/10.1016/j.cognition.2004.07.001>
- House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. *Science*, 241(4865), 540–545. <https://doi.org/10.1126/science.3399889>
- Huang, K., Yeomans, M., Brooks, A. W., Minson, J., & Gino, F. (2017). It doesn't hurt to ask: Question-asking increases liking. *Journal of Personality and Social Psychology*, 113(3), 430–452. <https://doi.org/10.1037/pspi0000097>

- Huang, Y. (Ed.). (2017). *The Oxford handbook of pragmatics*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199697960.001.0001>
- Hutchby, I., & Wooffitt, R. (2008). *Conversation analysis*. Polity press. (Original work published 1998)
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, 2, 255. <https://doi.org/10.3389/fpsyg.2011.00255>
- Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1-2), 101-144. <https://doi.org/10.1016/j.cognition.2002.06.001>
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, 22(1), 39-44. <https://doi.org/10.1177/0956797610392928>
- Itzchakov, G., Reis, H. T., & Weinstein, N. (2022). How to foster perceived partner responsiveness: High-quality listening is key. *Social and Personality Psychology Compass*, 16(1), e12648. <https://doi.org/10.1111/spc3.12648>
- Itzchakov, G., Weinstein, N., Saluk, D., & Amar, M. (2023). Connection heals wounds: Feeling listened to reduces speakers' loneliness following a social rejection disclosure. *Personality and Social Psychology Bulletin*, 49(8), 1273-1294. <https://doi.org/10.1177/01461672221100369>
- Jaffe, J. & Feldstein, S. (1970). *Rhythms of dialogue*. Academic Press.
- Jefferson, G. (1984). Notes on a systematic deployment of the acknowledgement tokens "yeah"; and "mm hm." *Linguistics*, 17(2), 197-216. <https://doi.org/10.1080/08351818409389201>
- Jefferson, G. (1993). Caveat speaker: Preliminary notes on recipient topic-shift implicature. *Research on Language and Social Interaction*, 26(1), 1-30. https://doi.org/10.1207/s15327973rlsi2601_1 (Original work published 1983)
- Jones, E. E., & Kelly, J. R. (2007). Contributions to a group discussion and perceptions of leadership: Does quantity always count more than quality?. *Group Dynamics: Theory, Research, and Practice*, 11(1), 15-30. <https://doi.org/10.1037/1089-2699.11.1.15>
- Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, 111(33), 12002-12007. <https://doi.org/10.1073/pnas.1407479111>

Kardas, M., Kumar, A., & Epley, N. (2022). Overly shallow?: Miscalibrated expectations create a barrier to deeper conversation. *Journal of Personality and Social Psychology*, 122(3), 367–398. <https://doi.org/10.1037/pspa0000281>

Kardas, M., Schroeder, J., & O'Brien, E. (2022). Keep talking: (Mis)understanding the hedonic trajectory of conversation. *Journal of Personality and Social Psychology*, 123(4), 717–740. <https://doi.org/10.1037/pspi0000379>

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, 35(2), 157–175. <https://doi.org/10.1006/jmla.1996.0009>

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26, 22–63. [https://doi.org/10.1016/0001-6918\(67\)90005-4](https://doi.org/10.1016/0001-6918(67)90005-4)

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511807572>

Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes*, 52(4), 255–289. <https://doi.org/10.1080/0163853x.2014.955997>

Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*, 4(1), 71–84. <https://doi.org/10.1515/ip.2007.004>

Keysar, B., & Bly, B. (1995). Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans?. *Journal of Memory and Language*, 34(1), 89–109. <https://doi.org/10.1006/jmla.1995.1005>

Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science*, 13(3), 207–212. <https://doi.org/10.1111/1467-9280.00439>

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89(1), 25–41. [https://doi.org/10.1016/s0010-0277\(03\)00064-7](https://doi.org/10.1016/s0010-0277(03)00064-7)

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology* 2(1), 125–156. <https://doi.org/10.1515/labphon.2011.004>

Kirscht, J. P., Lodahl, T. M., & Haire, M. (1959). Some factors in the selection of leaders by members of small groups. *The Journal of Abnormal and Social Psychology*, 58(3), 406–408. <https://doi.org/10.1037/h0038710>

- Knox, D., & Lucas, C. (2021). A dynamic model of speech for the social sciences. *American Political Science Review*, *115*(2), 649-666.
<https://doi.org/10.1017/S000305542000101X>
- Koudenburg, N. (2018). Regulating shared reality with micro-dynamics in the form of conversation. *Current Opinion in Psychology*, *23*, 47-51.
<https://doi.org/10.1016/j.copsyc.2017.12.002>
- Koudenburg, N., Postmes, T., & Gordijn, E. H. (2011). Disrupting the flow: How brief silences in group conversations affect social needs. *Journal of Experimental Social Psychology*, *47*(2), 512-515. <https://doi.org/10.1016/j.jesp.2010.12.006>
- Krauss, R. M., & Chiu, C. (1998). Language and social behavior. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 2, pp. 41-88). McGraw-Hill.
- Krauss, R. M., & Fussell, S. R. (1991a). Constructing shared communicative environments. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 172-200). American Psychological Association.
<https://doi.org/10.1037/10096-008>
- Krauss, R. M., & Fussell, S. R. (1991b). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition*, *9*(1), 2-24.
<https://doi.org/10.1521/soco.1991.9.1.2>
- Krauss, R. M., & Fussell, S. R. (1996). Social psychological models of interpersonal communication. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: A handbook of basic principles* (pp. 655-701). Guilford.
- Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*, 113-114. <https://doi.org/10.3758/bf03342817>
- Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*(3), 343-346. <https://doi.org/10.1037/h0023705>
- Kuhlen, A. K., Allefeld, C., & Haynes, J. D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in Human Neuroscience*, *6*, 266.
<https://doi.org/10.3389/fnhum.2012.00266>
- Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes*, *47*(7), 567-587. <https://doi.org/10.1080/01638530903441339>

- Kumar, A. (2022). Some things aren't better left unsaid: Interpersonal barriers to gratitude expression and prosocial engagement. *Current Opinion in Psychology*, 43, 156–160. <https://doi.org/10.1016/j.copsyc.2021.07.011>
- Kumar, A., & Epley, N. (2018). Undervaluing gratitude: Expressers misunderstand the consequences of showing appreciation. *Psychological Science*, 29(9), 1423–1435. <https://doi.org/10.1177/0956797618772506>
- Lahnakoski, J. M., Glerean, E., Jääskeläinen, I. P., Hyönä, J., Hari, R., Sams, M., & Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*, 100, 316–324. <https://doi.org/10.1016/j.neuroimage.2014.06.022>
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2-3), 259–284. <https://doi.org/10.1080/01638539809545028>
- Lassiter, D., & Goodman, N. D. (2017). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, 194, 3801–3836. <https://doi.org/10.1007/s11229-015-0786-1>
- Leary, T. (1957). *Interpersonal diagnosis of personality*. Ronald Press.
- Lerner, G. H. (2004). Collaborative turn sequences. In G.H. Lerner (Ed.) *Conversation analysis: Studies from the first generation* (pp. 225–256). John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.125.12ler>
- Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41–104. [https://doi.org/10.1016/0010-0277\(83\)90026-4](https://doi.org/10.1016/0010-0277(83)90026-4)
- Levelt, W. J. (1993). *Speaking: From intention to articulation*. MIT Press. <https://doi.org/10.7551/mitpress/6393.001.0001>
- Levine, E. E., & Cohen, T. R. (2018). You can handle the truth: Mispredicting the consequences of honest communication. *Journal of Experimental Psychology: General*, 147(9), 1400–1429. <https://doi.org/10.1037/xge0000488>
- Levine, E. E., Roberts, A. R., & Cohen, T. R. (2020). Difficult conversations: Navigating the tension between honesty and benevolence. *Current Opinion in Psychology*, 31, 38–43. <https://doi.org/10.1016/j.copsyc.2019.07.034>
- Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511813313>
- Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.

- Levinson, S. C. (2016). Turn-taking in human communication – Origins and implications for language processing. *Trends in Cognitive Sciences*, 20(1), 6–14.
<https://doi.org/10.1016/j.tics.2015.10.010>
- Levinson, S. C. (2019). Interactional foundations of language: The interaction engine hypothesis. In P. Hagoort (Ed.), *Human language: From genes and brain to behavior* (pp. 189–200). MIT Press. <https://doi.org/10.7551/mitpress/10841.003.0018>
- Levinson, S. C. (2020). On the human “interaction engine.” In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39–69). Routledge. <https://doi.org/10.4324/9781003135517-3> (Original work published 2006)
- Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130302. <https://doi.org/10.1098/rstb.2013.0302>
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6, 731.
<https://doi.org/10.3389/fpsyg.2015.00731>
- Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In P. Cosi, R. De Mori, G. Di Fabbrizio, & R. Pieraccini (Eds.), *Proceedings of Interspeech* (pp. 3081–3084).
<https://doi.org/10.21437/interspeech.2011-771>
- Lewin, K. (1951). Problems of research in social psychology. In D. Cartwright (Ed.), *Field theory in social science: Selected theoretical papers by Kurt Lewin*. Harper & Bros.
- Lewis, D. (1969). *Convention: A philosophical study*. Harvard University Press.
- Li, J., Zhong, Z., & Mo, L. (2020). Negative deviation effect in interpersonal communication: Why people underestimate the positivity of impression they left on others. *Psychology Research and Behavior Management*, 13, 733–745.
<https://doi.org/10.2147/PRBM.S258057>
- Lieberman, M. D. (2022). Seeing minds, matter, and meaning: The CEEing model of pre-reflective subjective construal. *Psychological Review*, 129(4), 830.
<https://doi.org/10.1037/rev0000362>
- Lindsay, L., Gambi, C., & Rabagliati, H. (2019). Preschoolers optimize the timing of their conversational turns through flexible coordination of language comprehension and production. *Psychological Science*, 30(4), 504–515.
<https://doi.org/10.1177/0956797618822802>

- Liu, J., Nikolaus, M., Bodur, K., & Fourtassi, A. (2022). Predicting backchannel signaling in child-caregiver multimodal conversations. In R. Tumuluri, N. Sebe, G. Pingali, D. B. Jayagopi, A. Dhall, R. Singh, L. Anthony, & A. A. Salah (Eds.), *Companion publication of the 2022 International Conference on Multimodal Interaction* (pp. 196–200). Association for Computing Machinery. <https://doi.org/10.1145/3536220.3563372>
- Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PloS One*, *13*(5), e0196391. <https://doi.org/10.1371/journal.pone.0196391>
- Local, J., & Walker, G. (2012). How phonetic features project more talk. *Journal of the International Phonetic Association*, *42*(3), 255–280. <https://doi.org/10.1017/s0025100312000187>
- MacKinnon, D. W. (1944). The structure of personality. In J. M. Hunt (Ed.), *Personality and the behavior disorders* (pp. 3–48). Ronald Press.
- MacLaren, N. G., Yammarino, F. J., Dionne, S. D., Sayama, H., Mumford, M. D., Connelly, S., Martin, R. W., Mulhearn, T. J., Todd, E. M., Kulkarni, A., Cao, Yiding., & Ruark, G. A. (2020). Testing the babble hypothesis: Speaking time predicts leader emergence in small groups. *The Leadership Quarterly*, *31*(5), 101409. <https://doi.org/10.1016/j.leaqua.2020.101409>
- Magyari, L. (2022). Predictions in conversation. In J. Gervain, G. Csibra, & K. Kovács (Eds.), *A life in cognition. Language, cognition, and mind* (Vol 11). Springer, Cham. https://doi.org/10.1007/978-3-030-66175-5_5
- Magyari, L., Bastiaansen, M. C., De Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience*, *26*(11), 2530–2539. https://doi.org/10.1162/jocn_a_00673
- Magyari, L., & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, *3*, 376. <https://doi.org/10.3389/fpsyg.2012.00376>
- Malinowski, B. (1946). The problem of meaning in primitive languages. In C. K. Ogden, & I. A. Richards (Eds.), *The meaning of meaning: A study of the influence of language upon thought and of the science of symbolism* (pp. 296–336). Harcourt, Brace & World. (Original work published 1923)
- Marková, I., Graumann, C. F., & Foppa, K. (Eds.). (1995). *Mutualities in dialogue*. Cambridge University Press.

- Martin, A. E., & Slepian, M. L. (2021). The primacy of gender: Gendered cognition underlies the big two dimensions of social cognition. *Perspectives on Psychological Science*, 16(6), 1143–1158. <https://doi.org/10.1177/1745691620904961>
- Martz, M. J., & Welkowitz, J. (1977). WELMAR—Computer programs to analyze dialogic time patterns. *Perceptual and Motor Skills*, 45(2), 531–537. <https://doi.org/10.2466/pms.1977.45.2.531>
- Maschler, Y., & Schiffrin, D. (2015). Discourse markers: Language, meaning, and context. In D. Tannen, H. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis* (pp. 189–221). John Wiley & Sons. <https://doi.org/10.1002/9781118584194.ch9>
- Mastroianni, A. M., Cooney, G., Boothby, E. J., & Reece, A. G. (2021). The liking gap in groups and teams. *Organizational Behavior and Human Decision Processes*, 162, 109–122. <https://doi.org/10.1016/j.obhdp.2020.10.013>
- Mastroianni, A. M., Gilbert, D. T., Cooney, G., & Wilson, T. D. (2021). Do conversations end when people want them to?. *Proceedings of the National Academy of Sciences*, 118(10), e2011809118. <https://doi.org/10.1073/pnas.2011809118>
- Matarazzo, J. D., Saslow, G., & Guze, S. B. (1956). Stability of interaction patterns during interviews: A replication. *Journal of Consulting Psychology*, 20(4), 267–274. <https://doi.org/10.1037/h0044937>
- Matarazzo, J. D., Saslow, G., & Matarazzo, R. G. (1956). The interaction chronograph as an instrument for objective measurement of interaction patterns during interviews. *The Journal of Psychology*, 41(2), 347–367. <https://doi.org/10.1080/00223980.1956.9713009>
- Matarazzo, J. D., Weitman, M., Saslow, G., & Wiens, A. N. (1963). Interviewer influence on durations of interviewee speech. *Journal of Verbal Learning and Verbal Behavior*, 1(6), 451–458. [https://doi.org/10.1016/s0022-5371\(63\)80031-6](https://doi.org/10.1016/s0022-5371(63)80031-6)
- Matarazzo, J. D. & Wiens, A. N. (1972). *The interview: Research on its anatomy and structure*. Routledge. <https://doi.org/10.4324/9781315132723>
- Matarazzo, J. D., Wiens, A. N., Saslow, G., Dunham, R. M., & Voas, R. B. (1964). Speech durations of astronaut and ground communicator. *Science*, 143(3602), 148–150. <https://doi.org/10.1126/science.143.3602.148>
- Mayo, O., & Gordon, I. (2020). In and out of synchrony—Behavioral and physiological dynamics of dyadic interpersonal coordination. *Psychophysiology*, 57(6), e13574. <https://doi.org/10.1111/psyp.13574>

- Mead, G. H. (1934). *Mind, self, and society: From the standpoint of a social behaviorist*. University of Chicago Press.
- Meyer, A. S., Alday, P. M., Decuyper, C., & Knudsen, B. (2018). Working together: Contributions of corpus analyses and experimental psycholinguistics to understanding conversation. *Frontiers in Psychology*, 9, 525. <https://doi.org/10.3389/fpsyg.2018.00525>
- Micklos, A., Walker, B., & Fay, N. (2020). Are people sensitive to problems in communication?. *Cognitive Science*, 44(2), e12816. <https://doi.org/10.1111/cogs.12816>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv:1301.3781 [cs.CL]. <https://doi.org/10.48550/arXiv.1301.3781>
- Mills, G. (2014). Dialogue in joint activity: Complementarity, convergence and conventionalization. *New Ideas in Psychology*, 32, 158–173. <http://doi.org/10.1016/j.newideapsych.2013.03.006>
- Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2019). Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1), 18–31. <https://doi.org/10.1109/taffc.2017.2740923>
- Mondada, L. (2009). Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of Pragmatics*, 41(10), 1977–1997. <https://doi.org/10.1016/j.pragma.2008.09.019>
- Monroe, W., Hawkins, R. X., Goodman, N. D., & Potts, C. (2017). Colors in context: A pragmatic neural model for grounded language understanding. *Transactions of the Association for Computational Linguistics*, 5, 325–338. https://doi.org/10.1162/tacl_a_00064
- Moore, C., Dunham, P. J., & Dunham, P. (Eds.). (2014). *Joint attention: Its origins and role in development*. Psychology Press. <https://doi.org/10.4324/9781315806617> (Original work published 1995)
- Moreland, R. L. (2010). Are dyads really groups?. *Small Group Research*, 41(2), 251–267. <https://doi.org/10.1177/1046496409358618>
- Mullen, B., Salas, E., & Driskell, J. E. (1989). Salience, motivation, and artifact as contributions to the relation between participation rate and leadership. *Journal of Experimental Social Psychology*, 25(6), 545–559. [https://doi.org/10.1016/0022-1031\(89\)90005-x](https://doi.org/10.1016/0022-1031(89)90005-x)

Nguyen, M., Vanderwal, T., & Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*, 184, 161–170.

<https://doi.org/10.1016/j.neuroimage.2018.09.010>

Nguyen, T., Zimmer, L., & Hoehl, S. (2023). Your turn, my turn. Neural synchrony in mother–infant proto-conversation. *Philosophical Transactions of the Royal Society B*, 378(1875), 20210488.

<https://doi.org/10.1098/rstb.2021.0488>

Nguyen, V., Versyp, O., Cox, C., & Fusaroli, R. (2022). A systematic review and Bayesian meta-analysis of the development of turn taking in adult–child vocal interactions. *Child Development*, 93(4), 1181–1200.

<https://doi.org/10.1111/cdev.13754>

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, 125(6), 737–759.

<https://doi.org/10.1037/0033-2909.125.6.737>

Nickerson, R. S., Baddeley, A., & Freeman, B. (1987). Are people's estimates of what other people know influenced by what they themselves know?. *Acta Psychologica*, 64(3), 245–259.

[https://doi.org/10.1016/0001-6918\(87\)90010-2](https://doi.org/10.1016/0001-6918(87)90010-2)

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, 33(6), 417–433.

<https://doi.org/10.1017/s0140525x10000865>

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., & Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences*, 109(24), 9599–9604.

<https://doi.org/10.1073/pnas.1206095109>

Ostrand, R., & Chodroff, E. (2021). It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. *Journal of Phonetics*, 88, 101074.

<https://doi.org/10.1016/j.wocn.2021.101074>

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382–2393.

<https://doi.org/10.1121/1.2178720>

Park, H. W., Gelsomini, M., Lee, J. J., & Breazeal, C. (2017). Telling stories to robots: The effect of backchanneling on a child's storytelling. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 100–108). Association for Computing Machinery.

<https://doi.org/10.1145/2909824.3020245>

Parker, D. (1942). Here we are. In *The collected stories of Dorothy Parker* (pp. 51–66). Random House. (Original work published 1931)

- Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 1532–1543). Association for Computational Linguistics. <https://doi.org/10.3115/v1/d14-1162>
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. arXiv:1802.05365 [cs.CL]. <https://doi.org/10.48550/arXiv.1802.05365>
- Pfänder, S., & Couper-Kuhlen, E. (2019). Turn-sharing revisited: An exploration of simultaneous speech in interactions between couples. *Journal of Pragmatics*, 147, 22–48. <https://doi.org/10.1016/j.pragma.2019.05.010>
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169–190. <https://doi.org/10.1017/s0140525x04000056>
- Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4, 203–228. <https://doi.org/10.1007/s11168-006-9004-0>
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347. <https://doi.org/10.1017/s0140525x12001495>
- Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge University Press. <https://doi.org/10.1017/9781108610728>
- Pika, S., Wilkinson, R., Kendrick, K. H., & Vernes, S. C. (2018). Taking turns: bridging the gap between human and animal communication. *Proceedings of the Royal Society B: Biological Sciences*, 285(1880), 20180598. <https://doi.org/10.1098/rspb.2018.0598>
- Pinel, E. C. (2018). Existential isolation and I-sharing: Interpersonal and intergroup implications. *Current Opinion in Psychology*, 23, 84–87. <https://doi.org/10.1016/j.copsyc.2018.01.002>
- Pinker, S. (2007). The evolutionary social psychology of off-record indirect speech acts. *De Gruyter Mouton*, 4(4), 437–461. <https://doi.org/10.1515/IP.2007.023>
- Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences*, 105(3), 833–838. <https://doi.org/10.1073/pnas.0707192105>
- Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In J. M. Atkinson & J. Heritage (Eds.), *Structures*

- of social action: *Studies in conversation analysis* (pp. 57–101). Cambridge University Press. <https://doi.org/10.1017/cbo9780511665868.008>
- Pomerantz, A., & Heritage, J. (2013). Preference. In J. Sidnell & T. Stivers (Eds.), *Handbook of conversation analysis* (pp. 210–228). Cambridge University Press. <https://doi.org/10.1002/9781118325001.ch11>
- Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E., & Mihalcea, R. (2019). MELD: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 527–536). Association for Computational Linguistics. <https://doi.org/10.18653/v1/p19-1050>
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences*, 1(4), 515–526. <https://doi.org/10.1017/s0140525x00076512>
- Quine, W. V. O., & Ullian, J. S. (1978). *The web of belief*. Random House. (Original work published 1970)
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. *OpenAI*. Retrieved from https://cdn.openai.com/research-covers/language_unsupervised/language_understanding_paper.pdf
- Ram, A., Prasad, R., Khatri, C., Venkatesh, A., Gabriel, R., Liu, Q., Nunn, J., Hedayatnia, B., Cheng, M., Nagar, A., King, E., Bland, K., Wartick, A., Pan, Y., Jayadevan, S., Hwang, G., & Pettigrew, A. (2018). Conversational AI: The science behind the Alexa prize. arXiv:1801.03604 [cs.AI]. <https://doi.org/10.48550/arXiv.1801.03604>
- Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in multimodal interaction: An integrative framework. *Cognitive Science*, 44(11), e12911. <https://doi.org/10.1111/cogs.12911>
- Ray, M. L., & Webb, E. J. (1966). Speech duration effects in the Kennedy news conferences. *Science*, 153(3738), 899–901. <https://doi.org/10.1126/science.153.3738.899>
- Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, 20(8), 495–505. <https://doi.org/10.1038/s41583-019-0179-4>
- Reece, A., Cooney, G., Bull, P., Chung, C., Dawson, B., Fitzpatrick, C., Glazer, T., Knox, D., Liebscher, A., & Marin, S. (2023). The CANDOR corpus: Insights from a large

- multimodal dataset of naturalistic conversation. *Science Advances*, 9(13), eadf3197. <https://doi.org/10.1126/sciadv.adf3197>
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5), 368–373. <https://doi.org/10.1111/j.1467-9280.1997.tb00427.x>
- Richards, I. A. (2001). *Principles of literary criticism*. Routledge. <https://doi.org/10.4324/9780203164747> (Original work published 1924)
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science*, 18(5), 407–413. <https://doi.org/10.1111/j.1467-9280.2007.01914.x>
- Richardson, D. C., Dale, R., & Shockley, K. (2008). Synchrony and swing in conversation: Coordination, temporal dynamics, and communication. In I. Wachsmuth, M. Lenzen, & G. Knoblich (Eds.), *Embodied communication in humans and machines* (pp. 75–94). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199231751.003.0004>
- Ritchie, D. (1986). Shannon and Weaver: Unravelling the paradox of information. *Communication Research*, 13(2), 278–298. <https://doi.org/10.1177/009365086013002007>
- Roberts, G., Langstein, B., & Galantucci, B. (2016). (In) sensitivity to incoherence in human communication. *Language & Communication*, 47, 15–22. <https://doi.org/10.1016/j.langcom.2015.11.001>
- Roller, S., Dinan, E., Goyal, N., Ju, D., Williamson, M., Liu, Y., Xu, J., Ott, M., Smith, E. M., Boureau, Y., & Weston, J. (2021). Recipes for building an open-domain chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 300–325). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.eacl-main.24>
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13(3), 279–301. [https://doi.org/10.1016/0022-1031\(77\)90049-x](https://doi.org/10.1016/0022-1031(77)90049-x)
- Rossano, F. (2012) Gaze in conversation. In J. Sidnell & T. Stivers (Eds.), *Handbook of conversation analysis* (pp. 308–329). Wiley-Blackwell. <https://doi.org/10.1002/9781118325001.ch15>
- Rossignac-Milon, M., Bolger, N., Zee, K. S., Boothby, E. J., & Higgins, E. T. (2021). Merged minds: Generalized shared reality in dyadic relationships. *Journal of Personality and Social Psychology*, 120(4), 882–911. <https://doi.org/10.1037/pspi0000266>

- Rühlemann, C., & Gries, S. T. (2020). Speakers advance-project turn completion by slowing down: A multifactorial corpus analysis. *Journal of Phonetics*, 80, 100976. <https://doi.org/10.1016/j.wocn.2020.100976>
- Rumsey, A. & Sidnell, J. (2011). Part two. The ethics of speaking. In M. Lambek (Ed.), *Ordinary ethics: Anthropology, language, and action* (pp. 103–140). Fordham University Press. <https://doi.org/10.1515/9780823292318-005>
- Sacks, H. (1985). Notes on methodology. In J. M. Atkinson (Ed.), *Structures of social action* (pp. 21–27). Cambridge University Press. <https://doi.org/10.1017/cbo9780511665868.005>
- Sacks, H., & Schegloff, E. A. (2007). Two preferences in the organization of reference to persons in conversation and their interaction. In N. J. Enfield & T. Stivers (Eds.), *Person reference in interaction: Linguistic, cultural and social Perspectives* (pp. 23–28). Cambridge University Press. <https://doi.org/10.1017/CBO9780511486746.003> (Original work published 1979)
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735. <https://doi.org/10.2307/412243>
- Sanchez, K. L., Kalkstein, D. A., & Walton, G. M. (2022). A threatening opportunity: The prospect of conversations about race-related experiences between Black and White friends. *Journal of Personality and Social Psychology*, 122(5), 853–872. <https://doi.org/10.1037/pspi0000369>
- Sandstrom, G. M., & Boothby, E. J. (2021). Why do people avoid talking to strangers? A mini meta-analysis of predicted fears and actual experiences talking to a stranger. *Self and Identity*, 20(1), 47–71. <https://doi.org/10.1080/15298868.2020.1816568>
- Sandstrom, G. M., Boothby, E. J., & Cooney, G. (2022). Talking to strangers: A week-long intervention reduces psychological barriers to social connection. *Journal of Experimental Social Psychology*, 102, 104356. <https://doi.org/10.1016/j.jesp.2022.104356>
- Sandstrom, G. M., & Dunn, E. W. (2014a). Is efficiency overrated? Minimal social interactions lead to belonging and positive affect. *Social Psychological and Personality Science*, 5(4), 437–442. <https://doi.org/10.1177/1948550613502990>
- Sandstrom, G. M., & Dunn, E. W. (2014b). Social interactions and well-being: The surprising power of weak ties. *Personality and Social Psychology Bulletin*, 40(7), 910–922. <https://doi.org/10.1177/0146167214529799>

- Saslow, G., Matarazzo, J. D., & Guze, S. B. (1955). The stability of interaction chronograph patterns in psychiatric interviews. *Journal of Consulting Psychology*, 19(6), 417–430. <https://doi.org/10.1037/h0047036>
- Savitsky, K., & Gilovich, T. (2003). The illusion of transparency and the alleviation of speech anxiety. *Journal of Experimental Social Psychology*, 39(6), 618–625. [https://doi.org/10.1016/s0022-1031\(03\)00056-8](https://doi.org/10.1016/s0022-1031(03)00056-8)
- Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist*, 70(6), 1075–1095. <https://doi.org/10.1525/aa.1968.70.6.02a00030>
- Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of ‘uh huh’ and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 71–93). Georgetown University Press.
- Schegloff, E. A. (1989). Harvey Sacks — lectures 1964–1965: An introduction/memoir. In G. Jefferson (Ed.), *Harvey Sacks lectures 1964–1965*. https://doi.org/10.1007/978-94-015-6853-1_2
- Schegloff, E. A. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 97(5), 1295–1345. <https://doi.org/10.1086/229903>
- Schegloff, E. A. (1998). Body torque. *Social Research*, 65(3), 535–596.
- Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1), 1–63. <https://doi.org/10.1017/s0047404500001019>
- Schegloff, E. A. (2007). *Sequence organization in interaction*. Cambridge University Press.
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53(2), 361–382. <https://doi.org/10.1353/lan.1977.0041>
- Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica* 7(4), 289–327. <https://doi.org/10.1515/semi.1973.8.4.289>
- Schiffrin, D. (1987). *Discourse markers*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511611841>
- Schippers, M. B., Roebroek, A., Renken, R., Nanetti, L., & Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proceedings of the National Academy of Sciences*, 107(20), 9388–9393. <https://doi.org/10.1073/pnas.1001791107>

Schmid Mast, M. (2001). Gender differences and similarities in dominance hierarchies in same-gender groups based on speaking time. *Sex Roles*, 44, 537–556.

<https://doi.org/10.1023/a:1012239024732>

Schmid Mast, M. (2002). Dominance as expressed and inferred through speaking time: A meta-analysis. *Human Communication Research*, 28(3), 420–450.

<https://doi.org/10.1111/j.1468-2958.2002.tb00814.x>

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21(2), 211–232. [https://doi.org/10.1016/0010-0285\(89\)90008-x](https://doi.org/10.1016/0010-0285(89)90008-x)

Schramm, W. L. (1963). Communication research in the United States. In W. L. Schramm (Ed.), *The science of human communication: New directions and new findings in communication research* (pp. 1–16). Basic Books.

Schroeder, J., Lyons, D., & Epley, N. (2022). Hello, stranger? Pleasant conversations are preceded by concerns about starting one. *Journal of Experimental Psychology: General*, 151(5), 1141–1153. <https://doi.org/10.1037/xge0001118>

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press. https://doi.org/10.1007/978-3-658-13213-2_61

Serban, I. V., Lowe, R., Henderson, P., Charlin, L., & Pineau, J. (2015). A survey of available corpora for building data-driven dialogue systems. arXiv:1512.05742 [cs.CL]

<https://doi.org/10.48550/arXiv.1512.05742>

Shannon, C. E. (1948a). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>

Shannon, C. E. (1948b). A mathematical theory of communication. *The Bell System Technical Journal*, 27(4), 623–656. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.

Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, 1(2), 305–319. <https://doi.org/10.1111/j.1756-8765.2009.01021.x>

Sidnell, J., & Stivers, T. (Eds.). (2012). *The handbook of conversation analysis*. John Wiley & Sons. <https://doi.org/10.1002/9781118325001>

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentive blindness for dynamic events. *Perception*, 28(9), 1059–1074. <https://doi.org/10.1068/p281059>

- Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, 5, 644–649. <https://doi.org/10.3758/bf03208840>
- Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, 9(1), 16–20. <https://doi.org/10.1016/j.tics.2004.11.006>
- Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition*, 136, 304–324. <https://doi.org/10.1016/j.cognition.2014.10.008>
- Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, 32(1), 25–8. <https://doi.org/10.1006/jmla.1993.1002>
- Sorrentino, R. M., & Boutillier, R. G. (1975). The effect of quantity and quality of verbal interaction on ratings of leadership ability. *Journal of Experimental Social Psychology*, 11(5), 403–411. [https://doi.org/10.1016/0022-1031\(75\)90044-x](https://doi.org/10.1016/0022-1031(75)90044-x)
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Wiley-Blackwell. (Original work published 1986)
- Stang, D. J. (1973). Effect of interaction rate on ratings of leadership and liking. *Journal of Personality and Social Psychology*, 27(3), 405–408. <https://doi.org/10.1037/h0034940>
- Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32), 14425–14430. <https://doi.org/10.1073/pnas.1008662107>
- Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, 41(1), 31–57. <https://doi.org/10.1080/08351810701691123>
- Stivers, T. (2012) Sequence organization. In J. Sidnell & T. Stivers (Eds.), *Handbook of conversation analysis* (pp. 191–209). Wiley-Blackwell. <https://doi.org/10.1002/9781118325001.ch10>
- Stivers, T. (2021). Is conversation built for two? The partitioning of social interaction. *Research on Language and Social Interaction*, 54(1), 1–19. <https://doi.org/10.1080/08351813.2020.1864158>
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592. <https://doi.org/10.1073/pnas.0903616106>

- Stivers, T., & Rossano, F. (2010). Mobilizing response. *Research on Language and Social Interaction*, 43(1), 3–31. <https://doi.org/10.1080/08351810903471258>
- Stolk, A., Basnakova, J., & Toni, I. (2023). Joint epistemic engineering: The neglected process in human communication. In A. M. García & A. Ibáñez (Eds.), *The Routledge handbook of semiosis and the brain* (pp. 259–278). Routledge. <https://doi.org/10.4324/9781003051817-21>
- Stolk, A., Noordzij, M. L., Verhagen, L., Volman, I., Schoffelen, J. M., Oostenveld, R., Hagoort, P., & Toni, I. (2014). Cerebral coherence between communicators marks the emergence of meaning. *Proceedings of the National Academy of Sciences*, 111(51), 18183–18188. <https://doi.org/10.1073/pnas.1414886111>
- Stolk, A., Verhagen, L., Schoffelen, J. M., Oostenveld, R., Blokpoel, M., Hagoort, P., van Rooij, I., & Toni, I. (2013). Neural mechanisms of communicative innovation. *Proceedings of the National Academy of Sciences*, 110(36), 14574–14579. <https://doi.org/10.1073/pnas.1303170110>
- Sun, J., Harris, K., & Vazire, S. (2020). Is well-being associated with the quantity and quality of social interactions?. *Journal of Personality and Social Psychology*, 119(6), 1478–1496. <https://doi.org/10.1037/pspp0000272>
- Ta, V. P., Babcock, M. J., & Ickes, W. (2017). Developing latent semantic similarity in initial, unstructured interactions: The words may be all you need. *Journal of Language and Social Psychology*, 36(2), 143–166. <https://doi.org/10.1177/0261927X16638386>
- Tannen, D. (Ed.). (1993). *Framing in discourse*. Oxford University Press. <https://doi.org/10.1017/s0047404500019047>
- Tannen, D. (2007). *Talking voices: Repetition, dialogue, and imagery in conversational discourse*. Cambridge University Press. (Original work published 1989)
- Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences*, 119(4), e2116915119. <https://doi.org/10.1073/pnas.2116915119>
- Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2023). Long gaps between turns are awkward for strangers but not for friends. *Philosophical Transactions of the Royal Society B*, 378(1875), 20210471. <https://doi.org/10.1098/rstb.2021.0471>
- Templeton, E. M., & Wheatley, T. (2023). Listening fast and slow. *Current Opinion in Psychology*, 53, 101658. <https://doi.org/10.1016/j.copsyc.2023.101658>

- Terasaki, A. K. (2004). Pre-announcement sequences in conversation. In G. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 171–223). John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.125.11ter>
- Tickle-Degnen, L. (2006). Nonverbal behavior and its functions in the ecosystem of rapport. In V. Manusov & M. Patterson (Eds.), *The SAGE handbook of nonverbal communication* (pp. 381–400). SAGE Publications. <https://doi.org/10.4135/9781412976152.n20>
- Tolins, J., & Fox Tree, J. E. (2014). Addressee backchannels steer narrative development. *Journal of Pragmatics*, 70, 152–164. <https://doi.org/10.1016/j.pragma.2014.06.006>
- Tolins, J., Namiranian, N., Akhtar, N., & Fox Tree, J. E. (2017). The role of addressee backchannels and conversational grounding in vicarious word learning in four-year-olds. *First Language*, 37(6), 648–671. <https://doi.org/10.1177/0142723717727407>
- Tomasello, M. (2008). *Origins of human communication*. MIT Press. <https://doi.org/10.7551/mitpress/7551.001.0001>
- Tomasello, M. (2022). The coordination of attention and action in great apes and humans. *Philosophical Transactions of the Royal Society B*, 377(1859), 20210093. <https://doi.org/10.1098/rstb.2021.0093>
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5), 675–691. <https://doi.org/10.1017/S0140525X05000129>
- Tomlinson, J. M., Jr., & Fox Tree, J. E. (2011). Listeners' comprehension of uptalk in spontaneous speech. *Cognition*, 119(1), 58–69. <https://doi.org/10.1016/j.cognition.2010.12.005>
- Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity: Research, theory, and clinical applications. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(1), 3–48. <https://doi.org/10.1111/1469-7610.00701>
- Trujillo, J. P., Dideriksen, C., Tylén, K., Christiansen, M. H., & Fusaroli, R. (2023). The dynamic interplay of kinetic and linguistic coordination in Danish and Norwegian conversation. *Cognitive Science*, 47(6), e13298. <https://doi.org/10.1111/cogs.13298>
- Tylén, K., Fusaroli, R., Østergaard, S. M., Smith, P., & Arnoldi, J. (2023). The social route to abstraction: Interaction and diversity enhance performance and transfer in a rule-based categorization task. *Cognitive Science*, 47(9), e13338. <https://doi.org/10.1111/cogs.13338>

- van Arkel, J., Woensdregt, M., Dingemanse, M., & Blokpoel, M. (2020). A simple repair mechanism can alleviate computational demands of pragmatic reasoning: Simulations and complexity analysis. In *Proceedings of the 24th Conference on Computational Natural Language Learning* (pp. 177–194). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.conll-1.14>
- Wachsmuth, I., De Ruiter, J., Jaecks, P., & Kopp, S. (2013). *Alignment in communication: Towards a new theory of communication*. John Benjamins Publishing Company. <https://doi.org/10.1075/ais.6>
- Ward, N., & Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, 32(8), 1177–1207. [https://doi.org/10.1016/s0378-2166\(99\)00109-5](https://doi.org/10.1016/s0378-2166(99)00109-5)
- Warner, R. M. (1979). Periodic rhythms in conversational speech. *Language and Speech*, 22(4), 381–396. <https://doi.org/10.1177/002383097902200407>
- Warner, R. (1988). Rhythm in social interaction. In J. E. McGrath (Ed.), *The social psychology of time* (pp. 63–88). Sage.
- Warner, R. M. (1992a). Cyclicity of vocal activity increases during conversation: Support for a nonlinear systems model of dyadic social interaction. *Behavioral Science*, 37(2), 128–138. <https://doi.org/10.1002/bs.3830370204>
- Warner, R. M. (1992b). Speaker, partner and observer evaluations of affect during social interaction as a function of interaction tempo. *Journal of Language and Social Psychology*, 11(4), 253–266. <https://doi.org/10.1177/0261927x92114004>
- Warner, R. M. (2002). Rhythms of dialogue in infancy: Comments on Jaffe, Beebe, Feldstein, Crown, and Jasnow (2001). *Journal of Psycholinguistic Research*, 31, 409–420. <https://doi.org/10.1023/a:1019573825857>
- Warner, R. M., Malloy, D., Schneider, K., Knoth, R., & Wilder, B. (1987). Rhythmic organization of social interaction and observer ratings of positive affect and involvement. *Journal of Nonverbal Behavior*, 11(2), 57–74. <https://doi.org/10.1007/bf00990958>
- Watzlawick, P., & Beavin, J. (1967). Some formal aspects of communication. *American Behavioral Scientist*, 10(8), 4–8. <https://doi.org/10.1177/0002764201000802>
- Weaver, W. (1949). Recent contributions to the mathematical theory of communication. In *The mathematical theory of communication* (pp. 1–28). University of Illinois Press.
- Welker, C., Walker, J., Boothby, E., & Gilovich, T. (2023). Pessimistic assessments of ability in informal conversation. *Journal of Applied Social Psychology*, 53(7), 555–569.

<https://doi.org/10.1111/jasp.12957>

Wertsch, J. V. (2003). Ragnar Rommetveit: His work and influence. *Mind, Culture, and Activity*, 10(3), 183–185. https://doi.org/10.1207/s15327884mca1003_1

Wheatley, T., Boncz, A., Toni, I., & Stolk, A. (2019). Beyond the isolated brain: The promise and challenge of interacting minds. *Neuron*, 103(2), 186–188. <https://doi.org/10.1016/j.neuron.2019.05.009>

Wheatley, T., Thornton, M. A., Stolk, A., & Chang, L. J. (2024). The emerging science of interacting minds. *Perspectives on Psychological Science*, 19(2), 355–373. <https://doi.org/10.1177/17456916231200177>

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, 37(3), 395–412. <https://doi.org/10.1037/0022-3514.37.3.395>

Wilson, K. S., DeRue, D. S., Matta, F. K., Howe, M., & Conlon, D. E. (2016). Personality similarity in negotiations: Testing the dyadic effects of similarity in interpersonal traits and the use of emotional displays on negotiation outcomes. *Journal of Applied Psychology*, 101(10), 1405–1421. <https://doi.org/10.1037/apl0000132>

Winkielman, P., Niedenthal, P., Wielgosz, J., Eelen, J., & Kavanagh, L. C. (2015). Embodiment of cognition and emotion. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology, Vol. 1. Attitudes and social cognition* (pp. 151–175). American Psychological Association. <https://doi.org/10.1037/14341-004>

Wohltjen, S., & Wheatley, T. (2021). Eye contact marks the rise and fall of shared attention in conversation. *Proceedings of the National Academy of Sciences*, 118(37), e2106645118. <https://doi.org/10.1073/pnas.2106645118>

Wolf, W., Nafe, A., & Tomasello, M. (2021). The development of the liking gap: Children older than 5 years think that partners evaluate them less positively than they evaluate their partners. *Psychological Science*, 32(5), 789–798. <https://doi.org/10.1177/0956797620980754>

Wynn, C. J., & Borrie, S. A. (2022). Classifying conversational entrainment of speech behavior: An expanded framework and review. *Journal of Phonetics*, 94, 101173. <https://doi.org/10.1016/j.wocn.2022.101173>

Yeomans, M., Boland, F. K., Collins, H. K., Abi-Esber, N., & Brooks, A. W. (2023). A practical guide to conversation research: How to study what people say to each other. *Advances in Methods and Practices in Psychological Science*, 6(4). <https://doi.org/10.1177/25152459231183919>

- Yeomans, M., Brooks, A. W., Huang, K., Minson, J., & Gino, F. (2019). It helps to ask: The cumulative benefits of asking follow-up questions. *Journal of Personality and Social Psychology*, 117(6), 1139–1144. <https://doi.org/10.1037/pspi0000220>
- Yeomans, M., Kantor, A., & Tingley, D. (2018). The politeness package: Detecting politeness in natural language. *R Journal*, 10(2), 489–502. <https://doi.org/10.32614/rj-2018-079>
- Yeomans, M., Schweitzer, M. E., & Brooks, A. W. (2022). The Conversational Circumplex: Identifying, prioritizing, and pursuing informational and relational motives in conversation. *Current Opinion in Psychology*, 44, 293–302. <https://doi.org/10.1016/j.copsyc.2021.10.001>
- Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017). Same story, different story: The neural representation of interpretive frameworks. *Psychological Science*, 28(3), 307–319. <https://doi.org/10.1177/0956797616682029>
- Yngve, V. H. (1970). On getting a word in edgewise. In *Papers from the sixth regional meeting Chicago Linguistic Society, April 16-18* (pp. 567–578). Chicago Linguistic Society.
- Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2020). Polite speech emerges from competing social goals. *Open Mind*, 4, 71–87. https://doi.org/10.1162/opmi_a_00035
- Yzerbyt, V. Y. (2018). The dimensional compensation model: Reality and strategic constraints on warmth and competence in intergroup perceptions. In A. E. Abele & B. Wojciszke (Eds.), *The agency-communion framework* (pp. 126–141). Routledge. <http://dx.doi.org/10.4324/9780203703663-11>
- Zajonc, R. B. (1960). The process of cognitive tuning in communication. *The Journal of Abnormal and Social Psychology*, 61(2), 159–167. <https://doi.org/10.1037/h0047987>
- Zhao, X., & Epley, N. (2021). Kind words do not become tired words: Undervaluing the positive impact of frequent compliments. *Self and Identity*, 20(1), 25–46. <https://doi.org/10.1080/15298868.2020.1761438>
- Zhao, X., & Epley, N. (2022). Surprisingly happy to have helped: Underestimating prosociality creates a misplaced barrier to asking for help. *Psychological Science*, 33(10), 1708–1731.
- Zimmerman, D. H. (1993). Acknowledgment tokens and speakership incipency revisited. *Research on Language and Social Interaction*, 26(2), 179–194. https://doi.org/10.1207/s15327973rlsi2602_4

