AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 3, 864–889
https://doi.org/10.1037/xge0001301

Quantifying the Interplay of Conversational Devices in Building Mutual Understanding

Christina Dideriksen[1], Morten H. Christiansen[1, 2, 3], Kristian Tylén[1, 2], Mark Dingemanse[4], and Riccardo Fusaroli[1, 2]

[1] School of Communication and Culture, Aarhus University
[2] The Interacting Minds Center, Aarhus University
[3] Department of Psychology, Cornell University
[4] Centre for Language Studies, Radboud University

Humans readily engage in idle chat and heated discussions and negotiate tough joint decisions without ever having to think twice about how to keep the conversation grounded in mutual understanding. However, current attempts at identifying and assessing the conversational devices that make this possible are fragmented across disciplines and investigate single devices within single contexts. We present a comprehensive conceptual framework to investigate conversational devices, their relations, and how they adjust to contextual demands. In two corpus studies, we systematically test the role of three conversational devices: *backchannels*, *repair*, and *linguistic entrainment*. Contrasting affiliative and task-oriented conversations within participants, we find that conversational devices adaptively adjust to the increased need for precision in the latter: We show that low-precision devices such as backchannels are more frequent in affiliative conversations, whereas more costly but higher-precision mechanisms, such as specific repairs, are more frequent in task-oriented conversations. Further, task-oriented conversations involve higher complementarity of contributions in terms of the content and perspective: lower semantic entrainment and less frequent (but richer) lexical and syntactic entrainment. Finally, we show that the observed variations in the use of conversational devices are potentially adaptive: pairs of interlocutors that show stronger linguistic complementarity perform better across the two tasks. By combining motivated comparisons of several conversational contexts and theoretically informed computational analyses of empirical data the present work lays the foundations for a comprehensive conceptual framework for understanding the use of conversational devices in dialogue.

Keywords: backchannels, common ground, conversational dynamics, interactive alignment, repair

Supplemental materials: https://doi.org/10.1037/xge0001301.supp

Conversation lies at the heart of language. Conversational exchanges are foundational to our social lives, critical for learning language and establishing friendships, as well as key to information-sharing, the acquisition of new knowledge, and the coordination of tasks and problem-solving across a range of contexts (Schegloff, 2006). Conversations are a joint undertaking, but one that seems surprisingly easy (Pickering & Garrod, 2004). Not unlike experienced dancers, interlocutors engage with each other in seamless coordination, synchronizing and complementing each other, as well as quickly solving problems when things go wrong (Clark, 1985; Clark & Brennan, 1991; Dale et al., 2013; Duran & Fusaroli, 2017; Fusaroli et al., 2016; Fusaroli, Gangopadhyay, et al., 2014). How are people able to conduct this "conversational dance" with such ease, ensuring mutual understanding through entrainment and coordination of subtle behavioral cues?

This article was published Online First December 15, 2022.

Christina Dideriksen https://orcid.org/0000-0002-0528-5530

Preliminary analyses of parts of this corpus, overlapping with Part 1, were presented at CogSci 2019 (see Dideriksen, Fusaroli, et al., 2019). This study was not preregistered. All data have been made publicly available at Open Science Framework and can be accessed at https://doi.org/gqw3.

This study was supported by the Danish Council for Independent Research through FKK Grant DFF-7013-00074 awarded to Morten H. Christiansen and seed funding grants from the Interacting Minds Center at Aarhus University. Mark Dingemanse acknowledges funding from NWO Grant 016.vidi.185.205. Moreover, we thank Patrick Healey and Sophie Skach for providing expert contributions to the informed priors. We also thank Jakob Steensig, and Kathrine Garly for defining and discussing coding schemes for repair and backchannel. Last, we thank all the research assistants, for their invaluable work transcribing and coding the datasets (see the online supplemental materials for a full list of names).

Correspondence concerning this article should be addressed to Christina Dideriksen, School of Communication and Culture, Aarhus University, Jens Chr. Skous Vej 2, 8000 Aarhus C, Denmark. Email: dideriksenchristina@gmail.com

The question of how people successfully coordinate with one another in conversations has been widely debated (Bangerter & Clark, 2003; Clark & Brennan, 1991; Fusaroli, Rączaszek-Leonardi, et al., 2014; Pickering & Garrod, 2004), with individual accounts typically pointing to different cognitive mechanisms and conversational devices. An integrative and detailed exploration of these diverse findings has potential implications and applications in several fields. A thorough understanding of how we create mutual understanding in conversation can accommodate the design of more successful human-computer interactions, as well as building a more detailed understanding of how effective team collaboration is facilitated through conversation. Furthermore, knowledge of how conversational devices are structured in conversation can help us better understand how atypical social functioning in neurodiverse people is expressed during interaction.

A key account of conversational coordination involves the notion of *common ground*, defined as "the sum of information that people assume they share" (Clark, 2009). Common ground is built and maintained as interlocutors share positive or negative evidence of understanding in conversation, a process also referred to as *grounding of situation models* (Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). The grounding process is thought to involve pragmatic top-down mechanisms and a variety of conversational devices such as backchannels (positive evidence), repairs (negative evidence, or better, evidence of lack of mutual understanding), as well as lexical entrainment reflecting agreed on representations, or conceptual pacts (positive evidence; Brennan & Clark, 1996; Clark & Brennan, 1991).

Another prominent account of mutual understanding in conversation is *interactive alignment* (Pickering & Garrod, 2004). This account eschews the conception of common ground as explicitly shared representations, in favor of more implicit and automatic mechanisms (but see also Pickering & Garrod, 2021, for an integration of more explicit processes). When processing each other's linguistic forms, interlocutors are primed to reuse them (*linguistic alignment*). And as they converge on similar lexical expressions, interlocutors also come to align their syntax and semantics, and thereby end up sharing similar representations of the situation (Pickering & Garrod, 2004), which makes them able to communicate and interact successfully.

A third account, *interpersonal synergy* (Fusaroli et al., 2016; Fusaroli, Rączaszek-Leonardi, et al., 2014; Riley et al., 2011), takes an orthogonal stance and focuses on the cognitive challenge posed by conversations: as one interlocutor speaks, the other has to monitor and process their behaviors (from words to posture), produce and monitor their own reactions (from an attentive gaze to a backchannel), and plan their own turns (including anticipating the other's reactions) in a way that is sensitive to the goals of the interaction. The cognitive challenge to control all this complexity is facilitated by synergies: interlocutors locally couple and constrain their own and each other's degrees of freedom, greatly reducing the amount of control needed. In other words, the different behaviors enacted in conversations are made interdependent. Linguistic entrainment is an obvious example: by reusing each other's words we are reducing the space of possible utterances, simplifying the coordination. However, although linguistic entrainment might be pervasive and play an important role, just repeating each other is not enough to successfully coordinate, especially when trying to solve a task together. Indeed, interpersonal synergies are crucially constituted by complementary behaviors and routines sensitive to contextual demands. Just like muscular synergies in our arms are assembled on the fly to carefully hold a precious cup, or more forcefully take hold of a handrail while slipping, different conversational devices, including linguistic entrainment, are modulated and interconnected to fit contextual demands.

Mutual understanding between speakers has often been treated as a one-size-fits-all concept. The vast majority of quantitative studies have thus focused only on one conversational context assessing whether, for instance, linguistic entrainment was a good predictor of the success of the conversation, in terms of rapport or performance in a specific task (for a review, see, e.g., Duran et al., 2019). However, people continuously adjust their expectations and behaviors to the particular context of a conversation (e.g., see the *grounding principle* in Clark & Brennan, 1991, and the notion of *functional assemblies* in Fusaroli, Rączaszek-Leonardi, et al., 2014; Dale, 2015). For example, conversations between pilots and control towers are likely to require high levels of precision and are explicitly structured to ensure more frequent checks for mutual understanding than a casual dinner party conversation (Prinzo & Britton, 1993). By contrast, in more casual conversations, precise mutual understanding might not always be necessary, or even sought for. Indeed, Galantucci and colleagues repeatedly found interlocutors to be insensitive to miscommunication and to avoid repair as much as possible when engaged in casual conversation (Galantucci et al., 2020; Galantucci & Roberts, 2014).

Given these differences in contextual demands, we expect the use of conversational devices to vary across contexts (Fusaroli et al., 2017; Healey et al., 2014; Reitter & Moore, 2014). Specifically, in conversational contexts demanding a high degree of precision in information sharing, we should expect variation in the use of conversational devices to be associated with variation in performance. For instance, proponents of the interactive alignment model have shown that the more interlocutors entrain in a task-oriented task, the better they perform on this task (Fusaroli & Tylén, 2016; Gries, 2005, p. 20; Reitter & Moore, 2006, 2014; Slocombe et al., 2013). Nonetheless, approaches based on the interpersonal synergy account of conversation have argued that effective conversations do not rely on similarity alone. On the contrary, effective coordination requires complementary contributions adapted to the current context (Fusaroli, Rączaszek-Leonardi, et al., 2014). Indeed, too much entrainment might be deleterious: highly entrained interlocutors might end up merely repeating each other without contributing new information, which will impede performance (Fusaroli et al., 2012; Fusaroli, Gangopadhyay, et al., 2014; Tylén et al., 2020; Xu & Reitter, 2017).

Inspired by such accounts, in this article we combine these different foci of analysis to critically investigate how conversational devices contribute to the mutual understanding between speakers. In particular, we focus on three prominent linguistic devices and their interrelations: *backchannels*, *conversational repair*, and *linguistic entrainment* (Bangerter & Clark, 2003; Clark & Brennan, 1991; Dale et al., 2013; Dingemanse & Enfield, 2015; Fusaroli et al., 2017; Gardner, 2001; Mills et al., 2017; Pickering & Garrod, 2004; Schegloff, 1982, 2000; Yngve, 1970). Backchannels are subtle cues that signal understanding or agreement (positive evidence of understanding; Bangerter & Clark, 2003; Clark & Brennan, 1991; Schegloff, 1982; Yngve, 1970).[1]

[1] Whereas backchannels frequently feature multimodal aspects (as in head nods or eye blinks that may accompany or replace verbal forms (Hömke et al., 2018), our empirical scope in the studies reported here is on verbal backchannels.

Conversational repair refers to linguistic resources that signal a potential misunderstanding and request its resolution (negative evidence of understanding; Dingemanse et al., 2015; Dingemanse & Enfield, 2015; Schegloff, 2000; Schegloff et al., 1977). Lastly, linguistic entrainment—independently of the underlying mechanism assumed—indicates the reuse of one's interlocutor's lexical choices, syntax and semantic topics. Linguistic entrainment has been variously argued to facilitate and reflect mutual understanding in conversation sometimes by signaling and promoting a common representation of the situation, sometimes by indicating a lack of diverse contributions to the problem (Brennan & Clark, 1996; Duran et al., 2019; Garrod & Anderson, 1987; Pickering & Garrod, 2004). We refer to these central linguistic processes as *conversational devices* motivated by their function as facilitators for the establishment of mutual understanding in conversations between speakers. Although a large number of studies have looked at individual conversational devices in isolation, in the current article, we investigate them together in a systematic and quantitative manner.

By experimentally manipulating contextual demands, in particular the need for precise information sharing, we present theoretically motivated constraints of how conversational devices facilitate mutual understanding in conversations. By measuring the relationship between the use of conversational devices and task performance across two diverse tasks, we provide robust and generalizable hypotheses for their functional role.

The aim of this article is to develop a conceptual framework to better understand and gain new insights into the use and function of fundamental conversational devices underpinning the collaborative dynamics of conversation. We do this by (a) providing empirically grounded, theoretically informed, and standardized ways to identify, quantify, and model three conversational devices from the current literature; (b) investigating these three devices and their relations as they systematically vary across two well-motivated types of conversation, and (c) quantifying their relation to performance across two diverse tasks. Moreover, we aim to develop an open science pipeline to enable replicable and transparent progress in the scientific study of conversational dynamics. In particular, we provide reproducible methodological tools, such as manual and automated coding schemes, appropriate distributional modeling of the data, and full release of the preprocessing and analytical code. In this way, we ensure our conceptual contributions are seamlessly integrated and support methodological and empirical contributions.

Next, we detail the different kinds of conversational devices before reporting on our two-part study of Danish conversations in the subsequent sections (building on an exploratory study reported in the online supplemental materials) and conclude with a broad discussion of the dynamics of conversation.

General Hypotheses

The following study investigates backchannels, conversational repair, and linguistic entrainment, in two different conversational contexts: task-oriented conversations and affiliative conversations. Based on the notion that task contexts are often more informationally dense, we hypothesize that conversational devices which enable higher referential precision in the construction and maintenance of shared knowledge, will be more frequent in task-oriented conversations (H1). In addition, we explore how the individual mechanisms are related, and whether they are interdependent. Furthermore, since we argue that conversational devices have an adaptive function, we hypothesize that pairs of interlocutors that promote relevant conversational devices will perform better in the tasks (H2). In other words, if we observe task-oriented conversations to be associated with, for instance, higher entrainment levels than affiliative conversations, we would also expect pairs with higher entrainment levels to perform better in that task than those with lower entrainment levels. We investigate H1 in a controlled within-subject corpus collected for the explicit purpose of task comparison. H2 is then assessed on a subset of the same corpus, including only the task-oriented conversations.

It should be noted that none of these hypotheses should be taken to the extreme: a conversation consisting only of backchannels or with no reuse of linguistic forms from the interlocutor is unlikely to be successful. We are only stipulating these hypotheses within relatively naturalistic adult-adult conversations and the typical ranges of backchannels, repairs, and entrainment they entail. Further, although we manipulate the contextual demands of the conversations (H1), the assessment of the adaptive role of conversational devices (H2) is purely observational and will need further experimental evidence to enable more robust inferences of causality (see Discussion).

Conversational Devices

Research on conversational devices—although productive—has been challenged by a disconnect between the various disciplines involved. The sociologically oriented field of Conversation Analysis has carried out qualitative work identifying and describing conversational devices such as backchannels and repair, and studying their function in interaction (for instance, Gardner, 2001; Schegloff et al., 1977; Yngve, 1970). In contrast, the field of psycholinguistics has experimentally investigated linguistic entrainment in various forms (for instance, Brennan & Clark, 1996; Garrod & Anderson, 1987; Pickering & Garrod, 2004; Reitter & Moore, 2006, 2014). Although these studies could be argued to address related phenomena, the use of radically different methods makes it difficult to compare results across studies, thereby restricting more general theoretical claims and hypotheses (de Ruiter & Albert, 2017). Here, we combine insights from both fields, to investigate the role of conversational devices within a unified framework.

We focus on three different conversational devices: backchannels, conversational repair, and linguistic entrainment and how they are differentially engaged contingent on varying contextual demands, such as the need for more or less precise information sharing across different contextual settings. It is important to note that these are not the only conversational features that promote mutual understanding or shared conceptualization (see Clark & Brennan, 1991, for a number of processes whereby one can assume that the grounding criterion has been met), and the incorporation of additional devices should be included in future studies.

Backchannels

Backchannels are subtle cues produced by the recipient in response to a current speaker (Yngve, 1970). They mostly consist of short words like "*yes*," "*uh-huh*" or "*okay*" but they can also be short utterances that repeat what the speaker just said (A: "I'll bring cake, then you can bring coffee." B: "I'll bring coffee!"). Even though backchannels are typically produced while someone else is speaking, they are not considered an interruption. Rather,

they act as an affirmation that the listener is paying attention to the conversation and that there are currently no problems of mutual understanding between the interlocutors (Gardner, 2001).

Backchannels can also act as "continuers" (Schegloff, 1982), signaling both that some input has been received and that more is expected, and that the recipient has no intention of taking the floor (Bangerter & Clark, 2003; Goodwin, 1986; Jefferson, 1984; Schegloff, 1982; Yngve, 1970). This continuer use is illustrated here in Danish examples (see Example 1) with English gloss (using boldface to highlight the particular device).

Example 1
Backchannel

A	og dreje mod øst efter guldminen sådan at man - at guldminen kommer til at gå—være nord om for en and turn east after the goldmine so that you - that the goldmine is going to - will be North of you
B	**ja** **yes**
A	og så skal man efter guldminen gå cirka en centimeter længere mod øst og så dreje en lille smule mod nord and, after the goldmine you move about one centimeter further East and then turn a bit towards the North

Previous studies have found backchannels to be more prevalent in task-oriented conversations, where they support precise information sharing in grounding processes by continuously confirming mutual understanding between speakers (Fusaroli et al., 2017; see also the Preliminary Exploratory Study in the online supplemental materials). Based on this, we expect to see a higher rate of backchannels in task-oriented conversations, and an even higher prevalence of them in the better performing pairs.

Repair

Whereas backchannels serve as positive feedback to the speaker, conversational repair comes into play when mutual understanding is potentially compromised and needs to be reestablished or indeed repaired. By using a repair request, an interlocutor indicates they have not fully heard or understood the previous sentence and that they need a clarification. Repair requests (see Example 2) can take different forms and vary in degree of specificity (Schegloff et al., 1977), but they all share the main function of mending possible break-downs in mutual understanding between interlocutors.

Example 2
Repair (Restricted Offer)

A	fra fra runestenene altså sådan lidt højere oppe end der hvor der står guldmine from from the runic stones, like just a little higher up than where it says goldmine
B	okay hvad sagde du det var? **En kirkegård**? okay what did you say it was? **A cemetery**?
A	en kirkegård, ja a cemetery, yes

Previous work has identified three basic ways to initiate repair, on a scale from general to specific (Dingemanse et al., 2015): *open requests* call for repair of a prior utterance while leaving open what or where the problem is. They are often initiated with short expressions like "*huh?*" or "*what?*." *Restricted requests*, by contrast, point to a specific part of the utterance that needs clarification. They often feature question words like "*where?*," "*who?*" or "*which?*" along with expressions that repeat or refer to some element of the prior utterance. Finally, *restricted offers* put forward a possible interpretation of the prior utterance: "*the main street?*" or "*the Labrador?*" They restrict the problem space by pointing directly to the troublesome aspect and offer a possible solution that enables the original speaker to respond, in many cases, with a simple confirmation.

In line with the *principle of least collaborative effort* (Clark & Wilkes-Gibbs, 1986), comparative cross-linguistic work has found a general preference for using more specific repair types: people tend to choose the most specific repair initiation type possible given the circumstances (Dingemanse et al., 2015). Assuming that interlocutors try to minimize collaborative effort, initiating a repair by offering the most general repair type would be less helpful and require more effort.

The contextually adaptive nature of repair initiation has only rarely been studied contrastively: one study finds a higher frequency of repairs in task-oriented conversations compared with affiliative conversations (Mills et al., 2017), whereas others suggest that repairs might facilitate the construction of shared abstract conventions (Mills, 2014). Nonetheless, the importance of repair in the negotiation of mutual understanding and the sharing of information (in particular with restricted requests and restricted offers) indicates that they might be most prominent in contexts that require a high degree of referential precision, such as task-oriented conversations. In other words, interlocutors might be less likely to check their mutual understanding, and more likely to overlook minor signs of miscommunication when the need of precise information sharing is lower, as in affiliative conversations (Galantucci et al., 2020; Galantucci & Roberts, 2014).

Linguistic Entrainment

A third mechanism that has been suggested to be crucial for mutual understanding between interlocutors is linguistic entrainment (Bangerter & Clark, 2003, 2003; Brennan & Clark, 1996; Dale et al., 2013; Duran et al., 2019; Fusaroli et al., 2012; Fusaroli & Tylén, 2016; Mills, 2014; Pickering & Garrod, 2004). Linguistic entrainment has been conceptualized in a variety of ways (Rasenberg et al., 2020). For example, Pickering and Garrod construe interactive alignment as an implicit mechanism leading to mutual understanding (Goudbeek & Krahmer, 2012; Pickering & Garrod, 2004). An interlocutor's linguistic production is mutually primed by the partner's production, leading to the production of more similar linguistic forms that, in turn, automatically and implicitly promote a shared representation of the situation (alignment). By sharing a mutual understanding of the situation, then, interlocutors can more easily coordinate their efforts on a joint task or problem. Other approaches have conceptualized linguistic entrainment as a more explicit mechanism, whereby individuals reuse their interlocutors' expressions in a deliberate and targeted way, to signal and express mutual understanding (e.g., "conceptual pacts," Brennan & Clark, 1996; Carbary & Tanenhaus, 2011). For the purpose of this study, we use the more general term of "linguistic entrainment" and will not address the implicit or explicit nature of the underlying cognitive mechanism.

No matter the mechanism, interlocutors have been observed to entrain on several parameters, from lexical and syntactic choices to relatively lower-level features such as bodily postures, facial expression, and accent (Branigan et al., 2007; Brennan & Clark, 1996; Chartrand & Bargh, 1999; Chartrand & Lakin, 2013; Dumas & Fairhurst, 2021; Garrod & Anderson, 1987; Healey et al., 2014; Louwerse et al., 2012; Misiek et al., 2020; Rasenberg et al., 2020; Scheflen, 1964; Shockley et al., 2003). However, the function of entrainment, and linguistic entrainment in particular, has been debated, with different studies focusing either on the importance of alignment and similarity (Pickering & Garrod, 2004) or on that of dissimilarity and complementarity of contributions (Fusaroli, Rączaszek-Leonardi, et al., 2014; Tylén et al., 2020).

Here, we focus on the verbal aspects of entrainment: the interlocutors' propensity to reuse each other's linguistic forms in adjacent conversational turns; specifically, lexical items, syntactic constructions, or semantic content. Entrainment could, of course, happen over longer time scales and across multiple turns. Here we focus on adjacent turns as they represent a particularly salient and often investigated phenomenon (for a broader discussion of the time scales at work, see Duran et al., 2019). Notice that episodes of entrainment may occasionally overlap with both backchannel and restricted offers, where part of the previous utterance is often repeated to explicitly signal understanding or request clarification (Fusaroli et al., 2017).

Lexical entrainment (see Example 3) is the perhaps most studied form of linguistic entrainment, whereby interlocutors reuse each other's lexical material (Brennan & Clark, 1996; Fusaroli et al., 2012).

Example 3
Lexical Entrainment

A	**jeg har** ikke nogen altså jeg har en stor klippe men den ligger **I** don't **have** any, I mean, I have a big rock, but it's
B	**jeg har** to **I have** two

In the above example, the lexical forms "I" and "have" uttered by speaker A are repeated by speaker B, constituting a case of lexical entrainment. Note that different words have different baseline rates of occurrence and that could confound measures of entrainment (Healey et al., 2014; we return to this issue in the Method section).

Interlocutors can also entrain on the syntactic level. For the current purposes, the repetition of consecutive sequences of parts of speech (bigrams or trigrams) from a previous utterance acts as a proxy for syntactic entrainment (see Example 3; including cases of whole or partly lexical repetition, see Example 4) and has been shown to yield analogous results to the explicit modeling of syntactic trees (Reitter & Moore, 2006). For instance, an interlocutor can repeat the use of pronouns and verb forms.

Example 4
Syntactic Entrainment

A	det er faktisk slet ikke godt PRON VB ADV ADV ADV ADJ that's actually not very good
B	det er overhovedet ikke sundt PRON VB ADV ADV ADJ that's not healthy at all

Here speaker A's sequence of pronoun (PRON), verb (VB), adverbs (ADV) and adjective (ADJ) are repeated in the identical order by speaker B, although most of the specific words change. Syntactic entrainment is more abstract than lexical repetition, and we expect it to be pervasive (as the inventory of parts of speech is limited, making n-grams more likely to co-occur across utterances). It has been argued to play an important role in conversational success, as it catalyzes and/or reflects a shared understanding of the situation (Reitter & Moore, 2014; it has also been connected to language development: Fusaroli et al., 2021; Rowland et al., 2012).

We also investigate semantic entrainment (see Example 5): that is, the extent to which people stay with the similar topics and themes although not necessarily using the same words and syntax.

Example 5
Semantic Entrainment

A	den dér er anderledes end det første **rumvæsen** that one is different from the first **alien**
B	det ville sikkert være nemmere hvis vi havde et **rumskib** it would probably be easier if we had a **spaceship**

Here the two interlocutors, although not entraining lexically, use words that are semantically related and therefore entrain on their topic, namely "outer space."

Several studies have found relations between linguistic entrainment and performance in joint tasks (positive relations: Fusaroli & Tylén, 2016; Gries, 2005; Reitter & Moore, 2014, 2006; Slocombe et al., 2013; negative relations: Fusaroli et al., 2012; Xu & Reitter, 2017). However, going beyond previous studies, we will distinguish between two components of lexical and syntactic entrainment: rate and level. *Rate of entrainment* is defined as the frequency of utterances that contain any entrainment relative to an immediately prior utterance, no matter how much. *Level of entrainment* is defined as the relative amount of repeated items (for instance, words) within adjacent utterances displaying entrainment. This is motivated conceptually and supported by the data. We might imagine conversations with lots of "touching base" by frequent repetitions of single key lexical or syntactic forms, but without any extensive repetition of full utterances (high rate, low level, see Example 6a). On the other hand, we can also imagine conversations with very infrequent entrainment, but when entrainment happens, full utterances are repeated (low rate, high level, see Example 6b). This distinction is further supported by the data: many utterances present no lexical and syntactic entrainment (zero inflation), whereas utterances containing entrainment have more normally

Example 6a
High Rate, Low Level in an Affiliative Conversation (Words That Are the Same Color Are Entrained)

	Danish	English translation
A	men- men det vel lige så fint, fordi det jo både din lønseddel og en- altså det jo både en lønseddel	but- but it is just as good, because it is both your paycheck and a- you know it's both a paycheck
B	skal jeg ikke bare vedhæfte årsopgørelsen og så se om den ikke går	shouldn't I just attach the tax return and then see if it goes through?
A	jo nemlig. den går da, fordi det j- den har jo begge dele	exactly. It will go through, because it- it has both
B	altså jeg har ikke nogen	well, I don't have any

Note. See the online article for the color version of this table.

distributed entrainment scores (centered somewhere around .30 and decreasing in frequency at both sides). This can be described as a bimodal distribution of lexical and syntactic entrainment, which requires us to separately deal with the two modes (rate and level, see the Method section and Figure 3).

Example 6a displays an affiliative conversation where there is some entrainment in all adjacent utterances (high rate), however, only few words are repeated (low level).

Example 6b

Low Rate, High Level in Task-Oriented Conversations (Words That Are the Same Color Are Entrained)

	Danish	English translation
A	det var den der ikke?	it's that one, right?
B	ja mm	yes uh-huh
A	den har ikke pletter	it does not have spots
B	den har ikke pletter det er næsten det samme som ja	it does not have spots, it's almost the same as, yes
A	ja det var det det var blå okay måske kan man ikke få lov så	yes it was it it was blue, okay, maybe you're not allowed then
B	nej ha ha	no ha ha
A	er det også sådan en blå?	is it also this kind of blue?
B	ja	yes
A	det gjorde vi også til den der	we did the same for that one

Note. See the online article for the color version of this table.

Example 6b displays a task-oriented conversation in which relatively few utterances are aligned (low rate), but more words from the previous utterance are aligned (high level).

Given the novelty of these measures and the contradictory previous findings we rely on a single study for hypotheses for entrainment (see Section 1 in the online supplemental materials). Based on this, we expect to see a contextual difference with higher entrainment rates in affiliative conversations, but low levels and the opposite pattern in task-oriented conversations: high entrainment levels, but low rates. These hypotheses attempt to combine previous theoretical approaches (e.g., the need for similarity and for complementarity) with the empirical findings established in the exploratory analysis reported in the online supplemental materials.

Relations Between Conversational Devices

One important question is also how independent these conversational devices are from each other, and whether they can meaningfully be analyzed in isolation. For instance, it is a tenet of the Interactive Alignment framework that linguistic entrainment at any level catalyzes the entrainment of all other levels (Pickering & Garrod, 2004). Therefore, higher lexical entrainment should go hand in hand with syntactic entrainment (as also shown in Rowland et al., 2012). Analogously, Fusaroli et al. (2017) have shown that specific forms of repairs often involve lexical repetition, and therefore should covary with entrainment. The interpersonal synergy account would also suggest that there are multiple ways in which interlocutors can reduce the complexity of the interaction potentially resulting in trade-offs in the use of conversational devices (Fusaroli, Rączaszek-Leonardi, et al., 2014). For instance, Schegloff (1982) conceptualized backchannels as passing opportunities for repair, which would imply that more backchannels could be related to fewer repairs. Furthermore, interlocutors relying more on lexical entrainment as a conversational device for mutual understanding have been shown to also display fewer backchannels and repairs (Fusaroli et al., 2017). It seems therefore crucial to better understand the network of interdependencies in which conversational devices are used.

Preliminary Exploratory Study

Prior to the current study we conducted an exploratory between-subjects study relying on two already existing corpora, the DanTIN (Steensig et al., 2013) and the DanPASS corpora (Grønnum, 2009). The DanTIN corpus consists of affiliative conversations, whereas the DanPASS corpus consists of task-oriented conversation, where participants solved the Map Task (Anderson et al., 1991). The two corpora were small (DanTIN: $N = 18$; DanPASS: $N = 22$) and varied along a number of properties, which made them suboptimal for direct contrastive comparison between conversations. However, this exploratory study was aimed at developing and testing the methods to analyze our within-subject corpus, and to identify meaningful priors. The analyses are the same as those described in Part I.

We found that backchannels, restricted offers, and lexical and syntactic entrainment level were more frequent in task-oriented conversations, whereas open requests, lexical and syntactic entrainment rate, and semantic entrainment were more frequent in affiliative conversations. Thus, the key findings suggested that the increased need for referential precision in task-oriented conversations would be related to a more frequent occurrence of conversational devices. See the online supplemental materials for a detailed overview of the corpus, methods, and findings.

Part I: Conversations

Part I sets out to systematically investigate the hypotheses in a large and controlled within-subject experimental design. This approach allows us to directly test whether the same participants modulate their use of conversational devices as contextual demands change.

Most previous studies rely on a single conversational context: for instance, informal phone conversations, or a specific collaborative task involving navigating a map or jointly making a decision. In an attempt to maximize the comparability of our findings to the previous literature and to address the heterogeneity of real-world conversational contexts, we opted for collecting four different conversations from each pair of participants: two sessions of affiliative conversations and of task-oriented conversations covering a number of tasks used in previous studies.[2] Note that the distinction between task-oriented and affiliative conversations likely belongs on a continuum. Affiliative concerns, such as not losing face, being polite, etc., are also present during task-oriented conversations although they are likely to have a relatively lower salience, as other concerns—such as performing well on the task—will be at the forefront.

The two affiliative conversations varied by the degree of participants' familiarity, being situated at the beginning and at the end of

[2] Preliminary analyses of parts of this corpus were presented at CogSci 2019 (Dideriksen, Fusaroli, et al., 2019).

the data collection. The two task-oriented conversations varied in the particular tasks assigned. Previous studies investigating conversational strategies in task-oriented conversations have often applied tasks from one of two categories: more asymmetric director-matcher tasks (e.g., the Map Task, Anderson et al., 1991; or the visual perspective taking task, Hawkins et al., 2021; Keysar et al., 2000), or more symmetric tasks with both participants having access to equal amounts of information (e.g., the Optimally Interacting Minds task, Fusaroli et al., 2012; the Maze Game, Garrod & Anderson, 1987; and the tangram task, Clark & Wilkes-Gibbs, 1986; Hawkins et al., 2020). To be able to more meaningfully compare our findings to previous studies, as well as to assess the generality of our findings beyond one specific task, we chose to include two tasks that vary along these dimensions. The first task—the Alien Game—implemented a symmetric joint decision situation: the participants had to jointly assess whether different aliens were dangerous or not (see Figure 1 and Tylén et al., 2020). The second task, the Map Task, was more asymmetrical. Here one participant has access to information (a path on a map) that they have to share with the other (Anderson et al., 1991).

Given the studies reviewed above, we predict the increased need for accurate information sharing in task-oriented conversations to be related to an increased frequency of backchannels, and repairs compared with affiliative conversations. Moreover, we expect the type of repairs to be affected by the contexts, with a higher frequency of the more specific kinds of repair types in task-oriented conversations compared with affiliative conversations. Linguistic entrainment has shown more mixed results, with some studies arguing for a positive relation to performance (the more similar, the easier to coordinate, see Gries, 2005; Ireland & Henderson, 2014; Reitter & Moore, 2014; Slocombe et al., 2013) and others for a negative relation (the more similar, the less we contribute to each other's solutions, see Fusaroli et al., 2012; Tylén et al., 2020; Xu & Reitter, 2017). To refine our hypotheses, we explored two existing corpora of affiliative and task-oriented conversations (see Section 1 in the online supplemental materials). This led us to restate our hypotheses on backchannels and repairs and to hypothesize lower lexical and syntactic entrainment rates and semantic entrainment, but higher lexical and syntactic entrainment levels in task-oriented conversations. In other words, supported by exploratory analyses on a different corpus, we hypothesize that when more precise and complementary information sharing is required, entrainment will be less frequent, but involve more substantial repetitions of the previous utterance. This strongly resonates with our previous work on conversations as synergies, where complementarity and entrainment both play a role (Dale et al., 2013; Fusaroli, Rączaszek-Leonardi, et al., 2014; Fusaroli & Tylén, 2016).

Last, we explore whether the occurrences of conversational devices are related to each other and display interdependencies across mechanisms. For instance, we expect repair and entrainment to be related, because repair often involves the repetition of linguistic forms, but backchannel and repair to be negatively related, as interactions where mutual understanding is continuously confirmed should require fewer repairs (see Fusaroli et al., 2017).

Method

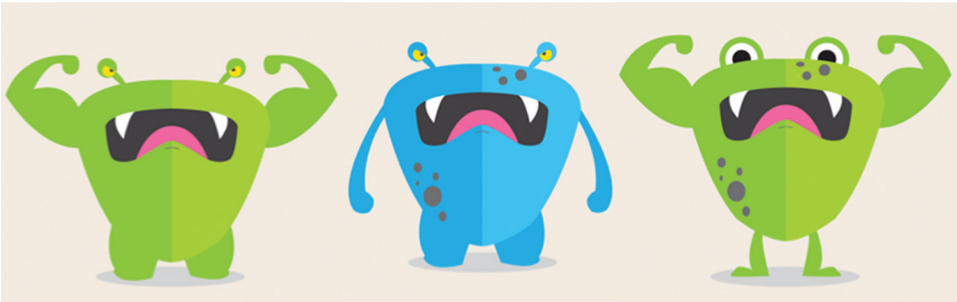
Corpora and Participants

We elicited conversations from 40 pairs (80 Danish individuals). Each pair produced four conversations: 2 affiliative conversations and 2 task-oriented conversations (for dataset see https://doi.org/gqw3, Dideriksen, Christiansen, et al., 2019). Prior to the conversations, participants filled out a questionnaire reporting their age (mean age = 23.2, *SD* = 3.5), gender (58% female participants), and education (5 participants had finished a high school degree, whereas the remaining 75 were students at Aarhus University). All participants had Danish as their first language, although 8 participants also had early exposure to at least one other language from birth (Italian, Cantonese, Tamil, Turkish, German, and English).

For the first affiliative conversation, we provided the participants with a sheet of open-ended conversation starters (e.g., "Discuss and agree on two superpowers that you would both like to have," see Section 2 in the online supplemental materials for the full list) and asked them to get acquainted for a while. Afterwards they were asked to complete the two collaborative tasks.

First, they played the Alien Game (Tylén et al., 2020), a joint decision task, where participants jointly make decisions about how to categorize a series of stimuli (in this case, aliens; see Figure 1). The stimulus categories related to visual features of the aliens, where for instance, the combination of spots and eyes on stalks would indicate

Figure 1
Sample of the Aliens Appearing in the Alien Game

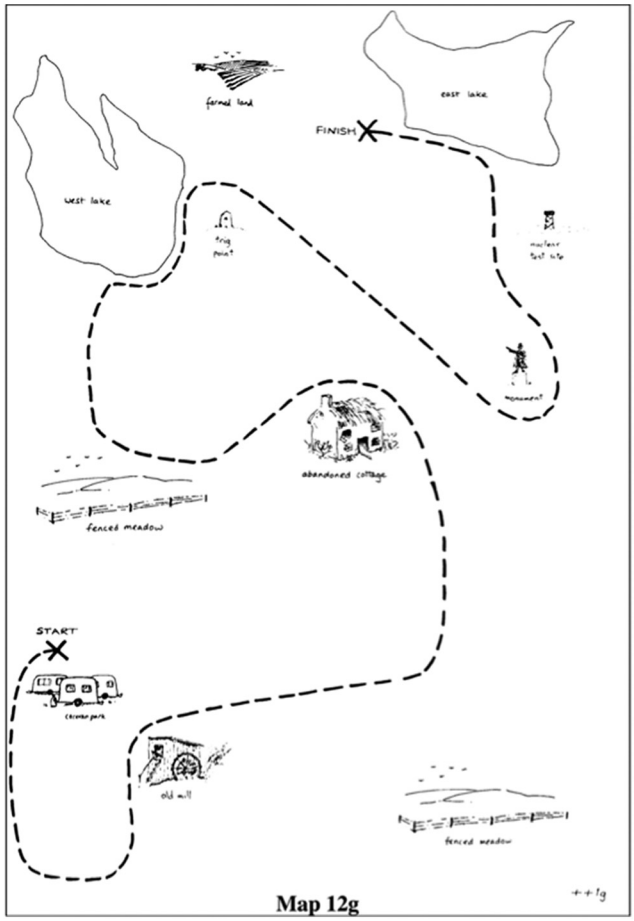


Note. "The Social Route to Abstraction: Interaction and Diversity Enhance Rule-Formation and Transfer in a Categorization Task," by K. Tylén, R. Fusaroli, P. Smith, & J. Arnoldi, 2020, *PsyArvix*, p. 8 (https://doi.org/10.31234/osf.io/qs253). CC-BY 4.0. Adapted with permission. See the online article for the color version of this figure.

whether the alien is friendly or hostile. Participants were explicitly instructed that different species of aliens might behave differently, without specifying that they should attend to visual features to classify them. Before the task started, examples of different aliens were visually displayed including the available choices and the participants were given time to discuss them. Then participants were presented with one alien at a time on a joint screen. After 3 s the alien disappeared and participants were instructed to jointly decide whether they thought the alien was hostile, friendly or valuable. The procedure made sure that decisions relied on participants' propensity to interactively share their experiences of the stimulus.

At first, participants would rely on a trial-and-error strategy to identify the correct choice, but gradually they seemed to learn the association between specific features deciding the category and became able to correctly identify also never-seen-before aliens. Correct answers were rewarded with 100 points, whereas wrong answers cost the pair a penalty of −100 points. The game terminated after 10 min, yielding a variable number of trials depending on the pace and collaborative style of the participants.

Figure 2
Director's Map in the Map Task



Note. "The HCRC Map Task Corpus," by A. H. Anderson et al., 1991, *Language and Speech, 34*(4) (https://doi.org/10.1177/002383099103400404). Copyright 1991 by SAGE. Reprinted with permission.

Second, participants were instructed to solve the Map task (see Figure 2 and Anderson et al., 1991), which is a director-matcher task. The task is an asymmetric game, where a director gives directions to a matcher as to how to draw a path on a map, with the matcher being free to interact, ask for explanations, etc. (see Figure 2). The participants each had their own individual monitor on which the maps appeared but were separated by a screen limiting their view of each other from the chest down; however, they were still able to see each other's facial expressions. The participants were introduced to the task and shown examples of the maps. At the beginning of the task a map would appear on both screens. Although the landmarks and a "start" marking were the same on both screens, the route only appeared on the director's screen. The task was for the matcher to draw a route as similar as possible to the route on the director's map guided only by the director's verbal instructions. Again, the game terminated after 10 min, yielding a variable number of maps solved by each pair. For the second affiliative conversation, if no conversation arose spontaneously, the participants were instructed to continue discussing the conversation starters provided for the first affiliative conversation. Often, however, the participants naturally continued talking casually about the games without any need for the experimenter to prompt them (23 of 39 pairs). This second affiliative conversation differed from the first not only in terms of simple increased familiarity between the participants but also because it took place after an extended conversational interaction with potential long term priming effects, and because the actual topics discussed were typically not the same. Each conversation lasted for about 10 min and participants were seated next to each other with visual access to each other's facial expressions throughout all conversations.

Due to technical issues, data from one pair and three conversations from other pairs had to be discarded. The final corpus consists of 39 pairs, 153 conversations, and a total of 38,259 conversational turns. The first affiliative conversation had an average of 226 utterances per conversation (min: 52; max: 351) and the second of 212 (min: 86; max: 395). The first task-oriented conversation (the Alien Game) had an average of 311 utterances per conversation (min: 241; max: 459) and the second (the Map Task) of 249 (min: 135; max: 384). All conversations were transcribed orthographically and coded manually for backchannel and repair. Twenty research assistants were trained to code the data aided by coding schemes with explicit examples explaining backchannels types and repair types (coding schemes can be retrieved here: https://doi.org/gqw3; see also example 1 and 2 for instances of backchannels and repairs). The Preliminary Exploratory study was used to validate the coding schemes and ensure intercoder reliability (see Section 1 in the online supplemental materials). Moreover, the data were coded for the entrainment analysis, as explained below.

Equipment

Conversations were recorded with a GoPro Hero 5 camera and sound was recorded with an omni-directional microphone connected to the camera. The stimulus presentation and response collections were set up in Psychopy3 (Peirce & MacAskill, 2018).

Statistical Analyses

To test whether backchannels are modulated by conversational context (task-oriented conversations vs. affiliative conversations), we built a Bayesian multilevel logistic regression. The presence or absence of backchannel for any given turn was

predicted by conversation type (task-oriented vs. affiliative). Individual and pair-level propensities to use backchannels were accounted for using separate but correlated varying intercepts for each conversational type for the rate parameter. Full details on the implementation, including priors, prior impact assessment and other quality checks are reported in the online supplemental materials (see Section 3).

Estimates from the model are reported as mean and 95% compatibility intervals (CI) of the posterior estimates. We tested whether conversational contexts modulate the use of conversational devices by calculating an evidence ratio (ER) in the form of the posterior probability of the directed hypothesis against the posterior probability of all the alternatives. That is, if we expected increased backchannels in affiliative conversations, we would count the posterior samples compatible with this hypothesis and divide them by the number of posterior samples compatible with a null or negative effect. For analyses without a directed hypothesis we assessed both directions of effects. We also report the credibility of the estimated parameter distribution, that is, the probability that the true parameter value is above 0 if the mean estimate is positive, or below 0 if it is negative. When our hypotheses were not supported by the data (evidence ratio below 3) or when our hypothesis was of no difference, we also tested for evidence in favor of the null. We also performed analyses at individual level, that is, if we observe the same effects at the population (grand average) and at the group (estimate by individual) level. Note that individual level effects are made more robust by our experimental setup, since keeping the order of the task constant makes any order effects the same for all participants. Because the results of the individual level analyses were in agreement with the population level analysis (all individuals displayed the same direction of effects as the population), the details are only reported in the online supplemental materials.

To test whether repair is modulated by conversational demands, we used the same procedure as for backchannels, but with absence or presence of repair as predicted outcome. To test the more specific hypothesis that the type of repair is affected by the context, we included only utterances coded as repair, and repeated the same procedure as for backchannels, with the specific type of repair as predicted outcome.

To test whether entrainment is modulated by contextual demands, we first had to quantify the degree of turn-to-turn lexical, syntactic and semantic entrainment between speakers. We relied on the standardization work and ALIGN package provided by Duran et al. (2019).

To prepare for lexical entrainment measures, all words were lemmatized, that is, we reduced all inflected forms to their dictionary form, or lemma, relying on the Danish language model on udpipe (Strømberg-Derczynski et al., 2020). In this way, "dogs" becomes "dog" and "are" becomes "be." To prepare for syntactic entrainment, we inferred parts-of-speech tags (Adverb, Noun, etc.) from the original (nonlemmatized) transcripts and produced for each utterance a list of *n*-grams (2-grams, 3-grams and 4-grams), that is, of contiguous sequences of parts of speech of length *n* within the same utterance. If an utterance contained fewer parts-of-speech than necessary to create the *n*-gram, syntactic entrainment to that sentence was considered impossible and therefore excluded from the analysis. Since the results are analogous using different *n*-grams, we only report those related to 2-grams (see Table 1 for an example), because 3- and 4-gram analyses excluded more utterances owing to insufficient sentence length. Note that our approach deviates from some previous approaches (e.g., Hopkins et al., 2016; Reitter & Moore, 2014) that removed words that were repeated across two turns before calculating syntactic entrainment. We chose not to do so because removing repeated words would most often break the parts-of-speech sequence and create artifactual 2-grams (see Example 1).

Example 7

A Depiction of How Removing Repeated Words Between Utterances Can Create Artifactual 2-Grams

Without removing repetitions		Removing repetitions	
A:	så skal man bare kende en masse	A:	så skal man bare kende en masse
PoS:	[CONJ; VB; PRP; ADV; VB; ART; N]	PoS:	[CONJ; VB; PRP; ADV; VB; ART; N]
eng	*then you just have to know a lot*	*eng*	*then you just have to know a lot*
B:	give dem eliksir så de ikke øh dør	B:	give dem eliksir ~~så~~ de ikke øh dør
PoS:	[VB; PRP; N; CONJ; PRP; ADV; INJ; VB]	PoS:	[VB; PRP; N; ~~CONJ~~; PRP; ADV; INJ; VB]
eng	*give them elixir, so they do not eh die*	*eng*	*give them elixir, ~~so~~ they do not eh die*

Note. The parts-of-speech sequence of Speaker b in the second column is artifactual because it creates n-grams with words that were not close to each other before the removal. In particular, we would now observea [N; PRP] 2-gram that does not reflect the original sentence structure. One could remove the spurious n-grams to overcome the issues, but that would generate much shorter sequences of 2-grams (five 2-grams instead of seven) with a nontrivial impact on the quantification of entrainment.

To prepare for semantic entrainment, we relied on word embeddings trained on the Danish version of Wikipedia (Bojanowski et al., 2017). Word embeddings encode the meaning of a word as a vector of values, such that words that appear in similar contexts (co-occurring with similar words) are closer in the vector space (they have similar values) and thus are assumed to have similar meanings. Word embeddings are widely and effectively used for text processing purposes such as automated translation. Identifying word embeddings requires much larger amounts of text than the corpora in our study provide. Although the genre of conversations arguably is not akin to that of online encyclopedias, no sufficiently large-scale corpus of conversational Danish is yet available to create reliable word embeddings representations (Kirkedal et al., 2019; Strømberg-Derczynski et al., 2020). Each word was associated with the 300 values identifying its position in the 300-dimension vector space of the word, and we averaged word embeddings within each utterance.

To calculate entrainment, we first identified all pairs of successive utterances spoken by the two interlocutors and transformed them into numerical vectors for lexical, syntactic, and semantic forms. The lexical vector included all unique lemmas present in at least one of the utterances in the pair. Each lemma constitutes a "dimension" of the vector, and the number of occurrences of that lemma in a given utterance, the value under that dimension. The syntactic vector included all unique n-grams of parts of speech present in at least one of the utterances in the pair. Each unique

Table 1
Example of Transcript Preprocessing

Original	Lemmatized	Parts of speech	2-Grams
A: she wants to play	[“she,” “want,” “to,” “play”]	[(“she,” “PRP”), (“want,” “VB”), (“to,” “IN”), (“play,” “VB”)]	[(“PRP,” “VB”), (“VB,” “IN”), (“IN,” ’VB”)]
B: she gets to play with the bath toys	[“she,” “get,” “to,” “play,” “with,” “the,” “bath,” “toy”]	[(“she,” “PRP”), (“gets,” “VB”), (“to,” “IN”), (“play,” “VB”), (“with,” “PP”), (“the,” “DT”), (“bath,” “NN”), (“toys,” “NN”)]	[(“PRP,” “VB”), (“VB,” “IN”), (“IN,” ’VB”), (“VB,” “IN”), (“IN,” “DT”), (“DT,” “NN”), (“NN,” “NN”)]

Note. Parts of speech reported here are abbreviated: Personal Pronouns (PRP), Verb (VB), Preposition (PP), Determiner (DT), Noun (NN), Infinitive Marker (IN).

n-gram constitutes a “dimension” of the vector, and the number of occurrences of that *n*-gram in a given utterance, the value under that dimension. The semantic vector was the utterance-level word embeddings representation described above. See Table 2 for an example.

Linguistic entrainment was calculated as cosine similarity (that is, the cosine of the angle between two vectors) between successive conversational turns according to the following formula:

$$Similarity(A,B) = \frac{A \cdot B}{||A|| \times ||B||} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}},$$

where A and B represent the vectors respectively for the first and the second interlocutor’s utterance, and *i* indicates the *i*th dimension in the vector. Cosine similarity is a common measure of similarity in natural language processing. Although it is not always preferable to other methods, for instance, the proportion of words or part-of-speech *n*-grams reused (see Xu & Reitter, 2015, for a systematic comparison), cosine similarity has the advantage that it can be consistently applied across lexical, syntactic and semantic entrainment. Two utterances with no elements in common (A: “look!,” B: “where?”) will have cosine scores of 0 for lexical and syntactic entrainment. On the contrary, two utterances with similar lexical but not syntactic choices (A: “pet pet pet,” B: “pet the alien!”) will yield a relatively high cosine of .58 for lexical entrainment, but 0 for syntactic entrainment.

Alternative ways to calculate entrainment are available in the literature. Some focus only on content words or on selected word classes. Others ignore turn-by-turn dynamics to focus on overall similarity over longer stretches of conversation (see Duran et al., 2019, for a review). These different choices are likely to lead to different results. However, we choose to stick to the standardized procedure described and motivated in Duran et al. (2019). Turn-by-turn dynamics of entrainment is a more consistent operationalization of the theoretical construct of linguistic entrainment for our study because it is more akin to the way we analyze backchannels and repairs, and including all words reflects the notion that function words play important roles in linguistic style and style matching (Ireland et al., 2011; Ireland & Henderson, 2014; Tausczik & Pennebaker, 2010).

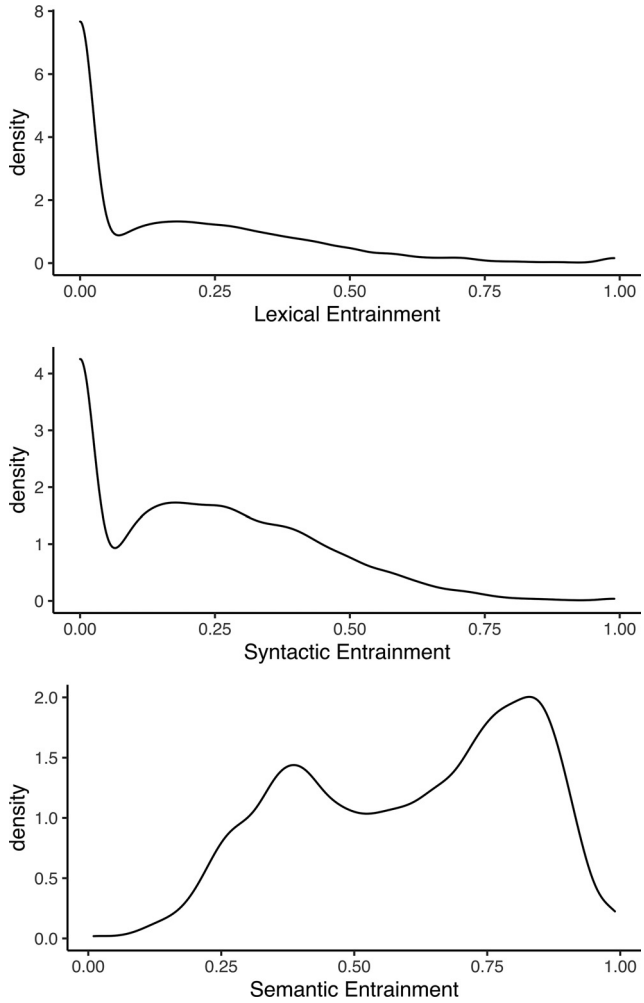
Entrainment scores had a clear peak at 0 (no entrainment) and a second peak within the 0 and 1 boundaries for lexical and syntactic, but not semantic, entrainment (see Figure 3). In other words, a high number of turns do not align with the previous turn. However, when there is any entrainment, the distribution is centered above zero and decreases steadily at both sides.

We chose to use a Bayesian multilevel Zero Inflated Beta regression to model lexical and syntactic entrainment, and a Bayesian multilevel Beta regression to model semantic entrainment. By separating the occurrence of zeros (zero inflation) from the actual level of entrainment once an utterance is entraining, we respect the distributional nature of the data, avoiding model misspecifications (for instance, assuming that a mean entrainment would be an adequate description of the bimodal distribution). Moreover, this distinction makes sense from a conceptual perspective: we want to know both how relevant it is to entrain at all in a given context (entrainment rate, or the inverse of the zero inflation), and how much entrainment there is within an utterance (entrainment level). We report entrainment rate as the negative of the inflation term, indicating the log odds rate of any entrainment instead of the rate of no entrainment. We report entrainment level as the log odds of the cosine similarity for the utterances containing entrainment. We conditioned zero inflation and average level of entrainment on the conversation context and individual and pair level variability. The remaining procedures of model

Table 2
Examples of Lexical, Syntactic, and Semantic Vectors

Original	Lexical vector	Syntactic vector	Semantic vector
	She, Want, To, Play, Get, With, The, Bath, Toy	(“PRP,” “VB”), (“VB,” “IN”), (“IN,” ’VB”), (“VB,” “PP”), (“PP,” “DT”), (“DT,” “NN”), (“NN,” “NN”)	300 dimensions
She wants to play	1, 1, 1, 1, 0, 0, 0, 0, 0	1, 1, 1, 0, 0, 0, 0	0.0,232, 0.02,515, 0.027185, . . .
She gets to play with the bath toys	1, 0, 1, 1, 1, 1, 1, 1, 1	1, 1, 1, 1, 1, 1, 1	0.027013, 0.0117625, −0.01,897 . . .

Figure 3
Zero Inflated Distributions of Linguistic Entrainment



Note. *x* axes indicate probability on a 0–1 scale (cosine similarity, with 0 being no entrainment, and 1 being perfect repetition). *y* axes indicate both the probability of the different rate (height of the density plot).

fitting, model quality checking, model comparison, and hypothesis testing were analogous to the previously described analyses.

It has been debated whether methods to measure entrainment adequately consider the chance level of entrainment, given the baseline frequency of different words and parts of speech as well as the constraints a conversational context gives to those baseline frequencies (Healey et al., 2014). Accordingly, in the online supplemental materials, we report a control analysis involving surrogate pairs (see Figure S54). We built surrogate pairs by artificially interleaving the utterances of speakers from two different pairs within the same condition. We compared entrainment in real and surrogate pairs via plots and statistical models including an interaction between task and type of pairs. We find that all forms of entrainment are credibly higher in real pairs compared with surrogate pairs. Detailed results are reported in the online supplemental materials (see Figure S54).

Networks

To explore how the use of conversational devices covaries across pairs and contexts, we constructed marginal correlation networks. Estimates of individual propensities to use each of the mechanisms were extracted from the models described above and a correlation matrix was estimated. The separate conversational devices were used as nodes, and the correlation scores as the weight of the connection between each pair of nodes. Given the small sample size of the corpus for a network analysis, we adopted a conservative threshold for considering connections to avoid a proliferation of false positives. Only correlations with an absolute coefficient above .5 (explaining at least 25% of the variance in the two variables) were included.

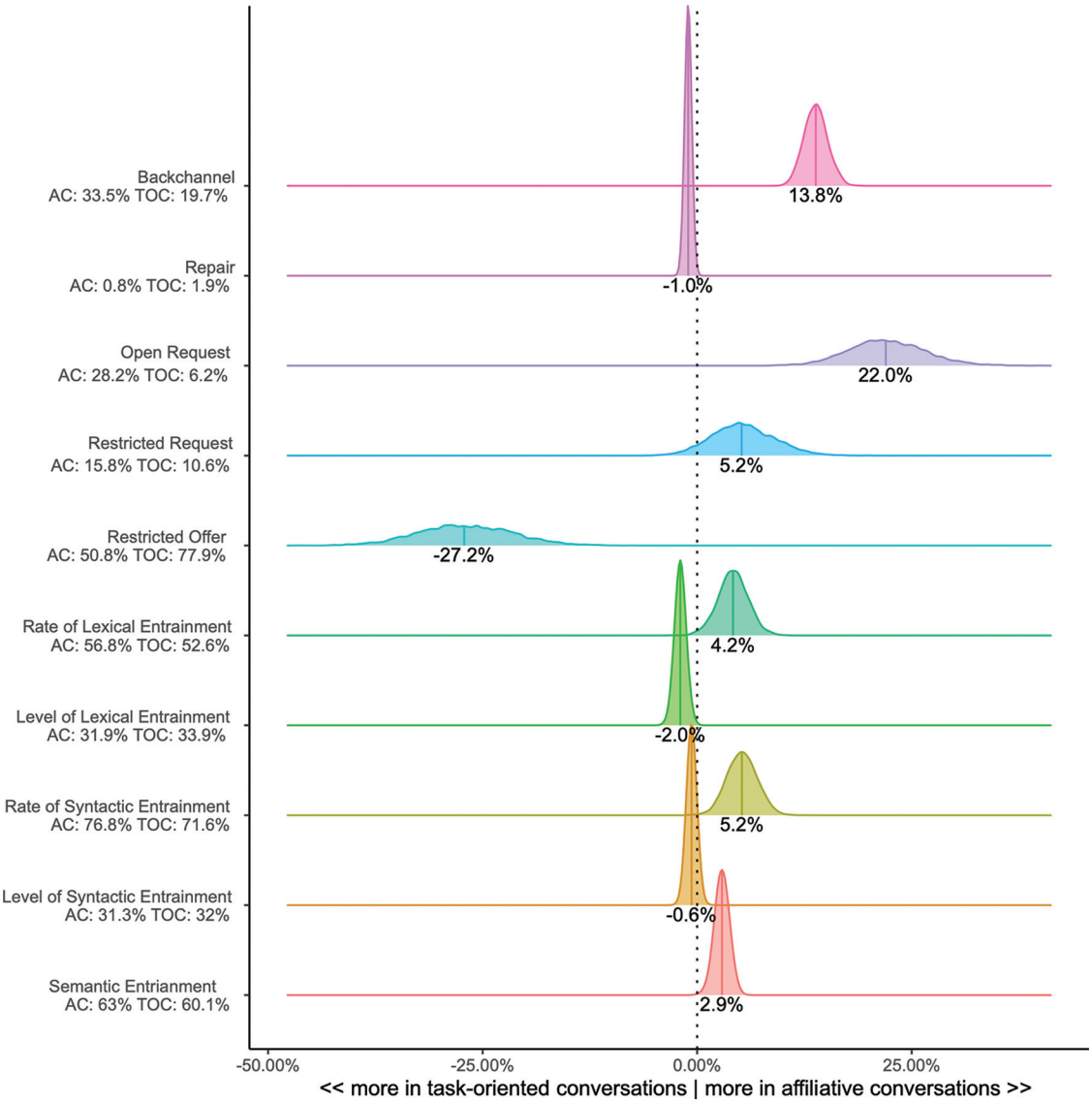
Results

The data revealed some of the patterns expected from previous literature, but also some more surprising ones. Based on previous studies and the exploratory study reported in the online supplementary materials, our main hypothesis (H1) was that backchannels, repairs, and lexical and syntactic entrainment levels would be more frequent in task-oriented conversations than in affiliative conversations, whereas lexical and syntactic entrainment rate and semantic entrainment would be more frequent in affiliative conversations. This was partially supported by the data. As shown in Figure 4 and Table 3, we find that task-oriented conversations show a lower frequency of backchannels (against H1), a higher frequency of repair and in particular specific forms of repair (supporting H1), a lower rate of lexical, syntactic, and semantic entrainment, as well as higher levels of lexical and syntactic entrainment (supporting H1). The results indicate a more complex pattern than expected where each conversational device seems to be modulated differently by the conversational context.

Differences could, however, also be observed within the single conversational contexts (Figure 5 and Table 4). Open requests and lexical entrainment levels slightly increased in affiliative conversations with increased familiarity (the second affiliative conversation > the first affiliative conversation). Furthermore, the asymmetric task-oriented conversation (the Map Task) presented higher frequencies of repair—particularly restricted offers—than the more symmetric Alien Game. Similarly, we found a more frequent use of backchannels in the Map Task, again pointing to conversational devices being modulated by specific demands in a particular conversational context.

In summary, the results do not accommodate a simple explanation of conversational coordination, where, for instance, linguistic entrainment enables mutual understanding across contexts and is driven bottom-up independently of contextual demands. On the contrary, the use of conversational devices is modulated by context within the same individuals. Note that although we expected to see relations between repair and entrainment, as well as between backchannel and repairs, we did not find any strong interdependence between these different conversational devices. However, the network analysis (see Figure 6) displayed positive and negative correlations within the same types of conversational devices, and showed at least two clusters of relations, as well as differences contingent on contextual demands. Note the strong positive correlations between entrainment types, in particular in affiliative

Figure 4
Effects of Task on Conversational Devices



Note. Posterior values for each mechanism are depicted in the left column (affiliative conversations [AC]; task-oriented conversations [TOC]), whereas the plot illustrates the posterior estimates of differences between the two conditions. Note that backchannel and repair are reported relative to the total number of utterances, whereas repair types are reported relative to repair utterances. See the online article for the color version of this figure.

conversations, while repairs display more sparse and negative correlations.

An exploratory analysis was also conducted (see Figure 7) lowering the threshold for relations to .25 (about 6% of the variance in common). There are indications of some trade-offs. These results will be further explored and interpreted in detail in the discussion.

Part I Interim Discussion

Part I set out to compare the use of conversational devices in a large and controlled within-subject corpus, including a diverse set of affiliative conversations (increasing familiarity) and task-oriented conversations (Map Task and Alien Game). We expected conversational patterns to be credibly affected by the context, and in particular that the increased need for precise information sharing in task-oriented conversations would be related to more frequent backchannels, repair, and entrainment compared with affiliative conversations. Further, we expected repairs to be more specific in task-oriented conversations, again to serve the higher need for precise information sharing. This was generally supported by the data. The need for precise information sharing in task-oriented conversations increases—within participants—the use of repair, particularly more specific forms of repair, and lexical and syntactic entrainment levels. Task-oriented conversations also display decreased within-participant lexical and syntactic entrainment

Table 3
Effects of Conversational Devices by Context

Parameter	Affiliative conversations	Task-oriented conversations	Difference (in % points)
Backchannel % of utterances	33.52%, β = −0.68, 95% CI [−0.81, −0.57]	19.69%, β = −1.41, 95% CI [−1.53, −1.28]	13.8%, β = 0.72, 95% CI [0.61, 0.84], ER > 1,000
Repair % of utterances	0.81%, β = −4.81, 95% CI [−5.02, −4.61]	1.85%, β = −3.97, 95% CI [−4.16, −3.78]	1%, β = −0.84, 95% CI [−1.07, −0.61], ER > 1,000
Open requests % of repairs	28.22%, β = −0.93, 95% CI [−1.37, −0.52]	6.18%, β = −2.72, 95% CI [−3.12, −2.35]	22%, β = 1.79, 95% CI [1.29, 2.28], ER > 1,000
Restricted requests % of repairs	15.84%, β = −1.67, 95% CI [−2.15, −1.22]	10.61%, β = −2.13, 95% CI [−2.52, −1.78]	5.2%, β = 0.46, 95% CI [−0.04, 0.96], ER = 14.27
Restricted offer % of repairs	50.76%, β = 0.03, 95% CI [−0.34, 0.4]	77.93%, β = 1.26, 95% CI [0.92, 1.61]	27.2%, β = −1.23, 95% CI [−1.65, −0.81], ER > 1,000
Lexical entrainment rate % of utterances	56.77%, β = 0.27, 95% CI [0.14, 0.41]	52.58%, β = 0.1, 95% CI [−0.01, 0.22]	4.2%, β = −0.17, 95% CI [−0.28, −0.05], ER = 89.91
Lexical entrainment level % of entrained forms in entrained utterances	31.94%, β = −0.76, 95% CI [−0.8, −0.71]	33.91%, β = −0.67, 95% CI [−0.71, −0.62]	2%, β = −0.09, 95% CI [−0.13, −0.04], ER = 1,999
Syntactic entrainment rate % of utterances	76.83%, β = 1.2, 95% CI [1.02, 1.37]	71.58%, β = 0.92, 95% CI [0.77, 1.08]	5.2%, β = −0.28, 95% CI [−0.43, −0.13], ER = 399
Syntactic entrainment level % of entrained forms in entrained utterances	31.32%, β = −0.79, 95% CI [−0.83, −0.74]	31.97%, β = −0.76, 95% CI [−0.78, −0.73]	0.6%, β = −0.03, 95% CI [−0.07, 0.01], ER = 9.13
Semantic entrainment % of entrained forms in entrained utterances	62.98%, β = 0.53, 95% CI [0.46, 0.6]	60.09%, β = 0.41, 95% CI [0.35, 0.46]	2.9%, β = 0.12, 95% CI [0.06, 0.18], ER = 999

Note. Percentages indicate the proportion of utterances identified as either backchannel, repair or entrainment (for entrainment rate). Repair types (open request, restricted request, restricted offers) are percentages within the subset of repair utterances only. For entrainment level we report the percentages of linguistic structures that are repeated from the previous utterance. β estimates and compatibility intervals are reported in the original scale of the model (log odds).

rates and semantic entrainment. However, contrary to our hypothesis, we observe an overall higher frequency of backchannels in affiliative conversations. Last, we see consistent patterns of covariation modulated by context. Interestingly, we do not observe conversation-level covariations across, for instance, repairs and entrainment, or backchannels and repairs, but only within-mechanism covariations (for instance, between different forms of entrainment).

Session Effects

These results generally hold across sessions, but with some important variations, especially in how the use of conversational devices vary between the two tasks in the task-oriented conversations. In particular, we see that effects of repair, restricted offers, and syntactic entrainment level (where the mean frequency is higher in task-oriented conversations than in affiliative conversations) are primarily driven by a high frequency of use in the Map Task, whereas the Alien Game results are closer to affiliative conversation levels. Similarly, the effects on lexical entrainment level (where task-oriented conversations are higher than affiliative conversations) are driven by the Alien Game, whereas restricted requests are lower in task-oriented conversations than in affiliative conversations, but only because of low levels in the Map Task.

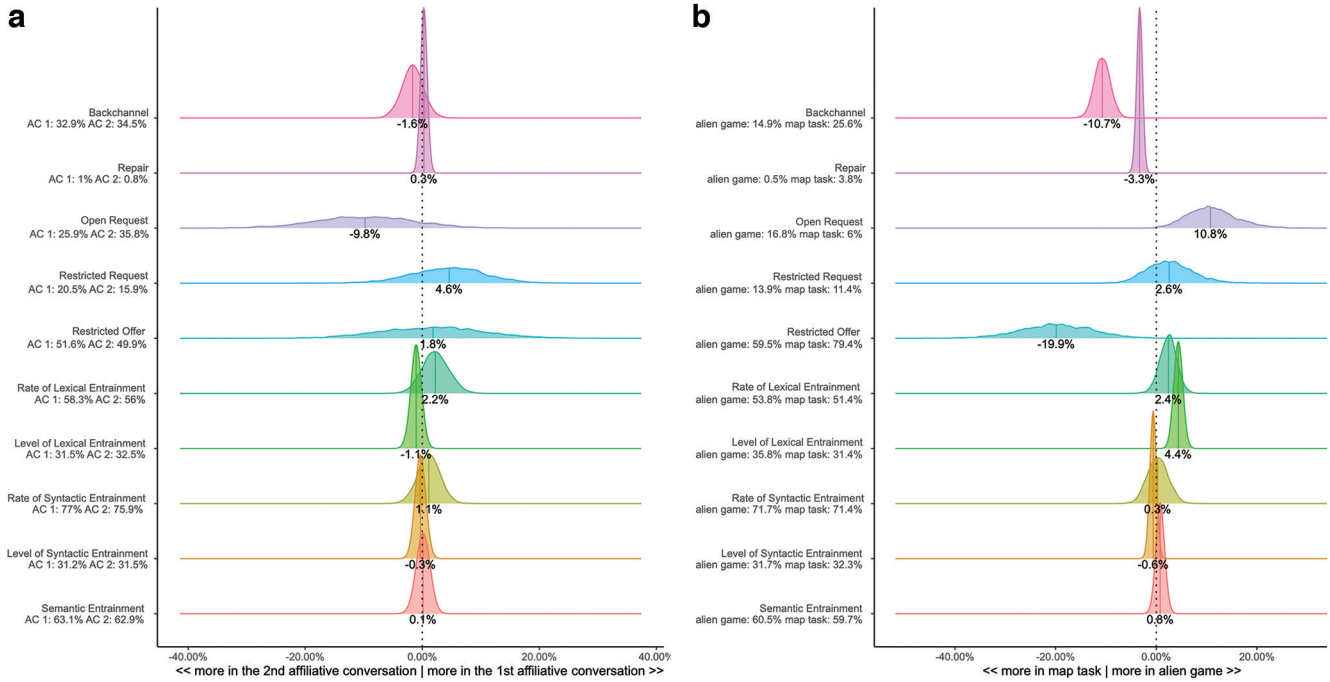
This suggests that, beyond general differences between affiliative and task-oriented conversations, these mechanisms are modulated by the specific task demands. For instance, the two tasks differ in the amount of information that interlocutors share, which is likely to modulate the requirements for mutual understanding. In the Alien Game, the participants have perceptual access to the same information, which likely decreased their need for restricted offers, the most specific and precise repair type. In the Map Task, on the contrary, there is an asymmetry in the information that the participants have access to (the director sees a path which is not visible to the matcher). Therefore, participants have to be more specific about potential problems and give more detailed information to their partner to facilitate mutual understanding and increase their chance of success in the task. In this sense, the differences between the two tasks further support our hypothesis that the use of conversational devices are specifically affected by the contextual need for precise information sharing. Further, the results also seem to support the model of conversation as interpersonal synergy, with its emphasis on contextually sensitive coordination. Pure bottom-up models of linguistic coordination as reciprocal priming would be challenged by systematic within-interlocutors changes in the use of conversational demands across tasks. By adapting the rate and level of entrainment to the conversational context, interlocutors vary the conversational devices they use to build mutual understanding while avoiding redundancy of contributions in contexts that afford a high degree of precise information sharing. At the same time, entrainment is widespread and does seem to play a role, which would also exclude pure top-down models of coordination.

The observed heterogeneity in the use of conversational devices between the two tasks calls for the exploration of a wider variety of tasks, carefully designed to systematically vary contextual demands. Moreover, the frequency of individual conversational devices is only an indirect cue to their actual function. To systematically address the more functional aspect of mutual understanding, we need to assess whether pair variability in the use of conversational devices is related to variability in task performance; that is, whether for instance, a higher frequency of repair in the Alien Game also implies a higher performance in the task.

Summing Up

Our findings support the hypothesis that the use of conversational devices is indeed modulated by contextual demands. When tasks demand high precision information sharing, we see fewer

Figure 5
Effects of Session on Conversational Devices



Note. Panel a (right) shows posterior values of the difference between the two affiliative conversations, while panel b (left) shows posterior values of the difference between the two task-oriented conversations. Posterior values for each mechanism are depicted in the left column (affiliative conversations [AC]; task-oriented conversations [TOC]), whereas the plot illustrates the posterior estimates of differences between the two conditions. Note that backchannel and repair are reported relative to the total number of utterances, whereas repair types are reported relative to repair utterances. See the online article for the color version of this figure.

backchannels, more selective entrainment (lower rates, but higher levels), and more precisely targeted interactive repair (restricted offers). By being more precise, interlocutors might better match the contextual demands of task-oriented conversations and this precision might assist them in performing well on a task. These observations also raise the question of whether the differences are simply a reflection of the context or, crucially, whether they are functionally adaptive. In other words, does the increase in, for example, levels of entrainment in task-oriented conversations imply that pairs with higher levels also achieve higher task performance, because levels of entrainment enable more effective and precise coordination? The questions also apply across tasks: If the Map task displays higher use of backchannels than the Alien Game, do we also observe a more positive relation between backchannels and performance in the former than in the latter? We explore these and related questions in Part 2.

Part 2: Task Performance

The literature reports several attempts at more directly relating specific conversational devices to collaborative performance. Backchannels have mostly been related to social or affiliative functions (Bavelas et al., 2000; Cutrone, 2005, 2011; Gardner, 2001), consistent with the observations in Part I reporting a higher frequency in affiliative conversations. However, a few studies also suggest that backchannels may play a role in structuring and sharing information: interlocutors differentially use backchannels to navigate joint projects, and the use of backchannels facilitate both the precision of information produced and its comprehension (Bangerter & Clark, 2003; Bangerter et al., 2004; Kuhlen & Brennan, 2010; Tolins et al., 2017; Tolins & Fox Tree, 2016; Tolins & Tree, 2014). Similarly, we see a higher use of backchannels in the Map Task (compared with the Alien Game), which indicates that they might serve to compensate for the asymmetries in access to information of the director and matcher.

Conversational repair has often been highlighted on conceptual grounds as a tool to fix miscommunication and restore mutual understanding. One study manipulated the frequency of repairs (by artificially adding repair requests in chat conversations), which was found to facilitate the formation of more abstract representations and to aid pairs' performance (Mills, 2014). In line with this, we see more specific restricted offers, but fewer generic repairs (open and restricted requests) in task-oriented conversations, in particular in the Map Task. However, no study to our knowledge has directly and quantitatively assessed the relationship between performance in task-oriented conversations and backchannels or repairs.

As mentioned before, the literature is richer when assessing the relation between entrainment and performance. There are conceptual arguments suggesting that entrainment should facilitate performance (Pickering & Garrod, 2004), and indeed some studies positively relate different forms of lexical, syntactic and semantic

Table 4
Effects of Conversational Devices by Session

Parameter	The first affiliative conversation	The second affiliative conversation	Difference (in % points) between the first and the second affiliative conversation	Alien game	Map task	Difference (in % points) between the alien game and the map task
Backchannel % of utterances	32.87%, β = −0.71, 95% CI [−0.85, −0.58]	34.52%, β = −0.64, 95% CI [−0.77, −0.5]	1.7%, β = −0.07, 95% CI [−0.21, 0.06], ER = 4.93	14.93%, β = −1.74, 95% CI [−1.91, −1.58]	25.65%, β = −1.06, 95% CI [−1.21, −0.92]	10.7%, β = −0.68, 95% CI [−0.83, −0.51], ER > 1,000
Repair % of utterances	1.03%, β = −4.57, 95% CI [−4.81, −4.34]	0.75%, β = −4.88, 95% CI [−5.15, −4.6]	0.3%, β = 0.31, 95% CI [0.02, 0.61], ER = 25.32	0.49%, β = −5.32, 95% CI [−5.64, −5.01]	3.79%, β = −3.23, 95% CI [−3.45, −3.01]	3.3%, β = −2.09, 95% CI [−2.39, −1.77], ER > 1,000
Open requests % of repairs	25.87%, β = −1.05, 95% CI [−1.55, −0.58]	35.76%, β = −0.59, 95% CI [−1.16, −0.02]	9.9%, β = −0.47, 95% CI [−1.12, 0.18], ER = 7.39	16.77%, β = −1.6, 95% CI [−2.29, −0.96]	6.04%, β = −2.74, 95% CI [−3.17, −2.35]	10.7%, β = 1.14, 95% CI [0.48, 1.78], ER = 665.67
Restricted request % of repair	20.46%, β = −1.36, 95% CI [−1.89, −0.85]	15.85%, β = −1.67, 95% CI [−2.35, −1.01]	4.6%, β = 0.31, 95% CI [−0.4, 1.01], ER = 3.19	13.86%, β = −1.83, 95% CI [−2.55, −1.17]	11.43%, β = −2.05, 95% CI [−2.46, −1.68]	2.4%, β = 0.22, 95% CI [−0.43, 0.86], ER = 2.52
Restricted offer % of repair	51.59%, β = 0.06, 95% CI [−0.38, 0.5]	49.89%, β = 0, 95% CI [−0.56, 0.54]	1.7%, β = 0.07, 95% CI [−0.53, 0.66], ER = 1.38	59.54%, β = 0.39, 95% CI [−0.16, 0.94]	79.42%, β = 1.35, 95% CI [1, 1.71]	19.9%, β = −0.96, 95% CI [−1.5, −0.42], ER = 570.43
Lexical entrainment rate % of utterances	58.26%, β = 0.33, 95% CI [0.18, 0.49]	55.99%, β = 0.24, 95% CI [0.09, 0.39]	2.3%, β = −0.09, 95% CI [−0.25, 0.05], ER = 5.48	53.82%, β = 0.15, 95% CI [0.05, 0.25]	51.45%, β = 0.06, 95% CI [−0.07, 0.19]	2.4%, β = −0.09, 95% CI [−0.2, 0.02], ER = 12.38
Lexical entrainment level % of entrained forms in entrained utterances	31.49%, β = −0.78, 95% CI [−0.82, −0.73]	32.53%, β = −0.73, 95% CI [−0.79, −0.67]	1%, β = −0.05, 95% CI [−0.11, 0.01], ER = 11.08	35.82%, β = −0.58, 95% CI [−0.63, −0.54]	31.45%, β = −0.78, 95% CI [−0.84, −0.72]	4.4%, β = 0.2, 95% CI [0.14, 0.26], ER > 1,000
Syntactic entrainment rate % of utterances	77.03%, β = 1.21, 95% CI [1.04, 1.39]	75.9%, β = 1.15, 95% CI [0.96, 1.33]	1.1%, β = −0.06, 95% CI [−0.23, 0.11], ER = 2.71	71.65%, β = 0.93, 95% CI [0.77, 1.08]	71.36%, β = 0.91, 95% CI [0.73, 1.09]	0.3%, β = −0.01, 95% CI [−0.18, 0.15], ER = 1.23
Syntactic entrainment level % of entrained forms in entrained utterances	31.18%, β = −0.79, 95% CI [−0.84, −0.74]	31.49%, β = −0.78, 95% CI [−0.83, −0.73]	0.3%, β = −0.01, 95% CI [−0.07, 0.04], ER = 2.1	31.69%, β = −0.77, 95% CI [−0.8, −0.74]	32.3%, β = −0.74, 95% CI [−0.78, −0.7]	0.6%, β = −0.03, 95% CI [−0.07, 0.01], ER = 8.26
Semantic entrainment % of entrained forms in entrained utterances	63.08%, β = 0.54, 95% CI [0.45, 0.62]	62.94%, β = 0.53, 95% CI [0.45, 0.61]	0.1%, β = 0.01, 95% CI [−0.07, 0.08], ER = 1.22	60.48%, β = 0.43, 95% CI [0.37, 0.48]	59.68%, β = 0.39, 95% CI [0.33, 0.46]	0.8%, β = 0.03, 95% CI [−0.02, 0.09], ER = 5.92

Note. Percentages indicate the proportion of utterances identified as either backchannel, repair or entrainment (for entrainment rate). Repair types (open request, restricted request, restricted offers) are percentages within the subset of repair utterances only. For entrainment level we report the percentages of linguistic structures that are repeated from the previous utterance. β estimates and compatibility intervals are reported in the original scale of the model (log odds).

Figure 6

Correlational Network of Each of the Sessions, With a Threshold of 0.5 (Only Correlations With an Absolute Value > 0.5 Are Visualized)

Note. Blue lines (darker grey) represent positive correlations, whereas red lines (lighter grey) represent negative. Line thickness is proportional to the absolute effect size of the relation. See the online article for the color version of this figure.

entrainment with task performance (Fusaroli et al., 2012; Gries, 2005; Ireland & Henderson, 2014; Reitter & Moore, 2014; Slocombe et al., 2013). However, alternative models of conversation (for instance, interpersonal synergy, Fusaroli, Rączaszek-Leonardi, et al., 2014) argue that simply repeating each other is not a productive form of cooperation and emphasize the need for complementary contributions (i.e., with low entrainment). These perspectives have also found empirical support in negative relations between entrainment and performance (Fusaroli & Tylén, 2016; Tylén et al., 2020; Xu & Reitter, 2017). In Part 1, we observe a first tentative synthesis of the two perspectives with both entrainment and complementarity playing a role: entrainment levels are higher in task-oriented conversations, but rates and semantic entrainment are lower.

With the above findings in mind, we investigate the frequency of conversational devices, and explore the function of these mechanisms with regard to task performance. This relation has not yet been examined across different tasks in a within-subjects experimental design. Specifically, we hypothesize that the distribution of conversational devices will be predictive of task performance (H2). Given the widespread inconsistencies in the prior literature, we base our hypothesis on Part 1: if a mechanism was more frequently used in task-related conversations, we expect it to be positively related to performance. Additionally, we also expect that variations in frequency between tasks should lead to variations in performance. For instance, given that backchannels are more frequent in the Map task than in the Alien Game, we expect a more positive relation between backchannels and performance in the Map task than in the Alien Game (see Table 4 for specific hypotheses).

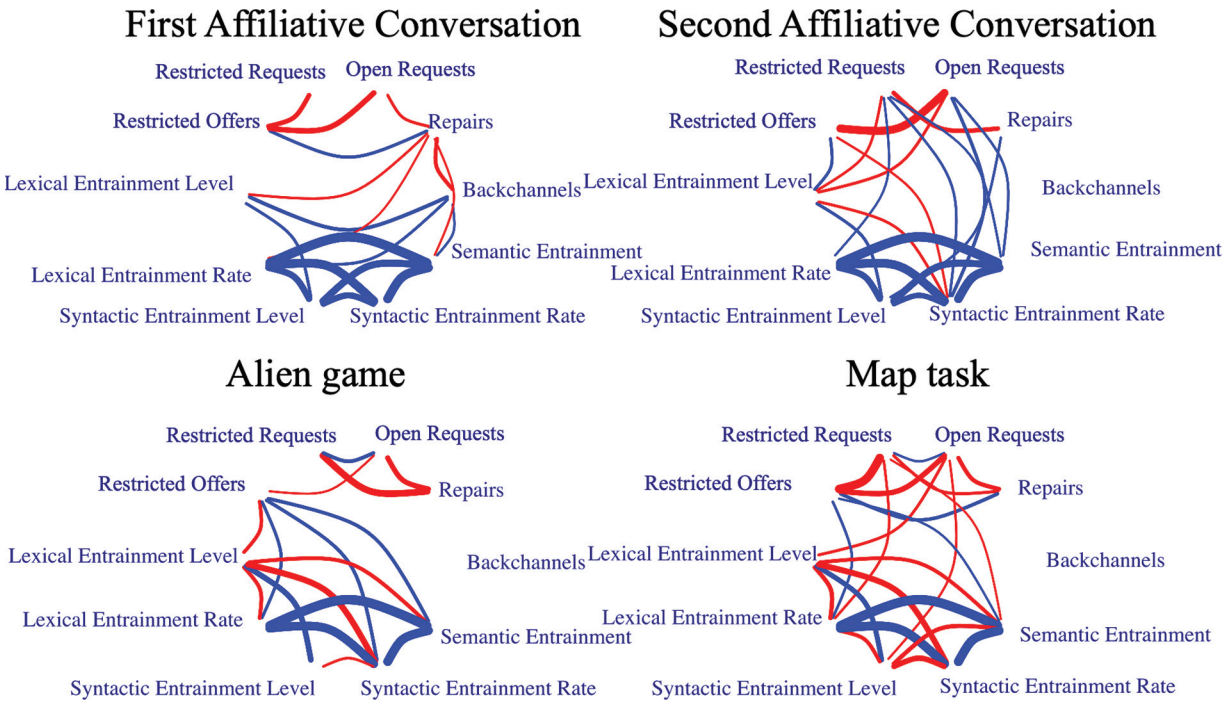
It should be noted that these analyses are observational. However, we do provide theoretical reasons and previous findings to motivate our hypotheses. The findings thus constrain our theoretical framework and provide a more robust and potentially generalizable basis for future direct manipulations or other forms of causal inference (see Discussion).

Method

Part II uses the data from Part 1. Individual differences in prevalence of conversational devices (within session) were extracted from the models reported in Part 1 (adding up variation by participant and variation by pair). From the Alien Game, we collected the cumulative score achieved by each pair across the whole session. For the Map Task, we measured the area between the route of the directors' map and the route drawn by the matcher, using the package "pathmapping" for RStudio (Mueller et al., 2016). The smaller the area, the more similar the routes are and the higher the performance. Map performance was calculated from the second last map that the participants managed to finish, or from the only map if

Figure 7

Correlational Network of Each of the Sessions, With a Threshold of 0.25 (Only Correlations With an Absolute Value > 0.25 Are Visualized)



Note. Blue lines (darker grey) represent positive correlations, whereas red lines (lighter grey) represent negative. Line thickness is proportional to the absolute effect size of the relation. See the online article for the color version of this figure.

only one was completed. This was done to avoid performance penalties for maps that were unfinished due to the time constraint, and to measure performance at the stage where participant pairs had had most time to establish mutual understanding. Some pairs only managed to finish the first map, so their final score is based solely on one map; for others it is based on their fourth map (average number of completed maps: 1.6). We invert the deviance score (how different the routes are) to yield a more intuitively interpretable performance score: lower deviance is higher performance. Note that other performance measures would be possible: for instance, the number of completed maps. Because the instructions emphasize the need for precise reconstruction of the path, the deviation score is the most direct measure of that. Pairs with a high number of completed maps might have chosen to prioritize speed over accuracy against instructions.

To test the hypothesis that the use of conversational devices, for instance the frequency of backchannels, is systematically related to task performance, we first calculated a performance score per each pair across the two tasks. Frequency of use of each conversational device for each pair was extracted as the estimate from the models in Part 1 (rather than a raw count). We then built two Bayesian linear regression models with performance as an outcome predicted by one conversational device at a time (for instance, backchannel, repair, or lexical entrainment rate). Performance and predictor scores were transformed on a *z*-score scale—which made the regression coefficients comparable to a Pearson correlation score—to facilitate comparison between tasks and model convergence. Note that the distributions maintain the same shape and are not otherwise stretched.

Results

We hypothesized that if a mechanism was more frequently used in task-related conversations, we would see it positively related to performance, and if less frequent, negatively related. In support for this hypothesis, we find that lower rates, and higher levels of lexical and syntactic entrainment, as well as lower semantic entrainment were related to higher performance across the task-oriented conversations, with moderate absolute mean effect sizes (mostly between .16 and .4). In other words, pairs that entrained less often, except for occasional lengthy repetitions, also performed better in the two different tasks.

Contrary to our hypotheses, some devices that were generally more frequent in task-oriented conversations presented positive relations with performance in only one task (higher frequency of backchannels and restricted offers in the Map Task, lower frequency of restricted requests in the Alien Game); and others showed potentially opposite patterns. In the Map Task, backchannels were positively related to performance, and repairs were mostly unrelated or negatively related to performance. For the full details see Table 5 and Table 6, as well as the discussion section for interpretation of the findings.

Table 5
Effects of Conversational Devices on Performance

Device	Alien game		Map task	
	Estimate [95% compatibility intervals]	ER	Estimate [95% compatibility intervals]	ER
Backchannel	0.05 [−0.13, 0.23]	2.16	0.25 [0.07, 0.42]	70.43
Repair	−0.19 [−0.36, −0.01]	27.99	−0.06 [−0.24, 0.12]	2.58
Open request	0.07 [−0.11, 0.24]	2.55	−0.13 [−0.3, 0.05]	8.22
Restricted request	−0.07 [−0.25, 0.1]	2.85	−0.08 [−0.26, 0.11]	3.02
Restricted offer	−0.15 [−0.33, 0.03]	11.74	0.16 [−0.02, 0.34]	11.66
Lexical entrainment rate	−0.32 [−0.48, −0.17]	>1,000	−0.25 [−0.43, −0.08]	124
Lexical entrainment level	0.24 [0.07, 0.42]	75.92	0.17 [−0.01, 0.35]	17.18
Syntactic entrainment rate	−0.32 [−0.48, −0.15]	>1,000	−0.27 [−0.44, −0.09]	116.65
Syntactic entrainment level	0.08 [−0.09, 0.26]	3.65	0.16 [0, 0.33]	19
Semantic entrainment	−0.39 [−0.56, −0.22]	>1,000	−0.31 [−0.49, −0.15]	>1,000

Note. Estimates correspond to how much the increase of 1 standard deviation in the predictor changes performance (on a standard deviation scale). They are comparable across tasks, and similar to Pearson correlation scores.

A model including all predictors showed that not all predictors maintained evidence of relation to performance.[3] This indicates that not all conversational devices contribute unique information to our understanding of effective conversational coordination. However, backchannels, repair, open request, and semantic entrainment retain moderate to strong correlations to performance. Full details are presented in the online supplemental materials (Table S67).

Part II Interim Discussion

Whereas Part I highlighted the role of contextual demands in relation to the differential use of conversational devices, Part 2 more directly investigates whether these differences are related to performance in the tasks. We expected that pairs with, for instance, a high frequency of backchannels in the Map task would also show higher performance scores, thus indicating a functional dimension of the conversational strategy. Table 6 shows that, although we generally find widespread associations between conversational devices and performance, our directional hypotheses are strongly supported only for entrainment. Lexical and syntactic entrainment rates, as well as semantic entrainment, are less frequent in task-oriented conversations and indeed negatively correlated with performance, whereas lexical and syntactic levels are higher in task-oriented conversations and mostly positively related to performance. This further supports our argument that complementarity (contributing different content and perspectives) and entrainment (using the same material and building on each other's contributions) both have a place in successful task-oriented interactions (see additional discussion in the General Discussion).

Whereas entrainment shows relatively consistent patterns across the two tasks, repairs and backchannels appear to play different roles, thus highlighting more nuanced task-dependent contextual demands. For instance, although in Part I, affiliative conversations displayed higher frequencies of backchannels compared with task-oriented conversations, this difference was primarily driven by a low frequency in the Alien Game, while backchannels are rather frequent in the Map Task. Accordingly, Part 2 found a positive correlation between the frequency of backchannels and performance only in the Map Task. Analogously, open requests are less frequent in the Map task and negatively correlated with performance. The same pattern is evident for restricted offers, which are more frequent in the Map task and positively related to performance only in that task. This indicates a pattern in which pairs flexibly modulate their conversational structure in a nuanced way according to the local contextual demands, to achieve a successful outcome. These findings indicate that for the Map Task, an increase in performance is related to matchers providing more positive evidence of understanding (frequent backchannels) and maximally cooperative requests for clarification (frequent restricted offers). The picture is not as clear-cut for the Alien Game as more specific repair formats, like restricted offers, are negatively related to performance, and backchannels do not present a credible relation (see Table 5). However, this pattern is not consistent for overall repair and restricted requests, both of which are negatively correlated with performance in the task in which they are more frequent.

In summary, we find a general trend where conversational devices which are more frequent in a conversational context (task-oriented conversations) are also consistently positively related to performance scores (and vice versa, if less frequent they are also negatively related to performance). In particular, we find that performance is consistently tied to entrainment patterns, whereas the link to backchannels and general repair is less clear. However, the relation between high frequency in Part I and performance in Part II is credible for restricted offers in the Map Task. The differences between the two tasks indicate that variability in contextual demands within task-oriented conversations might be more pronounced than assumed in previous studies. Future research needs to further disentangle the causes of this variability. For instance, a productive approach could be to manipulate whether there are shared visible referents, as in the current Alien Game, and whether these might disappear before the conversation starts, or only allow one interlocutor to have access to them.

General Discussion

We set out to systematically investigate how conversational devices, such as backchannels, repairs, and linguistic entrainment,

[3] For the Alien Game, lexical entrainment level did not provide credible evidence of relation to performance once other predictors were included, and in the Map Task, this was the case for backchannels, lexical entrainment rate and syntactic entrainment level.

Table 6
Indications of Whether the Findings of Part 1 (Which Conversation Type Shows Higher Frequencies of Conversational Devices in Task-Oriented Conversations) Are Further Supported by Those of Part 2 (How Are Conversational Devices Related to Performance)

Measure	Affiliative conversations versus task-oriented conversations contrast-based hypothesis: Relation to performance across task:	Supported	Map task vs. alien game contrast-based hypothesis: More positive relation to performance in:	Supported
Backchannels	Negative	No	Map Task	Yes
Repair	Positive	No	Map Task	No
Open request	Negative	Partially (Map)	Alien Game	No
Restricted request	Negative	Partially (Map)	Alien Game	No
Restricted offer	Positive	Partially (Map)	Map Task	Yes
Lexical entrainment rate	Negative	Yes	Alien Game	No
Lexical entrainment level	Positive	Yes	No credible difference	Yes
Syntactic entrainment rate	Negative	Yes	No credible difference	Yes
Syntactic entrainment level	Positive	Yes	Map Task	Yes
Semantic entrainment	Negative	Yes	No credible difference	Yes

relate to diverse contextual demands and the extent to which that relates interpersonal coordination and performance. Part I found largely consistent patterns with general repairs, restricted offers, and lexical and syntactic entrainment level being more frequent in task-oriented conversations, contingent on affordances for higher precision when establishing and maintaining mutual understanding in these contexts compared with affiliative conversations. Notably, these findings are consistent (with the exception of backchannels) with the exploratory between-subjects study described in the online supplemental materials. The consistency across studies is a nontrivial cue in assessing the generalizability of the results. Finding analogous patterns even when corpora diverge across several dimensions, in particular one being more controlled, and the other more naturalistic, suggests that the patterns are robust across a wide range of variations.

Different tasks, varying in contextual demands, modulated the frequency of the conversational devices used, with restricted offers and backchannels being more frequent in the asymmetric director-matcher context, the Map Task, than when making joint decisions in the Alien Game. These findings provide the foundation for a more systematic framework for understanding conversational devices in the interplay between their informativeness, cost and contextual demands. Crucially the way in which mechanisms are adjusted to contextual demands also affects performance in the task. Part II found that entrainment in particular was more frequent and associated with higher performance in task-oriented conversations. However, the picture seemed more complex for backchannels and repair.

Backchannels

Backchannels were generally very frequent in the data: they occurred on average every 8.4 s—in every third utterance in affiliative conversations, and every 10.81 s—or every fifth utterance—in task-oriented conversations.[4] The findings are compatible with the suggestion in the literature that backchannels provide tools for face management and affiliation (Bavelas et al., 2000; Brown & Levinson, 1978; Cutrone, 2005). However, backchannels can also help maintain and develop mutual understanding: they are more frequent in the Map task than in the Alien Game and indeed are positively related to performance in the former but not the latter. Although these findings should be replicated and further elaborated, we suggest that backchannels might be important for developing mutual understanding in those situations where there is a clear perceptual asymmetry in the conversation. In the Map Task, a director has exclusive access to the relevant information (the path to retrace) and accordingly does most of the talking. The matcher does not have much information to share but can provide continuous low-cost feedback (backchannels) steering the flow of information, reassuring the director of continuous attention, and when needed, signaling potential gaps or misunderstandings (repairs). In the same way, affiliative conversations often do not offer a joint perceptual display, indicating that backchannels might be used as a verbal device to manage shared attention. This role of backchannels is supported by previous work on storytelling both in naturalistic and experimental contexts highlighting the importance of backchannels in shaping the interlocutor's production. For instance, these studies find that in presence of attentive listeners providing backchannels, speakers produce more extensive and detailed utterances (Bavelas et al., 2000; Kuhlen & Brennan, 2010; Roger & Nesshoever, 1987; Sacks, 1992). One could also speculate that conversations composed by longer turns (as in the Map task where a director provides instructions) would provide more opportunity for backchannels compared with interactions characterized by shorter and more dynamic turns (as in the Alien Game where there is more equal information sharing).[5] This is an interesting target for further studies.

[4] Calculations are based on Part 1 as the more controlled dataset. In the Preliminary Exploratory Study (see Section 1 in the online supplemental materials), we find analogous overall frequencies, but with task-oriented conversations having more frequent backchannels than affiliative conversations.

[5] In the current study the Map Task has a longer mean turn length and fewer speaker changes than the Alien Game suggesting that this was the case.

Repairs

Cross-linguistic work has found that repair occurs roughly once per 1.4 min in affiliative conversations across 12 languages (not including Danish; Dingemanse et al., 2015). In the current study, we observe a lower range of frequencies. In affiliative conversations, .81% of all utterances are repairs, which is equivalent to once per 5 min and 53 s (or .17 repairs per minute) in a conversation. However, we should note that this study involved more controlled and homogeneous affiliative conversations than the cross-linguistic study, which included dyadic as well as multiparty interactions in a wide range of indoor and outdoor settings, often with concurrent activities and various background noises and distractions.[6]

In the task-oriented conversations in our study, 1.85% of all utterances were repairs, which is equivalent to .52 repairs per minute (once per 1 min and 55 s). We suggest that the higher frequency of repair in task-oriented conversations is associated with the increased need to maintain and/or reestablish mutual understanding in these kinds of contexts. Importantly, task demands also make a difference in the use of repairs: We see that repairs are slightly more frequent in the Map task than in the Alien Game. This difference resonates with the backchannel results, suggesting that it might be driven by the asymmetry of knowledge between the participants: The matcher must consistently update their mental representation of the route, which increases the need of repairing.

Earlier studies on affiliative conversations suggest that the more specific (and thereby less general) the repair, the more preferred it would be (Dingemanse et al., 2015), supporting the principle of least collaborative effort (Clark & Schaefer, 1987; Clark & Wilkes-Gibbs, 1986). Listeners seem to prefer to provide as much information as possible to the speaker when requesting a repair, to minimize the effort of the speaker when reestablishing mutual understanding. This is replicated in our study with restricted offers being the most frequent repair type across all contexts and even more so in task-oriented conversations, where 65.68% of all repairs consist in restricted offers—the more specific repair type (although it is only 45.85% in affiliative conversations). This supports the idea that repairs are used to establish and maintain mutual understanding, especially when accurate information sharing is needed, and that the specificity of the repair format follows the need for precision. Indeed, both experimental and simulation work suggest a positive role for repairs in developing and maintaining effective communication conventions (Arkel et al., 2020; Mills, 2014).

The investigation of task performance in Part II reveals a nuanced relation between repair and communicative success. The overall frequency of repairs is negatively related to performance in both tasks, something that is not unexpected given that repair is a sign of possible trouble. In the Alien Game, specific repair types are likewise negatively related to performance. In the Map Task, however, we find the opposite pattern, with restricted offers being the only positive predictor of performance, while open requests are a negative predictor. As for backchannels, the difference between the tasks could be explained by the absence of a joint reference point in the Map Task, which might increase the need for a continuous confirmation of mutual understanding, whereby reference specific mechanisms such as restricted offers are used more frequently to update the mutual understanding during the task. Still the negative relations are not easily explained. We speculate (following Mills, 2014) that the temporal distribution of the repairs over the task should be further investigated: Early repairs may help establish mutual understanding, whereas repairs occurring in later phases of the conversation might signal enduring problems in reaching mutual understanding.

Linguistic Entrainment

Linguistic entrainment has been conceptualized within different frameworks, from which contrasting predictions have been advanced and supported by empirical studies. Whereas the interactive alignment framework suggests a positive relation to task-oriented conversations and performance, the interpersonal synergy framework is also compatible with a negative one. Distinguishing between different forms of entrainment enables us to get a more nuanced picture of entrainment and may contribute to solving apparent inconsistencies. We found linguistic entrainment to be pervasive, with more than half of all utterances containing some lexical repetitions (around 45% in surrogate pairs) and more than 70% containing syntactic repetitions (around 50% in surrogate pairs). These results hold even when controlling for surrogate pairs (see Figure S54) and are consistent with a priming mechanism. However, entrainment rate and levels also vary consistently with contextual demands, which is indicative of context-sensitive top-down modulation of entrainment (for a similar take, see also Wang & Hamilton, 2012), compatible with a more explicit strategy (e.g., I need to decrease/increase entrainment), or the consequence of higher order goals (e.g., I need to get those landmarks right). Lexical and syntactic rates, as well as semantic entrainment, were lower in task-oriented conversations and accordingly negatively related to performance in both tasks. Lexical and syntactic levels were higher in task-oriented conversations and consistently positively related to performance in both tasks (except for syntactic levels in the Alien Game). This indicates that interlocutors repeat each other less often in task-oriented conversations than in affiliative conversations, but when they do so they repeat longer sequences, possibly to increase precise information sharing. This behavior seems adaptive as pairs following this pattern also perform better in the tasks. Differences in task-specific demands within task-oriented conversations seem to play a minor role for linguistic entrainment, with most frequencies being equal across tasks and in relation to performance.

Our findings provide support for a synergistic model of conversation where both entrainment and complementarity (of content and perspective) play a role (Dale et al., 2013; Fusaroli, Rączaszek-Leonardi, et al., 2014; Fusaroli & Tylén, 2016). By entraining their language, interlocutors might facilitate the establishment of mutual understanding, especially when the entrained utterances concern central aspects of the task (Fusaroli et al., 2012; Hawkins et al., 2020). At the same time, generic entrainment is less pervasive in contexts involving the need for more precise information sharing, where interlocutors rather provide complementary contributions to the conversations. This also suggests that entrainment might play different roles in different phases of the interaction, as we discuss further in the section on limitations.

[6] The Preliminary Exploratory Study (see Section 1 in the online supplemental materials)—involving more varied affiliative conversations, including dinners and other activities—showed a frequency of repairs comparable to the cross-linguistic study (1 repair every 1 min and 38 s).

Networks

Whereas previous studies predominantly investigated one conversational device at a time, here we set out to explore their interdependencies across conversational contexts. The previous literature motivated the expectation that linguistic entrainment form a clear cluster, with each form of entrainment catalyzing other forms of entrainment (Pickering & Garrod, 2004); trade-offs between backchannel and repairs (Schegloff, 1982), as well as between entrainment, repair, and backchannel (Fusaroli et al., 2017; Fusaroli, Rączaszek-Leonardi, et al., 2014). We observed two main clusters of relations and clear variations between conversational contexts.

The stronger cluster was within the measures of linguistic entrainment, supporting the general idea that different forms of entrainment relate to and perhaps facilitate each other, as hypothesized by Pickering and Garrod (2004), and partially supported by studies showing positive relations between syntactic and lexical entrainment (Hopkins et al., 2016; Rowland et al., 2012). However, the network of relations is markedly different between conversational contexts. Whereas affiliative conversations present a strong cluster of positive correlations between entrainment types, task-oriented conversations display a sparser cluster also including negative correlations. This indicates both that there is context sensitive top-down modulation of the mechanisms underlying entrainment (against a pure priming mechanism) and that task-oriented conversations might shape the use of conversational devices beyond a general increase of coordination.

The second cluster was—predictably—within measures of repairs, which directly relates to the way we express specific repair types as a proportion of the total repairs observed: the more open requests are observed, the fewer specific repairs. We do not observe—with the current conservative threshold—any trade-offs between different categories of conversational devices: for instance, between repairs and entrainment, or repair and backchannels. However, an exploratory analysis lowering the threshold to .25 (about 6% of the variance in common, see Figure 7) indicates that some trade-offs exist, but they appear to be very context dependent and should be investigated in larger corpora. Future work should more systematically assess these networks across contexts and also explore the possibility of trade-offs at different scales, for instance in turn or sequence level covariance (for an example of this, see Fusaroli et al., 2017).

Overview of the Conceptual Contributions

Based on previous literature and a within-participant experimental design that manipulated the conversational contexts, we contribute to the foundations of a more comprehensive conceptual and methodological framework for investigating the role of conversational devices in mutual understanding.

We argue that there is a role for both bottom-up mechanisms (priming of linguistic entrainment) as well as top-down mechanisms (the use of conversational devices is modulated by contextual demands). Both aspects seem to provide complementary insights as to the functioning of conversations.

We also argue that conversational devices seem to be used adaptively. Linguistic entrainment in particular show robust cross-task relations to joint performance. Although these findings are still largely observational, they provide a more robust and nuanced foundation for future studies that more directly assess causal impact. In particular, previous studies have argued for contradictory functions of linguistic entrainment and provided conflicting evidence. Here we show across tasks that a lower frequency of entrainment (rate) might be indicative of complementary contributions to the task and therefore higher performance; but once entrainment happens (level) more extensive reuse of the linguistic forms might provide a better check of mutual understanding and is positively related to performance. This might help resolve the contradictions in the literature, as both similarity and dissimilarity of contributions are seen to play a role.

Furthermore, the literature has often opposed more explicit forms of conversational coordination—grounding, based on top-down conversational devices, such as backchannels and repairs—to more implicit ones—interactive alignment, based on more bottom-up entrainment. Our findings suggest that natural conversations might defy a mutual exclusion between the two approaches, and mutual understanding relies on complex dynamics of implicit and explicit mechanisms. Indeed, both explicit conversational devices, such as restricted offers, and implicit conversational devices, such as entrainment, display independent correlations with performance. More research is needed to identify the fine-grained articulation of how these devices interact on a turn-by-turn level.

Limitations

The current series of studies presents some limitations. First, while we manipulated contextual demands, we did so with qualitatively different conversational tasks. Our initial concerns were to assess the generalizability of patterns across varied tasks inspired by the existing literature. This has obvious limitations. Our selection is not extensive and including additional—and more diverse—tasks is a necessary next step. But more crucially, future work will have to directly and systematically address the dimensions underlying the diversity of these tasks to more precisely identify the mechanisms underlying variations in conversational dynamics. For instance, one could parametrically manipulate the need for precise information sharing, and amount of shared visual references within the same task. This would permit stronger claims regarding the relation between the use of conversational devices and contextual demands.

Another important limitation is that the order of the tasks was fixed: the Alien Task appeared first, the Map Task second. Order did not seem to matter for task differences in our pilot data, and we note that the two affiliative conversations—the first and last of the conversations—are more similar to each other than the two tasks are. Note also, that having a fixed order makes the within-subject comparison less noisy because any order effects will be similar across participants. This is particularly important for the analyses in Part II. However, order effects are an important concern and observed differences between the two tasks should be treated with caution.

Second, a more fine-grained qualitative assessment of the concrete use of backchannels, repairs, and entrainment in context is still missing. For instance, backchannels might be used in qualitatively different ways in affiliative and task-oriented conversations, or within different phases of a conversation (Bangerter & Clark, 2003), or they may represent different subtypes with their own

interactional functions (O'Keeffe & Adolphs, 2008). Analogously, entraining to one's interlocutor might fulfil a variety of functions: requesting a confirmation or expressing incredulity (repair), elaborating on the other's utterance, displaying shared understanding, etc. Furthermore, entrainment might involve linguistic forms more or less relevant to the task (e.g., as highlighted in Fusaroli et al., 2012; Hawkins et al., 2020). Purely automated analyses are not, at this stage, able to capture this complexity, and future work should complement them with a more nuanced functional assessment of the conversations.

Third, although the current studies offer new and interesting insights about the correlation between conversational devices and performance, it is important to stress that these are observational relations, and that any causal relation cannot be inferred from this experimental setting. Conversational devices are notoriously difficult to manipulate, although some promising directions have been identified. For instance, some studies have manipulated the degree of shared information between interlocutors either by providing the same or different background information (Richardson, 2007) or manipulating the presence/absence of backchannels (Brown-Schmidt, 2012). In other studies, written chats and virtual reality settings can be used to surreptitiously manipulate conversations, for example, by adding repair utterances, or changing the frequency of backchannels (Hale et al., 2020; Mills, 2014). At the same time, conversations can be described as complex systems (Fusaroli, Rączaszek-Leonardi, et al., 2014), implying that manipulations of one element are likely to cascade across other elements and any future experiment should acknowledge this. Additionally, computational modeling of conversational devices could provide a much needed check on the theoretical accounts of conversational coordination, as it forces them to be specific and can more precisely assess their implications (e.g., Arkel et al., 2020; Dale et al., 2013, 2016; Hawkins et al., 2020, 2021).

Fourth, we investigated conversations as whole units, assessing rate and level of lexical entrainment across the entire conversation. However, there is increasing evidence that the use of conversational devices changes over the course of a conversation: entrainment has been reported to decrease (Duran et al., 2019), and repair has been reported to play different functions depending on when it is used in a conversation (Mills, 2014). Future work should further investigate the temporal dynamics and timescales in the use of conversational devices and how they covary.

Fifth, the current work focused on verbal aspects of backchannels, repairs, and entrainment. However, these three mechanisms do not exhaust the complexity of conversations. For instance, several studies have highlighted the pervasiveness and importance of multimodality in conversations: from the fine-grained temporal dynamics of turn taking to eye blinks, gaze, pointing and other bodily gestures (Holler & Levinson, 2019; Louwerse et al., 2012). Differential affordances of interactional settings across tasks might also have influenced participants' reliance on nonverbal versus verbal signals of mutual understanding. Investigating more than one modality would enhance our understanding of the complex interdependencies of language in interaction, and the mechanisms behind establishing and maintaining mutual understanding. Multimodal semiotic resources add another level of complexity to the pattern of individual differences and contextual dependencies that characterizes language in interaction.

Finally, it is important to note that our studies investigated conversational dynamics in Danish, a medium-sized language with about six million speakers, using Danish university students. To better assess frequencies and variability in frequency in the use of conversational devices, we need more diverse samples within Danish and across languages. Besides furthering a more basic understanding of how these mechanisms are used and vary, a more systematic integration of diverse samples can shed light on the interplay between cognitive and linguistic variation and the way it is accommodated by and compensated for in conversation. For example, Danish is known for its strong consonantal reduction compared with Norwegian (Trecca et al., 2021). It is conceivable that the properties of specific languages, like Danish, might impact conversational behaviors in idiosyncratic ways (Dideriksen et al., 2022). It will therefore be important to study conversational devices across a typologically broad set of languages and across a wider spectrum of participants (cf. Christiansen et al., in press). For a list of recommendations for future studies, summing up the previous paragraphs, see Section 6 in the online supplemental materials.

Broader Implications

With the current study we have combined knowledge from several fields and investigated three distinct conversational devices together to assess interdependencies between them. This has led to a better understanding of conversational devices and their interactional functions, with potential applications in at least three domains: human-computer interaction, team collaboration, and clinical research.

With the increasing demand for well-functioning speech technology, research in conversational devices has become more relevant. Human-computer interaction is already an integrated part of many people's lives, with voice-controlled software and hand-held devices. Despite the staggering progress in this area, many artificial conversation systems are characterized by simple turn-taking structures, which struggle to create context-appropriate sentences and sequences, and often require a predetermined specific use of words or word order to execute a command (Loth et al., 2015; Sugiyama et al., 2018). As the complexity of the situations in which we use these systems increases (for instance, noisy environments or planning of more complex activities such as travels), more knowledge about conversational structure is necessary to ensure a smooth and accurate interaction. The current study supplies a number of pieces to this puzzle and contributes to a more comprehensive understanding of the structure of conversations. Future studies could focus on how to implement more specific repairs, for example, in automated voice assistants to increase the chances of successful outcomes. Likewise, attention to how higher levels of entrainment in voice assistants might also have an effect on how well commands are carried out (Fischer, 2014) could benefit the development of automated speech systems, as well as furthering the development of multisentence conversations.

Research in team collaboration (for instance, involving disciplines such as cognitive science and management) aims to uncover the mechanisms of successful team coordination to promote effective and efficient collaboration (Fusaroli & Tylén, 2016; Pentland, 2012; Wiltshire et al., 2018). The finding that some conversational devices are positively correlated with performance could be used

to better understand performance in group efforts and further develop team coordination monitoring tools.

Last, a range of neuropsychiatric disorders, such as schizophrenia and autism spectrum disorders, are characterized by atypical social functioning (Dwyer et al., 2019; Hopkins et al., 2016); however, there is little knowledge about how it concretely unfolds in social situations (Bottema-Beutel, 2017; Fusaroli et al., 2019, 2021). A better general understanding of conversational mechanisms can help us characterize the varied ways in which neurodiverse people conduct interaction. The methods we have developed here are agnostic to people's neuropsychiatric profiles and so can be applied to any kind of social interaction to deliver insights into forms and functions of conversational devices.

Conclusions

By reviewing previous research findings, and implementing a quantitative experimental approach, we developed a theoretically and empirically motivated framework for the study of conversational dynamics based on how conversational devices, providing different levels of precision, are adaptively used to meet contextual demands. By carrying out a principled comparison of affiliative and task-oriented conversations, sharing data and code for others to build on, and formulating new insights into the relations between backchannels, repair, and entrainment, our work provides the empirical, methodological and conceptual foundations for a framework that unifies research into the ways people cocreate meaning in interaction.

Context of the Research

This article is part of a larger research program aimed at understanding how people manage to successfully coordinate with one another in conversations and what mechanisms they use when communication fails (Christiansen & Chater, 2016; Dideriksen, Fusaroli, et al., 2019; Dingemanse et al., 2015; Fusaroli et al., 2017). Groundbreaking work—like Clark & Brennan (1991) and Pickering and Garrod (2004)—has highlighted the importance of conversational devices in conversations, but much of this research has been reported in separate literatures, relying on post hoc synthetic attempts to compare studies. We aim at building a more comprehensive, integrative, and quantitative framework for understanding coordinative mechanisms, their relation to each other, and to context. Specifically, our work compares between- and within-subject study designs, systematically manipulates the goal of the conversations, and provides open access to the corpora and scripts to enable future research to incorporate and improve on the current work.

References

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

Arkel, J., van Woensdregt, M. S., 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). https://doi.org/10.18653/v1/2020.conll-1.14

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

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

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146. https://doi.org/10.1162/tacl_a_00051

Bottema-Beutel, K. (2017). Glimpses into the blind spot: Social interaction and autism. *Journal of Communication Disorders*, *68*, 24–34. https://doi.org/10.1016/j.jcomdis.2017.06.008

Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, *104*(2), 163–197. https://doi.org/10.1016/j.cognition.2006.05.006

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Brown, P., & Levinson, S. C. (1978). Universals in language usage: Politeness phenomena. In E. N. Goody (Ed.), *Questions and politeness: Strategies in social interaction* (pp. 56–311). Cambridge University Press.

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, *27*(1), 62–89. https://doi.org/10.1080/01690965.2010.543363

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Carbary, K., & Tanenhaus, M. (2011). Conceptual pacts, syntactic priming, and referential form. *Proceedings of the CogSci Workshop on the Production of Referring Expressions: Bridging the Gap Between Computational, Empirical and Theoretical Approaches to Reference (PRE-CogSci 2011)*, 1–6.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

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

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual Review of Psychology*, *64*(1), 285–308. https://doi.org/10.1146/annurev-psych-113011-143754

Christiansen, M. H., & Chater, N. (2016). *Creating language: Integrating evolution, acquisition, and processing*. MIT Press. https://doi.org/10.7551/mitpress/10406.001.0001

Clark, H. H. (1985). Language use and language users. In G. A. Lindzey Elliot (Ed.), *The handbook of social psychology* (Vol. 2, 3rd ed. pp. 179–231). Harper and Row.

Clark, H. H. (2009). Context and common ground. In M. L. Jacob (Ed.), *Concise encyclopedia of pragmatics* (pp. 116–119). Elsevier.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition*, *13*, 127–149. https://doi.org/10.1037/10096-006

Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, *2*(1), 19–41. https://doi.org/10.1080/01690968708406350

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*(2), 259–294. https://doi.org/10.1207/s15516709cog1302_7

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

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*(5), 1–9.

Cutrone, P. (2005). A case study examining backchannels in conversations between Japanese–British dyads. *Multilingua - Journal of Cross-Cultural and Interlanguage Communication*, *24*(3), 237–274. https://doi.org/10.1515/mult.2005.24.3.237

Cutrone, P. (2011). Politeness and face theory: Implications for the backchannel style of Japanese L1/L2 Speakers. *Language Studies Working Papers*, *3*, 51–57.

Dale, R. (2015). An integrative research strategy for exploring synergies in natural language performance. *Ecological Psychology*, *27*(3), 190–201. https://doi.org/10.1080/10407413.2015.1068649

Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). The self-organization of human interaction. *Psychology of Learning and Motivation*, *59*, 43–95. https://doi.org/10.1016/B978-0-12-407187-2.00002-2

Dale, R., Fusaroli, R., Tylén, K., Raczaszek-Leonardi, J., & Christiansen, M. H. (2016). Exploring mechanisms for interaction in a connectionist framework. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 901–906). Cognitive Science Society.

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

Dideriksen, C., Christiansen, M. H., Dingemanse, M., Højmark-Bertelsen, M., Johansson, C., Tylén, K., & Fusaroli, R. (2022). *Language specific constraints on conversation: Evidence from Danish and Norwegian*. PsyArXiv. https://doi.org/10.31234/osf.io/t3s6c

Dideriksen, C., Christiansen, M. H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2019). Danish dialogues. https://osf.io/2r4h3/

Dideriksen, C., Fusaroli, R., Tylén, K., Dingemanse, M., & Christiansen, M. H. (2019). Contextualizing conversational strategies: Backchannel, Repair and linguistic alignment in spontaneous and task-oriented conversations. In A. Goel, C. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 261–267). Cognitive Science Society.

Dingemanse, M., & Enfield, N. J. (2015). Other-initiated repair across languages: Towards a typology of conversational structures. *Open Linguistics*. Advance online publication. https://doi.org/10.2478/opli-2014-0007

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, 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), Article e0136100. https://doi.org/10.1371/journal.pone.0136100

Dumas, G., & Fairhurst, M. T. (2021). Reciprocity and alignment: Quantifying coupling in dynamic interactions. *Royal Society Open Science*, *8*(5), Article 210138. https://doi.org/10.1098/rsos.210138

Duran, N. D., & Fusaroli, R. (2017). Conversing with a devil's advocate: Interpersonal coordination in deception and disagreement. *PLoS ONE*, *12*(6), Article e0178140. https://doi.org/10.1371/journal.pone.0178140

Duran, N. D., Paxton, A., & Fusaroli, R. (2019). ALIGN: Analyzing linguistic interactions with generalizable techNiques-A Python library. *Psychological Methods*, *24*(4), 419–438. https://doi.org/10.1037/met0000206

Dwyer, K., David, A. S., McCarthy, R., McKenna, P., & Peters, E. (2019). Linguistic alignment and theory of mind impairments in schizophrenia patients' dialogic interactions. *Psychological Medicine*, *50*(13), 2194–2202.

Fischer, K. (2014). Alignment or collaboration? How implicit views of communication influence robot design. *2014 International Conference on Collaboration Technologies and Systems (CTS),* (pp. 115–122).

Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive Science*, *40*(1), 145–171. https://doi.org/10.1111/cogs.12251

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., Gangopadhyay, N., & Tylén, K. (2014). The dialogically extended mind: Language as skilful intersubjective engagement. *Cognitive Systems Research*, *29*, 31–39. https://doi.org/10.1016/j.cogsys.2013.06.002

Fusaroli, R., Rączaszek-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

Fusaroli, R., Tylén, K., Garly, K., Steensig, J., Christiansen, M. H., & Dingemanse, M. (2017). Measures and mechanisms of common ground: Backchannels, conversational repair, and interactive alignment in free and task-oriented social interactions, 2055–2060.

Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2019). Hearing me hearing you: Reciprocal effects between child and parent language in autism and typical development. *Cognition*, *183*, 1–18. https://doi.org/10.1016/j.cognition.2018.10.022

Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2021). *Caregiver linguistic alignment to autistic and typically developing children*. Psyarxiv. https://doi.org/10.31234/osf.io/ysjec

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), Article e103182. https://doi.org/10.1371/journal.pone.0103182

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

Gardner, R. (2001). *When listeners talk*. Benjamins. https://doi.org/10.1075/pbns.92

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218. https://doi.org/10.1016/0010-0277(87)90018-7

Goodwin, C. (1986). Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies*, *9*(2–3), 205–217. https://doi.org/10.1007/BF00148127

Goudbeek, M., & Krahmer, E. (2012). Alignment in interactive reference production: Content planning, modifier ordering, and referential overspecification. *Topics in Cognitive Science*, *4*(2), 269–289. https://doi.org/10.1111/j.1756-8765.2012.01186.x

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, *34*(4), 365–399. https://doi.org/10.1007/s10936-005-6139-3

Grønnum, N. (2009). A Danish phonetically annotated spontaneous speech corpus (DanPASS). *Speech Communication*, *51*(7), 594–603. https://doi.org/10.1016/j.specom.2008.11.002

Hale, J., Ward, J. A., Buccheri, F., Oliver, D., & Hamilton, A. F. C. (2020). Are you on my wavelength? Interpersonal coordination in dyadic conversations. *Journal of Nonverbal Behavior*, *44*(1), 63–83. https://doi.org/10.1007/s10919-019-00320-3

Hawkins, R. D., Frank, M. C., & Goodman, N. D. (2020). Characterizing the dynamics of learning in repeated reference games. *Cognitive Science*, *44*(6), Article e12845. https://doi.org/10.1111/cogs.12845

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), Article e12926. https://doi.org/10.1111/cogs.12926

Healey, P. G., Purver, M., & Howes, C. (2014). Divergence in dialogue. *PLoS ONE*, *9*(2), Article e98598. https://doi.org/10.1371/journal.pone.0098598

Heldner, M., Hjalmarsson, A., & Edlund, J. (2013). Backchannel relevance spaces. In E. L. Asu & P. Lippus (Eds.), *Nordic Prosody XI, Tartu, Estonia, 15–17 August, 2012* (pp. 137–146). Peter Lang Publishing Group.

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Hömke, P., Holler, J., & Levinson, S. C. (2018). Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLoS ONE*, *13*(12), Article e0208030. https://doi.org/10.1371/journal.pone.0208030

Hopkins, Z., Yuill, N., & Keller, B. (2016). Children with autism align syntax in natural conversation. *Applied Psycholinguistics*, *37*(2), 347–370. https://doi.org/10.1017/S0142716414000599

Ireland, M. E., & Henderson, M. D. (2014). Language style matching, engagement, and impasse in negotiations. *Negotiation and Conflict Management Research*, *7*(1), 1–16. https://doi.org/10.1111/ncmr.12025

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

Jefferson, G. (1984). Notes on a systematic deployment of the acknowledgement tokens "yeah"; and "mm hm." *Papers in Linguistics*, *17*(2), 197–216.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38.

Kirkedal, A., Plank, B., Derczynski, L., & Schluter, N. (2019). The Lacunae of Danish Natural Language Processing. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics* (pp. 356–362).

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

Loth, S., Jettka, K., Giuliani, M., & de Ruiter, J. P. (2015). Ghost-in-the-Machine reveals human social signals for human-robot interaction. *Frontiers in Psychology*, *6*, Article 1641. https://doi.org/10.3389/fpsyg.2015.01641

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, *36*(8), 1404–1426. https://doi.org/10.1111/j.1551-6709.2012.01269.x

Mills, G. (2014). Establishing a communication system: Miscommunication drives abstraction. In *Evolution of Language: Proceedings of the 10th International Conference (EVOLANG10)* (pp. 193–194). https://doi.org/10.1142/9789814603638_0023

Mills, G., Groningen, C., & Redeker, G. (2017). Amplifying signals of misunderstanding improves coordination in dialogue. In C. Howes & H. Rieser (Eds.), FADLI 2017 *Formal Approaches to the Dynamics of Linguistic Interaction 2017: Proceedings of the Workshop on Formal Approaches to the Dynamics of Linguistic Interaction 2017 co-located within the European Summer School on Logic, Language and Information (ESSLLI 2017)* (Vol. 1863, pp. 52–54). (CEUR Workshop Proceedings). CEUR-WS.org

Misiek, T., Favre, B., & Fourtassi, A. (2020). Development of multilevel linguistic alignment in child-adult conversations. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 54–58).

Mueller, S. T., Perelman, B. S., & Veinott, E. S. (2016). An optimization approach for mapping and measuring the divergence and correspondence between paths. *Behavior Research Methods*, *48*(1), 53–71. https://doi.org/10.3758/s13428-015-0562-7

O'Keeffe, A., & Adolphs, S. (2008). Response tokens in British and Irish discourse: Corpus, context and variational pragmatics. *Pragmatics and beyond New Series*, *178*, Article 69.

Peirce, J., & MacAskill, M. (2018). *Building experiments in PsychoPy*. Sage.

Pentland, A. (2012). The new science of building great teams. *Harvard Business Review*, *90*(4), 60–69.

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. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge University Press. https://doi.org/10.1017/9781108610728

Prinzo, O., & Britton, T. (1993). ATC/pilot voice communications-a survey of the literature: DTIC Document.

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in multimodal interaction: An integrative framework. *Cognitive science*. Advance online publication. https://doi.org/10.1111/cogs.12911

Reitter, D., & Moore, J. D. (2006). Priming of Syntactic Rules in Task-Oriented Dialogue and Spontaneous Conversation. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society, Vancouver, 2006*.

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, *76*, 29–46. https://doi.org/10.1016/j.jml.2014.05.008

Richardson, J. E. (2007). Analysing texts: Some concepts and tools of linguistic analysis. In J. E. Richardson (Ed.), *Analysing newspapers: An approach from critical discourse analysis* (pp. 46–74). Palgrave Macmillan. https://doi.org/10.1007/978-0-230-20968-8_3

Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in Psychology*, *2*, Article 38. https://doi.org/10.3389/fpsyg.2011.00038

Roger, D., & Nesshoever, W. (1987). Individual differences in dyadic conversational strategies: A further study. *British Journal of Social Psychology*, *26*(3), 247–255. https://doi.org/10.1111/j.2044-8309.1987.tb00786.x

Rowland, C. F., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. M. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition*, *125*(1), 49–63. https://doi.org/10.1016/j.cognition.2012.06.008

Sacks, H. (1992). *Lectures on conversation* (edited by G. Jefferson). Blackwell.

Scheflen, A. E. (1964). The significance of posture in communication systems. *Psychiatry*, *27*(4), 316–331. https://doi.org/10.1080/00332747.1964.11023403

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh'and other things that come between sentences. *Analyzing Discourse: Text and Talk*, *71*, Article 93.

Schegloff, E. A. (2000). When'others' initiate repair. *Applied Linguistics*, *21*(2), 205–243. https://doi.org/10.1093/applin/21.2.205

Schegloff, E. A. (2006). Interaction: The infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted. In S. C. L. N. J. Enfield (Ed.), *Roots of human sociality: Culture, cognition and interaction* (pp. 70–96). Berg.

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

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 326–332. https://doi.org/10.1037/0096-1523.29.2.326

Slocombe, K. E., Alvarez, I., Branigan, H. P., Jellema, T., Burnett, H. G., Fischer, A., Li, Y. H., Garrod, S., & Levita, L. (2013). Linguistic

alignment in adults with and without Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *43*(6), 1423–1436. https://doi.org/10.1007/s10803-012-1698-2

Steensig, J., Brøcker, K. K., Grønkjær, C., Hamann, M. G. T., Hansen, R. P., Jørgensen, M., Kragelund, M. H., Mikkelsen, N. H., Mølgaard, T., & Pedersen, H. F. (2013). The DanTIN project–creating a platform for describing the grammar of Danish talk-in-interaction. In J. Heegård & P. J. Henrichsen (Eds.), *New perspectives on speech in action* (pp. 195–225). Samfundslitteratur Press.

Strømberg-Derczynski, L., Baglini, R., Christiansen, M. H., Ciosici, M. R., Dalsgaard, J. A., Fusaroli, R., Henrichsen, P. J., Hvingelby, R., Kirkedal, A., & Kjeldsen, A. S. (2020). The Danish Gigaword Project. Preprint ArXiv:2005.03521.

Sugiyama, H., Meguro, T., Yoshikawa, Y., & Yamato, J. (2018). Improving dialogue continuity using inter-robot interaction. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 105–112). https://doi.org/10.1109/ROMAN.2018.8525542

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Tolins, J., & Fox Tree, J. E. (2016). Overhearers use addressee backchannels in dialog comprehension. *Cognitive Science*, *40*(6), 1412–1434. https://doi.org/10.1111/cogs.12278

Tolins, J., & Tree, J. E. F. (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

Trecca, F., Tylén, K., Højen, A., & Christiansen, M. H. (2021). Danish as window onto language processing and learning. *Language Learning*, *71*(3), 799–833. https://doi.org/10.1111/lang.12450

Tylén, K., Fusaroli, R., Smith, P., & Arnoldi, J. (2020). *The social route to abstraction: Interaction and diversity enhance rule-formation and transfer in a categorization task.* Psyarxiv. https://doi.org/10.31234/osf.io/qs253

Wang, Y., & Hamilton, A. F. D. C. (2012). Social top-down response modulation (STORM): A model of the control of mimicry in social interaction. *Frontiers in Human Neuroscience*, *6*, Article 153. https://doi.org/10.3389/fnhum.2012.00153

Wiltshire, T. J., Butner, J. E., & Fiore, S. M. (2018). Problem-solving phase transitions during team collaboration. *Cognitive Science*, *42*(1), 129–167. https://doi.org/10.1111/cogs.12482

Xu, Y., & Reitter, D. (2015). An evaluation and comparison of linguistic alignment measures. *Proceedings of the 6th Workshop on Cognitive Modeling and Computational Linguistics*, 58–67.

Xu, Y., & Reitter, D. (2017). Spectral analysis of information density in dialogue predicts collaborative task performance. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 623–633). https://doi.org/10.18653/v1/P17-1058

Yngve, V. (1970). On getting a word in edgewise. In *Chicago Linguistics Society, 6th meeting* (pp. 567–577).

Received November 25, 2020
Revision received July 14, 2022
Accepted August 4, 2022 ■