COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 47 (2023) e13298

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13298

The Dynamic Interplay of Kinetic and Linguistic Coordination in Danish and Norwegian Conversation

James P. Trujillo,[a,b] Christina Dideriksen,[c] Kristian Tylén,[c,d]
Morten H. Christiansen,[c,d,e] Riccardo Fusaroli[c,d,f]

[a]*Donders Institute for Brain, Cognition and Behaviour, Radboud University*
[b]*Max Planck Institute for Psycholinguistics*
[c]*Linguistics, Cognitive Science and Semiotics, School of Communication and Culture Aarhus University*
[d]*The Interacting Minds Center, School of Communication and Culture Aarhus University, Aarhus University*
[e]*Department of Psychology, Cornell University*
[f]*Linguistic Data Consortium University of Pennsylvania*

Received 4 May 2022; received in revised form 22 February 2023; accepted 2 May 2023

Abstract

In conversation, individuals work together to achieve communicative goals, complementing and aligning language and body with each other. An important emerging question is whether interlocutors entrain with one another equally across linguistic levels (e.g., lexical, syntactic, and semantic) and modalities (i.e., speech and gesture), or whether there are complementary patterns of behaviors, with some levels or modalities diverging and others converging in coordinated fashions. This study assesses how kinematic and linguistic entrainment interact with one another across levels of measurement, and according to communicative context. We analyzed data from two matched corpora of dyadic interaction between—respectively—Danish and Norwegian native speakers engaged in affiliative conversations and task-oriented conversations. We assessed linguistic entrainment at the lexical, syntactic, and semantic level, and kinetic alignment of the head and hands using video-based motion tracking and dynamic time warping. We tested whether—across the two languages—linguistic alignment correlates with kinetic alignment, and whether these kinetic-linguistic associations are modulated either by the type of conversation or by the language spoken. We found that kinetic entrainment was positively associated with low-level linguistic (i.e., lexical) entrainment, while negatively associated with high-level linguistic (i.e., semantic) entrainment, in a cross-linguistically robust way. Our findings suggest that conversation makes use of a dynamic coordination of similarity and complementarity both

Correspondence should be sent to James P. Trujillo, Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6525 GD Nijmegen, The Netherlands. E-mail: james.trujillo@mpi.nl

between individuals as well as between different communicative modalities, and provides evidence for a multimodal, interpersonal synergy account of interaction.

Keywords: Conversation; Multimodal; Communication; Language; Alignment

1. Introduction

Social interaction frequently involves discussions, idle chats, or cooperative decision-making, all of which require us to get on the same page as our interlocutor and ensure that we stay in tune. As such, conversations can be seen as a form of joint action where interlocutors work together in order to achieve their communicative goals (Clark, 1996). Depending on the specific goal of the interaction (e.g., to solve a problem, plan a trip together, or exchange information), the collaborative dynamics of this process involves various degrees of both complementary behaviors, as well as entrainment or mimicry of behaviors.

Previous research has shown that interpersonal coordination occurs across communicative modalities. This may include linguistic entrainment, such as in the reuse of one's interlocutor's lexical and syntactic structure (Dideriksen, Christiansen, Tylén et al., 2022; Pickering & Garrod, 2004), as well as kinetic entrainment, such as synchronization of body and head movements, and the use of similar gestures (Louwerse, Dale, Bard, & Jeuniaux, 2012; Paxton & Dale, 2017; Ramseyer & Tschacher, 2011; Rasenberg, Dingemanse, & Özyürek, 2020; Rasenberg, Özyürek, Bögels, & Dingemanse, 2022; Tsuchiya et al., 2020). However, despite multimodality being the natural context of human language—with gestures, nods, and ever-changing facial expressions accompanying speech—coordination through entrainment is typically not studied across modalities. This has led researchers to call for a more integrative approach to studying how people coordinate in communication (Rasenberg et al., 2020; de Ruiter & Albert, 2017), with the current work being part of a growing movement of studying cross-modal entrainment of communicative behaviors (Alviar, Dale, Dewitt, & Kello, 2020; Condon & Ogston, 1966; Paxton, Dale, & Richardson, 2016; Pouw et al., 2021).

Interestingly, there are opposing theories about how coordination is likely to occur across modalities. Some accounts have focused on the importance of alignment and similarity in facilitating coordination (K. A. Duffy & Chartrand, 2015; Garrod & Pickering, 2009; Ireland & Henderson, 2014; Pickering & Garrod, 2004), wherein interlocutors automatically entrain with one another throughout the course of an interaction, across modalities and across levels of analysis. For example, as the conversation proceeds, interlocutors are expected to become more and more similar in the words they use to refer to something (Fusaroli et al., 2012; Hawkins, Frank, & Goodman, 2020), in their average speech rate (Manson, Bryant, Gervais, & Kline, 2013), as well as their posture (Paxton & Dale, 2017). By aligning more with one another, interlocutors have an easier time predicting each other's behavior and sharing conceptualizations of the activity at hand (Pickering & Garrod, 2013).

Louwerse et al. (2012) observed the entrainment of multiple visual and linguistic behaviors, as would be predicted by the interpersonal alignment account. Yet, they also suggest that this alignment could be more of a "stable background" that allows coordination at higher levels (Louwerse et al., 2012). However, these "higher levels" have remained somewhat underspecified, and there was no evidence for a difference in entrainment between levels. Other accounts have taken this argument further, construing dialogue as interpersonal synergy (Dale, Fusaroli, Håkonsson, Healey, & Mønster, 2013; Fusaroli & Tylén, 2016; Fusaroli, Rączaszek-Leonardi, & Tylén, 2014), with the implication that similarity and complementarity in behavior both play equally important roles. In this view, "lower-levels" of behavior (e.g., postural sway) likely entrain as a stable background for the interaction (Shockley, Santana, & Fowler, 2003), while "higher-levels" (e.g., semantic or syntactic structures) may diverge as interlocutors avoid redundant contributions and coconstruct their conversation from diverse and complementary contributions and perspectives. Here, we define high- and low-level as relating to the unit of measurement that is required to capture the behavior. For example, phonemes, pitch, and posture would be considered low-level, while semantics, which is abstracted away from the specific speech sounds, would be high-level. This is in line with a recent study showing that speakers do not necessarily align on all levels, or features, during conversation (Ostrand & Chodroff, 2021), as well as earlier work showing that efficient communication likely involves greater degrees of freedom in how behaviors and behavioral complexity are coordinated between and within interactants (Paxton, Abney, Kello, & Dale, 2014). Combining these perspectives within a multimodal perspective brings to the fore the important issue of how coordination *within modality* relates to coordination *across modalities* (e.g., whether the frequent interpersonal reuse of linguistic expressions is accompanied by a frequent reuse of similar gestures, or whether it makes the latter superfluous).

Investigating the multimodal dynamics of human interaction allows us to directly contrast the imitation and the interpersonal synergy accounts of dialogue. Following similarity-focused models (e.g., Pickering & Garrod, 2004, 2009; Duff & Chartrand, 2015; Ireland & Henderson 2014), we would expect people to entrain both kinetically and linguistically, as this would reduce the complexity of the interaction and facilitate shared conceptualizations. Indeed, entrainment should spread across modalities (i.e., linguistic and kinetic) and linguistic levels, for instance, starting with kinetic and lexical entrainment and leading to syntactic and semantic entrainment. This mechanism should enable high-level cognitive alignment, which ultimately facilitates coordination. By contrast, an interpersonal synergy account would suggest at least one crucial difference. Higher kinetic entrainment might lead to lower—and not higher—linguistic entrainment, as it establishes enough similarity at one level, allowing linguistic contributions to become more complementary without an excessive risk of misunderstanding. Further, such correlations between linguistic and kinetic entrainment may also differ depending on the level of linguistic entrainment (e.g., lexical, syntactic, or semantic).

Given that these two accounts of entrainment are oriented toward effective interpersonal coordination, it is also important to take the communicative context into account (Fusaroli et al., 2014; Yeomans, Schweitzer, & Brooks, 2022). In both accounts, we expect to observe higher levels of entrainment in task-oriented conversations when compared to conversations without a specific task goal, in line with the finding that different types of conversation (i.e., affiliative vs. argumentative) affect the degree of entrainment, both kinetically (i.e., head movement) and in speech (i.e., speech rate, Paxton & Dale, 2013a, 2013b; see also Duran, Paxton, & Fusaroli, 2019; Duran & Fusaroli, 2017). However, according to the interpersonal synergy account, we would expect communicative context to affect the association between linguistic and kinetic entrainment, as the multimodal communicative system adapts dynamically to meet contextual demands.

In order to address the question of how multimodal communicative entrainment is realized in conversation, we analyzed two matched cross-linguistic corpora of dyads involving either native speakers of Danish or Norwegian and comprising interactions in both affiliative and task-oriented contexts. The two contexts of conversations are not assumed to elicit entirely different forms of behavior, given that they still involve dialogue between the participants. However, a task-oriented context is likely to push participants toward a more goal-oriented focus of their coordination. We hypothesize that such a shift in goals will be reflected in the kinetic and linguistic dynamics of our participants. We used manual transcriptions and automated annotations of the corpora to measure linguistic entrainment, and video-based motion tracking to capture kinetic entrainment. This allowed us to quantify how linguistic and kinetic entrainment are related to one another (Research Question 1) and how any cross-modal associations are modulated by these two communicative contexts (Research Question 2). Moreover, the cross-linguistic nature of the corpora provides a first assessment of the generalizability of our findings (Research Question 3) through a fine-grained comparison across two closely related languages.

2. Methods

2.1. Participants

We elicited conversations from 160 participants in 80 dyads (40 Danish and 40 Norwegian). The Danish corpus has been previously analyzed in Dideriksen, Christiansen, Tylén et al. (2022) and Dideriksen et al. (2019). Three dyads were excluded due to technical issues, leaving a total data set of 77 dyads. The Danish participants were recruited through the SONA recruitment system hosted by Cognition and Behavior lab at Aarhus University, Denmark, and consisted mainly of university students (five participants had completed a high school degree, and 75 participants were university students). The Norwegian participants were recruited through student groups and by handing out flyers on the campus of the University in Bergen (one participant had completed a high school degree and the remaining 79 either had a university degree or were university students). Conversations took place at the university. Participants were asked to complete a questionnaire prior to the conversations, stating their age (Danish corpus: mean 23.2, SD = 3.5; Norwegian corpus: mean = 23, SD = 3.4), gender (Danish corpus: 58% female, 42% male; Norwegian corpus: 51% female, 49% male), education, and dialectal affiliation (Danish corpus: Aarhusiansk: 38% Københavnsk: 6%, Vestjysk: 6%, other: 50%; Norwegian corpus: Østlandsk 35%, Bergensk 20%, Vestlands 13.75%, other 31.25%). Both Danish and Norwegian participants went through the same experimental design. Of the Danish participants, eight dyads knew each other before taking part in the experiment, while from the Norwegian participants, 19 dyads knew each other in advance. Approval was obtained at the IRB at Aarhus University and confirmed by the COBElab ethical committee, also at Aarhus University.

2.2. Procedure

Each dyad participated in four conversations: two affiliative conversations (AC) and two task-oriented conversations (TOC). Participants were seated next to each other, facing computer screens. Participants could adjust their position throughout in order to engage with tasks on the computer or sit more face-to-face to converse. For the first affiliative conversation, participants were provided with a sheet of open-ended conversation starters (e.g., "Discuss and agree on two superpowers that you would like to have") and instructed to get acquainted for a while.

Next, they were asked to complete two collaborative tasks. First, they played the Alien Game (Tylén, Fusaroli, Smith, & Arnoldi, 2016, 2020), a joint decision task, where the participants have to make decisions about how to categorize stimuli (in this case, a series of aliens) presented on a joint screen. Each alien appeared on the screen for 3 s after which it disappeared. The participants then had to jointly decide whether the alien was friendly, hostile, or valuable based on combinations of the aliens' physical traits (such as spotted skin or raised arms). A correct answer was rewarded with 100 points, while a wrong answer was penalized by the deduction 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.

Participants were then instructed to solve the Map Task (Anderson et al., 1991; Fay et al., 2018). This task is an asymmetric game, where participants take turns giving directions to one another about how to draw a path on a map, with the matcher being free to interact, ask for explanations, and so on. The participants sat next to each other, but used individual monitors. They were separated by a screen from the chest down, limiting their view while they still could see each other's facial expressions. 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 about the games without any need for the experimenter to prompt them. This conversation also lasted for about 10 min.

The fixed order of the tasks was introduced to make the within-subject comparison of task effects less noisy, as any order effects would be the same across participants.

2.3. Corpora

The Norwegian corpus consisted of 38 dyads, 150 conversations, and a total of 37,569 conversational turns. The average number of utterances per dyad was 234 for the first affiliative conversation and 219 for the second. The first task-oriented conversation (the Alien Game) had an average of 311 utterances per dyad, whereas the second (the Map Task) had 236.

The Danish corpus consisted of 39 dyads, 153 conversations, and a total of 38,646 conversational turns. The average number of utterances per dyad was 226 for the first affiliative conversation and 212 for the second. The first task-oriented conversation (the Alien Game) had an average of 321 utterances per dyad, whereas the second (the Map Task) had 236.

All conversations were transcribed orthographically (and are available at: https://osf.io/xaqvy/?view_only = 30eaf105cc9d4e77829e3da0fb5dedb5). The transcription is utterance-based and each turn was defined as a vocal production by one speaker that did not contain a pause longer than 1 s. A turn would continue as long as the participant was speaking even though they were interrupted by their conversational partner. Often backchannels or repairs would occur as overlapping speech. Camera recording setup was identical between the two corpora in terms of recording quality as well as distance and angle with respect to the participants.

2.4. Linguistic entrainment

Interactive entrainment was automatically measured relying on the standardized approach and software in Duran et al. (2019). All utterances were preprocessed to identify lemmas ("dog" and "dogs" becoming the same lemma "dog"), parts of speech (article, adverb, noun, etc.), and word-embedding-based semantic similarity. Word embeddings were based, respectively, on the Danish and Norwegian versions of Wikipedia and applied to the current corpora (Bojanowski, Grave, Joulin, & Mikolov, 2017). Word embeddings encode the meaning of a word as a vector of values, where words that appear in similar contexts are closer in vector space, and thus have similar values. This similarity in vector space is taken to represent a similarity in semantic meaning. Wikipedia data were used as there are currently no sufficiently large-scale corpora of conversational Danish and Norwegian that can be used to create reliable word embeddings. In our analyses, each word is identified in a 300-dimensional vector space, and we averaged word embeddings within each utterance.

After calculating word embeddings, we identified all pairs of successive utterances produced 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 constituted a dimension of the lexical vector, and the number of occurrences of that lemma in a given utterance constituted the value of that dimension. The syntactic vector included all unique parts of speech n-grams present in at least one of the utterances in the pair. Each unique n-gram constituted a dimension of the syntactic vector, and the number of occurrences of that n-gram in a given utterance determined the value of that dimension. The semantic vector was the utterance-level word embeddings representation described above. Linguistic entrainment was then calculated as the cosine similarity (i.e., 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 for the first and second interlocutor's utterances, and i indicates the i'th dimension in the vector. See Fig. 1 for a graphical representation of linguistic entrainment.

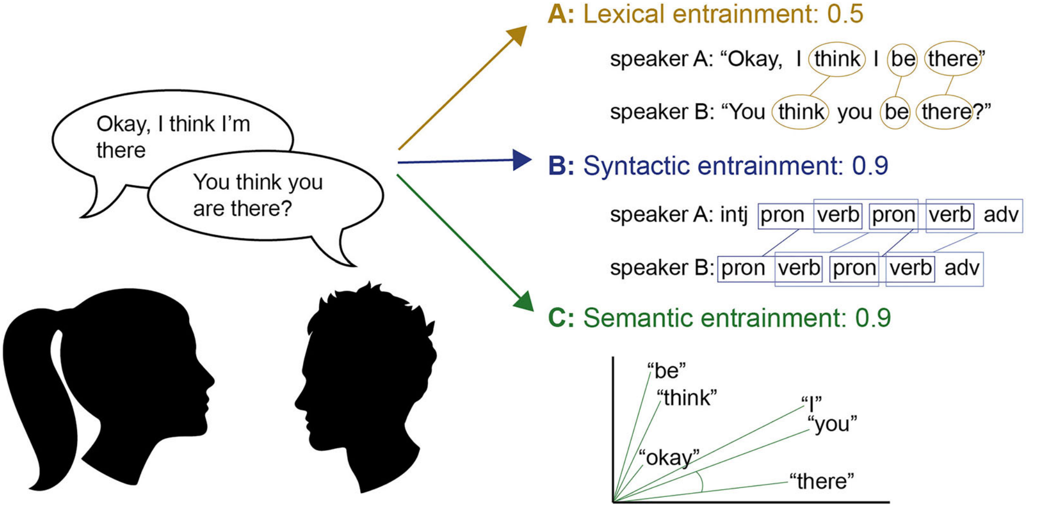

Fig. 1. Graphical representation of linguistic entrainment between two utterances (in lemma form), from speaker A and speaker B. The text **bubbles** provide the transcribed **(and translated)** utterances from the two speakers **(original Danish: Speaker A: "Okay, jeg tror, jeg er der"; Speaker B: "Du tror du er der")**. Semantic entrainment is based on the distances between word embeddings, syntactic entrainment is based on n-grams of parts of speech, while lexical entrainment is based on the occurrence of the same lemma between the two speakers. **Note that the text under A, B, and C is lemmatized du match how these features were calculated**.

2.5. *Kinetic entrainment*

In order to calculate kinetic entrainment, we first processed all videos using OpenPose (Qiao, Wang, & Li, 2017) to extract keypoint positions (in two-dimensional space) of the left and right hands and for the head. For the purpose of these analyses, we did not track or extract key points of the individual fingers, since these would not be used in the present analyses. After calculating keypoint positions for the videos, we extracted data separately, utterance by utterance (as described in the subsection Linguistic entrainment). Tracked keypoints were manually checked, by plotting two-dimensional displacement, to ensure there were no artifacts in the extracted data sets.

We calculated kinetic entrainment of the head and hands by applying *dynamic time warping* to the pairs of utterances relying on the dtw-python package (Giorgino, 2009). Dynamic time warping (DTW; Mueen & Keogh, 2016) is a method of determining the distance (or dissimilarity) between the shape of two time-series, with the advantage of being unaffected by differences in the actual length of the time-series. DTW has previously been applied to gesture velocities in order to show similarity between groups of gestures (Pouw & Dixon, 2019), and has even shown that semantically related gestures show lower DTW distance scores (i.e., lower dissimilarity; Pouw et al., 2021). We, therefore, utilized (centered) x,y coordinates for the nose in order to calculate head movement similarity between each pair of participants. For the hands, we first calculated the dissimilarity between the left hands in each pair, and the right hands in each pair. We then took the mean of these two

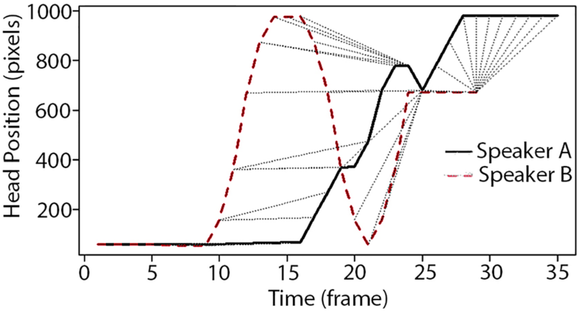

Fig. 2. Graphical representation of kinetic entrainment using dynamic time warping. The plot depicts the (centered) head position along the horizontal plane of the two speakers, recorded in two consecutive utterances. The x-axis shows the time in frames, while the y-axis shows the position of the head in pixels. Speaker A is denoted by the solid black line, while speaker B is indicated by the dashed red line. The dotted gray lines depict how the two time-series are aligned by the dynamic time series (DTW) calculation. Note that for simplicity, only alignment along the x-dimension of the motion data is depicted here, although our DTW calculation was performed on the two-dimensional (x,y) data.

values as our measure of hand movement dissimilarity. We did not consider cross-articulator entrainment (hand to head, or right-hand to left-hand) for simplicity, and following previous studies (Holler & Wilkin, 2011a; Pouw et al., 2021). Future work should consider more complex cross-articulator dynamics. As we used the spatial coordinates of the articulators, rather than a velocity trace, to calculate dissimilarity, the measure is effectively matching the actual movements produced, such as leaning forward, tilting the head, raising the hands in a particular way in front of the body, and so on. Note that DTW provides Euclidean distance as an output, which is inherently a measure of *dis*similarity. We, therefore, invert these values to get a measure of kinetic similarity, which we take in this case to represent kinetic *entrainment*. See Fig. 2 for a graphical representation of kinetic entrainment.

2.6. Statistical analyses

To analyze the relation between linguistic and kinetic entrainment, we constructed generalized linear mixed effects models with the two primary kinetic entrainment variables (i.e., head and hand distance) as the dependent variables. Our statistical models are constructed to predict the kinetic variables in order to directly assess our hypothesis that kinetic entrainment would correlate with some levels of linguistic entrainment, but not others. Additionally, these models always contained pair/video as a nested random term. Before assessing any model fits, we ensured that there were no convergence or singularity issues, and checked that the models adequately captured the data by checking the residuals. When convergence or singularity issues arose, the nested random effect was replaced by a simpler term, “file,” which was a unique identifier for each pair-video combination. Assessing model residuals also revealed that, for all models described here, a gamma distribution with log link function provided the best fit to the data.

Table 1
Number of included datapoints across conditions and model sets

	Danish			Norwegian		
	Affiliative	Task-oriented	Total	Affiliative	Task-oriented	Total
Semantic+lexical models	4869	6434	11,303	2676	3114	5790
Semantic-only models	5183	6748	11,617	3428	2990	6104

We built and tested our models across three stages, in order to test our three main research questions. First, we tested whether there was any evidence of *cross-modal coordination between kinetic and linguistic entrainment*. For this set of analyses, we started from a null model containing only random intercepts and subsequently tested whether adding context, language, and/or linguistic entrainment improved model fit (see below for model comparison method). Linguistic variables were added in the order of Semantic, Syntactic, then Lexical. Our next set of analyses aimed to answer whether there is evidence for *the association between kinetic and linguistic entrainment being modulated by context* (Task-Oriented vs. Affiliative conversation). For these analyses, we started with the final best-fit model from the previous step. We then tested whether adding an interaction term between context and each linguistic variable contributed to a better model fit. Finally, we test whether the *kinetic-linguistic associations are modulated by language* (Danish vs. Norwegian). For these analyses, we test the addition of language as an interaction term with any linguistic entrainment $\times$ context interaction term already present in the model. If this three-way interaction leads to fitting or convergence issues, we include the linguistic entrainment $\times$ language interaction terms separately in the model.

To determine the best model fit to our data, and to subsequently test whether there was a significant association between the linguistic entrainment variables and kinetic entrainment, we used chi-square tests of model comparison. This was done by first comparing a simple model containing only semantic/syntactic entrainment as independent variable (plus random slope, when possible) against a more complex model that also contained lexical entrainment. This was repeated again by adding communicative context as an additional independent variable. Whichever model was revealed to be the best fit to the data, based on the likelihood ratio test comparison, was considered the "final model." This test determined if the linguistic entrainment variables were significantly associated with the kinetic entrainment variable. We report the p-value for the final model comparison. p-Values for model parameters were calculated using the Anova function of the R package "car" (Fox, Friendly, & Weisberg, 2012). However, note that there is much discussion about whether parameter p-values can be accurately or confidently estimated from mixed effects models. We, therefore, suggest caution when interpreting these values. We provide an estimate of effect size via conditional coefficient of determination for generalized mixed effects models (R^2_{GLMM}) for mixed models, implemented using the package MuMIn (Barton, 2020). The number of included datapoints for each condition, and within the model sets, is provided in Table 1. All model equations are provided in Table S1.

In order to visualize multivariable associations, we used network plots, which depict the strength and direction of correlation between sets of variables (Csardi & Nepusz, 2005). As these associations are calculated across separate models, for visualization, we simply take the partial correlations (calculated using the partial.r function of R package *psych*, Revelle, 2020) between all of the entrainment variables. Partial correlations are calculated separately for the two conversation contexts. This was done instead of using extracted model parameters because these parameter values are specific to the model from which they are extracted, and cannot be easily combined across lexico-syntactic and lexico-semantic models. To provide a complete overview of the associations, we also generated network plots for each communicative context, both for Danish and Norwegian, including all entrainment variables, regardless of the final model outcomes. These plots are included in the Supplementary Materials.

Finally, we ran three control analyses on our data. First, we assessed whether kinetic entrainment was greater than chance level, in order to ensure that we are actually capturing entrainment, rather than random fluctuation in movement data. Therefore, control analyses were conducted in order to ensure that any entrainment measure used in the present study is above such baseline, chance-level entrainment. Results of the linguistic control analyses are reported by Dideriksen, Christiansen, Tylén, Dingemanse, and Fusaroli (2022) and show that all linguistic levels of entrainment are above the chance level. For the kinetic variables, we performed an analysis using shuffled data. Specifically, for each dyad, we randomly shuffled the order of each participant's utterances and then calculated the turn-to-turn entrainment as described above. Shuffling within dyads ensured that the control analysis took into account potential similarities between how participants in a dyad move, while testing whether the similarity is a function of the dynamic flow of the conversation. We then fit linear mixed models, with hand or head entrainment as the dependent variable, and "file" and "context" as random intercepts (language led to a singular model fit). Using likelihood ratio tests, we compared this model against a model that additionally included a data set (shuffled vs. real) as a fixed effect. We found that both head (χ^2 (1) = 25.26, $p < .001$) and hand (χ^2 (1) = 64.16, $p < .001$) entrainment are higher in the real data compared to the shuffled data (see Supplementary Material on the OSF).

In the second control analysis, we assessed whether potential discrepancies between turn durations could be influencing our main statistical models. To this end, we reran all of the statistical tests described above after removing turn sequences that differed substantially in duration. We, therefore, calculated the proportional difference between speaker A and speaker B in each turn sequence, and then defined a cutoff point as the third quartile plus 1.5 times the interquartile range. We then reran all of our statistical tests to ensure that model comparisons, and model parameter estimates remained the same in terms of outcome.

Finally, given that some dyads knew one another prior to participation, we additionally reran all final models (reported below) while also including partner familiarity (known vs. unknown) as an additional fixed effect. We again used likelihood ratio test comparisons to determine if familiarity explained a significant portion of the variance. We found no evidence for partner familiarity explaining additional variance in the kinetic entrainment data of the head (χ^2 (1) = 0.755, $p = .385$) or the hands (χ^2 (1) = 0.118, $p = .732$).

Table 2
Parameter estimates for final cross-modal coordination models

Head entrainment					Hand entrainment				
Entrainment parameter	β	Std. error	t	p	Entrainment parameter	β	Std. error	t	p
Semantic	–0.058	0.009	–5.991	<.001	Semantic	–0.049	0.007	–7.195	<.001
Lexical	0.017	0.005	3.562	<.001	Lexical	0.018	0.004	3.914	<.001
Context	0.012	0.005	2.651	.004	Context	0.013	0.005	2.432	.015
Semantic × Context	0.061	0.017	3.581	<.001	Semantic × Context	0.045	0.012	3.852	<.001

All processed data and analysis scripts are openly available on the Open Science Framework. https://osf.io/qy7v5/

3. Results

3.1. Cross-modal coordination

Our first analysis assessed whether kinetic entrainment is associated with linguistic entrainment. We found that a model predicting kinetic entrainment of head ($R^2 = 0.056$) and hand ($R^2 = 0.061$) from semantic and lexical entrainment, as well as communicative context was the best model ($p < .001$, thus excluding syntactic entrainment as meaningful additional predictor). In addition, for both kinetic variables, we found no evidence that adding a main effect of language (Danish/Norwegian) would improve the model fit, indicating that Danish and Norwegian present comparable levels of kinetic alignment within the current modeling approach. For detailed estimates for each of the final models, see Table 2, and for an overview, see Fig. 3. For a graphical representation of the correlations between factors, see Fig. 4. For an overview of the correlations additionally split by language, see Fig. S1.

3.2. Contextual modulation of kinetic-linguistic association

Our second analysis assessed whether the cross-modal associations discovered in the previous analyses were further modulated by context; in other words, whether there is evidence for an interaction effect between communicative context and linguistic entrainment. We found that the association between kinetic entrainment and semantic entrainment was modulated by communicative context for both head ($R^2 = 0.068$) and hand ($R^2 = 0.063$) ($p < .001$). In both cases, the negative semantic-kinetic entrainment association became weaker in the task-oriented conversation compared to the affiliative conversation. See Fig. S2 for the model predictions of this interaction effect.

3.3. Cross-linguistic modulation of kinetic-linguistic associations

For the final analysis, we assessed whether the cross-modal associations observed in the previous analyses were modulated by the language (Danish vs. Norwegian) being spoken. We found no evidence of language improving model fit neither for head or hand in the current

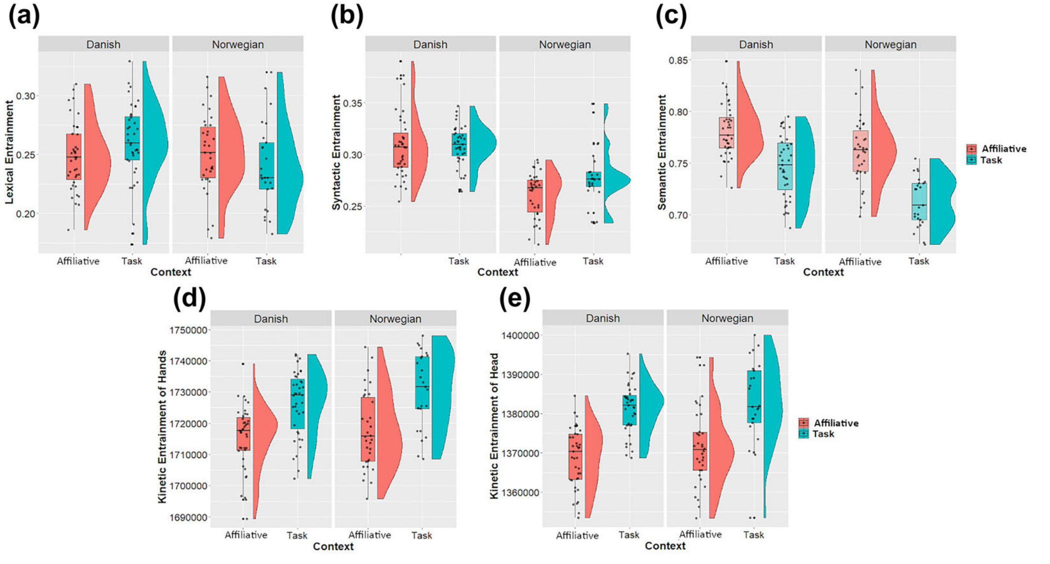

Fig. 3. Overview of data distributions, split by Context and Language. In each plot, boxplots provide the median (notch in box), first and third quartiles (lower and upper bounds of the box), and the furthest datapoint that is maximally 1.5 times the interquartile range from the nearest bound (whiskers). The half-violin shapes depict the density of data distribution. Mean values for each dyad are given as circles overlaid on the boxplots. Panel A shows Lexical Entrainment, panel B shows Syntactic Entrainment, panel C shows Semantic Entrainment, panel D shows Kinetic Entrainment of the Hands, and panel E shows Kinetic Entrainment of the Head. Each panel is additionally split into Danish (left plot) and Norwegian (right plot). Finally, red plots depict data from the Affiliative Conversations, while blue plots depict data from the Task-Oriented conversations.

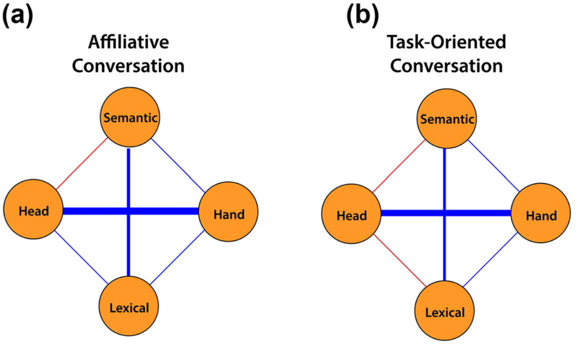

Fig. 4. Correlation networks between linguistic, kinetic, and communicative entrainment variables. Width of the connecting lines indicates the strength of the association between two variables, while color indicates the direction of association with **blue** corresponding to a positive correlation and **red** to a negative correlation. Values are calculated as partial correlations between each pair of nodes, taking all other variables as covariates. Panel A depicts entrainment in the Affiliative conversation, and panel B depicts entrainment in the Task-Oriented conversation.

stepwise approach (respectively, $p = .544$ and $p = .603$). We provide the partial correlation network plots for the two languages separately in the Supplementary Material.

3.4. Network representation

Fig. 4 provides a network representation of the relationships between each of the variables that were found to be significantly associated with one another in the previous analyses. We include both head and hand entrainment variables in order to provide a more complete representation. The figure shows how communicative context primarily affects lexical entrainment.

3.5. Control analyses

Our first control analyses confirmed that kinetic entrainment in our data was higher than what would be expected by chance (i.e., in a shuffled data set), both for the hand entrainment $\chi^2\,(1) = 64.16, p < .001$ and head entrainment $\chi^2\,(1) = 25.26, p < .001$.

In our second set of control analyses, after removing turn pairs with outlying duration differences, we confirmed that all model comparisons showed the same outcome, and all model parameters showed the same direction of effect and approximate magnitude.

4. Discussion

This study assessed how interpersonal entrainment is dynamically coordinated across linguistic and kinetic modalities. We were specifically interested in whether interpersonal entrainment spreads across different linguistic levels (e.g., lexical, syntactic, and semantic) and modalities (i.e., speech and kinetic), as predicted by *interpersonal synchrony/alignment* accounts (Garrod & Pickering, 2009; Ireland & Henderson, 2014; Pickering & Garrod, 2004), or whether there is a divergence (or complementarity) of entrainment levels, as would be predicted by *interpersonal synergy* accounts (Dale et al., 2013; Dideriksen, Christiansen, Tylén et al., 2022; Fusaroli et al., 2014; Fusaroli & Tylén, 2016; Riley, Richardson, Shockley, & Ramenzoni, 2011). To more fully contextualize these findings, we assessed whether this multimodal entrainment would generalize across contextual constraints and—cross-linguistically—across matched Danish and Norwegian conversations.

Similar to previous research by, for instance, Paxton and Dale (2013a), we thus address multimodal coordination across task contexts, but with some important differences: The present analysis uses motion tracking to capture hand and head motion specifically, as opposed to the pixel-differencing method that captures whole body movement, and we additionally utilize multiple levels of linguistic entrainment, rather than the presence or absence of speech. This allows us to approach a more fine-grained level of analysis of communicative behaviors. We hypothesized, based on accounts of multimodal coordination (Rasenberg, Özyürek, & Dingemanse, 2020), that kinetic and linguistic entrainment would correlate with one another, effectively showing that individuals entrain with one another across communicatively relevant modalities. Specifically, based on the interpersonal synergy account, we hypothesized that interpersonal kinetic entrainment would correlate with some levels of linguistic entrainment, but not others. In particular, lower-level "background" similarity, or entrainment, between individuals (e.g., in postural sway or lexical choice), could allow interlocutors to complement each other at higher levels (e.g., semantic and syntactic structure). Note that this would be in contrast to the interpersonal synchrony/alignment accounts that would predict entrainment at all levels of analysis. In line with the interpersonal synergy account, we found that the higher the kinetic entrainment of the head and hands, the lower the semantic entrainment. The pattern is reversed for lexical entrainment, which is also high when kinetic entrainment is high.

Our results show that kinetic entrainment was differentially associated with higher (e.g., semantic) compared to lower (e.g., lexical) levels of linguistic entrainment. This pattern is in line with our hypothesis that individuals would coordinate their entrainment across modalities. Specifically, we suggest that speakers of Danish and Norwegian may entrain kinetically, in terms of movement of the head and hands, when their linguistic contributions are semantically more complementary, and thus less similar to one another. This could manifest as reusing manual gestures (or their kinematic style) or head gestures to signal and maintain mutual understanding with one another (see Holler & Wilkin, 2011b, for a similar discussion).

It should be noted that our study did not assess the semantic content of the gestures in this data set. However, a recent study similarly showed that lexical entrainment in Dutch speakers is correlated with gestural entrainment (defined as two individuals using an iconic gesture to refer to the same referent; Rasenberg et al., 2020). And using the same dynamic time-warping approach as used in the present study, Pouw et al. (2021) found that semantically related gestures are also kinetically similar to one another. These findings are compatible with our observation that lexical and kinetic entrainment are positively correlated with one another. Reusing specific words appears to accompany the repetition of specific movements or movement styles. This multimodal coordination moreover fits well with the interpersonal synergy account (Dale et al., 2013; Fusaroli et al., 2014; Fusaroli & Tylén, 2016), as well as the background-versus-high-level-coordination framing by Louwerse et al. (2012), and suggests that the synergistic coordination among communicative features also applies across communicative modalities. Future research should directly assess the form of gestures, both produced by the hands and head, in order to test whether this kinetic-level entrainment indeed translates to the repetition of similar gesture forms. Another question concerns the functional efficacy of these dynamics, however, as we did not assess task success, we cannot draw conclusions regarding how synergistic coordination of multimodal signals potentially contributes to the relative joint performance of interlocutors.

One interesting, and somewhat unexpected finding, is that syntactic entrainment was not associated with kinetic entrainment in our models (beyond what was already explained by lexical and semantic entrainment). The most likely explanation is that semantic entrainment explains all the variance in kinetic entrainment that would be explained by syntactic entrainment. Indeed, in a control analysis, in which we included syntactic entrainment in our model first, we observed a similar pattern of negative cross-modal entrainment as the semantic model (see Supplementary Materials—OSF files), providing some evidence for this interpretation. This suggests that these two higher-level linguistic variables play similar roles in the overall coordination of linguistic-kinetic entrainment.

4.1. Modulation by communicative context

We additionally found that kinetic entrainment was higher during task-oriented compared to affiliative (i.e., unrelated to an experimental task) conversation. This is directly in line with the findings of Paxton and Dale (2013a, 2013b, 2017; see also Duran et al. 2017) who found that the type of conversation (e.g., affiliative vs. argumentative) affects kinetic entrainment. Previous studies of this same corpora also showed higher lexical entrainment during task-oriented conversation, and more complementarity at semantic and syntactic levels (Dideriksen et al., 2019). The current findings build on these results by suggesting that repetition of not only words, but potentially also gestures or head movements is increased during task-oriented conversation. As suggested by Dideriksen and colleagues (2019), this may be a way to maintain mutual understanding while allowing coordination on higher linguistic levels to be characterized by complementarity. Our findings also resonate with the work by Duran et al. (2019). Here, participants were instructed to actively deceive their interactional partner by espousing a different view than their own, thus leading to an artificial agreement or disagreement on particular topics. However, Duran et al. did not observe the effects of these communicative contexts on linguistic entrainment (Duran et al., 2019). The contrary findings in the present study thus suggest that different communicative contexts, or conversational intentions, can shape (multimodal) entrainment in different ways.

Interestingly, we also found an interaction between semantic entrainment and communicative context. Specifically, task-oriented conversation was associated with a weaker negative correlation between semantic and kinetic alignment. This could suggest that while task-oriented conversation leads to more complementarity at the linguistic level (Dideriksen et al., 2019), kinetic entrainment actually becomes generally more pronounced beyond its association with individual differences in semantic entrainment. One possible explanation is that this is due to kinetic entrainment increasing substantially in the task-oriented condition, while the change in semantic entrainment is less pronounced. Comparing these two forms of entrainment, which are necessarily on different scales of measurement, was outside the scope of the present manuscript. Therefore, future studies will be needed to systematically test this possibility.

4.2. Cross-linguistic replication

Finally, we assessed whether the kinetic-linguistic associations differed between Danish and Norwegian speakers. For both the head and hands, we found no evidence for kinetic entrainment differing between the languages, nor did we find evidence for the kinetic-linguistic association being modulated by language. This is interesting because it means that, despite the difference in sound structure between the two languages likely causing increased linguistic entrainment in Danish (Dideriksen et al., 2022; Trecca et al., 2019), the patterns of kinetic entrainment and relations between linguistic and kinetic entrainment remain similar. It should be noted that while we did not find statistical differences between Danish and Norwegian in our models, the complete network plots (Supplementary Material) suggest that there may be some differences between the two languages. Most prominently, these plots indicate that kinetic entrainment is positively associated with semantic entrainment during affiliative conversation, and negatively associated with one another. This may suggest that Norwegian speakers coordinate between kinetic and high-level linguistic entrainment differently than Danish speakers. However, this is only based on the partial correlation patterns, rather than the more constrained statistical testing with our stepwise mixed effects models. Overall, our statistical results show substantial cross-linguistic generalization of the associations between kinetic and linguistic entrainment. Future studies will be needed to more systematically compare Danish and Norwegian, and to assess whether similar patterns hold across other languages.

4.3. Limitations

The current study has several important limitations. First, we did not include measures of task success, which prevents us from drawing any conclusions about how the coordinative patterns that we observe in this study may affect the outcome of an interaction. Future studies should, therefore, investigate the functional role of these intra- and interpersonal synergies in communicative success (Fusaroli et al., 2014; Riley et al., 2011). Second, we utilized raw movement data rather than annotated communicative signals, such as hand or head gestures. This makes it more difficult to ascertain whether or how specific visuo-motor behaviors are entrained between turns. Nonetheless, our approach allows us to use all of the movement behavior that is present in the interactions, including even subtle shifts in head or hand position, as well as copying of specific communicative movement qualities. Still, future studies should build on these results by relating multimodal coordination patterns to task success, and quantifying how specific visual signals, such as manual gestures, contribute to multimodal, interpersonal coordination. In this work, we investigated broad conversational contexts. However, single conversations are likely to involve shifting activities and goals, such as small-talk, strategy building, and turn-to-turn social actions, such as checking mutual understanding or providing meta-commentary on the discourse. Investigating how coordinative dynamics shift and adjust as local activities change will be crucial for future studies.

4.4. Conclusions

Our study provides evidence that (1) in speakers of Danish and Norwegian, kinetic and linguistic entrainment do not necessarily follow one another, but rather are dynamically coordinated between higher and lower linguistic levels, and (2) task-oriented conversation is associated with higher kinetic entrainment more generally, but also a relative reduction in the semantic-kinetic complementarity (i.e., the negative association). These results are in line with a multimodal, interpersonal synergy account of interactional entrainment as flexibly adapted to the communicative context.

References

Alviar, C., Dale, R., Dewitt, A., & Kello, C. (2020). Multimodal coordination of sound and movement in music and speech. *Discourse Processes*, *57*(8), 682–702. https://doi.org/10.1080/0163853X.2020.1768500

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

Barton, K. (2020). *MuMIn: Multi-Model Inference* (R package version 1.43.17). Retrieved from https://CRAN.R-project.org/package=MuMIn

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

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Condon, W. S., & Ogston, W. D. (1966). Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease*, *143*(4), 338–347.

Csardi, G., & Nepusz, T. (2005). The igraph software package for complex network research. *InterJournal, Complex Systems, 1695(5), 1-9*.

Dale, R., Fusaroli, R., Håkonsson, D. D., Healey, P., & Mønster, D. (2013). Beyond synchrony: Complementarity and asynchrony in joint action. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*, 3.

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. (2022). Quantifying the interplay of conversational devices in building mutual understanding. *Journal of Experimental Psychology: General*, 152(3), 864-889. https://doi.org/10.31234/osf.io/a5r74

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* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/fd8y9

Duffy, K. A., & Chartrand, T. L. (2015). Mimicry: Causes and consequences. *Current Opinion in Behavioral Sciences*, *3*, 112–116. https://doi.org/10.1016/j.cobeha.2015.03.002

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

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

Fay, N., Ellison, T. M., Tylén, K., Fusaroli, R., Walker, B., & Garrod, S. (2018). Applying the cultural ratchet to a social artefact: The cumulative cultural evolution of a language game. *Evolution and Human Behavior*, *39*(3), 300–309. https://doi.org/10.1016/j.evolhumbehav.2018.02.002

Fox, J., Friendly, M., & Weisberg, S. (2012). Hypothesis tests for multivariate linear models using the car package. *R Journal*, *5*(1), 39–52.

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., 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. (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

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292–304. https://doi.org/10.1111/j.1756-8765.2009.01020.x

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in *R*: The **dtw** package. *Journal of Statistical Software*, *31*(7), 1-24. https://doi.org/10.18637/jss.v031.i07

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

Holler, J., & Wilkin, K. (2011a). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior*, *35*(2), 133–153. https://doi.org/10.1007/s10919-011-0105-6

Holler, J., & Wilkin, K. (2011b). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics*, *43*(14), 3522–3536. https://doi.org/10.1016/j.pragma.2011.08.002

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

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

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, *34*(6), 419–426. https://doi.org/10.1016/j.evolhumbehav.2013.08.001

Mueen, A., & Keogh, E. (2016). Extracting optimal performance from dynamic time warping. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. https://doi.org/10.1145/2939672.2945383

Ostrand, R., & Chodroff, E. (2021). It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. *Journal of Phonetics*, *88*, 101074. https://doi.org/10.1016/j.wocn.2021.101074

Paxton, A., Abney, D. H., Kello, C. T., & Dale, R. (2014). Network analysis of multimodal, multiscale coordination in dyadic problem solving. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*(36), 7.

Paxton, A., & Dale, R. (2013a). . *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*(35), 1121-1126.

Paxton, A., & Dale, R. (2013b). Argument disrupts interpersonal synchrony. *Quarterly Journal of Experimental Psychology*, *66*(11), 2092–2102. https://doi.org/10.1080/17470218.2013.853089

Paxton, A., & Dale, R. (2017). Interpersonal movement synchrony responds to high- and low-level conversational constraints. *Frontiers in Psychology*, *8*, 1135. https://doi.org/10.3389/fpsyg.2017.01135

Paxton, A., Dale, R., & Richardson, D. C. (2016). Social coordination of verbal and nonverbal behaviours. In Passos, P., Davids, K., & Chow, J.Y. (Eds.) *Interpersonal coordination and performance in social systems* (pp. 277–292). Routledge.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. Scopus.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525x12001495

Pouw, W., de Wit, J., Bögels, S., Rasenberg, M., Milivojevic, B., & Ozyurek, A. (2021). Semantically related gestures move alike: Towards a distributional semantics of gesture kinematics. In V. G. Duffy (Ed.), *Digital human modeling and applications in health, safety, ergonomics and risk management. Human body, motion and behavior* (Vol. *12777*, pp. 269–287). Springer International Publishing. https://doi.org/10.1007/978-3-030-77817-0_20

Pouw, W., & Dixon, J. A. (2019). *Gesture networks: Introducing dynamic time warping and network analysis for the kinematic study of gesture ensembles*. https://doi.org/10.31234/osf.io/hbnt2

Pouw, W., Proksch, S., Drijvers, L., Gamba, M., Holler, J., Kello, C., Schaefer, R. S., & Wiggins, G. A. (2021). Multilevel rhythms in multimodal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1835), 20200334. https://doi.org/10.1098/rstb.2020.0334

Qiao, S., Wang, Y., & Li, J. (2017). Real-time human gesture grading based on OpenPose. 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). https://doi.org/10.1109/CISP-BMEI.2017.8301910

Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*, *79*(3), 284–295. https://doi.org/10.1037/a0023419

Rasenberg, M., Dingemanse, M., & Özyürek, A. (2020). Lexical and gestural alignment in interaction and the emergence of novel shared symbols. Evolution of Language Conferences.

Rasenberg, M., Özyürek, A., Bögels, S., & Dingemanse, M. (2022). The primacy of multimodal alignment in converging on shared symbols for novel referents. *Discourse Processes*, *59*(3), 209–236. https://doi.org/10.1080/0163853X.2021.1992235

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in multimodal interaction: An integrative framework. *Cognitive Science*, *44*(11), e12911. https://doi.org/10.1111/cogs.12911

Revelle, W. (2020). *psych: Procedures for Personality and Psychological Research (2.0.12) [R]*. Northwestern University.

Riley, M., Richardson, M., Shockley, K., & Ramenzoni, V. (2011). Interpersonal synergies. *Frontiers in Psychology*, *2*, 38.

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

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. https://doi.org/10.1037/0096-1523.29.2.326

Trecca, F., McCauley, S. M., Andersen, S. R., Bleses, D., Basbøll, H., Højen, A., Madsen, T. O., Ribu, I. S. B., & Christiansen, M. H. (2019). Segmentation of highly vocalic speech via statistical learning: Initial results from Danish, Norwegian, and English: Segmenting vocalic speech in statistical learning. *Language Learning*, *69*(1), 143–176. https://doi.org/10.1111/lang.12325

Tsuchiya, A., Ora, H., Hao, Q., Ono, Y., Sato, H., Kameda, K., & Miyake, Y. (2020). Body movement synchrony predicts degrees of information exchange in a natural conversation. *Frontiers in Psychology*, *11*, 817. https://doi.org/10.3389/fpsyg.2020.00817

Tylén, K., Fusaroli, R., Smith, P., & Arnoldi, J. (2016). The social route to abstraction. *Paper Presented at the Cognitive Science*, *2016*, 1.

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

Yeomans, M., Schweitzer, M. E., & Brooks, A. W. (2022). The Conversational Circumplex: Identifying, prioritizing, and pursuing informational and relational motives in conversation. *Current Opinion in Psychology*, *44*, 293–302. https://doi.org/10.1016/j.copsyc.2021.10.001

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary Figure 1. Partial correlation networks across conversation contexts and languages.

Supplementary Figure 2. Context-specific modulation of kinetic-semantic entrainment association.

Supplementary Table 1.