

# Cooperation in Sound and Motion: Complexity Matching in Collaborative Interaction

Drew H. Abney<sup>1</sup>, Alexandra Paxton<sup>2, 3</sup>, Rick Dale<sup>4</sup>, and Christopher T. Kello<sup>5</sup>

<sup>1</sup> Department of Psychology, University of Georgia

<sup>2</sup> Department of Psychological Sciences, University of Connecticut

<sup>3</sup> Center for the Ecological Study of Perception and Action, University of Connecticut

<sup>4</sup> Department of Communication, University of California, Los Angeles

<sup>5</sup> Cognitive and Information Sciences, University of California, Merced

Complex behaviors are layered with processes across timescales that must be coordinated with each other to accomplish cooperative goals. Complexity matching is the coordination of nested layers of behaviors across individuals. We hypothesize that complexity matching extends across individuals and their respective layers of processes when cooperating in joint tasks. We measured coordination in a joint tower building task through the layers of sound and movement patterns produced by partners and found that partners built higher towers when their sound patterns fell into more similar relations with each other across timescales, as measured by complexity matching. Our findings shed light on the function of complexity matching and lead to new hypotheses about multiscale coordination and communication. We discuss how complexity matching encompasses flexible and complementary dynamics between partners that support complex acts of human coordination.

*Keywords:* cooperative behavior, nested temporal structure, complexity matching, normalized wavelet variance

Humans and other intelligent species work together with their conspecifics to accomplish complex tasks and goals. Such forms of cooperation can occur through primarily physical interactions, as when two or more individuals work together to move heavy objects (Isenhower et al., 2010) or herd animals (Nalepka et al., 2017). Cooperation can also occur through the exchange of information via a system of communication, as when individuals search an area together by signaling information about locations to each other or when individuals talk to each other and coordinate tasks in large, extended projects. In all these cases, cooperation requires the timing of behaviors among two or more individuals in the service of some goal, and the nature of cooperation is reflected in the temporal patterns that arise through behavioral interactions.

The present study challenges an important theoretical assumption about the coordination among two individuals that can be emergent from and functional to collaboration. To be clear about

terminology, we share Bedwell et al.'s (2012) description of cooperation with respect to collaboration as “an attitude or predisposition held by the involved parties to be concerned about the overall collaborative goal rather than their own individual goal” (p. 136). That is, collaboration is a superordinate process that includes both goal-oriented cooperation and temporal coordination. Moreover, our use of the term coordination should be considered as a part of cooperation, not an entirely separate construct.

Most of the prior research suggests cooperation is a *local* matter (but see Mayo & Gordon, 2020), such as through synchronized body movements or the immediate “push and pull” between interacting persons. But such an account is incomplete, because it focuses on the moment-to-moment temporal matching of behaviors and misses the important structure of the wider interaction episode itself. It may also be critical for participants working together to match not just individual acts of behavior, but also the temporal structure of their behavior across a range of timescales, from seconds to the minutes that make up the “rhythm” of interaction. In this paper, we show that matching in temporal structure may indeed be an important ingredient of successful interaction. The theoretical implication is that cooperation patterns among humans are not just local affairs but also span more subtle behavioral contingencies over time.

In the present study, we employ new measures of cooperation based on temporal patterns in behavioral interactions. Our aim is to test whether the degree of success in a cooperative task can be measured in the coordination of movements and vocalizations produced by the people cooperating. Thus, we seek to quantify coordination in terms of temporal patterning—and correlate it with

---

Drew H. Abney  <https://orcid.org/0000-0003-3040-2927>

Data and code for analyses are available at the Open Science Framework: <https://osf.io/53e7m/>.

We thank Evelyn Alvarez, Kelli Eggleston, Jamie Faria, Alexis Luna, and Keith Wilson for assistance in data collection. We also thank Ramesh Balasubramaniam and Anne Warlaumont for helpful discussions. A portion of the dataset used in the current article was published in Abney et al. (2015).

Correspondence concerning this article should be addressed to Drew H. Abney, Department of Psychology, University of Georgia, 125 Baldwin Street, Athens, GA 30602, United States. Email: [Drew.abney@uga.edu](mailto:Drew.abney@uga.edu)

an assay of performance—in a naturalistic cooperative task. We chose a task that requires cooperation to build a physical structure (Abney et al., 2015). The task is designed to engage pairs of participants in a collaborative effort to build the highest tower they can in 15 min, using only uncooked spaghetti and marshmallows. Each pair must work together, coordinating their movements and verbal interactions to build the tower. The task is relatively unconstrained in the sequence and timing of steps and interactions that any given pair of individuals might perform. The open-ended nature of the task allows for complex interaction dynamics to emerge as partners explore and exploit different techniques and plans for building the tower.

In studies of interpersonal coordination, many tasks have been designed to elicit *synchronization* in behaviors, as a simple form of cooperation that can be more readily measured and analyzed (Bernieri, 1988; Fusaroli et al., 2012; Fusaroli et al., 2016; Hove & Risen, 2009; Louwerse et al., 2012; Shockley et al., 2003). Synchronization is one example of the more general construct of *co-occurrence*, which has proven to be fruitful in the design and analysis of many experiments on interpersonal coordination (Oullier et al., 2008; Richardson et al., 2007; Schmidt et al., 2012; Schmidt et al., 2014; Shockley et al., 2007). For instance, studies have shown that the co-occurrence of movements happens spontaneously between interlocutors and is correlated with rapport (Hove & Risen, 2009) and mental connectedness (Raffard et al., 2015).

Synchronization and co-occurrence continue to be useful constructs for studies of interpersonal coordination, but they are unlikely to tell the whole story about all forms of interaction—especially when considering tower building and other cooperative tasks that yield complex, complementary dynamics. In our case, the task of cooperating to build a tower does not require synchronized identical movements or co-occurring speech acts. Instead, successful performance is more likely to depend on complementary movements to coordinate subtasks and turn-taking to communicate information.

Task demands can influence or constrain the types of coordination that are possible, probable, and fruitful. As discussed above, some experimental tasks like rhythmic tapping (Hove & Risen, 2009) were designed to elicit synchronous movements whereas other tasks were designed to be more open-ended which could elicit more complex coordination patterns. For example, Louwerse et al. (2012) designed a map task that required an instruction giver and an instruction follower to communicate with each other to complete a map puzzle. Louwerse et al. (2012) observed that both synchronization and leader-follower dynamics emerged across a number of modalities (e.g., many communicative behaviors of the instruction giver preceded the same behaviors of the instruction follower). The plurality of coordination patterns observed in Louwerse et al. (2012)—and, specifically, the complex leader-follower dynamics—suggests that complementary coordinative dynamics can also emerge in naturalistic tasks. In another example, Fusaroli et al. (2016) statistically compared the contribution of different types of coordination patterns such as linguistic alignment and communicative complementarity (referred to as interpersonal synergies) during dialog to a measure of collaborative performance on a joint perceptual decision-making task. Although there was evidence for the presence of both types of coordination patterns, the results suggested the complementary patterns were the best statis-

tical predictors of task performance. Even with this short list of examples from previous work, it is clear that different tasks can afford different types of coordination patterns among dyads and that more open-ended tasks can afford more possible coordination patterns to emerge.

In the present study, we investigate new measures of cooperation based on *nested temporal structures* that characterize many different types of complex behaviors including speech and bodily movements during conversations (Abney, Paxton, et al., 2014; Abney, Warlaumont, et al., 2014; Abney et al., 2017; Falk & Kello, 2017; Kello et al., 2017; Ramirez-Aristizabal et al., 2018; Schloesser et al., 2019). In the case of cooperative tower-building, the concept of nested temporal structure is illustrated by conceptualizing different phases of construction at different timescales, with each shorter timescale nested within a longer one. For instance, at a long timescale, the task might unfold in several phases, with a planning phase, followed by a trial-and-error phase, followed by one or two routinized building phases. Nested within each of these long timescale phases, there will be shorter phases of subtask planning, level setting, establishment of mutual understanding, and so on. Further nested phases will be even shorter, perhaps relating to individual building actions and utterances.

Activities across timescales will be associated with nested periods of movement and speech activity. Rather than identifying and coding phases in terms of the semantics of movements and speech, we hypothesize that the efficacy of cooperation can be measured in the overall temporal nesting of movement and speech activity. Instead of identifying individual phases, we need only measure the statistics of clustering activity in motion and acoustic energy signals; the signatures of coordination should be recoverable within the *nested dynamics* of these observable behaviors.

Our hypothesis does not entail synchronization or any other phase-based relationship in the dynamics of cooperation. It is grounded in theories of how nested temporal structures emerge from complex systems and converge during interaction (West et al., 2008) and empirical results from prior studies. Moreover, we do not identify specific timescale ranges in which cooperation emerges (e.g., a range of short timescales or a range of long timescales). Instead, we zoom out to measure how *clusters* of activity—here, vocalizations or movements—across timescales converge during interaction.

### Nested Temporal Structures Underlie Human Interaction

Nested temporal structure has been found and quantified in various types of speech signals and in movements during conversations (Abney, Paxton, et al., 2014; Kello et al., 2017). In these studies, temporal events were measured as peaks or troughs in the amplitudes of speech and motion signals, and temporal clustering of events was quantified across a range of timescales using normalized wavelet variance, known as Allan Factor (AF) analysis (Allan, 1966). The relationship between timescale and AF variance was quantified as a measure of nested temporal structure (see Method) and shown to distinguish between monologs versus dialogs (Kello et al., 2017), between affiliative versus argumentative conversations (Abney, Paxton, et al., 2014), and even bilingual conversations that switched between English and Spanish (Schneider et al., 2020).

In Abney, Paxton, et al. (2014), the nested temporal structure of vocalizations, as measured using the AF analysis, was steeper during argumentative conversations relative to affiliative conversations. A closer inspection of the aggregated AF slopes suggested that there was very little differences in the clustering dynamics at short timescales across conversational contexts, but that there was more clustering variability during the argumentative conversations at longer timescales, where the context of the conversation would be more impactful. The differences in the longer timescales impacted the overall temporal relationship of clustering activity of the vocalizations across all of the timescales. If there were no temporal structure in events across timescales, there would be no relationship between timescale and AF variance. In other words, we would expect no linear (or any other higher-order) relationship between timescale and AF variance.

Nested temporal structure has been found to converge between interlocutors having an affiliative conversation but *not* for those having an argumentative conversation (Abney, Paxton, et al., 2014). We termed this *complexity matching* because nested clustering in speech reflects their complexity of structure and because theories of complex systems predict this convergence to occur when similar complex systems interact (West et al., 2008). Complexity matching is a convergence in the *overall* statistics of temporal structure rather than a measure of co-occurrence in *specific* measurements (i.e., synchronization). Moreover, synchronization of events occurs at a particular timescale, whereas complexity matching as a measure of multiscale convergence takes into account the similarity of clustering of events that occur at different timescales of the interaction. The complexity matching in affiliative conversations suggests that complexity matching may reflect the establishment of common ground and shared points of view between interlocutors (Abney, Paxton, et al., 2014)—a hypothesis that could be effectively tested in explicitly cooperative interactions.

From a general perspective, complexity matching and synchrony measure the convergence of behaviors at different ranges of timescales. Complexity matching is a summary statistic of the convergence of behavior (e.g., vocalizations or movements) *across multiple nested* timescales: stronger complexity matching means more similarity in the clustering of behavioral events across the measured timescales between two interacting agents. Synchronization is a summary statistic of the convergence of behavior at *one, instantaneous timescale*: stronger synchrony means more instances of a behavior being produced at the same time by two interacting agents.

### The Present Study

Previous work has not yet explored the connection between complexity matching and intentionally cooperative language interactions (for an example in perceptual-motor cooperation, see Schloesser et al., 2019). However, existing studies suggest a number of hypotheses: The convergence of nested temporal structure in individual behaviors will be greater than random pairs of individuals (H1), and the efficacy of cooperation between two people will be reflected in the extent to which those patterns converge (H2). Finally, given the open-ended nature of interactive context of the dyadic task, we would also expect that the functional relationship between task performance and complexity matching will not be

explained by other factors, including the nested temporal structure of individual behaviors, work rate, and—importantly—patterns of only local convergence. We test these hypotheses in the following experiment.

For H1, the alternative hypothesis is that the extent to which the nested temporal structure of behavioral patterns converge does not differ from the overall convergence of nested temporal structure of behavioral patterns from random pairings of individuals. For H2, the alternative hypothesis is that the extent to which the nested temporal structure of behavioral patterns converge is not associated with performance in cooperative interaction and can be explained by other factors related to individual behavioral complexity, building properties, and local coordination patterns. Both alternative hypotheses would suggest that the nested structure of behavioral patterns is not consequential for cooperative interactions.

## Method

### Corpus

The analyses presented herein were conducted using an expanded corpus of audio and video recordings reported previously (Abney et al., 2015). We outline the relevant parts of this study below. Additional information is available in the original article. This study was approved by the IRB at the University of California, Merced (030311–347), titled *Body and Voice Synchrony During Interaction*.

### Participants

Seventy-four (74) undergraduate students from the University of California, Merced, participated in 37 dyads in return for extra course credit. Participants individually signed up using an anonymous online subject pool system and could not see their partner's identity before arriving at the study location. Sample size was based on previously published data with a medium effect size ( $\sim 0.30$ ) and a sample of  $n = 27$  (Abney, Paxton, et al., 2014) when testing for within-dyad differences in complexity matching scores between empirical and surrogate pairings.

### Materials and Procedure

Participants were asked to sit in one of two stationary chairs near a square table (76.2 cm by 76.2 cm W by 71.1 cm H). Seating arrangement was participant-initiated, and experimenters were careful to not provide any explicit direction toward either of the two chairs. The two chairs and table were oriented such that the chairs were placed adjacent to each other, with the table rotated 45 degrees in line of sight of the camcorder (Canon Vixia HF M31 HD camcorder with MPEG4-AVC codec). The experimental room was equipped with fluorescent lighting and window shades to keep out external light.

Once seated and oriented with a comfortable sitting posture, participants were outfitted with Shure Beta 54 supercardioid microphone headsets. Participants were then instructed to construct the tallest tower structure possible within 15 mins using only the materials provided: one box ( $\sim 10$  oz) of large marshmallows and one box ( $\sim 1$  lb) of raw spaghetti. To increase the need for

coordination and communication, the participant seated on the right was only allowed to handle the marshmallows, and the participant on the left was only allowed to handle the spaghetti (both positions given relative to the video frame). Participants were instructed to use only whole spaghetti strands and marshmallows and to remove any pieces of material that broke during the construction. Experimenters monitored the dyads throughout the construction phase and corrected participants if they strayed from the task rules.

Experimenters answered any questions before the task began, and they provided 5- and 1-min warnings before their time expired. At the end of each 15-min building session, experimenters recorded the height of the tower and how much spaghetti and marshmallows were used.<sup>1</sup> See Figure 1 for an example picture of the experimental setup.

### Vocalization and Movement Data Coding

Audio and video files were truncated to contain only interactions occurring during the 15-min construction phase.

#### Vocalizations

The left- and right-side members of each pair were recorded on the left and right stereo channels using separate microphones for each member. Each channel was saved to a separate audio file (.wav) for subsequent audio cleaning and coding. A trained research assistant listened to each recording and omitted crosstalk and noise from noncommunicative breaths. Criteria for omitting crosstalk and noise included listening for the nontarget dyad member's voice and visually inspecting the acoustic amplitude visualization provided by Audacity. Given the quality of the recordings, it was very clear which dyad member was the target and which member was the nontarget for each .wav file.

Each file was then analyzed using the Audacity "sound finder" to locate acoustic onset/offset intervals. The threshold of acoustic intensity was set at  $-30$ dB for all audio files (Abney, Paxton, et al., 2014). Due to low acoustic intensity values for a majority of

recordings from the left channel, acoustic intensity was amplified by 6dB for all audio files from the left channel to make average acoustic intensity approximately equal across the left and right channels. Acoustic intensity was not consistent across all audio files. Therefore, where appropriate, acoustic intensity was amplified by 6dB for an entire .wav file.

Vocalization onsets were used in subsequent analyses. Binary spike trains of vocalization events were calculated by using the onsets estimated by the Audacity "sound finder" function. Vocalization onset events were coded as "1," and other states were coded as "0."

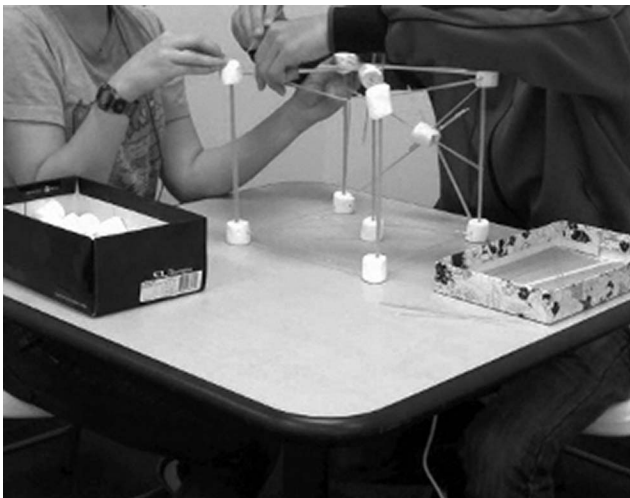
#### Movements

The truncated video files were analyzed using a frame-differencing method (FDM) to obtain time series of standardized movement indices computed from pixel-to-pixel changes from frame-to-frame converted .jpg images (Paxton & Dale, 2013a, 2013b; Ramseyer & Tschacher, 2011; Schmidt et al., 2012) for the left and right side of the video, respectively. To be clear, this analysis provides two separate time series of standardized movements: one time series for the member on the left side and one time series for the member on the right side. The FDM provided an objective measure of overall body movement. Higher FDM values indicated higher amounts of overall movement for each participant.

Binary spike trains of movement events were calculated by creating a threshold of movement indices. A movement event was defined as a sample in the FDM movement series that exceeded 2 standard deviations above the mean movement index. Movement events defined by this threshold were coded as "1," and other states were coded as "0" (see Figure 2). Note that because the FDM movement series were standardized  $z$ -scores, a threshold of 2 standard deviations leads to identifying the largest 5% pixel change frames in the time series. Notably, the motivation of this analysis was to identify the largest gross movement events during the 15-min construction phase and *not* to identify specific actions like goal-directed actions or non-goal-directed actions. This motivation parallels the protocol for generating binary spike trains of vocalization events, where the loudest auditory events were identified (rather than identifying semantically meaningful events).

To build an intuition about the movement events being identified by the threshold parameters, reference the bottom panel in Figure 2. Specifically, around frame number 5600, there is a high density of movement events. At one timescale, this is evidence that there are a number of large movements occurring in a short period of time. At a longer timescale, this is evidence that there is a bout of movement activity. Notably, we do interpret any bouts of movement activity differently (e.g., through qualitative event coding), and this level of analysis does not include any information about action duration or goal.

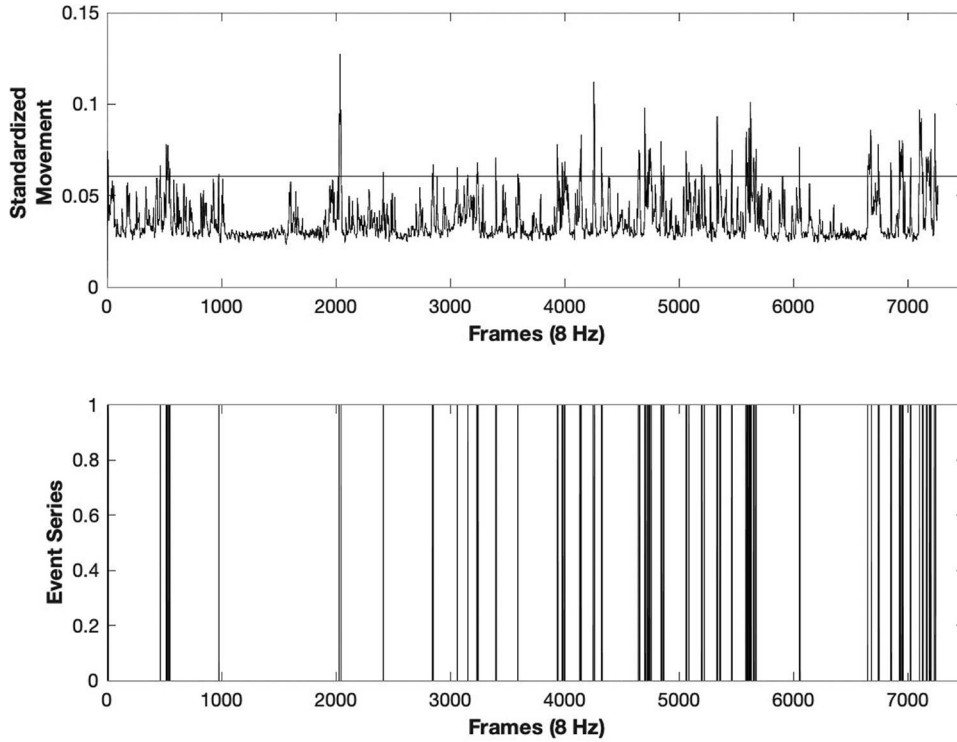
**Figure 1**  
*Example of the Experimental Setup*



<sup>1</sup> Note that height/weight ratio was used in Abney et al. (2015) to index building efficiency. The decision to use height as the main dependent in the current paper reflects a more accurate index of task performance as per the instructions provided to the participants.



**Figure 2**  
*Example Movement Time Series and Subsequent Event Series*



*Note.* (Top) An example time series of standardized movements estimated by the FDM (Paxton & Dale, 2013b) and after a 2nd-order Butterworth filter has been applied. Horizontal line represents an index of 2 standard deviations above the average movement. (Bottom) An example event series of body movement onsets. Each vertical line corresponds to a movement onset identified using the thresholding protocol.

### Deriving Nested Temporal Structure With AF Analyses

AF analysis (Allan, 1966) was used to estimate the multiscale clustering of vocalization and movement dynamics for each interlocutor's vocalizations and body movement time series. The AF analysis estimated the variance of vocalization or body movement events (i.e., onsets of vocalization or body movement) at particular timescales and computed correlation estimates,  $\alpha$ , across those multiple time scales.

The AF analysis is a point process analysis that inputs a binary spike train of events (1) and nonevents (0). A point process is segmented into  $M$  adjacent windows of size  $T$ , and the number of events  $N_j$  is counted within each window indexed by  $j = 1$  to  $M$ . The differences in event counts between adjacent windows of a given size  $T$  is computed as  $d(T) = N_{j+1}(T) - N_j(T)$ .  $d(T)$  values are computed for each window size  $T$ . The AF variance  $A(T)$  for a given timescale  $T$  is the expected value of the squared differences, normalized by mean counts of events per window,

$$A(T) = \frac{\langle d(T)^2 \rangle}{2\langle N(T) \rangle}$$

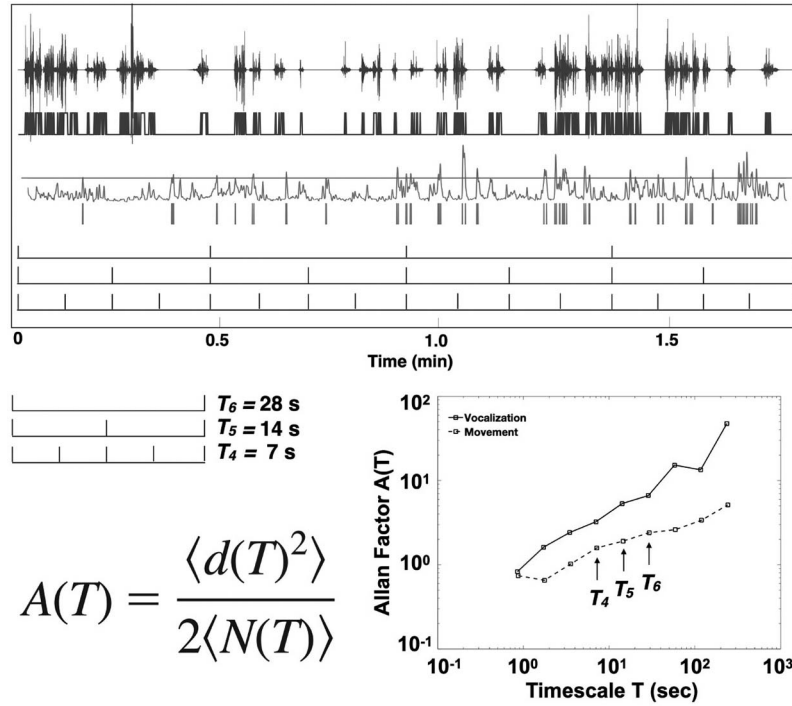
Poisson processes yield  $A(T) \sim 1$  for all  $T$ , whereas power law clustering yields  $A(T) \sim (T/T_1)^\alpha$ , where  $T_1$  is the smallest time scale considered and  $\alpha$  is the exponent of the scaling relation. Point

processes with  $\alpha \sim 0$  are Poisson distributed, whereas  $\alpha > 0$  indicates power-law clustering across the measurable range of timescales (Turner et al., 1997).

$A(T)$  was computed for vocalization and body movement event time series from each member of each dyad. Each participant's time series was  $\sim 15$  min, and time windows varied as a power of 2,  $T = 2^t$ , where  $t$  ranged from 2 to 10 samples. The resulting timescales ranged from  $\sim 900$ ms to  $\sim 220$ s. Smaller timescales were excluded because they are heavily affected by measurement error and sample rates, and larger timescales could not be reliably estimated given the length of the time series (see Figure 3). Allan Factor (AF) estimates were computed by estimating the linear fit between  $\log(A(T))$  and  $\log(T)$ .

Certain ranges of timescales in the AF analysis approximate important phases or timescales of dyadic construction task. For example, for vocalizations, the shortest timescale ( $\sim 1$ s) approximates phonetic and lexical levels during the production of a vocal utterance. For movements, the middle range of timescales (7–14s; see Figure 3) include short bouts of construction, approximating the duration of large movements needed to pick up and place building materials on the tower structure. For vocalizations, the longest timescales of  $\sim 2.5$  min would include longer bouts of conversation, like the planning of a particular segment of the tower or discussion about how best to keep the tower from falling.

**Figure 3**  
 Illustration of Allan Factor Analysis Using an Example Waveform (Top Signal) and Movement Time Series (Bottom Signal)



*Note.* The waveform and movement time series are accompanied by corresponding spikes representing events, along with the time scale (in number of samples  $T_n$  for scale). The AF variance equation is also shown, along with the corresponding  $A(T)$  function for the waveform and movement time series.

To compute a measure of complexity matching of a dyad, we used an AF similarity metric introduced in Abney, Paxton, et al. (2014). The AF similarity metric was the summed absolute difference between two AF functions, with a negative log transformation:

$$D_{a,b} = -\sum_T \log |A(T_a) - A(T_b)| + 10$$

For each dyad, a complexity matching metric was computed for vocalizations and body movements. To increase interpretability, we added a constant ( $c = 10$ ) to each score so that all scores were above 0. A linear translation of the scores is merely for interpretation and does not affect statistical outcomes of the models. The log transformation takes into account the scaling law over  $T$ , which is used to estimate the AF function. The negative sign makes larger values of  $D_{a,b}$  correspond to greater complexity matching. Figure 4 provides a visual depiction of what types of similarities and differences in the shape of the AF functions would lead to either having high or low complexity matching.

## Results

Data and code for analyses are available at the Open Science Framework (<https://osf.io/53e7m>). The average height of the spaghetti-marshmallow towers was 46.6 cm ( $SD = 15.40$ ). AF estimates for body movements were higher ( $M = 0.36$ ,  $SD = 0.25$ ,

95% CI [0.30, 0.41]) relative to AF estimates for vocalizations ( $M = 0.28$ ,  $SD = 0.14$ , 95% CI [0.25, 0.32]),  $t(146) = 2.24$ ,  $p = .027$ , 95% CI for the difference between  $M$ s = [0.008, -0.14],  $d = .33$  (see Figure 5). Differences in AF estimates across modalities can be influenced by a number of factors. We expand upon these factors in the Discussion section.

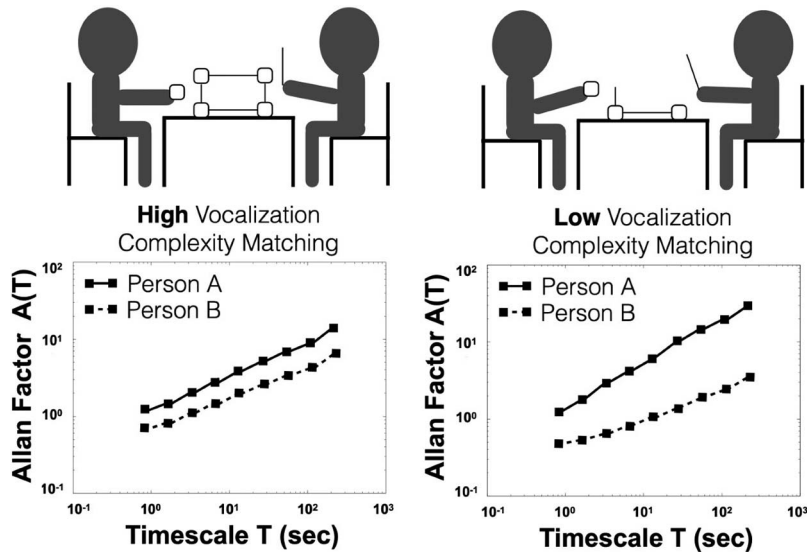
## Testing H1: Complexity Matching in Vocalizations and Body Movement

To determine if complexity matching occurred beyond random pairings of interlocutors, average complexity matching scores were computed for each participant and a random surrogate pairing of a participant who was not their partner. We conducted a linear mixed-effects model (Baayen et al., 2008) to examine the effects of pairing type (empirical vs. surrogate), modality (vocalization vs. body movement), and the interaction term on complexity matching scores. We included dyad membership as a random slope with the maximally permitted random intercept to achieve model convergence. The model was built in R (Team, 2016) using the lme4 library (Bates et al., 2014).

Overall, the full model was significantly better than a more parsimonious model that included only the random effect structure, likelihood ratio test:  $\chi^2(2) = 58.56$ ,  $p < .001$ . As predicted, there was a significant effect of pairing type ( $b = -2.57$ ,  $\beta = -0.42$ ,

**Figure 4**

*Toy Illustration of What the AF Functions Look Like and How the Similarities and Differences in the Shapes of the Functions Would Lead to Different Degrees of Complexity Matching*



$SE = 0.71$ , 95% CI  $[-3.96, -1.17]$ ,  $d = 0.37$ ): Empirical complexity matching scores ( $M = 12.92$ ,  $SD = 6.56$ , 95% CI  $[11.42, 14.41]$ ) were greater than surrogate complexity matching scores ( $M = 10.34$ ,  $SD = 5.24$ , 95% CI  $[9.15, 11.54]$ ). Modality also significantly predicted complexity matching scores ( $b = 5.61$ ,  $\beta = 0.93$ ,  $SE = 0.71$ , 95% CI  $[4.21, 6.99]$ ,  $d = 0.83$ ), such that body movement complexity matching scores ( $M = 14.43$ ,  $SD = 5.74$ , 95% CI  $[13.12, 15.74]$ ) were greater than vocalization complexity matching scores ( $M = 8.83$ ,  $SD = 4.99$ , 95% CI  $[7.69, 9.97]$ ); see Figure 6). These results suggest that complexity matching across multiple modalities occur above and beyond random pairings of individuals and that the nested temporal structure of movements

among pairs in a dyad converge more across multiple timescales than the nested temporal structure of vocalizations.

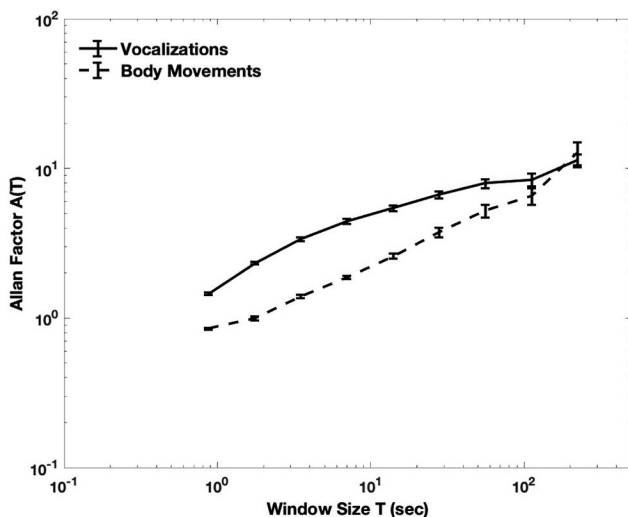
## Testing H2: Connecting Complexity Matching to Task Performance

To determine whether complexity matching predicted task performance, we conducted linear regression models for complexity matching scores from each modality predicting tower height. Models including complexity matching estimates were significantly better than a more parsimonious model that included only the intercept (null vs. null + vocalization complexity matching:  $\chi^2(1) = 251.62$ ,  $p = .004$ ; null vs. null + movement complexity matching:  $\chi^2(1) = 198.7$ ,  $p = .013$ ). Moreover, including an interaction term did not improve model fit (addition model vs. interaction model:  $\chi^2(1) = 5.73$ ,  $p = .658$ ), so we therefore included vocalization and body movement complexity matching scores as only additive predictors into the model. For all subsequent statistical tests, we report standardized beta weights, a measure of effect size (Keith, 2014).

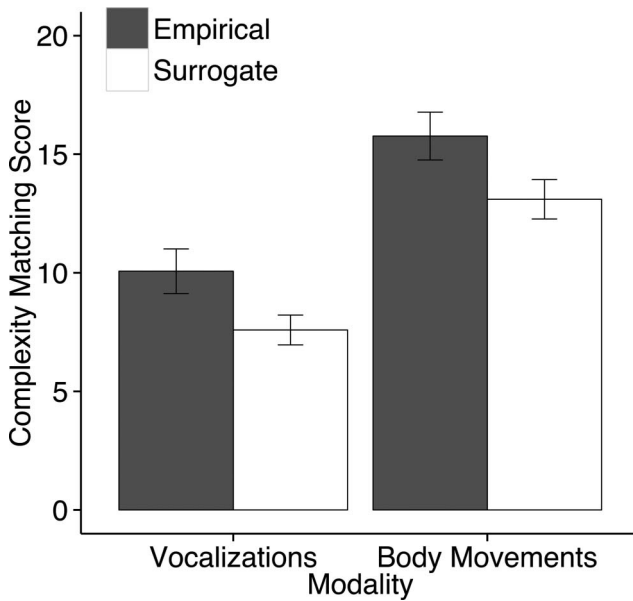
Vocalization complexity matching significantly predicted tower height ( $b = 0.38$ ,  $\beta = 0.36$ ,  $SE = 0.16$ , 95% CI  $[0.05, 0.71]$ ,  $p = .025$ ). Although it did not reach statistical significance, body movement complexity matching trended towards predicting tower height ( $b = 0.29$ ,  $\beta = 0.29$ ,  $SE = 0.15$ , 95% CI  $[-0.02, 0.59]$ ,  $p = .065$ ). These results suggest that greater complexity matching predicted better performance on the tower-building task differently by modality.

We also conducted two additional linear models of complexity matching scores for each modality predicting tower height variable after residualizing averaged AF estimates and an estimate of the work rate of the dyad. These analyses allowed us to determine whether the main result of complexity matching predicting tower height was impacted by the overall amount of complexity of the

**Figure 5**  
*Mean Allan Factor Functions for Vocalizations (Solid) and Body Movements (Dashed), With Standard Error Bars*



**Figure 6**  
Average Complexity Matching Scores for Vocalizations (Left) and Body Movements (Right), Separately for Empirical (Black) and Random (White) Surrogate Pairings



Note. Error bars reflect standard error.

dyad (as measured by the averaged AF estimates). For example, if the relationship between vocalization complexity matching and performance could be explained by the overall amount of vocalization complexity of the dyad, we would not be justified in claiming that the multitimescale convergence of clustering activity between members of a dyad facilitates better performance. Moreover, it is possible that speed of the dyad in building a tower

structure could explain not just task performance but also the extent to which a dyad's vocalizations or movements converge across multiple timescales.

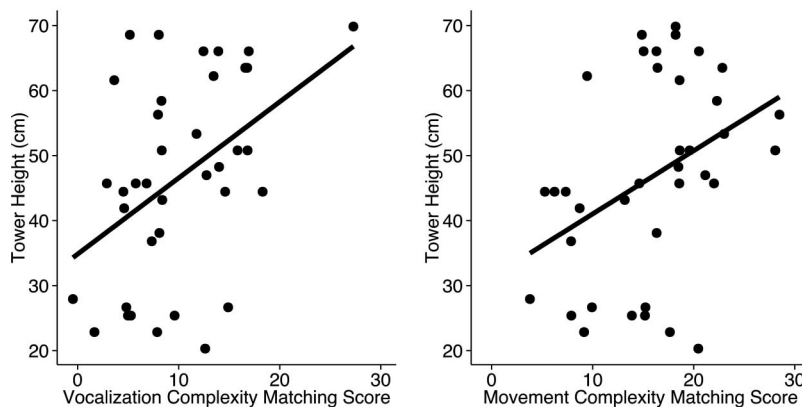
With the assumption that faster dyads would be able to utilize more building materials, we constructed an index of work rate which was the total amount of materials (both marshmallows and spaghetti pieces) that were used to build the final tower structure at the end of the construction phase. Vocalization complexity matching significantly predicted residual tower height ( $b = 0.45$ ,  $\beta = 0.43$ ,  $SE = 0.19$ , 95% CI [0.13, 0.78],  $p = .007$ ). Body movement complexity matching did not reliably predict residual tower height ( $b = 0.06$ ,  $\beta = 0.07$ ,  $SE = 0.15$ , 95% CI [-0.25, 0.36],  $p = .698$ ; see Figure 7). The results from these two models provide further evidence that vocalization complexity matching predicts tower height over and above the overall vocalization complexity and work rate of the dyad.

### Controlling for Synchronization or Behavioral Matching

Overall complexity estimates influenced the relation between body movement complexity matching and task performance, but no such effect was found for vocalization complexity matching. Despite this independence, the relationship between vocalization complexity matching and tower building performance may have been influenced by local coordination patterns—that is, vocal behavioral matching—at shorter temporal scales. Our next analyses directly tested whether vocal behavioral matching could account for these effects.

The most directly related measure of vocal behavioral matching would be to cross-correlate the same event series (spike trains) that were used to compute AF estimates and subsequent complexity matching scores. However, because spike trains are instantaneous and do not include the durational information important for cross-correlation, we computed a square signal for each dyad member by

**Figure 7**  
Scatter Plots of Vocalization (Left) and Body Movement (Right) Complexity Matching Scores and Tower Height (cm)



Note. Each point corresponds to a dyad. Solid lines are linear regression fit. The relationship between vocalization complexity matching and tower height is still significant when omitting the highest vocalization complexity matching score,  $b = 0.44$ ,  $\beta = 0.37$ ,  $SE = 0.19$ , 95% CI [0.05, 0.82],  $p = .027$ .



assigning a “1” for event durations (acoustic energy between vocalization onset and vocalization offset) and “0” for samples in the event series when no acoustic energy was detected (Abney, Paxton, et al., 2014). These square signals for each participant in each dyad were then cross-correlated.

Inspection of the averaged cross-correlation profile across all dyads revealed a distinct positive peak around lag 0, indicating that dyads tended to exhibit the same vocalization states at the same time (i.e., simultaneous speech or simultaneous silence). We used the cross-correlation estimate at lag 0 as our measure of vocal behavioral matching. Vocalization complexity matching and vocal behavioral matching were moderately positively correlated,  $r(35) = 0.38$ ,  $t = 2.41$ ,  $p = .02$ . In a linear regression model predicting tower height with vocal behavioral matching, vocal behavioral matching did not reliably predict tower height ( $b = 0.87$ ,  $\beta = 0.03$ ,  $SE = 4.27$ , 95% CI  $[-7.80, 9.55]$ ,  $p = .840$ ).

Finally, to determine if the relationship between vocalization complexity matching and tower height held independently of vocal behavioral matching, we conducted a final linear regression model on vocalization complexity matching predicting tower height after residualizing out vocal behavioral matching. Vocalization complexity matching significantly predicted residual tower height ( $b = 0.45$ ,  $\beta = 0.42$ ,  $SE = 0.16$ , 95% CI  $[0.12, 0.78]$ ,  $p = .009$ ). These results yield two important implications: First, in-the-moment matching of vocalizations does not predict performance, and second, although vocalization complexity matching and behavioral matching are moderately positively correlated, the functional relationship between vocalization complexity matching and performance is independent of vocalization behavioral matching.

## Discussion

Taken together, our findings build on previous work to support the idea that multimodal nested temporal dynamics are predictive of collaborative performance. Dyads with more similar nested patterns of vocalizations performed better on a collaborative task. We found that the convergence of the nested temporal structure of vocalizations was positively related to collaborative performance. This relationship was even stronger when we controlled for overall dyad-level vocal complexity, work rate, and vocal behavioral matching. Compared to surrogate pairs, original pairs had greater amounts of complexity matching for body movements and vocalizations.

The present results continue to affirm that complexity matching is not reducible to other forms of coordination such as behavioral matching and synchronization. We also replicate a growing body of research suggesting that individuals in interactive contexts match the complexity of their vocalizations and movements above and beyond random pairs (Abney, Paxton, et al., 2014; Abney, Warlaumont, et al., 2014; Marmelat & Delignières, 2012; Schloesser et al., 2019).

### The Functionality of Complexity Matching in Vocalizations and Body Movement

Although we observed evidence for movement complexity matching (above and beyond surrogate pairing), we did not observe a functional relationship between movement complexity matching and task performance. There are a few possible reasons for why the degree of complexity matching of movements would

not impact tower building performance. First, one possibility is that task constraints directly restricted participants' movements (especially by dictating spaghetti- and marshmallow-holder roles) without restricting participants' vocalizations. If so, we would expect lower variability of AF estimates for movement complexity. However, upon inspecting the variability of individual AF estimates and complexity matching estimates across the two modalities, we found more variability in movement complexity and movement complexity matching estimates compared to vocalizations.

It is also possible that movement complexity matching did not vary with task performance because movement complexity matching will occur in any kind of joint construction and therefore, the lack of any relationship with task performance is due to a ceiling effect. This explanation is likely considering there was greater movement complexity matching than vocalization complexity matching. Future research is needed to determine the variability of movement complexity matching across a diverse set of interaction contexts.

Another possible explanation is that vocalization complexity matching and movement complexity matching are conceptually different processes or serve different functional purposes. For vocalization complexity matching, the convergence of vocal clustering across multiple nested timescales reflects the emergence of common ground (Pickering & Garrod, 2004) between the two members of the dyad. This common ground affords the dyad to produce dialogue that includes plans, strategies, and opportunities to repair miscommunication. By contrast, movement complexity matching—in this context—is the *instantiation* of the continuous output of ideas generated by the vocal discourse. The observation that vocalization complexity matching, but *not* movement complexity matching, has a functional relationship with task performance is consistent with the overall concept of complexity matching introduced in statistical mechanics literature. Specifically, the outcome of convergence of complexity between two complex systems is increased information transmission between the two systems. Therefore, our results point to a functional relevance of vocalizations as a medium for information transmission.

### The Timescales of Complexity Matching in Vocalization and Movement

The degree of nested temporal structure differed across modalities: AF estimates were higher for movements relative to vocalizations. Further inspection of the aggregated AF functions (see Figure 5) showed that the main reason AF estimates were higher for movements was due to the lower variability in clustering at shorter timescales compared to vocalizations—which, in turn, steepened the linear fit estimates. This suggests that at shorter timescales, movement events cluster in more ordered sequences compared to vocalization events at the same timescales.

The nested organization of vocalizations likely originates from the constraints of temporal variations across linguistic levels. For example, phonetic variations occur within a hundred milliseconds, syllabic variations occur within a few hundred milliseconds, phrasal variations occur within hundreds of milliseconds to a few seconds, and utterance-level variations occur within few seconds. The variations across these time scales and corresponding linguistic units cluster within each other—or, in other words, are hierarchically organized within each other (Abney, Paxton, et al., 2014). The relationship between AF functions and variability in the

durations of linguistic units was demonstrated by Falk and Kello (2017) in a study of hierarchical temporal structure in infant-directed and versus adult-directed speech.

Compared with the nested structure of speech, less is known about the origins of nested organization of body movements and the hierarchical alignment of body movements. Variations in speech can be conceptually linked to linguistic units, but variations in body movements across multiple time scales have fewer context-independent conceptual mappings. While multilevel movement dynamics have been less of a focus than multilevel language dynamics, the present work showed that temporal variation of movements also clusters across multiple time scales.

A preliminary interpretation of the organization of body movements is that the nested movements may reflect bouts of activity at different levels of goal-directed behavior. Although highly context-specific, human interaction typically involves overarching goals (e.g., moving company relocates a household) that include subgoals (e.g., get futon from second floor bedroom into moving truck) and sub-subgoals (e.g., make sure futon does not scratch walls or flooring). Research on joke-telling provides the most relevant evidence for our interpretation of nested structure of body movements. Schmidt et al. (2012) observed that the periodicities of body movements related to the time scales specific to the temporal nature of joke-telling sequences: Stronger body movement periodicities were associated with the whole joke sequence ( $\sim 8$  s) and with the lines of the joke ( $\sim 1$ – $2$  s). In a reanalysis of Schmidt et al. (2012); Schmidt et al. (2014) observed higher entrainment across the multiple periodicities of body movements for joke teller and joke responder relative to chance, providing initial evidence for the notion of hierarchical alignment of body movements during social coordination. More recent work has demonstrated that the performance on a collaborative team problem solving task is constrained by low-level movement coordination at multiple temporal scales (Wiltshire et al., 2019).

The context-specific nature of movement timescales certainly adds complexity to our ability to interpret movement timescales, as compared with the (fairly) context-independent interpretations of vocalization timescales. However, we believe that grounding our understanding of the system in the *multiple* timescales at which it lives—that is, moving toward a more “meaning-full” quantification of the system—may provide critical insights into the functional role of body movement complexity matching. It may be that complexity matching across different modalities may be functionally connected to different outcomes (Paxton et al., 2016)—an important question for future work.

## Limitations

One limitation of the current study is that our vocalization and movement binary spike trains do not provide qualitative information about specific events (e.g., requests, backchanneling, gestures, postures). This limitation is a natural consequence of the decision to automate the onset detection protocols. Therefore, our interpretations of the results are restricted to the fact that our vocalization and movement data are based on magnitudes of acoustic energy and pixel change, respectively, and not based on other higher-level information like semantic content or goal-directed actions.

Another limitation is that our analysis for measuring clusters of activity across timescales does not directly map onto particular phases of interaction but rather maps onto approximate timescales

of when these phases (e.g., planning, execution) occur. Future work should focus on more accurately identifying the timescales of specific actions and phases that occur during interactions across various contexts.

In a similar vein, the structure of human interaction is context specific (Dale et al., 2013). Although the contributions of this study provide new insights into the consequences of convergent nested behaviors during a free-form cooperative task, the generalizability of our findings will require future research. For example, previous work has found that some dyadic tasks vary in the degree to which “local” versus “global” coordination processes explain behavioral phenomena (Fine et al., 2015; Marmelat & Delignières, 2012). We suggest that a systematic framework for describing the dimensions of human interactive contexts should be developed and used to determine the generalizability of different coordination patterns during contexts of human interaction that systematically vary along multiple dimensions.

## Treating Interacting Humans as Interacting Complex Systems

A key finding from the current study is that the convergence of two people’s vocalization complexity is associated with increased task performance during an explicitly cooperative task. The current study was inspired by models of information exchange between complex networks. Research in statistical physics has demonstrated that information exchange rates between complex networks are optimized when the complexity of the networks, akin to AF estimates in the current paper, converge (West et al., 2008). To some, it might be considered obvious that humans should be viewed as complex systems, but not all approaches to studying human interaction have taken this approach. For example, most alternative models of human interaction have formalized how people interact as the outcome of local interactions such as “mimicry,” “mirroring,” and “matching” that do not include the inclusion of temporal information like dependencies and histories of behaviors across timescales (for review, see Paxton et al., 2016). The assumption of viewing humans as complex systems requires a formalization that includes how behaviors interact across components of a system over time.

As discussed in the Limitations section, future research should focus on the generalizability of our contributions to other performance-based interactive contexts. Relatedly, stronger tests of the complexity matching hypothesis, as introduced by mathematical models of interacting complex networks, will require more in-depth quantifications of information transmission. Advances in computational linguistics using information-theoretic measures of communicative properties is one possible avenue (Genzel & Charniak, 2002; Levy & Jaeger, 2007), especially considering the present study’s observation of a functional relationship between the convergence of nested clusters of acoustic energy and task performance.

We interpret the convergence of the nested, temporal structure of behaviors in humans to conceptually equate to West et al.’s (2008) notion of complexity matching as the convergence of network complexity. Treating the human dyadic system as the coordination of two complex networks stimulates new questions and hypotheses about how context sensitivity constrains human cooperation, and we will end the current work by suggesting new

empirical questions that link the observations of the current study with concepts from complex adaptive systems.

One principle from complex adaptive systems and specifically, a set of coupled systems is the ability to respond to intrinsic and extrinsic perturbations. In a complex system of interdependent parts, a response to perturbation (commonly referred to as reciprocal compensation; see Kelso et al., 1984) refers to the ability of a component of the system to react to changes in the overall system or changes in other components of the system. Reciprocal compensation is a signature condition of a synergy (Kelso et al., 1984; Latash, 2008; Scholz & Schöner, 1999), which has gained increased attention in the study of human interaction by viewing the human dyad as a synergy (Riley et al., 2011).

Under the complexity matching framework, the response to perturbation principle suggests a key hypothesis for how dyads adapt to changes in the interaction. Specifically, this suggests that a dyad with a higher degree of complexity matching should be able to respond and adapt faster in the face of a perturbation. Importantly, this hypothesis would even be testable under the same overall task used in the current paper. For example, halfway through the tower building phase, the dyad could be asked to change the goal of the construction from building the tallest tower structure to building a tower structure that could hold the most weight. Under this approach, we would hypothesize that dyads with higher degrees of complexity matching during the introduction of the perturbation should be able to be more flexible and adaptable, and this compensation should be reflected in subsequent measures of complexity matching.

## Conclusion

Here, we sought to test hypotheses that have consequences for how to study human interaction and coordination. By adapting a model of interacting complex systems—complexity matching—to human interaction, we observed that the complexity of human vocal and motor behavior is organized across nested temporal scales during cooperative task-related interactions. Moreover, we found that the nesting of vocalizations between cooperating partners converges during interaction and provided new evidence of the function of complexity matching of vocalizations: higher rates of joint success.

## Context

The hypothesis that complexity matching could be a model for human interaction shares a similar history with previous research that applied mathematical models of synchronization to human interaction. Models from statistical physics suggested that information transmission between two complex systems is optimized when those systems have matching activation patterns across multiple temporal scales. Similar to decades of research applying mathematical models of synchronization to models of human interaction, psychologists have started developing hypotheses about human interaction motivated by mathematical models of complexity matching. Previous work from our research program provided some of the initial evidence for complexity matching in human interaction (Abney, Paxton, et al., 2014, *Journal of Experimental Psychology: General*). Despite a growing collection of evidence for complexity matching in psychological phenomena, no research has yet to test a critical *functional* hypothesis of com-

plexity matching: Increased complexity matching of behaviors across multiple temporal scales should lead to increased information transmission. The current study provides evidence for the hypothesis that higher rates of complexity matching are associated with higher rates of cooperation. Future research should focus on replicating our findings in diverse contexts and across various measures of information transmission.

## References

- Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General, 143*(6), 2304–2315. <https://doi.org/10.1037/xge0000021>
- Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2015). Movement dynamics reflect a functional role for weak coupling and role structure in dyadic problem solving. *Cognitive Processing, 16*(4), 325–332. <https://doi.org/10.1007/s10339-015-0648-2>
- Abney, D. H., Warlaumont, A. S., Haussman, A., Ross, J. M., & Wallot, S. (2014). Using nonlinear methods to quantify changes in infant limb movements and vocalizations. *Frontiers in Psychology, 5*, 771. <https://doi.org/10.3389/fpsyg.2014.00771>
- Abney, D. H., Warlaumont, A. S., Oller, D. K., Wallot, S., & Kello, C. T. (2017). Multiple coordination patterns in infant and adult vocalizations. *Infancy, 22*(4), 514–539. <https://doi.org/10.1111/inf.12165>
- Allan, D. W. (1966). Statistics of atomic frequency standards. *Proceedings of the IEEE, 54*(2), 221–230. <https://doi.org/10.1109/PROC.1966.4634>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (R package version 1.1–7). [https://www.google.com/search?q=Bates%2C+D.%2C+Maechler%2C+M.%2C+Bolker%2C+B.%2C+%26+Walker%2C+S.++\(2014\).+lme4%3A+Linear+mixed-effects+models+using+Eigen+and+S4.+R+package+version+1.1%E2%80%937.&aq=chrome.69i57.341j0j4&sourceid=chrome&ie=UTF-8](https://www.google.com/search?q=Bates%2C+D.%2C+Maechler%2C+M.%2C+Bolker%2C+B.%2C+%26+Walker%2C+S.++(2014).+lme4%3A+Linear+mixed-effects+models+using+Eigen+and+S4.+R+package+version+1.1%E2%80%937.&aq=chrome.69i57.341j0j4&sourceid=chrome&ie=UTF-8)
- Bedwell, W. L., Wildman, J. L., DiazGranados, D., Salazar, M., Kramer, W. S., & Salas, E. (2012). Collaboration at work: An integrative multilevel conceptualization. *Human Resource Management Review, 22*(2), 128–145.
- Bernieri, F. J. (1988). Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal Behavior, 12*(2), 120–138. <https://doi.org/10.1007/BF00986930>
- Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). *The self-organization of human interaction. Psychology of learning and motivation* (Vol. 59, pp. 43–95). Academic Press.
- Falk, S., & Kello, C. T. (2017). Hierarchical organization in the temporal structure of infant-direct speech and song. *Cognition, 163*, 80–86. <https://doi.org/10.1016/j.cognition.2017.02.017>
- Fine, J. M., Likens, A. D., Amazeen, E. L., & Amazeen, P. G. (2015). Emergent complexity matching in interpersonal coordination: Local dynamics and global variability. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 723–737.
- 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 Experimen-*



- tal Psychology: Human Perception and Performance*, 42(9), 1297–1310. <https://doi.org/10.1037/xhp0000207>
- Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. *Proceedings of the 40th Annual meeting of the Association for Computational Linguistics* (pp. 199–206).
- Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, 27(6), 949–960. <https://doi.org/10.1521/soco.2009.27.6.949>
- Isenhower, R. W., Richardson, M. J., Carello, C., Baron, R. M., & Marsh, K. L. (2010). Affording cooperation: Embodied constraints, dynamics, and action-scaled invariance in joint lifting. *Psychonomic Bulletin & Review*, 17(3), 342–347. <https://doi.org/10.3758/PBR.17.3.342>
- Keith, T. Z. (2014). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling*. Routledge. <https://doi.org/10.4324/97811315749099>
- Kello, C. T., Bella, S. D., Médé, B., & Balasubramaniam, R. (2017). Hierarchical temporal structure in music, speech and animal vocalizations: Jazz is like a conversation, humpbacks sing like hermit thrushes. *Journal of the Royal Society, Interface*, 14(135), 20170231. <https://doi.org/10.1098/rsif.2017.0231>
- Kelso, J. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10(6), 812–832. <https://doi.org/10.1037/0096-1523.10.6.812>
- Latash, M. L. (2008). *Synergy*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195333169.001.0001>
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems 19* (pp. 849–856). MIT Press.
- 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>
- Marmelat, V., & Delignières, D. (2012). Strong anticipation: Complexity matching in interpersonal coordination. *Experimental Brain Research*, 222(1–2), 137–148. <https://doi.org/10.1007/s00221-012-3202-9>
- Mayo, O., & Gordon, I. (2020). In and out of synchrony—Behavioral and physiological dynamics of dyadic interpersonal coordination. *Psychophysiology*, 57(6), e13574.
- Nalepka, P., Kallen, R. W., Chemero, A., Saltzman, E., & Richardson, M. J. (2017). Herd those sheep: Emergent multiagent coordination and behavioral-mode switching. *Psychological Science*, 28(5), 630–650. <https://doi.org/10.1177/0956797617692107>
- Oullier, O., de Guzman, G. C., Jantzen, K. J., Lagarde, J., & Scott Kelso, J. A. (2008). Social coordination dynamics: Measuring human bonding. *Social Neuroscience*, 3(2), 178–192. <https://doi.org/10.1080/17470910701563392>
- Paxton, A., & Dale, R. (2013a). Argument disrupts interpersonal synchrony. *Quarterly Journal of Experimental Psychology*, 66(11), 2092–2102.
- Paxton, A., & Dale, R. (2013b). Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior Research Methods*, 45(2), 329–343. <https://doi.org/10.3758/s13428-012-0249-2>
- Paxton, A. E., Dale, R., & Richardson, D. C. (2016). Social coordination of verbal and nonverbal behaviors. In P. Passos, K. Davids & J. Y. Chow (Eds.), *Interpersonal coordination and performance in social systems* (pp. 259–273). Routledge.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169–190.
- Raffard, S., Salesse, R. N., Marin, L., Del-Monte, J., Schmidt, R. C., Varlet, M., Bardy, B. G., Boulenger, J.-P., & Capdevielle, D. (2015). Social priming enhances interpersonal synchronization and feeling of connectedness towards schizophrenia patients. *Scientific Reports*, 5, 8156. <https://doi.org/10.1038/srep08156>
- Ramirez-Aristizabal, A. G., Médé, B., & Kello, C. T. (2018). Complexity matching in speech: Effects of speaking rate and naturalness. *Chaos, Solitons, and Fractals*, 111, 175–179. <https://doi.org/10.1016/j.chaos.2018.04.021>
- 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>
- Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26(6), 867–891. <https://doi.org/10.1016/j.humov.2007.07.002>
- Riley, M. A., Richardson, M., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in Psychology*, 2, 38. <https://doi.org/10.3389/fpsyg.2011.00038>
- Schloesser, D. S., Kello, C. T., & Marmelat, V. (2019). Complexity matching and coordination in individual and dyadic performance. *Human Movement Science*, 66, 258–272. <https://doi.org/10.1016/j.humov.2019.04.010>
- Schmidt, R., Morr, S., Fitzpatrick, P., & Richardson, M. J. (2012). Measuring the dynamics of interactional synchrony. *Journal of Nonverbal Behavior*, 36(4), 263–279. <https://doi.org/10.1007/s10919-012-0138-5>
- Schmidt, R., Nie, L., Franco, A., & Richardson, M. J. (2014). Bodily synchronization underlying joke telling. *Frontiers in Human Neuroscience*, 8, 633. <https://doi.org/10.3389/fnhum.2014.00633>
- Schneider, S., Ramirez-Aristizabal, A. G., Gavilan, C., & Kello, C. T. (2020). Complexity matching and lexical matching in monolingual and bilingual conversations. *Bilingualism: Language and Cognition*, 23(4), 845–857. <https://doi.org/10.1017/S1366728919000774>
- Scholz, J. P., & Schöner, G. (1999). The uncontrolled manifold concept: Identifying control variables for a functional task. *Experimental Brain Research*, 126(3), 289–306. <https://doi.org/10.1007/s002210050738>
- Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1), 201–208. <https://doi.org/10.1037/0096-1523.33.1.201>
- 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>
- Team, R. C. (2016). *R: A language and environment for statistical computing* [Computer software manual]. R Foundation for Statistical Computing.
- Turner, M. G., Dale, V. H., & Everham, E. H., III. (1997). Fires, hurricanes, and volcanoes: Comparing large disturbances. *Bioscience*, 47(11), 758–768. <https://doi.org/10.2307/1313098>
- West, B. J., Geneston, E. L., & Grigolini, P. (2008). Maximizing information exchange between complex networks. *Physics Reports*, 468(1–3), 1–99. <https://doi.org/10.1016/j.physrep.2008.06.003>
- Wiltshire, T. J., Steffensen, S. V., & Fiore, S. M. (2019). Multiscale movement coordination dynamics in collaborative team problem solving. *Applied Ergonomics*, 79, 143–151. <https://doi.org/10.1016/j.apergo.2018.07.007>

Received March 27, 2020

Revision received September 17, 2020

Accepted November 1, 2020 ■