Correlations between speech behavior and social network dynamics in a constrained vocabulary game

Sam Tilsen

1. Introduction

The speech and social network dynamics experiment investigates interpersonal social influences on speech in an ad-hoc network, focusing on a timescale of days/weeks. Generally, we want to understand what can we predict about change in speech behavior on this timescale. The generic hypotheses of the study are that some temporal variation in speech is caused by interactions between people, and that interpersonal social relations modulate this effect. A detailed presentation of hypotheses, background, and methodology appears in Tilsen (2015), and is not repeated in this paper, which focuses on preliminary analyses of experimental data. These analyses demonstrate that macroscopically, variation in social relations correlates with variation in speech behavior, specifically in regards to vowel quality, sibilant fricative quality, and syntactic patterns. These findings prompt an analysis of social modulation of behavior on the scale of individual interactions, which provides less conclusive results.

To contextualize the rationale for the experimental design, let's consider three major problems in analyzing social influences on speech. First, one problem is the vast range of contexts in which speech typically occurs and the enormous variability in conversational goals. These characteristics of speech are obvious from a bit of self-reflection: consider for each utterance you make in a given day, where you were, who was there, and what the direct goal of that specific utterance was. There is very little consistency in our answers to these questions, which is problematic for several reasons: the occurrence rates of most lexical items are relatively low, the lexical/syntactic contexts in which those items occur are quite diverse, and word productions are influenced by hard-to-quantify paralinguistic factors. Hence statistical power is low: if we choose to study the production of some particular word, we will tend to have to wait a long time to observe multiple tokens, and even longer if we hope to observe them in relatively similar contexts. Thus it is desirable to minimize linguistic-contextual variation and goal/task-contextual variation.

A second major obstacle is the complexity of the social networks in which speakers participate. None of our networks are isolated. Most of us participate in numerous social networks, which can be defined on a range of spatial and temporal scales. The structure of these networks and our positions in them vary, many networks overlap or are embedded/hierarchically related, we occasionally leave old networks and join new ones, and different networks carry different contextual associations and social valences. Even defining the social networks of a given speaker requires many arbitrary analytical decisions. The problem is that if we want to understand the influence of social relations in a given network on linguistic behavior, we have to quantify those relations. To do that, we have assume that social relations in the networks that we have *not* sampled have negligible effects. This seems to be a pretty miserable assumption, but we might lessen such effects by constructing an ad-hoc network, i.e. selecting speakers who have no prior acquaintance with each other.

A third obstacle is the logistical and methodological difficulty in obtaining information regarding speech behavior and social relations with sufficiently frequent spatial

and temporal sampling. It is fairly obvious that a decent spatiotemporal sampling of speech behavior requires recording all of the utterances from all of the speakers in a network over an extended period of time. In contrast, it is far from obvious how to frequently sample social relations: the sampling procedure should provide quantitative measures of social relations between all dyads in the network, and should obtain those measures frequently in time—yet the sampling procedure should not unduly interfere with the collection of speech data, and should not be so invasive that it drives the social dynamics of network.

How have these problems—i.e. contextual variation, network complexity, social sampling—been addressed by previous studies of speech on extended timescales? The network complexity problem has been addressed by using a corpus of data from members of a coherent social network, e.g. the reality television show Big Brother (Bane, Graff, & Sonderegger, 2010) or the U.S. Supreme Court (Yu, Abrego-Collier, Phillips, Pillion, & Chen, 2015). However, data from such contexts are suboptimal because of high variability in interaction contexts, sparse sampling of speech behavior (i.e. "off-camera" and "off-the-record" interactions), and lack of quantitative information regarding social relations.

The contextual variation problem has been addressed by controlling interaction contexts, for example using a map task in which dyadic interactions are highly goal-oriented (Anderson et al., 1991; Pardo, 2006). Such approaches have demonstrated pairwise convergence in dyadic interactions, but have not investigated those patterns on longer timescales. The problem of how to sample social relations both frequently and non-invasively has not been addressed to my knowledge.

The current experiment is an attempt to combine some of the desirable features of the above studies. Specifically, network complexity was minimized by the use of a small, ad-hoc network. Task and linguistic contextual variation was minimized by allowing speech interactions only during a map game and by enforcing the use of a limited vocabulary when playing the game. Frequent sampling of social relations was achieved by eliciting full-network teammate preference rankings from each player in each round of the game. Below I briefly describe the design of the experiment, report on several types of analyses, and discuss future analysis directions.

2. Method

2.1 Experiment design

A 10-week longitudinal study was conducted with an ad-hoc group of eight native English speakers (all college freshman/sophomores, four females/four males). The participants played a total of 134 rounds of a map game over the course of the experiment, which amounted to 535 total games. In each game, a giver and receiver sit facing one another, in front of laptops with a randomly generated map on the screen. The maps are identical for the giver and receiver, except that the giver has a route on their map. The goal is for the giver to verbally communicate to the receiver how to draw the route, which the receiver accomplishes by clicking on map locations in the correct order. Fig. 1A shows part of a map (20% of map area) and provides an example of typical giver instructions for the first three route segments (19 total segments were present on each route). The vocabulary of the map task was highly constrained: players were allowed to say only location names (8 unfamiliar nonwords), location properties (3 colors: *red*, *green*, *blue*; 3 shapes: *circle*, *triangle*, *square*; 2

sizes: *big*, *small*; 2 textures: *filled*, *unfilled*), and a small set of function words (*up*, *down*, *left*, *right*, *and*, *okay*, etc.).

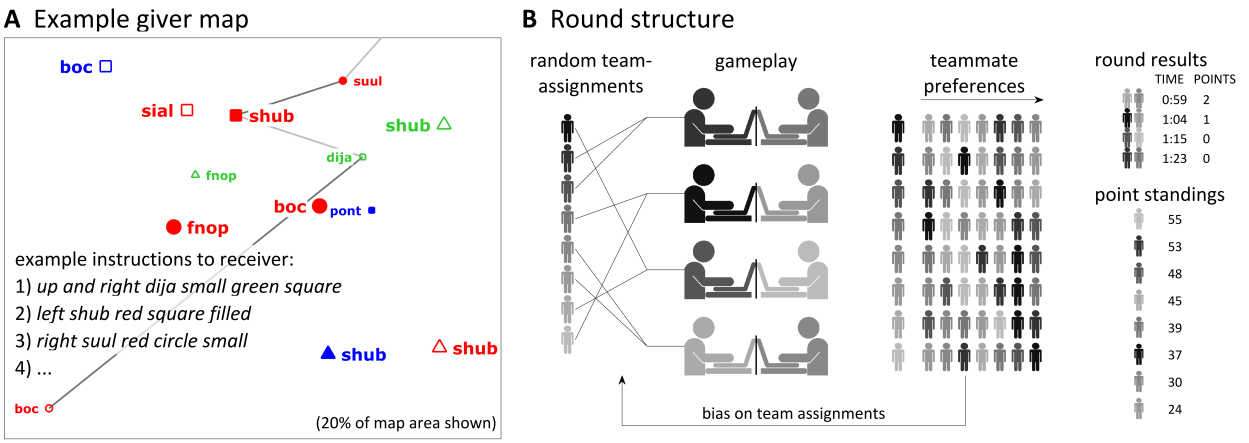

Fig. 1. Example map and round structure. (A) Giver map with route from starting location. Receiver has an identical map, initially without the route. Example instructions to receiver for the first three route segments are shown. 20% of map area is shown. (B) Round structure: each round begins with random team assignments; after completing a game, participants produce confidential teammate preference rankings. Points are awarded to players on the two fastest teams in the round and cumulative point standings are displayed. The teammate preference rankings bias the team assignments in the next round.

The structure of each round in the experiment is schematized in Fig. 1B. Each round began with a random assignment of the eight players to four teams. Two teams played the game simultaneously in separate rooms, and the other two teams stayed in a waiting room until their turn to play. Immediately before and after each game, players completed a four question survey. After each game, players also produced a teammate preference ranking: drop-down lists were used to order the other seven players according to whom they most/least wanted to be on a team with in the next round (cf. Fig. 2A). The teammate preference rankings were used to bias the random team assignments in the next round. After all four games in a round had been completed, players were gathered in a lobby and the game completion times were presented. Two/one points were awarded to players on the fastest/second fastest teams, and cumulative player point standings were displayed. Subsequently new teams were randomly generated (biased by the teammate preference rankings from the preceding round), and this process was iterated for 90 min in each of the ten sessions.

Additional details are reported in (Tilsen, 2015). The reader is encouraged to consult this prior work for more methodological information, as the focus here is on preliminary analyses. Nonetheless, several specific features of the design are worth emphasizing:

First, outside of gameplay, participants were forbidden from speaking to each other during the experimental sessions, and were not able to hear other teams playing the game while waiting for their turn. This ensured that spoken communicative interactions were always dyadic and were restricted to the context of the map game. Second, participants were explicitly instructed several times in the initial session that their teammate preference rankings influenced but did not fully determine the team assignments in the next round. This ensured that the rankings were of consequence for the participants. Third, the game

vocabulary was highly constrained in order to elicit many tokens of the same words. Moreover, the location names were unfamiliar nonwords, and some of these had ambiguous grapheme-phoneme mappings. These aspects of the design increased statistical power, increased the opportunity for phonetic drift, and increased the potential for variation in pronunciation. The nonwords are also unlikely to have been uttered between experimental sessions, reducing the potential for unobserved datapoints. It should also be noted that assistant experimenters monitored all games for word violations and 5 second penalties were applied for each violation, during the results presentation; after the first several rounds violations were extremely rare.

2.2 Social distance metric

The teammate preference rankings are used to construct a social distance metric, which plays an important role in analyses. Most generally, the social network is conceptualized as a fully connected network of speakers, with bidirectional connections. Each connection is associated with a scalar variable that represents a "social distance". The social distance from player X to Y is assumed to correlate with behaviorally relevant dimensions of the myriad interpersonal attitudes that player X has toward player Y; no attempt is made here to analyze the composition of those attitudes.

The social distance metric can also be understood as a projection of the complex cognitive processes which underlie the teammate preference ranking behavior. As shown in Fig. 2A, this behavior was prompted by the instruction: "rank all players by who you most want to play with in the next round", and was accomplished by the selection of 7 unique players in 7 drop-down lists arranged vertically on the laptop screen. The initial ordering of player names in each of the drop-down lists always corresponded to the player point standings (highest standing player listed first), with alphabetical ordering for ties.

Hence after each round an asymmetric ranking matrix is obtained, as in Fig. 2B. These rankings were symmetrized by averaging player pairs, and the symmetrized rankings are used as a social distance metric, examples of which are shown for a subnetwork of players in Fig. 2C. The distances d(X,Y) are labeled on the connections between player-nodes. An alternative social distance metric can be defined using the asymmetric rankings. The asymmetric social distance space is 56-dimensional, while the symmetrized social distance space is 28-dimensional, with one dimension for each unique player-pair. It is important not to mistakenly infer from two-dimensional projections of the network (such as Fig. 2C) that the distances might constitute a metric space: this is not the case because the triangle inequality is not observed, i.e. d(X,Z) is not necessarily less than d(X,Y) + d(Y,Z). It is also worthwhile to note that in the asymmetric bidirectional network, the total social distance from any given player is constant, but not in the symmetric, unidirectional network. Nonetheless, in both cases the total social distance in the network is constant.

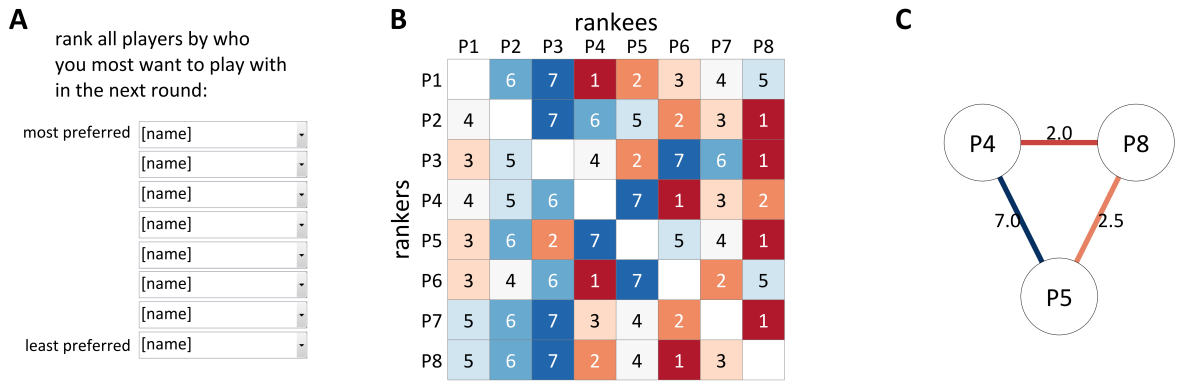

Fig. 2. Social distance metric obtained from teammate preference rankings. (A) Players used seven drop-down lists to rank the other players. (B) Asymmetric ranking matrix, values represent teammate preference order. (C) Symmetrized social distances for the subnetwork of P4, P5, and P8, derived by averaging pairwise asymmetric rankings.

It was important for analytical purposes that players produce non-arbitrary teammate preference rankings, because these are used to quantify social relation dynamics. One concern was that social preferences would not influence in the teammate preference rankings. If that were the case, then players would most likely adopt the most expedient ranking strategy, which would be to rank players in the default list order (i.e. according to the current player point standings). Of course, it is ambiguous whether a teammate preference ranking that mirrors the standings is the result of expedience, or a strategy for improving the chances of winning. Another concern was that preference rankings would not exhibit a high degree of variability: in the absence of fluctuations in the social distance metric, analysis of correlations between social distance and linguistic behavioral distance would not be possible.

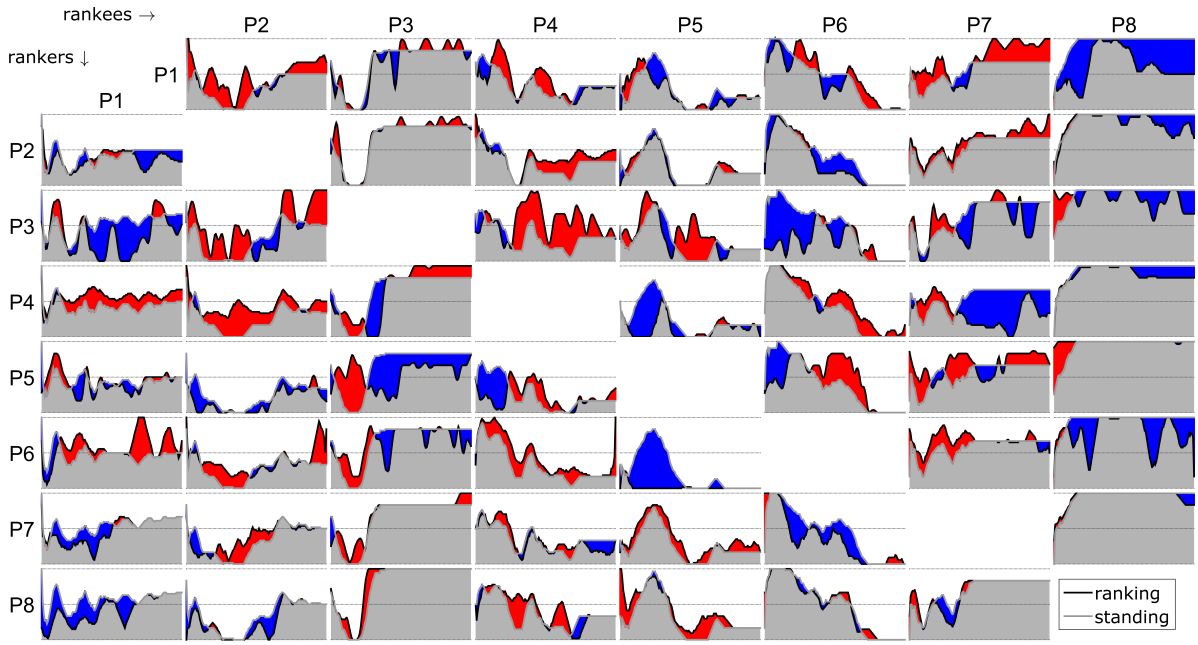

Fig. 3. Ranking time series. Rows correspond to rankers, columns to rankees. Gray lines show standings (default ranking in drop-down lists), black lines show selected rankings; red/blue indicate positive/negative differences between ranking and standing.

Fortunately, all of the players typically produced teammate preference rankings that deviated from point standings and that fluctuated substantially over time. These properties are evident in Fig. 3, which shows the full time series of asymmetric rankings for each player-pair. In each panel of the figure, the gray lines represent the standing of the rankee, and the black line represents the teammate preference ranking of that rankee. Red/blue areas correspond to rounds in which the ranker ranked the rankee more/less highly than their standing.

3. Analyses

In this section I report three types of analyses. First, a macroscopic analysis of relations between social distance and linguistic behavior reveals substantial correlations between social and linguistic variables. This analysis is macroscopic in both a spatial and temporal sense: spatially the analysis is large-scale because it pools the entire set of 28 player-pairs; temporally the analysis is large-scale because it makes use of a temporal coarse-graining procedure in which observations are averaged over non-overlapping periods of time.

Second, a microscopic analysis is conducted, in which the effects of social distance on the interaction scale are examined. Although the macroscopic results suggest that the microscopic analysis should also demonstrate socially driven modulation of behavior, the interaction-scale analysis is inconclusive. Several possible reasons for this are considered further in Section 4: interaction-scale measures may be too variable, a nonlinear relation between social distance and behavioral change may be required, and a social modulation factor may need to the incorporate additional information. Third, I present first-pass analyses of social and behavioral fluctuations and discontinuities; these inform future development of the interaction-scale model.

3.1 Macroscopic analysis of correlation between vowel quality distance and social distance

A macroscopic analysis of social distance and vowel quality distance in location names shows a substantial correlation between social and acoustic information. Nine different vowel categories were analyzed from the nonword location names: *boc*, *dija*, *fnop*, *pont shub*, *sial*, *sond*, and *suul* (note that there are two vowels in *dija*). To quantify vowel quality the following procedures were used:

[1] The central 50% of each vowel waveform was transformed into an auditory spectrogram using a gammatone filterbank (Ellis, 2009) with the following parameters: 64 erb filters over 70-10000 Hz, 20 ms window, 10 ms step.

[2] Auditory spectrograms were linearly time-warped to the median frame length in each vowel category. Auditory spectra from midpoint frames of *boc* are shown in Fig. 4A, but the reader should note that each vowel is associated with several dozen such frames.

[3] For each vowel category and player, tokens were excluded from subsequent analyses when more than 25% of the time/frequency bins had z-scores greater than ±2.0.

[4] For each vowel category, auditory spectra were pooled across speakers and converted to principal components, as shown in Fig. 4B.

[5] Coarse-grained time-varying states for each player/vowel category were defined by averaging principle component vectors of datapoints from non-overlapping windows on all scales (i.e. 1 round, the interaction scale, to 67 rounds, the maximal coarse grain). The trajectories in Fig. 4C show examples of vowel state evolution, although these were calculated with overlapping Gaussian windows to emphasize continuity.

[6] Time-varying vowel similarities were defined by computing the Euclidean distances between the coarse-grained states of each vowel category for each player-pair. The distance calculation was restricted to the first 6 principal components, which generally contained 90-95% of the variance in each category.

[7] Mutual information between social distance time-series and vowel similarity time-series was calculated for each analysis scale, vowel, and player-pair. Mutual information between variables X and Y is the sum of the entropies of the probability distributions of X and Y minus the joint probability distribution of X and Y. Probability distributions were estimated by standardizing both variables and estimating the one- and two-dimensional distributions with 30 equally spaced bins over the range [-3,3] in each dimension, with a Gaussian kernel and bandwidth following Silverman (1986).

[8] A Monte Carlo permutation procedure was used to estimate the expected distribution of mutual information for independent variables for each scale/vowel/pair. The procedure in [7] was applied to 2000 random permutations of the vowel similarity time-series.

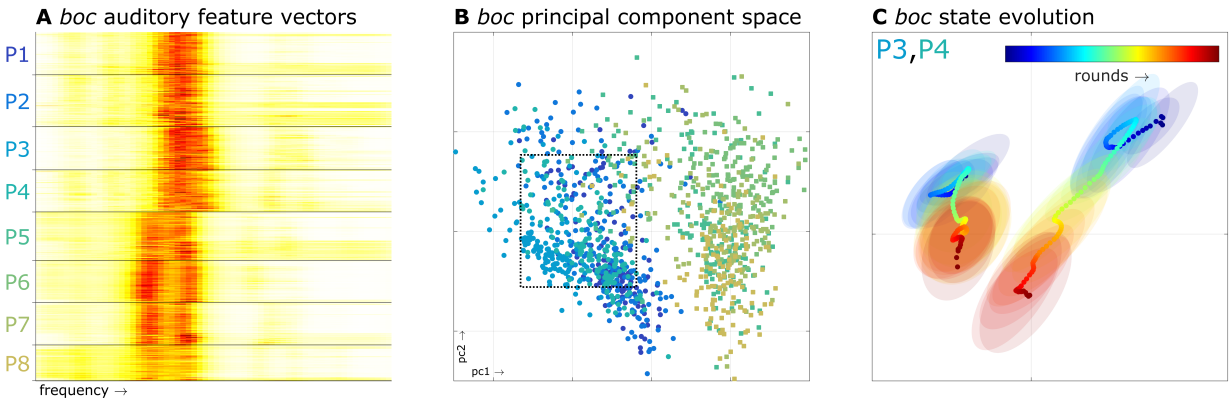

Fig. 4. Example of vowel state estimation. (A) Vowel-midpoint auditory spectra for each token of *boc*. Actual feature vectors consist of a temporal series of such spectra, i.e. auditory spectrograms. (B) First two principal components of auditory spectrograms of *boc*. The first component encodes much of the gender-related variation. (C) Smoothed state-space trajectories over the experiment for players P3 and P4, exemplifying convergence; region shown corresponds to dashed box in panel (B).

Social distance and vowel distance exhibit a substantial degree of excess mutual information, in comparison to randomly permuted variables. Fig. 5A shows the cumulative densities of mutual information from the empirical and randomized data on a 7-round analysis scale. The distributions describe the mutual information estimates from all vowels/player pairs (9 × 28 = 252 empirical datapoints). The difference between the empirical and random distributions shows that vowel quality distance and social distance have more redundancy than expected by chance. Notably, excess mutual information is present on all analysis scales (Fig. 5B), plateauing on scales of 4-9 rounds.

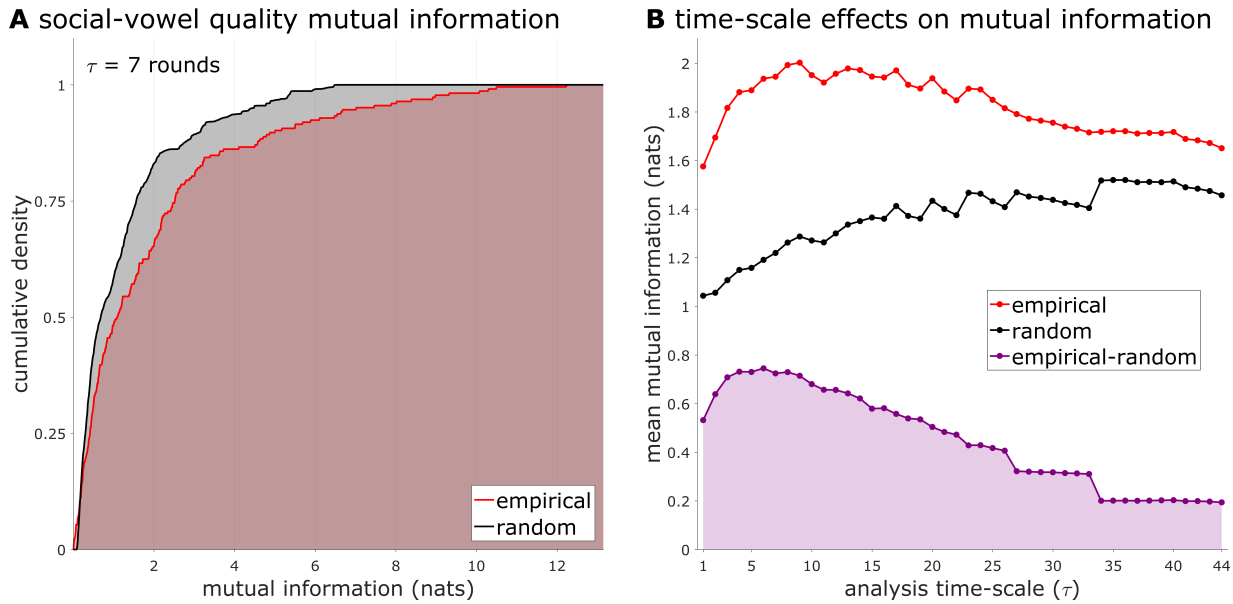

Fig. 5. Mutual information between social distance and vowel quality distance. (A) Cumulative density function of empirical and random mutual information, estimated on a scale of 7 rounds. (B) Mean mutual information of empirical and randomly permuted data for scales of 1-44 rounds. The difference between means, i.e. excess average mutual information, is shown in purple.

Mutual information is preferred over linear correlation as a measure of correlation here because the social distances and vowel similarities are not stationary and their relation is unlikely to be linear. Nonetheless it should be noted that excess linear correlation between social and vowel distance obtained with identical coarse-graining and permutation procedures is mostly positive, as would be expected from the hypothesis that players converge to a greater degree with players they prefer as teammates, and vice versa diverge from dispreferred teammates. The presence of excess negative correlation is not expected and may be an artifact of the constant sum of social distance in the network.

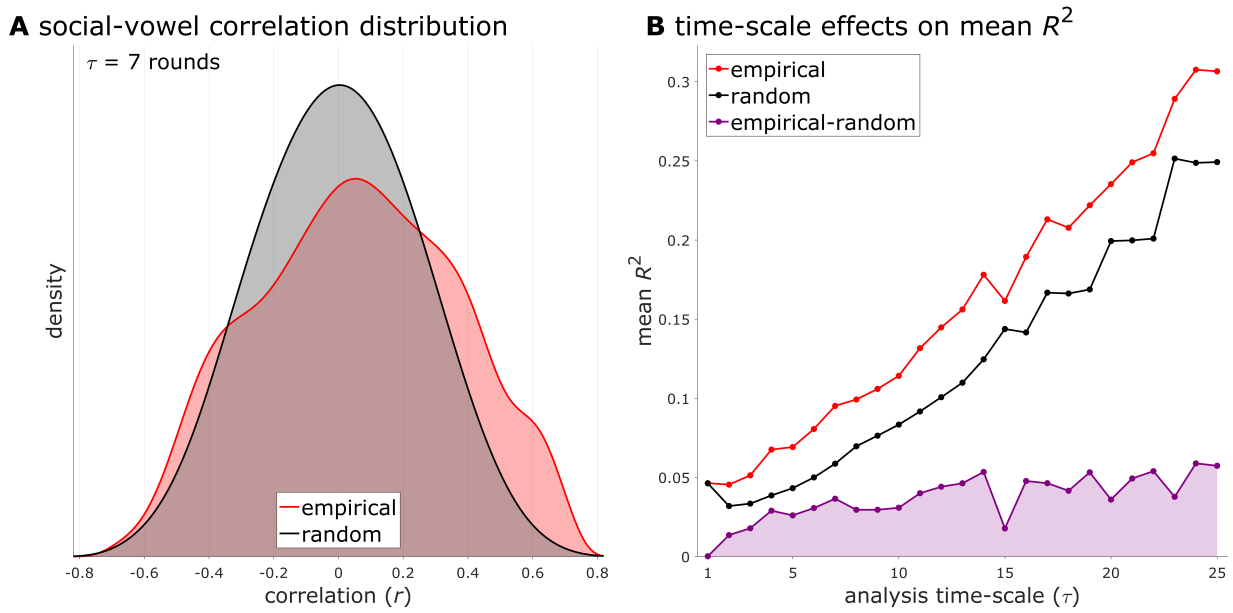

Fig. 6. Correlation between social distance and vowel quality distance. (A) Kernel estimate of probability density function of empirical and random correlations, estimated on an analysis scale of 7 rounds. (B) Mean R^2 of empirical and randomly permuted data for analysis scales of 1-25 rounds. The difference between means, i.e. excess average correlation, is shown in purple.

3.2 Macroscopic analysis of correlation between sibilant quality distance and social distance

Analysis of social distance and sibilant quality distance shows a substantial correlation between these two variables. Four different sibilant categories were analyzed from the nonword location names *shub*, *sial*, *sond*, and *suul*. To quantify sibilant quality distance, nearly identical procedures as for vowel quality were applied to sibilant waveforms, except that a step size of 5 ms was used for the auditory spectrograms. Fig. 7A shows the cumulative densities of mutual information from the empirical and randomized data on a 7-round analysis scale. As with vowel quality, excess mutual information is present on all analysis scales (Fig. 7B), but in this case plateauing over wider range of scales. Fig. 8 shows that the excess linear correlation between social and sibilant distance is almost entirely positive.

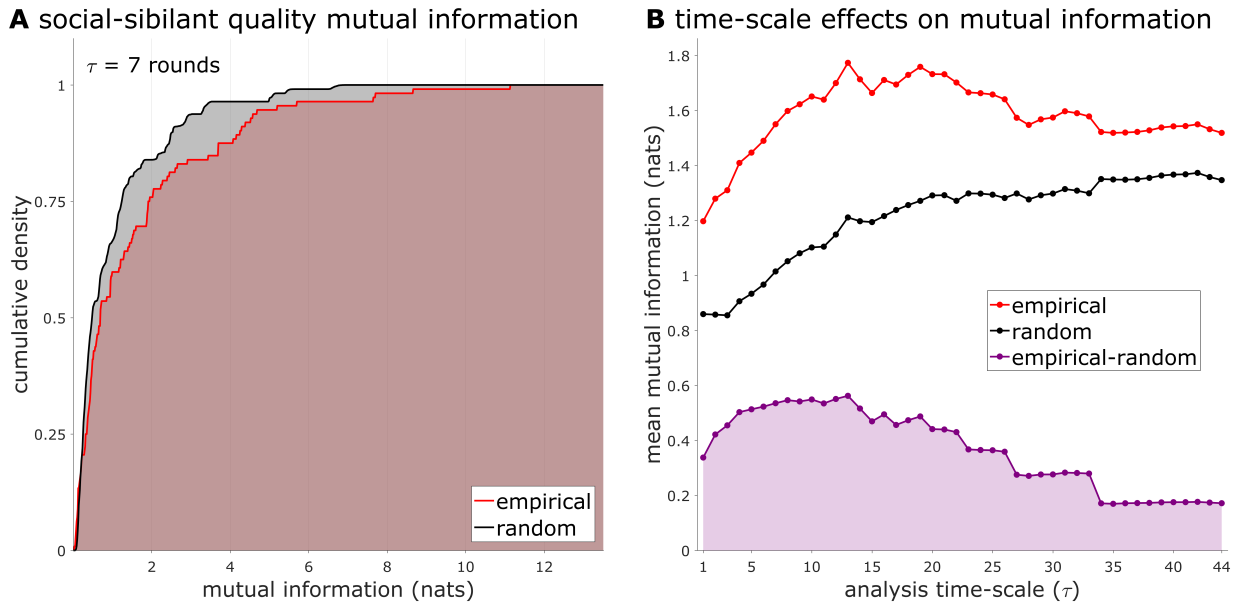

Fig. 7. Mutual information between social distance and sibilant quality distance. (A) Cumulative density function of empirical and random mutual information, estimated on a scale of 7 rounds. (B) Mean mutual information of empirical and randomly permuted data for scales of 1-44 rounds. The difference between means, i.e. excess average mutual information, is shown in purple.

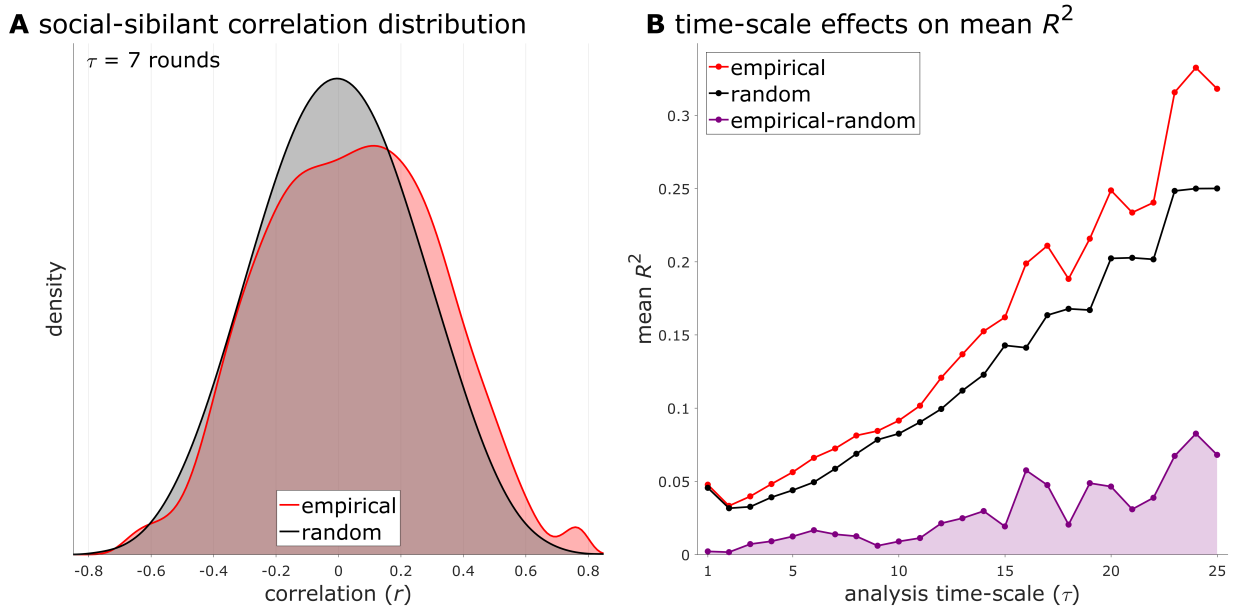

Fig. 8. Correlation between social distance and sibilant quality distance. (A) Kernel estimate of probability density function of empirical and random correlations, estimated on a scale of 7 rounds. (B) Mean R^2 of empirical and randomly permuted data for analysis scales of 1-25 rounds. The difference between means, i.e. excess average linear correlation, is shown in purple.

3.3 Macroscopic analysis of correlation between syntactic distance and social distance

A macroscopic analysis of social distance and syntactic distance in giver instructions shows a substantial correlation between these variables. To quantify syntactic distances, the following procedures were used:

[1] All instruction sequences (n=10,151) were converted to forward and backward word-category transition count matrices. Source and sink (START, END) states were imposed on each instruction sequence.

[2] On all analysis scales (1 round to 67 rounds), for each player, first-order Markov chain forward transition probability matrices were calculated from the count matrices, with examples shown in Fig. 9. Cells with zero probability were given a value of 1.0^{-5} to avoid singular entropy estimates and matrix rows were renormalized to sum to unity.

[3] The Jensen-Shannon distance, which provides a metric of the similarity between two probability distributions, was calculated for each word category between each player-pair at each time step, on all analysis scales. Examples are shown on the right of Fig. 9, expressed as a proportion of the maximum distance. Each row of the transition probability matrix is a separate probability distribution, describing the probability of the next word category (column headers) given the preceding word category (row headers). The Jensen-Shannon distance is the square root of the Jensen-Shannon divergence (JSD), which is the average of the Kullbak-Leibler divergences D(X||Z) and D(Y||Z), where Z is 0.5*(X+Y). Hence the JSD is a symmetric, smoothed version of the Kullbak-Leibler divergence.

[4] Syntactic distance was defined as the average of the Jensen-Shannon distances associated with all of the categories in each Markov chain. The END category was excluded because it always transitions to START.

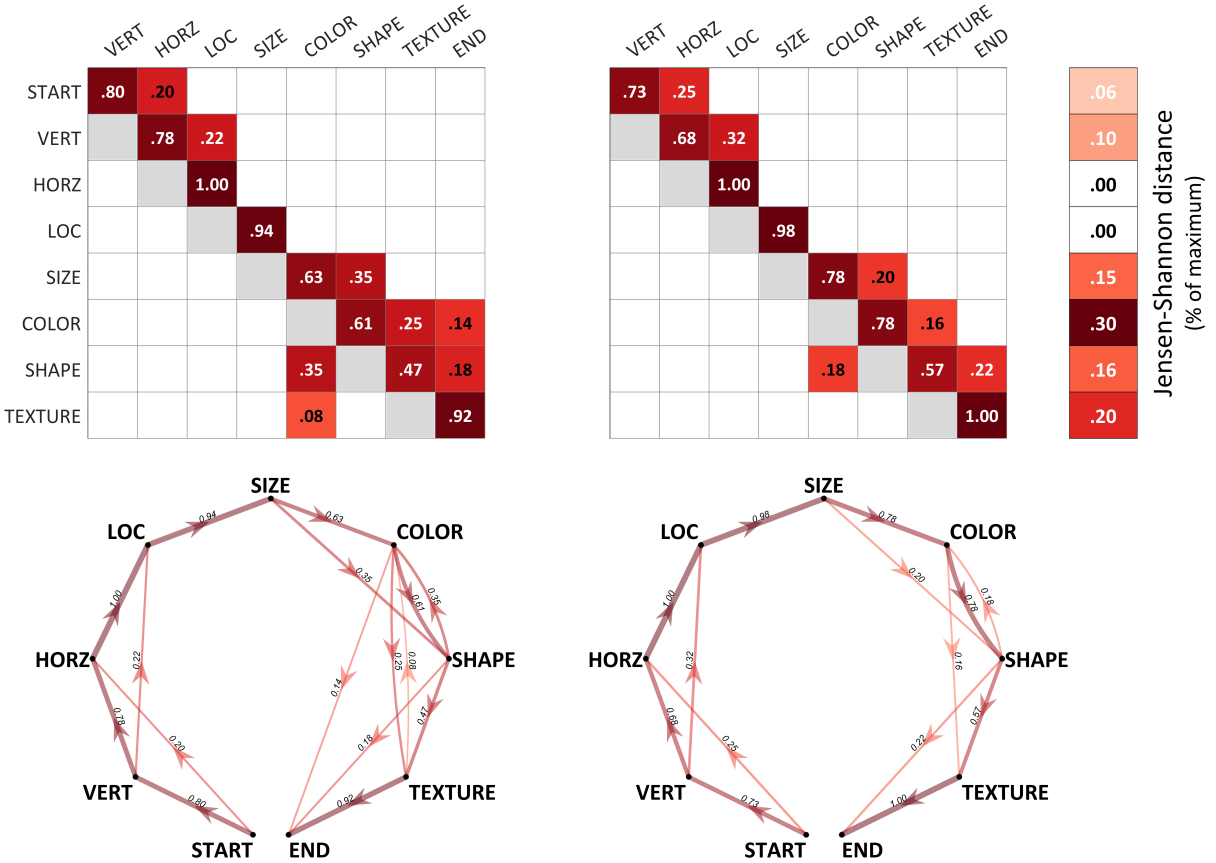

Fig. 9. Examples of transition probability matrices and Jensen-Shannon distance. (Top) Transition probability matrices. (Top-right) Jensen-Shannon distance (proportion of theoretical maximum) for each row. (Bottom) directed graph representation of Markov chains.

Social distance and syntactic distance exhibit a substantial degree of excess mutual information, in comparison to their randomly permuted counterparts. Fig. 10A shows the cumulative densities of mutual information from the empirical and randomized data on a 5-round analysis scale. The distributions describe the mutual information estimates from all player-pairs. The difference between the empirical and random distributions shows that syntactic and social distance have more redundancy than expected by chance.

Excess social-syntactic mutual information is present on all analysis scales (Fig. 10B), peaking at an analysis scale of 2 rounds. It should be noted that the scale-related fluctuations in mutual information and discontinuities in the cumulative density functions result from the invariance of initial window offset for a given analysis scale, which is introduced in the coarse-graining procedure. In other words, the estimated probability distributions are influenced by the decision to set the first window at round 1. This may be addressed in future analyses by estimating distributions with all initial window offsets 1-n for analysis scale n, and by implementing random permutations prior to the coarse-graining procedure.

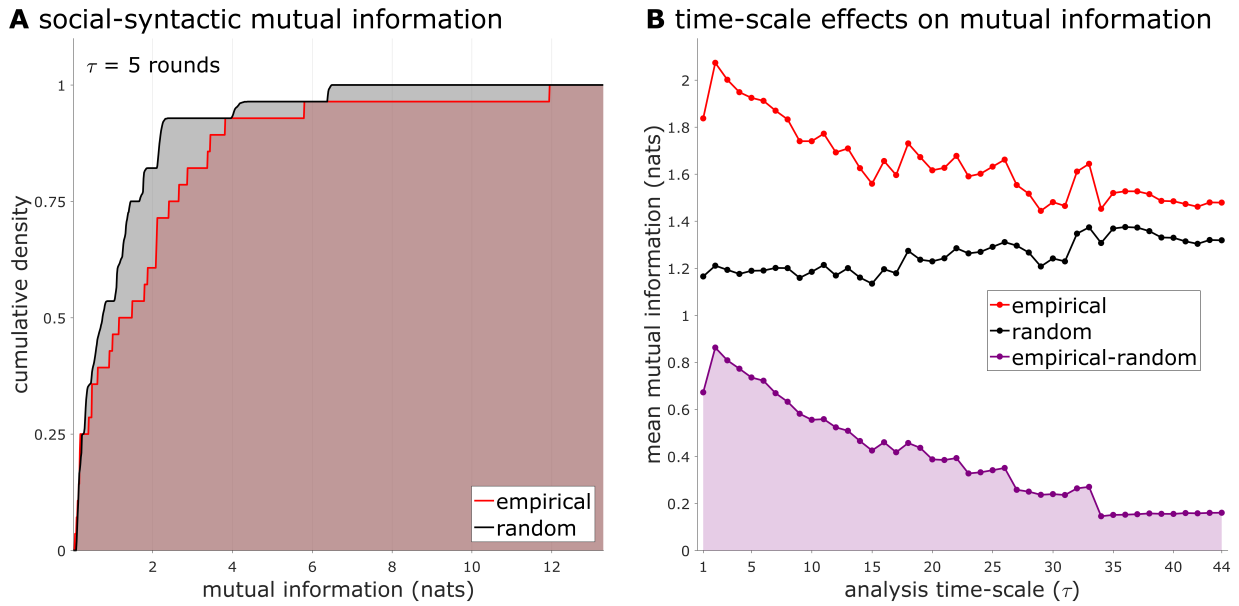

Fig. 10. Social-syntactic mutual information. (A) Cumulative density function of empirical and random mutual information, estimated on an analysis scale of 5 rounds. (B) Mean mutual information of empirical and randomly permuted data for analysis scales of 1-44 rounds. The difference between means, i.e. the excess average mutual information, is shown in purple.

3.4 Interaction-scale analysis of vowel quality

Given the strong evidence above for the influence of social forces on speech behavior on a macroscopic scale, it is reasonable to model these influences on a spatially and temporally smaller scale, namely the interaction scale. In this perspective, each game is considered an asymmetric interaction in which the linguistic systems of the receiver experience forces through perceiving the states of linguistic systems of the giver. The effects of these interaction forces are estimated by comparing observations of receiver behavior before and after the interaction.

To focus on interaction scale phenomena, we define a behavioral space for each player-pair: the dimensions of these *pairspaces* depend on the specific variable under consideration. In the case of vowel quality, each pairspace is defined by recentering the auditory spectrograms for each vowel category/player and then conducting separate principle component transformations for each vowel category and player-pair. Trajectories of the first two principal components of the vowel *boc* in each pairspace are illustrated in Fig. 11A.

The reader should note that the motivation for recentering player data is that listeners are known to perceptually normalize talker variation, which includes anatomically-related variation. Indeed, sex-related variation in vocal tract length appears to be the largest factor in vowel quality distance, and this can be factored out by recentering. Note that recentering is not always necessary or desirable, but to the extent that listeners compensate for such variation it is a useful transformation. It is also important to note that conducting analyses in principal component pairspaces is a strategy to improve distance estimates by emphasizing the variation that is specific to a given player-pair.

Each interaction can be conceptualized as shown in Fig. 11B. For an interaction in round *i*, S_{i-1} is the pre-interaction state of the receiver, G_i is the state of the giver during the interaction, and S_{i+1} is the post-interaction state of the receiver. The vector from S_{i-1} to G_i, which represents the difference between the receiver and giver, is the interaction force vector. The vector from S_{i-1} to S_{i+1} is the displacement vector, which represents the effect of the interaction on the receiver. Several aspects of this conceptualization are of interest: relations between the magnitudes of the components of the interaction force and displacement vectors indicate that the vowel quality of the giver had an influence on vowel quality of the receiver. Similarly, any non-uniformity in the distribution of the angle θ indicates an influence of the interaction; under a null hypothesis that interactions have no effect, the direction of the displacement vector should be uncorrelated with the interaction force vector.

In the analyses presented below, only simplex interactions are considered. These are sequences in which a player is a giver, then a receiver, then a giver (i.e. a *GRG* series). The behavioral change in the player from round *i*-1 to *i*+1 is assumed to result solely from the force experienced as receiver in round *i*. Complex interactions occur, for example, in a *GRRG* series, which involves two consecutive experienced interactions. The nature of the force experienced in the second consecutive game as a receiver is less certain because the behavioral state displacement from the previous round is more difficult to estimate. Furthermore, as a strategy to diminish the influence of behavioral variability in the analysis, vowel quality states were estimated with three-round half-Gaussian windows, using a backward-decreasing window for S_{i-1} and G_i, and a forward-decreasing window for S_{i+1}.

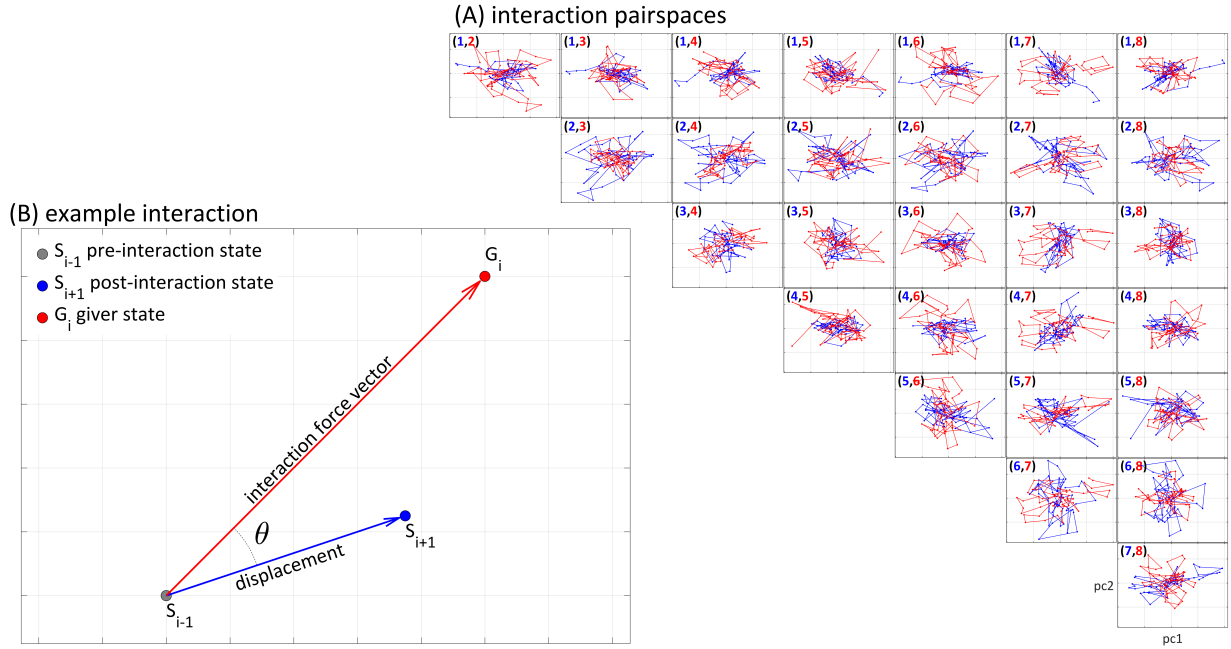

Fig. 11. Interaction pairspaces and example of interaction variables. (A) All 28 interaction pairspaces for the vowel in *boc*, first two principal components. (B) Example of states and vectors associated with each interaction.

Inspection of the distributions of angles between the interaction and displacement vectors shows a strong tendency for the displacement to be biased in the direction of the

interaction force. This is exemplified in Fig. 12, which shows semipolar histograms of θ for each vowel category and for data grouped across vowel categories. The red lines indicate uniform distributions under the null hypothesis that the directions of the interaction force and displacement vectors are independent. Kolmogorov-Smirnov tests for differences between empirical and uniform distributions were significant for all vowel categories and for the pooled data.

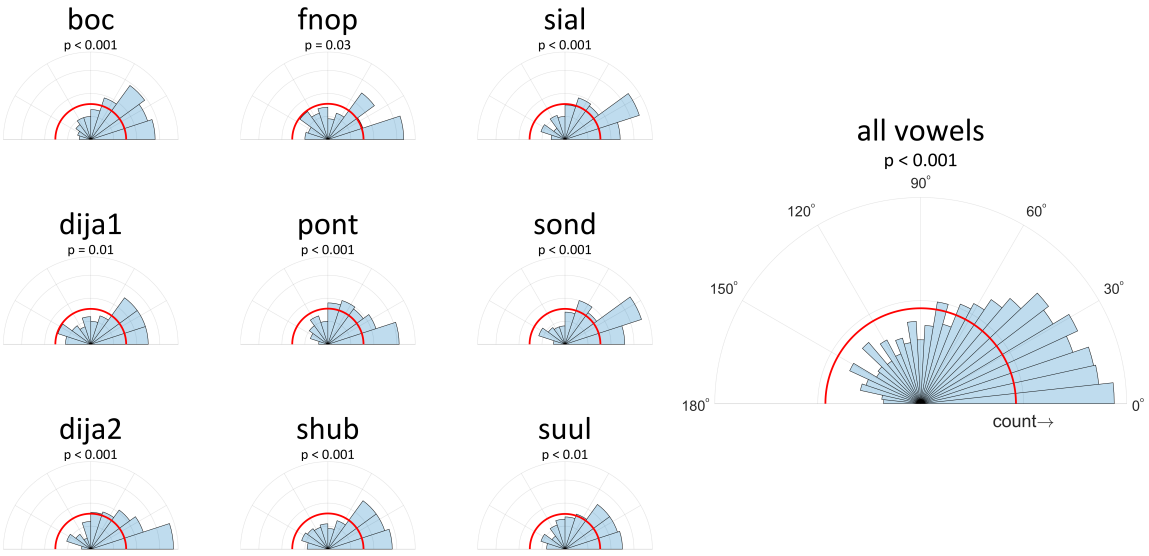

Fig. 12. Distributions of angles between the interaction and displacement vectors. The null hypothesis uniform distribution is shown with a red line. The *p*-value of each Komolgorov-Smirnov test for departure from a uniform distribution is shown.

The bias toward small θ show that the interaction force vector and displacement are not independent. Regarding the social modulation hypothesis, the prediction is that higher rankings (of the giver by the receiver) will be associated with smaller θ. Fig. 13 shows regression lines, R^2 and p-values for each vowel category: the regressions of θ and ranking show almost no linear effect. Furthermore, spline fits of the two variables suggest no consistent nonlinear relation. The apparent lack of a relation between ranking and θ may be attributable in part to a high degree of nonlinearity and player-specificity in the relation between teammate preference rankings and effects on behavior; these and other possible issues are considered in Section 4.

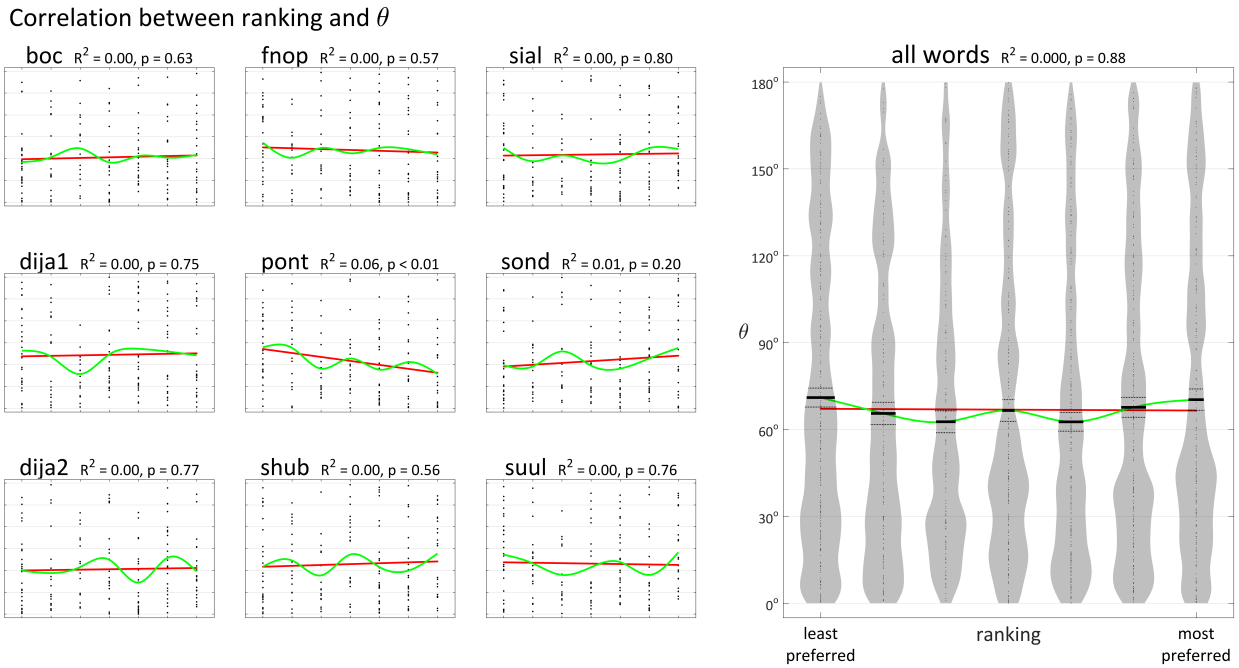

Fig. 13. Correlations between ranking and θ. R^2 values of each linear regression are shown, along with p-values. Linear regressions are shown with red lines, spline fits shown with green lines.

As shown in Fig. 14 and Fig. 15, there are significant correlations between the individual components of the interaction force vectors and the displacement vectors, showing that interaction forces influence the receiver, without necessarily being socially modulated. The R^2 of these correlations range from 0.03 to 0.34 for the vowel categories, and are 0.12 and 0.13 for the first and second principal components over all vowels. In all cases the correlations are statistically significant, although they may not seem to be very substantial. However, given the complexity of the systems responsible for vowel production, and noting the numerous factors which influence vowel production, accounting for about 15% of the change in vowel production in a receiver from an estimate of the state of the giver is somewhat more impressive.

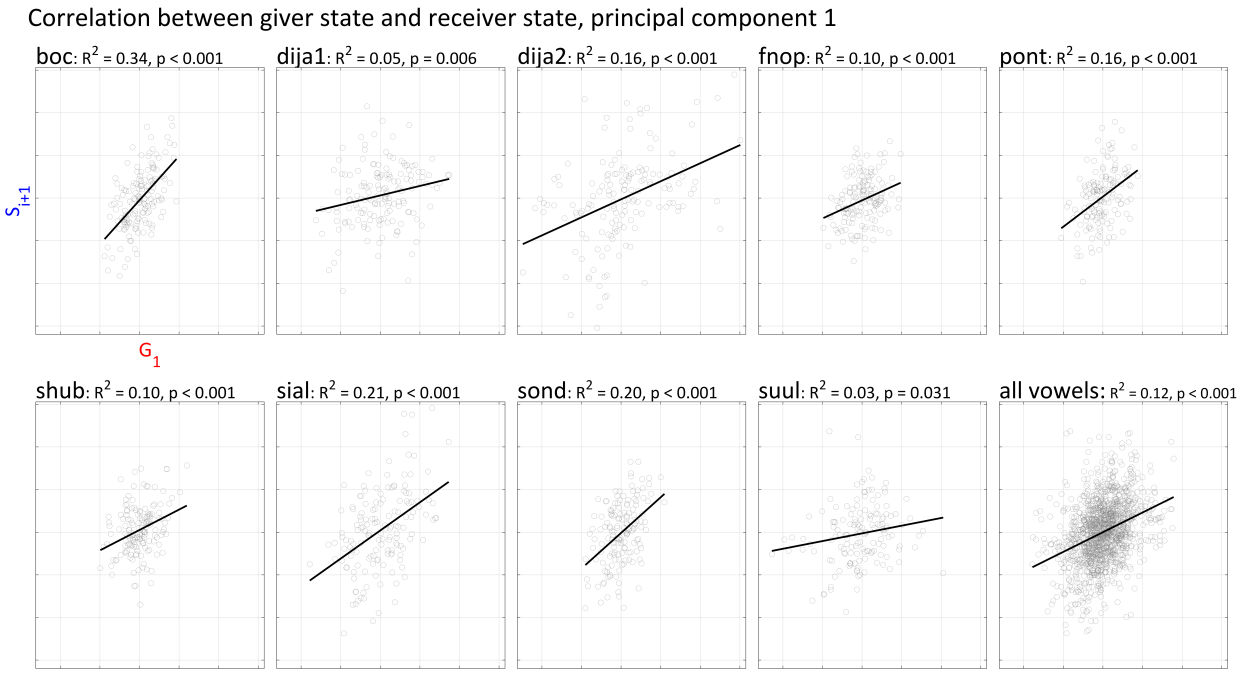

Fig. 14. Correlations between the first principal components of the interaction force and displacement vectors. R^2 and p-values of linear regressions are shown.

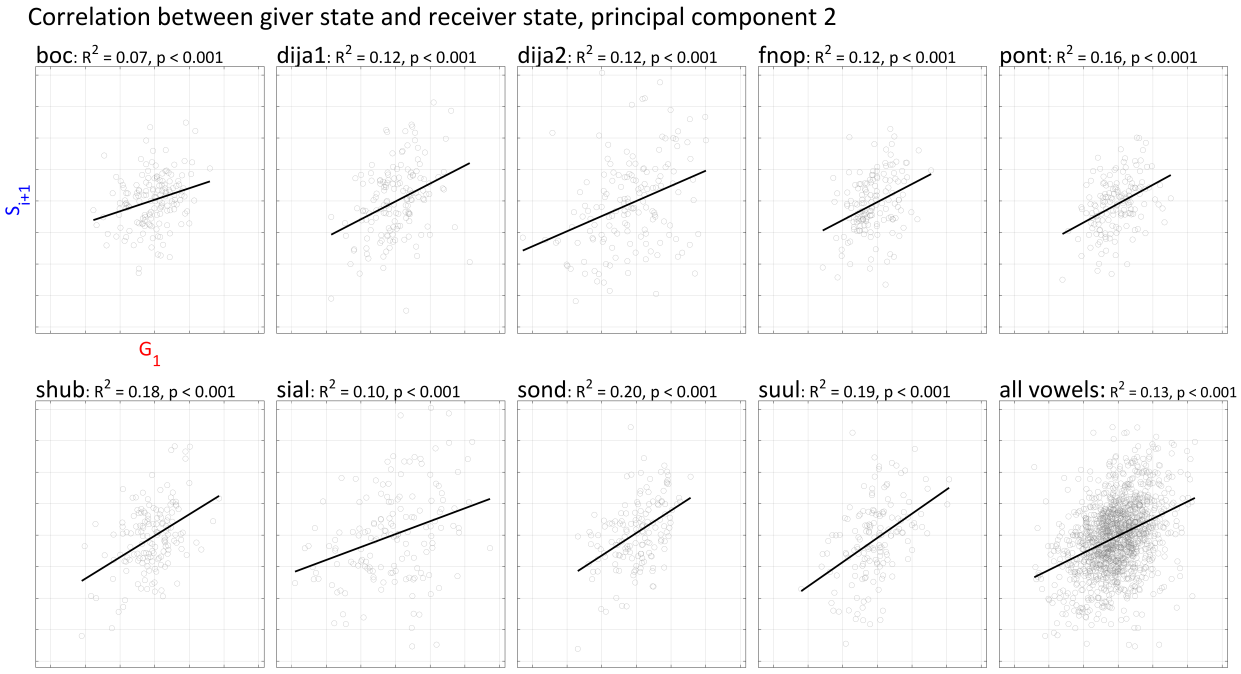

Fig. 15 Correlations between the second principal components of the interaction force and displacement vectors. R^2 and p-values of linear regressions are shown.

Under the hypothesis that social distances modulate interaction forces, the linear model in the preceding analysis should be significantly more predictive when a ranking-by-displacement interaction term is included. However, comparisons of these two models generally showed very small gains in R^2. Although this might call into question the social

modulation hypotheses, a more savvy conclusion may be that the effect of the interaction force vector and giver ranking is not suitably described by a linear-interaction model. We discuss other possibilities in detail in Section 4.

3.5 Fluctuations and discontinuities

An important consideration in analyzing the experimental data is that non-stationarity is manifested not only in the central tendencies of social and linguistic variables but also in their variances. In other words, the magnitudes of the fluctuations in social and linguistic systems change over the course of the experiment. In this regard, an interesting contrast appears to exist between the fluctuations in syntactic and vowel states: like ranking variance, syntactic entropy decreases with a quasi-exponential decay over the experiment, but vowel quality variance does conform to a decay-like pattern.

Fig. 16A shows three behavioral variance time series, all of which were estimated with 10-round moving-windows. The syntactic entropy time series is the average over players and word classes of the entropies of the single-round transition probability distributions in overlapping 10-round windows. The ranking variance is the average over rankers of the by-rankee variance also on a 10-round scale, and the vowel variance is the average over vowel categories, players, and the first two principal components. The interesting contrast is that vowel variance returns to its early-experiment levels in late rounds (circa round 100), whereas syntactic entropy and ranking variances maintain a global decay pattern. The contrast is also salient when examining spectral coherence between variance time-series, as in Fig. 16B. (Coherence spectra were estimated with a 67-round window, using Welch's averaged modified periodogram method). There is a strong peak in coherence at a period of about 22 rounds between syntactic entropy and ranking variance, but this peak is relatively diminished in the spectral coherence between vowel and ranking variance.

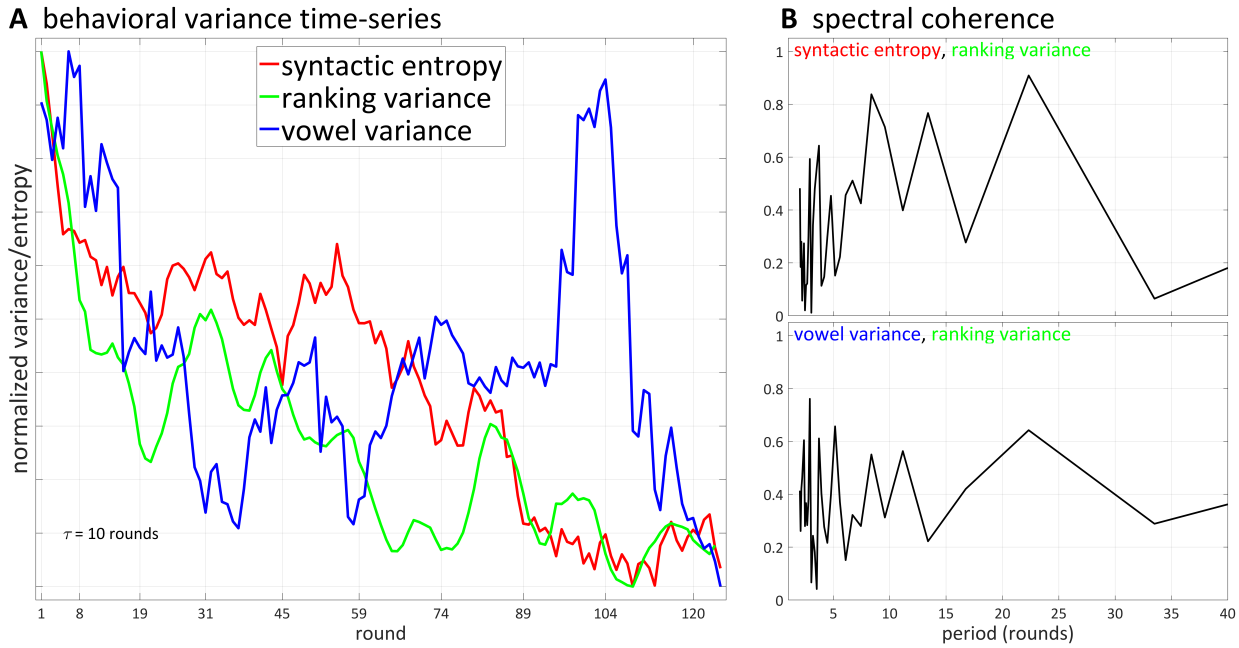

Fig. 16. Fluctuation magnitudes and spectral coherence. (A) Normalized variance of rankings, averaged over participants; normalized variances of vowels, averaged over vowel categories; normalized entropy of transition probability matrices, averaged over word categories and participants. Variances and entropies estimated at each time step with 10-round windows. (B) Spectral coherence of variance/entropy time-series, as a function of period.

Although a comprehensive analysis of behavioral discontinuities is still underway, I provide an example here to support later discussion. For any given state variable, a threshold can be defined to distinguish anomalously large changes from more typical changes. In the case of syntactic behavior, changes in syntactic patterns can be quantified as an auto-JSD (Jensen-Shannon distance), i.e. the distance between the Markov chain transition probability matrices at successive times for the same player, for all analysis scales. Fig. 17 depicts two examples of auto-JSDs estimated over a range of scales (here smoothing is used rather than coarse-graining). Bright vertical stripes in the auto-JSD scalograms are indicative of abrupt changes in syntactic patterns.

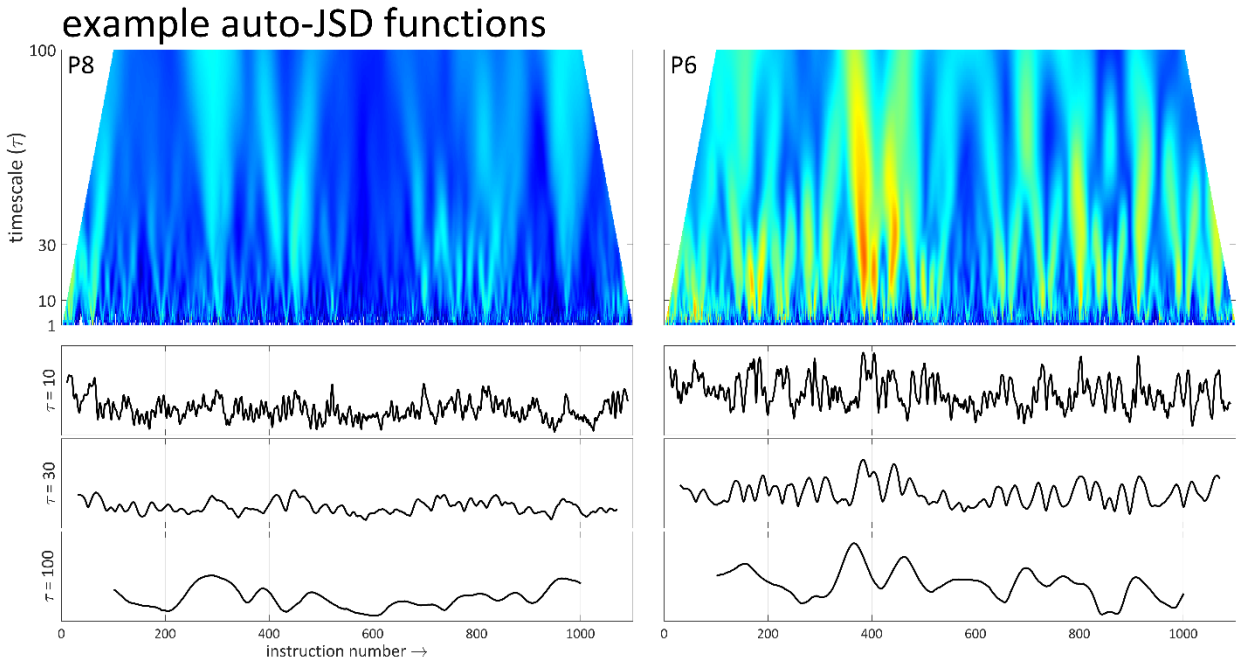

Fig. 17. Examples of auto-JSD scalograms. Auto-JSD can be used to detect behavioral discontinuities. Auto-JSD functions from 10, 30, and 100 round timescales are shown below each scalogram.

4. Discussion

The analyses reported above show that on a macroscopic scale, changes in social network structure and linguistic behavior are correlated. This was observed for vowel qualities, sibilant qualities, and Markov models of syntactic behavior. These macroscopic correlations presumably arise from the spatial and temporal integration of biases that are present on the interaction scale. However, linear analyses of modulatory effects of social distance on vowel quality displacement did not show evidence of social modulation at the scale of individual interactions. Here I discuss some of the shortcomings of the analysis and possible remedies.

4.1 Rethinking the interaction scale model

In the context of the interaction scale, we want to model the displacement vector. Because this vector is derived from estimated states at two successive observation times, the model is effectively a difference equation. Suppose that the model takes the following general form for a simplex interaction (i.e. a giver-receiver-giver series):

$$\vec{x}_{i+1} = \Delta\vec{x} = f(x_{i-1}, \vec{g}_i, \sigma, i)$$

In other words, the displacement vector (state change from round i to i+1) is, generically, a function of the past state of the receiver, the interaction force vector g_i, social forces σ, and time. In the case of the simple linear/interaction analysis conducted in section 3.5, the models that were compared are shown below, where σ = r_i is the teammate

preference ranking, transformed and rescaled to the interval [0,1] with higher values corresponding to higher teammate preference.

$$\Delta\vec{x} = b_1\vec{g}_i r_i \quad \textit{vs.} \quad \Delta\vec{x} = b_1\vec{g}_i$$

Hence in the social modulation model, each component of the displacement vector is modeled as the product of the corresponding component of the interaction force vector, the ranking, and a free parameter. The socially modulated model did not to perform better than its counterpart without social modulation. The was also the case when the interaction force vector was replaced with a unit vector, effectively discarding information about the distance between states. Hence we are led to ask why social modulation fails to improve the model?

One obvious problem is that the social modulation effect likely differs across players. This suggests that player-specific parameters are needed. A deeper problem is that the modulating effects of social variables are likely nonlinear. The linear model presupposes that the effect of a one-unit increase in ranking is independent of where it occurs on the ranking scale. Other possibilities are more plausible; for example, perhaps a strong modulation is present only for the two highest ranks, and is negligible for the remaining ranks, i.e.:

$$r_i \geq b_2, \qquad \sigma_i = b_1$$

$$r_i < b_2, \qquad \sigma_i = 0$$

or: $$\sigma_i = \frac{b_1}{1+b_3 e^{-(r_i - b_2)}}$$

Such non-linearity would not be surprising, since the design of the ranking task forced a linear rank-ordering of teammate preference. This linear ordering is highly unlikely to faithfully reflect the actual sentiments of the players regarding their future teammates.

Furthermore, the modulation might not only act as a nonlinear scaling of the interaction force, but could reverse its direction. Perhaps the modulation becomes negative for very low ranking, so that behavioral divergence can occur when the receiver has a negative social relation to the giver. For example:

$$r_i \geq b_{21}, \qquad \sigma_i = b_{11}$$

$$b_{22} \leq r_i < b_{21}, \qquad \sigma_i = 0$$

$$r_i < b_{22}, \qquad \sigma_i = b_{12}$$

Another problem is that the appropriate nonlinear relation between teammate preference ranking and the social modulation factor is likely time-varying. This suggests that each parameter in the model be expanded into two or more parameters, at least one of which describes the time-dependence. Suitable replacements might be, for example:

$$r_i \geq (b_{21} + b_{22}i), \qquad \sigma_i = b_{11} + b_{12}i$$

$$r_i \geq (b_{21} + b_{22}e^{-b_{21}}), \qquad \sigma_i = b_{11} + b_{12}e^{-b_{11}i}$$

The possibilities are many, but we may look to the global fluctuation patterns for some guidance. Notably, in Section 3.5 we saw that ranking fluctuations exhibit a long-term exponential decay. This suggests that the rankings become more influential later in the experiment, since the ratio of a unit change in ranking to the local variance in ranking becomes larger later on. Yet the more parameters we use to describe the time-dependence of the social modulation, the more we run the risk of overfitting the data. This problem becomes even more severe if we allow the time-dependence to be player-specific, and finding a global optimum becomes more challenging.

Further issues arise when fitting data over a set of categories. For example, when modeling vowel quality data, which parameters should be vowel-specific and which should be fixed across all vowels? On the one hand, a model in which social modulation is independent of vowel category is desirable on the basis of parsimony. On the other hand, different vowel categories may have different degrees of social salience, and there are likely forces that derive from physiological constraints and from the organization of perceptual space that exert different constraints on different vowels. These considerations argue for expanding the parameter space to include vowel-category specificity.

In all likelihood, the interaction-scale model of social modulation could substantially benefit from including information from the pre- and post-game surveys. The questions in each survey are shown in the table below; these were answered on a seven-point scale with the endpoints labeled as shown. It is important to note that in contrast to the teammate preference rankings, which provide a full-network social distance measure for every round, the survey questions are specific to a player and teammate in a given round; this sparse sampling of player-pair social information may be useful for the interaction scale analysis, but not for macroscopic analyses.

Pre-game survey

	Endpoint Labels	
How enthusiastic are you to play the game with [teammate]:	very enthusiastic	not very enthusiastic
How enthusiastic do you think [teammate] is to be playing the game with you:	very enthusiastic	not very enthusiastic
How likely do you think you are to win this round:	not very likely	very likely
How good is [teammate] at [giving/receiving] directions:	not very good	very good

Post-game survey

	Endpoint Labels	
How enthusiastic would you be to play the game with [teammate] in the future:	very enthusiastic	not very enthusiastic
How enthusiastic do you think [teammate] will be to play the game with you in the future:	very enthusiastic	not very enthusiastic
How likely do you think you and [teammate] would be to win in the future:	not very likely	very likely
How well did [teammate] [give/receive] directions:	not very well	very well

Incorporating the data from survey responses into the social modulation term is not entirely straightforward. Because there are eight survey questions associated with each round, we can view the responses as an eight-dimensional vector, so that a general model of the social modulation is:

$$\sigma_i = f(R_{1\ldots i}, \vec{S}_{1\ldots i}, i)$$

This general model incorporates the full history of player-player rankings R, the full history of survey responses S, and the current round. Even if the ranking and survey histories are ignored, there are still nine variables plus time in this expression for social modulation. If each of these variables is treated as an independent source of information whose contribution to social modulation is parameterized by player and time, the parameter space becomes hopelessly high-dimensional.

One promising solution to this problem is make use of strong correlations between the survey responses by converting them to principal components. For all of the participants, the first principal component accounts for more than 55% of the variance in pre- and post-survey responses, and for half of these cases the first component accounts for more than 75% of the variance. Hence it seems reasonable to incorporate only the first component for each player into the social modulation function. By combining this information with the teammate preference ranking the social modulation factor may become more explanatory, assuming that an appropriate nonlinear model can be identified. It is further likely that differences between pre- and post-game survey responses contain useful information, and that changes in teammate preference rankings between round *i*-1 and *i* could be useful to incorporate.

An alternative to attempting to reduce the dimensionality of the social information is to implement the model with high-dimensional input in a machine-learning framework, or similarly to employ a deep multi-layer network or reservoir neural network. In these approaches, evaluating the contribution of social information to the model would require training models on subsets of data and evaluating their performance on withheld data. Even though a substantial amount of data was collected, it is not clear whether it is sufficient for this sort of approach; moreover, the nonstationarity of the social relations and linguistic behaviors is potentially problematic for defining training and test subsets.

Another set of challenges in modeling changes in speech behavior involves factors which have unknown degrees of independence from social mechanisms. These include potential covariates such as speech rate, pitch register, hyperarticulation, and various paralinguistic communicative behaviors. If such variables were independent from the mechanisms responsible for social modulation, then it would be sensible to analyze social modulation effects in the residuals of behavioral variables, after those variables have been regressed by the extraneous factors. But problematically, speech rate, hyperarticulation, etc. are not likely to be weakly correlated with the social context: for example, speakers may be likely to speak more quickly when playing the game with a partner when the players are closer socially.

A substantially different approach to modeling social forces on the interaction scale is to focus solely on more abrupt/discontinuous changes in behavioral states. Perhaps thresholds can be used to identify anomalous state changes, and then one can examine

whether social factors have predictive power for the occurrence of a discontinuous state change. This simplified form of analysis relieves the burden of predicting the magnitude of a state change, while maintaining a focus on the interaction timescale. It is worth noting that the anomaly-prediction approach implies a different perspective on the relation between the macroscopic behavior of the system and the microscopic origins of that macro-behavior. We might hypothesize that the macroscopic patterns are in fact attributable to a small set of anomalous events, rather than the aggregate effects of many interactions. Although a daunting prospect from an analytical perspective, a model which combines both anomalies and aggregated interactions may achieve maximal predictive success.

4.2 Physical systems-conceptualization of the speech and social network experiment

The reader has likely noticed an analogy lurking in this manuscript, namely the idea that we can understand speech behavior in ways that are used to understand phenomena associated with physical systems. In this section I briefly describe some of the mappings in this analogy, hoping to convey their potential utility. Recall that generally we aim to understand what we can predict about the evolution of speech behaviors on a timescale of days/weeks; and that our generic hypotheses are that some temporal variation in speech is caused by interactions and that interpersonal social relations modulate this effect. My contention is that even beginning to answer this question and test this hypothesis requires us to think about speech differently.

First, linguistic behaviors are associated with systems, and the states of those systems determine our observations of the associated behavior. For example, there are neural systems responsible for controlling vowel production, and the states of these systems partly determine our observations of vowel qualities. Of course, each of these vowel production systems is composed of a network of microscopic, neuronal systems, but the states of any one of those subsystems is not crucial because the macroscopic behavior—vowel quality state—emerges from the collective behaviors of the neuronal systems. Likewise, observations of syntactic behaviors, sibilant quality behaviors, social behaviors, etc. reflect the outputs of systems with time-varying states. All of these "systems" are analytical constructs that integrate over a lot of microscopic degrees of freedom.

Second, our methods for analyzing systems, and our construction of systems for analytical purposes, depend crucially on analysis space- and time-scales. The scale-dependence of our analyses should be evident from the contrast between the macroscopic analyses conducted in sections 3.1-3.3 and the microscopic, interaction-scale analyses conducted in section 3.4 and discussed in the preceding section. The scale-dependence of system construction is less obvious, but is nonetheless a deep issue. For example, the concept of a social network is not coherent on short spatial and temporal scales, or at least its properties are not readily quantifiable. Indeed, for meaningful characterization of network dynamics, a substantial observation time period and space scale is required. Although not particularly relevant to the current experiment, other examples of a scale-dependent analytical constructs are those of "language" and "dialect": quantitative differentiation of these categories depends strongly on the spatial and temporal scales over which they are analyzed.

Third, by constructing systems analytically, we have implied a surroundings for those systems, and this begs the question of how the surroundings interface with those systems.

Focusing on the social network and on speakers as systems, all of our systems are quite open in a thermodynamic sense. Yet there may be important ways in which they are weakly coupled to the surroundings, and even ways in which they are closed. For example, the social network in the experiment is closed with respect to speakers, since new speakers do not enter the network. Each speaker in the network is strongly coupled to their surroundings by virtue of needing energy flows to support movement and neural activity, and these undoubtedly fluctuate, but what are the influences of the physical apparatus of the experimental surroundings, i.e. the chairs, laptops, lab rooms, etc.? Most likely these couplings impose constraints on speakers, but they do not fluctuate so much. The surroundings that are spatially and temporally more remote from gameplay, i.e. unobserved events and interactions in the lives of the speakers outside of the experimental sessions, may be strongly coupled to their gameplay behavior, but we hope that they are not. In contrast, during gameplay the interactions between speakers correspond to strong coupling, and the hypothesis of social modulation is a hypothesis that the strength of the coupling between speaker-systems is related to an abstract "distance" between them.

Fourth, nearly all of the systems we have constructed appear to be far from equilibrium, at least when analyzed on the experimental timescale. This accords with the observed nonstationarity in central tendencies and variances of nearly all observed states. Yet here a profound and interesting question arises: given enough time, and given sufficiently weak coupling to fluctuations in the surroundings, would behaviors approach a steady state? It is evident from the analyses of fluctuations in section 3.5 that 10 weeks/134 rounds/535 games/10,500 instructions are not sufficient to demonstrate a steady state. Certainly a longer sampling period is required to probe for stability. Yet the fact that syntactic and social distance fluctuations seem to bottom out by the 10^{th} week is suggestive of stabilization, and vowel quality variance initially exhibited a decay-like pattern before its anomalous increase in late rounds.

The observation of transient exponential decays in fluctuations of social and linguistic systems from the beginning of the experiment provokes some speculative extensions of the physical systems analogy. Suppose we view social ranking variance as a proxy for energy flow to linguistic systems; this is not so far-fetched if we assume that various autonomic nervous system functions that regulate arousal and motivation are tied to our social relations and influence speech perceptuo-motor systems. Note also that to some extent we observed a decrease in the entropy of the distributions of linguistic behaviors. These two observations suggest that the temperatures of our systems and their immediate surroundings cool over time, through dissipation, and that those systems adopt a more ordered, glass-like configuration. As with physical glasses, we would expect with more cooling that intermittent reorganizations would occur and would tend to further reduce entropy—unless, for unknown reasons, the social energy flux transiently increases.

There may be little hope of exploring these ideas from speech observed in the natural world, outside of the lab. It seems self-evident that steady states of behavior in the wild are non-existent on all but the shortest timescales. The reason for this is the multiscale nature of the relevant systems and their strong coupling: neurons, populations of neurons, articulatory movements, words, individuals, social networks, dialects, languages—the range of spatial and temporal scales over which these systems interact is vast, and the patterns on any given scale are always emergent phenomena arising from many interactions on smaller

scales. Indeed, natural speech sometimes seems to be *too complex* to study with the traditional tools used to study complex systems.

Yet all hope may not be lost. The finding that some of the behavioral variation in an experimentally controlled context can be explained by social variation is encouraging. More data and more intelligent models—of social mechanisms, of physiological systems, of cognitive systems, and of surroundings—are needed.

References:

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., … others. (1991). The HCRC map task corpus. *Language and Speech*, *34*(4), 351–366.

Bane, M., Graff, P., & Sonderegger, M. (2010). Longitudinal phonetic variation in a closed system. *Proc. CLS*, *46*, 43–58.

Ellis, D. (2009). *Gammatone-like spectrograms* [2009]. Retrieved from http://www.ee.columbia.edu/~dpwe/resources/matlab/gammatonegram/

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

Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26). CRC press. Retrieved from https://books.google.com/books?hl=en&lr=&id=e-xsrjsL7WkC&oi=fnd&pg=PR9&dq=silverman+kernel+density&ots=iwRmnq7HWo&sig=mEloXROWkOK_Y2LkcOD61L2XSyE

Tilsen, S. (2015). Speech and social network dynamics in a constrained vocabulary game: design and hypotheses. *Cornell Working Papers in Phonetics and Phonology*.

Yu, A. C., Abrego-Collier, C., Phillips, J., Pillion, B., & Chen, D. (2015). Investigating variation in English vowel-to-vowel coarticulation in a longitudinal phonetic corpus, *ICPhS 2015*.