Dual-EEG Evidence for a Shared Cognitive Space Underlying Human Communication
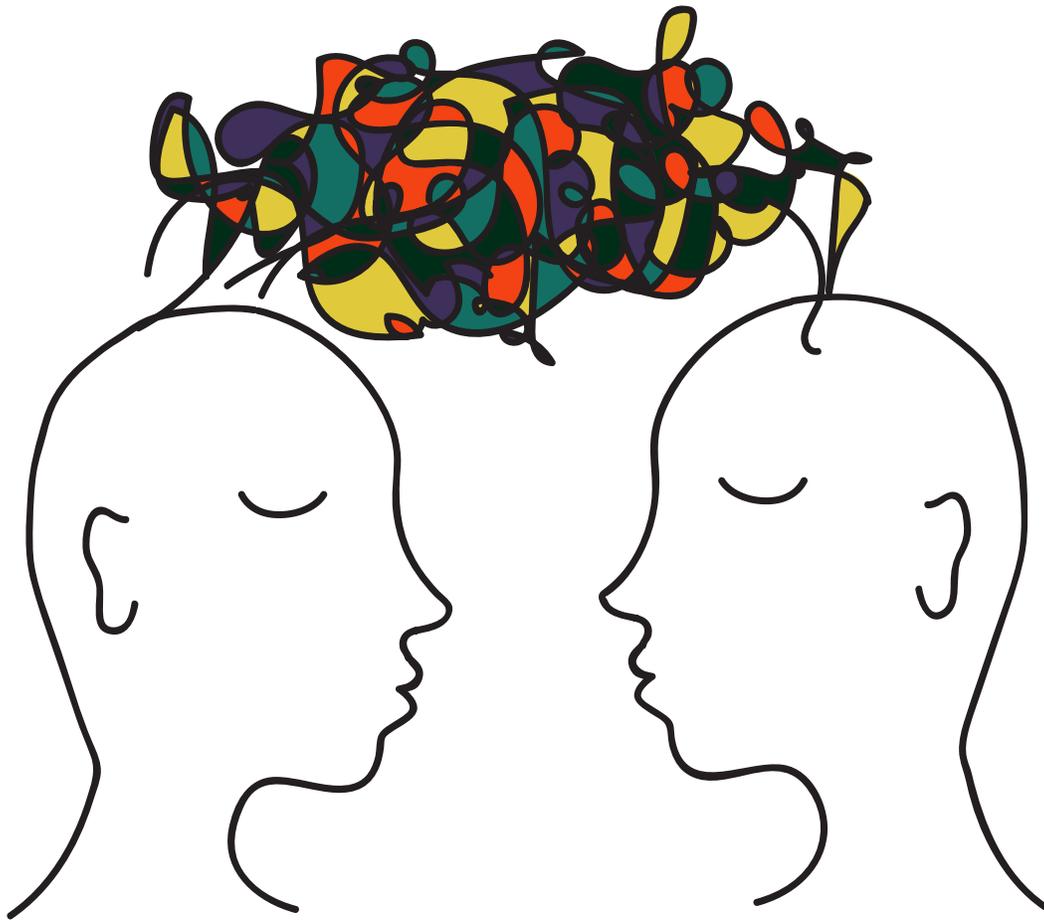
by
Ilina Bhaya-Grossman

A thesis submitted in satisfaction
of the Honors Program option in pursuit of the Bachelor of Arts
in
Cognitive Science
in the
COLLEGE OF LETTERS AND SCIENCE
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Faculty Advisor:
Professor Robert Knight

Co-advisor:
Dr Arjen Stolk

Faculty Reader:
Professor Richard Ivry

Fall 2018

Abstract

There is little known about the neural intricacies and ambiguities of human referential communication; cognitive neuroscience has yet to definitively account for the extreme flexibility with which we employ our words and gestures during everyday social interactions. The fMRI study conducted by Stolk, et al. "Cerebral coherence between communicators marks the emergence of meaning" is one of the few recent projects attempting to explore this area of research, proposing that the right frontotemporal region of the brain is the site of pair-specific cerebral coherence representing the conceptualization dynamics of shared communicative history. Evidence from this study supports the theoretical hypotheses that claim mutual understanding and shared common ground develop through negotiated communication contexts, a process that is both highly adaptive in nature and partner-specific. In this study, we examine direct neural evidence for shared representations in communication, testing the notion that the production and comprehension of mutually understood signals are reliant on a shared cognitive space. Using electroencephalographic (EEG) data acquired simultaneously during live communicative interactions, we show that communicators' neural representations align to one another over the course of the communicative exchange. This communicative alignment process is neurophysiologically supported by broadband high-frequency activity prefrontal cortex. Shared representational dynamics occurred at a rate slower than the occurrences of individual communicative behaviors and were better accounted for by the accumulation of shared knowledge than by sensorimotor properties of the communicative behaviors. These findings challenge current signal-centered accounts of human communication, highlighting a mechanistic process that does not appear to be bound to and triggered by sensorimotor stimuli. Though data from intracranial EEG and fMRI participants performing a similar dual task is still being

analyzed, this project serves as a significant stepping stone that will refine and hone the direction of future experimentation in human communication.

Keywords: social interaction, hyperscanning, conceptual knowledge, common ground, high gamma band activity, representational similarity analysis

Introduction

In our day to day lives, we constantly share our thoughts and emotions with those around us, both verbally and non-verbally, and reasonably expect that our expressions will be met with appropriate and relevant responses. Relevance here refers specifically to an intimate understanding of a shared history and context as this serves as a foundation for most of our future communication. Even if we may not be aware, a majority of our interactions with others are context dependent, take for example, if I am gesturing to my friend Mary, I may be indicating that she should turn around or perhaps I am making a gesture used frequently in the TV show "Friends" that I know we both watch (Tomasello, 2010). Though the act of referencing context in dialogue is a frequent occurrence, there is some debate as to the mechanistic and neural processes responsible for our incredible ability to maintain contextual, rather than simply attentional, synchrony. This thesis aims to marry psycholinguistic theory and modern-day neurophysiology in order to develop a compelling and robust story about the ways in which we communicate.

Discussion of Theoretical Background Literature

Though we are far from understanding all the theoretical intricacies surrounding the process in which ideas, thoughts, and abstract elements of communication are translated into appropriate semantic content based on our current contexts, there exist myriad hypotheses suggesting how this process might occur.

Early prominent figures in the field have proposed that successful communication implies that semantic transfer of information is taking place in the form of a cipher message or conduit (Akmajian, Farmer, Bickmore, Demers, & Harnish, 2017). In the semantic transfer model of communication (that has previously been primarily applied to linguistic dialogue), phonetically encoded linguistic and/or gestural messages are sent between interlocutors and decoded through the application of rules which only the interlocutors are privy to. This theory of communication implies that there is an inherent bijection from the set of linguistic/gestural units to the set of semantic meanings; an intercepting communicator would only need access to this finite set of mappings in order to decode the semantic content flowing between the two original interlocutors. Though this model of communication is intuitive and simple to understand, it does not account for the immense flexibility and ambiguity of everyday communication (Akmajian et al., 2017; Susan E. Brennan, Galati, & Kuhlen, 2010).

More recent theories in the field posit that the translation process between thought and semantic content occurs through a relatively automatic procedure of *interactive alignment* in which the coherence of a *situational model* develops and allows for shared understanding through semantic priming (Pickering & Garrod, 2004). The primary claim is that dialogue is fundamentally different from monologue, requiring that interlocutors be synchronized across an abstract notion of internal situational model. A process of *routinization* occurs when elements or expressions are fixed to a meaning and become routine or easily understandable in the conversation (Pickering & Garrod, 2006). Many examples of linguistic priming, especially syntactic and lexical, have been demonstrated by empirical linguistic studies. In one such example, Cleland and Pickering found that participants were more likely to use a complex noun phrase structure after recently encountering a similar structure than to use a simple noun phrase

structure (Cleland & Pickering, 2003). Additionally, arguments have been made to support the priming of non-verbal situational contexts. These arguments claim that there exist both automatic priming and adaptive strategic components that play a role in co-gestural communication (Wachsmuth, Ruiter, Jaecks, & Kopp, 2013). Criticisms of Garrod and Pickering's account of automatic cognitive alignment argue that the mechanistic procedure that they define does not necessarily address how the process of rapidly updating the situational context takes place nor does it acknowledge the pair-specificity of day-to-day dialogue (Susan E. Brennan & Metzing, 2004).

A third set of theories postulates the existence of *conceptual pacts*. Language processing in this framework exists as a joint collaboration in which two or more interlocutors "share or synchronize aspects of their private mental states and act together in the world" (S. E. Brennan & Clark, 1996; Susan E. Brennan et al., 2010; Herbert H. Clark, 1996). A conceptual pact in this context refers to the deliberately agreed upon relationship between a communicative signal and semantic content. This theory advocates for an alignment process of grounding, a separate process from that of priming**.** Grounding exists as the mechanism to interactively link communicative signals, that on their own have no "contained" meaning, to a mutually understood semantic concept. Grounding occurs through partner-specific negotiations that can be both implicit and explicit and serves to expand the space of successfully linked signals. It is also used to simplify and delineate the links that already exist (Susan E. Brennan et al., 2010; H. H. Clark & Wilkes-Gibbs, 1986). Behavioral evidence for the grounding model of communication include studies that observe acknowledgment as finite positive feedback (Clark & Brennan 1992); elements of acknowledgement, like the paralinguistic utterance 'uh huh' or 'm' (Schegloff, 1982), are used as negotiation tools in the process of grounding. Additional evidence

for the grounding model in linguistic systems is provided by a study in which interlocutors copy their partners' over-informative distinctive descriptions of semantic elements in a matching card game (Clark & Brennan, 1996).

Discussion of Background Neuroscience Work

Despite the breadth of theories proposed to explain context-dependent communication in humans, there have been few studies that examine the procedural relation between the theoretical bases for successful communication and corresponding patterns in neural activity. It is evident, even to those developing behavioral models of communication, that considering neurophysiological data in conjunction with behavioral data will allow for stronger and more holistic hypotheses to be tested (Brennan et al. 2010). A handful of fMRI experiments have attempted to shed light on this area of study by establishing and testing links between psycholinguistic theories of communication and the corresponding expected neurophysiological behavior. Noordzij et al. found evidence of overlapping fMRI brain activity in (non-verbally) communicating pairs (right posterior superior temporal sulcus); overlap was sensitive not to sensorimotor difficulty but rather to the inherent communicative difficulty (ambiguity) of the task itself. This study supports the conclusion that there is a localized brain region responsible for communicative planning, though it does not rule out the potential engagement of other brain regions. Perhaps more importantly, this fMRI study demonstrates that shared activity of communicating pairs does not adhere to the "mirror neuron" hypothesis of communication since activity is regulated by the complexity of the communicative task rather than simply the act of communicating (Noordzij et al., 2009). Stolk et al. found that the cerebral activity produced through the accumulation of shared contextual knowledge in non-verbal communication synchronizes across pairs of participants at a temporal scale that is not time locked to the

presentation of stimuli. This dual-fMRI study provides convincing evidence supporting the argument that alignment (defined through the interactive alignment model) is taking place. Further, the process by which alignment happens in the brain more closely resembles what one would expect to observe in the case of the conceptual pact model (Stolk et al., 2014).

EEG and Representational Similarity Analysis

This experiment was conducted using EEG, a brain imaging modality that measures electrical activity originating from ionic currents in the cortex. Since EEG has a high temporal resolution on the order of milliseconds, this modality affords trial singularity and sensitivity. In contrast to fMRI and other techniques that measure the brain's hemodynamic response, EEG signal is able to capture the idiosyncratic spectral fingerprints of neural phenomena. Ultimately, EEG provides experimenters with unique insights into the neurophysiological mechanisms underlying complex action or thought.

In the last decade, a rash of multivariate analysis techniques have been applied to fMRI, MEG, and EEG data with significant success (Chadwick et al., 2016; Cichy, Pantazis, & Oliva, 2014; James V. Haxby, 2012; James V. Haxby et al., 2011; Kaneshiro, Perreau Guimaraes, Kim, Norcia, & Suppes, 2015; Proklova, Kaiser, & Peelen, 2018; Salmela, Salo, Salmi, & Alho, 2018). At the forefront of this movement into multivariate analyses is representational similarity analysis (RSA), pioneered most recently by Haxby and Kriegeskorte (J. V. Haxby et al., 2001; Kriegeskorte, Mur, & Bandettini, 2008). Representational similarity analysis, at a fundamental level, is simply the comparison of representational dissimilarity matrices (RDMs). An RDM is a square symmetric matrix in which each element represents the dissimilarity between two stimulus responses; the diagonal of an RDM is all zero as each of the diagonal entries represent the dissimilarity of response between a stimulus and itself. Though dissimilarity measures can be

computed in multiple ways, this study defines the notion of dissimilarity to be correlation

distance. Equivalently, representational analysis can compare representational similarity matrices

(RSMs); an RSM with corresponding RDM, X, can be calculated as $1 - X$. RSA is an appealing

method of analysis since it is able to abstract away from the brain imaging modality and

anatomy, instead capturing the similarities between underlying brain activity patterns

corresponding to experimental stimuli (Kriegeskorte et al., 2008). It is also a technique that is

well-designed to relate behavioral and conceptual explanatory models to neurophysiological

data, testing the correlations between latent neural representation of stimuli rather than vectors of

inconsistent spectral data.

　　　RSA, under the larger umbrella of multivariate decoding models, is particularly well-

suited to our paradigm as it allows for the exploration of an abstract neural space that is no

longer reliant on source-specific features of communication (Konvalinka & Roepstorff, 2012).

Through the comparison and analysis of neural RSMs, we were able to investigate the nature of

the shared neural structures that indirectly reflect communicative context and cognitive space. It

is important to note that any mention of neural representation or representational space in the

following sections refers to the manipulation of or interaction between neural RSMs, and does

not necessarily apply to broader definitions of the term 'representation'.

　　　The use of spatially and spectrally resolved RSA allowed us to further speculate about the

locations and frequencies at which the development of shared communicative context could be

taking place. One important technical term to define, related to these ideas, includes high gamma

band (HGB) activity. HGB activity corresponds to the neuronal activity that occurs in the range

of 60 to 140 Hz, though this range is subject to some debate. HGB activity is most commonly

observed in intracranial recordings of electrical activity and has been shown to reflect local

neural spiking activity in the cortex (Lachaux, Axmacher, Mormann, Halgren, & Crone, 2012; Manning, Jacobs, Fried, & Kahana, 2009). However, recent works have also shown that Ca^{2+} dependent dendritic spikes in supragranular cortical layers contribute to HGB activity as well (Leszczynski et al., 2019).

HGB activity has only recently been observed in scalp-EEG studies and there is some evidence to support HGB frequencies also playing a functional role in perception and cognition (Castelhano et al., 2017; Darvas et al., 2010). However, significant studies have cautioned against the assumption that induced gamma band response (iGBR) recorded through EEG is reflecting neural oscillation or representation, rather it may be reflecting saccadic or muscle activity (Yuval-Greenberg, Tomer, Keren, Nelken, & Deouell, 2008). Though our study does not explicitly study the effects of stimulus-locked gamma band activity, it is a crucial reminder to carefully control for spurious activity that may be contaminating the recorded data.

Context and Goals of this Study

This thesis aims to address the following questions:

- Is EEG, as a neuroimaging modality, able to produce substantial evidence to support the notion of shared representation reflecting communicative context? In other words, is the alignment of spectral dynamics more significant between communicating pairs than between pairs that were not communicating?

- How can we effectively characterize the neurophysiological underpinnings of communication, that is, what features do the representations mentioned above consist of and how resolving are these features?

- Once the previous two questions have been sufficiently addressed, can we determine which theoretical framework best explains how the emergence of a pair-specific shared

communicative history manifests in the brain? Does the neurophysiological data support an account of communication that is defined by sensorimotor associations or does it support an account of communication that highlights the dynamics of conceptual agreements?

The Stolk et al. study provides neural evidence for a shared production of context-dependent meaning through interactive grounding processes. What this study fails to explore, however, is the diversity of modality that may be able to capture this phenomenon and the underlying representational structures (RSMs) that define the neural correlates of mutual understanding. The purpose of this thesis is to shed light on the structure of communication and shared context at a neurophysiological level. Furthermore, it will contribute to our understanding of the imaging capabilities of EEG especially in the high frequency bands through the use of representational similarity analysis.

Materials and Methods

Participants

Participants were 42 undergraduates and graduate students ($M = $ 22.4 years; range = 18 - 38 years; $SD = 4.5$; 22 female) recruited from the University of California, Berkeley, participating in exchange for class credit. No participant reported a history of psychological or neurological disorders. Study procedures were in ethical accord with the University of California guidelines and all participants involved in the study gave written informed consent.
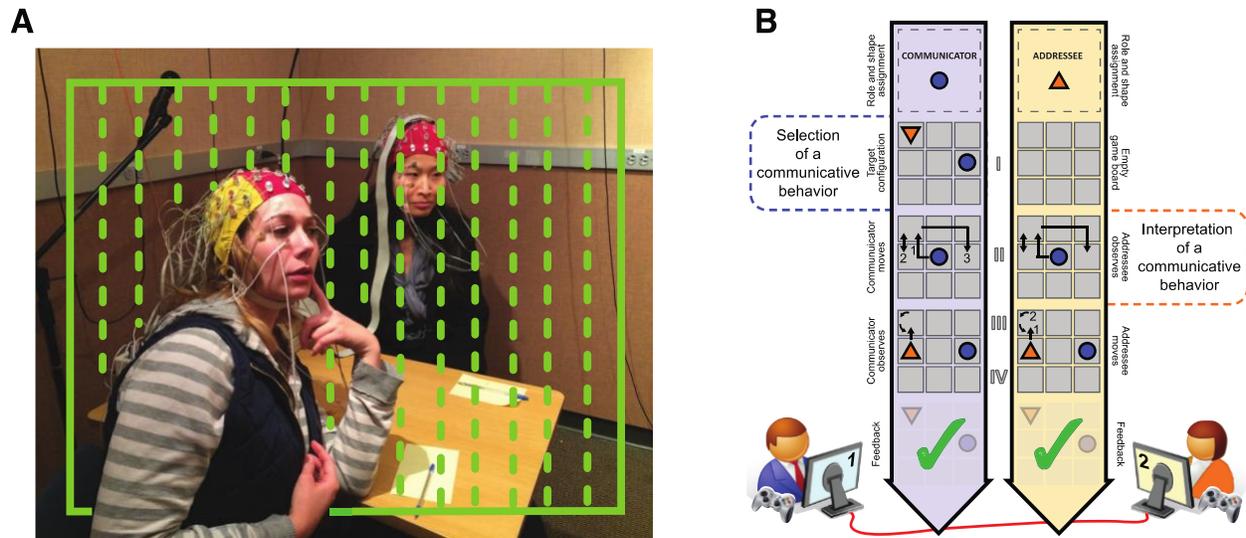
Figure 1: *(A)* Depicts experimental setup; the computer screen is not shown here but is located in front of both subjects (front corresponding to the direction in which they are looking). The green rectangle indicates where a curtain was placed during the experiment to keep subject pairs from communicating verbally or gesturally. *(B)* Shows the discrete segments of a communicative interaction task (figure modified from Stolk et al. 2014). Segment number is located between the two blue and orange arrows.

Materials and Experimental Design

During the task, experimenters recorded EEG signals continuously from both participants with 64 Ag–AgCl pin-type electrode EEG caps (BioSemiTM). Eye movements and blinks were recorded with bipolar vertical and horizontal electrooculography (EOG) through two pairs of electrodes located at external canthi and the infraorbital and supraorbital regions of the right eye. EEG and EOG signals were digitally amplified with the BioSemi Active II system and sampled at 1024 Hz. The pair of participants sat in the same room with a curtain dividing the room in two halves (Fig. 1A). Each participant sat approximately a meter from a screen displaying the communication task. Participants were instructed to abstain from any verbal or visual communication apart from the interactions with the computer screen and controls.

Chronology of an Interaction Event and Details of Communication Task

Since one of the primary goals of this study was to examine the extent to which differing neuroimaging modalities can capture previously established neural results, the experimental design of the tasks used very closely resembled that of the Stolk et al. paper (Stolk et al., 2014). Here, we recap the details of the task and highlight the sections in which this experiment differs from the previous designs.

Similar to the Stolk et al. work, the task was organized into 80 distinct trials, each consisting of one main communicative interaction. For each pair one subject was labeled the Communicator and the other the Addressee. These roles did not alternate between trials (Fig. 1B). At the beginning of each interaction, both players were assigned a token each that only they were able to manipulate during their designated turn in a trial. The Communicator would be assigned a blue token and the Addressee an orange token. Following the assignment of the tokens, the Communicator only was shown a goal configuration in which both tokens (blue and orange) were located at some arbitrary position on the grid and, depending on shape, were facing some distinct direction (segment 1 in Fig. 1B). The ultimate objective of the interaction was for the Communicator to successfully communicate this goal configuration to the Addressee such that at the end of the trial, both tokens (orange and blue) were at locations and orientations identical to that of the goal configuration shown to the Communicator at the onset of the trial. Since the Communicator could only manipulate the blue token, the Communicator must have communicated the Addressee's final token configuration (orange) using only his/her own token (blue).

The Communicator had an unlimited period of time to plan his/her future movement while being shown the goal configuration. The Communicator indicated the end of his/her

'planning period' by pressing the start button (segment 1). On both screens, the Communicator's

token appeared at the center of the screen (the orange token disappears) and he/she could

manipulate the blue token in this 'communication period' by moving it horizontally, vertically,

or rotating it in increments of 90 degrees (segment 2). This 'communication period' would end

after 5 s but could have ended earlier if the Communicator pressed the start button again.

Following the Communicator 'communication period', the orange token appeared at a random

location on the grid that signals the Addressee that he/she now must infer, based on the

Communicators' movements, where to move the orange token. The Addressee would press the

start button to indicate that he/she was ready to begin manipulating the orange token and

effectively end the Addressee 'planning period' (segment 3, not shown explicitly in the

diagram). The Addressee had, at most, 5 s to move and rotate the orange token to the position

that he/she believed was the final goal configuration (segment 4). Once 5 s had elapsed or if the

Addressee pressed the start button again, a feedback screen was presented to both players

signifying to the pair whether the final configuration of the tokens at the end of the trial matched

that of the goal configuration (green check if correct, red cross if not).

Of the 80 trials, 40 presented goal configurations that had not been seen previously in the

task (novel) and 40 presented goal configurations that the participant pair had encountered

during a training session before the experiment (known); novel and known trials were pseudo-

randomly interspersed throughout the task with the difficulty of novel trials increasing as the task

progressed. We ordered the tasks in this way to enable the tracking of mutual understanding over

the course of the experiment. See Fig. 4A for the behavioral dynamics of mutual understanding

in the sample population indexed by communicative success.

Strategy Classification

Communicators used a set of common communicative strategies to relay the goal configuration details to his/her partner. A set of strategies and corresponding descriptions that had been produced in a previous study with an identical communicative paradigm was modified to include all novel strategies encountered in this study. A total of 25 unique signals were observed and categorized by researchers, see Table S1.

Behavioral analysis

Behavioral data were aggregated on a single subject pair automatically through the computer program operating during the experiment. Custom Matlab scripts were developed to analyze and visualize this data over the course of the experiment.

Behavioral Metric and Description Table

Behavioral Metric Name	Description
Communicator Plan Time	Duration of segment 1
Communicator Movement Time	Duration of segment 2
Communicator Number of Moves	Number of moves recorded in segment 2
Addressee Plan Time	Duration of segment 3
Addressee Movement Time	Duration of segment 4
Addressee Number of Moves	Number of moves recorded in segment 4
Success	Whether the subject pair was able to successfully reach the goal configuration following segment 4

Table 1: Relevant behavioral data consisted of the above metrics recorded on a trial-by-trial basis. Descriptions of the behavioral metrics refer to the segments labeled in Fig. 1B.

EEG Data Cleaning and Preprocessing

Once neural data had been collected from the 21 subject pairs, standard filtering and re-referencing procedures were performed with the Fieldtrip toolbox in order to preserve data

hygiene (Oostenveld, Fries, & Jensen, 2009). Initial filtering procedures included band-stop filtering at the line-noise frequency and its harmonics (i.e. 60, 120, and 180 Hz) as well as band-pass filtering between 1 and 200 Hz to remove high frequency noise and suppress slow wave fluctuations. A semi-automatic procedure was then designed to identify and eliminate all signal components caused by oculo-muscular activity; the screening procedure consisted of three main steps. First, data (unre-referenced and continuous) was decomposed into spatiotemporally independent signal components using Independent Component Analysis (Makeig, Bell, Jung, & Sejnowski, 1996; Ungureanu, Bigan, Strungaru, & Lazarescu, 2004). Second, components correlated to electrooculographic (EOG) activity (threshold = 0.5) were removed from the signal. Third, spectral content was estimated for each component and components with power spectral curves closer to that characteristic of muscle sources (diminished fall-off or even an increase of content in higher frequencies) rather than that of neural sources (exponential fall-off of frequency content with increasing frequency) were removed from the signal (Yuval-Greenberg et al., 2008). On average, this screening procedure rejected 32.9 components from each Communicator and 33.0 components from each Addressee data (range = 17 - 42). After identifying and rejecting non-neural components, we manually screened the re-composed data using a data browser tool and excluded any remaining contaminated trials from further analysis (6 out of 80 trials on average; range = 0 - 14).
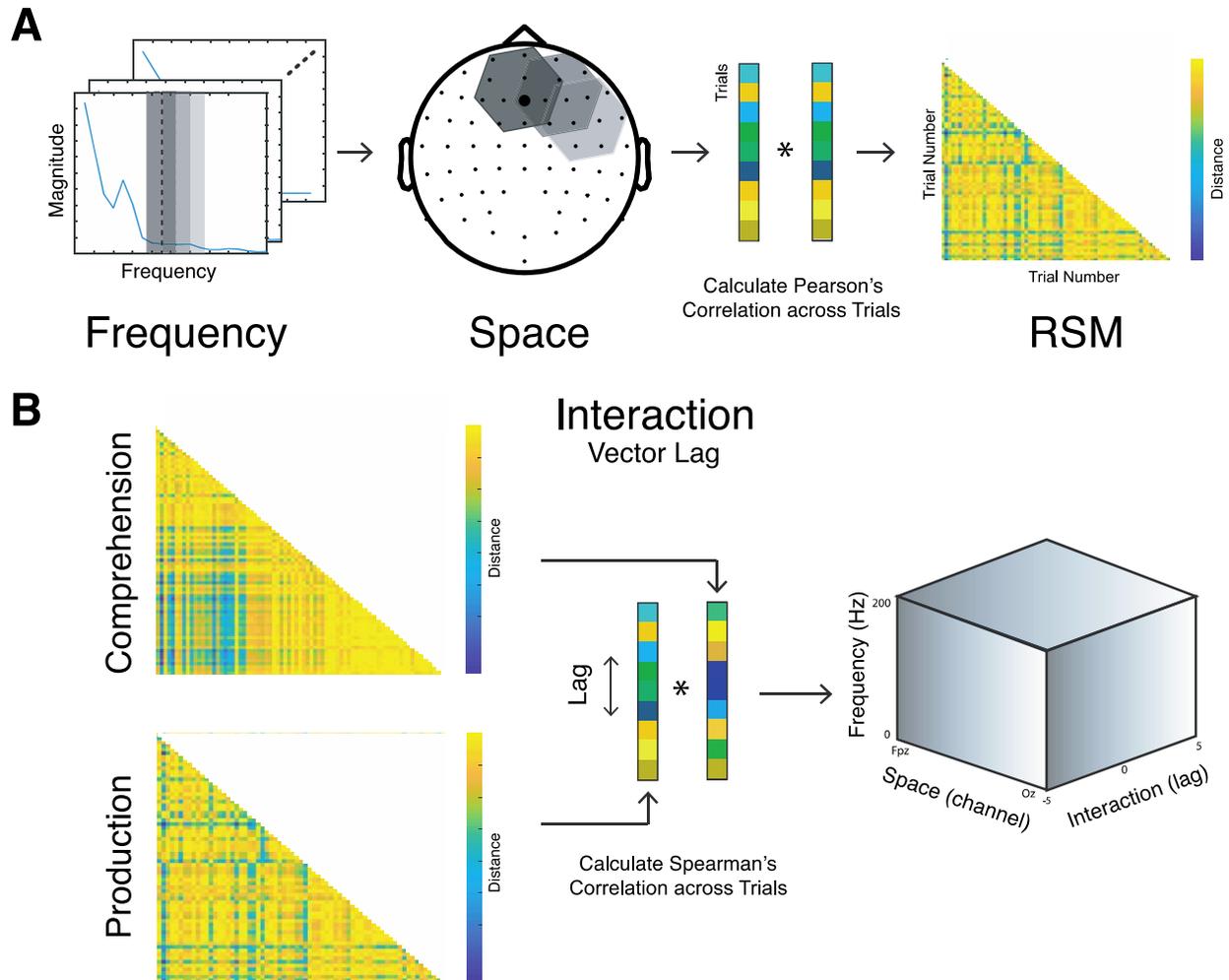
Figure 2: *(A)* Illustration of the process of transforming a series of power spectral densities associated with space and stimuli into an RSM through Pearson correlation along the trial dimension. *(B)* Once RSMs had been computed for each subject, pairwise correlations between Communicator (Production) and Addressee (Comprehension) subject RSMs were calculated iteratively. Time-lagged correlations were produced through a process of shifting one of the two trial vectors and performing Spearman's Rank correlation on the resulting vectors. The lag dimension allowed experimenters to determine the effects of temporal change on average correlation. The final cubic product of a pairwise correlation preserved the spectral, spatial, and temporal dimensions of the communicative interaction.

Representational Similarity Analysis

RSA was used to explore the neural mechanisms of mutual understanding through the comparison of subject specific trial-to-trial resemblances. At a fundamental level, the purpose of this analysis was to determine whether subjects' neural representations of communicative interaction were similar and if so, to what extent. After data cleaning procedures, single subject

RSMs were created through a custom Matlab pipeline (Fig. 2A). Single subject EEG data were segmented into 500 ms windows with 0.9 overlap. Frequency analysis was performed on the segmented data with a hanning taper to mitigate the effects of spectral leakage; the frequency of interest was specified to be in the range of 1 to 200 Hz with a step size of 2 Hz. Resulting power spectra, averaged over the time dimension, produced structures of the shape [trial x channel x frequency]. These three-dimensional structures were transformed into searchlight representational similarity matrices through the Pearson correlation method over the trial dimension (stimuli) as described by Kriegeskorte (Kriegeskorte et al., 2008).

Neural RSM Correlations

Single subject RSMs were compared to each other through a process of pairwise correlation. Spearman's Rank correlation was performed on the lower triangular elements (or equivalently, upper) of the symmetrical RSMs excluding the diagonal (Fig. 2B), as including the diagonal has been shown to result in illusory correlation effects (Ritchie et al. 2016). Distinctions between real pair correlations and cross pair (all sets consisting of one Communicator and one Addressee that did not participate in the experiment together; a total of 420 cross pair correlations were calculated) correlations can be found in results below.

Neural Off-Diagonal RSM Correlations

To further explore the nature of pair-specific neural RSM overlap, we correlated the off-diagonal vectors (the sequence of diagonal elements immediately below, or equivalently above, the main diagonal) of each neural RSM instead of correlating the entire RSM. The off-diagonal vector represents the correlation values between the brain activity patterns recorded from adjacent trials only. In a process identical to that of the neural RSM correlations, Spearman's Rank correlation was performed on the off-diagonal vectors of every subject pair combination.
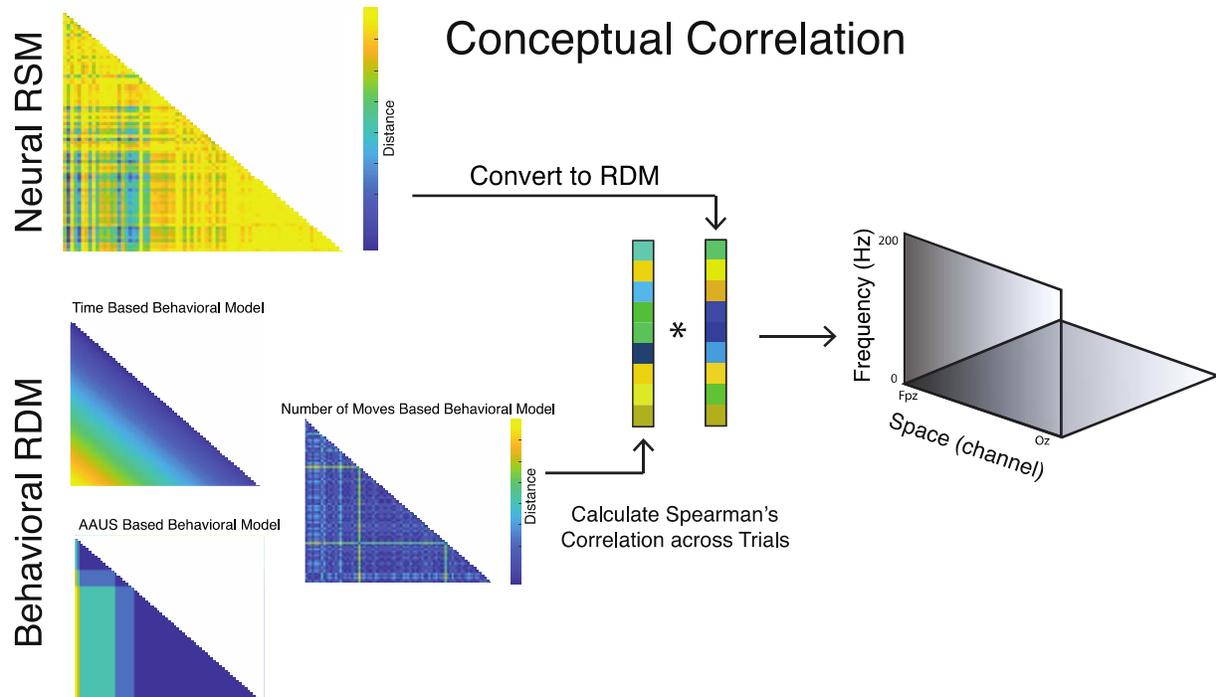
Figure 3: Once neural RSMs had been computed for each subject, they were transformed into RDMs through subtraction. Pairwise correlations between neural subject RSMs and behavioral model RDMs were calculated iteratively. The final product of a pairwise correlation preserved the spectral and spatial dimensions of the communicative interaction.

Neural RSM to Behavioral Correlations

In order to investigate the content of neural representations rather than simply the alignment between communicating pairs, we compared neural RSMs to behavioral based dissimilarity structures. The purpose of this analysis was to determine the extent to which behavioral models could explain or account for the neural representations of a particular brain region (Kriegeskorte et al., 2008). Single subject RSMs were transformed into RDMs and compared to behavioral-based dissimilarity structures through an additional process of pairwise correlations. Behavioral-based dissimilarity structures were produced through the aggregation of behavioral data and subsequent behavior-type dependent manipulation over the trial dimension.

Behavioral-based dissimilarity structures consisted of a pair-specific *number of moves*-based model, an *accumulation of strategy*-based model, and a control *time*-based model. The

number of moves-based model matrix was computed as follows; a cell contained the absolute difference between the number of moves made by the Communicator in each trial. This coarse behavioral model was constructed to resemble the neural representational geometries provided that sensorimotor associations governed neural activity. The accumulation of strategy (cf. the notion of negotiated conceptual pacts) based model matrix was computed as follows; first a correspondence was made between trial number and the accumulation of strategy measure. Accumulation of strategy represents the number of strategies (categorized in Fig. S1) that have been used successfully up till the current trial. A cell in the accumulation of strategy based model matrix contained the absolute difference between the accumulation of strategy measure of each trial. This behavioral model was constructed to imitate the behavioral representational geometries that may develop over the course of a communicative task, expanding with each new negotiated and agreed upon strategy (Diedrichsen & Kriegeskorte, 2017; Kriegeskorte & Kievit, 2013). The time-based model matrix was computed as follows; a cell contained the absolute difference between the trial order numbers. For example, the cell corresponding to trial five and trial eight would contain the number three.

Statistical Inference at the Group Level

Group statistics were performed over the entirety of each type of correlation data to provide descriptive and inferential information; specifically, determining the statistical significance of single subject correlations. Since we were primarily interested in the effects induced by novel trials in comparison to known trials and real pairs in comparison to cross pairs, statistical inferences shown in the topo-plots reported below have been produced by computing the results of independent sample t-tests over either of these two sets. We performed cluster

statistics on the resulting t-values with a cluster alpha of 0.05, in order to determine the locations and frequencies of significant positive and negative clusters (Maris & Oostenveld, 2007).

<div align="center">

Results

</div>

Behavioral Evidence for Shared Knowledge Updating

To gain insight into the shared dynamics of mutual understanding, we first examined behavioral results collected over the course of the experiment. Both the theoretical and experimental analysis of behavioral data, as a precursor to the analysis of neurophysiological data, are considered to be integral to the investigation of lower level neural mechanisms, correcting for the common "reductionist" bias described by Krakauer et al. (Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017).

Behavioral results were analyzed for Communicator and Addressee planning times as well as trial success; these metrics were selected for analysis since they embody some of the most critical features of human communication: duration of communication-related preparation time and whether or not a communicative interaction is successful. Descriptions of these results are categorized into classes of evidence either relating to the pair-specificity of communication or the comparison of known and novel trial dynamics.
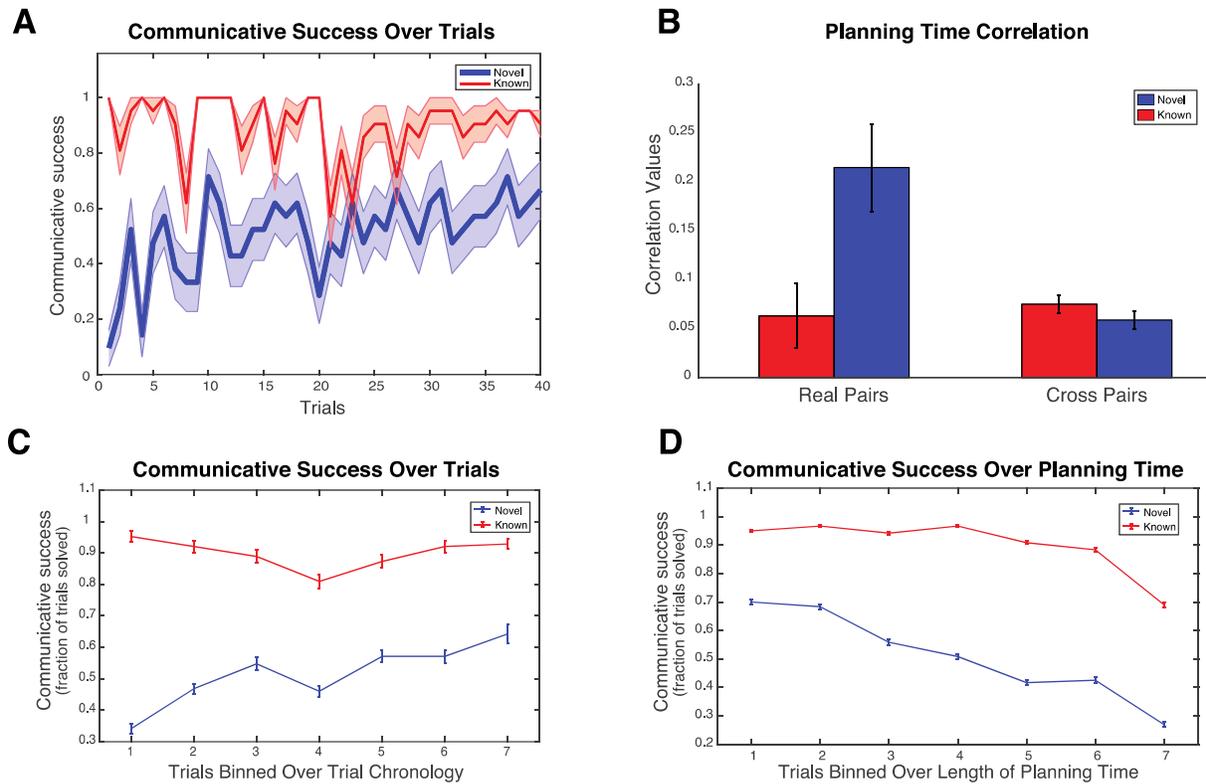
Figure 4: *(A)* Percentage of successful trials over all subject pairs at each trial. Error clouds indicate the standard error of the mean (SEM) for each trial. *(B)* Correlation of plan time values between real and cross pairs as well as between known and novel trials. Error bars indicate the SEM for each correlation grouping. *(C)* Percentage of successful trials at binned trial times averaged over all subject pairs. Error bars indicate the SEM for each set of binned trials. *(D)* Percentage of successful trials corresponding to increasing binned sender plan times averaged over all subject pairs.

Pair Specific Alignment

Correlating Communicator Planning Time with Addressee Planning Time

The results of correlating Communicator planning time with Addressee planning time (Fig. 4B) reveal that pairs of real communicators exhibit a high degree of planning time correlation in novel trials (M = 0.2147, SD = 0.2050). Every other combination of planning time correlation exhibits roughly the same positive average, real communicators in known trials (M = 0.06301, SD = 0.1521), cross communicators in known trials (M = 0.07495, SD = 0.1880), and cross communicators in novel trials (M = 0.05876, SD = 0.1813). Analysis of variance statistic (ANOVA) models support these claims (JASP 2018). ANOVA for Novelty (known or novel

trial) X Group (real or cross pair) revealed a significant effect for Group [$F(1, 878) = 6.093$, $P = 0.014$], Novelty [$F(1, 878) = 5.395$, $P = 0.020$], and a significant interaction between Novelty * Group [$F(1, 878) = 8.281$, $P = 0.004$]. These observations suggest that both members of real communicating pairs found novel trials similarly challenging.

Alignment Dynamics

Visualizing Success over Trials

We examined the average success rate of each trial in chronological order over all subjects (Fig. 4A). In order to demonstrate the general trend effect of time on the percentage of average success, we additionally plotted the average success rate over binned trials (each mark on the x-axis represents the average of approximately six trials, Fig. 4C). In both plots, the average success of known trials stays relatively constant while the average success of novel trials increases as the experiment progresses. Taken together these plots suggest that the exposure to each new novel trial allows subject pairs to develop strategies and contexts that will contribute to success in later novel trials.

Visualizing Success over Planning Time

We examined the average success rate of each set of trials binned over increasing Communicator planning time. This plot illustrates the trend relating duration of Communicator planning time to percentage of average success (Fig. 4D). In both cases of known and novel trials, average success decreased as Communicator planning time increased, indicating a strong relationship between the difficulty of a trial and the effort or time put into communicating the goal configuration.

Neural Evidence for Shared Knowledge Updating

To better comprehend the neurophysiological framework that allows for successful communication, we examined the recorded EEG data through a series of RSA analyses. RSA, as described by Kriegeskorte, was first used to determine pair-specific overlap in the neural representation of the communicative exchange (Kriegeskorte et al., 2008). Expanding upon the traditional whole-matrix approach to RSA analyses, we also correlated off-diagonal vectors that had been extracted from subject-specific neural RSMs. Next, we constructed behavioral model RDMs and compared these models to the neural RSMs in order to explore the content of neural activity and its relation to behavior.

We categorized neural results into classes of evidence either relating to the pair-specificity of communication or the comparison of trial-type dynamics.

Pair Specific Representational Alignment
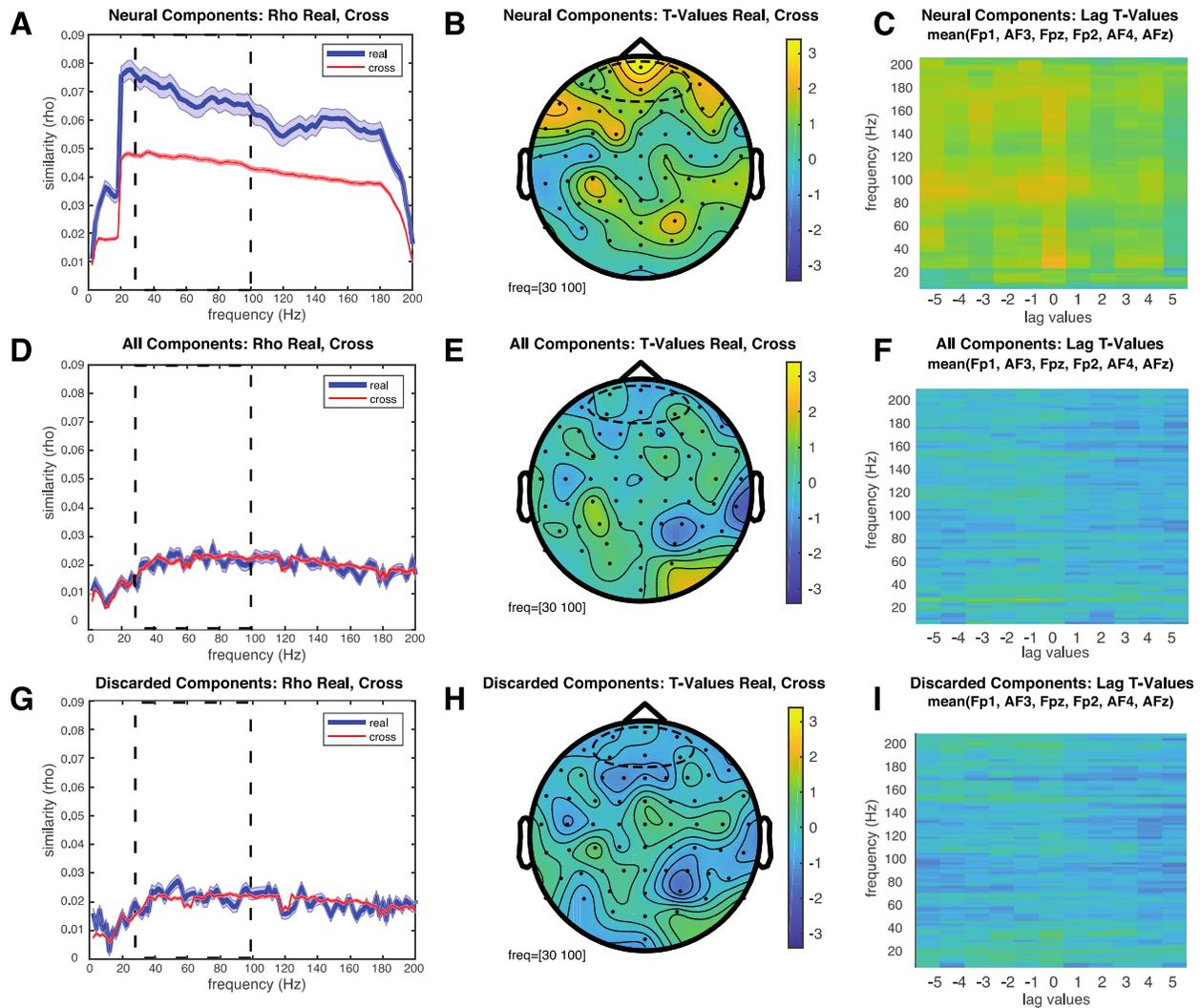
Figure 5: *(A)* Averaged Spearman correlations of RSMs over the entire scalp and all subject pairs, separated into real and cross pairs, produced from the EEG power-spectra once all non-neural contaminant components had been removed. *(B)* Corresponding topographic map of t-values produced by a two sample t-test between real pair RSM correlations and cross pair RSM correlations (in that direction) averaged over the frequency range 30 to 100 Hz. *(C)* Lag t-values across all frequency bins averaged over frontal electrodes, Fp1, AF3, Fpz, Fp2, AF4, AFz. Same scaling legend as the topographic maps. *(D)* Averaged Spearman correlation of RSMs over the entire scalp and all subject pairs, separated into real and cross pairs, produced from the EEG power-spectra with contaminant and neural components included. *(E)* Corresponding topographic map of t-values produced by a two sample t-test, as described above. *(F)* Corresponding lag t-values across all frequency bins averaged over frontal electrodes. Same scaling legend as the topographic maps. *(G)* Averaged Spearman correlations of RSMs over the entire scalp and all subject pairs, separated into real and cross pairs, produced from the EEG power-spectra with only the contaminant components included. *(H)* Corresponding topographic map of t-values produced by a two sample t-test, as described above. Shaded regions around all line plots indicate SEM. *(I)* Corresponding lag t-values across all frequency bins averaged over frontal electrodes. Same scaling legend as the topographic maps.

RSA Correlations: Between Neural and Neural

Subject-specific RSMs produced from uncontaminated EEG power-spectra were correlated across real and cross pairs. Plotting average correlation values across subject pairs clearly illustrates that across all frequency bins, real pairs (M = 0.0583, SD = 0.0027) exhibit higher correlations than cross pairs (M = 0.0390, SD = 0.0019) (Fig. 5A). The positive correlations between cross pairs is compatible with behavioral results, likely tapping into the pair agnostic shared dynamics of the task.

Control Analysis for Oculo Muscular Activity

Subject-specific RSMs produced from all EEG components (including contaminated components) were correlated across real and cross pairs. There exists no distinct difference between average correlation values in real pairs (M = 0.0193, SD = 0.0016) and cross pairs (M = 0.0194, SD =0.0009) across all frequency bins (Fig. 5C).

Subject-specific RSMs produced from discarded EEG components were correlated across communicating and cross pairs. There exists no distinct difference between average correlation values in real pairs (M = 0.0193, SD = 0.0023) and cross pairs (M = 0.0191, SD =0.0011) across all frequency bins (Fig. 5E).
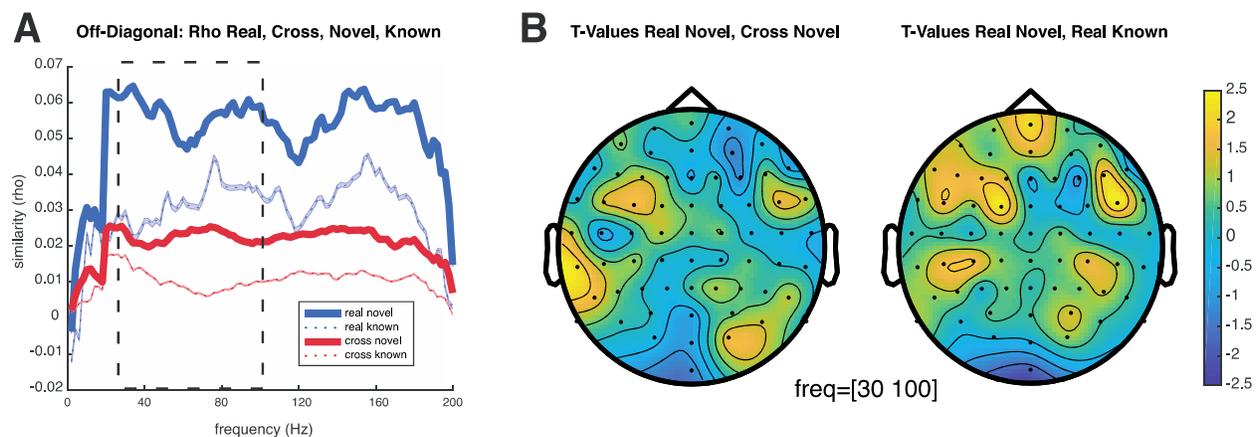
Figure 6: *(A)* Averaged Spearman correlations of off-diagonal vectors over the entire scalp and all subject pairs, separated into real and cross pairs as well as into Known or Novel trial-types. *(B)* Corresponding topographic map of t-values produced by a two sample t-test between novel trial-type real

pair and cross pair correlations as well as real pair novel trial-type and known trial-type correlations (in that direction) averaged over the frequency range 30 to 100 Hz.

RSA Off-Diagonal Correlations: Between Neural to Neural

The off-diagonal vector from subject-specific RSMs (the elements immediately below the main diagonal of the RSM) produced from uncontaminated EEG power-spectra were correlated across communicating and cross pairs as well as known and novel trials. Plotting average correlation values across each subject pair illustrates that that across all frequency bins, real pairs participating in novel trials (M = 0.0518, SD = 0.0045) exhibit higher correlations than real pairs in known trials (M = 0.0294, SD = 0.0049), cross pairs in novel trials (M = 0.0210, SD = 0.0030), and cross pairs in known trials (M = 0.0101, SD = 0.0048) (Fig. 6A).

Representational Alignment Dynamics and Content
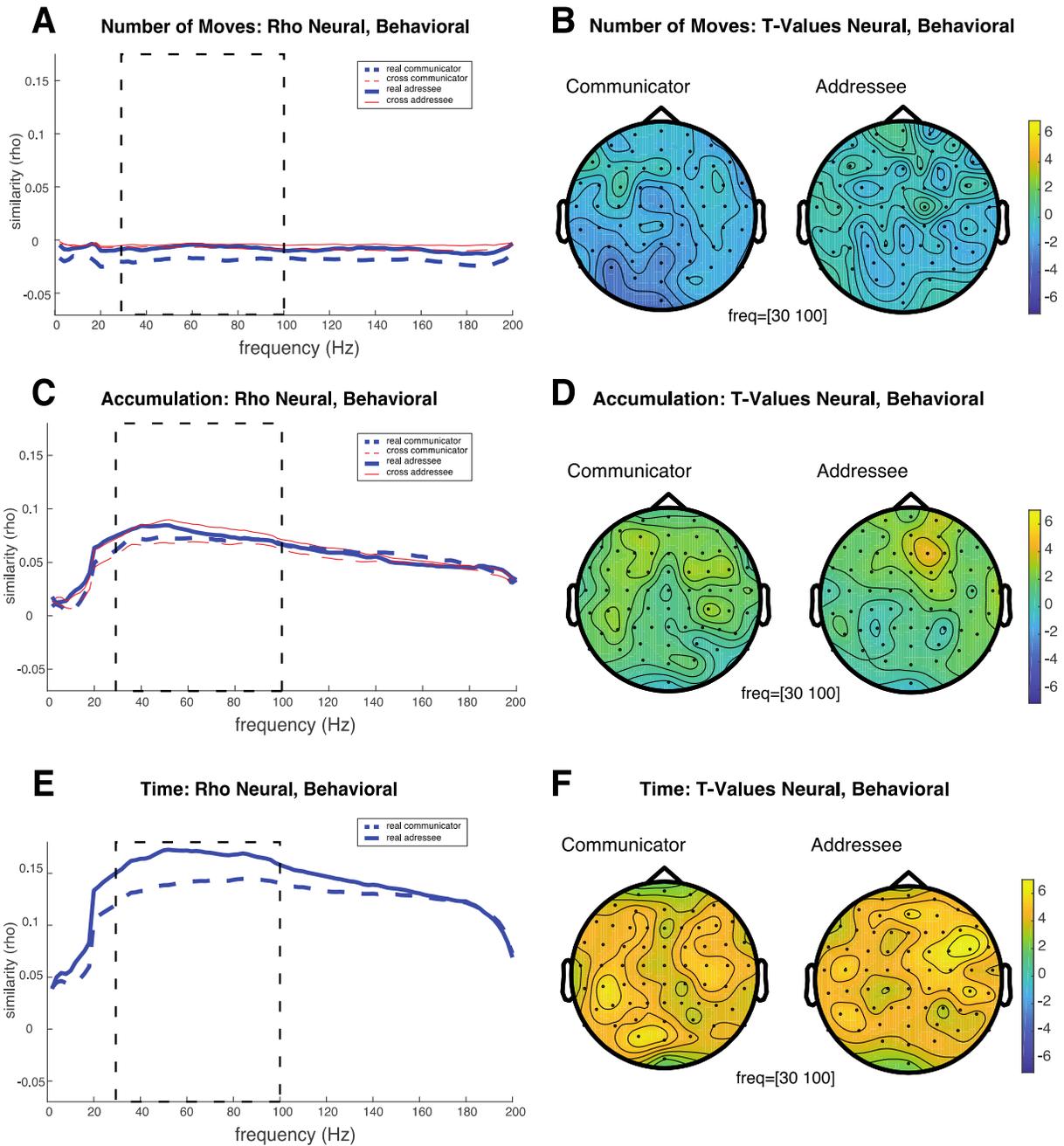
Figure 7: *(A)* Averaged Spearman correlation of neural RSMs to pair-specific behavioral number of moves matrix over the entire scalp, separated into both corresponding behavioral (real) and non-corresponding behavioral pairs (cross) as well as Communicators and Addressees. *(B)* Corresponding topographic maps of t-values produced by a one sample t-test over real pair correlations. *(C)* Averaged Spearman correlation of neural RSMs to pair-specific behavioral accumulation of strategy matrix over the entire scalp, separated into corresponding behavioral and non-corresponding behavioral pairs as well as Communicators and Addressees. *(D)* Corresponding topographic maps of t-values produced by a one sample t-test over real pair correlations. *(E)* Averaged Spearman correlation of neural RSMs to pair

agnostic conceptual time matrix over the entire scalp, separated into Communicator and Addressee correlations. *(F)* Corresponding topographic maps of t-values produced by a one sample t-test over real pair correlations.

RSA Correlations: Between Neural and Behavior

Averaged Spearman correlations between neural RSMs and the pair-specific number of moves based model matrices indicate the overlapping relationship between sensorimotor features of behavior and neural representations in Communicators (M = -0.0190, -0.007 SD = 0.0011, 0.0002) and in Addressees (M = -0.0083, -0.0043 SD =0.0007, 0.0002). Independent one sample t-tests performed on the correlations between all neural RSMs and corresponding behavioral matrices produced 0 significant positive clusters and 0 significant negative clusters.

Averaged Spearman correlations between neural RSMs and the accumulation of strategy based model matrix indicate the overlapping relationship between Addressees (M = 0.0588, 0.0618 SD = 0.0018, 0.0005) and Communicators (M = 0.0568, 0.0519 SD = 0.0063, 0.0011). An independent one sample t-test performed on the correlations between Communicator neural RSMs and the corresponding behavioral matrices produced 1 significant positive cluster containing 1456 points and 0 significant negative clusters. The positive cluster spanned the entire frequency range with the peak spanning 40 to 100 Hz. An independent one sample t-test performed on the correlations between Addressee neural RSMs and the corresponding behavioral matrices produced 1 significant positive cluster containing 1213 points and 0 significant negative clusters. The positive cluster spanned the entire frequency range with the peak spanning 40 to 80 Hz.

Averaged Spearman correlations between neural RSMs and the time based model matrices indicate the overlapping relationship between Addressees (M = 0.1381 SD = 0.0032) and Communicators (M = 0.1225 SD = 0.0039). An independent one sample t-test performed on the correlations between Communicator neural RSMs and the corresponding behavioral matrices

produced 1 significant positive cluster with 6286 points and 0 significant negative clusters. The positive cluster spanned the entire frequency range. Addressee neural RSMs and the corresponding behavioral matrices produced 1 significant positive cluster, containing 6304 points and 0 significant negative clusters. The positive cluster spanned the entire frequency range. It is important to note that since the time based model matrix is inherently not pair specific, it is not possible to correlate a subject's neural RSM to a pair-specific behavioral RSM.

Discussion

Current psycholinguistic accounts of communication outline three primary theoretical models: semantic transfer, automatic interactive alignment, and conceptual pact based (Susan E. Brennan et al., 2010; Pickering & Garrod, 2004). Several prominent studies from the field of cognitive neuroscience have attempted to make sense of these theories in the context of neurophysiological activity and space, determining that the degree of partner-specific alignment of Blood Oxygen Level Dependent (BOLD) signal is linked to the abstract experience of a trial rather than specific perceptual stimuli (Noordzij et al., 2009; Stolk et al., 2014, 2013). These seminal works support the conception of a shared communicative context that is modified by a series of abstract negotiations between communicating partners, a schema that most closely resembles the conceptual pact based model of communication. However, these studies have yet to provide direct neural evidence for the pair-specific representational overlap that would serve as the site of dynamic context. This study, exploring the nature of communication through a dual-EEG experimental paradigm, attempted to both address this salient issue and further strengthen the link between notions of communication in psycholinguistics and in neuroscience.

Behavioral evidence for shared knowledge updating

Returning to the questions posed in the introduction, we first address whether there is sufficient behavioral evidence to support the existence of a shared communicative context. The consistent correlation between real pair planning times during novel trials far exceeds that of real pairs during known trials and that of cross pairs during both known and novel trials. This comparison indicates that there does exist a strong pair-specific pattern in planning time during novel trials and is consistent with previous related studies (Stolk et al., 2014). If we accept that behavior reflects the state of communicative context, this result also emphasizes the distinct character of each pair's dynamic contexts when exposed to novel stimuli.

As for the expansion of a shared context, an intuitive interpretation of the plot Fig. 4A, 4C is that average success rate over novel trials steadily increases as subject pairs develop joint communicative history. This suggests that subject pairs become more skilled at completing novel trials successfully because they negotiate mutually understood strategies for communicating effectively. The high degree of variance across the success rate of novel trials suggests that every pair develops useful communicative contexts at a different rate.

One could argue that the difficulty or ambiguity of the trial strongly affected both planning time alignment and trial success rate trends (Noordzij et al., 2009), since it is likely that novel trials were inherently more difficult than known trials; this would mean that it is difficulty rather than novelty that predicts an increase in pair-specific communicative alignment. However, the assertion that the inherent difficulty of a trial is a more influential modulator of planning time correlation and success rate trends than novelty alone does not necessarily disprove the existence of a shared cognitive space, rather it calls into question the delineation of trial properties and prompts us to further explore other dimensions of the trial data.

Neural evidence for shared knowledge updating

Though we are still in the process of analyzing the neural data further for additional

trends and patterns, preliminary evidence from the analyses performed thus far supports the

existence of a dynamic, pair-specific communicative context. We use this space to address the

questions posed at the beginning of the introduction and speculate about which theoretical

interpretations are best able to explain our results.

We first address whether there is neural evidence to support the existence of a shared

communicative context. It is widely assumed that high frequency range electrical activity (both

gamma and HGB ranges) recorded through EEG is too weak to measure reliably and is heavily

confounded by muscle activity. However, recent studies have shown that through thorough

artifact identification and rejection procedures, neural high frequency activity can be extracted

from EEG recordings and the result shown to be associated with cognitive function

(Muthukumaraswamy, 2013; Onton & Makeig, 2009). Fig. 5A, 5B, 5C, 6A, 6B all show higher

correlation between real pair RSMs than cross pair RSMs; focal positive clusters are localized to

the medial prefrontal region of the brain at frequency range 30 to 100 Hz. Despite the poor

spatial resolution of EEG recording, our findings are in line with previous studies that have

shown the medial prefrontal cortex (mPFC) to be involved in social and predictive cognition

(Alexander & Brown, 2011; Amodio & Frith, 2006). Ultimately, our results indicate that there is

an overlap in subject-pair neural high frequency activity patterns that occurs independent of any

single trial property; this striking effect persists across neighbouring trials (Fig. 5C). Importantly,

from Fig. 5A, we observe that this difference is a sustained broadband effect. Given the results

illustrated in Fig. 5D, 5E, 5F, 5G, 5H, and 5I, we also know that it is likely not a result of

significant artifact contamination. These results suggest that the neural representations

corresponding to the entire communicative exchange are similarly encoded in communicating

subject pairs, that is, to some extent real pairs share abstract patterns of neural activity that cross

pairs do not; these results are neuroanatomically supported by brain activity in the medial

prefrontal areas.

Next, we attempt to characterize the neurophysiological processes that underlie the

expansion of shared contextual knowledge, describing the features of neural representation that

prove most resolving. Fig. 6 illustrates the results of comparing neural RSM off-diagonal vectors

to one another. Fig. 6A demonstrates two important findings: first, real pair off-diagonal vectors

were more closely correlated to one another than cross pairs were; second, novel-type off-

diagonal vectors were more closely correlated to one another than known-type vectors were.

Interestingly, trial-type sensitivity (Novel vs. Known) took precedence over pair-specificity

(Real vs. Cross). In other words, in real pairs, the neural resemblances in adjacent novel trials

were more similar to one another than adjacent known trials. To relate these findings to our

theoretical framework, we may speculate that the strong correlations between real pair neural

activity patterns indicate the existence of pair-specific alignment mechanisms. Stronger

correlations between novel-type off-diagonal vectors suggest that novel trials require

Communicators and Addressees to engage in more explicit negotiation. Subject pairs must

ensure that their knowledge space takes into account the context of the current situation (cf.

internal situational models) in order to communicate successfully; it is this dynamic and shared

process of adjustment that likely allows communicators to produce and comprehend novel

signals (Brennan et al. 2010; Pickering and Garrod 2004).

From Fig. 7A, 7B, we can conclude that the sensorimotor trial property, number of

moves, is not significantly or consistently correlated to neural RSMs in the high frequency

ranges. If we assume that the number of moves behavioral model resembles the neural activity patterns resulting from automatic interactive alignment, then these results challenge prominent brain-to-brain coupling theories (Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012), complicating the model of communication that relies on simple motor priming and mirror neurons (Pickering & Garrod, 2004). It is important to note that we have performed identical analysis on other sensorimotor trial properties (i.e. Communicator movement time) which all resulted in similar outcomes. From Fig. 7C, 7D , we observe that the accumulation of strategy based model matrix is more closely correlated to the neural RSMs than the number of moves based model matrix, the correlation is most pronounced in the frontal region of the brain at 30 to 100 Hz. From Fig 7E, 7F, we can definitively conclude that the time based model matrix is most closely correlated to the neural RSMs in comparison to the two other behavioral models that we tested; the correlation is pronounced throughout the whole scalp at frequency ranges 30 to 100 Hz. These results reveal an obvious property about our dynamic neural representations: the similarity between two single trial neural representations is almost directly proportional to the distance between the corresponding trials in time. Perhaps, the passage of time corresponds to the evolution and expansion of a shared knowledge space; this would mean that the discrete accumulation of knowledge occurs at a rate similar to that of the trial length. However, this interpretation may be reading a complicated problem too simplistically, and undoubtedly calls for further evidence and analysis.

How can we fairly interpret this large set of complex neural data through the lens of communication? It is evident from the correlations between neural RSMs that there are neural activity patterns shared between real communicating partners. It does not necessarily follow, from this alone, that these patterns constitute the shared, dynamic cognitive space responsible for

successful communication. In a study as intricate and unique as this one, it is crucial to consider alternate explanations for the results presented. One important piece of the puzzle that we have neglected to address is the role of working memory in our experimental paradigm. Working memory, a cognitive system responsible for holding and manipulating information, may contribute to shared neural activity patterns assuming that working memory demands for each trial were roughly equal for both Communicator and Addressee. This interpretation of our results may lead us to believe that subjects pairs do not interactively negotiate communicative strategies, but rather the Communicator chooses a single strategy and, over the course of 80 trials, learns how to execute it successfully. However, it has been shown that Communicators take into account the nature of Addressee misunderstanding; Communicators will emphasize or clarify features in a communicative strategy that the Addressee did not understand in previous trials (Blokpoel et al., 2012). This study strongly suggests that there is a negotiation or deliberate interaction taking place, the Addressee is not an entirely passive member of the communicative pair. Another explanation for our results may involve levels of engagement or shared attentional mechanisms driving high correlation values (Dikker et al., 2017). Though it is possible that attention plays a role in correlation, we see from Fig. 5A that our results indicate a sustained effect, one that likely cannot be explained through attentional effects alone. Possible analyses that may clarify some of these confounding issues include exploring the effect of task difficulty on neural activity pattern overlap as well as examining the weight or contribution of each trial to overall neural activity pattern overlap.

Taken in aggregate, these neural results provide preliminary evidence to support the notion of a shared representation space that adjusts in response to novel or difficult experiences; the semantic transfer and automatic interactive alignment models of communication cannot

account for these results. Though our findings can best be explained by the conceptual pact based model of communication, it is clear that these questions are difficult to answer definitively and almost certainly demand further exploration.

Limitations

Scalp-EEG data is widely accepted to be a relatively poor indicator of spatially focal effects. For this reason, a significant limitation of almost all EEG studies is the spatial resolution of specific sources of activity. There are several techniques to improve EEG spatial resolution; some notable examples include using the Surface Laplacian technique, which reduces volume conduction effects thereby improving spatio-temporal resolution (Burle et al., 2015; Carvalhaes & de Barros, 2015), and using dipole source localization, in which measured electrical potential patterns are used to reverse engineer the activity values of an embedded source (He, Yang, Wilke, & Yuan, 2011). Future analysis of this dataset will likely employ these techniques to focus in on specific regions of interest.

Another significant limitation of our study involves the structure of the task paradigm. Other studies that have conducted similar experiments chose a format of the task such that the role of Communicator and Addressee alternate by trial (Stolk et al., 2014, 2013). This modified task format allowed the experimenters to examine the evolution of strategies employed by both members of a subject pair rather than just one. New analyses could easily be fashioned to incorporate *strategy alignment,* a measure that indicates when the communicating pairs (both subjects) have employed the same strategy in consecutive trials; this measure would likely provide unique insights into the mechanistic process of negotiating communicative strategies (Wadge, Brewer, Bird, Toni, & Stolk, 2018).

Interpretational limitations of this study involving the assumptions imposed onto both behavioral and neurophysiological results have largely been addressed in previous sections.

Conclusion

Building upon the framework provided by previous communication studies (Noordzij et al., 2009; Stolk et al., 2014, 2013), the results of this work support the hypothesis that communicative contexts, or the fragments of information we draw on when participating in interpersonal social interaction, are represented in the brain to be both pair-specific and synchronized to abstract notions of task nature rather than simple sensorimotor stimuli. This work uses RSA in high frequency band ranges to make preliminary conclusions about the nature of communication; clarifying which theoretical models are unlikely to adequately explain the observed neurophysiological representations. Future experimentation and analyses of the neural correlates of communication will ideally outline a procedural mechanism that takes into account both theoretical scaffolding and neurophysiological evidence, painting a holistic and interdisciplinary picture of what it means to communicate.

Supplementary Methods

A   **T-Values Between Real and Cross**   **B**   **T-Values Between Novel and Known**
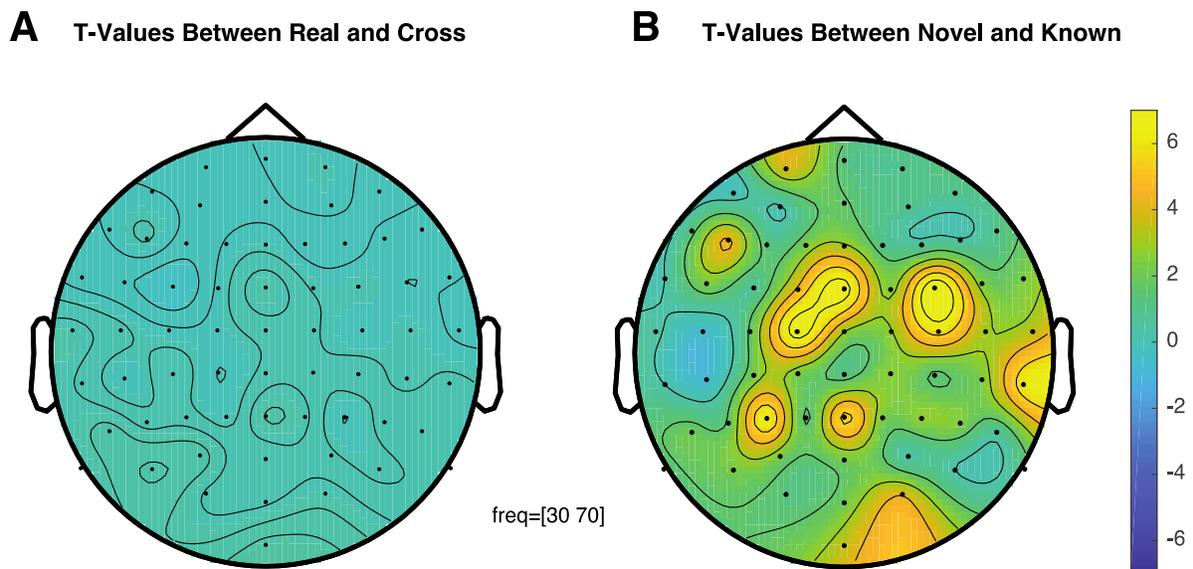
Figure S1: *(A)* Topographic map of t-values produced by an independent samples t-test between the neural power correlations of real pairs and cross pairs. No significant clusters were reported in any frequency bins tested, 1 to 200 Hz (p > 0.05). *(B)* Topographic map of t-values produced by an independent samples t-test between the neural power correlations of novel and known trials. One significant positive cluster was reported, with 13210 points in the high frequency bins, 30 to 100 Hz. The above maps exclude low frequency t-values, showing the average t-values in the range of 30 to 100 Hz.

Neural Power Correlations

As a final step to validate the use of RSA as a means to distinguish features of brain activity in communicating and cross pairs as well as in novel and known trials, we examined the power correlations between Communicators and Addressees. Neural state correlation analysis was performed using cross-correlation of subject-specific time series of cerebral activity estimated independently for each task state. The t-values produced as a result of running a spatially resolved independent samples t-test on the result of the power correlation values yielded the above figure (Fig. S1). No statistically significant clusters were reported in Fig. S1 A, one significant positive cluster was reposted in Fig. S1 B (Oostenveld et al., 2009).
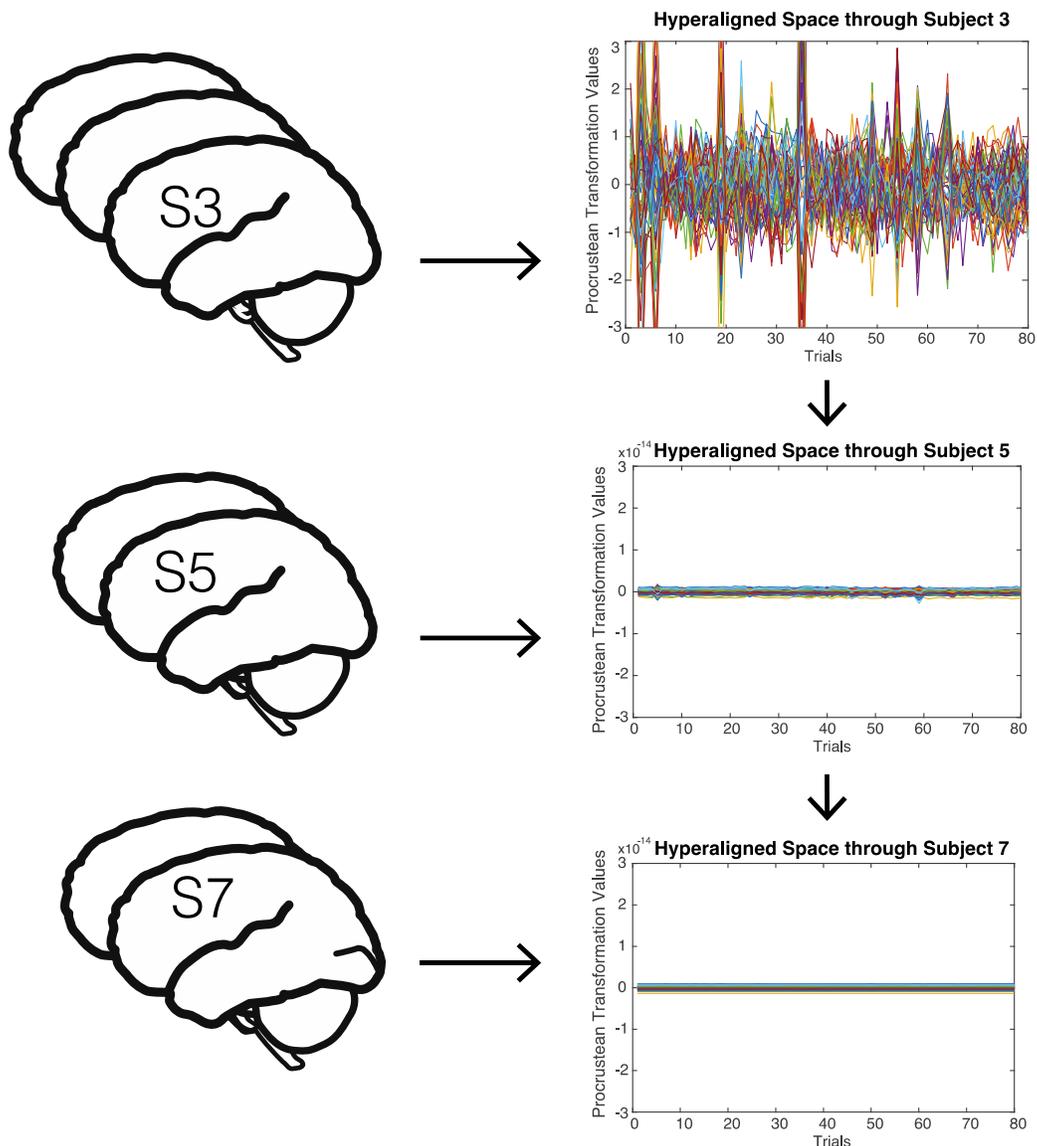
Figure S2: Process and results of running the searchlight hyperalignment algorithm through repeated Procrustean transformation on expanding sets of neural data.

Hyperalignment Methods and Results

In an effort to further explain representational similarity space, we implemented the

searchlight hyperalignment algorithm detailed in the study conducted by Guntupalli et al.

(Guntupalli et al., 2016; James V. Haxby et al., 2011). This algorithm mainly consisted of using

the Procrustean transformation algorithm to derive a common model space that could then be

used to compare single subject model spaces. However, this analysis relies on attentional driven

dynamics; since there were few attentional driven dynamics in the dual-EEG dataset, we noticed

that channel specific temporal dynamics were washed out by the hyperalignment

transformations. This washing-out phenomenon occurred even in the cases in which we

interpolated missing trial data and performed hyperalignment transformations on neural

representations rather than the neural spectral data.

Supplementary Tables

Strategy Classification Discussion and Table

Communicative Signal	Description
A - pause	The Communicator spends more time at the Addressee's target location than at any other visited location on the game board.
B - prolonged pause	The Communicator pauses on the Addressee's target location for periods of time proportionate to the number of rotations the Addressee needs to make.
C - match	The Communicator matches the Addressee's target orientation by rotating in place at the Addressee's target location (the players' shapes are identical).
D - rotate	The Communicator goes to the Addressee's target location and rotates in place however many times required for the Addressee to reach the target orientation.
E - entry	The Communicator enters the Addressee's target location from the direction the Addressee's shape should be pointing.
F - exit	The Communicator exits the Addressee's target location along the direction the Addressee's shape should be pointing.
G - line	The Communicator steps out multiple steps from the Addressee's target location to indicate the direction the Addressee's shape should be pointing.
H - single wiggle from target	The Communicator steps out of the Addressee's target location in the direction of the Addressee's target orientation.
I - multiple wiggles from target	The Communicator steps in and out of the Addressee's target location multiple times in the direction of the Addressee's target orientation (more emphatic version of H).

J - wiggle count from target	The Communicator goes to the Addressee's target location and steps in and out of that location however many times the Addressee needs to rotate to reach the target orientation.
K - wiggle count from target neighbor	The Communicator goes to a square adjacent to the Addressee's target location and steps into and out of the Addressee's target location however many times the Addressee needs to rotate to reach the target orientation. This signal will result in one less visit to the target location than J.
L - wiggle from center	The Communicator visits the Addressee's target location followed by the game board's center, and steps in and out of the central location however many times the Addressee needs to rotate to reach the target orientation.
M - exit from center	The Communicator exits the central start location along the direction the Addressee's shape should be pointing, before going to the Addressee's target location.
N - wiggle elsewhere	The Communicator steps out in the direction of the Addressee's target orientation at a game board location other than the Addressee's target location or the central location.
O - circle target count	The Communicator goes to the Addressee's target location and circles around it however many times the Addressee needs to rotate.
P - circle target direction	The Communicator goes to the Addressee's target location and circles around it in the direction of the Addressee's target orientation.
Q - circle board direction	The Communicator goes to the Addressee's target location and circles along the border of the entire board in the direction of the Addressee's target orientation.
R - circle board count	The Communicator goes to the Addressee's target location and circles along the border of the entire board however many times the Addressee needs to rotate.
S - draw	The Communicator uses a large section of the game board to sketch the Addressee's overall target configuration.
T - opposite	The Addressee goes to the location diagonally opposite the Communicator's end location on the game board, i.e. a complete absence of a communicative signal by the Communicator.
U -	The Communicator goes to the Addressee's target location and rotates to match own orientation.
V – Clicks (cheating)	The Communicator goes to the Addressee's target location and clicks with the controller however many times the Addressee need to rotate.

W	**The Communicator goes to the Addressee's target location and rotates to indicate target location.**
X	**The Communicator circles from the center of the grit and circles around it however many times the Addressee needs to rotate. Clockwise is right turn, counter clockwise is left turn.**
Y	**The communicator first rotates in place however many times required for the Addressee to reach the target orientation before going to the target location**

Table S1: Communicative strategy labels and corresponding descriptions of strategy type. These signal classifications were used to construct the accumulation of strategy behavioral dissimilarity model matrix. Bolded strategy labels indicate all strategy types that were not categorized in previous studies.

Acknowledgements

References

Akmajian, A., Farmer, A. K., Bickmore, L., Demers, R. A., & Harnish, R. M. (2017). *Linguistics: An Introduction to Language and Communication*. MIT Press.

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, *14*(10), 1338–1344.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews. Neuroscience*, *7*(4), 268–277.

Blokpoel, M., van Kesteren, M., Stolk, A., Haselager, P., Toni, I., & van Rooij, I. (2012). Recipient design in human communication: simple heuristics or perspective taking?

Frontiers in Human Neuroscience, *6*, 253.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Chapter 8 - Two Minds, One Dialog: Coordinating Speaking and Understanding. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 53, pp. 301–344). Academic Press.

Brennan, S. E., & Metzing, C. A. (2004). Two steps forward, one step back: Partner-specific effects in a psychology of dialogue. *The Behavioral and Brain Sciences*, *27*(2), 192–193.

Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., & Vidal, F. (2015). Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *97*(3), 210–220.

Carvalhaes, C., & de Barros, J. A. (2015). The surface Laplacian technique in EEG: Theory and methods. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *97*(3), 174–188.

Castelhano, J., Duarte, I. C., Abuhaiba, S. I., Rito, M., Sales, F., & Castelo-Branco, M. (2017). Cortical functional topography of high-frequency gamma activity relates to perceptual decision: an Intracranial study. *PloS One*, *12*(10), e0186428.

Chadwick, M. J., Anjum, R. S., Kumaran, D., Schacter, D. L., Spiers, H. J., & Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(36), 10180–10185.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and

time. *Nature Neuroscience*, *17*(3), 455–462.

Clark, H. H. (1996). *Using Language by Herbert H. Clark*. Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, *49*(2), 214–230.

Darvas, F., Scherer, R., Ojemann, J. G., Rao, R. P., Miller, K. J., & Sorensen, L. B. (2010). High gamma mapping using EEG. *NeuroImage*, *49*(1), 930–938.

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS Computational Biology*, *13*(4), e1005508.

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., … Poeppel, D. (2017). Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Current Biology: CB*, *27*(9), 1375–1380.

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A Model of Representational Spaces in Human Cortex. *Cerebral Cortex* , *26*(6), 2919–2934.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, *16*(2), 114–121.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: the early beginnings. *NeuroImage*, *62*(2), 852–855.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430.

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., … Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–416.

He, B., Yang, L., Wilke, C., & Yuan, H. (2011). Electrophysiological imaging of brain activity and connectivity-challenges and opportunities. *IEEE Transactions on Bio-Medical Engineering*, *58*(7), 1918–1931.

Kaneshiro, B., Perreau Guimaraes, M., Kim, H.-S., Norcia, A. M., & Suppes, P. (2015). A Representational Similarity Analysis of the Dynamics of Object Processing Using Single-Trial EEG Classification. *PloS One*, *10*(8), e0135697.

Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Frontiers in Human Neuroscience*, *6*, 215.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480–490.

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Lachaux, J.-P., Axmacher, N., Mormann, F., Halgren, E., & Crone, N. E. (2012). High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG

research. *Progress in Neurobiology*, *98*(3), 279–301.

Leszczynski, M., Barczak, A., Kajikawa, Y., Ulbert, I., Falchier, A., Tal, I., Schroeder, C. (2019). Dissociation of broadband high-frequency activity and neuronal firing in the neocortex. *BioRxiv*, 531368. https://doi.org/10.1101/531368

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent Component Analysis of Electroencephalographic Data. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in Neural Information Processing Systems 8* (pp. 145–151). MIT Press.

Manning, J. R., Jacobs, J., Fried, I., & Kahana, M. J. (2009). Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(43), 13613–13620.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Muthukumaraswamy, S. D. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Frontiers in Human Neuroscience*, *7*, 138.

Noordzij, M. L., Newman-Norlund, S. E., de Ruiter, J. P., Hagoort, P., Levinson, S. C., & Toni, I. (2009). Brain mechanisms underlying human communication. *Frontiers in Human Neuroscience*, *3*, 14.

Onton, J., & Makeig, S. (2009). High-frequency Broadband Modulations of Electroencephalographic Spectra. *Frontiers in Human Neuroscience*, *3*, 61.

Oostenveld, R., Fries, P., & Jensen, O. (2009). Fieldtrip toolbox. Obtido de http://www. ru. nl/fcdonders/fieldtrip.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, *27*(2), 169–190; discussion 190–226.

Pickering, M. J., & Garrod, S. (2006). Alignment as the Basis for Successful Communication. *Research on Language and Computation*, *4*(2), 203–228.

Proklova, D., Kaiser, D., & Peelen, M. V. (2018). *MEG sensor patterns reflect perceptual but not categorical similarity of animate and inanimate objects*. *bioRxiv*. https://doi.org/10.1101/238584

Salmela, V., Salo, E., Salmi, J., & Alho, K. (2018). Spatiotemporal Dynamics of Attention Networks Revealed by Representational Similarity Analysis of EEG and fMRI. *Cerebral Cortex* , *28*(2), 549–560.

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

Stolk, A., Noordzij, M. L., Verhagen, L., Volman, I., Schoffelen, J.-M., Oostenveld, R., … Toni, I. (2014). Cerebral coherence between communicators marks the emergence of meaning. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(51), 18183–18188.

Stolk, A., Verhagen, L., Schoffelen, J.-M., Oostenveld, R., Blokpoel, M., Hagoort, P., … Toni, I. (2013). Neural mechanisms of communicative innovation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(36), 14574–14579.

Tomasello, M. (2010). *Origins of Human Communication*. MIT Press.

Ungureanu, M., Bigan, C., Strungaru, R., & Lazarescu, V. (2004). Independent component analysis applied in biomedical signal processing. *Measurement Science Review*, *4*(2), 18.

Wachsmuth, I., Ruiter, J. de, Jaecks, P., & Kopp, S. (2013). *Alignment in Communication : Towards a new theory of communication*. Amsterdam, NETHERLANDS: John Benjamins Publishing Company.

Wadge, H., Brewer, R., Bird, G., Toni, I., & Stolk, A. (2018). Communicative Misalignment in

    Autism Spectrum Disorder. https://doi.org/10.31234/osf.io/7nbms

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., & Deouell, L. Y. (2008). Transient

    induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron*,

    58(3), 429–441.