

# Personality Differences Drive Conversational Dynamics: A High-Dimensional NLP Approach

Julia R. Fischer and Nilam Ram

Stanford University

{juliafischer, nilamram}@stanford.edu

## Abstract

This paper investigates how the topical flow of dyadic conversations emerges over time and how differences in interlocutors' personality traits contribute to this topical flow. Leveraging text embeddings, we map the trajectories of  $N = 1655$  conversations between strangers into a high-dimensional space. Using nonlinear projections and clustering, we then identify when each interlocutor enters and exits various topics. Differences in conversational flow are quantified via *topic entropy*, a summary measure of the "spread" of topics covered during a conversation, and *linguistic alignment*, a time-varying measure of the cosine similarity between interlocutors' embeddings. Our findings suggest that interlocutors with a larger difference in the personality dimension of openness influence each other to spend more time discussing a wider range of topics and that interlocutors with a larger difference in extraversion experience a larger decrease in linguistic alignment throughout their conversation. We also examine how participants' affect (emotion) changes from before to after a conversation, finding that a larger difference in extraversion predicts a larger difference in affect change and that a greater topic entropy predicts a larger affect increase. This work demonstrates how communication research can be advanced through the use of high-dimensional NLP methods and identifies personality difference as an important driver of social influence.

## 1 Introduction

Each conversation we take part in is an opportunity for others to influence us and for us to influence others. Among other categorizations, this social influence can be cognitive, such as introducing someone to a new topic to think about, or affective (related to emotion), such as making someone feel more positive by the end of a conversation as compared to before. However, not every conversation will induce changes in interlocutors' cognition or affect. How can we determine which ones will? We

might consider *whom* one talks to as an important predictor of the conversation's resulting social influence. Aspects of an interlocutor's personality are reflected in the language they use (Pennebaker and King, 1999). Thus, conversation can serve as a mechanism through which differences in personality drive social influence.

Large conversational corpora and the computational tools developed for working with linguistic data open new opportunity to test theories of social influence at a large scale. Identifying the mechanisms and drivers of social influence in real-world data is necessary for furthering our basic understanding of how conversations serve as vehicles for cognitive and affective change. This foundational knowledge can help inform downstream theories of applied social influence tasks, such as negotiation (Di Stasi et al., 2024; Glenn and Susskind, 2010) and persuasion (Humă et al., 2010; Wood, 2000).

In this paper, we investigate how differences in interlocutors' personality traits relate to the content and outcomes of a conversation. We focus on how personality differences relate to the topical flow of a conversation, the linguistic alignment between its interlocutors, and the interlocutors' subjective ratings of their affect. Through this analysis, we set forth a method of characterizing conversational behavior through high-dimensional text embeddings, projection into a low-dimensional space, and topic clustering.

## 2 Related Work

### 2.1 Conversation Analysis

Traditional methods of analyzing conversation are primarily qualitative and focus on the manual identification of several instances of a conversational phenomenon (Hoey and Kendrick, 2017; Silverman, 2006). While this approach generates rich data about a specific conversational behavior, it requires the phenomenon of interest to be clearly

defined in advance. The study of the broader topics that emerge during a conversation has largely been done through traditional conversation analysis (Todd, 2011; Yang, 2019). Recent advancements in natural language processing (NLP) tools makes possible the automated identification and quantitative analysis of conversation topics (Wallach, 2006).

Understanding social influence phenomena requires not only identifying conversation topics, but also how interlocutors move among them over time. For example, two interlocutors may begin a conversation in starkly different topics, then become more synchronized in the topics they visit toward the end of the conversation. This could indicate that some form of social influence occurred during the conversation to bring the two into closer alignment. Temporal analysis of conversation topics can be facilitated by defining a topic “space” within which interlocutors move (Templeton et al., 2024). High-dimensional text embeddings have been used to project conversational turns into a semantic space (O’Neill and Finn, 2024; Vakulenko et al., 2018), which can then be used to track the topical flow of a conversation over time.

## 2.2 Personality in Dialogue Systems

Although our work focuses on human-human conversations, personality is relevant in other kinds of interactions, including those with automated dialogue systems, and how they facilitate social influence. In human-machine interactions, a machine system can detect a human’s personality traits through analysis of their conversational behavior (Ivanov et al., 2011; Mairesse et al., 2007). Machine agents can also express personality traits through the language they generate, as detected by human interlocutors (Mairesse and Walker, 2008). Thus, it is plausible that personality functions in human-machine conversations similarly to how it does in human-human conversations.

With their generativity and flexibility, large language models (LLMs) are especially capable of adopting personality traits. In fact, LLMs may already exhibit particular traits without needing to be prompted. Hilliard et al. (2024) demonstrate that LLMs generally exhibit high openness and low extraversion and that newer models with more parameters exhibit a broader range of personality traits. Thus, the degree of social influence in a human-machine conversation may vary based on the model

with which one interacts. LLMs also show promise for adopting personality traits through prompting, then displaying these traits through their text outputs (Jiang et al., 2023; Serapio-García et al., 2023). However, when an LLM agent interacts with another LLM agent, it may struggle to maintain a consistent personality, instead aligning to produce utterances similar to that of the agent they are conversing with (Frisch and Giulianelli, 2024). A better understanding of personality’s role in human-human conversations may help us strengthen the social influence capabilities of LLMs.

## 3 Data

### 3.1 Overview

To investigate conversational dynamics in relation to interlocutors’ personality traits, we make use of the CANDOR (Conversation: A Naturalistic Dataset of Online Recordings) corpus, collected by Reece et al. (2023). The corpus comprises 1656 dyadic conversations that were facilitated through online video chat in the year 2020. The corpus provides multimodal data on a rich set of unscripted, naturalistic conversations in which interlocutors influence and respond to each other without specific constraints or goals.

### 3.2 Participants & Procedure

Participants were 1456 unique individuals ages 19 to 66 years who were located all across the United States (Reece et al., 2023). In brief, participants were matched, based on scheduling availability (without using any demographic information), with other participants to have human-human dyadic conversations online. All participants consented to having their conversation’s audio and video recorded and released for research purposes. At the scheduled time, pairs of participants joined a video meeting and chatted with their conversation partner for at least 25 minutes. They were not given any specific guidelines about what to discuss. Each member of the dyad was compensated up to \$15 for participating in the recorded conversation and completing pre- and post-conversation surveys. Although more than half of the participants engaged in multiple conversations, all 1656 conversations in the corpus were obtained from unique dyad pairs. Our analysis is based on 1655 conversations, after removing one conversation that contained a very small number of utterances. For each conversation in the CANDOR corpus, there are two main data

components: the transcript and the survey.

### 3.3 Transcriptions

The transcript data consists of turn-by-turn transcriptions of each conversation. The 850+ hours of recorded conversations were transcribed and parsed into conversational turns using three different turn segmentation algorithms that differ in how they track when the floor is passed back and forth from one interlocutor to the other: Audiophile, Cliffhanger, and Backbiter (Reece et al., 2023). In brief, Audiophile, the most basic algorithm applied to the data, initiates a new turn each time an interlocutor starts speaking. In contrast, the Cliffhanger algorithm ends the current turn and starts a new turn when the interlocutor reaches a terminal punctuation mark (i.e., a period, exclamation point, or question mark). Thus, whereas the Audiophile algorithm passes the floor back and forth whenever an interlocutor uses a backchannel acknowledgement like "mm-hmm," the Cliffhanger algorithm embeds backchannels within more substantive utterances. Pushing the conceptual meaning of backchannel utterances further—as meaningful utterances that can signal affiliation and understanding—the Backbiter algorithm identifies backchannel responses and separates them from the main transcript into a separate backchannel transcript. Thus, Backbiter produces two transcripts (main and backchannel) that run in parallel. An excerpt from one conversation, as segmented into turns by Audiophile, Cliffhanger, and Backbiter, respectively, is shown in Table 1.

Following our interest in parsing conversational flow, we sought a transcript structure that captured the conversational moves made by each interlocutor as they moved through different topics. As noted by Reece et al. (2023), the Audiophile algorithm provides a rather aggressive division of turns. As can be seen in Table 1, all instances in which the current non-speaking interlocutor uses a backchannel acknowledgement like "okay" or "yeah" are considered new speaking turns. Thus, Audiophile-based transcripts tend to have many very short speaking turns that do not include topical words and also fracture interlocutors' conveyance of thoughts or topics across multiple turns. In contrast, the Backbiter algorithm completely separates the backchannel turns from the conversational flow. We thus chose to analyze conversational flow using the transcripts produced by the Cliffhanger algorithm, as these transcripts provided smoother cover-

age of the topics engaged during each conversation (e.g., via longer turns) while retaining some of the social influence (e.g., rapport) provided through backchannel utterances.

### 3.4 Survey Measures

Survey data were collected in a three-part process. Participants completed a screening questionnaire when enrolling in the study where they provided basic demographic information. Immediately before each conversation, participants completed a pre-conversation survey where they reported on their current affective state. Then, immediately after each conversation, participants completed a post-conversation survey where they reported on a variety of psychological states, including their current affective state, psychological traits, and their perceptions of the conversation partner. The specific measures used in our analysis are described here.

#### 3.4.1 Personality

Participants' personality traits were measured during the post-conversation survey using the Big Five Inventory. Participants indicated their level of agreement (5-point Likert scale) with 15 statements related to the personality traits of openness, conscientiousness, extraversion, agreeableness, and neuroticism. Personality trait scores for each interlocutor were calculated as the average of the relevant item ratings for each of the five dimensions.

#### 3.4.2 Affect and Affect Change

Participants' affective states were measured immediately prior to and again after the conversation. In our analysis we specifically make use of responses to the item, "To what extent do you feel positive affect (e.g., good, pleasant, happy) or negative affect (e.g., bad, unpleasant, unhappy) right now?" that were provided on a 9-point scale ranging from "extremely negative" to "extremely positive". In addition to pre-conversation and post-conversation affect valence scores, we computed for each interlocutor in each conversation an *affect change* score as the difference between the post- and pre-conversation scores, where more positive scores indicate larger increases in positive affect.

## 4 Method

Conversational flow and dynamics captured in the Cliffhanger-based transcripts were summarized in

Audiophile Algorithm	Cliffhanger Algorithm	Backbiter Algorithm
<p><b>A:</b> So are you from like the Chicago area or elsewhere?</p> <p><b>B:</b> Uh, Chicago is about an hour away from us. . .</p> <p><b>A: Okay.</b></p> <p><b>B:</b> from. . .</p> <p><b>A:</b> That’s cool.</p> <p><b>B:</b> I don’t know what the, not, not, not downstate, but like, you know the mm. . . near there. . .</p> <p><b>A: Yeah.</b></p> <p><b>B:</b> basically.</p> <p><b>A: Yeah. Sure.</b></p> <p><b>B:</b> I’ve been to Chicago. My dad, um, lived there for like, you know, he grew up there, he met my mom there, you know?</p>	<p><b>A:</b> So are you from like the Chicago area or elsewhere?</p> <p><b>B:</b> Uh, Chicago is about an hour away from us from. . .</p> <p><b>A:</b> That’s cool.</p> <p><b>B:</b> I don’t know what the, not, not, not downstate, but like, you know, the mm. . . near there basically. I’ve been to Chicago. My dad, um, lived there for like, you know, he grew up there, he met my mom there, you know?</p>	<p><b>A:</b> So are you from like the Chicago area or elsewhere?</p> <p><b>B:</b> Uh, Chicago is about an hour away from us from. . . I don’t know what the, not, not, not downstate, but like, you know, the mm. . . near there basically.</p> <p><b>A: Okay. That’s cool. Yeah. Yeah. Sure.</b></p> <p><b>B:</b> I’ve been to Chicago. My dad, um, lived there for like, you know, he grew up there, he met my mom there, you know?</p>

Table 1: Audiophile vs. Cliffhanger vs. Backbiter algorithms’ turn segmentation of the same portion of a conversation transcript. Backchannel utterances bolded.

a multi-step process that made use of a variety of computational methods and tools. We characterized the flow and dynamics of each conversation by computing several metrics based on both high- and low-dimensional representations of the conversations.

#### 4.1 Text Embeddings: Mapping Conversation in High-Dimensional Space

First, we mapped the conversation transcripts into numerical vectors using a standard set of text embeddings that were developed on other corpora, specifically the SentenceTransformers Python framework (Reimers and Gurevych, 2019). In particular, we used SentenceTransformers’ all-mpnet-base-v2 model (Song et al., 2020) to compute a 768-dimension sentence embedding for each utterance in each conversation. These embeddings thus provide a collection of 768-dimensional dyadic time series that chronicle the turn-by-turn evolution of each of the 1655 conversations—specifically how interlocutors A and B led, followed, and moved with each other through the high-dimensional space.

##### 4.1.1 Conversation Metric: Linguistic Alignment across the Conversation

Using the 768-dimensional vectors, we calculated the time-varying *linguistic alignment*, or the degree of similarity between interlocutors’ language, as

the cosine similarity (1) between successive speaking turns: the first embedding representing an utterance from Interlocutor A and the second embedding representing the consecutive utterance from interlocutor B.

$$\begin{aligned}
 S_c(A, B) &= \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} \\
 &= \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)
 \end{aligned}$$

We then summarized how linguistic alignment changed across each conversation by modeling the cosine similarity scores as a function of time, specifically, turn in conversation. We ran this regression (2) for each of the 1655 conversations separately to obtain three summary linguistic alignment (LA) metrics: *LA intercept*, *LA linear change*, and *LA quadratic change*. We elected to include time as a quadratic polynomial predictor to capture potential nonlinearity in linguistic alignment over time. Note that turns are typically shorter near the beginning of a conversation and become longer as the conversation progresses (Edwards, 2024), so these coefficients should not be interpreted as exactly linear with respect to time.

$$\begin{aligned}
 \text{CosineSimilarity}_t &= \beta_0 + \beta_1 \text{Turn}_t \\
 &\quad + \beta_2 \text{Turn}_t^2 + \epsilon_t \quad (2)
 \end{aligned}$$

## 4.2 Topics: Projection and Clustering Conversation in Low-Dimensional Space

To identify when interlocutors entered and exited areas of the space that might hold specific and human-interpretable meaning, we used nonlinear projection to cast the locations of each speaking turn in the 768-dimensional embedding space into a two-dimensional space. Specifically, we computed the projection using Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2020), a technique that reduces dimensionality while preserving the global topological structure of the data. The minimum distance parameter was set at 0.2 (above the 0.1 default) so that the two-dimensional projection would be more spread out, and thus facilitate topological separation and identification of conversation topics.

After randomly sampling 10 utterances from each of the 1655 conversations and projecting them into the two-dimensional space obtained via UMAP, we identified discernable areas of the space using cluster analysis. In particular, we used Mclust (Scrucca et al., 2023), an R package that fits Gaussian finite mixture models for model-based clustering using an expectation-minimization (EM) algorithm. The optimal number of clusters, selected by minimizing the Bayesian information criterion (BIC) (3), was nine.

$$BIC = k \ln(n) - 2 \ln(\hat{L}) \quad (3)$$

Locations of the random subset of speaking turns in the two-dimensional space and their cluster assignments are shown in Figure 1. This optimal cluster solution was then used to compute cluster assignments for all speaking turns in all 1655 conversations.

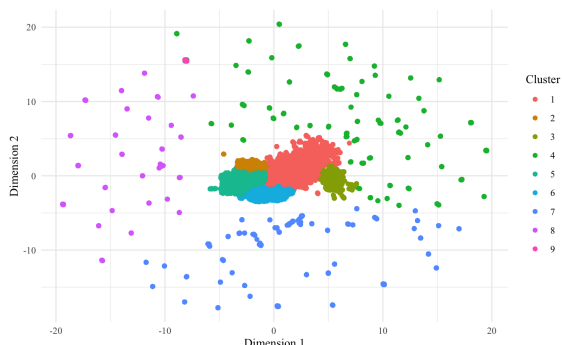


Figure 1: Topic clusters determined by model-based clustering on a thinned set of 10 utterances sampled from each conversation.

Face validity of the topic clustering was examined by extracting the top differentiating keywords from each cluster. After filtering out stop words, as well as words occurring at very low and very high frequencies, keywords for each topic were identified as those word stems with the greatest "keyness," a measure of the differential occurrence of a word stem in one topic versus in the remaining topics (Benoit et al., 2018; Bondi, 2010). As shown in Table 2, the keywords representative of several of the clusters were interpretable as relatively cohesive topics. Other clusters, however, were less interpretable, "catch-all"-type topics.

Topic A	Topic B	Topic C
dog	live	school
mask	citi	survey
cat	famili	class
wear	york	play
vote	california	onlin
trump	move	watch
pet	state	studi
elect	area	money
hair	florida	job
breath	place	prolif

Table 2: Top 10 keywords for three of the nine topic clusters.

### 4.2.1 Conversation Metric: Topic Entropy

Using the cluster assignments for each speaking turn, we introduce and use *topic entropy* as a summary measure quantifying the "spread" of topics covered during a conversation. Specifically, topic entropy was computed for each conversation as the Shannon entropy (4) of the cluster assignments of all speaking turns in that conversation.

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (4)$$

### 4.3 Example Conversations

Illustrations of the conversational flow of two randomly selected conversations are shown in Figure 2. The graphical representations show both how the linguistic alignment (position on y-axis) of successive speaking turns and the topic (color) of each interlocutor's speaking turn changed as the conversation unfolded over time (position on x-axis). Summary measures of the level, linear change, and quadratic change of linguistic alignment across

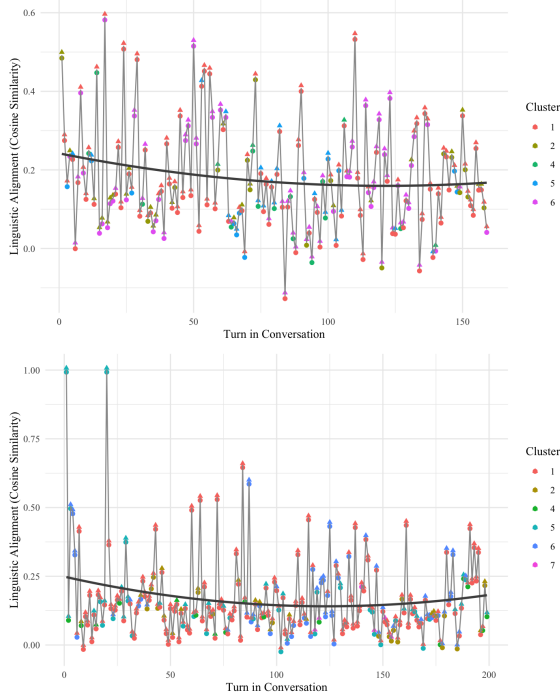


Figure 2: Two example conversations from two different interlocutor dyads. Linguistic alignment (cosine similarity) vs. turn in conversation overlaid with smoothed quadratic curve. Interlocutors’ topic cluster locations indicated by color of symbol (Interlocutor A = circle, Interlocutor B = triangle).

each of the conversations were derived from the fitting of the bold black lines to the cosine similarity time series. The summary topic entropy measure indicates the breadth and relative abundance of the different colors across the time series.

#### 4.4 Survey Features

Alongside the summary measures of conversational flow for each of the 1655 conversations derived from the transcript data (as described above), we also calculated several features about the dyads participating in those conversations from the survey data. For each of the 1655 dyads of interlocutors, we summarized the personality and affect data as follows. For each of the Big Five personality traits, we calculated the mean level of personality (e.g., average of Interlocutor A’s and Interlocutor B’s extraversion scores) and difference in personality (e.g., absolute value of the difference between Interlocutor A’s and Interlocutor B’s neuroticism scores). For affect, we calculated the mean level of pre-conversation affect valence (average of Interlocutor A’s and Interlocutor B’s pre-conversation affect scores); mean level of post-conversation affect valence; mean level of affect change scores

(average of Interlocutor A’s and Interlocutor B’s affect change scores), and difference in affect change scores (absolute value of the difference between Interlocutor A’s and Interlocutor B’s affect change scores).

## 5 Data Analysis

Through application of text embeddings, projection, clustering, and calculations described above, we obtained a set of summary measures that describe aspects of conversational flow (topic entropy; level, linear change, and quadratic change in linguistic alignment), the personalities the interlocutors brought with them into the dyadic conversation (mean and difference in five personality traits), and the affective states of the dyads (pre-post mean and change in affect valence) across 1655 conversations. Descriptive statistics of the summary dataset used for subsequent statistical modeling are shown in Table 3.

Variable	Mean	SD	Min.	Max.
topic_entropy	7.88	0.39	6.47	9.56
LA_intercept_term	0.18	0.02	0.12	0.47
LA_linear_term	-0.09	0.20	-0.72	1.79
LA_quad_term	0.25	0.19	-0.76	1.19
extra_mean	3.15	0.69	1.00	5.00
agree_mean	3.88	0.56	1.67	5.00
consc_mean	3.45	0.69	1.50	5.00
neuro_mean	2.89	0.81	1.00	5.00
open_mean	4.00	0.57	1.50	5.00
extra_diff	1.09	0.81	0.00	4.00
agree_diff	0.90	0.70	0.00	4.00
consc_diff	1.13	0.81	0.00	3.67
neuro_diff	1.27	0.92	0.00	4.00
open_diff	0.89	0.69	0.00	4.00
pre_aff_mean	6.10	1.05	1.50	9.00
post_aff_mean	7.30	1.02	2.00	9.00
aff_chg_mean	1.20	1.16	-3.50	6.50
aff_chg_diff	1.53	1.34	0.00	10.00

Table 3: Descriptive statistics for the dataset used for statistical modeling.

Using these summary data, we investigated the relations among conversation metrics, personality, and affect using linear regression models. Data analysis and statistical modeling were done using version 4.3.3 of the R programming language (R Core Team, 2021). Key findings are reported in the next section.

## 6 Results

### 6.1 Model 1: Topic Entropy as a Function of Personality Differences

In Model 1 we examined if and how the topic entropy conversation metric was related to differences in interlocutor dyads' personality traits, controlling for dyad-level personality trait means. As conveyed in Table 4, larger between-interlocutor difference in openness predicts a conversation with greater topic entropy ( $\hat{\beta} = 0.03, p = 0.04$ ).

Variable	Est.	SE	Pr(> t )
(Intercept)	7.76	0.14	< 2e-16 ***
extra_mean	-0.00	0.02	0.82
agree_mean	0.05	0.02	0.02 *
consc_mean	-0.04	0.02	0.01 **
neuro_mean	-0.02	0.01	0.12
open_mean	0.03	0.02	0.10 .
extra_diff	-0.01	0.01	0.51
agree_diff	0.02	0.02	0.25
consc_diff	-0.00	0.01	0.76
neuro_diff	-0.01	0.01	0.43
<b>open_diff</b>	<b>0.03</b>	<b>0.02</b>	<b>0.04 *</b>

Table 4: Summary of topic entropy vs. personality differences linear model. Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '.' 1.

### 6.2 Model 2: Linear Change in Linguistic Alignment as a Function of Personality Differences

In Model 2 we examined if and how extent of linear change in linguistic alignment across the conversation was related to differences in interlocutor dyads' personality traits, controlling for dyad-level personality trait means. As conveyed in Table 5, larger between-interlocutor difference in extraversion was associated with steeper decrease in linguistic alignment across the conversation ( $\hat{\beta} = -0.02, p = 0.01$ ). We also examined the relationship between quadratic change in linguistic alignment and differences in interlocutor dyads' personality traits but found no significant associations.

Variable	Est.	SE	Pr(> t )
(Intercept)	-0.11	0.07	0.10
extra_mean	-0.01	0.01	0.36
agree_mean	0.01	0.01	0.61
consc_mean	0.00	0.01	0.73
neuro_mean	0.01	0.01	0.23
open_mean	0.01	0.01	0.53
<b>extra_diff</b>	<b>-0.02</b>	<b>0.01</b>	<b>0.01 **</b>
agree_diff	-0.00	0.01	0.92
consc_diff	-0.00	0.01	0.95
neuro_diff	-0.00	0.01	0.67
open_diff	-0.01	0.01	0.18

Table 5: Summary of linear change in linguistic alignment vs. personality differences linear model. Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '.' 1.

### 6.3 Model 3: Affect Change Difference as a Function of Personality Differences

In Model 3, we examined if and how extent of difference in affect change (pre-to-post-conversation) was related to differences in interlocutor dyads' personality traits, controlling for dyad-level affect change and personality trait means. As conveyed in Table 6, larger between-interlocutor difference in extraversion was associated with larger between-interlocutor difference in pre-to-post-conversation affect change ( $\hat{\beta} = 0.11, p = 0.01$ ).

Variable	Est.	SE	Pr(> t )
(Intercept)	0.32	0.46	0.49
aff_chg_mean	0.15	0.03	2.41e-07 ***
extra_mean	0.05	0.05	0.33
agree_mean	-0.03	0.07	0.62
consc_mean	0.08	0.06	0.17
neuro_mean	0.15	0.05	0.00 **
open_mean	-0.00	0.06	0.99
<b>extra_diff</b>	<b>0.11</b>	<b>0.04</b>	<b>0.01 **</b>
agree_diff	0.08	0.05	0.14
consc_diff	0.03	0.04	0.47
neuro_diff	0.07	0.04	0.06 .
open_diff	-0.02	0.05	0.72

Table 6: Summary of affect change difference vs. personality differences linear model. Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '.' 1.

### 6.4 Model 4: Mean Affect Change as a Function of Conversational Flow

Finally, in Model 4, we examined if and how extent of interlocutor dyads' mean pre-to-post-

conversation affect change was related to conversational flow, as quantified by topic entropy and the three summary linguistic alignment metrics, controlling for dyad-level personality trait means and differences. As conveyed in Table 7, greater topic entropy was associated with larger dyad-level mean pre-to-post-conversation affect change ( $\hat{\beta} = 0.43$ ,  $p < 0.001$ ).

Variable	Est.	SE	Pr(> t )
(Intercept)	-3.39	0.74	5.51e-06 ***
<b>topic_entropy</b>	<b>0.43</b>	<b>0.07</b>	<b>3.02e-09 ***</b>
intercept_term	-1.34	1.15	0.24
linear_term	0.22	0.14	0.13
quad_term	0.22	0.15	0.14
extra_mean	-0.11	0.04	0.02 *
agree_mean	0.14	0.06	0.02 *
consc_mean	-0.01	0.05	0.91
neuro_mean	0.29	0.04	4.42e-13 ***
open_mean	0.11	0.05	0.03 *
extra_diff	0.00	0.03	0.93
agree_diff	-0.03	0.04	0.53
consc_diff	0.03	0.04	0.42
neuro_diff	-0.05	0.03	0.12
open_diff	-0.01	0.04	0.85

Table 7: Summary of mean affect change vs. conversational flow linear model. Significance codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘.’ 1.

## 7 Discussion

The CANDOR corpus (Reece et al., 2023) opens unique opportunity to examine unscripted communicative behavior between strangers. The diversity in the participants’ personality traits and linguistic behavior allowed us to investigate some of the potential pathways through which people influence each other via dyadic conversations. The large sample size (over 1450 participants taking part in over 1650 conversations) makes the findings reasonably generalizable (e.g., to the U.S. population). Further, the relatively long conversation length (25+ minutes) supports a robust analysis of how discussion topics and linguistic behavior change over the course of a conversation.

Our analysis supports several key findings about how between-interlocutor differences in personality influence conversational dynamics, as well as how those conversational dynamics relate to interlocutors’ affect. Findings from Model 1 indicate that interlocutors who differ in their openness will spend more time talking about more topics. This result

may be a consequence of the more open interlocutor influencing the less open interlocutor to explore new topics. Findings from Model 2 indicate that interlocutors who differ in their extraversion will experience a more pronounced divergence in language use over the course of a conversation. This result may reflect a lack of effective social influence between such pairs of interlocutors. Findings from Model 3 indicate that interlocutors who differ in their extraversion also differ in how much their affect changes from before to after their conversation. This result supports the theory that such interlocutors do not subjectively experience the conversation in the same way. Finally, findings from Model 4 indicate that when interlocutors spend more time discussing more topics, they experience a greater boost in affect from the conversation. This result highlights the positive affective social influence of conversations with high topic entropy.

Taken together, the findings support the hypothesis that the dynamics of a conversation, measured by topic entropy and linguistic alignment, mediate the influence of interlocutors’ personality differences on their affective responses to the conversation. In other words, personality differences drive social influence, especially affective social influence, through the mechanism of conversation. We identify openness and extraversion as Big Five personality dimensions that are particularly consequential for social influence. LLMs often exhibit high openness and low extraversion, thus it will be important to consider personality’s role in LLM-based social influence as human-machine conversations become more commonplace.

## 8 Conclusion

In this paper, we examined if and how differences in interlocutors’ personality traits were related to differences in the dynamics of naturalistic conversations. Our findings suggest that personality differences are associated with a conversation’s topic entropy and how interlocutors’ linguistic alignment changes over time. We also illustrated a process by which personality differences influence conversational dynamics, which in turn influence interlocutors’ affective states. At a more general level, this work demonstrates new possibilities to engage in quantitative conversation analysis by leveraging text embeddings, projection, and clustering to track interlocutors’ movements throughout semantic space over the course of a conversation.



## Limitations

One limitation of this work is that topic entropy and linguistic alignment are theoretical constructs for which we do not have ground truth values. There are potentially many different ways to measure these constructs, and it is unclear how to determine which approach is more accurate. However, it is promising that we find several intuitive relationships between these two constructs and Big Five personality traits, which have demonstrated validity and reliability (Hahn et al., 2012).

In addition, specific design choices we made during the analysis may obscure the significance of some predictors in our models. In particular, the use of text embeddings to quantify conversational behavior appears to focus more on "big picture" aspects of a conversation, and may thus obscure some important aspects of conversational flow that manifest in specific linguistic features. Those features may also be related to personality differences and affect, but this is not reflected in our models.

## Ethics Statement

We believe that our analysis promotes social good by highlighting the novel (and often positive) outcomes of conversing with people different from ourselves. We also note that the data we used were collected from participants who shared highly personal information, such as their appearance, voice, traits, and emotions. The original data collection study was approved by Ethical & Independent Review Services, protocol #19160-01. All participants gave informed consent and were compensated appropriately. Nevertheless, we took additional caution to protect participants, in this case by opting to only analyze the raw text data (not the raw video and audio data) and by aggregating and de-identifying all data used in our analyses.

## Acknowledgements

Thank you to the CANDOR participants for providing data and to the CANDOR team for their work on preparing the dataset.

## References

Kenneth Benoit, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. [quanteda: An R package for the quantitative analysis of textual data](#). *Journal of Open Source Software*, 3(30):774.

Marina Bondi. 2010. *Keyness in texts*. John Benjamins Publishing Company.

Matteo Di Stasi, Templeton, and Jordi Emma Quoidbach. 2024. [Zooming out on bargaining tables: Exploring which conversation dynamics predict negotiation outcomes](#). *Journal of Applied Psychology*, 109(7):1077–1093.

David Edwards. 2024. [The time course of the rate of speaker transitions in conversation](#). *Proceedings of the Linguistic Society of America*, 9(1).

Ivar Frisch and Mario Giulianelli. 2024. [LLM agents in interaction: Measuring personality consistency and linguistic alignment in interacting populations of large language models](#). *arXiv:2402.02896 [cs.CL]*.

Phillip Glenn and Lawrence Susskind. 2010. [How talk works: Studying negotiation interaction](#). *Negotiation Journal*, 26(2):117–123.

Elisabeth Hahn, Juliana Gottschling, and Frank M. Spinath. 2012. [Short measurements of personality – Validity and reliability of the GSOEP Big Five Inventory \(BFI-S\)](#). *Journal of Research in Personality*, 46(3):355–359.

Airlie Hilliard, Cristian Munoz, Zekun Wu, and Adriano Soares Koshiyama. 2024. [Eliciting personality traits in large language models](#). *arXiv:2402.08341 [cs.CL]*.

Elliott M. Hoey and Kobin H. Kendrick. 2017. [Conversation analysis](#). In Annette M. B. de Groot and Peter Hagoort, editors, *Research methods in psycholinguistics and the neurobiology of language: A practical guide*, pages 151–173. Wiley-Blackwell.

Bogdana Humă, Elizabeth Stokoe, and Rein Ove Sikveland. 2010. [Putting persuasion \(back\) in its interactional context](#). *Qualitative Research in Psychology*, 17(3):357–371.

Alexei V. Ivanov, Giuseppe Riccardi, Adam Sporcka, and Jakub Franc. 2011. [Recognition of personality traits from human spoken conversations](#). In *Proceedings of the Annual Conference of the International Speech Communication Association (INTER-SPEECH)*, pages 1549–1552.

Guangyuan Jiang, Manjie Xu, Song-Chun Zhu, Wenjuan Han, Chi Zhang, and Yixin Zhu. 2023. [Evaluating and inducing personality in pre-trained language models](#). In *Advances in Neural Information Processing Systems (NeurIPS)*.

François Mairesse, Marilyn A. Walker, Matthias R. Mehl, and Roger K. Moore. 2007. [Using linguistic cues for the automatic recognition of personality in conversation and text](#). *Journal of Artificial Intelligence Research*, 30:457–500.

François Mairesse and Marilyn A. Walker. 2008. [Can conversational agents express Big Five personality traits through language?: Evaluating a psychologically-informed language generator](#).

- Leland McInnes, John Healy, and James Melville. 2020. [UMAP: Uniform manifold approximation and projection for dimension reduction](#). *arXiv:1802.03426 [stat.ML]*.
- Kathryn O’Neill and Emily S. Finn. 2024. [Pink noise in speakers’ semantic synchrony dynamics as a metric of conversation quality](#). In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 46, pages 1016–1022.
- James W. Pennebaker and Laura A. King. 1999. [Linguistic styles: Language use as an individual difference](#). *Journal of Personality and Social Psychology*, 77(6):1296–1312.
- R Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Andrew Reece, Gus Cooney, Peter Bull, Christine Chung, Bryn Dawson, Casey Fitzpatrick, Tamara Glazer, Dean Knox, Alex Liebscher, and Sebastian Marin. 2023. [The CANDOR corpus: Insights from a large multimodal dataset of naturalistic conversation](#). *Science Advances*, 9(13).
- Nils Reimers and Iryna Gurevych. 2019. [Sentence-BERT: Sentence embeddings using siamese BERT-networks](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Luca Scrucca, Chris Fraley, T. Brendan Murphy, and Adrian E. Raftery. 2023. *Model-Based Clustering, Classification, and Density Estimation Using mclust in R*. Chapman and Hall/CRC.
- Greg Serapio-García, Mustafa Safdari, Clément Crepy, Luning Sun, Stephen Fitz, Marwa Abdulhai Peter Romero, Aleksandra Faust, and Maja Matarić. 2023. [Personality traits in large language models](#). *arXiv:2307.00184 [cs.CL]*.
- David Silverman. 2006. *Interpreting qualitative data: Methods for analyzing talk, text and interaction*, 3rd edition. SAGE.
- Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. [MPNet: Masked and permuted pre-training for language understanding](#). *arXiv:2004.09297 [cs.CL]*.
- Emma Templeton, Luke Chang, and Thalia Wheatley. 2024. [Conversational launch pads: Strangers start their conversations with topics that lead to many other topics](#). In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 46, pages 2010–2016.
- Richard Watson Todd. 2011. [Analyzing discourse topics and topic keywords](#). *Semiotica*, 184(4):251–270.
- Svitlana Vakulenko, Maarten de Rijke, Michael Cochez, Vadim Savenkov, and Axel Polleres. 2018. [Measuring semantic coherence of a conversation](#). In *International Semantic Web Conference (ISWC)*.
- Hanna M. Wallach. 2006. [Topic modeling: Beyond bag-of-words](#). In *Proceedings of the 23rd International Conference on Machine Learning*, pages 977–984.
- Wendy Wood. 2000. [Attitude change: Persuasion and social influence](#). *Annual Review of Psychology*, 51:539–570.
- Yuening Yang. 2019. *The management of topics in ordinary conversation*. Ph.D. thesis, University of York.