# Dynamics of topic exploration in conversation

**Helen Schmidt**
Department of Psychology & Neuroscience
Temple University
helen_schmidt@temple.edu

**Claire Augusta Bergey**
Department of Linguistics
Stanford University
cbergey@stanford.edu

**Changyi Zhou**
Department of Psychology
UW-Madison
czhou249@wisc.edu

**Chelsea Helion**
Department of Psychology & Neuroscience
Temple University
chelsea.helion@temple.edu

**Robert D. Hawkins**
Department of Linguistics
Stanford University
rdhawkins@stanford.edu

## Abstract

Conversations are intricately structured forms of social interaction in which talkers move through interconnected topics with nested levels of semantic specificity. What principles govern how conversational partners jointly navigate an expansive topic space? To characterize these dynamics, we introduce a new dataset of annotated topic shifts from $N = 1,505$ annotators on 200 distinct video call conversations between strangers (Reece et al., 2023). Conversational dyads made stochastic but systematic transitions between topics, and within individual topics, we find that dyads begin concentrated in semantic space before dispersing to more idiosyncratic regions as topics progress. The same dispersion pattern also holds over entire conversations, providing quantitative evidence for nested levels of increasing specificity over conversations. Overall, our findings suggest that strangers get to know one another through systematic exploration of topic space, revealing hierarchical structure in idle talk.

**Keywords:** conversation; NLP; topic structure; dialogue; naturalistic conversation

## Introduction

When meeting someone for the first time, we engage in an intricate conversational dance: sharing information about ourselves, seeking out what we have in common, and building rapport – all while following implicit conversational norms about turn-taking, topic selection, and appropriate levels of self-disclosure (Goffman, 1959; Sacks, Schegloff, & Jefferson, 1974). This dance requires coordination across multiple timescales: moment-to-moment as speakers alternate turns (Stivers et al., 2009), minute-to-minute as they propose and exhaust topics (Maynard, 1980; Drew & Holt, 1998; Garcia & Joanette, 1997), and over the length of the entire conversation as they build up shared knowledge and establish common ground (Clark, 1991).

These processes intuitively lend conversation a rich hierarchical structure, with topics and subtopics nested within different phases of the conversation (Schegloff, 1968). This hierarchical structure is shared with other forms of joint action (Clark & Schaefer, 1987; Bratman, 1992). For example, consider building a chair with a partner (Deutsch, 1974; Clark, 2020; Zheng & Tversky, 2024). You first have to agree to build a chair and not some other piece of furniture. But then you can move to more and more specific commitments, such as building the back, affixing the rungs to the back, and finding a screw to affix those rungs. Once the back is built and these lower-level commitments are complete, you return to the higher-level commitment of building the chair and can take on another series of more specific commitments (e.g., building the seat, finding pieces for the seat, and so on). In this way, commitments are nested: more general commitments precede and encompass more specific ones, so when specific commitments are complete, partners can 'pop' back out to the higher-level plan.

The nested structure of commitments observed in joint action suggests a specific hypothesis about conversational dynamics: we should find that utterances not only become more specific over the course of a conversations, but also should do so in a nested manner (Clark, 1996; Brown, 1983; Abney, Paxton, Dale, & Kello, 2014). Just as building a chair involves cycles of general-to-specific planning, segments of conversation following the introduction of a new topic should *also* show a pattern of increasing specificity, which then roughly resets at topic boundaries. We might expect, for instance, that when discussing hobbies, people start with broader statements (e.g., "I like outdoor activities") before diving into content more specific to that conversation (e.g., "I hike in the Cascades", "I saw a plover on the shore"), and then may return to broader statements when transitioning to a new topic. That is, conversations should have a fractal-like structure with increasing specificity at multiple scales.

Though this kind of structure has been extensively theorized, it has proven challenging to test empirically. Shifts in content are idiosyncratic to each conversation, so granular data from many conversations is necessary to make generalizations about conversational structure. While computational approaches like topic modeling have characterized topical content across texts (Yeh, Tan, & Lee, 2016; Jelodar et al., 2019; Ashby et al., 2024), they typically provide static topic mixtures instead of capturing temporal dynamics. And while some models have attempted to segment texts temporally based on topic context, such as Hidden Topic Markov Models (Gruber, Weiss, & Rosen-Zvi, 2007) or TextTiling (Hearst, 1997), they have not been thoroughly validated on natural conversations. Human annotations of topic shifting in everyday conversations are therefore necessary to provide a ground truth standard for modeling topic shifts and to answer empirical questions about the topic structure of natural conversation (Decker, Amblard, & Breitholtz, 2023).



In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* 

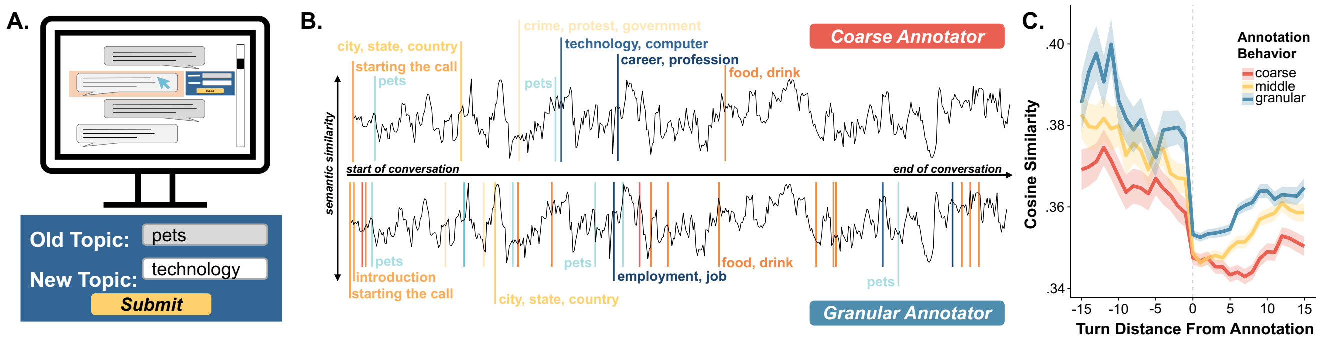


Figure 1: *Overview of Procedure and Analytic Approach.* **(A.)** Annotation task interface. **(B.)** Example annotated conversation by two annotators (coarse, top; granular, bottom). Black line reflects within-transcript semantic similarity across adjacent sliding windows of 10 utterances; colored lines indicate annotated topic shifts. Text reflects a selection of cluster labels and colors indicate clusters of topics identified across transcripts (see Topic Clustering). **(C.)** Semantic similarity of adjacent sliding windows across all transcripts around annotated topic shifts: semantic similarity sharply drops across topic boundaries. Line color reflects annotation behavior (coarse, middle, granular). Shading reflects $\pm 1$ standard error.

In this paper, we introduce an annotation dataset of topic shifts from 1,505 human annotators on 200 transcripts of casual video call conversations from the CANDOR Corpus (Reece et al., 2023) with the intention of releasing this data publicly as a complement to CANDOR. First, we validate our annotation method, examining how human annotators label topics in naturalistic conversation and characterizing different levels of annotation granularity. We further examine how topic shifts coincide with shifts in semantic content, modeled through sentence embeddings. Then, we characterize how conversational partners explore the conceptual space of topics, both on a macro scale – how they navigate between-topic transitions – and a micro scale – how their utterances become more specific within topics. Finally, we examine the hypothesis that interlocutors explore more specific content in a nested way, both within topics and across the course of entire conversations.

## Methods

### CANDOR Corpus

CANDOR is a large, publicly available dataset of video call conversations between pairs of English-speaking strangers located in the United States (Reece et al., 2023). In CANDOR data collection, participants were instructed to have a conversation with each other for at least 25 minutes and were not given further instructions for their conversation; rather, they were simply told that they could talk about any topic and imagine that they were getting to know the other person as if meeting at a social event. The conversations were transcribed using automated transcription models, and segmented using several different turn-taking schemas. We selected the ‘backbiter’ version to use in the experiment, which removes short words and utterances listeners might use to acknowledge what another person is saying without taking the floor (*backchannels*, e.g., “yeah”, “mhm”). We randomly selected 201 conversation transcripts for the topic annotation task.

### Procedure

**Annotation Task** Participants were directed to complete an annotation task developed on Empirica, an open-source virtual platform for hosting participant-facing experiments (Almaatouq et al., 2021). Our custom-built web interface mimicked the design of popular messaging apps to present conversation transcripts as if they were mobile text exchanges (Figure 1A). Utterances appeared sequentially in alternating chat bubbles in a scrollable window. Speakers from the original CANDOR transcript were randomly assigned to a chat bubble color and alignment. Each participant annotated one entire transcript, and multiple participants annotated each transcript [1].

Participants were instructed to read the chat conversation and mark topic shifts by clicking on chat bubbles to indicate when a new conversation topic began (Figure 1B). All conversations included a researcher-determined first topic (“starting the call”) and participants were responsible for locating and labeling all subsequent topics. Each time they selected a chat bubble to annotate, participants were prompted to type a few words describing the new topic and were shown the previous topic label (either “starting the call” for the first topic shift, or the participant-provided label from the previous topic shift). After a chat bubble was annotated, a box appeared around it to visually mark it as an annotated turn. Participants also had navigation buttons at the bottom of the chat window to quickly review the locations of prior annotation labels.

**Participants** We recruited 1,850 participants across three rounds of data collection to complete an online annotation

[1] Data & Code: https://github.com/SocialInteractionLab/Candor.

task. Of these, 1,757 participants provided demographic information ($mean_{age}$ = 35.16, $SD_{age}$ = 12.33, 861 female, 881 male, 12 preferred not to disclose, 795 non-white, 946 white). Participants were recruited from Prolific (www.prolific.com) and were required to be located in the United States, the United Kingdom, or Canada at the time of participation. Participants also had to pass Prolific's built-in English fluency screening. Each participant was allowed to complete the study only once to maintain data integrity. All procedures were approved by UW-Madison's Institutional Review Board and participants received $5 for completing the task (based on an estimated completion time of 20 minutes).

Of the 1,850 participants recruited, 345 participants provided too few (fewer than 7) or too many (40 or more) annotations for their assigned transcript, according to the bottom and top 10th percentile of all participant annotations, respectively. These participants were discarded from our final analytic sample, and one transcript was dropped due to all annotators being excluded. In the final analytic sample, we had 1,505 annotators across 200 transcripts, with between 3 and 11 annotators assigned to each transcript.

**Transcript Assignment** We selected a sequential subsample of 200 transcripts from the CANDOR corpus, based on the random alphanumeric transcript ID assigned by developers of the CANDOR corpus. To ensure multiple participants annotated the same dense subsample of transcripts, we randomly assigned 1,646 participants to these transcripts across 3 rounds of data collection. In the third round of data collection, we manually allocated 204 participants to transcripts in the subsample that had been randomly assigned few or no annotators in the first two rounds to ensure all transcripts had multiple annotators.

## Pre-Processing Pipeline

**Extracting Embeddings** Using a pre-trained sentence transformer model (`all-MiniLM-L6-v2`; Reimers & Gurevych, 2019), we embedded all transcript utterances and participant-provided topic labels into a 384-dimensional dense vector space. Each embedding is a point in this multidimensional space, and points that are closer together reflect utterances and labels that are more semantically similar to each other. In analyses, we quantitatively represent this semantic similarity by calculating the cosine of the angle between points (i.e., cosine similarity). In visualizations, we represent these embeddings in a two-dimensional space reduced using t-SNE (Van der Maaten & Hinton, 2008).

**Standardizing Conversation Trajectories** CANDOR participants were instructed to converse for at least 25 minutes (Reece et al., 2023), and our 200 conversations varied in length ($m$ = 32.54 minutes, $SD$ = 8.93 minutes) and in number of turns ($m$ = 357 turns, $SD$ = 192 turns) between speakers. Turns were numbered sequentially as they occurred in the conversation, and to account for variation in the number of turns, we rescaled each conversation to reflect how far along each utterance fell relative to the total number of utterances. We rounded percent completion values to the nearest integer percentile (e.g., 26% complete) and grouped utterances into concatenated bins. This standardization allows us to compare across transcripts and account for conversation length variation. In subsequent analyses, *turn* reflects the sequentially numbered utterances extracted from the conversation dataset while *conversation completion* reflects the standardized bins tracking progress through a conversation.

**Topic Clustering** Annotators provided a wide variety of topic labels, and to compare topic shifts at the aggregate level across transcripts, we used k-means clustering on topic label embeddings to group topics into higher-level clusters. To determine the number of clusters to use, we calculated inertia (a measure of clustering quality, also known as within-cluster sum of squares) across all clusters integers between 0 and 200. The bend in inertia (i.e., "elbow") fell between 50 and 75 clusters; our goal was to select a cluster number that prioritized interpretability and minimized topic overlap, so we selected 50 clusters. Following clustering, we crafted researcher-generated cluster labels to describe the participant-generated topic labels included in each cluster. We also qualitatively grouped cluster labels into 10 themes (topic collections) to use for illustrative purposes when discussing broad topic shifts within and across conversations.

# Validating Annotation Data

**Annotator Agreement** To evaluate agreement in annotation placement between participants who annotated the same transcript, we calculated Jaccard similarity coefficients between all pairwise sets of participants for each transcript. Specifically, we took the intersection of annotated turns divided by the union of annotated turns to calculate participant agreement, and calculated an average for each transcript and an overall average for all transcripts. Given the fuzzy nature of topic transitions, we counted annotations as agreeing if they fell within a tolerance window of 5 turns of each other. The mean Jaccard similarity across transcripts was 0.293 ($SD = 0.06$), indicating that annotators showed meaningful consensus about major topic transitions, while also indicating some subjective variation in how they perceive topics to shift in conversation.

**Semantic Similarity at Topic Shifts** To examine whether local shifts in semantic similarity are related to topic shifts, we concatenated utterances into windows of 10 utterances, extracted sentence embeddings for each window, and calculated cosine similarity between adjacent windows sliding through each conversation. That is, at each point in the conversation, we extracted a measure for how semantically similar the 10 prior utterances were to the 10 subsequent utterances (Figure 1C). We next examined whether there were systematic shifts in semantic content at topic boundaries. Centering the sliding windows on topic boundaries (i.e., the final utterance of the *prior* window was labeled as a topic shift),

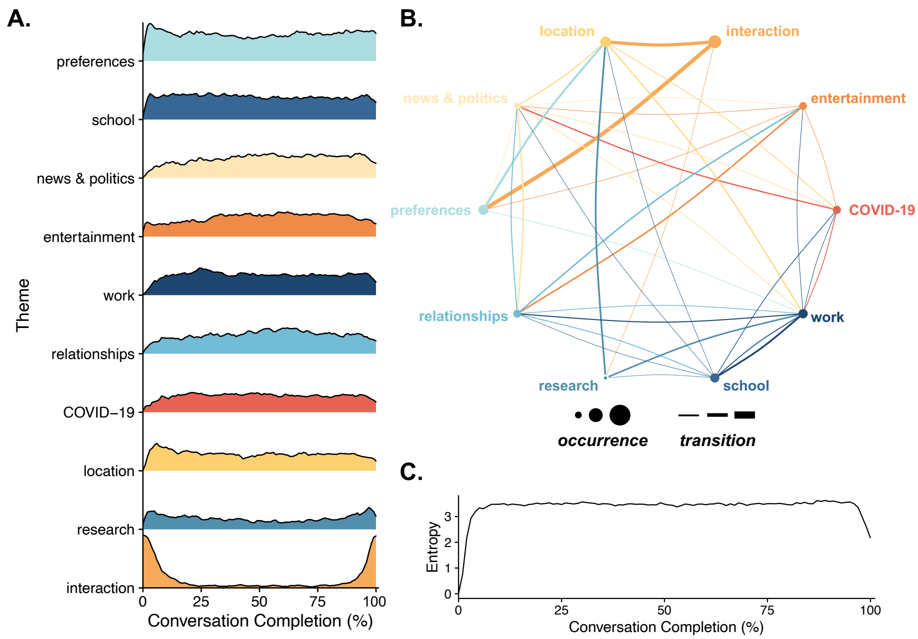


Figure 2: *Topic Shifts*. **(A.)** Thematic occurrence over the course of transcripts. Topic cluster labels are grouped into 10 thematic groups, and curve height represents the density of clusters within each theme across 200 transcripts. **(B.)** Thematic cluster transitions. Cluster labels grouped into thematic groups and probability of transitioning from each theme to another reflected in line weight. Probabilities less than 0.125 not shown. Line color indicates which theme was transitioned from. **(C.)** Entropy. Average topic entropy over the course of a conversation over all 200 transcripts.

we found that there is a sharp fall in semantic similarity at topic shifts. That is, stretches of conversation that cross a topic boundary are much less similar than equally adjacent stretches that are within a topic.

To assess whether these sharp drops in semantic similarity were associated with annotated topic shifts, we compared cosine similarities of 10-utterance spans that crossed topic boundaries compared to the average of fifteen 10-utterance spans that preceded topic boundaries (and were within-topic). We found that spans that crossed topic boundaries had significantly lower semantic similarities than those that did not cross a boundary ($t(302.86) = -3.81$, $p < .001$). For robustness, we repeated these analyses with window sizes of 15- and 20-utterance spans (increasing preceding average spans by 5 for each); results were numerically similar but not significant, which may be because as window size increases, it is more likely a topic shift occurs within a given window ($t_{15}(267.2)$ = -0.70, $p_{15}$ = 0.482, $t_{20}(276.2)$ = -1.94, $p_{20}$ = 0.054).

We also explored variation in this analysis based on the total number of annotations each participant made, grouping annotators into "coarse", "middle", and "granular" annotation behaviors. Coarse annotators made fewer than 17 (33rd percentile) topic shift annotations, granular annotators made more than 26 (66th percentile), and middle annotators made between 18 and 25 annotations. The sharp fall in semantic similarity was similar for each annotator group (Figure 1C), suggesting this analysis is robust to individual differences in annotation style.

## How Interlocutors Explore the Topic Space

To quantitatively evaluate how interlocutors explore topic space across conversations, we examined three measures: *cluster occurrence*, or the density of thematic content across transcripts; *transition probability*, or the conditional likelihood of transitioning from a given topic to another; and *entropy*, or how much variability in topic space exists across transcripts as conversations progress.

**Cluster Occurrence** To examine how often topic clusters appeared across transcripts and where clusters occurred over the course of an individual conversation, we tagged participant-provided topic labels with corresponding researcher-generated cluster labels (see *Topic Clustering* section above). Because multiple participants annotated the same 200 transcripts, cluster length (number of turns) and order (sequence of labels) varied at the participant level. We therefore calculated the conditional probability of transitioning between topics (e.g., given a prior topic, what is the probability of transitioning to the current topic) for each participant ($N$ = 1,505) and each thematic group of cluster labels ($N$ = 10), before calculating an overarching average across participants. Finally, we calculated the proportion of transcripts (out of 200) that contained cluster labels within each theme at each conversation completion bin (0-100%) and plotted density curves for each theme (Figure 2A).

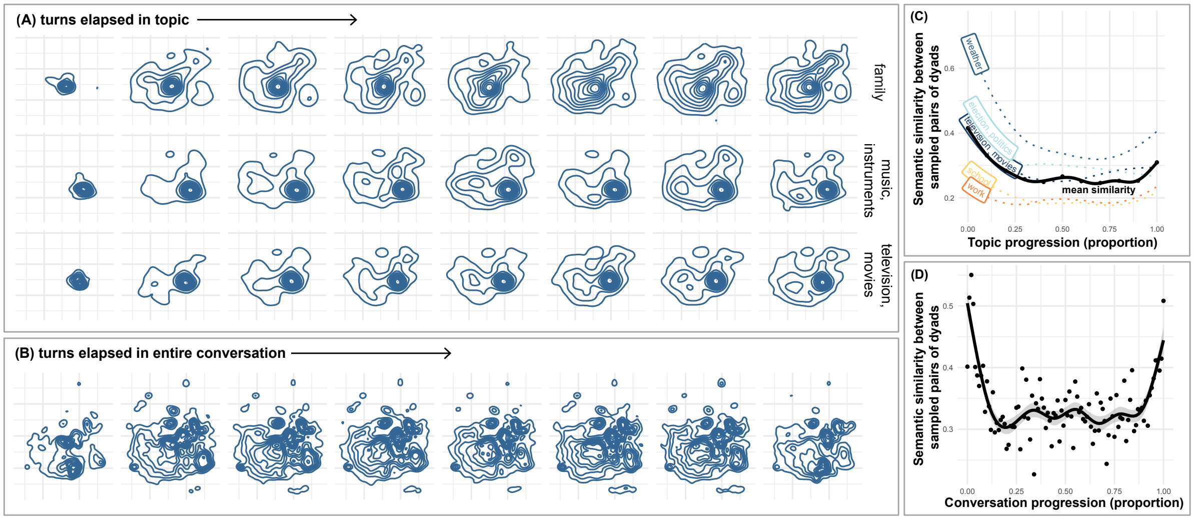


Figure 3: *Topic and Conversation Specificity*. How conversational partners explore the semantic space over the course of a given topic (**A.** three example topics shown) and entire conversations (**B.**). Plots show density of utterance embeddings in t-SNE reduced semantic space; facets show progression through topic or conversation length. Semantic similarity between sampled pairs of dyads over the course of topics (**C.**) and entire conversations (**D.**). Across dyads, people tend to start a given topic densely concentrated in one part of semantic space (high similarity), and disperse to more dissimilar parts of the semantic space over the course of the topic (with a return to similarity at the end). The same pattern holds across entire conversations.

Density curves highlight varying occurrences of conversation topics across the course of a conversation. Interaction topics, including introductions and goodbyes, tend to occur at the beginning and end of the conversation, while topics related to school, entertainment, and preferences occur frequently throughout. Location and work topics fall earlier in the conversations, but conversely, discussions of news, politics, and relationships tend to occur later. These results suggest a loose temporal structure in conversations, by which certain topics tend to be explored earlier than others.

**Transition Probability** We were also interested in thematic transitions between cluster labels aggregated across entire transcripts. Using the conditional probabilities of topic transitions, and the 10 thematic groups describing the 50 cluster labels, we used `ggraph` (Pedersen, 2024) to illustrate base rate and transition probability. Node sizes reflect the base rate occurrence (e.g., commonality) across transcripts and edge weight reflects the transition probability between cluster themes. Edge color reflects the prior topic and the connection node reflects the current topic; for example, a red line from the "COVID-19" node to the yellow "news & politics" node would reflect a high transition probability of talking about the pandemic and then talking about the news. We dropped lines with weights below 0.125 to emphasize more common topic transitions and increase readability (Figure 2B).

To compare our observed transition matrices to a null model in which speakers wander randomly between topics, we calculated Jenson-Shannon divergences (JSD) between empirical transition probabilities and the base rates estimated across all conversations (excluding researcher-provided "starting the call" topic). We compared observed values to a distribution of 1000 permuted values by randomly shuffling the sequence of topics within each conversation and calculating a null distribution of JSDs expected using transition probabilities estimated under each permutation. Empirical transition probabilities for both themes ($JSD = 0.15$, $p < .001$) and clusters ($JSD = 4.89$, $p_{cluster} < .001$) differed significantly from base rates, demonstrating systematic structure in the sequence of topics as interlocuters explore topic space.

**Entropy** Do conversations spread to more varied topics as they progress? One coarse-grained way to test whether conversations become more idiosyncratic as they progress is to measure the entropy (i.e., diversity or unpredictability) of topics across conversations. To evaluate whether entropy increases as conversations progress, we extracted the most common cluster label at every time bin (0-100) per transcript across all participants who annotated the transcript, establishing a consensus label for each time bin in each conversation. We then computed the relative appearance of each consensus label across all 200 transcripts at each time bin and used this proportion to calculate entropy (Figure 2C). We find that entropy increases at the beginning and decreases at the end of conversations (Bayesian quadratic model: $\beta = 0.03$, 95%*CrI* [0.02,0.04], $\beta^2 = -0.0003$, 95%*CrI* [-0.0004, -0.0002].

Qualitatively, at the start of conversations, entropy is low (less than 1), but before approximately 5% of the conversation has elapsed, we observed a dramatic increase in entropy (greater than 3). Entropy remains high over the course of the conversation and does not further increase as conversations progress, with a similar decrease toward the final 5% of the conversation. Low entropy at the beginnings and ends of conversations is consistent with stereotyped opening and closing interactions, including introductions and goodbyes. These results suggest that openings and closings tend to be similar across conversations, but talk in between varies much more.

**Increasing Specificity Within and Across Topics** If people talk about more specific content over the course of a conversation and over the course of topics, we should see that their utterances become more idiosyncratic over these nested time periods. If they also return to more general content before transitioning to new topics (or ending conversations), we should see that they talk about slightly more similar things at the ends of these time periods. To examine this, we calculated how concentrated or dispersed people's utterances were in semantic space over the course of stretches talking about a given topic and over entire conversations. If conversations follow this repeating pattern of similarity and idiosyncrasy, we should see their distribution in semantic space become more dispersed over both of these timescales.

This is just what we see: across dyads, utterance embeddings are more densely concentrated at the beginnings of topics and conversations and become more dispersed as topics (Figure 3A) and conversations (Figure 3B) progress. To statistically test this, we measured the average cosine similarity in embedding space between pairwise samples of utterances across conversations, within conversation and topic segments. If conversations become more dispersed, randomly sampled pairs of utterances across conversations should be more similar at the beginnings of topics and of conversations, become less similar, and may have an uptick at the end as they transition to new topics (or end conversations).

Modeling the cosine similarity of pairwise sampled utterances within topic cluster and within topic time increment using a quadratic regression, we find a convex relationship ($\beta_t$ = -0.40, $p<.001$; $\beta_{t^2}$ = 0.33, $p<.001$), indicating that similarity starts high, drops, and increases again (Figure 3C). Modeling the same relationship over the course of entire conversations (Figure 3D), we also find a convex relationship ($\beta_t$ = -0.45, $p<.001$; $\beta_{t^2}$ = 0.43, $p<.001$). This pattern is consistent with the hypothesis that conversations have a nested structure of increasingly specific content. That is, conversations and topics often look similar at the start, but they disperse to distinctive places as they progress.

## Discussion

In this paper, we introduce a dataset of human topic annotations of unstructured naturalistic conversations between strangers, and investigate how people explore the space of topics as they talk. Our analyses reveal several key patterns in conversational dynamics. First, we were able to validate that annotators are picking up on genuine structure in the conversations, as annotators showed substantial agreement and their annotations coincided with significant shifts in embeddings of semantic content. Second, we demonstrated that conversations exhibit nested levels of increasing specificity: both within individual topics and across entire conversations, speakers begin with more general, semantically similar content before diverging into more idiosyncratic territory. This pattern is evidenced by the initial clustering of utterances in semantic space followed by greater dispersion of utterances over time. Our findings suggest that while conversations can appear unpredictable, they follow systematic patterns of joint exploration and navigation of conceptual space.

Our use of the CANDOR corpus presents a few limitations. First, conversations were limited to English-speaking, mostly white participants located in the United States (Reece et al., 2023), and there may be cultural and linguistic variability in how people navigate conversational topics. Second, the nature of the relationship between speakers may influence conversational dynamics, such that friends and family members navigate topic space differently than strangers. Finally, annotators only saw the written transcripts of the CANDOR video calls, and did not have access to non-verbal cues that may affect annotation decisions.

A possible concern is that our findings could be explained by a simpler null model: a dyad initializes somewhere in semantic space at the beginning of a conversation and drifts randomly from there. On this account, the apparent nested structure we observed arises as an artifact of the annotation process: annotators are simply 're-centering' topic labels around continuously drifting content once it wanders too far from the previous label. However, our analyses contradict this null 'drift' alternative: there are in fact sharp change points in underlying semantic content at topic boundaries (Figure 1C). Further, topic transition probabilities significantly diverge from base rates: navigation between topics is non-random.

An alternative account that makes more sense of these results draws from information foraging theory (Pirolli, 2007): dyads in conversation may explore the topic space by searching for regions of fruitful connection or commonality, analogous to how animals forage for food (Charnov, 1976; O'Nell & Finn, 2024). Early topics may offer indicators of potential common ground, similar to how foragers learn to recognize reliable cues for resource-rich patches (Cohen & Todd, 2018). And the hierarchical distribution of topic transitions, with many small local shifts and occasional large jumps, may mirror the area-restricted search found in both animal foraging and human search behaviors (Todd & Hills, 2020; Dorfman, Hills, & Scharf, 2022; Pacheco-Cobos et al., 2019; Garg & Kello, 2021). This search pattern is optimal when searching a space in which resources (in this case, opportunities for connection) are sparsely and randomly distributed. More broadly, these results suggest that the apparent messiness of casual conversation masks a deeper underlying structure – one by which goal-directed speakers are jointly traversing semantic space in search of common ground, starting with reliable indicators of similarity before venturing into more specific, idiosyncratic territory.

## Acknowledgments

We thank Pranav Dronavalli for initial development of the messaging app web interface. We also thank members of the Stanford Social Interaction Lab for helpful feedback.

## References

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General*, *143*(6), 2304.

Almaatouq, A., Becker, J., Houghton, J. P., Paton, N., Watts, D. J., & Whiting, M. E. (2021). Empirica: a virtual lab for high-throughput macro-level experiments. *Behavior Research Methods*, *53*, 2158–2171.

Ashby, T., Kulkarni, A., Qi, J., Liu, M., Cho, E., Kumar, V., & Huang, L. (2024). Towards effective long conversation generation with dynamic topic tracking and recommendation. In *Proceedings of the 17th International Natural Language Generation Conference* (pp. 540–556).

Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review*, *101*(2), 327–341.

Brown, G. (1983). *Discourse analysis*. Cambridge University Press.

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, *9*(2), 129–136.

Clark, H. H. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (p. 127 - 149).

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

Clark, H. H. (2020). Social actions, social commitments. In N. J. E. Stephen C. Levinson (Ed.), *Roots of human sociality* (pp. 126–150).

Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, *2*(1), 19–41.

Cohen, S. E., & Todd, P. M. (2018). Relationship foraging: Does time spent searching predict relationship length? *Evolutionary Behavioral Sciences*, *12*(3), 139.

Decker, A., Amblard, M., & Breitholtz, E. (2023). Analysing topic shifts in task-oriented dialogues. In *Journées scientifiques du GDR Lift - LIFT 2023* (p. 48–54).

Deutsch, B. G. (1974). *The structure of task oriented dialogs*. Artificial Intelligence Center, SRI International.

Dorfman, A., Hills, T. T., & Scharf, I. (2022). A guide to area-restricted search: a foundational foraging behaviour. *Biological Reviews*, *97*(6), 2076–2089.

Drew, P., & Holt, E. (1998). Figures of speech: Figurative expressions and the management of topic transition in conversation. *Language in Society*, *27*(4), 495–522.

Garcia, L. J., & Joanette, Y. (1997). Analysis of conversational topic shifts: A multiple case study. *Brain and Language*, *58*(1), 92–114.

Garg, K., & Kello, C. T. (2021). Efficient lévy walks in virtual human foraging. *Scientific Reports*, *11*(1), 5242.

Goffman, E. (1959). *The presentation of self in everyday life.* Doubleday.

Gruber, A., Weiss, Y., & Rosen-Zvi, M. (2007). Hidden topic markov models. In *AISTATS* (pp. 163–170).

Hearst, M. A. (1997). Text tiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, *23*(1), 33–64.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia tools and applications*, *78*, 15169–15211.

Maynard, D. W. (1980). Placement of topic changes in conversation. *Semiotica*, *30*(3-4).

O'Nell, K., & Finn, E. S. (2024). Pink noise in speakers' semantic synchrony dynamics as a metric of conversation quality. In *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

Pacheco-Cobos, L., Winterhalder, B., Cuatianquiz-Lima, C., Rosetti, M. F., Hudson, R., & Ross, C. T. (2019). Nahua mushroom gatherers use area-restricted search strategies that conform to marginal value theorem predictions. *Proceedings of the National Academy of Sciences*, *116*(21), 10339–10347.

Pedersen, T. L. (2024). ggraph: An implementation of grammar of graphics for graphs and networks [Computer software manual].

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*. Oxford University Press.

Reece, A., Cooney, G., Bull, P., Chung, C., Dawson, B., Fitzpatrick, C., ... Marin, S. (2023). The candor corpus: Insights from a large multimodal dataset of naturalistic conversation. *Science Advances*, *9*(13), eadf3197.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 3982–3992).

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist*, *70*(6), 1075–1095.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... others (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592.

Todd, P. M., & Hills, T. T. (2020). Foraging in mind. *Current Directions in Psychological Science*, *29*(3), 309–315.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, *9*(11).

Yeh, J.-F., Tan, Y.-S., & Lee, C.-H. (2016). Topic detection and tracking for conversational content by using conceptual dynamic latent dirichlet allocation. *Neurocomputing*, *216*, 310–318.

Zheng, C., & Tversky, B. (2024). Putting it together, together. *Cognitive Science*, *48*(2), e13405.