Miscommunication:

How prediction and egocentricity increase progressivity but decrease intersubjectivity

A dissertation submitted by

Julia Mertens

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Psychology

Tufts University

August 2022

Advisor: Dr. Jan P. de Ruiter

Abstract

Miscommunication contributes to adverse medical outcomes, transportation disasters, inequality, and other problems. While researchers study the types of problems in conversation, we still know little about why people do not resolve those problems more efficiently. Therefore, I propose and evaluate one possibility: interlocutors prioritize progressivity over intersubjectivity, causing miscommunication.

Chapter I introduces progressivity (the social goal of moving forward) and intersubjectivity (the social goal of achieving mutual understanding), and outlines how interlocutors balance the two. Chapter II describes the data, defines the key terms, and provides examples of the phenomena relevant to both studies.

Chapter III shows that listener predictions – thought to increase progressivity – interfere with intersubjectivity. First, I hypothesized that listeners would struggle to comprehend surprising turns. I found that more surprising turns were more likely to be targeted by other-initiations of repair (OIRs, e.g., “What did you just say?”). Second, I hypothesized that listeners would “give up” on processing surprising trouble sources, so surprising trouble sources would be followed by earlier OIRs. However, I found no effect of surprisal on the turn transition times after trouble sources. Finally, an exploratory analysis found a complex, unexpected relationship between prediction and miscommunication.

In Chapter IV, I investigated whether egocentricity, the tendency for interlocutors to produce and interpret language from their own perspective, increases progressivity but decreases intersubjectivity. People are more egocentric when talking with friends than with strangers in experimental settings, but this finding has never been extended to natural conversation. Therefore, I investigated whether friends experience more frequent lapses in intersubjectivity and are more progressive than strangers. First, I found that friends experienced more than two times as many communication problems as strangers. Second, I found that friends had higher values for some, but not all, quantifications of progressivity. Third, I found that some progressivity metrics were correlated with the rate of communication problems, but more granular metrics and analyses are needed. Finally, an exploratory analysis found that egocentricity and social closeness contribute to miscommunication in myriad ways. Finally, Chapter V explores these findings and proposes directions for future research.

Acknowledgements

First and foremost, I'd like to thank my chair, Dr. JP de Ruiter, for guiding me through the past five years. I would also like to thank the other members of my committee, Dr. Ariel Goldberg, Dr. Gina Kuperberg, and Dr. Kobin Kendrick. I have been privileged to received feedback from four experts in four areas of language research.

I would be negligent if I didn't mention the many social networks that supported me especially through a COVID-era dissertation. My family kept me grounded, reminding me that no matter the degree, I'm still a Mertens. It's been especially exciting to finish my PhD along with my mother, Beret Ravenscroft, who defended her PhD in Nursing one week after I defended this dissertation. My cohort – Alia Wulff, Paige Hickey, Tom McWiliams, and Tori Sharpe – provided great group therapy and support. My fellow HI Lab graduate students, Muhammad Umair, Lena Warnke, and Chas Threlkeld, sat through clumsy practice talks, provided feedback on disoriented chapters, and provided structure during an incredibly unstructured time. I've also received some non-human social support: about the same time that I started writing this dissertation, I adopted two kittens: Mojito and Mimosa Mertens. They grew even faster than my page count.

And finally, I would like to thank Nicholas Cage, Aubrey Gordon, Michael Hobbes, Marc Rebillet, Robert Evans, and J.B. Smoove for creating the media content I've devoured over the course of the COVID pandemic.

Table of Contents

Table of Equations

Table of Excerpts

Table of Figures

Table of Tables

Chapter 1: Introduction

On December 1998, NASA launched the Mars Climate Orbiter. Nine months later, the orbiter suddenly disappeared. Tragically, the engineers designed the orbiter to use metric standard units (Newton-seconds), but the navigation team sent commands in English units (pound-seconds; Siddiqi, 2018). Instead of entering the intended orbit, the orbiter entered Mars' atmosphere and disintegrated. This miscommunication cost $327.6 million in materials and labor.

Less well known are the ways miscommunication costs money, lives, and energy every day. Miscommunication contributes to airplane (Hamzah & Fei, 2018; Mcmillan, 1998) and ship (Boström, 2021) disasters. It increases employee stress (McKenzie & Qazi, 1983). Doctor-patient (Britten et al., 2000) and clinician-clinician (Chang et al., 2010) miscommunication is frequent and leads to adverse outcomes. In one survey, approximately 38% of medical dispute cases in Japan involved miscommunication (Aoki et al., 2008). Anywhere communication takes place, so does miscommunication – sometimes with life-altering consequences.

Researchers have outlined many sources of communication problems, including differences in background knowledge, noisy environments, and strong emotional states (e.g., Paxton et al., 2021; Sayer, 2013). However, this knowledge is not enough to prevent miscommunication. Some organizations attempt to avoid sources of communication problems by implementing strict communication protocols. These protocols eliminate some sources of miscommunication, but they are also notoriously difficult to follow. For example, pilots are trained to repeat their interlocutors' turns (e.g., Hamzah & Fei, 2018). This strategy allows the original speaker to correct any mishearings. However, pilots frequently break this protocol.

Further, some sources of communication problems are unavoidable. For example, linguistic ambiguity may be a feature, not a flaw, of communication (Piantadosi et al., 2012), and therefore is impossible to avoid entirely. I hope to move past categorizing the types of problems in conversation and move towards explaining why interlocutors do not cope with problems more effectively. Specifically, I will explore whether the way interlocutors prioritize *progressivity* over *intersubjectivity* (first described in Heritage, 2007) creates the opportunities for miscommunication.

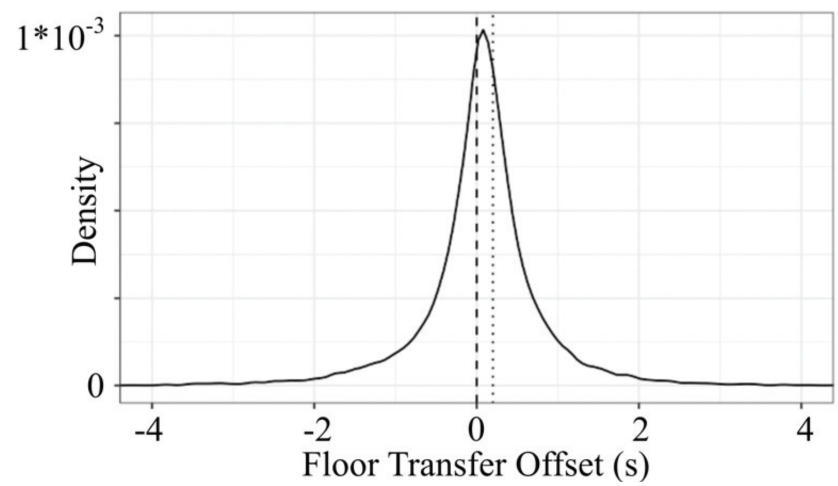
Schegloff (2007) defined progressivity as the movement “from one element to a hearably-next-one with nothing intervening.” Sequences, turns, words, and gaps can all intervene between one element and the expected next element, therefore decreasing progressivity. For example, the expected turn after an information request (e.g., “When do you defend again?”) might be an answer (e.g., “On June 7th”). An intervening turn (e.g., “Defend what?”) would decrease progressivity. Further, progressive turns are *minimized*, meaning the speaker says as few words as necessary to communicate their message (Levinson, 1987). For example, speakers minimize their references (e.g., “Julia” instead of “the fifth year student defending her PhD”) whenever they can (Sacks & Schegloff, 1979).

Interlocutors also minimize the gaps between turns in conversation. For example, if the selected speaker (e.g., Julia in “How’s your dissertation going Julia?”) does not respond quickly, someone else may jump in to re-establish progressivity (Stivers & Robinson, 2006). Across languages, turn transition times are frequently less than 200ms long (Jefferson, 1989; Sacks et al., 1974; Stivers et al., 2009). This is especially impressive given that it takes participants at approximately 500ms to initiate the name of a single object in a picture (Indefrey & Levelt, 2004). Turn transitions in the In Conversation Corpus (ICC, see Chapter II), were similarly short

(Figure 1). In the ICC, interlocutors exchanged the floor approximately 21.4 times per minute (*95%CI*: 19.9 – 22.8); if every speaker transition were just 100ms longer, every minute of conversation would be approximately 2.14 seconds (3.5%) slower.

Figure 1

The distribution of turn transition times in the In Conversation Corpus



Note: A turn transition time of 0ms represents a perfectly timed turn transition. The highest mode turn transition time across languages was 200ms in Stivers et al. (2009), represented here as the vertical dotted line.

Interlocutors also attempt to achieve intersubjectivity, a state of apparent mutual knowledge: when each interlocutor believes they understand the others and believes the others understand them (Clark & Marshall, 1981; Raymond, 2019). Interlocutors achieve intersubjectivity by *grounding* (Clark & Schaefer, 1987). During grounding, interlocutors can provide positive evidence of (non)understanding that varies in strength (Clark & Brennan, 1991). The recipient of a turn provides weak positive evidence when they continue to attend to their social partner, stronger positive evidence when they produce a minimal positive response (e.g., “mhm”), and the strongest positive evidence when they respond appropriately. Similarly, interlocutors can provide negative evidence of understanding (or evidence of non-understanding) by producing a mismatched turn or an other-initiation of repair (OIR; Dingemanse et al., 2015; Kendrick, 2015a; Schegloff et al., 1977). OIRs signal trouble in the conversation and make a solution relevant in an upcoming turn. In Excerpt 1, line 1 contains the trouble source, line 3 is the OIR, and line 4 is the solution.

Excerpt 1

Candidate OIR sequence[1]

```
1   SP1: where do you l:ive,
2         (0.4)
3   SP2: um you mean on camp⌈us:? ⌉
4   SP1:                    ⌊yeah.⌋
5   SP2: uh I live in west hall.
```

---

[1] Unless otherwise mentioned, excerpts are from the ICC (see Chapter II). The symbols used in transcriptions are explained in

Appendix A. Grey boxes mark turns that contain a trouble source, orange boxes represent OIRs, and blue boxes represent solutions.

As first described in Heritage (2007), it is impossible to maximize both intersubjectivity and progressivity because they require opposing strategies. Providing stronger positive evidence of understanding decreases the likelihood intersubjectivity will lapse, because it provides more information that another interlocutor may correct. However, stronger positive evidence of understanding takes more time and decreases progressivity. Providing evidence of non-understanding can pause progressivity for even longer: OIRs halt progressivity for at least two turns. Excerpt 1 without lines 3 and 4 would have been as coherent and more progressive. However, without the OIR and solution, Speaker 2 may have answered an unintended question.

When I refer to progressivity, I refer to the progressivity of the next turn. Longer-term progressivity, like the progressivity over the course of a whole conversation or semester, is not well studied, but does have a different relationship with intersubjectivity. If intersubjectivity lapses early on in a project but the teammates do not notice, any short-term progress is erased once the interlocutors identify the miscommunication. However, I assume that interlocutors prioritize specifically the immediate progressivity of the sequence.

Interlocutors could maximize intersubjectivity by repeatedly producing OIRs, never making any progress. Alternatively, interlocutors could maximize progressivity by ignoring their, and their partners', understanding. Neither extreme is practical; likely, interlocutors balance these priorities depending on the context. Any time interlocutors sacrifice some amount of intersubjectivity for progressivity – so, all the time – they may *miscommunicate.* Therefore, when people prioritize progressivity more, they should miscommunicate more.

Heritage (2007) found evidence that interlocutors prioritized progressivity over intersubjectivity when producing or interpreting references. In most cases, recipients of references (e.g., "Julia Mertens") do not explicitly confirm their understanding. When prompted (e.g., “you know Julia Mertens?”), listeners typically confirm their understanding with a minimal response (e.g., “yes”). Listeners only provide strong, positive evidence of understanding of a reference (e.g., “yeah, the fifth-year student who just defended right?”) when speakers try for recognition at least three times. This evidence suggests that the recipient of a reference only relaxes their preference for progressivity when there is a clear need for more evidence of understanding. When intersubjectivity lapses, it creates the possibility for miscommunication. Therefore, the preference for progressivity may contribute to miscommunication.

Psycholinguists have long known that comprehenders predict language; Chapter III investigates whether prediction contributes to progressivity but decreases intersubjectivity in conversation. Similarly, Participants are egocentric in experimental environments; Chapter IV investigates whether egocentricity contributes to communication problems and progress in conversation. Together, these studies explore whether interlocutors prioritize progressivity over intersubjectivity, resulting in miscommunication.

Chapter II: Common Methods Across Studies

In this chapter, I will describe the methods that Chapters III and IV share. First, I will describe the In Conversation Corpus (ICC). Then, I will describe how I calculated turn transition times. Third, I will describe how I used the second Generative Pretrained Transformer (GPT-2) to calculate the surprisal and probability of words in the ICC. Fourth, I will describe the collection of other-initiated repair (OIR) and miscommunication sequences. Finally, I will describe Generalized Additive Mixed Models (GAMMs), a statistical technique used in both studies.

In Conversation Corpus

The ICC was collected by the Human Interaction Laboratory at Tufts University in Medford, Massachusetts. The ICC contains 100 conversations (47 hours) between dyads. Participants were Tufts University undergraduate students taking one of two required courses for a Psychology major (Introduction to Psychology and Statistics for Behavioral Sciences). Participants were recruited using the Tufts University Psychology SONA system. The only inclusion criterion was that the participants had to be fluent in English. Participants were either asked to recruit a friend to sign up for the same time slot or were allowed to sign up freely. Most participants asked to recruit a friend did so, but occasionally a stranger signed up for their time slot. In addition, since the participant pool was relatively small, some participants who were not asked to sign up with a friend ended up talking to someone they knew. Therefore, I categorized dyads as “friends” or “strangers” based on the beginning of their conversation. For example, conversations that began with the participants asking each other’s names were classified as strangers. Strangers made up 34 dyads. The data collection spanned three semesters: Fall 2017, Spring 2018 and Fall 2018.

When participants arrived, the researcher collected informed consent and led the participants to the recording rooms. The recording rooms were adjoining, and the shared wall was a large window. The rooms were mirror images of each other. Each room contained an artificial tree, a painting, a desk, and a Sony PMW-300K1 video camera that recorded the opposite participants' face. Participants spoke into Shure MX153T/O-TQG Omnidirectional Earsets and heard each other though Sennheiser PX 200-lli headsets. This configuration separated the audio from each participant while maintaining visual and auditory communication channels. The researchers instructed the participants to talk in English about anything they wanted for twenty to thirty minutes.

I sampled 42 conversations (19.6 hours) from the ICC, 18 stranger dyads (8.9 hours) and 24 friendly dyads (10.6 hours; Table 1). I used GailBot (Umair et al., 2022), an automated speech-to-text software that also represents overlaps and timing information, to produce a first-pass transcript and corrected the words and timings using CLAN (MacWhinney & Wagner, 2010) and Praat (Boersma, 2006). I refer to this collection as the ICC subset (Table 1).

Table 1

The ICC Subset

	Friends			Strangers	
	Fall 2017	Fall 2018	Spring 2018	Fall 2017	Fall 2018
Duration (min)	30.69	346.16	260.06	423.35	114.11
N	1	14	9	14	4

Floor Transfer Offset

Floor Transfer Offset (FTO; De Ruiter et al., 2006) quantifies the time between turns. FTOs are calculated by subtracting the end time of the first turn from the start time of the second turn (Equation 1).

Equation 1

General equation for Floor Transfer Offset

$$FTO = t_2\ start\ time - t_1\ end\ time$$

While this formula appears simple, it requires defining the "start time" and "end time" of tuns. People may signal the upcoming start of their turn, for example with an inbreath or lip smack. They may also start a turn with a prefaces (e.g., "well") or pause marker (e.g., "uh"), delaying their first content word. So where does a turn start? Kendrick & Torreira (2015) analyzed FTOs with three start times: (a) the first inbreath or turn-related noise, (b) the start of the first word of the turn, or (c) the start of the first turn while excluding prefaces (e.g., "well") or pause markers (e.g., "uh"). They found that FTOs using the start time of the first word (b) resulted in FTOs sensitive to the social aspects of the turn. Therefore, I defined "start time" as the beginning of the first word in a turn.

The "end time" of a turn is also difficult to operationalize. Speakers produce Turn Construction Units (TCUs), which are units of talk that are perceived as complete (Sacks et al., 1974). Therefore, one definition of a turn's end time is the end time of the last word in the last TCU before the speaker transition. However, listeners frequently respond at Transition Relevance Places (TRPs; Schegloff, 1982), places where speakers could finish a TCU but may or may not continue (Clayman, 2013; Sacks et al., 1974). The speech leading up to the TRP could be syntactically, pragmatically, prosodically, and otherwise complete. Especially common at TRPs are *backchannels*, ways that listeners participate in the unfolding turn, including “mhm,” “yeah,” or a head nod (Schegloff, 1982; Yngve, 1970). In Excerpt 2, lines 1-2, 4-5, and 7 make up one TCU that contains several TRPs, including those located at the ends of line 2 and 5. Speaker 2 jumps in to say “oh right,” slightly before the TRP at the end of line 2, and “oh really” slightly before the end of line 5. If I defined “end time” as the end of the TCU, these two turns would have extremely negative FTOs, indicating a long overlap. However, Speaker 2’s responses were placed soon after the speech that they responded to. Therefore, I defined the “end time” as the TRP closest to the start time of the second turn. In Excerpt 2, I calculated the following FTOs: (1) start time of line 3 – end time of line 2, (2) start time of line 4 – end time of line 3, (3) start time of line 6 – end time of line 5, and (4) start time of line 7 – end time of line 6.

Excerpt 2

Transcript with TCU broken up by TRPs for FTO calculation

```
1 SP1: so I have like a lot of u:m: (0.3) Δclasses i need to catch upΔ
2      ⌈o:n,⌉
3 SP2: ⌊oh   ⌋ri:gh⌈t,⌉
4 SP1:              ⌊a⌋nd I'll probably be taking like fi:ve every semester for
5      the rest of ⌈tufts, ⌉
6 SP2:              ⌊oh real⌋⌈ly.⌉
7 SP1:                      ⌊so ⌋i just don't feel like comfortable going abroad,
```

Even this refined definition of “end time” – the time of the TRP closest to the start time of the second turn – covered many, but not all cases. If the second turn started less than 200ms into the first turn, I considered the “end time of the first turn” to be the previous TRP, even if the next TRP was closer in time to the start time of the second turn. For example, in Excerpt 3, line 3 started only 37ms into line 2. This is not enough time for Speaker 2 to realize Speaker 1 is speaking and respond accordingly. I calculated no FTO before turn 3.

Excerpt 3

No Floor Transfer Offset calculated

```
1 SP2: the regular: psych building?
2 SP1: o⌈kay.          ⌉
3 SP2:  ⌊i think at ele⌋ven?
```

In Excerpt 4, Speaker 1 initiates “yeah” (line 4) almost immediately after Speaker 2 initiates “yeah” (line 3). Therefore, the FTO before line 4 was 991ms, the time between the end of line 1 and the start of line 4.

Excerpt 4

Floor Transfer Offset calculated from previous turn

```
1 SP2: yeah i: i couldn't do that,
2      (1.0)
3 SP2: ⌈yeah⌉
4 SP1: ⌊yeah⌋
```

Information, probability and surprisal

Shannon & Weaver (1964) presented a mathematical theory of communication that remains influential to this day. With the goal of measuring the maximum information a system can transmit without risking miscommunication, Shannon & Weaver (1964) quantified information as the number of binary digits required to encode a specific message. If the speaker chooses a message from two options (e.g., "yes" or "no"), they only require one binary digit to encode their message and therefore their message is less informative. If the speaker chooses a message from eight options, they require three binary digits to encode their message and therefore any message conveys more information.

In addition, *highly constraining* environments, which lead to strong predictions about upcoming language, result in different patterns of information than other environments. This is because predictable language is less informative. For example, a Declaration of Consent at the end of a wedding ("I do") contains little information because there is a close to 100% probability that the couple will say those words. At the same time, language that violates a prediction will convey a lot of information. For example, in the unlikely event that the wedding ends in "I don't," the audience learns a lot of new information. Overall, very constraining environments will typically lead to very uninformative language; rarely, a very constraining environment will lead to extremely informative language. The most informative turns are not only unlikely, but extremely unlikely alternatives to probable turns.

In *minimally constraining* environments, no strong predictions are possible. Instead, there are many equally probable possibilities for upcoming information. No matter what, a minimally constraining environment will never result in as much information as saying "I don't" at the end of a wedding ceremony. However, most of the time the speaker in an un-constraining environment will convey more information than a speaker in a constraining environment.

Early information theory research focused on quantifying the information contained in systems, like telegraphs. More recently, cognitive science and psycholinguistic research has focused on studying what information means for comprehenders of language. Hale (2001) proposed *surprisal theory*, which posits that cognitive processing is proportional to the probability of all disconfirmed possible models of the utterance. When the turn is very unlikely, the listener dramatically updates their predictions for the rest of the turn, requiring much more cognitive processing. Surprisal theory has recently been challenged by findings that cognitive processing is inversely proportional to the probability of the supported model of the utterance (Brothers & Kuperberg, 2021). However, the majority of the work in the field has used surprisal instead of probability (e.g., Hale, 2001; Levy, 2008; Levy & Jaeger, 2007). I will analyze surprisal (Equation 2) in Chapters III and IV.

Equation 2

Surprisal

$$surprisal(u) = -log(P(u|context)$$

While psycholinguists study the relationship between surprisal and cognitive processing, there is no work linking the amount of cognitive processing to most of the outcome measures in Chapter II. Therefore, I will analyze eight turn-level metrics of surprisal. I will calculate these metrics using word probabilities calculated by a language model (see the Generative Pre-trained Transformer Section). One common metric is *information density* (Jaeger, 2010; Frank & Jaeger, 2008; Levy & Jaeger, 2007), the surprisal of a unit divided by the duration of the unit. Speakers strive for a high, uniform information density. For example, speakers choose the shorter version of a word (e.g., chimp vs. chimpanzee) when that word is highly predictable (Mahowald et al., 2013), and lengthen words that are surprising (Bell et al., 2003). However, speakers may pause during their turn, reducing the information density even if they produce very informative words. Therefore, I will also compute mean word surprisal.

Averages are susceptible to outliers. Outliers will especially influence information density and average surprisal of short turns. Therefore, I will also calculate the median surprisal. It is also possible that listeners are more sensitive to the variance of information in the turn. I will calculate the maximum and minimum surprisal, as well as the surprisal range. Finally, I will analyze the surprisal of the first and last words. I will explore whether some metrics predict trouble source TCUs and/or FTOs more accurately than others.

Generative Pre-trained Transformer. I will use a Generative Pre-Trained Transformer (GPT; Radford et al., 2018; Vaswani et al., 2017) model to calculate the probability of words in the ICC. GPT is a language model, which means it determines the probability distribution over a series of words. "Generative" means that GPT returns the joint probabilities of the context and different possible outcomes, as opposed to labels (as in classification), or predicted values for the outcome variable (as in regression). The most common use of GPT is to produce language: it uses the existing linguistic context to create a distribution of probabilities for the next word and selects the next word accordingly. I will not use GPT to produce language; instead, I will use GPT to calculate the probability of every word.

I used GPT-2 large in Chapters III and IV (Radford et al., 2018). This version of GPT-2 has 762 million parameters and represents each token with a word embedding with a length of 1280. Since GPT-2 has so many parameters, it must be trained on a lot of data. It would be very time consuming to collect and transcribe enough conversations to train GPT-2 from scratch. Luckily, "Pre-trained" means that researchers can download a GPT-2 already trained on language scraped from 8 million websites. Specifically, 8 million websites linked to on Reddit posts that received positive responses from readers. Ideally, this pre-training allows GPT-2 to transfer its general knowledge to new tasks, with no or limited fine-tuning.

Finally, “transformer” refers to a type of language model that uses “attention” to focus on the most important parts of the context. Since transformers learn which parts of the context are most important, they know which parts to ignore. This means transformers can use more context than other models. Early work on information density in conversation used trigram language models (Levy & Jaeger, 2007), which use two words of context to estimate the probability of the current word. In contrast, GPT can process up to 1024 tokens (which map roughly onto words) of context.

GPT-2 achieved state-of-the-art performance on traditional tests for language models (Radford et al., 2018). In addition, growing evidence suggests that GPT-2 accurately estimates listener predictions. For example, GPT-2 probabilities correlated with the neurophysiological responses associated with the surprisal induced by listening to an audiobook (Caucheteux et al., 2021b, 2021a; Heilbron et al., 2019). They also correlated with reading times (Wilcox et al., 2020). While these results suggest that GPT-2 could represent the predictions of someone listening to an audiobook, there is no evidence showing GPT-2 can represent interlocutors during spoken, interactive language. Next, I describe how I fine-tuned GPT-2, and how I evaluated GPT performance when predicting language in conversation.

Fine-tuning GPT-2. Since GPT-2 was trained on written data from websites, it must be *fine-tuned* to be able to predict language in conversation. Fine-tuning type of training that allows models to retain what they have already learned while learning a new task (Czum, 2020). Therefore, fine-tuning requires much less data than training a model from scratch. First, the output layer is set to random weights. Then, that layer is trained from scratch on the fine-tuning data, and the other layers only slightly change their weights.

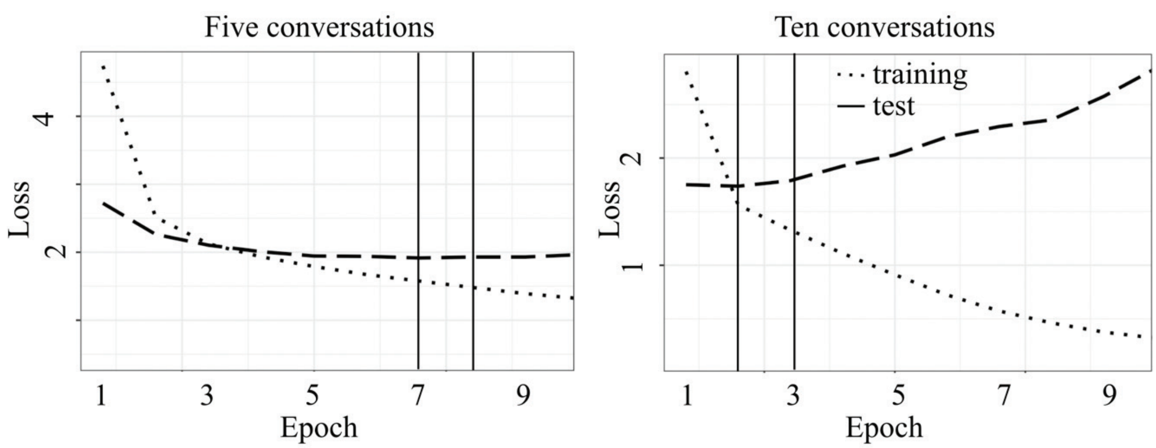
The data used to fine-tune GPT-2 cannot be used for hypothesis testing in Chapters III and IV because GPT-2 will always be less surprised by data it has already seen. I demonstrated this by finding that the information density was lower for five conversations used to fine-tune GPT-2 (Mdn = 5.68) than for the 37 unseen conversations (Mdn = 5.89, $W > 1000$, $p < 0.01$). In addition, if the model is fine-tuned for too long, it will overfit the training data and incorrectly predict the test data. Therefore, I tested GPT-2 performance when fine-tuned on different amounts of training data for different durations in the GPT-2 Test Sections below.

It may be that no amount of fine-tuning will be enough for GPT-2 to accurately model language in conversation. One important assumption of transfer learning, where a model learns on lower-cost data and that training supports performance on higher-cost data, is that the lower-cost data (written, online language) accurately represents aspects of the higher-cost data (transcribed language in interaction). There are large enough differences between language written on websites and language spoken in conversation that we cannot assume that transfer learning will work. Instead, GPT-2 may need to be trained from scratch on transcribed conversations. To ensure that a fine-tuned GPT-2 can model listeners in conversations, I designed two tests to determine whether GPT-2 can model speaker identity and preference structure. I used these tests to determine exactly how much and how long to fine-tune GPT-2, and whether any fine-tuned GPT-2 would be appropriate.

I trained GPT-2 on five and ten conversations, and tested GPT-2s performance on a test set of five conversations. The training and test data had similar proportions of friends and strangers as the ICC subset. I recorded the training and test cross-entropy loss after each epoch (round through the training data; Figure 2). Cross-entropy is a metric that analyzes the average number of bits needed for the model to produce the observed messages. Lower cross-entropy suggests that the model is less surprised by the language and is likely a better model. Training loss decreases with more training and approaches an asymptote. Test cross-entropy loss decreases initially but starts to increase when the model becomes overfit. As Figure 2 shows, when trained on five conversations, GPT-2s test loss started to increase slightly after seven epochs. When trained on ten conversations, GPT-2s test loss increased almost immediately, after just two epochs. Therefore, in the GPT-2 tests below, I tested four GPT-2 models: GPT-2 fine-tuned on five conversations for seven or eight epochs, and GPT-2 fine-tuned on ten conversations for two or three epochs.

Figure 2

GPT-2 loss when trained on five or ten conversations



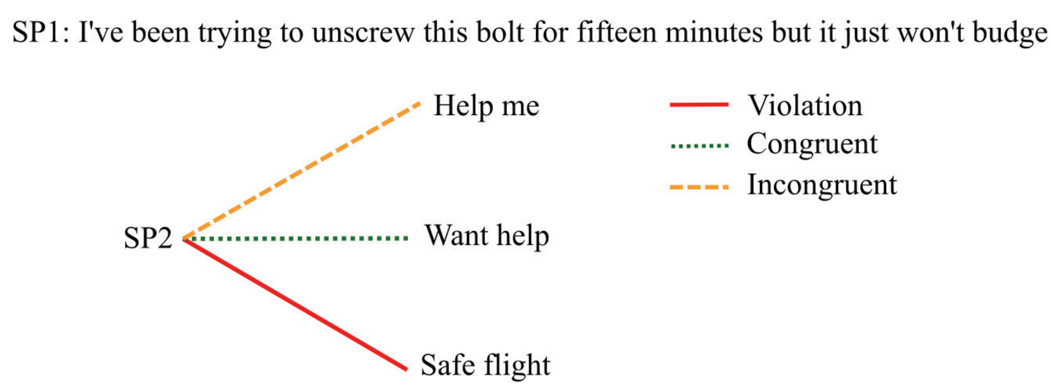
Note: Vertical lines represent the models tested below.

GPT-2 Test: Speaker Switches. Unlike language on websites or books, all language in conversation is produced by speakers in a *turn-taking system* (Sacks et al., 1974). If the current speaker does not specify who should speak next (e.g., “Julia what do you think?”), either the current speaker can continue speaking or another interlocutor can jump in. Depending on who takes the next turn, different types of turns will be appropriate; for GPT-2 to accurately model listener surprisal, it must know what is appropriate for the same or a different person to say.

To determine whether GPT-2 could use speaker switches to influence the probability of turns, I extracted GPT-2 output from 125 groups of stimuli from Warnke (2022). Each group contained three two-turn sequences that differed in congruence (violation, congruent and incongruent). Figure 3 shows an example group of stimuli, when the first turn is a complaint about unscrewing a bolt. The second turn “help me,” is *incongruent* because it would have made sense if spoken by Speaker 1 but does not make sense when spoken by Speaker 2. The second turn “want help,” is *congruent* because it makes sense when spoken by Speaker 2, but not speaker 1. Finally, the second turn “safe flight” is a violation because it wouldn’t make sense no matter who spoke the turn. The same turns were used in different stimulus groups to control for word-specific characteristics.

Figure 3

Stimuli for speaker test



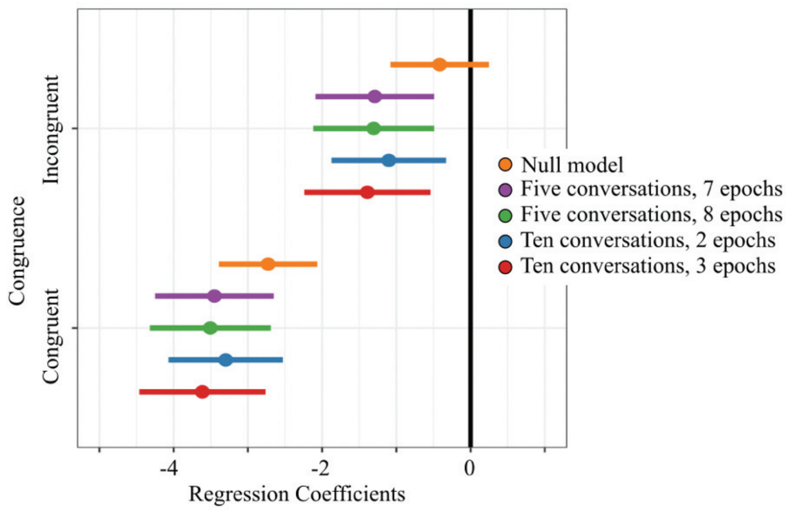
Human participants listened to these sequences and pressed a button when they believed the second turn was finished (Warnke, 2022). In this type of end-of-turn detection task, slower responses indicate that the turn was harder to process and/or less predictable. In Warnke (2022), participants responded slowest in the violation condition. This finding suggested that participants predicted the lexicosemantic content of the second turn, facilitating the processing of the congruent and incongruent second turns. In addition, participants responded fastest in the congruent condition. This finding suggested that participants predicted the set of plausible speech acts for the specific speaker, facilitating the processing of congruent second turns. GPT-2 probability values should be inversely related to the reaction time of participants in Warnke (2022): the congruent condition should be the most probable and the violation condition should be the least probable.

I passed the stimuli as input to each of the four fine-tuned GPT-2 models as well as the pretrained (null) model. I calculated the probability of the second turns by multiplying the probabilities of all words in the turn. The resulting turn probabilities were positively skewed, so I calculated the log transformed probabilities. I regressed the log probability of the second turn on congruence and speaker identity, while including a random effect for stimulus group.

Figure 4 shows the regression coefficients and 95% confidence intervals for the data produced by each model. Every trained model found the expected pattern: the congruent turns were more probable than the incongruent turns, and the incongruent turns were more probable than the violation turns. The null model found no difference between the congruent and incongruent turns, and a smaller difference between the congruent and violation turns than the trained models. This means that the effects found by the trained models, but not the null model, mimic those found by Warnke (2022). These findings show that – only once fine-tuned – GPT-2 can reproduce the effect of speaker switches on the probability of a turn.

Figure 4

Regression coefficients for speaker identity test

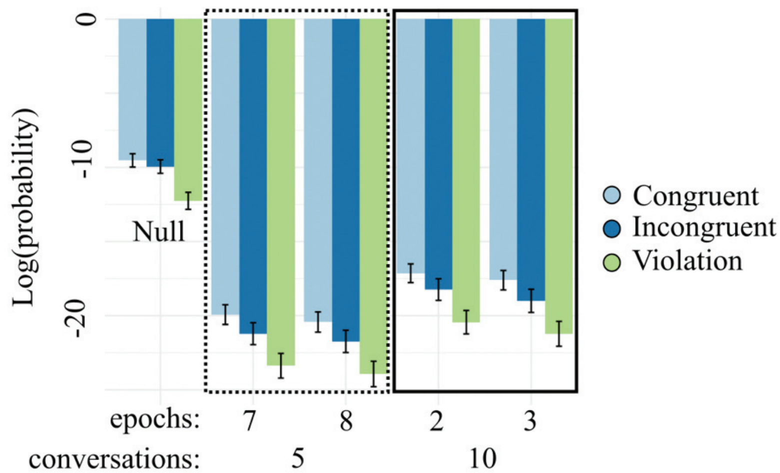


Note: A negative estimate represents a less likely second turn as compared to a turn in the congruent, different speaker condition.

Figure 5 displays the log probabilities for the three congruence conditions split by GPT-2 model. The null model provided significantly higher probability values than the trained models. While there was little difference in the effect size when analyzing probability values from GPT-2 trained on five or ten conversations, there was an apparent effect on the baseline probabilities: when the model was trained on five conversations, the model estimated every turn as less probable than when the model was trained on ten conversations.

Figure 5

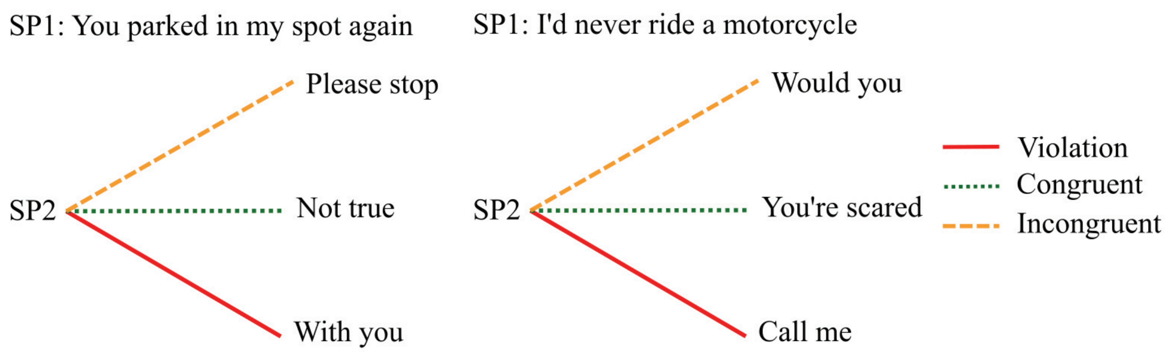
Log probabilities of congruent conditions from GPT models



Finally, I conducted a preliminary qualitative analysis of the stimuli that did or did not produce the expected pattern (congruent more likely than incongruent, incongruent more likely than violation). Only approximately 34% of the stimuli groups followed the expected pattern (e.g., Figure 6, left), according to the GPT-2 model trained on five conversations for seven epochs. The next most common pattern is displayed in Figure 6, right: in 28.8% of the groups, the incongruent second turn was more probable, and the congruent condition was the least probable. There are many reasons why this may be the case. One interesting possibility is that there may be differences in the stimuli, and participants may have responded similarly. Future research should directly correlate GPT-2 output to participant responses to see how well GPT-2 models participants. Perhaps responded later to the congruent conditions in the same stimulus groups that GPT-2 estimated to be improbable.

Figure 6

Stimuli that matched (left) or violated (right) speaker switch predictions



Notes: A stimulus group where the congruent condition was most probable and the violation condition was least probable (left), and a stimulus group where the incongruent condition was most probable, and the congruent condition was least probable (right).

GPT-2 Test: Preference Structure. When interlocutors respond to a turn, they usually choose between several actions. For example, when one person produces an assessment ("That was such a great meal" in 1a and 1b), the other can choose to agree (second turn in 1a) or disagree (second turn in 1b). Typically, one action is *preferred*, meaning that it promotes social affiliation more than the other(s) (Clayman, 2002; Pillet-Shore, 2017; Schegloff, 2007). Importantly, preference in this context does not refer to ones' opinion; the preferred response to an invitation is an acceptance, even if the invitee would rather not attend (or even if the inviter would secretly rather the invitee not attend). Approximately two thirds of responses to actions are preferred (Kendrick & Torreira, 2015), so listeners likely use the heuristic of predicting that any given response will be preferred. In other words, the second turn in 1a should be more likely than the second turn in 1b.

1a. Agreement

*SP1: That was such a great meal

*SP2: It was delicious

1b. Disagree condition

*SP1: That was such a terrible meal

*SP2: It was delicious

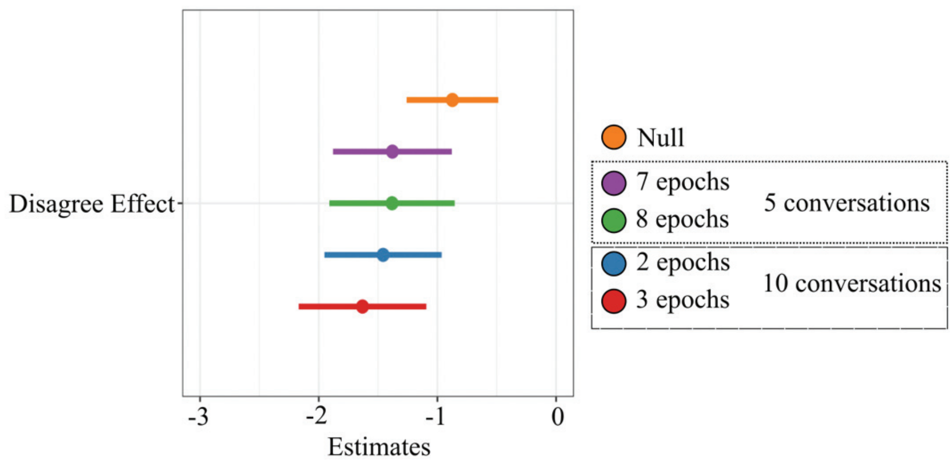
When speakers do produce a *dispreferred* turn (the second turn in 1b), they usually construct the turn in a way that mitigates the effect of the action (Pomerantz, 1984). The speaker may delay initiating the turn, produce turn-initial delays (“um”) or qualifiers (“well”), hedge (“maybe I can come by later?”), and/or apologize. A direct dispreferred response, like an immediate “no,” is relatively rare. Therefore, the second turn in 1b – a direct dispreferred response – should be especially unlikely. In this test, I determine whether GPT-2 is sensitive to preference: does GPT-2 produce lower probability values for direct dispreferred responses than for direct preferred responses?

The Human Interaction Laboratory designed 49 pairs of two-turn sequences like 1a and 1b. Within each pair, the second turn was the same. The first turn changed so the second turn either agreed or directly disagreed. If GPT-2 mimics listeners in conversation, it should find a direct preferred response more predictable than a direct dispreferred response.

I used GPT-2 to extract the probability of each word in the second turn and multiplied the words to get the probability of the entire second turn. Again, the probabilities were positively skewed, so I regressed the log probability on condition (agree vs. disagree) and included a random intercept for stimulus pair. All models, including the untrained model, found that the same turn was more surprising when it was a direct dispreferred response. Figure 7 displays the estimated effect of disagreement, along with the 95% confidence interval, for the data produced by each GPT model.

Figure 7

Regression coefficients from preference structure test

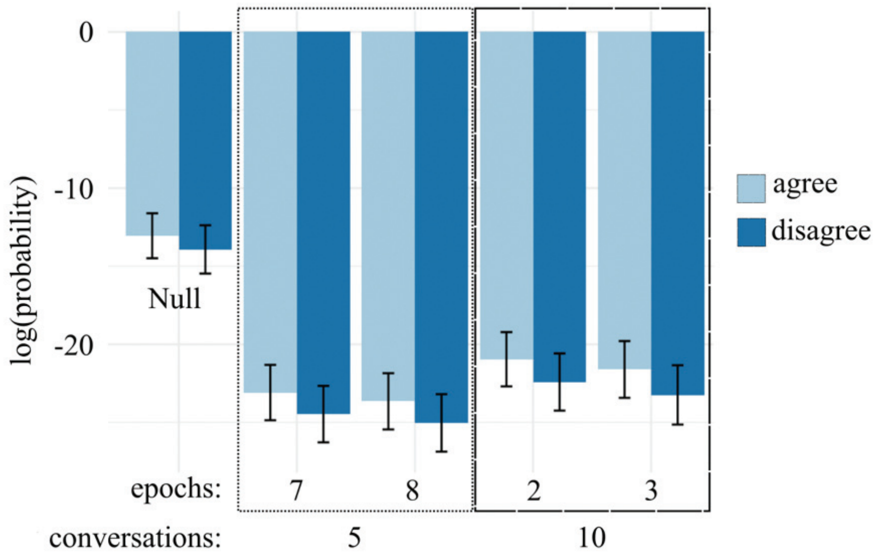


Note: A negative estimate represents that the disagree condition had a lower log probability than the agree condition for the model.

Figure 8 displays the log probabilities for the agree and disagree conditions, split by GPT-2 model. It shows that the null model estimated the experimental turns as more likely than the trained models. Further, while the null model did estimate that the disagree condition was less likely than the agree condition, the magnitude of that difference is less than for the trained models.

Figure 8

Log probabilities of (dis)agreements from each GPT-2 model

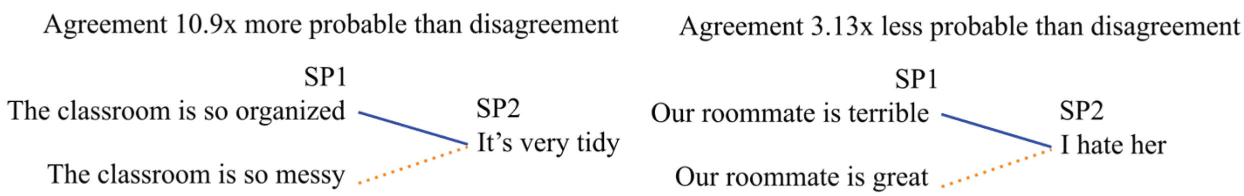


Note: The error bars represent 95% confidence intervals.

For most (39, 70.6%) stimulus pairs the GPT-2 model trained on five conversations for seven epochs found the agreement more probable than the disagreement. I examined the pairs with the biggest difference between the agreements and disagreements. GPT-2 found the agreement to be more probable (left) or less probable (right) than the disagreement. For stimuli pairs where the agreement was considered more probable than the disagreement, the top four biggest differences included an infrequent word (e.g., “tidy” in Figure 9, left). For stimuli pairs where the disagreement was considered more probable than the agreement, the pairs with the top three biggest differences were focused on emotions or opinions (e.g., “hate” in Figure 9, right). Exactly why GPT-2 produces these patterns is unclear. As hypothesized in the GPT-2 Test: Speaker Switches Section, it is not impossible that there is a difference in how acceptable a disagreement is when it involves emotions. However, more research is needed to explore whether human participants perceive these differences, or if GPT-2 is affected by an unknown confound.

Figure 9

Example stimuli where agreement was less/more probable than disagreement



Note: The blue line represents agreement, the orange dotted line represents disagreement.

GPT-2 Test: Word-level probabilities. Next, I determined whether GPT-2 output mapped onto qualitative assessments of probability. GPT-2 was the most surprised by uncommon proper nouns (e.g., 2a-b) and out-of-dictionary words that could not be split into meaningful subparts (e.g., 2c). When GPT-2 encounters out-of-dictionary words, it attempts to split the words into sub-tokens, which may or may not map onto morphemes. If those subparts map onto more predictable units, then an out-of-dictionary word will be less surprising. However, out-of-dictionary words are typically surprising and rarely predictable.

2a. my mom does a lot of um she used to do like genetic modification for um **meadwestvaco**

2b. he's my idol it was like it was a firm called **novantis**

2c. my dad will go in and be like **errerererer**

GPT-2 strongly predicted words in the recent context (3a) and at the end of common phrases (3b). GPT-2 occasionally produced extremely high probability values for predictable, but not inevitable words. For example, “couple” in 3c could also have been “week,” “day”, or “few months,” but GPT-2 assigned “couple” a probability close to 1.

3a. you couldn't tell what was trash and **what**

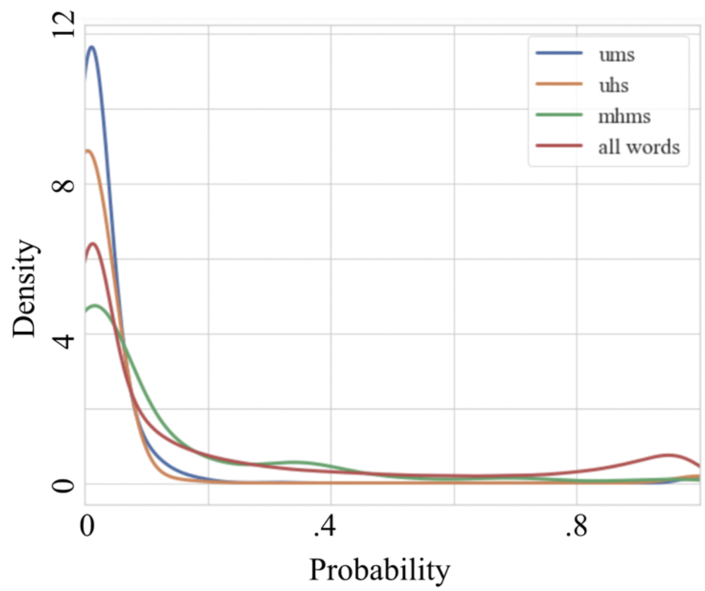
3b. i can't even imagine what that would feel **like**

3c. i still talk to my two best friends from it every **couple**

Another way conversation is different than most written language is that conversation is rife with speech particles. For example, “uh” and “um” project upcoming silence (Clark & Fox Tree, 2002) and “mhm” encourages the speaker to continue talking (Schegloff, 1982), even though these particles do not have standard dictionary entries. Figure 10 shows the probabilities of the “um”s, “uh”s, “mhm”s, and all words in the ICC subset. There were a few notable patterns: “um”s were less probable than “uh”s, even though there were more “um”s (n = 874) than “uh”s (n = 357). This may be because “um”s at the beginning of turns signal more trouble than “uh’s” (Clark & Fox Tree, 2002). In addition, both “um”s and “uh”s were less probable than the average word. In contrast, “mhm”s (n = 385) were more likely than the average word, perhaps because they provide evidence that the conversation is going well.

Figure 10

Probabilities of speech particles



Excerpt 5 illustrates an important limitation of GPT-2: it estimates the average overhearer, not the exact listener. GPT-2 gave “Deloitte” (line 2) a probability less than 0.01. While many people may not know of Deloitte, lines 3-4 show that Speaker 1 does know of Deloitte, and therefore was unlikely to be surprised. The second time “Deloitte” appears (line 4), GPT-2 assigned a probability of 0.09. While this is nine times more probable than the first instance of “Deloitte,” Speaker 2 could have strongly predicted that Speaker 1 was about to say “Deloitte.” This phenomenon introduces error into GPT-2 probability estimates: more often, participants produce a word they understand but GPT-2 does not, but it is also possible that GPT-2 understands rare jargon words that the participants do not. Future research should investigate how influential and problematic this phenomenon is; hopefully, the large amount of data in this study will mitigate the effect of this noise.

Excerpt 5

Probability example

```
SP2: i'm choosing between a firm called oliver weinmann: and
     another one called deloitte: (0.5) so ⌈ i ⌉
SP1:                                       ⌊oh.⌋ yeah. i know
     deloitte,
     (0.3)
SP2: ye⌈ah    i   just  ha⌉ve to make a decision some point in
     the next week
SP1:   ⌊i've heard of that⌋
```

Excerpt 5 also demonstrates that GPT-2 struggles to deal with overlapping talk. GPT-2 can only process one line at a time, so it processed the entirety of lines 6-7 before processing line 8. Therefore, GPT-2 is more surprised by "I've heard of that" than it should be, producing a median word probability value of 0.04. This problem introduces systematic error into the analyses in Chapters III and IV, in that it selectively affects the probability of overlapping turns. However, it likely affects the variability of the probability of overlapping turns, as opposed to systematically increasing or decreasing the probability of overlapping turns. However, these two caveats are important and should be investigated in further research.

GPT-2 Test Summary. GPT-2 passed the speaker identity and preference structure tests. The amount of training data (five or ten conversations) and the number of epochs did not meaningfully affect performance on those tests. Therefore, I decided to use GPT-2 trained on five conversations for seven epochs.

Next, I qualitatively analyzed the probability estimates produced by the final model. The words GPT-2 found most predictable or surprising did seem to be more predictable or surprising. However, this examination uncovered some downsides to using GPT-2, namely that it is a more accurate model of an overhearer than the specific listener of a turn. Further, GPT-2 cannot process the timing of turns, occasionally resulting in poor probability estimates for overlapping turns. However, these downsides are not unique to GPT but are downsides of using any current language model to estimate predictions.

Other-Initiated Repair Sequences

I identified and categorized all OIRs in the ICC subset. All OIR sequences have two elements (Dingemanse et al., 2016). First, one interlocutor produces the *trouble source TCU*, or the TCU targeted by the OIR. Then, the OIR requests a repeat, reformulation, or elaboration of the trouble source TCU. OIR sequences typically, but do not always, include a *repair solution*, speech that resolves the OIR. Any interlocutor can produce the repair solution. To identify trouble source turns, OIRs, and solutions, I used the *next-turn proof procedure* (Sacks et al., 1974): I labeled actions based on how the interlocutors treated the utterances. For example, I did not consider Excerpt 6**Error! Reference source not found.** an OIR sequence, even though repeats can initiate repair, because neither interlocutor treated line 3 as an OIR.

Excerpt 6

Repeat that is not an OIR

```
1   SP2: ≈was it good,
2   SP1: eh not goo⌈d,⌉
3   SP2:           ⌊n⌋⌈ot    good   ⌉
4   SP1:              ⌊but it was re⌋ally funny
```

In addition, some OIRs can accomplish other social tasks than initiating repair (Kendrick, 2015a). Response tokens (e.g., “really?”, Excerpt 7 lines 3 and 5), pro-repeats (e.g., “you did?”), ritualized expressions of disbelief (e.g., “for real?”) and corrections (e.g., “you mean tomorrow, not Wednesday”) all suggest that the person both heard and understood what the other person said, but that the content was surprising, newsworthy, or incorrect. I did not include these types of OIR sequences in this collection, because my goal was to analyze potential miscommunications, not newsworthy or surprising content. The resulting collection of OIR sequences included 368 trouble source TCUs (8.8 per conversation), 382 OIRs (9 per conversation), and 344 solutions (8.2 per conversation). In Chapter II, at different times I will analyze trouble source TCUs not used to train GPT-2 (n = 316; see Generative Pre-trained Transformer Section below) and trouble source TCUs not used to train GPT-2 that were followed by an OIR in the immediately next turn (n = 230). In Chapter III, at different times I will analyze all OIRs in the entire ICC subset (n = 382) and median values for each conversation (n = 42).

Excerpt 7

Response token that was not categorized as an OIR

```
SP2: it's a good area, dewick's better than (.) carm s:⌈o⌉
SP1:                                                   ⌊o⌋h
     rea:l⌈ly.⌉
SP2:      ⌊ye⌋ah,
SP1: rea:lly. that's ho:tly contested
```

There were more OIRs than trouble source TCUs because multiple OIRs can target the same trouble source TCU. This is usually because an interlocutor *upgrades* their OIR. For example, the trouble source TCU “Testing one two three” was followed first by “What?” and then by “One two three?”. There are also more trouble source TCUs than solutions. This may be, in part, because I only transcribed verbal behavior, and solutions, like nodding or shaking of the head, can be nonverbal. Occasionally, the interlocutor who produced the trouble source TCU will continue the conversation instead of solving the OIR. In Chapter III, I analyze the probability of trouble source TCUs, and in Chapter IV, I examine the frequency of OIRs.

Most (275, 74.73%) but not all trouble source TCUs were followed by an OIR in the next turn. Occasionally, an interlocutor produced a minimal response (e.g., “okay”) or expression of surprise (e.g., “oh really”) before producing an OIR. Other times, the interlocutor recognized the problem after beginning their turn (e.g., Excerpt 8). In Chapter III, I compare the turn transition times between trouble source TCUs and OIRs to the turn transition times between other turns.

Excerpt 8

Delayed recognition of trouble source

```
SP1: ≈so: (0.4) °like° (0.4) i don't know.
SP2: ⌈stick with oreos,⌉
SP1: ⌊    se- w- ye.    ⌋ah. we're just
     (0.6)
SP1: did you just say stick with o:reos?
SP2: yeah
     (1.4)
SP1: you know they're vegan.
```

Miscommunication Sequences

While the term *miscommunication* has a clear meaning in day-to-day speech, it is a controversial term in the study of conversation. In Chapters III and IV, I define a miscommunication sequence as any sequence where intersubjectivity lapses, but the interlocutors do not notice. In a corpus analysis, the only way to know a miscommunication has happened is if it is obvious based in the talk. Therefore, I included two types of sequences in the collection. First and more frequently, the interlocutors temporarily behave as if intersubjectivity has been achieved. Later in the conversation, they signal that intersubjectivity has lapsed (e.g., “Oh, I thought you said…”). Second and rarely, intersubjectivity lapsed but the interlocutors signal a communication problem. Instead, there are clear but implicit cues indicating that intersubjectivity lapsed. In Excerpt 9, Speaker 2 asks Speaker 1 about what Speaker 1 does in the summers. Speaker 1 states that they sublet “from a friend like on campus,” but soon after, Speaker 2 asks where they lived over the summer. This means that Speaker 2 asks for information that Speaker 1 just provided in response to the previous question. While Speaker 1 does not explicitly state that intersubjectivity has lapsed (e.g., “As I already said…”), on lines 15-18 they hesitate before responding, produce a delay marker (“um”), pause once again, and then substantially elaborated on lines 6-7. These behaviors suggest that while Speaker 1 did not explicitly signal a lapse in intersubjectivity, they have identified one. Interlocutors do not always challenge incorrect statements; for example, an recipient may not explicitly address a speaking problem if they do not consider the problem important (Jefferson, 2007).

Excerpt 9

Miscommunication without explicit signaling

```
SP2: what have you been doing in the summers. just like d- do
     you do like a lot of like (0.2) m- pre med kind of stuff
SP2: or ⌈do you: just like⌉
SP1:   ⌊ so: after my f:⌋reshman year: (0.4) i volunteered ∆at a
     hospital in my hometown so I was just at home,∆ (0.6) and
     the:n after my sophomore year I was (.) subletting from a
     friend (0.5) like on ca:mpus,
SP1: a:nd i just did research here. (0.2) in two hundred boston
     avenue, (0.6) just like for the whole summe⌈r,⌉
SP2:                                           ⌊o⌋h. you were
     here during the ⌈summer,⌉
SP1:                 ⌊ yeah  ⌋
SP2: where do you where do you stay. do you have to room here?
SP2: like ⌈uh,⌉
SP1:      ⌊ u:⌋m (.) ∆I was li:ving o:n Bo:ston Avenue,∆ (0.3) like
     one of my friends had an off campus house but he wasn't staying
```

```
17       here for the summer s:o I just like subletted his room from
18       hi:m,
19 SP2: oh:::.
```

In contrast to OIR sequences, miscommunication sequences require the interlocutors to (at least temporarily) respond as if they understand the information in the trouble source TCUs. The difference between the two is that in OIR sequences, the interlocutors signal the lapse in intersubjectivity almost immediately; in miscommunication sequences, they do not. Interlocutors may produce OIRs during a miscommunication sequence, but miscommunication sequences do not always include OIR sequences and most OIRs occur outside a miscommunication sequence.

There were 43 miscommunications in the ICC subset, approximately 2.2 miscommunications per hour. In this section, I describe some patterns in the miscommunications and compare them to other miscommunication collections.

The most well-studied and common type of miscommunication sequence is also the shortest: *third position repair sequences* (Schegloff, 1987, 1992). In these sequences, one interlocutor responds to a turn as if they understand, but instead exposes their lack of understanding (Excerpt 10, line 3). The speaker of the trouble source TCU (Excerpt 10, line 1) then notices and repairs the problem (Excerpt 10, lines 4-5). My collection contained 22 (51.16%; 1.12 per hour) third position repairs sequences.

Excerpt 10

Third position repair, problematic reference

```
1   SP2: ah d- (0.6) ah (0.3) i can't do this,
2        (0.3)
3   SP1: yes you (0.4) absolute⌈ly can,⌉
4   SP2:                       ⌊   i   ⌋mean like be awake right now
5        like i'd rather be asleep
6        (1.1)
7   SP1: ∆i thought you just meant∆ like talk for like thirty
8        minutes and i was like yes you can,
```

Schegloff (1987) categorized the problems leading to third position repairs into two groups: problematic references ("this" on line 1 of Excerpt 10) and sequential implicativeness problems. Sequential implicativeness problems occur when an interlocutor acts as if a turn performed a different action than the speaker intended to perform. In Excerpt 11, Speaker 2 asks "do you know how they work?" on line 2. Many "do you know" questions are ambiguous: they can be interpreted as a pre-telling, which paves the way for a telling ("They work by…"), but they can also be interpreted as a question that requests an answer. For this reason, "do you know" questions have already been connected to miscommunication (Schegloff, 1988). In this case, Speaker 1 treats line 2 as a pre-telling, but Speaker 2 states that line 2 was asking a question. Most (n = 37, 84%) miscommunications involved problems of reference, while only a few (n = 7, 16%) involved problems of sequential implicativeness.

Excerpt 11

Sequential implicativeness problem

```
1   SP2: this looks really thick that's why i- i really do think
2        this is a two way do you know how they work
3        (0.7)
4   SP1: um no how do they work huh
5        (0.9)
6   SP2: i was asking you
```

```
7       (3.0)
8  SP2: it just looks like
9       (1.6)
10 SP1: i feel like i should be able to see myself
```

I also categorized miscommunication sequences into 6 nonexclusive categories: problems of ambiguity, mishearing, mis-inference, joking/exaggeration, and context. In 18 (40.86%) of the miscommunication sequences, the trouble source involved an ambiguous word or phrase (e.g., “this” on line 1 of Excerpt 10). In 8 (18.60%; one every 2.45 hours) of miscommunication sequences, the listener either misheard or did not hear the speaker (Excerpt 12).

Excerpt 12

Mishearing

```
1  SP2: cause she lived in a forced tri:ple her freshman year?
2  SP1: mm⌈:,⌉
3  SP2:   ⌊so⌋ she had (0.4) you kn⌈ow:,     ⌉
4  SP1:                            ⌊a force⌋d tri:ple.
5       (0.9)
6  SP2: yeah. so it's like a double but three beds: (.) and three
7       ⌈    peo   ⌉ple
8  SP1: ⌊Δyou meanΔ⌋
9       (0.8)
10 SP1: yep (.) i heard forced si:ngl:e,
```

In 5 (11.63%; one every 3.92 hours) of miscommunications, the listener inferred incorrect information. For example, in Excerpt 13, Speaker 1 has “normal” (line 10) nail polish. Speaker 2 assumes that “normal” nail polish means that Speaker 1 painted her nails herself; on line 15 it becomes clear that Speaker 1 went to a salon, but just got “normal” nail polish.

Excerpt 13

Misinference

```
1  SP2: well your nails look nice at least,
2  SP1: my nails do?≈
3  SP2: ≈yeah they're shiny. is it shellac,
4       (0.3)
5  SP1: no.
6       (0.5)
7  SP2: what is it. (0.2) gel,
8       (0.6)
9  SP1: no. it's just (0.3) nai⌈l polish⌉
10 SP2:                         ⌊normal⌋?
11      (0.5)
12 SP1: yeah.
13      (0.3)
14 SP2: oh this is what happens when i do ⌈mine so⌉
15 SP1:                                   ⌊   oh   ⌋well i got them done.
16 SP2: oh.
```

In 6 (13.95%; one every 3.27 hours) miscommunication sequences, the speaker exaggerated or joked, but the listener perceived the turn to be serious. In Excerpt 14, Speaker 2 attempts to poke fun at the data collection process on line 2. Speaker 1 produces a minimal response ("yeah" on line 3) instead of laughing. Speaker 2 upgrades their joke to make it more extreme on line 5. Again, Speaker 1 takes them seriously, responding with an OIR on line 6.

Excerpt 14

Joking-serious problem

```
1  SP1: i've never done some shit like this.
2  SP2: yeah u- (0.3) oh i have l⌈ike,⌉
3  SP1:                          ⌊yea⌋h,
4       (0.3)
5  SP2: every weekend yeah,
6  SP1: what
7       (4.1)
```

```
8   SP2: i was just kidding. it was ⌈a joke,⌉
9   SP1:                            ⌊   oh w⌋hat
10       (0.4)
11  SP2: come on now.≈
```

In 8 (18.60%; one every 2.45 hours) miscommunication sequences, interlocutors used different context to produce and interpret an utterance. In each case, the speaker produced their turn in reference to more distant context, but listeners interpreted the turn in reference to more recent context. In Excerpt 15, “that’s unfortunate” was intended by Speaker 1 to respond to lines 1-4, as evidenced by their turn on line 13. However, Speaker 2 interprets that turn as referring to line 7, as indicated by lines 14-15.

Excerpt 15

Context problem

```
1   SP2: she was like walking around and she was like you're in my seat
2        you’re in fifteen and i was like (0.4) i'm in five i don't think
3        i'm in your seat (0.8) and then she tried to climb over and
4        spilled beer on someone
5        (0.6)
6   SP1: but it wasn't on you≈
7   SP2: ≈it wasn't on me
8        (0.3)
9   SP1: that's unfortunate
10  SP1: what time ⌈of day was it⌉
11  SP2:           ⌊what do you mean t⌋hat's unfortunate
12       (0.4)
13  SP1: well she spilled beer on som⌈ebody it's sticky and gross⌉
14  SP2:                             ⌊oh i thought you meant it's
15       unfo⌋rtunate that it wasn't on me
```

Generalized Additive Models

Utterances in conversation are not independent from each other. Speakers may produce all their utterances in a similar way; utterances from the same conversation involve similar topics. In addition, the relationships between variables analyzed in these studies were non-linear. When I fit linear null models, they produced skewed residuals. Non-linear, hierarchical data violate the assumptions needed for many statistical models. Therefore, I used Generalized Additive Mixed Models (GAMMs; Hastle & Tibshirani, 1992), with the mgcv package in R (Wood, 2017). GAMMs determine the optimal set of *smooth functions* to predict one outcome variable with a set of predictor variables (Hastle & Tibshirani, 1992).

The estimated degrees of freedom (edf) are the estimated number of parameters needed to model the data: an edf of 1 represents a linear relationship, an edf of 2 represents a quadratic relationship, and so on. The p-value for any GAMM predictor is the likelihood that the relationship between the predictor and outcome variable is a horizontal line. However, GAMMs do not provide an easy way to quantify the effect of any predictor. Instead, I will present plots of the effects estimated with GAMMs whenever I use them.

Chapter III: Prediction

Abstract

Listeners constantly predict incoming linguistic stimuli. Correct predictions reduce the cognitive load of comprehension, help the listener plan what to say earlier, and tell the listener exactly when to start speaking. These effects of correct predictions should help decrease gaps in conversation, increasing progressivity.

The conceptual link between incorrection predictions and miscommunication is clear. Depending on whether an incorrect prediction is challenged and how the comprehension system processes a challenge, an incorrect prediction may hypothetically lead to perceived or unnoticed lapses in intersubjectivity. However, no research has investigated this hypothesis using empirical data. In this study, I investigated whether surprising turns were followed by longer gaps, more likely to be trouble source TCUs, and followed by earlier other-initiations of repair. Finally, I explored the role of prediction in miscommunication sequences.

I found that more surprising turns were followed by longer gaps in conversation, unless they were followed by other-initiations of repair initiations. In addition, I found that surprising turns were more likely to become trouble source TCUs, indicating that listeners may struggle to understand information that challenges their incorrect predictions. Finally, I found that prediction contributed to miscommunication sequences in unexpected ways. These findings provide the first evidence that, while prediction increases progressivity, it can decrease intersubjectivity.

Listeners predict language in conversation

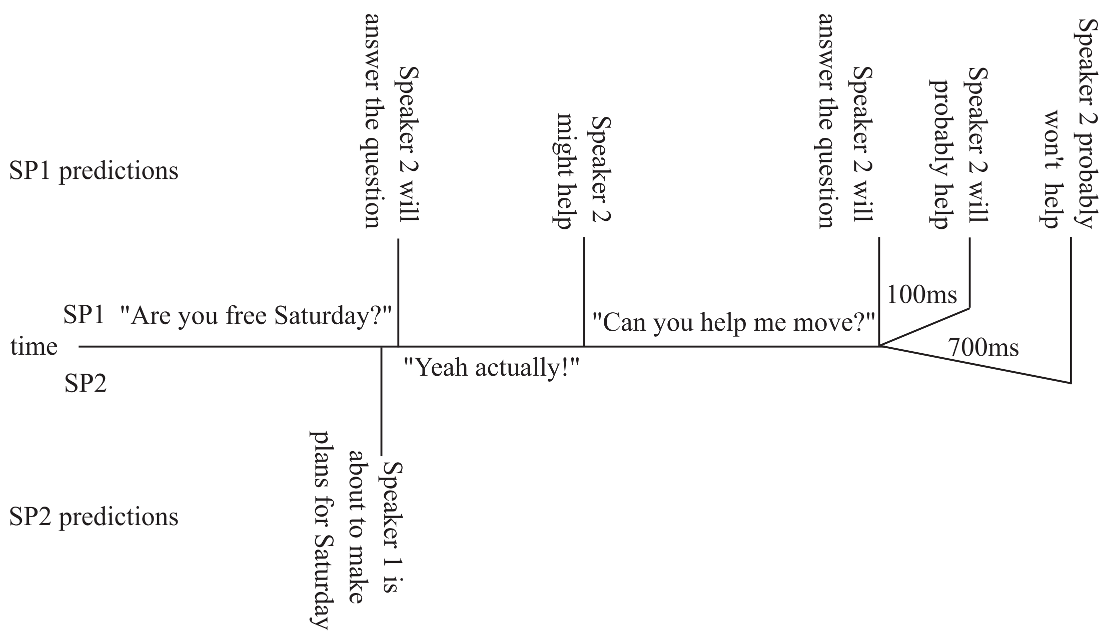
Listeners are busy multitaskers. They keep track of ambiguous information in the hopes that the speaker will clear up the confusion (Gwilliams et al., 2018). They react to the speakers' turn as it unfolds (Yngve, 1970). They also predict the words, syntax, and function of the unfolding turn (Kuperberg & Jaeger, 2016). Listeners live in the past, present, and future of the turn.

Predictions about the current turn can begin at least two turns before. Interlocutors can project their intentions by producing pre-sequences: sequences that clear a path for the next (Schegloff, 2007). In Figure 11, Speaker 1 starts a pre-sequence when they ask, "Are you free Saturday?" At this point, Speaker 2 can likely predict that Speaker 1 will ask them to do something on Saturday in the next sequence. In addition, many turns create expectations about the next turn (Sacks et al., 1974): questions should be followed by answers, invitations by acceptances or declinations. Experimental evidence shows that listener use these relationships to predict the social action (e.g., acceptance, answer, pre-offer) of the next turn (Gisladottir, et al., 2018; Gisladottir et al., 2015). Once Speaker 2 says they have no plans, Speaker 2 can predict the next turn will be an invitation and Speaker 1 can predict that the response to the invitation will be positive.

Exactly when interlocutors initiate their turns signal the valence of that turn. Dispreferred turns are turns that are potentially face-threatening (Clayman, 2002; Pillet-Shore, 2017). Speakers delay dispreferred turns (Pomerantz, 1984). Kendrick & Torreira (2015) quantified the delay before dispreferred responses to turns: most responses are preferred (not face-threatening), but most turns that start at least 700ms after the previous turn ends are dispreferred (Kendrick & Torreira, 2015). As a consequence, participants use the timing of a turn to predict whether it will be (dis)preferred (Bögels et al., 2020; Bögels et al., 2015). Figure 11 shows that if Speaker 2 begins speaking after 100ms, Speaker 1 can predict Speaker 2 will offer to help. If Speaker 2 begins speaking after 700ms, Speaker 1 can predict Speaker 2 will refuse to help.

Figure 11

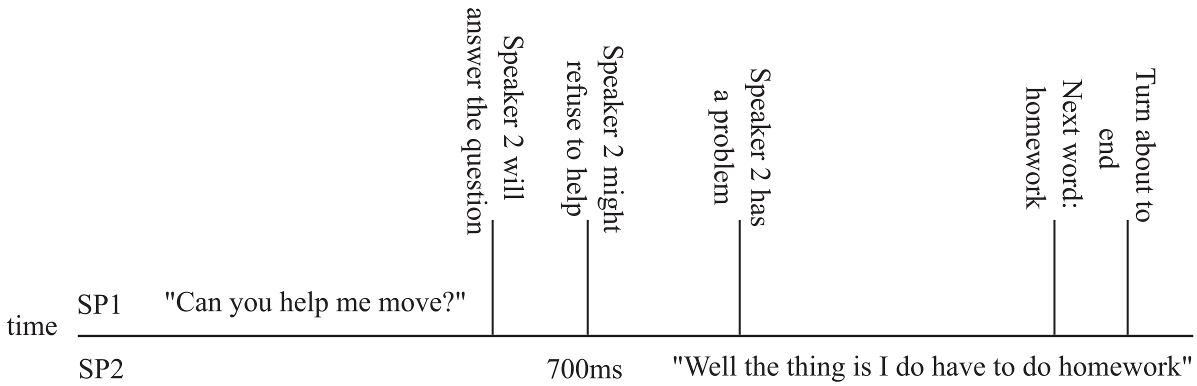
How interlocutors signal upcoming speech acts



Listeners also exploit the statistical relationships between words within the turn (e.g., Kutas & Hillyard, 1980). The first few words of an utterance strongly indicate the utterances' preference (Pomerantz, 1984). For example, a turn-initial "well" indicates that the turn will be dispreferred (Heritage, 2015). Words also project upcoming syntactic categories (e.g., "the" is typically followed by a noun), semantic features (e.g., "My favorite animal is a" is probably followed by the name of an animal), or exact words (e.g., "rake the" is typically followed by "leaves"). Figure 12 provides an example of some of these predictions.

Figure 12

Some finer grained signaling and prediction



Correct predictions increase progressivity

Why do interlocutors predict? One answer may be that correct predictions shorten turn transition times in conversation, therefore increasing progressivity. First, the predictive coding theory suggests that the brain predicts information at many different levels, and only feeds forward unpredicted information from lower levels (Rao & Ballard, 1999; Shain et al., 2020). Therefore, accurately predicted language requires less time and cognitive resources to comprehend. Second, correct predictions allow interlocutors to plan their next turn in advance – they free the cognitive resources needed for language planning and they increase the chance that any planned response will be appropriate. Planning a response is time-intensive; it takes participants approximately 600ms to initiate the name of a single object in experimental settings (Indefrey & Levelt, 2004). Since turns are so much more complex than the name of a single object, planning a response in conversation likely takes even longer. Luckily, listeners can start planning as soon as they can predict the remaining information in the turn (Bögels, Magyari, et al., 2015; Magyari et al., 2017). Finally, listeners predict the end time of the current turn, so they know exactly when to start speaking (De Ruiter et al., 2006). These three effects of prediction likely contribute to shorter gaps in conversation.

Incorrect predictions may decrease intersubjectivity

Predictions are not always correct. There are experimental findings showing that people take longer to comprehend unpredicted language in experimental settings (e.g., Brothers & Kuperberg, 2021); however, there are no studies investigating whether incorrect predictions decrease intersubjectivity in conversation.

The listener likely maintains any predictions unless those predictions are *challenged* – they notice information that contradicts those predictions. I hypothesize that depending on whether an incorrect prediction is challenged, and how the listener processes any challenge, incorrect predictions may interfere with intersubjectivity in different ways.

There are a few reasons why an incorrect prediction may not be challenged. First, language is ambiguous, so the speaker may never produce information that explicitly contradicts the incorrect prediction. Second, listeners can fail to notice seemingly obvious, contradictory information. One of the earliest examples of this is the *Moses Illusion* (Erickson & Mattson, 1981). Participants were asked questions that presupposed incorrect information, like "How many animals did Moses take on the arc?" In the biblical story, it was Noah – not Moses – who put animals on an arc. Nevertheless, many participants incorrectly responded with "two." In experiments using more interactive paradigms, participants did not notice an incoherent utterance or when their text-message partner was randomly switched (Roberts et al., 2016). These examples illustrate that listeners may ignore the information that, if processed, would have contradicted a prediction. If an incorrect prediction is not challenged or the listener does not process the challenge, the result is the same: the incorrect prediction is maintained, resulting in an unnoticed lapse in intersubjectivity. I predict that miscommunications will occur when the listener makes an incorrect prediction that is not challenged – either because the speaker is ambiguous or because the listener ignores or replaces surprising information with a more likely alternative.

The listener might also notice – perhaps but not necessarily consciously– a challenge to an incorrect prediction. In this case, the listener experiences surprisal, the amount of cognitive processing induced by unpredicted or violative information (See Chapter II; Hale, 2001; Levy, 2008). I hypothesize that part of this process is determining whether to trust the predicted information, the perception of the stimulus, or neither. Each option has risks, and the listener's reaction is likely affected by the strength of the prediction, the amount of surprisal, and the context. Maintaining an incorrect prediction would lead to an unnoticed lapse in intersubjectivity. However, the listener can mis-hear, and the speaker can mis-speak; overriding a correct prediction in favor of trusting a mis-hearing or -speaking will also lead to an unnoticed lapse in intersubjectivity. If the listener trusts neither their prediction nor their perception, they will perceive a lack of intersubjectivity and may choose to initiate repair – a face-threatening action (Lerner, 1996; Schegloff et al., 1977).

There is little research investigating these hypotheses. One paper found that some unexpected actions were followed by open OIRs, or OIRs that request a repeat or reformulation of the entire trouble source TCU (Drew, 1997). For example, after someone expresses gratitude, the expected action is an acknowledgement (e.g., "You're welcome"). Excerpt 16 is a transcript of a segment of a telephone conversation presented in Drew (1997). An expression of gratitude (line 5) was followed by turn that began a closing sequence (line 6), not an acknowledgement.

This unexpected action was followed by an open OIR on line 7. The second attempt at starting the closing sequence (line 8) was understood, even though lines 6 and 8 were almost identical. Drew (1997) argues that some open OIRs, which typically index larger problems than other types of OIRs, occur when a speech act does not match the listeners' expectations. In Excerpt 16, the closing sequence was also constructed in an unusual way, suggesting that surprising language may contribute as well. In this proposal, I will examine the evidence for a stronger claim: that surprising turns are more likely to be followed by OIRs.

Excerpt 16

Violative trouble source, adapted from Drew (1997)

```
1  GOR: are you going tonight,
2  NOR: mm.
3  GOR: would you mind giving me a lift[t.
4  NOR:                               [no that's alright
5  GOR: very kind of you.
6  NOR: caught me in the bath again.
7  GOR: pardon?
8  NOR: caught me in the bath.
9  GOR: oh i'm sorry well i should let you get back to it.
```

Experimental evidence shows that comprehenders take longer to process surprising information (Brothers & Kuperberg, 2021), suggesting that surprisal causes listener to re-adjust their cognitive model. However, but we still know little about the role of comprehension in (non)understanding. One possibility is that listeners have a surprisal threshold. If a turn is more surprising than that threshold, the listener may give up on attempting to understand the turn and may initiate repair. In contrast, the listener may spend more time extracting information from a trouble source turn that is not surprising. If so, then very surprising trouble source TCUs should be followed by earlier OIRs than less trouble source TCUs. Another possibility is that listener try as hard as possible to process a turn. If a turn is surprising, the listener may more thoroughly check the discourse context, the shared common ground, or other information sources before initiating repair. One study did find that surprising information elicited a late ERP response only when that surprising information was presented within a context, suggesting that listeners may indeed persevere in trying to understand surprising information (Brothers et al., 2020). If so, surprising trouble source TCUs would be followed by longer FTOs than probable trouble source TCUs. Finally, some turns may be trouble source TCUs because the listener fails to extract information from them at all. In these cases, surprisal – which is synonymous with the concept of information (see Chapter II) – should not affect the FTOs after trouble source TCUs at all. In addition to analyzing whether trouble source TCUs are more surprising than other turns, I will investigate how surprisal influences FTOs after trouble source TCUs.

Finally, I will analyze miscommunication sequences to determine whether prediction causes unnoticed lapses in intersubjectivity. As described in Chapter II, there were only 43 miscommunication sequences in the ICC subset. Since miscommunications are diverse, uncontrolled, and complex, this sample size would be inadequate for any inferential statistics. Further, a miscommunication can occur without either interlocutor explicitly flagging the trouble source TCU, so some miscommunications do not have clear trouble source TCUs. Finally, miscommunications are poorly understood. Therefore, I used primarily qualitative methods to examine miscommunication sequences, with the goal of exploring any possible connection between prediction and miscommunication. I hypothesized that listeners may develop and rely on incorrect predictions, even when speakers contradicted those predictions, leading to miscommunication.

Analysis Plan

For analyses involving surprisal, I will exclude the five conversations used to fine-tune GPT-2 (see Chapter II). These conversations contained 17.22 hours of conversation, 316 TCUs containing trouble sources (turns that were targeted by at least one OIR; TSTs), and 27,335 TCUs without trouble sources (TCU-NT).

I will analyze noticed lapses in intersubjectivity with two sets of eight GAMMs. Each set of GAMMs will analyze the surprisal of the first and last words, the maximum and minimum surprisal, the mean and median surprisal, information density and the range of surprisal values. First, I will use surprisal metrics to predict whether a turn was a TST. I hypothesize that TSTs will have higher surprisal values than other turns. Second, I will determine whether surprisal metrics predict the following Floor Transfer Offset (FTO; De Ruiter et al., 2006) for TSTs and non-TSTs. I hypothesize that more surprising non-TSTs will be followed by longer FTOs and will explore the relationship between surprisal and FTOs for TSTs.

I will analyze whether incorrect predictions lead to miscommunication. Specifically, I will explore whether listeners ignore/replace surprising information in favor of more predictable information, causing miscommunication.

Prediction in Noticed Lapses in Intersubjectivity

Are trouble source TCUs more surprising? I hypothesized that more surprising turns would be more likely to be followed by an OIR. Table 2 presents the median and median absolute deviation (MAD) for each surprisal metric, split by whether the turn was a TST.

Table 2

Descriptive statistics: Surprisal for trouble source and other turns

	Non-TSTs (*n* = 27,335)		TST (*n* = 316)	
	Median	MAD	Median	MAD
First	1.56	0.72	1.77	0.69
Last	1.27	0.87	1.62	1.14
Max	2.97	1.12	4.27	1.08
Min	0.43	0.4	0.16	0.16
Mean	1.56	0.43	1.74	0.34
Median	1.38	0.45	1.52	0.37
Range	2.21	1.79	3.8	1.01
Density	5.85	1.88	6.65	1.69

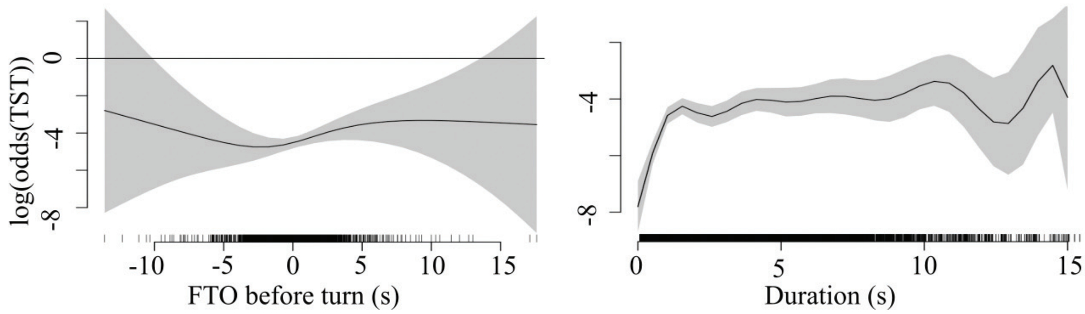
Note: Turns come from the 37 conversations not used to fine-tune GPT.

I included several control variables in the GAMMs to account for other factors that could be related to whether a turn was a TST. Previous research suggests that a longer FTO before a TST is associated with the OIR signaling a greater problem (Mertens & De Ruiter, 2021). Therefore, I included the FTO before the target turn as a predictor. In addition, I labeled the start and end times of TST turn construction units (TCUs; Sacks et al., 1974). TCUs are complete units of speech. Multiple TCUs can be part of the same turn; even if the TST was preceded or followed by another TCU by the same speaker, I identified the start and end time of the TST. I did not do this for every single TCU; consecutive TCUs by the same speaker may be counted as one long turn. Therefore, there was a chance that TSs would be shorter in duration than other turns, which in turn could affect the surprisal metrics. To account for this variance, I included the duration of the turn as a control variable. The null model contained a random intercept for the conversation as well as smoothed duration and previous FTO predictors.

Figure 13 displays the relationship between the likelihood a turn was a TST with the three predictors in the null model: the previous FTO and the duration of the turn. The previous FTO was a borderline, but statistically insignificant predictor of the log odds that a turn was a TST ($edf = 2.67$, $F = 8.26$, $p = 0.06$). When the previous FTO was between -1.5 seconds and 5s seconds, the previous FTO was linearly and positively related to the log odds that a turn was a TST (Figure 13, left). This was almost all FTOs – only 690 (2.5%) turn transitions had more than 1.5 seconds of overlap, and only 49 (0.18%) turn transitions had gaps longer than 5 seconds. Duration was a statistically significant predictor of the log odds that a turn was a TST ($edf = 8.988$, $F = 73.76$, $p < 0.01$). When the turn had a duration shorter than 1 second (47.6% of turns), duration was positively associated with the log odds that a turn would become a TST (Figure 13, right). Very short turns were extremely unlikely to be TSTs, because they were backchannels (e.g., "mhm" or "oh"; Bavelas et al., 2006; Schegloff, 1982; Yngve, 1970); the shortest trouble source TCU was 982ms long. Finally, the likelihood a turn was a TST differed by conversation ($edf = 17.3$, $F = 36.15$, $p < 0.01$). In one conversation, there were no TSTs. In the conversation with the most frequent TSTs, 2.28% of the turns were TSTs. These findings show that the previous FTO, duration of the turn, and the conversation are all useful predictors that should be included in the target models.

Figure 13

Effects in the null model predicting trouble source TCUs



Next, I created eight GAMMs to determine whether surprising turns were more likely to be TSTs. When creating the GAMMs, I included an interaction effect between the target metric (e.g., median surprisal) and the duration of the turn. I included this interaction effect because in the prescreening, I discovered that probable, short turns (e.g., “yeah”) were never TSTs, while surprising, short turns (e.g., a unique name or noise) were sometimes TSTs. Table 3 presents the main effects from of each target variable.

Table 3

Surprisal metrics predicting TSTs in GAMMs

Predictor	Edf	F	P	Exp. Deviance	R^2 (adj)
Null model	.	.	.	8.08%	0.008
First	4.24	13.986	0.02*	7.88%	0.008
Last	1.11	16.198	0.02*	8.88%*	0.010*
Max	2.65	33.646	<0.01**	8.78%*	0.009*
Min	1.79	2.534	0.28	7.64%	0.007
Mean	2.66	17.225	<0.01**	8.44%*	0.009*
Median	1	19.12	<0.01**	9.01%*	0.010*

Range	3.24	60.003	<0.01**	8.25%*	0.009*
Density	3.08	34.711	0.02*	8.73%*	0.008*

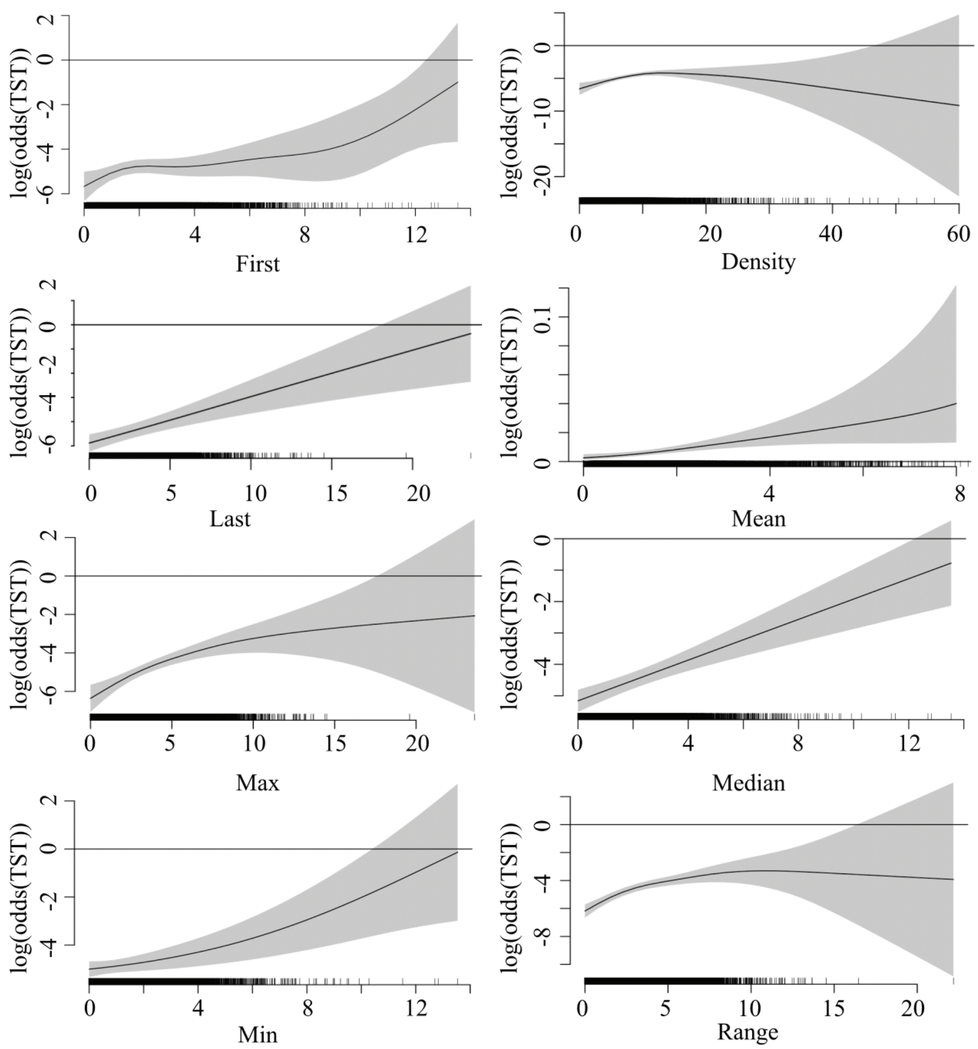
Note: For p-values, * means the p-value was less than 0.05, and ** means the p-value was less than 0.01. For R^2, * signifies that the model explained more deviance than the null model.

Figure 14 displays the GAMM plots relating surprisal metrics to the log odds that a turn was a TST. Most metrics had a straightforward relationship with the log odds that a turn was a TST. The surprisal of the last word and the median surprisal had a linear, positive relationship with the log odds that a turn was a TST. Mean surprisal was exponentially, positively related to the log odds a turn was a TST, although with a relatively gradual slope. The maximum surprisal, information density and surprisal range had positive and logarithmic relationships with the log odds that a turn was a TST. The minimum surprisal did not statistically significantly predict the log odds that a turn was a TST.

The surprisal of the first word had a more complex relationship with the log odds that a turn was a TST. When the first word had a surprisal less than 2, it was approximately linearly and positively related to the log odds that a turn was a TST. When the first word had a surprisal greater than approximately 9, it was approximately linearly and positively related to the log odds that a turn was a TST. For values between 2 and 9, there was no relationship between the surprisal of the first word and the log odds a turn was a TST. While each GAMM produced slightly different results, they all pointed in the same direction: greater surprisal values were associated with a greater chance of a turn being a TST. This finding supported my hypothesis that surprising turns would be more likely to be followed by OIRs.

Figure 14

GAMM results: Does surprisal predict the log odds the turn will be a TS?



Note: The black vertical lines at the bottom of each plot represent the density of data. When there are few vertical lines, there were few data points with that value of the predictor variable.

These findings – that more surprising turns were more likely to become TSTs – are illustrated in a subset of miscommunication sequences. In Excerpt 17 (also seen in Chapter II, Excerpt 14), Speaker 1 jokes that they participate in studies like this one on line 1. When

Speaker 2 does not respond as expected, Speaker 1 upgrades their joke on line 5. Jokes are surprising (Veale, 2004), so when Speaker 1 takes Speaker 2's joke as a serious statement, they are likely surprised – and then they initiate repair on line 6. In the miscommunication collection (see Chapter II), jokes and exaggerations that were treated as serious were typically followed by an OIR or an expression of surprise.

Excerpt 17

Listeners take jokes too seriously

```
1  SP1: i've never done some shit like this.
2  SP2: yeah u- (0.3) oh i have l⌈ike,⌉
3  SP1:                          ⌊yea⌋h,
4       (0.3)
5  SP2: every weekend yeah,
6  SP1: what
7       (4.1)
8  SP2: i was just kidding. it was ⌈a joke,⌉
9  SP1:                            ⌊   oh w⌋hat
10      (0.4)
11 SP2: come on now.≈
```

Are surprising trouble source TCUs followed by earlier OIRs? I hypothesized that most turns would show a positive relationship between surprisal and the following FTO, but that trouble source TCUs would have a negative relationship between surprisal and the following FTO. As described in the Other-Initiated Repair Sequences Section (in Chapter II), some TSTs are not immediately followed by an OIR. Sometimes the person who produced the trouble source TCU continued, or the person who produced the OIR responded to part of the turn (e.g., "oh really?") before initiating repair on another part of the turn. The FTOs of interest are those between TSTs and OIRs; in the analyses below, I excluded the FTOs after TSTs that were followed by a non-OIR. For these reasons, the N for each category in Table 4, which describes FTOs after non-TSTs, all TSTs. and TSTs immediately followed by OIRs, are much smaller than those in Table 2.

Table 4

Descriptive statistics: Floor Transfer Offsets

	Non-TST	TST	TST-OIR
N	19,309	268	230
Mean	62.81ms	221.01ms	276.67ms
Std. Deviation	929.08ms	918.66ms	925.13ms

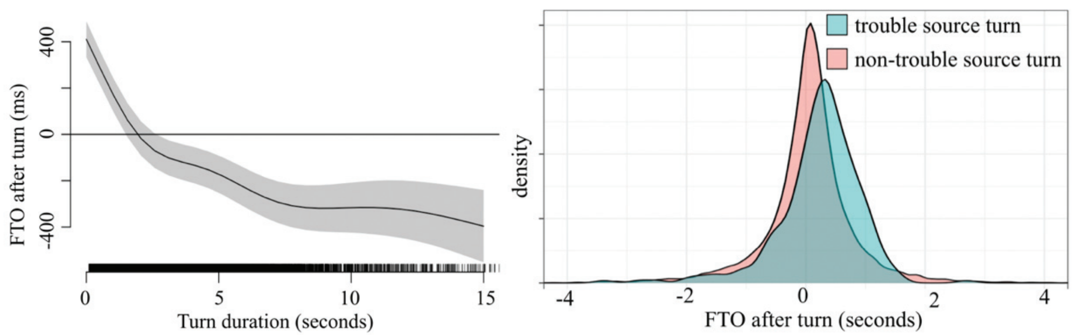
Note: Only turns followed by FTOs are represented.

Previous research found that a turns' duration had a non-linear relationship with the following FTO: turns shorter than 700ms were followed by longer FTOs, but for longer turns, duration and the following FTO were positively related (Roberts et al., 2015). Therefore, I decided to include duration as a predictor in the null model. Further, OIRs follow longer FTOs than other turns (Kendrick, 2015b). Therefore, I also included a binary variable for whether the turn was a TST in the null model.

The null model found that turn duration was a statistically significant predictor of the following FTO ($edf = 7.27$, $F = 114.69$, $p < 0.01$). When turn duration was between 0 seconds and approximately 8 seconds, longer turns were associated with shorter FTOs. However, once turns were at least 8 seconds long, there was no relationship between turn duration and the following FTO. In addition, the null model found that TSTs were followed by FTOs 218.95ms longer (t = 3.99, p < 0.01). These two effects are shown in Figure 15. In addition, the null model found a statistically significant effect of the random intercept for conversation (*edf* = 32.82, F = 31.99, p < 0.01).

Figure 15

Effects in the null model predicting FTO



Note: the left is the smoothed relationship between turn duration on the following FTO; on the right is a density plot of FTOs after trouble source TCUs and other turns.

In most cases, when a short turn was followed by a FTO longer than 300ms long, the sequence was like Excerpt 18. Typically, one interlocutor passes up an opportunity to take the floor from their social partner by producing a continuer ("Yeah" on line 4 in Excerpt 18; Schegloff, 1982). After a gap, the person talking before the continuer continues speaking.

Excerpt 18

Short turn followed by long FTO

```
1 SP1: i'm a big propone:nt of people just like (.) majoring in
2      whatever they like and then: just doing: something else random
3      w[ith it,]
4 SP2:  [    yea]h,
5      (0.8)
6 SP1: um (0.2) that's kind of (0.4) ∆what i'm doing with anthropology
7      i think,∆
```

After evaluating the null model, I created GAMMs that included the target surprisal metric. The effect for TSTs and non-TSTs was calculated separately (Table 5). All surprisal metrics were statistically significant predictors of FTOs after non-TSTs. No surprisal metrics were statistically significant predictors of FTO after TSTs.

Table 5

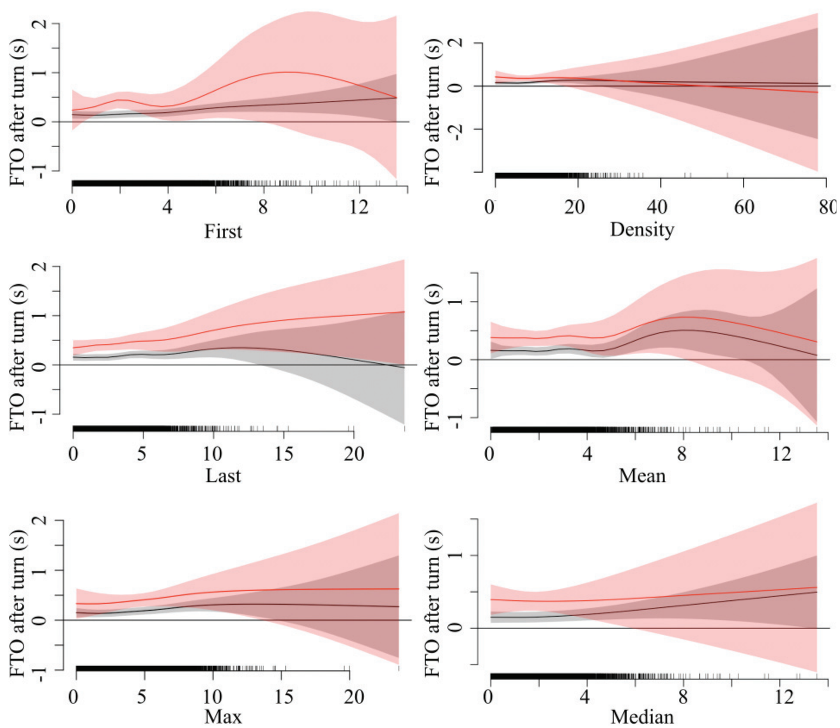
The effects of surprisal on FTOs after (non) trouble source TCUs

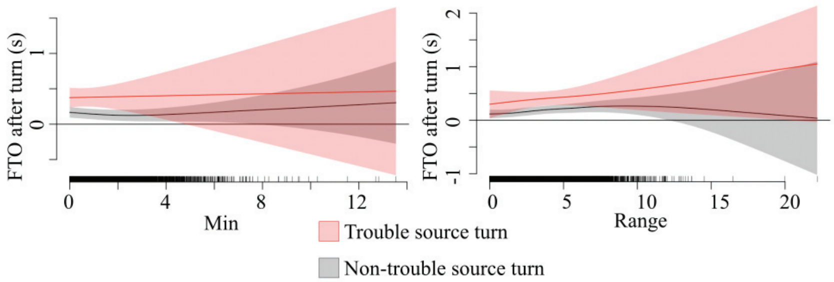
	Non-TSTs			TSTs			Model Performance	
	Edf	F	p	Edf	F	p	Exp. Deviance	R^2 (adj)
First word	1.99	6.73	<0.01**	4.05	0.65	0.66	10.60%	0.103
Last word	3.48	2.32	0.04*	1	1.5	0.22	10.20%	0.100
Min	2.12	3.14	0.04*	1	0.49	0.48	10.40%	0.102
Max	3.44	4.17	<0.01**	1	1.44	0.23	10.50%	0.103
Median	1	6.94	0.01**	1	0.08	0.78	10.50%	0.103
Mean	1	5.15	0.02*	1	0.27	0.61	10.60%	0.104
Range	3.6	2.51	0.03*	1	1.11	0.29	10.60%	0.103
Density	3.08	5.41	<0.01**	1	0.08	0.78	10.60%	0.104

Figure 16 displays the relationships between surprisal and FTO following TSTs (red) and non-TSTs (grey). For non-TSTs, all surprisal metrics was positively, approximately linearly related, with a gradual slope. In general, the TSTs had a similar relationship between surprisal and the following FTOs, but with more uncertainty. In summary, these analyses supported the hypothesis that for non-TSTs, surprisal would be positively correlated with the following FTOs. However, they did not support the hypothesis that for TSTs, surprisal would be negatively correlated with the following FTOs.

Figure 16

Surprisal and FTOs after trouble source TCUs and other turns





Note: the y-axis represents the FTO after the turn, in seconds.

As expected, non-trouble source turns were followed by longer FTOs when they were more surprising. Surprising turns likely contain more information and are harder to comprehend. However, there was no statistically significant effect of surprisal on the following FTO for trouble source TUCs. While more research is needed, this suggests that listeners of surprising trouble source TCUs do not ‘give up’ processing turns.

Prediction in Unnoticed Lapses of Intersubjectivity

I hypothesized two ways that incorrect predictions could result in miscommunication (temporarily unnoticed lapses of intersubjectivity). First, I hypothesized that speakers may never challenge incorrect predictions, instead leaving the interpretation of the turn ambiguous, resulting in miscommunications. Second, speakers may produce information that challenges the incorrect predictions, but listeners may either fail to notice the challenge, or (perhaps unconsciously) choose to believe their prediction over their perception of the turn. Neither hypothesis was supported by my analysis of unnoticed lapses in intersubjectivity.

Ambiguity did contribute to 18 (41.86%) miscommunication sequences. In Excerpt 19 (as shown in Excerpt 10), Speaker 2 uses an ambiguous pronoun, “this,” on line 1. However, for the listener to rely on strong but incorrect predictions to resolve ambiguity, the speaker must set up the context to project one piece of information but produce another piece of information. This was extremely rare, including for miscommunications contributed to by ambiguity.

Excerpt 19

Miscommunication contributed to by ambiguity

```
SP2: ah d- (0.6) ah (0.3) i can't do this,
     (0.3)
SP1: yes you (0.4) absolute⌈ly can,⌉
SP2:                       ⌊  i   ⌋mean like be awake right now
     like i'd rather be asleep
     (1.1)
SP1: Δi thought you just meantΔ like talk for like thirty
     minutes and i was like yes you can,
```

There was only one miscommunication that occurred within such a context. In Excerpt 20 occurred when *Downsizing* (Payne, 2017) was coming out in theaters. The title is a pun: the movie is about a technology that allows people to physically shrink and use fewer resources, so their savings are worth more. This pun starts the confusion. Speaker 2 says "it was called downsizing," on line 9, priming Speaker 2 for the standard meaning of the word – firing people at a company to reduce that company's size. When Speaker 2 says "get like shrinked down" (line 9), they produce information that contradicts Speaker 1's prediction. Speaker 1 responds as if Speaker 2 was attempting, and struggling, to explain the concept of downsizing. Speaker 2 does produce long pauses on lines 5 and 8, suggesting difficulty, and says "I don't know" on line 10. Speaker 2 requests confirmation of understanding with their raised turn-final intonation in "get like shrinked down," and by pausing. Speaker 1 claims they understand, repeating "downsizing," which is then confirmed by Speaker 2 on line 15.

Excerpt 20

Linguistic prediction implicated in miscommunication

```
1  SP2: i i wanna see: (.) this is i-
2       (1.4)
3  SP2: it was called d:ownsizing? (0.9) it's like it's like a new
4       mo:vie that like just came out? it it was u::h
5       (3.8)
6  SP2: mm:: i forget who was in it. (0.6) but basically the
7       concept was like
8       (1.7)
9  SP2: all these people who were like s:aving up like their savings
10      for their whole lives got to like (0.3) i: don't know l:ike
11      (0.4) like were middle aged and like (0.6) they got an offer
12      from this company: t:o get like shrinked down?
13      (0.6)
14 SP1: yeah, (.) downsizing.
15 SP2: downsizing? yea:h exactly and like they (0.2) they got their
16      offer to shrink d:own and then they're like (0.3) ∇they
17      tra:nslated their money in like big person terms to little
18      person terms∇ (0.8) so like their savings went from like
19      forty thousand to like five million?
20      (1.1)
21 SP2: because ⌈it's like⌉
22 SP1:         ⌊    h⌋o::w.
23 SP2: b- because when you're smaller like to build a hous↑e
24 SP2: i⌈t's just lik⌉e,
25 SP1:  ⌊    yeah,⌋
26      (0.7)
27 SP2: inexpensive. =like you you just like (0.3) don't ha- like ∆yo
28      know i mean,∆ (0.4) like a huge house might be like this big
29      (1.1)
30 SP1: yeah?
31      (0.6)
32 SP2: so it's like that might cost like a hundred dollars to make.
33      (1.4)
34 SP2: you know what i'm saying?
35      (0.7)
36 SP1: no. not at all,
37      (0.3)
38 SP2: okay. so it's kind of like,
39      (1.3)
40 SP2: oh ah ah wait. this is a g- this is a good way to th⌈ink about
41      it.⌉
42 SP1: ⌊is this l⌋ike an alternate wor↑ld?
43 SP2: no. it's like, (0.9) no it's like this company has figured ou
44      how to like shrink people down?≈
45 SP1: ≈oh sh- like physically shrink.≈
46 SP2: ≈yeah.≈
47 SP1: ≈oh. so it's like literal d:ownsiz⌈ing.⌉
```

```
48 SP2:                                ⌊ oh⌋ ye⌈ah⌉
49 SP1:                                          ⌊i⌋ thought you
50         meant like getting fi:red.≈
```

Interestingly, this confirmed prediction overrides multiple challenges. Repeatedly, Speaker 2 uses words that emphasize the physical meaning of downsized (lines 16, 17, 18, 23, 28). The interlocutors know there's a problem but misdiagnose the source of the problem. Specifically, they assume intersubjectivity for the word "downsizing," and act as if the problem is in the mechanics of downsizing. It isn't until lines 41-43 that the true problem is uncovered. While strong, incorrect predictions may rarely contribute to miscommunication, they may be difficult to correct.

As described above, I predicted that listeners would fail to update their incorrect predictions when challenged by surprising information. However, I found two miscommunications where the listener failed to notice predictable (or at least not surprising) information. In Excerpt 21, Speaker 2 asks Speaker 1 about their summers, and Speaker 1 talks about how they sublet from a friend for the summer after their sophomore year (lines 6-7). However, Speaker 2 asks about where Speaker 1 stayed over the summer on line 13, as if lines 5-7 never happened. Speaker 1's response makes it clear that Speaker 2's question was unexpected or not normative. Speaker 1 hesitates ("um" and a micropause), and instead producing a direct response, which would repeat lines 6-7 (e.g., "No, I subletted") Speaker 1 elaborates to provide information not contained in lines 6-7.

Excerpt 21

Unnoticed, predictable information

```
SP2: what have you been doing in the summers. just like d- do
     you do like a lot of like (0.2) m- pre med kind of stuff
SP2: or ⌈do you: just like⌉
SP1:    ⌊ so: after my f:ɹreshman year: (0.4) i volunteered ∆at a
     hospital in my hometown so I was just at home,∆ (0.6) and
     the:n after my sophomore year I was (.) subletting from a
     friend (0.5) like on ca:mpus,
SP1: a:nd i just did research here. (0.2) in two hundred boston
     avenue, (0.6) just like for the whole summe⌈r,⌉
SP2:                                           ⌊o⌋h. you were
     here during the ⌈summer,⌉
SP1:                 ⌊ yeah  ⌋
SP2: where do you where do you stay. do you have to room here?
SP2: like ⌈uh,⌉
SP1:      ⌊ u:ɹm (.) ∆I was li:ving o:n Bo:ston Avenue,∆ (0.3) like
     one of my friends had an off campus house but he wasn't staying
     here for the summer s:o I just like subletted his room from
     hi:m,
SP2: oh:::.
```

Speaker 2 may not have noticed lines 6-7, forgot their original request, or misinterpreted ‘subletting’ to refer to the school year and not the summer. Regardless, they failed to comprehend predictable information that they had requested.

Third, I hypothesized that listeners would override disconfirming or surprising information with a more likely alternative. Once again, I found no examples of this occurring in the miscommunication corpus, and instead found the opposite. First, listeners replaced probable information with surprising information. Before Excerpt 22 began, Speaker 2 had been listing different types of personal security information and had already mentioned “mother’s maiden name.” Therefore, “mother’s maiden name” in line 1 should have been predictable, or at least primed. However, there is some clear evidence that Speaker 2 replaced “mother’s maiden name” with “father’s maiden name,” a concept that does not even exist. First, Speaker 2 begins to laugh once Speaker 1 says “mother’s ma”. Second, Speaker 2 stresses “maiden” in line 3, marking that word as being specifically improbable. Finally, when Speaker 1 clarifies that they said "mother's," Speaker responds with a change-of-state token ("oh"; Heritage, 1986) and moves on by continuing listing personal security questions, instead of challenging the solution.

Excerpt 22

Speaker 2 perceives surprising instead of predictable information

```
1 SP1:  what are the like b:anking: questions like mother's
2       m⌈aiden na:me⌉ like first p⌈et,⌉
3 SP2:    ⌊((laughs))⌋        ⌊fa⌋ther's m:aiden name? ((laughs))≈
4 SP1: ≈↑no i said mother's.≈
5 SP2: ≈oh. (0.4) yeah favorite p⌈e:t,⌉
6 SP1:                           ⌊ ye⌋ah
```

In addition to replacing probable words with more surprising words, listeners also perceived the meaning of an ambiguous word that clearly did not fit into the linguistic context. In Excerpt 23, Speaker 1 discusses the weather in California, and specifically talks about how they like to "sit in front of a fire" (line 5). Speaker 2 responds to the appropriate meaning of "fire" in line 6. However, Speaker 1, Speaker 2 responds as if "fires" referred to forest fires. Either Speaker 2 incorrectly perceived lines 7-8 as referring to wildfires, or they attempted to shift the topic. Either way, Speaker 1 treats the sequence as a miscommunication by producing a solution. Speaker 2 accepts the solution with "oh yeah" and then produces a relevant response to lines 9-10.

Excerpt 23

Abrupt shift in the meaning of an ambiguous word

```
SP1: like it doesn't rain a lot in califo:rnia: and it's always
     like really nice ti- nice we:ather: so when it d:oes r:ain
     (0.4) you just li- you go insi:de and everything: and it's
     (.) you like (.) get a hot chocolate and you like bundle up
     and you sit it front of a ⌈fire⌉
SP2:                          ⌊i wi⌋:s⌈:h     ⌉
SP1:                                  ⌊there's⌋ no fires: at
     tufts: and it really pisses me of⌈f.⌉
SP2:                                  ⌊th⌋ere's too many fir:es
     in california,
     (0.5)
SP1: no like fire like firepla⌈ces⌉
SP2:                          ⌊oh ⌋yeah, (.) I haven't seen a
     single o:ne,
```

Finally, listeners interpreted turns to be more socially problematic than they were intended. In summary, there is ample evidence suggesting that Speaker 2 mis-interpreted a turn to be more face-threatening, even though one might expect socially problematic turns to be less likely than other turns. In Excerpt 24, Speaker 2 perceives line 1 as a version of "What're you doing this weekend," which could be perceived as a pre-invitation to an event. However, these interlocutors were only interacting because of the study – they did not know each other before this conversation. Consequently, a pre-invitation (or an invitation) would be socially awkward, as it could easily be perceived as asking Speaker 2 on a date in an environment where Speaker 2 is pressured to stay in the room. There is evidence that, in fact, Speaker 2 interprets line 1 in this way. Speaker 2 waits 1.2 seconds, a very long time, before responding. When she does respond, she produces a pause marker and pauses. In addition, the way Speaker 2 constructs line 3 is odd; "plans" is produced with a sharp rise and fall in pitch, and "yet" ends with an unusually sharp /t/. Interestingly, Speaker 1 does not explicitly signal that there has been a lapse in intersubjectivity; instead, Speaker 1 reformulates their question on line 5. With Speaker 1's reformulation, Speaker

2 realizes the problem for themselves. Even more evidence that Speaker 2 found line 1 socially problematic can be found in how Speaker 2 responds to the miscommunication. Twice, on lines 6 and 8, Speaker 2 produces a change-of-state token ("oh") and repeats the less face-threatening (and correct) interpretation of line 1. It may be that Speaker 2 responds this way to implicitly account for their uncomfortable turn on line 6 without admitting their face-threatening interpretation of line 1. In summary, there is ample evidence suggesting that Speaker 2 mis-interpreted a turn to be more face-threatening, and therefore perhaps more surprising, than it was.

Excerpt 24

A mishearing that turns a turn into a less appropriate turn

```
1   SP1: so: (0.4) what'd you do this weekend.
2        (1.2)
3   SP2: uh(h) (0.4) i have no ↑plans. (0.4) yet?
4        (0.4)
5   SP1: what about the past weekend.≈
6   SP2: ≈oh what did I do this ⌈weekend,⌉
7   SP1:                         ⌊ yeah   ⌋
8   SP2: oh like hallowe:eke:nd
9   SP1: yeah
10  SP2: yeah:: (.) um,
11       (1.8)
12  SP2: i:,
13       (3.5)
14  SP2: just like i ↑just hung out with my f:riends. i didn't really
15       (1.4) want to go to like the fra:t parties,
```

In summary, this qualitative analysis did not support my hypothesis that listeners would ignore or override surprising input in favor of a predictable language. Instead, I found tentative evidence for the opposite: that some miscommunications occur because listeners perceive more surprising or non-normative versions of a turn. While this conclusion is still preliminary and more research is needed, it suggests a more complex relationship between prediction and intersubjectivity than discussed previously.

Discussion

I investigated whether prediction contributed to noticed and unnoticed lapses in intersubjectivity. I hypothesized that a) trouble source TCUs would be more surprising than other turns, b) more surprising non-trouble source TCUs would be followed by longer FTOs, c) surprising trouble source TCUs could be followed by longer, shorter, or similar FTOs as predictable trouble source TCUs, and d) listeners would ignore or override surprising information in favor or more predictable information.

First, this study supported hypothesis a, that interlocutors were more likely to initiate repair on more surprising turns. While other work has explored surprisal in reading or listening to monologues, this study is the first to explore the impact of surprisal in naturalistic dialog. It complements the finding in Drew (1997), that one interlocutor may produce an open OIR when their social partner produces a different next action than expected.

However, it is still unclear whether surprisal causes communication problems. If trouble source TCUs are more surprising because the speaker uses the wrong words or is incoherent, then word choice and incoherence – not surprisal – may be responsible for the listeners' lack of understanding. In other words, surprisal could be an epiphenomenon associated with poorly produced turns. In addition, while trouble source TCUs were more surprising, the effect size was small; surprisal cannot be the only signal that an interlocutor uses to determine whether there is a comprehension problem. Future work is needed to fully understand if, and how, surprisal interferes with comprehension in conversation.

Complicating matters, this finding directly contradicts an experiment where participants failed to notice an extremely incongruent utterance (Galantucci & Roberts, 2014). Participants collaborated with a confederate on a ranking task. After a lapse in the conversation, the confederate stated the famous Chomskian sentence, "Colorless green ideas sleep furiously," at a reasonable volume. Participants did not respond to the utterance in the moment and responded in a survey afterwards that they were not even aware of utterance. Exactly what types of surprising utterances do people notice, and when?

Second, this study supported hypothesis b, that higher surprisal values would be related to longer FTOs for most turns. This supports the assumption that correct predictions contribute to shorter gaps in conversation, therefore increasing progressivity. It also supports experimental findings in non-interactive contexts, namely reading and eye tracking studies, that improbable turns took longer to process (Brothers & Kuperberg, 2021). However, it also contradicts a corpus analysis that found that surprisal did not predict FTOs (Roberts et al., 2015). Roberts et al. (2015) created a random forest model with 33 features that could affect FTOs in conversation. Mean surprisal was the 9th least important feature, and information density was the 3rd least important feature. There are a few differences between Roberts et al. (2015) and this study which may explain the discrepancy in results. First, Roberts et al. (2015) used the Switchboard corpus (Godfrey et al., 1992), a collection of telephone conversations. Telephone conversations have shorter gaps, and longer overlaps, than in-person conversation (Bosch et al., 2004). This is likely because interlocutors talking on the phone cannot see their social partner, who may have left the phone or even hung up. In contrast, even long gaps in face-to-face conversation can be accounted for by concurrent behaviors, like looking for a credit card to pay the bill. It could be that over the phone, interlocutors prioritize progressivity even more than in person, and therefore respond early regardless of whether they are surprised. The Switchboard corpus is also composed of five-minute long conversations and a specific topic to discuss. The conversations in the ICC were much longer and completely freeform. Second, Roberts et al. (2015) included many predictors, while I included only duration and the FTO before the current turn as predictors. There may be a multicollinearity problem in Roberts et al. (2015), or there may be an important confounding variable that I did not include in my GAMMs. Finally, we used very different statistical techniques. Random forest models do not account for nested relationships, like how utterances are nested within conversations. In every GAMM, I found a significant effect of conversation; in Roberts et al. (2015), the inability to capture a hierarchical effect could increase noise, making it more difficult to perceive a subtle effect of surprisal. Future research is needed to tease apart these possibilities.

Third, I found no statistically significant relationship between surprisal and FTO within trouble source TCUs. Figure 16 suggests that trouble source TCUs and non-trouble source TCUs had a similar relationship between surprisal and the following FTO, but that the estimates for trouble source TCUs had more error. It may be that more OIR sequences are needed to find an effect, or that FTOs after trouble source TCUs are more variable regardless of surprisal. Alternatively, it is possible that interlocutors do not extract all the information in a trouble source TCU, so whether that information is surprising does not affect the processing of the trouble source TCUs.

Finally, I investigated unnoticed lapses in intersubjectivity. Only one miscommunication occurred within a highly constraining environment, or one that could encourage very strong predictions. In Excerpt 20, the problem originated from the recipient of a turn predicting the literal meaning of a pun. It is worth noting that the miscommunication lasted much longer than most, perhaps because strong predictions can override evidence produced later.

In addition, I found several examples where listeners ignored predictable information. One possible explanation is that listeners were under stimulated by predictable information. In Excerpt 21, the target information would be easy to connect to the context. It was also embedded within a relatively long turn. Therefore, Speaker 2 may have stopped paying attention, or forgot the target information. Experimental evidence does show that people remember surprising information better than predictable information (Foster & Keane, 2019). Further, Uniform Information Density Theory states that interlocutors work to maintain a moderate level of surprisal (information) to maximize communication efficiency (Frank & Jaeger, 2008) – another effect may be to help the listener attend to and remember information. Similarly, optimum stimulation theory suggests that people and animals attempt to achieve a happy medium of perceptual stimulation (Hebb, 1955). This pattern is repeated in flow theory (Csikszentmihalyi & LeFevre, 1989), the theory that people are most fulfilled when they are challenged, but in a way that matches their skills. People do not thrive without challenge, or when they are confronted with an impossible task. This balance may be similarly true for comprehension.

I also found that listeners occasionally replaced predictable language with more surprising or non-normative information. It could be that social norms, more general knowledge, or someone's previous experience interferes with linguistic prediction. For example, people should produce new and notable turns. Before Excerpt 22, the interlocutors had already been discussing their mother's maiden names; perhaps Speaker 2 had expected Speaker 1 to say something new instead of repeating the same concept. Excerpt 23 was recorded in September of 2018, right after an extremely deadly and destructive fire season for California. Perhaps the statement “it doesn’t rain a lot in California” triggered the extremely salient and recent fire season. Speaker 2 may have recently seen coverage of the forest fires on the news or had conversations about the fires. Finally, in Excerpt 24, Speaker 2 presented as female, while Speaker 1 presented as male. It could be that Speaker 2 has had previous uncomfortable experiences with men asking them on dates, coloring their comprehension of line 1. All in all, these examples show that world knowledge, cognitive states, or previous experiences may cause one listener to override linguistically predictable information.

One important caveat to this study is that language models like GPT-2 cannot always accurately model the listener. This is especially problematic when the speaker uses slang, proper nouns that are common on the Tufts University campus or sounds that do not map onto words with dictionary meanings. In these cases, GPT-2 overestimates the surprisal of language that interlocutors do, in fact, understand. This likely inflates GPT-2 estimates of surprisal, especially for non-trouble source TCUs. It could be that the effects found in this study would be even greater (or weaker) with a more accurate model of surprisal. However, no language model can account for the cognitive model of the specific listener. In fact, even other human participants, for example in cloze studies where participants write down what they believe the next word would be, cannot estimate the specific listener’s cognitive state. The only way to estimate listener surprisal is through electroencephalography (EEG), electrodes placed on the skull that record the activity of the brain (Gevins et al., 1995). However, EEG is extremely sensitive to any movement – even blinking – and currently cannot be used during conversation. Researchers will have to get creative to achieve better estimates of surprisal.

In addition, I did not analyze how probability and surprisal unfold over time. It may be that listeners are not so much affected by the mean or median surprisal, but instead by patterns of surprisal. For example, the first half of a turn may be predictable, but if the last three words are surprising the listener might perceive a lapse in intersubjectivity. Future work should explore methods of analyzing surprisal on a word-by-word basis.

Even given these limitations, this study provides the first empirical evidence that prediction can contribute to perceived and overlooked lapses in intersubjectivity.

Conclusion

Surprisal may cause listeners to perceive a lapse in intersubjectivity, perhaps resulting in an other-initiation of repair. However, the relationship between prediction and miscommunication is less straightforward. Predictable and long turns may be forgettable, and world knowledge may override linguistic prediction, suggesting that surprisal could increase intersubjectivity. While many questions remain unanswered, one finding is clear: while prediction can decrease gaps in conversation, it interferes with intersubjectivity.

In many ways, linguistic prediction has been well-studied. Psycholinguists study the neural responses to predictable or surprising language (e.g., Kutas & Hillyard, 1980) while conversation analysts discuss the "projectability" of language (e.g., Sacks et al., 1974). However, the study of linguistic prediction is insulated. We know that listeners predict, but we know little about how predictions influence behavior or understanding in conversation. We know that when listeners listen to audiobooks or experimental stimuli, the linguistic context is extremely important for prediction. However, we know little about how other sources of information, like previous experiences, ones' train of thought, or world events affect – or even override – our linguistic predictions in the real world. Future research should push the borders of work on prediction, looking into how linguistic prediction affects behavior and interfaces with other types of knowledge.

Another limitation to this study is the lack of a comparison group. If all listeners predict linguistic stimuli in the same way, no such comparison group exists. However, Sinha et al. (2014) recently proposed a Predictive Impairment in Autism (PIA). They proposed that multiple symptoms of autism spectrum disorder (ASD) could be linked to differences in prediction. While PIA is still controversial, a recent systematic review of the empirical evidence for PIA suggests that people with ASD may indeed struggle to learn subtle predictive relationships (Cannon et al., 2021). If there is truly a PIA, then people with ASD should predict linguistic stimuli differently than neurotypical individuals. If interlocutors use surprisal to identify communication problems, then people with ASD should perceive lapses in intersubjectivity at different times than neurotypical individuals. A small literature has investigated conversational repair in conversations with interlocutors with ASD. In general, this literature focuses on determining whether people with ASD can initiate repair and respond to repair initiations. In general, the literature shows that people with ASD can deploy and react to conversational repair appropriately (Volden, 2004). However, this literature is not comprehensive, does not examine the types of problems in conversations with neurodiverse participants, and does not specifically examine the role of prediction in communication problems for people with ASD.

Corpora of conversations between neurodiverse interlocutors are few and far between, in part for privacy reasons. If such a corpus can be developed, it could prove an interesting comparison to the results in this study. Do people with ASD also initiate repair on turns that GPT-2 finds to be more surprising?

This study takes the first step towards understanding how prediction influences (mis)understanding in a conversational context. Prediction, a cognitive process fundamental to turn-taking in conversation as we know it, also causes communication problems.

Chapter IV: Egocentricity

Abstract

Interlocutors are egocentric. In the context of psycholinguistics, egocentricity refers to producing and interpreting language from one's own perspective. Egocentricity could easily interfere with intersubjectivity in conversation: speakers may produce ambiguous language without realizing listeners may be confused; listeners may comprehend ambiguous language from their perspective without realizing the speaker may have meant something else. It is also possible that egocentricity increases progressivity, as perspective taking requires effort and time. However, no one has investigated the relationship between egocentricity, communication problems, and progressivity in conversation.

In experimental settings, people are more egocentric when communicating with friends than when communicating with strangers. In this study, I compare dyads composed of friends to those composed of strangers to determine if and how egocentricity influences progressivity and intersubjectivity. First, I determined that friends have more communication problems than strangers and are more likely to use egocentric repair strategies than strangers. Second, I quantified progressivity and found that depending on the quantification, friends may be more or less progressive than strangers. Third, I found that communication problems were correlated with one quantification of progressivity. Finally, I discovered several ways that egocentricity and/or social closeness contributed to miscommunications. While many questions remain unanswered, these findings are consistent with the theory that egocentricity increases progressivity and decreases intersubjectivity.

Interlocutors are egocentric

Interlocutors can struggle to take their social partners' perspective. In the foundational experiments on egocentricity, researchers recruited pairs of participants to coordinate: one person instructed the other to select and move objects. However, some objects were only visible from one perspective. For example, the speaker may have seen a large and medium-sized mouse, and the listener may see both of those mice, as well as a third, small, mouse. If the speaker said, "move the small mouse," the listener should know to move the medium-sized mouse, because it was the smallest mouse the speaker can see. When children performed this task, they selected the incorrect (smallest) mouse; when adults performed this task, they were less likely to select the wrong mouse but still considered and reached towards the incorrect mouse (Epley, Keysar, et al., 2004; Epley, Morewedge, et al., 2004; Keysar et al., 2000).

These findings show that even adult interlocutors are *egocentric*. In psycholinguistics, egocentricity refers specifically to how the comprehension system plans and interprets linguistic stimuli – it does not refer to selfishness or arrogance. Egocentric speakers fail to tailor their turns to their listener, and over-estimate how well they communicate. For example, speakers overestimated how well they communicated ambiguous information, even though overhearers who knew the speakers' intentions could accurately estimate the speakers' communication success (Keysar & Henly, 2002). Similarly, authors of an email over-estimated how well readers understood the message (Kruger et al., 2005). Worryingly, pediatric residents overestimated their communication effectiveness when handing off patient files at shift transitions (Chang et al., 2010). If interlocutors share the same perspective, egocentricity may not cause problems. However, when interlocutors have different perspectives, egocentricity leads to lapses in intersubjectivity.

Keysar (2008) first explicitly connected egocentricity and miscommunication. An egocentric speaker may produce an utterance that makes sense to them but is confusing to the listener. The listener interprets that utterance from their own perspective, without realizing the speaker could have meant something else, and the conversation continues. This is a plausible model, and Deppermann (2015) provides some examples of naturally occurring miscommunications caused by egocentricity during driving classes. However, driving classes are a unique social environment, and could be more or less likely to elicit either egocentricity or miscommunication. First, there is an extreme (and known) epistemic gradient – the teacher knows much more about the immediate environment and how to drive than the student. This means that their cognitive perspectives are very different, and the teacher must constantly take the student's perspective to communicate effectively. There is also no way for the student to "fake" understanding, as they must demonstrate understanding through their actions. Therefore, miscommunications are more likely to be noticed quickly. Further, the topic under discussion is typically visually present in the immediate environment. In day-to-day conversations, interlocutors refer to objects, events, or abstract concepts outside of the immediate environment, which are likely harder to keep track of than cars or streets. Finally, the student and instructor face the same direction and, therefore, share much of the same perspective on the environment. The goal of this study is to develop a more detailed picture of egocentricity and miscommunication, especially in day-to-day, casual conversations.

Egocentricity increases progressivity

Why are interlocutors egocentric? A speaker whose primary goal is to maintain intersubjectivity would always engage in *recipient design* (Sacks et al., 1974), the process of tailoring one's utterances to the specific listener(s). Similarly, a listener whose primary goal is to maintain intersubjectivity would always check *common ground* (Clark & Brennan, 1991), the shared information between the specific interlocutors.

One answer is that perspective taking takes time and cognitive resources. Interlocutors likely perform *egocentric anchoring and adjustment* (Epley, Keysar, et al., 2004) – first, they "anchor" the planned or interpreted utterance to their own perspective. Then, they monitor that plan or interpretation for potential errors. If they find an error, they adjust their plan or interpretation until it is plausible and begin the process again. It is the monitoring and adjustment process that takes additional time. The egocentric anchoring and adjustment theory is supported the findings that interlocutors are more egocentric when under time pressure or had fewer available cognitive resources (Epley, Keysar, et al., 2004; Horton & Keysar, 1996).

Not only does perspective-taking require time, but it also may not be useful. Some turns may not include information that differs depending on the perspective. Other times, the interlocutors may share the same perspective. Finally, attempting to take another person's perspective is not guaranteed to be successful; for example, a person who lives in one place may be incapable of imagining the perspective of someone who lives somewhere very different. For these reasons, an interlocutor who prioritizes progressivity is likely to be more egocentric.

Friends are more egocentric than strangers

Intuitively, it may seem obvious that people communicate better when they know their interlocutor better. People may be motivated to communicate well with the people they spend more time with; miscommunicating with a spouse may lead to greater consequences than miscommunicating with a stranger. People also share more *common ground,* or mutual beliefs and experiences (Clark & Brennan, 1991), with people they spend more time with. Another possibility is that we may choose to develop relationships with people who we find easier to communicate with. People fill their lives with others that are similar to them (McPherson et al., 2001), and people are better at predicting speech (Hadley et al., 2020) and identifying emotions (Elfenbein & Ambady, 2002) when interacting with people similar to themselves. All these factors could mean that it is easier to communicate successfully with a friend than a stranger.

Interestingly, experimental evidence suggests that both listeners (Savitsky et al., 2011) and speakers (Schober & Carstensen, 2010) are more egocentric when interacting with a friend. Schober & Carstensen (2010) also recruited long-term romantic partners – even then, speakers were egocentric. However, these studies required participants to stick to a script or describe unrealistic environments to their social partner. To date, no one has investigated whether interlocutors are more egocentric when engaging in casual conversations with friends.

Social closeness may affect the prioritization of progressivity

Friends may be especially egocentric because they de-prioritize intersubjectivity more than strangers. Since friends share more experiences and are more similar to each other than strangers, friends may have a smaller risk for miscommunication – even if they are egocentric. In experimental settings, the researchers carefully design the environment and task so the interlocutors are forced to take each other's perspectives. In non-experimental settings, the greater common ground and similarity between friends may be enough to guarantee communication success, regardless of whether they are egocentric.

Friends may prioritize signaling that they are part of the same social group (Enfield, 2020) over intersubjectivity. Interlocutors may avoid producing other-initiations of repair (OIRs; Kendrick, 2015a; Schegloff et al., 1977), requests for repetition, reformulation, or elaboration of a turn (see Chapter II). This is because OIRs suggest the interlocutor does not understand their social partner, which in turn can imply greater social distance. For example, if an interlocutor does not understand an inside joke and produces an OIR, it becomes clear that they are not part of their social partner's inner circle – or at least did not share the experience that resulted in the inside joke.

Just as it is face-threatening to produce an OIR, it is also face-threatening to over-explain a concept the listener could know (Johnson, 2020). Doing so implies the listener is not perceived as knowledgeable or is outside the social group. So, interlocutors may avoid both ensuring the listener understands and initiating repair when they do not understand, to signal belonging. Alternatively, friends may prioritize intersubjectivity more than strangers, even if friends are more egocentric than strangers. For example, friends may be more comfortable admitting a lack of understanding and producing OIR than strangers.

In this study, I will determine whether friends have more OIRs and miscommunications than strangers. If so, then friends may be more egocentric than strangers, and the benefits of social closeness may not be enough to prevent communication problems when intersubjectivity lapses. If friends have fewer (or similar rates of) communication problems than strangers, than they may not be more egocentric, and/or the increased similarity and common ground may help prevent miscommunication. Further, I will quantify progressivity and investigate whether a) friends are more progressive than strangers and b) conversations that are more progressive have more problems.

Categories of other-initiations of repair

Different types of OIRs pause progressivity for different durations, distribute the linguistic work differently, and signal different sizes of communication problems. *Open* OIRs (e.g. Excerpt 25, line 6) target the entire trouble source TCU (Drew, 1997). *Specific* OIRs (e.g., Excerpt 26) provide some information about what the interlocutor is targeting. Finally, *candidates* (e.g., Excerpt 1 and Excerpt 8) request (dis)confirmation of an interpretation of the trouble source TCU. While future work may wish to use more detailed OIR categories (e.g., Dingemanse et al., 2016), I use these three broad categories for this study.

Excerpt 25

Open OIR sequence

```
SP2: think i got f::- screwed.
     (0.4)
SP1: ∆what do ⌈  you   mean∆   ⌉
SP2:         ⌊By the short answ⌋er,
     (0.3)
SP1: oh the short answer? i got the exact same score as last
     time.
```

Excerpt 26

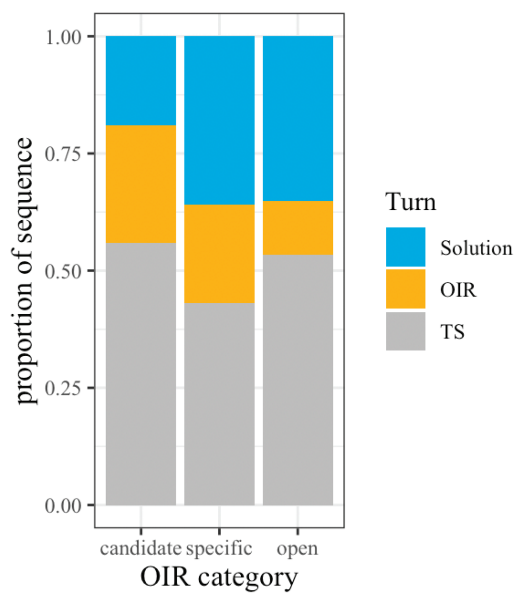
Specific OIR sequence

```
SP2: oh do you like to ski,
     (0.7)
SP1: yeah,
     (0.7)
SP1: i snowboard.
     (0.9)
SP2: and snowbo- wait which one. ski or ⌈snow.⌉
SP1:                                   ⌊ i sno⌋wboard (0.3)
     actually.
```

OIRs that target less of the trouble source TCU (or are "stronger", Schegloff et al., 1977) tend to be followed by shorter solutions (Dingemanse et al., 2015), and are therefore more progressive. In addition, stronger OIRs show a greater understanding of the trouble source TCU and perform more work. The strongest type of OIR in this study is the candidate OIR; the weakest type of OIR in this study is the open OIR. Figure 17 shows the relative duration of the trouble source TCUs, OIRs, and solutions in the ICC subset (consisting of 42 conversations; see Chapter II) after the OIRs were split by category (see the Categorizing other-initiations of repair section below). I used relative durations because when the turn construction unit (see Chapter II) containing the trouble source is longer, so will be the solution – regardless of the OIR category. Candidate OIR sequences have the shortest solutions, making up 18.9% of the sequences. Both specific and open OIRs are followed by solutions that are twice as long, making up approximately 36% of the sequence. However, OIRs are shortest in the open OIR sequences (11.4%), showing that open OIRs do the smallest proportion of the work in OIR sequences. For this reason, open OIRs have been described as "locally egocentric" (Dingemanse et al., 2015).

Figure 17

The relative durations of components of OIR sequences



Perhaps for these reasons, interlocutors prefer "stronger over weaker initiators" (Schegloff et al., 1977). This means people are more likely to use open OIRs in extenuating circumstances: when comprehension problems are very large (Clark & Schaefer, 1987), listeners are surprised (Drew, 1997), the environment is loud or distracting (Dingemanse et al., 2015), or the content of the trouble source TCU is face threatening (Svennevig, 2008). Interlocutors typically wait longer to initiate an OIR than another turn, but interlocutors wait especially long to initiate an open OIR (Kendrick, 2015b; Mertens & De Ruiter, 2021).

Interlocutors should produce more open OIRs when they are more egocentric. In this study, I will investigate whether OIRs in conversation between friends are more likely to be open than OIRs in conversations between strangers. If so, this would be additional evidence that friends are more egocentric than strangers.

Egocentric speakers and listeners

As described above, the literature on egocentricity suggests that both speakers and listeners are more egocentric when talking to people they know better. However, the responsibility for maintaining intersubjectivity may be placed differently when the interlocutors are more closely related. The more egocentric the speaker, the more potential communication problems they will produce, and the more OIRs and miscommunications in the conversation. The more egocentric the listener, the fewer potential communication problems they will notice, and the fewer OIRs and more miscommunication in the conversation. Table 6 displays how listener or speaker egocentricity would influence the rate of OIRs and miscommunications in conversations. If both speakers and listeners are both more egocentric when talking to friends, then both the rate of OIRs and miscommunications will increase in conversations between friends, and the rate of miscommunications will increase more than the rate of OIRs.

Table 6

Speaker and listener egocentricity affect the rate of communication problems

		Listener egocentricity		
		More	Similar	Less
	More	++Misc	+OIRs, +Misc	++OIRs
Speaker egocentricity	Similar	-OIRs, +Misc		+OIRs, -Misc
	Less	--OIRs	-OIRs, -Misc	--Misc

Summary of Hypotheses

I have a variety of hypotheses in this study. First, I hypothesize that friends will produce OIRs and miscommunications more often than strangers. Second, I hypothesize that the difference in the rate of miscommunications will be greater than the difference in the rate of OIRs. Third, I will define and analyze several measures of progressivity, described in the Methods Section. Fourth, I will explore whether conversations between friends and strangers differ in quantifications of progressivity. Fifth, I hypothesize that more progressive conversations will have more communication problems. Finally, I hypothesize that an examination of a corpus of miscommunications will show that egocentricity contributes to miscommunications.

Methods

Progressivity. Progressivity has never been quantified. I will attempt to quantify progressivity along two dimensions: the “speed” and “content” of turns. First, I will evaluate how quickly the interlocutors produce information. If everything else is the same, a conversation with shorter floor transition times (FTOs, see Chapter II; De Ruiter et al., 2006) and faster speech should be more progressive. FTO has previously been related to progressivity: Stivers &

Robinson (2006) argued that interlocutors prioritized progressivity when they jumped in to respond for another person when that person took too long to respond. Speech rate is a less traditional measure of progressivity. However, speech rate contributes to the "beat" of conversation, which in turn can affect the subjective length of a silence (Umair et al., 2022). In a conversation with slow speech, a longer gap may be perceived as shorter; in a conversation with fast speech, the same gap duration may be perceived as longer. In this study, I will operationalize speech rate as the number of words per second. This metric is flawed: a word may have just one syllable or many, resulting in different durations. However, calculating syllables per second is not always possible and much more computationally intensive than (and likely correlated with) words per second.

Second, I will attempt to quantify the progressivity of the content of turns. A turn with more words per second could be more progressive, but people also tend to produce less informative words more quickly (Mahowald et al., 2013). Therefore, a turn with many words per second could be less informative than a turn with fewer words per second. For that reason, I will also analyze information density (see Chapter II). I will also analyze the semantic similarity between turns. It could be that when some turns are extremely similar to the previous turn(s), they do not move the conversation forward. To calculate semantic similarity, I will convert each turn into a sentence embedding, a vector that represents the semantic meaning in the turn. I will produce those sentence embeddings with a Robustly Optimized BERT Pretraining Approach (RoBERTa) model (Liu et al., 2019). RoBERTa is a type of transformer model, meaning it learns which context to weigh more and less (like GPT; see Chapter II). It learns the relationship between dictionary words and the words that tend to come before and after the target word. Using this knowledge, it looks at the entire turn and creates a vector of ones and zeros that represent the semantics of the turn. For each turn, I will calculate its cosine similarity with the immediately previous turn (T-1 similarity) and the turn before the immediately previous turn (T-2 similarity). Cosine similarity ranges from -1 (opposite meaning) to 1 (same meaning). In Excerpt 27, the cosine similarity between lines 5-7 and lines 1-3 (a T-2 similarity) was 0.41 because they both discuss car use, Speaker 1s brother, and driving. Similarly, lines 8 and 4 had a cosine similarity of 0.47 (a T-2 similarity), because “mhm” and “yeah” are both backchannels that suggest understanding. In contrast, line 4 was dissimilar from lines 1-3 (a T-1 similarity), with a cosine similarity of -0.21.

Excerpt 27

Similarity example

```
SP1: uh yeah i had like the a car that was like designated for
     like me and my brother even though anyone in the family
     would drive it
SP2: mhm
SP1: and um my brother couldn't drive it until like the summer
     but once he could it got really annoying cause i just
     wouldn't have the car available
SP2: yeah
```

To validate these metrics, I calculated the median metric for each conversation and correlated them (Table 7). Since the medians were approximately normally distributed, I used Pearson correlations ($n_{converation}$ = 42). If all metrics represent progressivity as hypothesized, the median FTO of a conversation should be negatively correlated with the median speech rate and median information density, but positively correlated with T-1 and T-2 similarity. The median information density should be positively correlated with the median speech rate, but negatively correlated with the similarity metrics. Finally, the two similarity metrics should be positively correlated.

Table 7

Correlations between quantifications of progressivity

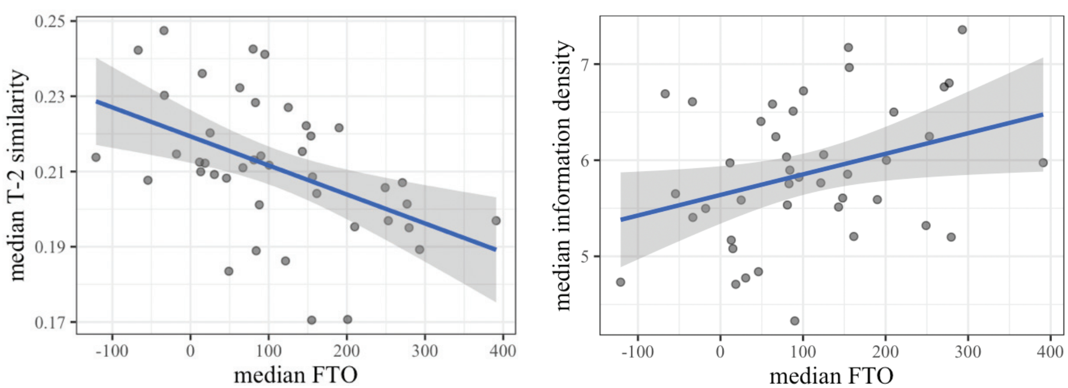
	Speech rate		Info Density		T-1 sim.		T-2 sim.	
	r	p	r	p	r	p	r	p
FTO	0.17	0.29	0.32	0.04*	0.24	0.13	-0.47	<0.01**
Speech rate	1	-	0.88	<0.01**	0.15	0.34	<0.01	0.97
Info Density	-	-	1	-	0.20	0.21	-0.18	0.27
T-1 sim.	-	-	-	-	1	-	0.08	0.63
T-2 sim.	-	-	-	-	-	-	1	-

Most statistically significant correlations were in the opposite direction than predicted. Median FTO was negatively correlated with median T-2 similarity (Figure 18, left). This means that in conversations where turns were more similar to the turn two turns ago, FTOs were shorter. This may be because the turns with the highest T-2 similarity were backchannels. When one person tells a story, the other may repeatedly backchannel, resulting in a high T-2 similarity. Backchannels are also associated with negative FTOs, which could explain the negative correlation between T-2 similarity and FTO. If a conversation has many backchannels, that conversation will likely have a shorter median FTO and a higher median T-2 correlation.

In addition, median FTO was positively correlated with the median information density of the conversation (Figure 18, right). This could be because more information takes longer to process. In fact, Chapter III found that more surprising turns (turns with more information) were followed by later FTOs. Both these findings go against the hypotheses outlined above, and there were no other statistically significant correlations with median FTO. The only other statistically significant correlation was the strong, positive correlation between median speech rate and median information density. This correlation shows that when people produce words more quickly, they produce information more quickly.

Figure 18

Floor transfer offset correlated with T-2 similarity and information density



In summary, these findings suggest that either some of the quantifications do not measure progressivity, that conversation-level measures are inappropriate, and/or there are multiple dimensions to progressivity that must be explored in greater detail.

The Denominator Problem. Frequency is typically defined as the number of occurrences over time. However, when quantifying phenomena found in natural conversation, researchers must consider the *denominator* or the "per what" problem (De Ruiter, 2013; Schegloff, 1993). Researchers can set the denominator to an objective metric like events per minute. However, these metrics are meaningless when considering conversation: a minute of conversation could contain 58 seconds of silence, one person talking for a minute straight, or many short turns where speakers are constantly exchanging the floor. Each of these possibilities change the expected base rate for any target behavior. More meaningful than a time-based metric would be setting the denominator to the number of *opportunities* for the behavior (De Ruiter, 2013; Schegloff, 1993). However, this solution also presents difficulties. For example, De Ruiter (2013) proposed identifying all the Transition Relevance Places (TRPs; Sacks et al., 1974), all the moments where it would be socially appropriate to take the floor. However, as De Ruiter (2013) points out, different researchers disagree on exactly where TRPs belong. Given the pros and cons of both options, I will perform analyses with a variety of denominators.

In this study, I will analyze the frequency of OIRs and miscommunications per minute of conversation, minute of speech, and speaker transition. I defined a minute of conversation as a minute that occurs between the start of the participants' first utterance directed at each other after the experimenter left the recording room and the end of the participants' last utterance before the researcher ended the conversation. I defined a minute of speech as a minute transcribed as part of a turn. This excludes between-turn gaps but includes within-turn pauses. In addition, if the participants spoke over each other, the duration of overlap would be counted twice – as speech produced by one speaker and as speech produced by the other. Finally, I defined a speaker transition as any time one interlocutor had been speaking, and another began to speak.

OIRs and miscommunications per minute of conversation maps on most directly to the raw data and is therefore easier to understand. OIRs and miscommunications per speaker transition is harder to interpret but is closest to solving the denominator problem. However, even this measure does not entirely solve the denominator problem. Speakers do not exchange the floor every time it is possible to do so: the same speaker may decide continue talking (Sacks et al., 1974). Listeners can also interrupt the current speaker to initiate repair, although doing so may be considered rude or nonnormative. OIRs and miscommunications per minute of speech is in the middle of the other two: it is more valid the per minute of conversation metrics and easier to interpret than the per speaker transition metric.

I calculated all the analyses in this study with all three denominators. All three metrics result in the same directionality of results, and almost always have the same statistical significance. Since OIRs and miscommunications per minute of conversation is the easiest to interpret and is the metric most commonly reported in other studies (e.g., Dingemanse et al., 2015; Kendrick, 2015a), I will present results using those metrics below. Where they differ from results using other metrics, I will report the other results as well.

Categorizing other-initiations of repair. I categorized all the OIRs in the ICC subset (see Chapter II). Open OIRs targeted the entire trouble source TCU, specific OIRs targeted a type of information, and candidate OIRs requested (dis)confirmation of a representation of the trouble source TCU. Table 8 displays the distribution of OIR categories in this dissertation, as well as OIR collections in peer-reviewed publications (Dingemanse et al., 2015; Kendrick, 2015a).

Table 8

Distribution of OIR categories

	ICC	Kendrick (2015a)	Dingemanse et al., (2015)
Duration	19.6 hours	7.4 hours	48.5 hours
Open	50 (13.02%), every 23.45 mins	42 (22.4%), every 10.5 mins	501 (33%), every 5.8 mins

Specific	102 (26.56%), every 11.50 mins	41 (21.9%), every 5.4 mins	371 (24.4%), every 7.8 mins)
Candidate	232 (60.42%), every 5.05 mins	104 (55.6%), every 4.2 mins	646 (42.5%), every 4.5 mins
All	384, every 3.05 mins	187, every 2.4 mins	1,518, every 1.9 mins

Note: The estimates of the Kendrick (2015a) and Dingemanse et al. (2015) collections are approximate, as they used different OIR taxonomies than the one used for this collection. For example, they categorized “repetition” as a separate class from open, candidate or specific. Since many repetitions request (dis)confirmation of an understanding, I considered repetitions in Kendrick (2015b) and Dingemanse et al. (2015) to be candidates. However, this may result in overrepresentation of candidates and an underrepresentation of specific OIRs in those distributions.

There are some differences between these collections. The conversations in the ICC were between two interlocutors, in a laboratory environment with soundproofed rooms, excellent audio and no distractions. The conversations in Kendrick (2015a) and Dingemanse et al. (2015) were between friends and family members and occurred in natural environments when interlocutors may have been playing games or preparing food together. The many differences in these corpora mean that any differences in OIRs cannot be attributed to a single cause. However, it is worth noting that the rates of OIRs are less frequent in the ICC than in other collections. This is especially true for open OIRs, which occurred more than twice as often in Kendrick (2015a), and four times as often in Dingemanse et al. (2015). This is likely because the ICC was recorded in a soundproofed laboratory with no noises or distractions, two common sources of open OIRs. In fact, given the environment, this may be more open OIRs than expected. Except for one a few sequences when interlocutors removed their headphones or admitted to "zoning out," the open OIRs were problems of understanding, especially during "deep" conversations or after silly turns (e.g., Excerpt 28). Perhaps the flexibility of the task in the ICC resulted in more open OIRs, in contrast to the task-oriented conversations typically recorded in experimental laboratories.

Excerpt 28

Open OIR after a silly turn

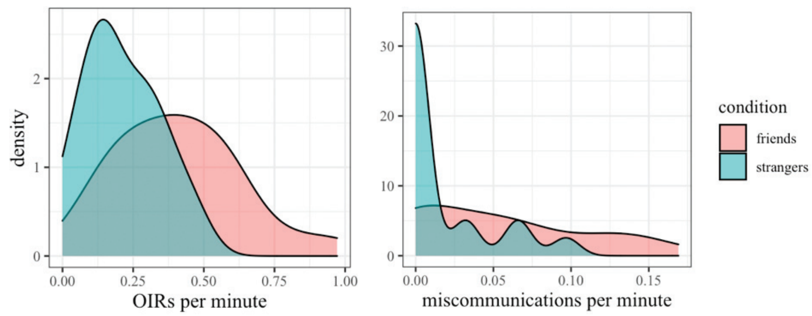
```
SP2: if like just by the way: i'm like really like (0.5) °i'm
     talking° (0.6) just like (.) i can't do a:nything,
     (1.1)
SP1: what?
SP2: i was talking to the ↑camera.
```

Does egocentricity contribute to progressivity and noticed lapses in intersubjectivity?

Do friends have more communication problems? Figure 19 shows the distribution of OIRs (left) and miscommunications (right) per minute in conversations between friends ($n_{friend\ conversations}$ = 24) and strangers ($n_{stranger\ conversations}$ = 18). Many conversations between strangers had low rates of OIRs and no miscommunication sequences. Therefore, I used Wilcox signed ranks tests to determine whether conversations between friends have more OIRs than conversations between strangers. Friends produced more OIRs per minute of conversation (Mdn = 0.37) than strangers (Mdn = 0.18, W = 94, p < 0.01). Friends had more miscommunications per minute (Mdn = 0.04) than strangers (Mdn = 0, W = 126.5, p = 0.01). These findings supported my hypothesis that friends would have more OIRs and miscommunications than strangers.

Figure 19

OIRs and miscommunications frequency for friends and strangers



Next, I determined whether friends and strangers differed in the type of OIRs they produced. Table 9 shows that friends produced more open and specific OIRs per minute of conversation. There was no statistically significant difference in the rate of candidate OIRs per minute of conversation. However, this difference became statistically significant when using different denominators. A Wilcoxon signed rank test found that friends (*Mdn* = 0.27, *MAD* = 0.09) had more candidate OIRs per minute of speech ($W = 129$, $p = 0.03$) than strangers (*Mdn* = 0.14, *MAD* = 0.15). Another test found that friends (*Mdn* = 4.5 thousand, *MAD* = 12 thousand) had fewer speaker transitions per candidate OIR than strangers (*Mdn* = 8.5 thousand, *MAD* = 7.8 thousand, $W = 125$, $p = 0.02$). These results suggest that friends conclusively produce open and specific OIRs more often but may or may not produce candidate OIRs more often.

Table 9

Frequency of OIR categories

	Open per min	Specific per min	Candidate per min

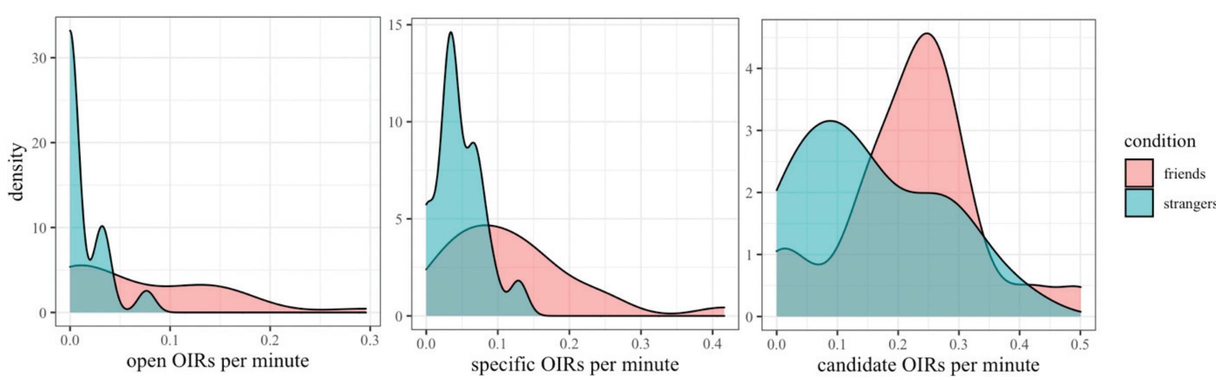
	Friends	Strangers	Friends	Strangers	Friends	Strangers
Median	0.04	0	0.10	0.03	0.23	0.13
MAD	0.06	0	0.09	0.04	0.09	0.14
Wilcox test	$W = 121$, $p = 0.01$*		$W = 96$, $p < 0.01$**		$W = 148$, $p = 0.08$	

Note: Wilcoxon signed ranks tests found no statistically significant difference between friends and strangers in the number of candidate OIRs per minute but found that friends had more candidates per minute of talk and fewer speaker transitions per candidate OIR.

Figure 20 displays the distribution of the frequency of OIR categories for friends and strangers. For open and specific OIRs, friends had a broad and flat distribution, while many strangers had very few (if any) open or specific OIRs. In contrast, more strangers produced more than one candidate OIR.

Figure 20

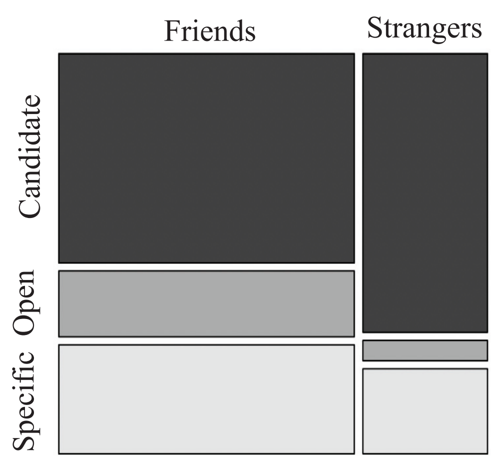
The rate of OIR categories in conversations between friends and strangers



Since the Wilcoxon signed-ranks tests suggested that friends produced more open and specific, but not necessarily candidate OIRs than strangers, I performed a follow-up chi squared test to determine whether friends and strangers chose different OIR categories when they produced an OIR. The distributions were statistically significantly different ($\chi^2 = 13.7$, $df = 2$, $p < 0.01$). As Figure 21 demonstrates, this effect was likely driven by the fact that most (72.57%) OIRs produced by strangers were candidate OIRs, where only half (55.76%) of OIRs produced by friends were candidate OIRs.

Figure 21

OIR distributions across friends and strangers



Are friends more progressive? I used Wilcox signed rank tests to examine whether turns in conversations between friends ($n_{turns\ in\ friendly\ conversations} = 16{,}724$) were more progressive than turns in conversations between strangers ($n_{turns\ in\ stranger\ conversations} = 14{,}814$). As shown in Table 10, friends had longer FTOs and a slightly higher T-1 similarity than strangers, indicating lower progressivity. At the same time, friends had a higher information density, faster speech rate, and lower T-2 similarity, indicating greater progressivity.

Table 10

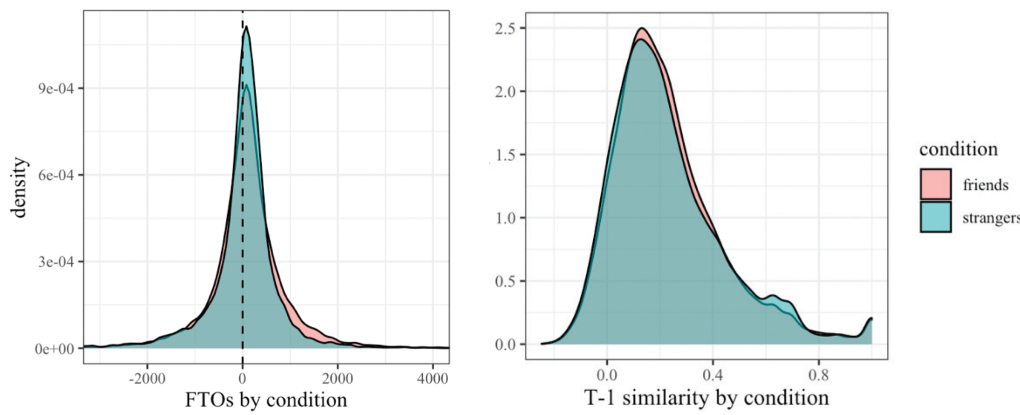
Wilcoxon signed rank tests of progressivity

	Friends		Strangers		Test	
	Median	MAD	Median	MAD	W	p
FTO	91ms	501.11ms	76ms	386.95ms	>1000	<0.01**
Words per second	3.7	1.39	3.58	1.31	>1000	<0.01**
Info density	6.09	2.78	5.61	2.8	>1000	<0.01**
T-1 similarity	0.193	0.17	0.187	0.18	>1000	0.05*
T-2 similarity	0.21	0.17	0.22	0.19	>1000	<0.01**

Figure 22 displays FTOs (left) and T-1 similarities (right) for friends and strangers. FTOs in conversations between friends were less likely to be between -160ms and 470ms than FTOs in conversations between strangers. This range is the range of most FTOs in conversation. At the same time, FTOs in conversations between friends were more likely to between approximately 470ms and 2 seconds than FTOs in conversations between strangers. Gaps within this range typically indicate upcoming troubles. The distributions of T-1 similarities were similarly shaped for friends and strangers, but the distribution for friends was shifted slightly to the right. However, strangers were more likely to have turns with a T-1 similarity between 0.55 and 0.75.

Figure 22

Metrics that indicated friends were less progressive than strangers

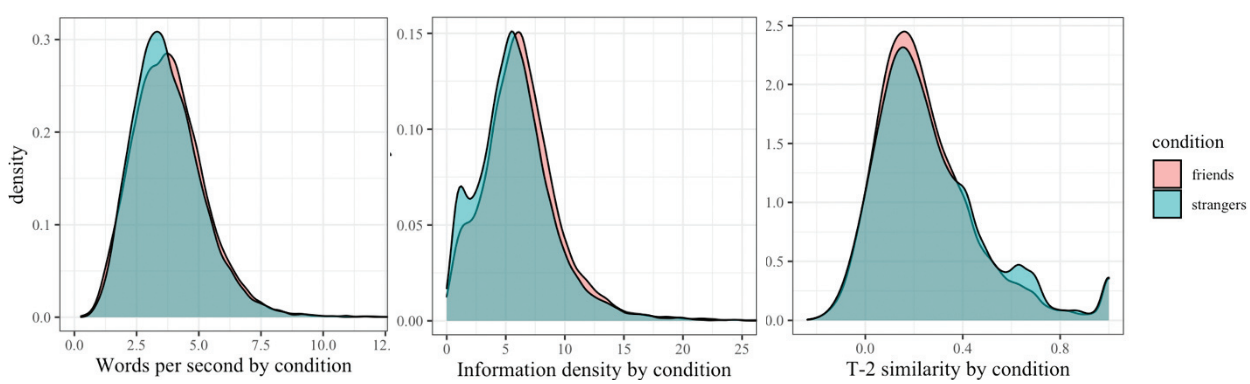


Note: Friends had longer FTOs (left) and higher T-1 similarity (right) than strangers.

Figure 23 shows the metrics that indicated friends were more progressive than strangers. In general, the distributions for friends were shifted slightly to the right for each metric. However, strangers were more likely to produce turns with an information density less than approximately 1.5 (Figure 23, middle) and a T-2 similarity between approximately 0.55 and 0.75 (Figure 23, right).

Figure 23

Metrics that indicated friends were more progressive than strangers



Note: Friends had more words per second (left), higher information density (middle) and lower T-2 similarity (right) than strangers.

I identified turns that fit into the “bump” shown by strangers in T-1 similarity, T-2 similarity, and information density distributions. These are turns with a low information density but high similarity with the previous turn and the turn before the previous turn. Many of these turns were like line 12 in Excerpt 29. Typically, one interlocutor is “telling” while the other is “listening” for multiple TCUs (Speaker 2 in lines 3, 5, and 8). When the first interlocutor finishes, both interlocutors tend to exchange minimal responses (lines 9, 11 and 12). It is usually one of these minimal responses that fits into the “bumps” in the graphs above. Sequences like these may be due to a difficulty in topic selection. A few turns before this sequence, the interlocutors were explicitly searching for topics (“what else should we talk about?”). In addition, in this part of the conversation, Speaker 1 has been asking Speaker 2 questions; when Speaker 2 finishes answering, Speaker 1 asks another question. In fact, immediately after this turn Speaker 1 asks Speaker 2 another question, starting the conversation again. These patterns suggest that while just one metric alone may not provide enough information to identify progressive or unprogressive sequences, the combination of multiple metrics may be able to identify at least some unprogressive sequences. Further, this provides some preliminary evidence that natural conversations between strangers may be less progressive because strangers struggle with topic selection.

Excerpt 29

Low information density and high similarity turn

```
SP1: do you have any fun p:lans,
     (0.5)
SP2: u::h (0.3) probably l:ater for the like not today.
SP1: ⌈ yeah.  ⌉
SP2: ⌊cause i l⌉ike I i'm w:orking today? i work ⌈monday we⌉dnesd:ays,
SP1:                                            ⌊ okay,  ⌋
     (0.4)
SP2: s⌈o: p:robably⌉ l:ike at the end of the weekend.
SP1:  ⌊ gotcha,  ⌋
     (1.5)
SP1: okay, that's awe⌈some,⌉
SP2:                 ⌊ yea⌋:h,
```

Do progressive conversations have more problems? I correlated the medians of quantitative metrics of progressivity (FTO, words per second, information density, T-1 similarity, T-2 similarity) and the rates of OIRs and miscommunications. A higher information density and a lower T-2 similarity (Figure 24), indicating greater progressivity, corresponded to more miscommunications and OIRs per minute. Finally, when there were more OIRs in a conversation, there were also more miscommunications.

Table 11

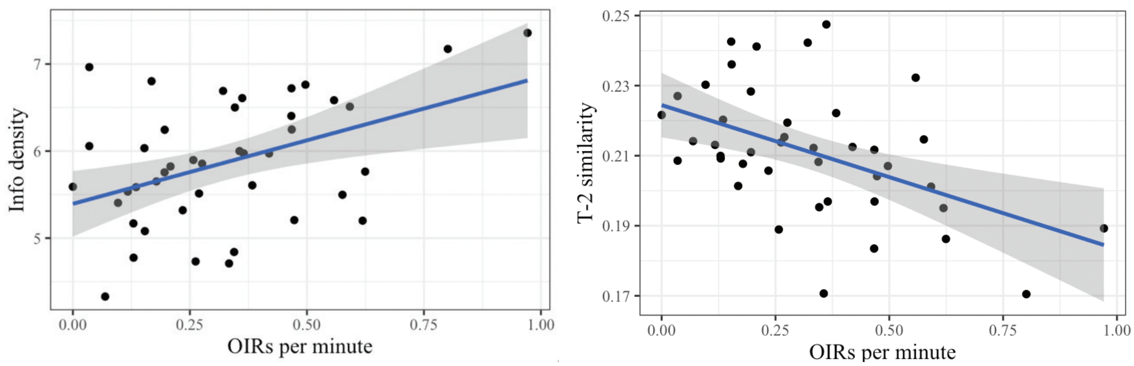
Progressivity metrics correlated with communication problems

	Misc. per min		OIRs per min	
	r	p	r	p
FTO	0.28 (-0.02 – 0.54)	0.07	0.28 (-0.03 – 0.53)	0.08
Speech rate	0.25 (-0.05 – 0.52)	0.10	0.28 (-0.03 – 0.54)	0.07
Info Density	0.31 (0.00 – 0.56)	0.05*	0.43 (0.17 – 0.65)	<0.01**
T-1 sim.	0.14 (-0.17 – 0.43)	0.38	0.11 (-0.20 – 0.40)	0.49

T-2 sim.	-0.36 (-0.60 – -0.07)	0.02*	-0.48 (-0.68 – -0.21)	<0.01**
Misc. per min	1	-	0.47 (0.19 – 0.67)	<0.01**

Figure 24

The relationships between progressivity metrics and OIRs



Egocentricity Contributes to Unnoticed Lapses in Intersubjectivity

In this section, I describe patterns in the miscommunication corpus (see chapter II) that suggest egocentricity may contribute to miscommunication. First, ambiguity was present in 18 (41%) of miscommunications. However, while experiments on egocentricity typically focus on the potential ambiguity of spatial words (e.g., “left” when the interlocutors face different directions) or objects (e.g., “move the mouse”), most ambiguity problems in this corpus were more abstract. For example, in Excerpt 30, the interlocutors miscommunicate because there are two meanings of “burlesque”: the idea of burlesque, and the Burlesque team at Tufts University.

Excerpt 30

Interlocutors assume smaller problems

```
SP2: um i ∆am a part of peer he:alth exchange?∆
     (0.4)
SP1: °pee:r health exchange oka⌈y,°⌉
SP2:                           ⌊ y ⌋eah and burle:sque, (0.3) oo
     (0.4)
SP1: i haven't heard of that,
     (0.7)
SP2: you haven't heard of burle:s⌈que?⌉
SP1:                             ⌊  n⌋o,
SP2: cr:a:zy:. you'll s⌈ee.        ⌉
SP1:                   ⌊what do you guy⌋s d⌈o.      ⌉
SP2:                                      ⌊you'll s⌋ee. co:me the end
     of the semester? (0.4) um,
     (0.4)
SP1: is that a (0.2) performing: ⌈  ↑thing,     ⌉
SP2:                             ⌊yeah ∆wait ha⌋ve you never heard of
     just like burlesque∆ in general?
SP2: okay
    (1.1)
SP2: wild okay
```

Excerpt 30 also demonstrates that when interlocutors do notice a problem, they may incorrectly assume the smallest problem possible. Speaker 2 is very knowledgeable about burlesque, so by assuming Speaker 1 knows at least the concept of burlesque, Speaker 2 assumes the interlocutors are more similar than they are. In Excerpt 20, the interlocutors knew there was an understanding problem with “downsizing,” but assumed the problem was in how the downsizing worked, not the basic meaning of the word downsizing. This suggests that interlocutors may be egocentric in diagnosing problems: they assume their social partner’s understanding is closer to theirs than it is. Seven (17.7%) miscommunications contained an OIR, expression of disbelief, or other behavior that suggested that someone misdiagnosed an intersubjectivity problem.

Excerpt 31

Listeners use more recent context to interpret utterances

```
SP2: she was like walking around and she was like you're in my seat
     you're in fifteen and i was like (0.4) i'm in five i don't think
     i'm in your seat (0.8) and then she tried to climb over and
     spilled beer on someone
     (0.6)
SP1: but it wasn't on you≈
SP2: ≈it wasn't on me
     (0.3)
SP1: that's unfortunate
SP1: what time ⌈of day was it⌉
SP2:           ⌊what do you mean t⌋hat's unfortunate
     (0.4)
SP1: well she spilled beer on som⌈ebody it's sticky and gross⌉
SP2:                             ⌊oh i thought you meant it's
     unfo⌋rtunate that it wasn't on me
```

Note: The solid black box contains the trouble source TCU, the dotted pink box contains the listener's context, and the grey dashed box contains the speaker's context.

Further, 7 (17.1%) of miscommunications occurred when listeners assumed a turn referred to more recent context, but speakers produced their turn using more distant context. In Excerpt 31, lines 1-4 told a story, a first pair part of an adjacency pair. Speaker 1 begins an insert sequence with a follow-up question. When that insert sequence is complete, Speaker 1 produces the second pair part of the adjacency pair, an emotional reaction to the first pair part (line 9). Speaker 2 perceives this second pair part as referring to the second turn of the insert sequence (line 7) instead of the first pair part (lines 1-4).

Excerpt 32 provides a useful counter-example. "The same class" on line 17 was ambiguous and could have referred to either "psych one" (line 4) or "psych stats" (line 11). Like in Excerpt 31, the recipient of the turn connected the ambiguity to the more recent context. Unlike Excerpt 31, the speaker intended the recipient to connect "the same class" to the more recent context ("psych stats"). Together, these examples show that listeners connect turns to their most recent context, but speakers may construct their turns in reference to more distant context.

Excerpt 32

Listeners and speakers connect to most immediate context

```
1  SP1: uh: psych one?
2       (0.6)
3  SP2: yes
4  SP2: p⌈sych on⌉e.
5  SP1:  ⌊okay. ⌋
6  SP1: cool. cause i know like some other psych classes have to do the
7       who:le (0.3) requirement,
8  SP2: ⌈  really  ⌉
9  SP1: ⌊of (.) stu⌋dies
10      (0.4)
11 SP1: yeah i have a friend who's in (0.2) psych stats and (0.4) he has
12      (0.3) n:ot done any of his credit⌈s ⌉
13 SP2:                                  ⌊o⌋⌈h wow⌉
14 SP1:                                     ⌊ ye:⌋⌈t,  ⌉
15 SP2:                                           ⌊tha⌋t's ba::⌈d.⌉
16 SP1:                                                      ⌊ye⌋ah,
17 SP2: i'm actually: t↑aking the same class l- um next year or (0.2)
18      not next year. Next semester

19 SP1: oh psych s⌈tats ⌉
20 SP2:           ⌊i lik⌋⌈e⌉
21 SP1:                  ⌊i⌋ thought you were just (0.3) say you're taking
22      psych one again.
```

Occasionally, friends incorrectly corrected each other or pushed the conversation forward by offering incorrect information. Before Excerpt 33, the interlocutors were discussing whether they should go to a retreat in Mexico for spring break. Speaker 2 assumes “ID” (line 1) refers to an ID one needs to purchase alcohol. In Mexico, the drinking age is 18; when Speaker 2 says “it’s legal there” they are referring to drinking alcohol between the ages of 18 and 21. Speaker 2 interrupts Speaker 1 to show they think they know where the conversation is going, and to get there faster. However, it turns out that Speaker 1 was referring to a school ID, to enter the retreat for UCLA. They correct Speaker 2 on line 6 though third position repair (Schegloff, 1992). Such incorrect corrections are rare, and only happen between friends.

Excerpt 33

Friends assume epistemic authority

```
SP1: they don't check i- it's not like an ID it'⌈s like  ⌉
SP2:                                          ⌊well yea⌋h. it's
     ⌈legal the:re?⌉
SP1: ⌊you just go: ⌋
     (0.7)
SP1: Δno no no but ID of like theΔ school. like you don't l:ike
     you go as a ⌈student of UCL⌉A but you don't actually have to be
     a student.
SP2:             ⌊     o::h.    ⌋
```

Finally, in one case, a friendly dyad miscommunicated but did not resolve the problem – perhaps to continue progressivity. In Excerpt 34, Speaker 2 mentions bringing their car up "senior year." The interlocutors are both freshman, so when Speaker 1 mentions "next year" on line 6, they are referring to "sophomore year." This is a not what Speaker 2 says on line 4, and Speaker 2 disconfirms that representation of line 4 when they say "no" on line 8. However, Speaker 1 then corrects their understanding to "junior year" (line 9) which is also incorrect; instead, of correcting this second incorrect representation, Speaker 2 says "yeah yeah."

Excerpt 34

Abdicated other-correction (Jefferson, 2007)

```
SP1: i kno:w. i need a c:a:r,
     (0.8)
SP2: ugh
SP2: i would wanna ↑bring my ca:r (.) u:m (0.3) senior year,
     (0.5)
SP1: next year,
     (0.7)
SP2: n⌈o,   ⌉
SP1:  ⌊oh j:⌋unior ye⌈ar.   ⌉
SP2:                 ⌊yeah y⌋eah,
SP1: yeah. i was like i don't thi:nk i would u:se it sophomore year cause
     like Davis is (0.7) so close,
```

Discussion

In this study, I investigated whether friends – shown to be more egocentric than strangers in experimental settings – have more, and worse, communication problems in conversation. Then, I investigated whether friends were more progressive than strangers, and if progressivity was associated with a higher frequency of communication problems. I hypothesized that a) friends would have more communication problems than strangers, b) friends would use more egocentric repair strategies than strangers, c) both listeners and speakers would be more egocentric when talking to friends, d) friends would be more progressive than strangers, and e) higher levels of progressivity would be associated with more communication problems. In addition, I examined miscommunication sequences to explore the relationships between egocentricity, social closeness, progressivity, and miscommunication.

First, I found that friends (M = 2.44 min per OIR) produced OIRs almost twice as often as strangers (M = 4.74 min per OIR). In addition, friends (M = 18.80 min per miscommunication) miscommunicated more than three times as often as strangers (M = 61.46 min per miscommunication). Further, an OIR produced in a friendly dyad was three times more likely to be open class than an OIR produced in a stranger dyad. Open OIRs have been described as “locally egocentric” because they place more work on the other interlocutor than specific or candidate OIRs. These findings complement the experimental findings that friends are more egocentric than strangers (e.g., Savitsky et al., 2011). In fact, the effect size in this study is greater than that reported in experimental contexts: Savitsky et al. (2011) found that friends were 1.6 times more likely to select the wrong object. Participants likely prioritize intersubjectivity more in a task-oriented experimental environment, where their accuracy is being recorded, than a casual conversation with no obvious task or success criteria. Alternatively, natural conversation may harder than tasks in experiments on egocentricity. In experimental settings, the interlocutors discuss objects in their physical environment; in most of the miscommunications in this corpus, the interlocutors discussed abstract concepts, events from the past, or other topics not physically present.

This finding also suggests that both speakers and listeners are more egocentric when talking to friends. Friendly speakers produce many more problems, indicated by higher rates of OIRs and miscommunications overall. Friendly listeners notice a smaller proportion of problems, indicated by a greater increase in miscommunications than in OIRs for friends.

Second, I quantified progressivity. I examined FTO, speech rate (words per second), information density, and similarity with the previous turn (T-1 similarity) or turn before the previous turn (T-2 turn). FTO and the similarity metrics should be higher when conversations are less progressive; speech rate and information density should be higher when conversations are more progressive. To evaluate how well these measures account for progressivity, I took the median of each metric for each conversation and correlated them. I found that median FTO was positively correlated with median information density and negatively correlated with the median

T-2 similarity. These results are not straightforward to interpret and suggest that either these metrics are inadequate or progressivity is multidimensional.

Future research should generate and validate better metrics for progressivity. A true measure for progressivity would require modeling adjacency pairs (Schegloff, 1995). Adjacency pairs are two turns where the first turn creates an expectation for the second. For example, a question creates the expectation for an answer, and an invitation creates the expectation for an invitation or rejection. In progressive sequences, the first in the pair is immediately followed by the second in the pair. In less progressive sequences, there are turns or sequences embedded within the adjacency pair.

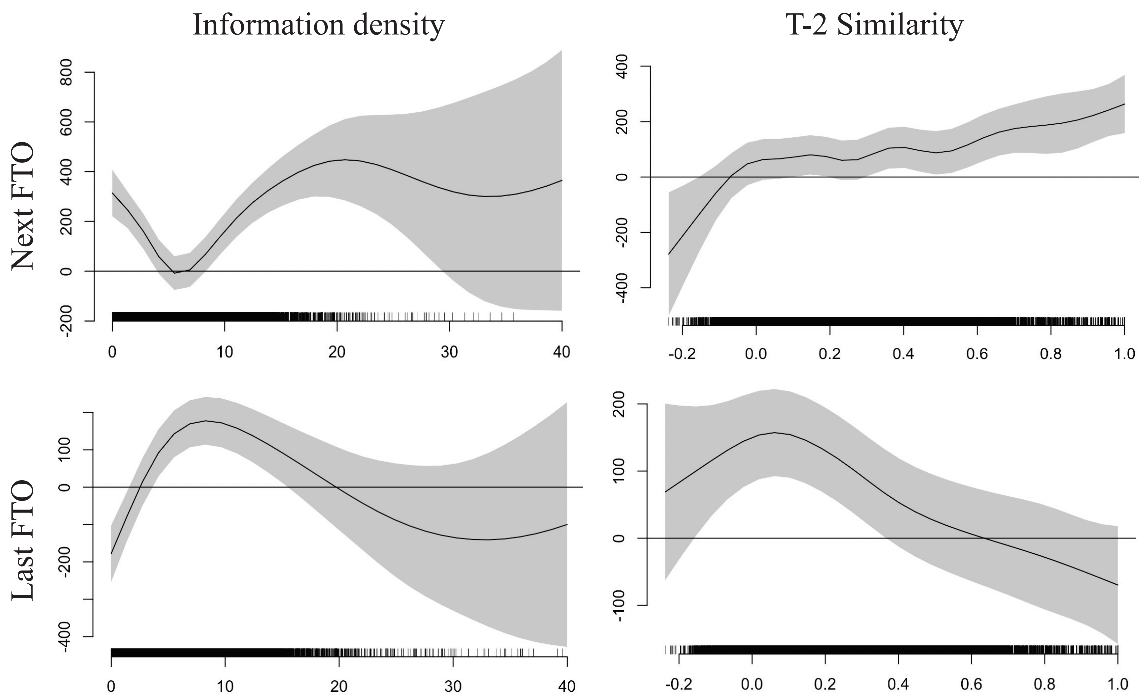
Only a few researchers have attempted to model adjacency pairs (e.g., Boyer et al., 2009). If a model could automatically identify adjacency pairs in conversation, we could develop a more sophisticated metric for the progressivity of a conversation. Specifically, researchers could calculate the number of turns between first and second pair parts.

In addition, it is unclear exactly how different metrics contribute to progressivity on the sequence level. For example, on the conversation level, FTO was positively correlated with information density and negatively correlated with T-2 similarity. However, the FTO before a turn had a different relationship with information density and T-2 similarity than the FTO after the turn. I computed GAMMs that included a random intercept for conversation and used either information density or T-2 similarity to predict the FTO after or before the current turn (Figure 25). Information density and T-2 similarity have almost opposite relationships with the FTO before or after the current turn. These findings show that interpreting conversation-level metrics is difficult, and we must identify exactly how these different quantifications relate to each other.

Future research should examine turn- or sequence-level progressivity metrics and use time series analyses to identify patterns of progressivity and repair in conversation.

Figure 25

GAMMs exploring progressivity metrics



Next, I determined whether friends were more progressive than strangers. Friends were more progressive in that they produced turns that had more words per second, higher information densities, and lower T-2 similarities. At the same time, friends were less progressive in that they had longer FTOs than strangers and had higher T-1 similarities. It is unclear exactly what this means for the hypothesis that friends are more progressive than strangers. To interpret these findings, researchers may need to improve the quantifications of progressivity, create more sophisticated models, or more comprehensively analyze the data. For example, many long gaps in conversations between friends could be caused by jokes – in my analyses, laughter was considered silence. It could also be that one component of progressivity, like information density, affects another, like the following FTO.

Third, I correlated the rate of communication problems with progressivity metrics. More frequent OIRs and miscommunications were associated with a higher information density and lower T-2 similarities. This provides some, albeit preliminary, evidence that more progressive conversations also have more communication problems.

Finally, I examined analyzed the miscommunication corpus to determine whether egocentricity, friendship, or progressivity could be connected to miscommunication sequences. A few interesting patterns emerged. First, 17 (41%) miscommunications involved ambiguity. This is interesting because ambiguous language opens the opportunity for egocentricity to cause misunderstandings. Speakers underestimate the ambiguity of messages (Keysar & Henly, 2002) and listeners process ambiguous turns from their own perspective (Savitsky et al., 2011). Perhaps because of the experimental environment – two adjoining rooms with very few objects – the ambiguities centered around pronoun use ("it" vs. "that") and abstract concepts (e.g., "fire" referring to fireplace or forest fire), instead of physical objects (e.g., "pass me the water bottle" when there are two water bottles). It may be that referring to physical objects is an easier task than keeping track of the ambiguity of abstract ideas or objects/events from the past.

Some types of miscommunications could be attributed to egocentricity. First, some miscommunications occurred when the interlocutors noticed a lapse in intersubjectivity but identified the problem as smaller than it was. Similarly, some miscommunications were caused by a difference in attention to context: interlocutors miscommunicated because listeners connected turns to more recent information, while speakers intended turns to be connected to more distant information. Finally, friends incorrectly assumed they knew what their social partner was about to say (e.g., Excerpt 33). These three patterns of miscommunication suggest that egocentricity directly contributes to miscommunication in casual conversation.

However, there were several important limitations to this study. First, social closeness a continuous construct, but was treated in this study as binary. The only requirement to be considered a friendly dyad was that the two participants signed up together. These dyads may have known each other for years, they could be romantic partners, or, they may have met in class just a few days ago. This study cannot determine whether there is a continuous effect of social closeness on egocentricity or communication problems.

In addition, there was no measure of how social closeness changed over the course of the study. A dyad may begin as strangers but become friendly by the end of the session, especially if they discover they have a lot in common. It was not rare for interlocutors in the stranger condition to realize they follow the same sports, know each other’s roommates, or are in the same classes, allowing them to quickly increase their common ground. Even “strangers” in this study that did not share acquaintances had a larger common ground than the average two strangers. All participants were Tufts undergraduate students in either Introduction to Psychology or Statistics for Behavioral Sciences. Tufts University is also a relatively small school, with an undergraduate enrollment of 6,114 in Fall of 2020. This means that – unlike a strange one meets on the street or in the airport – the interlocutors could reasonably expect to see each other again, perhaps in another course. In summary, the terms “friends” and “strangers” in these study are relatively imprecise. Future work should attempt to measure closeness, for example with surveys like the Inclusion of the Other in the Self Scale (Aron et al., 1992).

Finally, this study does not eliminate the possibility that some (or all) of the differences between friends and strangers are due to the structure of the conversation, and not due to egocentricity. Many conversations between strangers begin with standard get-to-know-you questions, like “What year are you?” or “What’s your major?” Further, at least in some conversations, participants fall into implicit social roles of “questioner” and “answerer.” However, even participants in some friendly conversations seem to fall into such roles. Future research should investigate whether social closeness or social role is a better predictor of communication problems in conversation: it may be that with more clearly defined social roles, there are fewer OIRs and miscommunications.

In summary, this study provides evidence that egocentricity causes communication problems in conversation. Friends, who are more egocentric than strangers, have more communication problems and use more egocentric repair mechanisms to resolve those problems. It also suggests that communication problems are associated with progressivity, although future research should work towards understanding how to quantify progressivity. Interlocutors fail to take each other’s perspectives in conversation, causing lapses in intersubjectivity but potentially increasing progressivity.

Chapter V: General Discussion

I investigated whether miscommunication occurs because interlocutors prioritize progressivity and/or deprioritize intersubjectivity. Maintaining intersubjectivity requires time but reduces the likelihood of miscommunicating. Maintaining progress requires accepting some uncertainty but moves conversation along.

I focused on two psycholinguistic concepts that have been extensively studied by psychologists: prediction and egocentricity. Prediction is thought to increase progressivity – it decreases the cognitive load of comprehension, allows listeners to plan their turns before the current turn is finished, and helps interlocutors time their turns more effectively. Each of these effects of prediction likely contribute to the very short gaps between turns in conversation; without prediction, turn transition times would likely be much longer. In Chapter III, I investigated whether prediction also interferes with intersubjectivity. Egocentricity, the tendency to produce and interpret linguistic information from ones' own perspective, has been hypothesized to cause miscommunication (Keysar, 2007). It may also increase progressivity because perspective-taking requires cognitive resources and time. In Chapter IV, I explored the relationship between egocentricity, progressivity and intersubjectivity.

These studies complement previous efforts to categorize problem sources in conversation. In fact, some common sources of communication problems may be so prevalent because of the cognitive processes discussed in this study. For example, ambiguity may contribute to so many problems because interlocutors are egocentric. However, neither research documenting communication problems nor exploring the cognitive mechanisms that contribute to miscommunication provide obvious interventions to reduce miscommunication. It is unclear how we could reduce either prediction or egocentricity, or if we would even want to. Not only does prediction increase progressivity, but it also helps listeners cope with incomplete stimuli, background noise, ambiguity, and other problems. Egocentricity may also have benefits to interlocutors that we do not know about. There is some evidence that multitasking or introducing time pressure can increase egocentricity and poor communication (Epley, Keysar, et al., 2004), so it may be wise to reduce time and resource constraints when communicating about important information. However, egocentricity and mis-predictions occur even when interlocutors have adequate time. Future work should investigate clinical interventions designed to prevent miscommunication, but must tread carefully, examining risks of interventions designed to prevent prediction or egocentricity.

The most immediately helpful clinical perspective may be to focus on adjusting how people perceive miscommunication. If people perceive miscommunication as inevitable event, they may be able to mitigate the negative consequences of miscommunication. For example, they may incorporate institutional procedures to check understanding before critical decisions of minutes. In personal relationships, miscommunication should be treated not as a personal affront, but as a natural and inevitable occurrence. In fact, in some relationships, miscommunications could be treated as neutral, since people miscommunicate more often with people they feel comfortable with.

Dialog systems, computers that communicate with people, are more and more commonplace every day. In fact, dialog systems are starting to take over fields where communication is exceedingly important, like healthcare (Laranjo et al., 2018). The designers of dialog systems – especially when dealing with important topics – must consider how progressivity interferes with intersubjectivity. First, many companies design chatbots to perform as efficiently as possible. Therefore, many chatbots strongly prioritize progressivity over intersubjectivity. This way, companies can maximize the number of customers they assist. However, this perspective will increase the frequency of communication problems and miscommunication. Especially unfortunate is that many dialog systems have poor conversational repair mechanisms. Typically, they have standard *fallback* messages, inflexible messages like "I'm sorry I didn't understand. Please repeat" whenever natural language understanding confidence decreases under a threshold. Along with more flexible repair mechanisms, computer scientists should start considering preference organization, including progressivity and intersubjectivity, in their design of dialog systems. A truly competent dialog agent needs to be able to prioritize progressivity or intersubjectivity depending on the context.

Dialog systems also provide a unique opportunity to reduce miscommunication amongst humans. Since they are not sensitive to preference structure, they do not care if they are socially awkward. A dialog system teammate could be used to slow down conversations when needed, for example by initiating repair on ambiguous language. This way, the humans on the team could avoid face-threat while maintaining intersubjectivity. Since it is likely difficult to change the cognitive processes that increase progressivity but decrease intersubjectivity, dialog systems may be an effective alternative intervention.

The findings in Chapter III and IV underlie the importance of analyzing both cognitive and social factors when researching conversation. Social goals (like progressivity) could arise from the human cognitive architecture: humans can time their turns precisely because they predict and are egocentric, and therefore they are expected to do so. Alternatively, social goals may affect the human language system: humans prioritize progressivity, and therefore use prediction and egocentricity to do so. Regardless of the directionality of this relationship, psychologists interested in conversation simultaneously analyze social and cognitive factors in conversation. For example, psycholinguists typically look at reaction time to determine whether one type of language is harder to produce or interpret than another. In conversation, floor transfer offset (FTO; De Ruiter et al., 2006) – the time between turns – is the analogous metric. However, FTO in conversation is affected by both social and cognitive factors (Mertens & De Ruiter, 2021; Roberts et al., 2015).

By necessity, these studies deeply simplified the concept of preference organization. First, interlocutors balance many more preferences than just the preferences for progressivity and intersubjectivity. Preferences can be *cross-cutting* and/or *hierarchical.* Cross-cutting preferences require the interlocutor to (at least sometimes) prioritize one over the other. These studies investigate the cross-cutting preferences of progressivity and intersubjectivity, but progressivity and intersubjectivity are cross-cutting with other preferences as well. Hierarchical preferences are more specific or general manifestations of another. For example, there is a preference for "stronger" repair initiators, that provide more information about the specific problem, over "weaker" repair initiators (Schegloff et al., 1977). Stronger repair initiations tend to request shorter solutions (Dingemanse et al., 2015), so the preference for stronger initiators may really be a manifestation of the preference for progressivity. Finally, I have assumed that interlocutors maximize more immediate progressivity and not longer-term progressivity. Longer-term progressivity, like the progressivity of a semester-long project, could increase if interlocutors prioritize intersubjectivity early on. This is because any short-term progress is erased once it becomes clear the interlocutors have miscommunicated. Future work investigating how progressivity affects miscommunication should take a more detailed, complex view progressivity.

These studies also simplified the cognitive processes that contribute to communication but have also been linked to miscommunication. First, there are clear ways that prediction and egocentricity may influence each other. For example, if a listener has a very strong prediction, they may use that prediction instead of taking their interlocutors' perspective. In addition, there are disparate phenomena that are likely connected to miscommunication but may or may not be independent from egocentricity and prediction. For example, a broad literature show that comprehenders use *shallow processing* – they only process the most important semantic features from language (e.g., Barnard et al., 2004; Sanford et al., 2011). They also aim for *good-enough representations*, or representations of utterances that are plausible, even if they are not accurate (Ferreira & Patson, 2007). However, these cognitive processes or strategies overlap significantly with each other, and likely interact. Comprehenders may use shallow processing more often when they have strong predictions, perhaps they accept good-enough representations only when they match up with their predictions, or egocentric processing may be a subtype of shallow processing. These are only examples of the myriad ways these findings may intersect. Determining how these findings and cognitive processes relate will be challenging, but vital.

Perhaps the biggest limitation to this work, and the field more generally, is a problem of data. I could only analyze miscommunications that were noticed and resolved within the conversation. Longer-lasting (or even permanent) miscommunications are likely different than the ones resolved in just a few turns. Further, it could be argued that the longest-lasting, hardest-to-identify miscommunications are the ones that have the greatest potential to cause problems, and therefore are the most important to study. Some studies have used retrospective methods, where interlocutors go over their conversation and identify and explain miscommunications (Mustajoki & Sherstinova, 2017). Another study asked participants to submit transcripts of their text-based miscommunications (Kelly & Miller-Ott, 2018). As the study of miscommunication develops, methods should focus specifically on longer-term miscommunications.

References

Aoki, N., Uda, K., Ohta, S., Kiuchi, T., & Fukui, T. (2008). Impact of miscommunication in medical dispute cases in Japan. *International Journal for Quality in Health Care*, *20*(5), 358–362. https://doi.org/10.1371/journal.pone.0145474

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4). https://doi.org/http://dx.doi.org/10.1037/0022-3514.63.4.596

Barnard, P. J., Scott, S., Taylor, J., May, J., & Knightley, W. (2004). Paying attention to meaning. *Psychological Science*, *15*(3), 179–186. https://doi.org/10.1111/j.0956-7976.2004.01503006.x

Bavelas, J. B., Coates, L., & Johnson, T. (2006). *Listener responses as a collaborative process: The role of gaze sign*. *September*, 566–580.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, *113*(2), 1001–1024. https://doi.org/10.1121/1.1534836

Boersma, P. (2006). *Praat: Doing phonetics by computer.* http://www.praat.org/

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never say no . . . How the brain interprets the pregnant pause in conversation. *PLoS ONE*, *10*(12), 1–15. https://doi.org/10.1371/journal.pone.0145474

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2020). Conversational expectations get revised as response latencies unfold. *Language, Cognition and Neuroscience*, 1–14. https://doi.org/10.1080/23273798.2019.1590609

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, *5*, 1–11. https://doi.org/10.1038/srep12881

Bosch, L. Ten, Oostdijk, N., & De Ruiter, J. P. (2004). Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, *3206*, 563–570. https://doi.org/10.1007/978-3-540-30120-2_71

Boström, M. (2021). Other-initiated repair as an indicator of critical communication in ship-to-ship interaction. *Journal of Pragmatics*, *174*, 78–92. https://doi.org/10.1016/j.pragma.2021.01.007

Boyer, K. E., Phillips, R., Ha, E. Y., Wallis, M. D., Vouk, M. A., & Lester, J. C. (2009). Modeling dialogue structure with adjacency pair analysis and hidden Markov models. *NAACL-HLT 2009 - Human Language Technologies: 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Short Papers*, *June*, 49–52. https://doi.org/10.3115/1620853.1620869

Britten, N., Stevenson, F. A., Barry, C. A., Barber, N., & Bradley, C. P. (2000). Misunderstandings in prescribing decisions in general practice: Qualitative study. *Bmj*, *320*(7223), 484–488. https://doi.org/10.1136/bmj.320.7233.484

Brothers, T., & Kuperberg, G. R. (2021). Word predictability effects are linear, not logarithmic: Implications for probabilistic models of sentence comprehension. *Journal of Memory and Language*, *116*, 104174. https://doi.org/10.1016/j.jml.2020.104174

Brothers, T., Wlotko, E. W., Warnke, L., & Kuperberg, G. R. (2020). Going the Extra Mile:

Effects of Discourse Context on Two Late Positivities During Language Comprehension. *Neurobiology of Language*, *1*(1), 135–160. https://doi.org/10.1162/nol_a_00006

Cannon, J., O'Brien, A. M., Bungert, L., & Sinha, P. (2021). Prediction in Autism Spectrum Disorder: A systematic review of empirical evidence. *Autism Research*, *14*(4), 604–630. https://doi.org/10.1002/aur.2482

Caucheteux, C., Gramfort, A., & King, J.-R. (2021a). Disentangling syntax and semantics in the brain with deep networks. *International Conference on Machine Learning*, 1336–1348. http://arxiv.org/abs/2103.01620

Caucheteux, C., Gramfort, A., & King, J.-R. (2021b). Long-range and hierarchical language predictions in brains and algorithms. *ArXiv Preprint ArXiv:2111.14232*, 1–10. http://arxiv.org/abs/2111.14232

Chang, V. Y., Arora, V. M., Lev-Ari, S., D'Arcy, M., & Keysar, B. (2010). Interns overestimate the effectiveness of their hand-off communication. *Pediatrics*, *125*(3), 491–496. https://doi.org/10.1542/peds.2009-0351

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on Socially Shared Cognition.* (pp. 127–149). https://doi.org/10.1037/10096-006

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Clark, H. H., & Marshall, C. R. (1981). *Definite reference and mutual knowledge*.

Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, *2*, 19–41. https://doi.org/10.1080/01690968708406350

Clayman, S. E. (2002). Sequence and solidarity. *Advances in Group Processes*, *19*, 229–253. https://doi.org/10.1016/S0882-6145(02)19009-6

Clayman, S. E. (2013). Turn-constructional units and the transition-relevance place. *The Handbook of Conversation Analysis*, 150–166.

Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology*, *56*(5), 815–822. https://doi.org/10.1037/0022-3514.56.5.815

Czum, J. M. (2020). Dive into deep learning. *Journal of the American College of Radiology*, *17*(5), 637–638. https://doi.org/10.1016/j.jacr.2020.02.005

De Ruiter, J. P. (2013). Methodological paradigms in interaction research. In *Alignment in communication. Towards a new theory of communication* (pp. 166–199).

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*(3), 515–535. https://doi.org/10.1353/lan.2006.0130

Deppermann, A. (2015). When recipient design fails: Egocentric turn-design of instructions in driving school lessons leading to breakdowns of intersubjectivity. *Discourse and Conversation Analysis*, *16*(16), 63–101.

Dingemanse, M., Kendrick, K. H., & Enfield, N. J. (2016). A coding scheme for other-initiated repair across languages. *Open Linguistics*, *2*(1), 35–46. https://doi.org/https://doi.org/10.1515/opli-2016-0002

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R. S., Kendrick, K. H., Levinson, S. C., Manrique, E., Rossi, G., & Enfield, N. J. (2015). Universal principles in the repair of communication problems. *PLoS ONE*, *10*, 1–15. https://doi.org/10.1371/journal.pone.0136100

Drew, P. (1997). 'Open' class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics*, *28*, 69–101. https://doi.org/10.1016/S0378-2166(97)89759-7

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Enfield, N. J. (2020). Social consequences of common ground. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality* (pp. 399–430). Routledge.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327–339. https://doi.org/10.1037/0022-3514.87.3.327

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, *40*(6), 760–768. https://doi.org/10.1016/j.jesp.2004.02.002

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 540–551.

Ferreira, F., & Patson, N. D. (2007). The “good enough” approach to language comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Florian Jaeger, T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Foster, M. I., & Keane, M. T. (2019). The role of surprise in learning: Different surprising outcomes affect memorability differentially. *Topics in Cognitive Science*, *11*(1), 75–87. https://doi.org/10.1111/tops.12392

Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*, 939–944.

Galantucci, B., & Roberts, G. (2014). Do we notice when communication goes awry? An investigation of people’s sensitivity to coherence in spontaneous conversation. *PLoS ONE*, *9*(7), 1–5. https://doi.org/10.1371/journal.pone.0103182

Gevins, A., Leong, H., Smith, M. E., Le, J., & Du, R. (1995). Mapping cognitive brain function with modern high-resolution electroencephalography. *Trends in Neurosciences*, *18*(10), 429–436. https://doi.org/10.1016/0166-2236(95)94489-R

Gisladottir, R. S., Bögels, S., & Levinson, S. C. (2018). Oscillatory brain responses reflect anticipation during comprehension of speech acts in spoken dialog. *Frontiers in Human Neuroscience*, *12*, 1–13. https://doi.org/10.3389/fnhum.2018.00034

Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLoS ONE*, *10*(3), 1–24. https://doi.org/10.1371/journal.pone.0120068

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. In *Acoustics, Speech and Signal Processing, IEEE International Conference* (Vol. 1, pp. 517–520). IEEE Computer Society.

Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In spoken word recognition, the future predicts the past. *Journal of Neuroscience*, *38*(35), 7585–7599. https://doi.org/10.1523/JNEUROSCI.0065-18.2018

Hadley, L. V., Fisher, N. K., & Pickering, M. J. (2020). Listeners are better at predicting speakers similar to themselves. *Acta Psychologica*, *208*(103094). https://doi.org/10.1016/j.actpsy.2020.103094

Hale, J. (2001). A probabalistic Early parser as a psycholinguistic model. *Second Meeting of the North American Chapter of the Association for Computational Linguistics.*, 1–8.

Hamzah, H., & Fei, W. F. (2018). Miscommunication in pilot-controller interaction. *3L: Language, Linguistics, Literature*, *24*(4), 199–213. https://doi.org/10.17576/3L-2018-2404-15

Hastle, T., & Tibshirani, H. (1992). Generalized additive models. *Statistical Science*, *10*(4), 354–363.

Hebb, D. O. (1955). Drives and the C. N. S. (conceptual nervous system). *Psychological Review*, *62*(4), 243–254. https://doi.org/10.1037/h0041823

Heilbron, M., Ehinger, B., Hagoort, P., & de Lange, F. P. (2019). Tracking naturalistic linguistic predictions with deep neural language models. *Conference on Cognitive Computational Neuroscience*. https://doi.org/10.32470/ccn.2019.1096-0

Hepburn, A., & Bolden, G. B. (2013). The conversation analytic approach to transcription. In J. Sidnell & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (Issue 1, pp. 32–56). Blackwell Publishing Ltd. https://doi.org/10.1002/9781118325001.ch4

Hepburn, A., & Bolden, G. B. (2017). *Transcribing for social research*. Sage.

Heritage, J. (1986). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in Conversation Analysis* (pp. 299–345). Cambridge University Press. https://doi.org/10.1017/CBO9780511665868.020

Heritage, J. (2007). Intersubjectivity and progressivity in reference to persons (and places). In T. Stivers & N. J. Enfield (Eds.), *Person reference in interaction: Linguistic, cultural, and social perspectives*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486746

Heritage, J. (2015). Well-prefaced turns in English conversation: A conversation analytic perspective. *Journal of Pragmatics*, *88*, 88–104. https://doi.org/10.1016/j.pragma.2015.08.008

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117. https://doi.org/10.1016/0010-0277(96)81418-1

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a “standard maximum” silence of approximately one second in conversation. *Multilingual Matters*.

Jefferson, Gail. (2007). Preliminary notes on abdicated other-correction. *Journal of Pragmatics*, *39*(3), 445–461. https://doi.org/10.1016/j.pragma.2006.07.006

Johnson, C. R. (2020). Mansplaining and illocutionary force. *Feminist Philosophy Quarterly*, *6*(4). https://doi.org/10.5206/fpq/2020.4.8168

Kelly, L., & Miller-Ott, A. E. (2018). Perceived miscommunication in friends’ and romantic partners’ texted conversations. *Southern Communication Journal*, *83*(4), 267–280. https://doi.org/10.1080/1041794X.2018.1488271

Kendrick, K. H. (2015a). Other-initiated repair in English. *Open Linguistics*, *1*(1), 164–190. https://doi.org/10.2478/opli-2014-0009

Kendrick, K. H. (2015b). The intersection of turn-taking and repair: The timing of other-initiations of repair in conversation. *Frontiers in Psychology*, *6*, 1–16.

https://doi.org/10.3389/fpsyg.2015.00250
Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes*, *52*(4), 255–289. https://doi.org/10.1080/0163853X.2014.955997
Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*, *4*(1), 71–84. https://doi.org/10.1515/IP.2007.004
Keysar, B. (2008). Egocentric processes in communication and miscommunication. *Intention, Common Ground and the Egocentric Speaker-Hearer*, 277–296.
Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38. https://doi.org/10.1111/1467-9280.00211
Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science*, *13*(3), 207–212. https://doi.org/10.1111/1467-9280.00439
Kruger, J., Epley, N., Parker, J., & Ng, Z. W. (2005). Egocentrism over E-mail: Can we communicate as well as we think? *Journal of Personality and Social Psychology*, *89*(6), 925–936.
Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/https://doi.org/10.1080/23273798.2015.1102299
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657
Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, *25*(9), 1248–1258. https://doi.org/10.1093/jamia/ocy072
Lerner, G. H. (1996). Finding "face" in the preference structures of talk-in-interaction. *Social Psychology Quarterly*, *59*(4), 303–321. https://doi.org/10.2307/2787073
Levinson, S. C. (1987). *Minimization and conversational inference* (p. 61). https://doi.org/10.1075/pbcs.5.10lev
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006
Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems*, *January 2006*, 849–856. https://doi.org/10.7551/mitpress/7503.003.0111
Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *ArXiv Preprint ArXiv:1907.11692*.
MacWhinney, B., & Wagner, J. (2010). Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprachsforschung*, *1*, 154–173.
Magyari, L., De Ruiter, J. P., & Levinson, S. C. (2017). Temporal preparation for speaking in question-answer sequences. *Frontiers in Psychology*, *8*, 1–14. https://doi.org/10.3389/fpsyg.2017.00211
Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313–318. https://doi.org/10.1016/j.cognition.2012.09.010

McKenzie, C. L., & Qazi, C. J. (1983). Communication barriers in the workplace. *Business Horizons*, *26*(2), 70–72. https://doi.org/10.1016/0007-6813(83)90088-5

Mcmillan, D. (1998). *Say again? Miscommunications in air traffic control*. Queensland University of Technology.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Mertens, J., & De Ruiter, J. P. (2021). Cognitive and social delays in the initiation of conversational repair. *Dialogue & Discourse*, *12*(1), 21–44. https://doi.org/10.5210/dad.2021.102

Mustajoki, A., & Sherstinova, T. (2017). The "retrospective commenting" method for longitudinal recordings of everyday speech. *International Conference on Speech and Computer*, *September*, 710–718. https://doi.org/10.1007/978-3-319-66429-3

Paxton, A., Roche, J. M., Ibarra, A., & Tanenhaus, M. K. (2021). Predictions of miscommunication in verbal communication during collaborative joint action. *Journal of Speech, Language, and Hearing Research*, *64*(2), 613–627. https://doi.org/10.1044/2020_JSLHR-20-00137

Payne, A. (2017). *Downsizing [Film]*. Ad Hominem Enterprises.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291. https://doi.org/10.1016/j.cognition.2011.10.004

Pillet-Shore, D. (2017). Preference organization. In *Oxford Research Encyclopedia of Communication*. https://doi.org/10.1093/acrefore/9780190228613.013.132

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In *Structures of social action* (pp. 57–101). Cambridge University Press. https://doi.org/10.1017/CBO9780511665868.008

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2018). Language models are unsupervised multitask learners. *OpenAI Blog*, *1*(8).

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79. https://doi.org/10.1038/4580

Raymond, C. W. (2019). Intersubjectivity, normativity, and grammar. *Social Psychology Quarterly*, *82*(2), 182–204. https://doi.org/10.1177/0190272519850781

Roberts, G., Langstein, B., & Galantucci, B. (2016). Surprising blindness to conversational incoherence in both instant messaging and face-to-face speech. *CogSci*, 1211–1216.

Roberts, S. G., Torreira, F., & Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: A corpus study. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00509

Sacks, H., & Schegloff, E. A. (1979). Two preferences in the organization of reference to persons in conversation and their interaction. *Everyday Language: Studies in Ethnomethodology*, 15–21. https://doi.org/10.2307/2066919

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Linguistic Society of America*, *50*(4), 696–735. https://doi.org/10.1353/lan.1974.0010

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, *23*(3), 514–523. https://doi.org/10.1162/jocn.2009.21370

Savitsky, K., Keysar, B., Epley, N., Carter, T., & Swanson, A. (2011). The closeness-communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology*, *47*(1), 269–273. https://doi.org/10.1016/j.jesp.2010.09.005

Sayer, I. M. (2013). Misunderstanding and language comprehension. *Procedia - Social and Behavioral Sciences*, *70*, 738–748. https://doi.org/10.1016/j.sbspro.2013.01.118

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. *Georgetown University Roundtable on Languages and Linguistics*.

Schegloff, E. A. (1987). Some sources of misunderstanding in talk-in-interaction. In *Linguistics* (pp. 201–218). https://doi.org/10.1515/ling.1987.25.1.201

Schegloff, E. A. (1988). Presequences and indirection: Applying speech act theory to ordinary conversation. *Journal of Pragmatics*, *12*, 55–62. https://doi.org/10.1016/0378-2166(88)90019-7

Schegloff, E. A. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, *97*(5), 1295–1345. https://doi.org/10.1086/229903

Schegloff, E. A. (1993). Reflections on quantification in the study of conversation. *Research on Language and Social Interaction*, *26*(1), 99–128. https://doi.org/10.1207/s15327973rlsi2601_5

Schegloff, E. A. (1995). Sequence organization. *Manuscript*.

Schegloff, E. A. (2007). Sequence organization in interaction: A primer in conversation analysis. In *Sequence Organization in Interaction: A Primer in Conversation Analysis I* (Issue January 2007). https://doi.org/10.1017/CBO9780511791208

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361. https://doi.org/10.2307/413107

Schober, M. F., & Carstensen, L. L. (2010). Does being together for years help comprehension? In *Expressing oneself/expressing one's self: Communication cognition, language and identity* (pp. 107–124).

Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., & Fedorenko, E. (2020). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, *138*, 107307. https://doi.org/10.1016/j.neuropsychologia.2019.107307

Shannon, C. E., & Weaver, W. (1964). *The mathematical theory of communication*. The University of Illinois Press. https://doi.org/10.4992/jjpsy.25.110

Siddiqi, A. A. (2018). Beyond Earth: A chronical of deep space exploration, 1958 - 2016. In *NASA History Program Office*.

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., Diamond, S. P., & Held, R. M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences*, *111*(42), 15220–15225. https://doi.org/10.1073/pnas.1416797111

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

Stivers, T., & Robinson, J. D. (2006). A preference for progressivity in interaction. *Language in Society*, *35*, 367–392. https://doi.org/10.1017/S0047404506060179

Svennevig, J. (2008). Trying the easiest solution first in other-initiation of repair. *Journal of Pragmatics*, *40*, 333–348. https://doi.org/10.1016/j.pragma.2007.11.007

Umair, M., Mertens, J., Albert, S., & Ruiter, J. P. de. (2022). GailBot: An automatic transcription system for Conversation Analysis. *Dialogue & Discourse*, *13*(1), 1–23. https://doi.org/10.5210/dad.2022.103

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *Nips*, 5998–6008. http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf

Veale, T. (2004). Incongruity in humor: Root cause or epiphenomenon? *Humor*, *17*(4), 419–428. https://doi.org/10.1515/humr.2004.17.4.419

Volden, J. (2004). Conversational repair in speakers with Autism Spectrum Disorder. *International Journal of Language and Communication Disorders*, *39*(2), 171–189. https://doi.org/10.1080/13682820410001663252

Warnke, L. (2022). *Speech act prediction across turn boundaries in conversation*. Tufts University.

Wilcox, E. G., Gauthier, J., Hu, J., Qian, P., & Levy, R. P. (2020). On the predictive power of neural language models for human real-time comprehension behavior. *ArXiv*, *2019*.

Wood, S. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall/CRC.

Yngve, V. (1970). On getting a word in edgewise. *Chicago Linguistics Society, 6th Meeting*, 567–578. https://doi.org/10.1215/9780822399896-005

Appendix A

Jeffersonian transcription symbols used in this paper

Symbol	Example	Meaning
_:	l:ive	Moderately higher pitch
(#.#)	(0.4)	Silence in seconds
:	camp⌈us:?⌉	Stretched sound
⌈ ⌉	camp⌈us:?⌉	Overlapping talk
⌊ ⌋	⌊yeah.⌋	
,	l:ive,	Flat turn-final intonation
?	camp⌈us:?⌉	Rising turn-final intonation
.	⌊yeah.⌋	Falling turn-final intonation
∇ ∇	∇I have no∇	Slower speech
↑	↑plans:	Significantly higher pitch
≈	weekend.≈	Latching – little to no gap
≈	≈oh	between speakers
_	did	Emphasis

Note: This is a subset of all the Jeffersonian transcription symbols. For a more complete list and description, see Hepburn and Bolden (2013, 2017).