2023

What Makes Conversation Good? How Responsivity, Topics, and Insider Language Predict Feelings of Connection

Emma M. Templeton
Dartmouth College, emma.m.templeton.gr@dartmouth.edu

WHAT MAKES CONVERSATION GOOD?
HOW RESPONSIVITY, TOPICS, AND INSIDER LANGUAGE PREDICT FEELINGS OF CONNECTION

A Thesis
Submitted to the Faculty
in partial fulfillment of the requirements for the
degree of

Doctor of Philosophy

in

Psychological and Brain Sciences

by Emma Templeton

Guarini School of Graduate and Advanced Studies
Dartmouth College
Hanover, New Hampshire

April 2023

Examining Committee:

(chair) Thalia Wheatley, Ph.D.

Luke Chang, Ph.D.

Arjen Stolk, Ph.D.

Riccardo Fusaroli, Ph.D.

F. Jon Kull, Ph.D.
Dean of the Guarini School of Graduate and Advanced Studies

Abstract

We spend our lives having conversations, and some inevitably go better than others. What happens in conversation that makes people feel connected? To explore this question, I recorded pairs of strangers and friends having unstructured conversations. In Chapter 1, I show that people who feel connected tend to respond quickly, creating short gaps between turns. However, long gaps are not necessarily bad. Although long gaps signal moments of disconnection and awkwardness for strangers, they mark moments of heightened connection for friends by providing space for enjoyment and mutual reflection. In Chapter 2, I examine how people start their conversations. Specifically, how do people who have never met before initiate their first interaction? And how do these approaches differ from people who are already robustly connected? I find that strangers start their conversations more similarly to each other, compared to friends. In particular, strangers tend to start with topics that can easily launch into many different topics, increasing the likelihood of finding common ground. Friends do not need to rely on this strategy and can instead immediately start their conversations with topics idiosyncratic to their relationship. In Chapter 3, I highlight another fundamental difference in how friends and strangers communicate by exploring the use of *insider language*, or words carrying specific meaning between some people but not others. I find that friends use insider language more than strangers and when they do, they feel more connected. When people know each other well, communication can move from spoken words to shared thoughts. Together, these findings reveal that people feel closer when they can respond quickly in conversation *and* feel comfortable not speaking, and that being able to jump right into a conversation and communicate using shorthand are hallmarks of friendship.

Acknowledgements

I got to work with my dream advising team—**Thalia Wheatley** and **Luke Chang**—in my dream department. I don't think it's an exaggeration to say that this dissertation work could not have been done anywhere else, with anyone else.

My first conversation with Thalia was amazing. We had a Skype call (remember those?) about my research interests. It was so fun! We bounced ideas around so easily and I had the sense that she really *got* me. For weeks afterward, I found my thoughts drift back to how much fun I had talking to her. Perhaps fittingly, I made my grad school decision based on one good conversation.

Luckily for me, that one conversation was not anomalous. Thalia is one of my favorite people to talk to. She is brilliant and inspiring and so, so funny. I have so many happy memories of truly transformative research meetings in her office, in Hanover, and during car rides to / from Boston—always complete with a McDonalds stop (my favorite), even though Burger King is her favorite.

I've loved being part of Thalia's lab. She makes doing science fun and fosters a really fantastic community to do science in. As an example, the lab went through a phase of weekly rock climbing followed by Worthy Kitchen and then pivoted to weekly dancing followed by Jewel of India (followed by more dancing). One summer, the lab flew to Budapest together and spent a week exploring, bonding, and playing word games. In the COVID era, Thalia organized some of the best picnics I've ever seen. Thalia is incredibly generous with her lab and always makes the time we spend together feel extra special.

Thalia is also a fantastic advisor! The conversation projects I pursued for my dissertation were intentionally open-ended. While this is exciting, it is also overwhelming. When a project

could go in a *million* different directions, it can be near impossible to pick *one*. This is where Thalia comes in. She has a nose for what questions are worth pursuing and what other people will think is interesting. Whenever I present analyses or results, she never fails to magically pluck out a brilliant thought or observation that never occurred to me. I spend so much time in the weeds with my data and am so lucky that Thalia is there to keep the 10,000-foot perspective. That perspective has been critical for understanding how to direct my analyses, frame my results, and plan for what comes next. As a bonus, Thalia also gives some of the best talks out there. She tells simple stories with bold messages that leave people feeling excited and energized to get to work. It has been an absolute joy to learn from her.

While Thalia is keeping the big picture in mind, Luke is in the data weeds with me, acting as my guide. When I started graduate school, I felt woefully out of my depth. I didn't feel smart enough to be in a department where everyone seemed to be effortlessly using the fanciest, cutting edge methods. In my very first meeting with Luke, all of this came out. He kept excitedly showing me different heatmaps to illustrate ongoing work in his lab, but all that was being illustrated to me was the fact that I had no idea how to make a heatmap! I was so overwhelmed that I started crying. Luke bought me an iced tea, told me to read "The importance of stupidity in scientific research," and generally calmed me down. He insisted that I would learn how to do really hard things and after I learned them they would feel easy. It felt impossible to believe, but he was right. Thanks, in large part, to his mentorship.

Working with Luke has made me become a better and more rigorous scientist. I've really honed my skills in thinking through how to answer a question and then build confidence around that answer. I've learned from Luke how to interrogate my data, perform robustness checks, think through alternative explanations, and interpret results. Luke is the rare advisor who can sit next to you during a meeting and just start coding a solution on your laptop. Every time he does this, it is shocking and then humbling and ultimately so helpful.

One of the things I appreciate most about Luke is that he always takes time to point out when I've gotten better at a particular research skill. Because he has worked through many of the same challenges in his own work, he has a real appreciation for just how difficult certain tasks are. So, when he notices that I've implemented something quickly or developed a good intuition about something, he points it out and explains exactly why I should feel proud. Luke always makes me stop and look back to see how far I've traveled. It is a skill that I try to practice on my own now, too.

His data analysis skills aside, I just *like* Luke! He always takes the time to check in with me as a person by asking how my relationships are going, talking about new restaurants in the area, and exchanging gossip. After so many hours of meetings, we've developed really efficient ways to communicate when we think through difficult analysis problems together. I find that I come up with ideas in Luke's presence that I can't come up with alone. It's the perfect example of conversation working well, by allowing two minds to produce something better than either mind can produce alone.

I was fortunate to get to meet with two other faculty members as part of my dissertation committee: **Arjen Stolk** and **Riccardo Fusaroli**. I remember watching Arjen's job talk and being blown-away by the connections between his communication work and the goals of my own work. I was thrilled when he joined our department. Every time I interact with Arjen, I learn something new and find myself scribbling down relevant citations. Arjen's knowledge of this field is so deep, and I look forward to learning more from him!

The magic of the (pre-Elon) Twitter algorithms introduced me to Riccardo. In addition to his own work being amazing, I noticed that he kept tweeting about all these interesting papers that weren't showing up anywhere else on my timeline. When my first paper came out, I couldn't believe my surprise when Riccardo reached out to me with one of the kindest emails I've ever received. He praised our paper, asked a series of super thoughtful questions about our work, and even included some unpublished results that he thought I might be interested in seeing. What a seriously gracious thing to do! I knew I wanted to have him on my committee and, boy, did that turn out to be a good decision. Riccardo is always the first to reply to my emails and is always so positive and encouraging. It has been an absolute joy to have him on my committee and I hope we can continue talking about research!

Now that I've thanked my dissertation committee, it's time to thank the members of the Wheatley Lab and Chang Lab. First up is my home base: **The Wheatlab**. Every iteration of the lab has been amazing. **Sophie Wohltjen** deserves an extra-special shout out. We started in the lab at the same time and went through this whole process together, bonding us for life (I hope!). Sophie is one of the few people I will cold call on the phone and she is always there to talk something through. I'm so grateful for her friendship, her wisdom, and her spirit. I could write paragraphs about every lab member, but I'm already feeling self-conscious about how long these acknowledgements are getting. So my compromise will be to list everyone I've worked with and share one thing about each of them. Here we go. **Beau Sievers** was my older lab brother and my main point of contact for how the lab and the department worked. **Laurel Symes** hosted the best science salons at her place during my early years of grad school. **Adam Boncz** was the happiest presence in the lab and facilitated an unforgettable lab trip to Central Europe. **Chris Welker** is my favorite conference buddy and hype man. **Adrienne Wood** does the most fun research with the most laughter (pun intended). **Kelly Finn** and I bonded over a project we worked on together during the pandemic years. **JD Knotts** is from Southern California like me, so we understand each other at a deep and fundamental level. **Caitlyn Lee** taught me how to fish (literally) and always tells me about her cool science ideas. As someone once wrote in the fMRI control room: Wheatlab Rulz!

The Chang Lab has always felt like a home away from home in Moore Hall. There have been so many members of the Chang Lab during my seven years of grad school, so I'm not going to list everyone's name alongside one memory (though all past and present members may request one offline). It's hard to get the Chang Lab out of the lab, but when they do leave it tends to be for epic adventures like Boda Borg in 2017, the Funspot arcade in 2019, and Ice Castles in 2022. Weekly lab meetings are equally epic. They go very (very) long, but are also filled with lively discussions, lots of laughs, and tons of useful information. Now seems like a good time to shout out Luke's wife, **Eunice Lee**, who contributes so much delicious food and much needed sophistication to so many Chang Lab gatherings.

I've had the fantastic opportunity to advise two honors thesis projects with two brilliant undergraduate students: **Eli Reynolds** and **Marie Cone LeBeaumont**. Eli collected many of the round-robin conversations (and handled the scheduling logistics like a pro!). Marie collected many of the friend conversations as well as the zoom conversations that appear in Chapter 3. I'm so grateful for the pivotal role they played in data collection for this dissertation and proud of them for winning thesis awards for their excellent work. I've been fortunate to work with many wonderful undergraduate research assistants and want to recognize the people who have contributed to results presented in this dissertation. **Emily Chen**, **Halla Hafermann**, and **Darley Sackitey** made ratings on long gaps in Chapter 1b. **Eliot Aiman**, **Grace Cason**, **Catherine**

Gorman, **Dawson Haddox**, and **Jonathan Lim** made insider language annotations for analyses in Chapter 3. Eliot and Dawson also created and coded insider language categories for a content analysis in Chapter 3. That was a *ton* of work that took much more careful attention and discussion than you would think!

I feel so lucky to have received strong mentorship in Psychology research even before graduate school. As an undergraduate student, I worked as a research assistant in **Jason Mitchell**'s lab for four years. The graduate student I worked with, **Diana Tamir**, taught me everything I know about how to conduct research. I still set up spreadsheets to run participants just as she taught me in 2010. I got to continue working with Diana through her postdoc and have loved being part of her life ever since—cheering her on as she set up her own lab, got married, and became a mother! I can always count on Diana for great advice, an infectious enthusiasm for science, and a willingness to walk to tacos. While working in that lab, I met a couple people that I ended up seeing again at Dartmouth. **Eshin Jolly** was the lab manager during that time and I looked up to him so much (and still do). When I saw him at Dartmouth's interview weekend, I felt instantly calm. I'm so glad I got the chance to become actual friends with one of my science idols. I've learned so much from Eshin, including which Netflix reality TV show I need to be watching next. Miraculously, I also got to meet **Mark Thornton** when he was a first year graduate student! He is now a professor at Dartmouth where he is a steadfast and comforting presence. After undergrad, I was a lab manager in **Jamil Zaki**'s lab for 2 years. Jamil was the best boss for the best job I've ever had. I learned so much about how to run a lab from him and I miss working with him every day! I met some of my favorite people while working in the SSNL including: **Erik Nook**, **Erika Weisz**, **Brent Hughes**, and **Yuan Chang Leong**. I'm

such a proud alumna of the SSNL. Conferences feel so much fuller and brighter by reconnecting with old lab members and meeting new ones.

For the past five years, I’ve been a Resident Fellow in **East Wheelock House**. This means that I live in an undergraduate dorm room, eat in the dining hall, and program activities for students. Doing this is literally a dream come true. Working with the Undergraduate Assistants (UGAs) is easily my favorite part of the job. They make me believe that the future is in good hands. I’m especially grateful to my boss **Josiah Proietti** for modeling great leadership and for always supporting me and advocating for me. I loved getting to work with our House Professor, **Sergi Elizalde**. Getting to know him and his family made the House truly feel like a Home. Thanks also to the fantastic admin staff in East Wheelock, **Lorrie Bosse** and **Bonghee Lis**, who always made me hot beverages and welcomed me into their office for long chats.

During dissertation crunch time, it was more important than ever to get in some physical activity. I didn’t expect to make so many friends in the process! I joined **The Studio** in White River Junction and left every class absolutely dripping in sweat, with all my thesis anxiety behind me. I’m so thankful for the wonderful community at The Studio. Case-in-point: while there, I met **Andy Friedland** who graciously let me use his Dartmouth Emeritus office for my final few months of dissertation writing. On Sundays, I played hours of volleyball with **DartSlam**, a graduate student volleyball club. This was the absolute highlight of my week, every week. I credit **Francois LeSage** for creating this group and turning it into quite the force on campus!

Thank you to my parents! My **mom** instilled in me an interest in human behavior at a young age and always takes the time to ask me lots of questions about my work. My **dad** routinely sends supportive messages and lots of treats, including fancy tea that brightens my

afternoons. They each take amazing care of me whenever I visit—I can always count on lots of delicious meals, drinks, and some serious TV time. When I'm not visiting, I can count on them to pick up their phones when I'm walking (I've done this so many times that they pick up and instinctually ask, "Where are you walking now?").

My **grandma** passed away during my last year of graduate school. She was my family on the East Coast up until then. I was so lucky to spend lots of Thanksgiving and Spring Breaks with her. She lived in a residential community that I've long described as "college for old people." It was such a fun place to visit. My grandma deeply understood the importance of being social, especially into old age. I saw first-hand how she made herself an integral member of that community and formed truly great friendships. She taught me my favorite saying ("a good cook cleans as they go"), bought me my first winter coat, fed me so many cookies and glasses of wine, and talked smack over rounds and rounds of Dominoes. My grandma must have been on my mind a lot as I worked on this dissertation. Only upon reading it all together did I realize just how many times I referenced her in my chapters. She would have been so proud that I finally finished this dissertation. I'm glad I got to include her.

And finally thanks to my partner, **Jason Hirschhorn**. I met Jason during my first week of grad school and he has made my life fuller and happier ever since. When Jason moved to the area and started teaching at Hanover High School, my world in the Upper Valley expanded. It was thrilling and comforting to get to know people outside of Dartmouth PBS. We suddenly got the best recommendations for seasonal events, invitations to eat tacos on porches, and so many house-sitting offers from people horrified to learn that we were living in a 350 square-foot dorm with no kitchen. Jason also expanded my world more generally. His family welcomed me onto their trips and into their group chat (complete with daily Wordle scores) and have been a source

of love and support throughout graduate school. Every day with Jason feels like an adventure! That’s not just a comment on his personality, he is literally always planning adventures for us. With an emphasis on *planning*. Jason and I have always gotten along so well because we are both incredibly organized, with impeccable email hygiene, thorough use of google calendars, and total bewilderment at the fact that some people are disorganized. Jason kicked his planning into overdrive during my final year of grad school. When I was feeling too overwhelmed to contemplate a possible defense date, Jason pulled out a calendar, weighed the many different factors, and picked the goal date for me. Public school teachers do not get a lot of time to rest, and Jason spent all of those precious moments preparing for my defense celebrations. He made a wedding website for “Emma & a PhD” to share information about the process and to collect RSVPs which was both hilarious and practical. I wish every PhD student had a Jason to plan their defense party. It was a gift to purely focus on my work, knowing that Jason had everything else taken care of. This example is obviously top-of-mind, but it is just one of many, many examples of Jason tangibly supporting me while also injecting plenty of whimsy, laughter, and joy along the way. Thank you, Jason, for helping me get here.

Table of Contents

List of Illustrations

List of Tables

General Introduction

When I first started graduate school, the question I wanted to answer was: *What makes conversation good?* At the time, I was embarrassed not to have a more "scientific" formulation of that question. I assumed that as I learned more and progressed as a scientist, I would develop a "smarter" way to phrase what I was interested in. Now, I embrace the question! I think it easily communicates the feeling that I am trying to capture in my research—the feeling of being in the midst of a really great conversation and realizing how much *fun* you are having and how effortlessly you and your conversation partner are building on each other. Those moments can happen with a stranger who you'll never see again, and they can happen with an old friend who you see multiple times a week. For me, there is no better feeling in the world than a great conversation.

In addition to being fun, it turns out that good conversations are also extremely important for maintaining good physical and mental health. Conversation is the primary way that we navigate our social world. It is how we form and maintain relationships, share information, and manage reputations (R. I. Dunbar et al., 1997; Emler, 1990). Social isolation is a risk-factor for early mortality, on par with the dangers of smoking and obesity (Holt-Lunstad et al., 2015). An entire body of research shows that people do not engage in as many conversations as they should (Epley et al., 2022; Kardas et al., 2022). And when people are instructed to have more conversations, they are happier because of it (Epley & Schroeder, 2014; Sandstrom & Boothby, 2021).

Of course, not all conversations are the same. You can talk to someone and click right away, potentially laying the foundation for a life-long friendship. You can also talk to someone and have an absolutely terrible time, potentially ruining the rest of your day. Other times, you can talk to someone and feel perfectly fine, neither particularly good nor particularly bad. What happens in a conversation that determines this outcome? This is what I set out to investigate in my dissertation.

Existing theoretical frameworks

Conversation can be conceptualized in many ways including as a means to share information, as a way to reach agreement, or as a way to feel connected. These different conceptualizations have inspired different theoretical frameworks that have, in turn, generated different insights about why and how we converse. In this section, I will review major theories that have guided conversation research.

According to information theory (Shannon, 1948), communication is a practice in reducing uncertainty. One person (a sender) transmits information to another person (a receiver) using language. If the sender and receiver understand the same language, the transmission of information should be successful and, as a consequence, reduce uncertainty. This framework has been used to study interactions between animals, humans, and even machines. However, it is easy to see how this simple picture gets complicated by common conversation behaviors. First, this theory assumes that the sender and receiver understand language in exactly the same way. In reality, words do not always have a 1:1 mapping and can instead take on different meanings depending on how they are being used and with whom (Brennan & Clark, 1996; Kull, 2020). Second, this account does not consider a speaker's *intentions*. Sometimes, the words that people say are not enough to decode their meaning; it is necessary to understand why they said it, how they said it, and what their goals in saying it were. This is a core tenet of Grice’s work on conversation (P. Grice, 1989), which emphasized the need to decide whether a signal was intended to communicate something in particular.

Later theoretical frameworks conceptualized conversation as a *joint activity*, where two people have the goal of getting into mental alignment with each other. In this way, dialogue is fundamentally different from monologue (Brennan et al., 2010). People in conversation influence each other. Many accounts agree with this basic premise, but disagree about the mechanism. One camp contends that the process of getting in alignment happens through a largely automatic process of priming. When someone uses a particular set of words or speaks in a particular way, their partner notices and incorporates that word choice and delivery into their speech as well. As a result, these choices spur changes in related behaviors that ultimately result in a shared mental representation between two people. Common accommodation behaviors in conversation, where people mutually adapt their body movements, language, or voice over the course of an interaction are taken as evidence of this view. These accounts do not require that people actively think about the mind of their conversation partner; doing so is cognitively demanding and would interrupt the natural flow of conversation that we so often experience. In these accounts, conversation is a fluid, easy process guided by automatic imitation (Garrod & Pickering, 2004). It is only in those clumsy (and infrequent) moments when conversational repair is needed that people need to consciously mentalize about their partner. Theories that propose this general structure go by different names, including interacting alignment (Pickering & Garrod, 2004), monitoring and adjustment (Horton & Keysar, 1996), and perspective adjustment (Keysar et al., 1998). A critique of this general framework is that two people in conversation need to be very similar to each other (if not direct copies of each other) for all this to work so smoothly.

A different camp emphasizes the need to consciously consider what your conversation partner is thinking and actively work together to achieve mental alignment. Conversation is not only a joint activity, it is a *collaboration* (H. H. Clark, 1992). This often plays out as a process referred to as *grounding* (H. H. Clark & Wilkes-Gibbs, 1986). Grounding is a series of back-and-forth check-ins between conversation partners that ensure concepts are being used in the same way (H. H. Clark & Schaefer, 1989). Once a concept has been grounded in this way, people can continue using it easily (Krauss & Weinheimer, 1964). Notably, grounding is partner-specific (R. D. Hawkins et al., 2021; Metzing & Brennan, 2003; Wilkes-Gibbs & Clark, 1992). If someone new were to enter the conversation, the concept would have to be grounded one again (though it may be abandoned altogether). This framework emphasizes the importance of finding and considering common ground in a conversation (Brennan et al., 2010; H. H. Clark & Brennan, 1991). A related account, proposed by Fusaroli and colleagues (Fusaroli & Tylén, 2016), de-emphasizes the route to achieving mental alignment (automatic priming vs consciously searching for common ground) and instead proposes that what matters most is the process of building synergy between two people. This has the advantage of reducing the complexity of the system overall.

A recent paper by Stolk and colleagues combines many of these accounts in acknowledgement of the fact that different frameworks can each be relevant in different contexts (Stolk et al., 2020). Specifically, they describe conversation in terms of three types of signals. Shannon-signals (a nod to Shannon's information theory; (Shannon, 1948)) act to reduce uncertainty and are at play when a sender and receiver can be assumed to have the same system for encoding and decoding information. Grice-signals (a nod to Grice's emphasis on speaker intention; (H. P. Grice, 1957)) are intended to induce changes that a speaker hopes to see in a receiver. Peirce-signals (a nod to Peirce's emphasis on meaning-making; (Pierce, 1931)) refers to the interpretation people make in reaction to an input. The authors propose that all of these signals work together to align *conceptual spaces* between people in conversation. These conceptual spaces are constructed by the members of the conversation and shaped by the ongoing interaction. New and exciting findings from brain imaging studies provide support for this framework (Stolk et al., 2020).

Other theoretical accounts focus on the subjective experience of being in a conversation. For example, Shared Reality Theory describes the motivations that people have to look for evidence that they have something in common with whom they are speaking (Echterhoff et al., 2009; Hardin & Higgins, 1996). This can be a powerful way to understand conversation. The emphasis on how connected people *feel* is a helpful reminder that conversation is more than an exercise of decoding signals and aligning mental states. People want to believe they are “on the same wavelength” and viewing the world in a similar way as others. Evidence of this can come out in the language that people use (“That’s exactly what I was thinking too!”) or in responses to survey items specifically designed to measure perceptions of shared reality (Rossignac-Milon et al., 2020).

Other theoretical approaches focus on how to *quantify* conversation. The Rational Speech Act framework (Frank & Goodman, 2012; Goodman & Frank, 2016) uses Bayesian models to formalize the need to infer intention from a speaker and to iteratively coordinate on creating shared meaning. This makes it possible to quantify and explain exactly how people incorporate feedback to develop shared understandings. Modeling an interaction makes it possible to see how the words that one person says changes beliefs in another person’s mind. Currently, using this approach necessitates using highly constrained paradigms that make some unrealistic assumptions (e.g., people perfectly form their thoughts before speaking) but it is easy to see how these models will be able to get more complex and sophisticated over time (Degen, 2023). Another such framework is the Social Relations Model (Kenny & La Voie, 1984) where conversation behaviors can be explained by the person of interest (the actor), the person they are talking to (the partner), and the unique combination of the actor and partner (the dyad). When someone smiles a lot in a particular conversation that behavior can be explained by a combination of actor effects (they tend to be a smiley person), partner effects (their partner tends to make people smile), or dyad effect (the combination of this actor and partner produces lots of smiles). At the heart of this approach is the understanding that conversations are not isolated events, but rather embedded in an individual's history, personality, and tendencies.

Certain theoretical approaches inspire particular ways of studying conversation, just as particular ways of studying conversation lend evidence to certain theories.

How has conversation been studied?

Conversation can be studied in many different ways. These different approaches have their own sets of advantages and disadvantages and are suited for different types of questions. In this section, I will review the various ways that researchers have studied conversation in prior work.

Eavesdropping. In 1992, Henry Moore spent several weeks walking up and down Broadway between Thirty-third Street and Fifty-fifth Street in New York City. As he walked, Moore wrote down fragments of all the conversations he was able to overhear and noted the gender composition of each group. Most fragments were only a few seconds long (NYC is loud and people walk fast!) but Moore kept walking until he collected 174 of them. He later categorized each fragment by topic to make claims about how the “original natures” of women differed from that of men (Moore, 1922). The motivation behind Moore’s original study comes across as undeniably sexist now and later replication studies suggest these differences are not as stark as he believed (Bischoping, 1993; R. I. Dunbar et al., 1997). Moreover, eavesdropping as a scientific technique became much more rigorous. A recent paper put forward a well-specified taxonomy that details how to unobtrusively code social interactions as they unfold (Mulwa & Kucker, 2022). This coding scheme was used to compare adult-adult and child-caregiver conversations in natural settings. Because participants do not know they are part of a research study, eavesdropping studies are typically used to compare conversations between categories that researchers think can be reasonably inferred, like gender or age. Sometimes, participants know that their conversations will be recorded, they just don't know exactly *when*. Studies using the Electronically Activated Recorder (EAR) device code audio recordings of conversations periodically sampled throughout participants' days (Mehl et al., 2001). When paired with self-report data, researchers can make claims about how the amount and types of conversations people have relate to their well-being (Mehl et al., 2010).

Passive Sensing. Passive sensing technology allows social interaction data to be collected much more frequently and at a larger scale. Participants are either given wearable devices (Eagle & (Sandy) Pentland, 2006; Onnela et al., 2014) or asked to download an app that runs in the background of their personal smartphones (Harari et al., 2017; Wang et al., 2014). This allows for many different variables to be collected simultaneously, over the span of several hours or several years. For ethical (and file storage) reasons, individual conversations are not typically recorded. However, meta-information about conversations can be stored. This includes when people are engaged in conversation, how long conversations tend to last, and the estimated group size. Researchers can then examine how these variables fluctuate over time (e.g., how conversation frequency changes over an academic term (Harari et al., 2020)) and how these variables fluctuate with each other (e.g., how stress on one day predicts conversation behavior on a future day (daSilva et al., 2021)).

Conversation Analysis. Conversation Analysis (Goodwin & Heritage, 1990) is a qualitative approach to studying conversation that typically takes a handful of conversation recordings and makes extremely detailed transcriptions and annotations about exactly what was said, how it was said, and what people were doing while they were talking (Hepburn & Bolden, 2012; Mondada, 2016). Researchers then look for patterns in these annotations to put forward claims about how interactions work. This careful, in-depth analysis has generated many important insights about turn-taking (Sacks et al., 1974), adjacency pairs (Schegloff & Sacks, 1973), and repair (Schegloff et al., 1977) in conversation. The approach gained prominence in the 1960s in sociology and is still alive and well today. People apply Conversation Analysis to many different settings, including doctor-patient interactions and high-profile interviews (Antaki, 2011). Of course, this detailed approach limits the scalability of the work. It is a tedious process to apply Conversation Analysis on a conversation and most work draws conclusions from a small number of examples, sometimes focusing entirely on a single conversation. However, insights from this careful work inspires work in other fields that can formalize and test specific claims in larger datasets.

Big Data. With the rise of social media and increased access to the Internet, the number of conversations recorded online has reached a scale that can be analyzed using "big data" approaches (Fan et al., 2014). For example, researchers quantify the influence of different accounts on Twitter (Bakshy et al., 2011), track the rise of political polarization (Van Bavel et al., 2021), and graph social media use (Brambilla et al., 2022). These big datasets can also train machine learning models to automatically extract information about language use, facial expression, and body movement in face-to-face conversations, yielding unprecedented opportunities to investigate questions about human behavior and to build tools that can be leveraged in other domains. That said, it is important to keep in mind that the kinds of conversations in Reddit communities, Facebook comments, and Twitter replies may be quite different from the kinds of conversations that happen face-to-face. Online conversations can be anonymous, asynchronous, and fraught with self-presentational concerns. Even with the increased opportunity to use social media to build more connections, research shows that the number of close friends we have and maintain has not increased (R. I. M. Dunbar, 2018) and that face-to-face communication is quite different from other types (Drijvers & Holler, 2022).

Communication Games. Conversation can also be studied by designing games that strip communication down to its essential components. The field of experimental semiotics is known for using games to examine how people develop novel forms of communication (Galantucci & Garrod, 2011). Participants always have a specific task to complete with a partner. Sometimes they solve that task by talking to each other (Fusaroli et al., 2012; R. X. D. Hawkins et al., 2017) and other times they are required to communicate without language, instead by manipulating tokens on a screen (Galantucci, 2005; Scott-Phillips et al., 2009; Stolk et al., 2014). Because these tasks are so constrained, researchers are able to model participant behavior in elegant and rigorous ways. Using communication games allow for precise predictions that can be tested and then built upon. Of course, constraining behavior so much can limit the generalizability of results beyond the confines of that particular communication game. Further, the way people behave in a game may be entirely different from how they behave outside of the lab (Levitt & List, 2007) and may be overly simplistic (Jolly & Chang, 2019).

Brain recordings. Conversation can also be studied using brain imaging. For example, fMRI studies find that brain activation when a speaker tells a story is similar to brain activation in a listener when they listen to that same story later on (Zadbood et al., 2017) and the degree of this coupling relates to comprehension (Stephens et al., 2010). Eye-tracking studies show that speakers and listeners have increased pupillary synchrony during salient moments of a story (Kang & Wheatley, 2017). This early work used study paradigms that separated speakers from listeners. However, newer *hyperscanning* methods record brain activity from two people at a time, permitting study designs that allow participants to interact with each other in real-time. Dual EEG and fNIRS paradigms have been used to show that neural synchrony is increased when people are physically oriented towards each other vs not (Drijvers & Holler, 2022; Jiang et al., 2012). A recent study using eye-tracking glasses measured pupillary synchrony between dyads as they engaged in unstructured conversations, finding that eye contact marks the rise and fall of shared attention (Wohltjen & Wheatley, 2021). Datasets from hyperscanning fMRI conversation tasks are currently being analyzed to further examine brain dynamics during conversation (Boncz, 2019; Tsoi et al., 2022). As is always the case with brain imaging, certain constraints need to be in place to deal with spatial and temporal resolution limitations of different technologies. Right now, brain activity recorded during totally unstructured conversation would be difficult to analyze and interpret.

Conversation Interventions. Another way to study conversation is to intervene on some dimension and then examine how that intervention impacts an outcome variable. For example, one study asked participants to have conversations with either lots of question-asking or limited amounts of question-asking, finding that participants who ask more questions are better liked by their conversation partners (Huang et al., 2017). Another study asked participants to vary their speaking time and found that participants who spoke more were more likable (Hirschi et al., 2022). A third study randomly assigned participants to engage in different *types* of conversations, finding that those asked to have more higher-quality conversations benefited from increased well-being (Hall et al., 2023). In other studies, *having a conversation* is the intervention. Commuters were randomly assigned to either connect with a stranger on public transport, keep to themselves, or commute as normal. Participants who initiated conversations reported a more positive commute experience (Epley & Schroeder, 2014). Other lines of work instruct people to have conversations and then collect impressions of those conversations, including when participants wanted them to end (Mastroianni et al., 2021) and discrepancies between how much participants liked their study partner and how much they believed their study partner liked them (Boothby et al., 2018). In these types of studies, the *impact* of the conversation is of primary interest. The conversation itself is often not even recorded or analyzed.

Naturalistic Studies. Taking a naturalistic approach to studying conversation means recording interactions with minimal intervention. This approach has been employed with speed-dates (McFarland et al., 2013), patient-therapist sessions (Ramseyer & Tschacher, 2011), negotiations (Curhan & Pentland, 2007), video calls (Reece et al., 2023), and strangers left alone in a waiting room (Cuperman & Ickes, 2009). Data from naturalistic study designs tend to get used repeatedly for many different purposes. One of the most well-known examples is the Switchboard Corpus (Godfrey & Holliman, 1997), a large collection of telephone conversations. Researchers have used this corpus to investigate linguistic convergence (Cohen Priva & Sanker, 2020), turn-taking (Pouw & Holler, 2022), speech rate (Cohen Priva, 2017), and word predictability (Shriberg & Stolcke, 1996), to name a few (see (Serban et al., 2015) for an overview of available corpora). There is so much rich information to be analyzed in natural conversation. Reusing the same corpus to study different questions is an efficient use of resources. However, relying on a single corpus for insights can be misleading. Conversations have different contexts and goals that can impact their dynamics. It is important to keep this in mind when collecting naturalistic data because decisions about what gets recorded will constrain future analyses and interpretations of results.

This dissertation

Although conversation is ubiquitous and important, it is difficult to study (H. H. Clark, 1996a; Wheatley et al., 2019). This difficulty is not surprising considering the *complexity* of conversation. When two people converse, they communicate using their words, body movement, gestures, voice, laughter, eye-contact, and more. Typical research studies aim to isolate one feature at a time to examine the impact of that feature on an outcome of interest. With so much happening all at once in conversation, this approach can become overwhelming. Beyond these logistical challenges is the fact that trying to study individual conversation features in isolation may fundamentally change the nature of the interaction—to the point where it no longer resembles the types of conversations that actually take place in our daily lives (Levitt & List, 2007; Lewin, 1939). Indeed, many of these features naturally co-vary over the course of a conversation. Further, many may be automatic processes that are disrupted by directing participants' attention to them.

In this dissertation, I take the approach of recording conversations as they happen naturally. Participants are not instructed to change or monitor a particular behavior. Instead, they are simply told to talk about whatever they want. This approach allows me to quantify different aspects of conversation behavior *after the fact*. I can then relate these behaviors to participant reports about how the conversation went.

The primary dependent variables. I am most interested in exploring what happens in initial conversation that leads to greater *enjoyment* and *connection*. Another major goal of this dissertation is to better understand how conversations differ between people who have never met (strangers) and close friends.

The datasets. To answer these questions, I collected two large datasets of unstructured conversations: one between people who were randomly paired together and one between close friends. Because these datasets appear in all three chapters of my dissertation, I wanted to explain them in detail here. Some chapters also contain additional datasets, collected to answer specific questions. Those are explained in their respective chapters.

Round Robin Dataset (Stranger dataset).

Participants. A total of 66 Dartmouth undergraduate students (33 female) were assigned to 11-person same-gender round robin groups. Each member of each round robin was scheduled to have 10 conversations, one with every other member of the round robin group. Participants never had more than three conversation sessions on any given day. We collected six round-robin groups with a goal of recording 330 conversation sessions. We were unable to complete eight sessions because of medical, scheduling, or technical issues. We collected a total of 322 conversation sessions.

Most participants had never met each other prior to their conversation. In response to the question “How well did you know your study partner before today?” (0 = Not well at all, 50 = Moderately well, and 100 = Extremely well), the mean response was 8.98 (SD = 20.55).

Conversation Session. In each study session, two participants entered the laboratory and had an unstructured, 10-minute conversation with each other that was video and audio recorded. Participants were told that they were free to talk about whatever they wanted. After the 10-minute conversation, participants were separated into private rooms where they completed a Qualtrics survey about the conversation they just had and about the conversation partner they just met. Participants then completed a second task that required them to watch the video recording of their conversation. As they watched, participants continuously rated how connected they remembered feeling to their conversation partner at each moment in time. Participants made these ratings by using a computer mouse to move a slider bar on the screen. Each conversation session took about 30 min to complete.

Benefits of the round-robin approach. The round-robin approach allows us to examine how a single individual behaves in 10 different conversations with 10 different people. It also allows us to examine how 10 different people feel about a single individual. There is a lot of rich information in a round robin structure (Kenny & La Voie, 1984; Wood et al., 2022). For example, in Figure 2B we can visualize how an entire round robin answered the question, "How much did you enjoy your conversation?". As we highlight, there are certain people who no one in the network enjoys talking to and others who everyone enjoys talking to. Recording conversations allows us to examine conversation features *within* a conversation and *across* conversations. Situating these conversations within a round-robin allows us to examine conversation features within and across *individuals*.

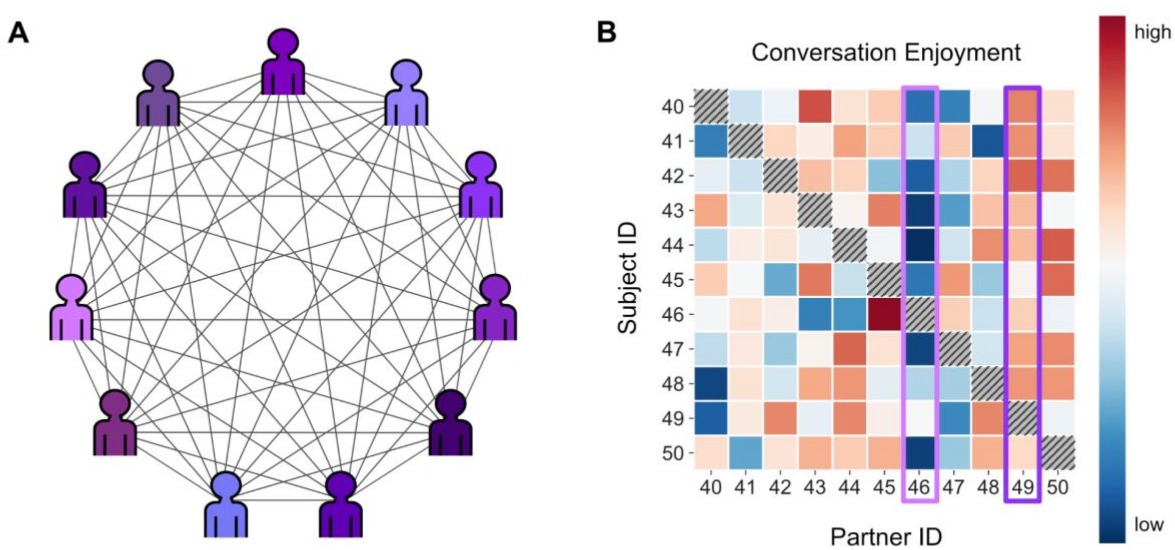


Figure 1. (A) Round robin network structure. (B) All responses to the question "How much did you enjoy your conversation?" for a single round robin. Responses have been z-scored within each subject. Two subjects are highlighted to show differences in partner impressions. People did not tend to enjoy conversations with Subject 46 whereas most people tended to enjoy conversations with Subject 49. Impressions for other subjects are varied; some people enjoy talking to them more than others.

Friend Dataset. In addition to studying what makes people feel connected in initial conversations, we also wanted to contrast conversations between people who were just getting acquainted with people who were already connected. To do this, we collected a second dataset of conversations between close friends.

Participants. We invited all 66 participants from the Round Robin dataset to participate in a follow-up study. Twenty-two of those participants were willing and able to participate. Participants completed the same conversation session as outlined above, with three of their close friends as their study partners. These conversational partners were someone they 1) considered to be a close friend, 2) interacted with regularly, and 3) were not romantically involved with. Dyads could be same or mixed gender for this study (female/female = 32, male/male = 20, and female/male = 13). We collected a total of 65 conversation sessions between friends.

Benefits of seeding participants from the round robin study. By comparing conversation behavior in this dataset to conversation behavior in the round-robin dataset, we can examine how friends and strangers converse differently. Because we seeded participants from the round-robin dataset, we also get to compare and contrast conversation behavior between friends and strangers for the *same people*. This allows us to examine a single person's behavior across multiple conversations, in two different contexts (friend and stranger). This opens up a whole host of interesting questions. For example, how much do people modulate their behavior when they are talking to strangers vs friends? Does a really good conversation with a stranger look like a conversation with a friend? Or are conversations with friends different in kind from conversations with strangers?

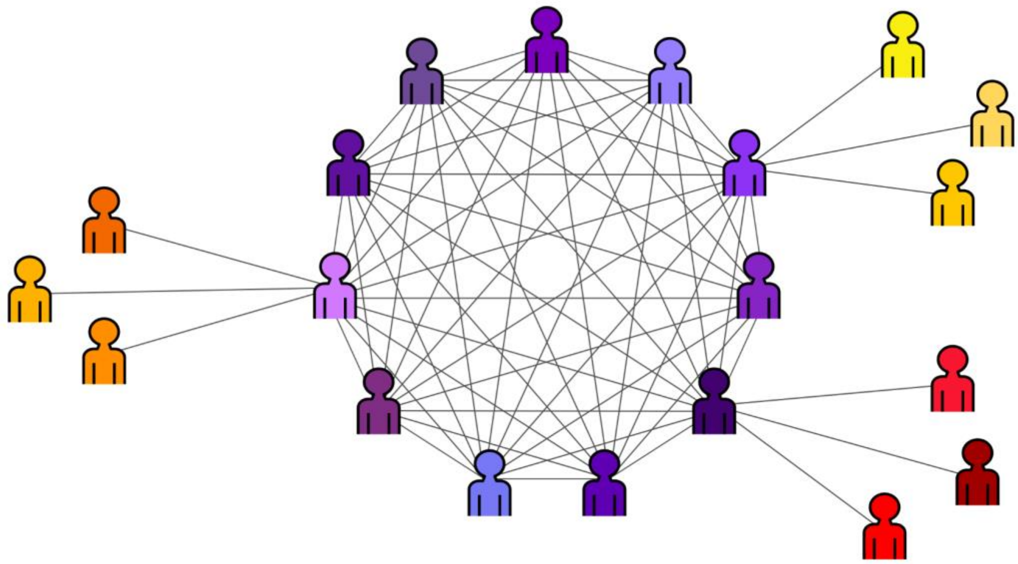


Figure 2. A subset of participants in the round-robin dataset also participated in the friend dataset. Each participant had three conversations with three different friends. In this visualization, the friends are in the periphery, in warmer colors.

An overview of the dissertation chapters

My dissertation will focus on three different aspects of conversation behavior (Fig 3). In Chapter 1, I examine whether response time predicts when people feel connected. I find that faster response times (shorter gaps between turns) act as an honest signal of connection in conversation. I also find that friends have more instances of particularly *long* gaps in their conversations and that these long gaps have different social consequences depending on relationship type—for strangers, long gaps mark moments of diminished connection whereas for friends, they mark moments of heightened connection. In Chapter 2, I examine how people *start* their conversations. I find that strangers start their conversations more similarly to each other, compared to friends. Further, the topics that strangers use to start their conversations may be particularly well-suited to "launch" them into more interesting topics later on. In Chapter 3, I explore the use of insider language, that is, words referring to something unspoken between two people (e.g., an inside joke or a past shared experience). I find that friends use more insider language compared to strangers and that insider language use corresponds to greater feelings of connection within a conversation.

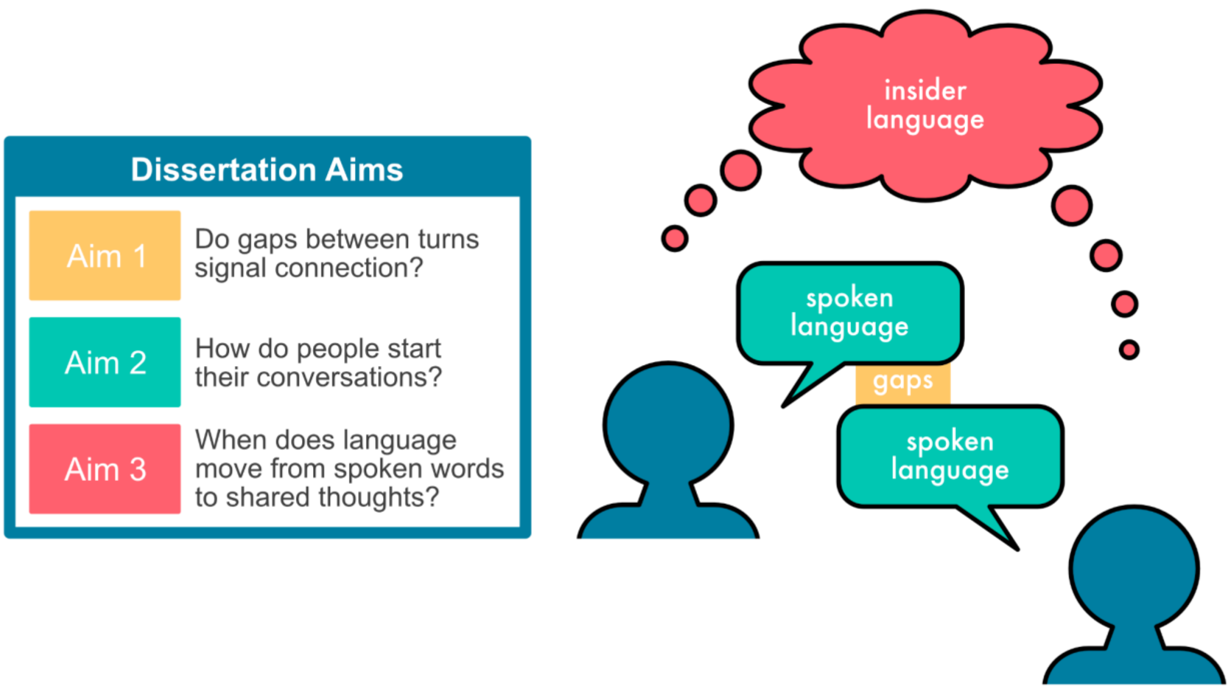


Figure 3. Dissertation aims, illustrated.

With these three chapters, I hope to better understand what happens in conversation that makes people feel connected. I also hope to highlight the benefits of using data-driven, naturalistic approaches to studying complex social phenomena like conversation.

Chapter 1a: Fast response times signal social connection in conversation

Emma M. Templeton, Luke J. Chang, Elizabeth A. Reynolds, Marie D. Cone LeBeaumont, & Thalia Wheatley. Published in *Proceedings of the National Academy of Sciences* (2022), Vol. 119 (4), e2116915119.

E. Templeton: Conceptualization, Data curation, Formal analysis, Investigation, Project administration, Software, Visualization, Writing - original draft, Writing - review & editing. **L. Chang**: Conceptualization, Resources, Supervision, Writing - review & editing. **E. Reynolds**: Investigation. **M. Cone LeBeaumont**: Investigation. **T. Wheatley**: Conceptualization, Funding acquisition, Resources, Supervision, Writing - original draft, Writing - review & editing.

Introduction

Turn-taking is a human universal (de Vos et al., 2015; Levinson, 2006; Pickering & Garrod, 2004; Stivers et al., 2009) that develops early (Bruner, 1975) and has deep evolutionary roots (Chow et al., 2015; Geissmann & Orgeldinger, 2000; Méndez-Cárdenas & Zimmermann, 2009; Snowdon & Cleveland, 1984; Takahashi et al., 2013). Months before words are uttered, infants engage in a communicative back and forth that helps establish a bond with their caregivers (Gratier et al., 2015; Jaffe et al., 2001). Within this ecological niche, language develops, adding the exchange of semantic meaning (Holler & Levinson, 2019; Schegloff, 2006). In a remarkable feat of coordination, turn-taking minimizes the time that one speaker stops and the other begins without sacrificing understanding (H. H. Clark, 1996b; H. P. Grice, 1975; Sacks et al., 1974). The modal conversational response time is extremely short, around 200ms (Heldner & Edlund, 2010; Levinson & Torreira, 2015)—three times faster than the average speed with which people can name an object (Indefrey, 2011; Indefrey & Levelt, 2004) and too rapid to rely on deliberative conscious control (Aron & Poldrack, 2006). Conversational response time is also extremely consistent across cultures and languages (Stivers et al., 2009), suggesting a universal optimum that balances efficiency and comprehension.

Minimizing time between turns requires multiple layers of prediction. People need to prepare an appropriate response in advance, notice when their partner is likely to end their turn, decide when to deliver their response, and anticipate their partner's reaction (Bögels et al., 2015; Bögels & Torreira, 2015; Levinson, 2016; Magyari et al., 2014; Riest et al., 2015; Sacks et al., 1974). Building an overarching mental model of the conversation further aids prediction, helping to anticipate not only *when* their partner is going to speak, but *where* their thoughts are headed (Barr & Keysar, 2006; Rossignac-Milon et al., 2020). As such, response time conveys how well one mind predicts another; a behavioral metric of being "heard and understood" (Gramling et al., 2016). As a marker of one mind understanding another, do fast response times also signal when two people feel connected?

Across three studies, we investigated whether response time provides a useful indicator of social connection in conversation. In Study 1, we leveraged a rich, naturalistic dataset to investigate the relationship between response time in real conversations to ratings of social connection at multiple levels of analysis—across and within conversations as well as individual differences. Unconstrained and naturalistic experimental contexts provide a representative design for the real conversations that we engage in every day (Brunswik, 1955). In Study 2, we determined whether these effects generalize to a different conversational context—conversations between close friends. In Study 3, we manipulated response times to investigate whether response times in conversation causally impact perceptions of social connection.

Results

Study 1: Social Connection and Response Time

In Study 1, we examined the relationship between response time and social connection across three levels of analysis: (i) across conversations, (ii) within conversations, and (iii) across individual participants. Participants (N = 66) completed ten 10-minute unstructured conversations within six same-gendered round-robin groups (322 conversations in total). The majority of participants had never met each other prior to their conversation. After their conversation, participants privately rated their overall conversation enjoyment and then watched a video-recording of their conversation while continuously rating how connected they felt to their partner at each moment in time. Response times were calculated by subtracting the start timestamp of a given speech turn from the end timestamp of the previous speech turn (Fig 4).

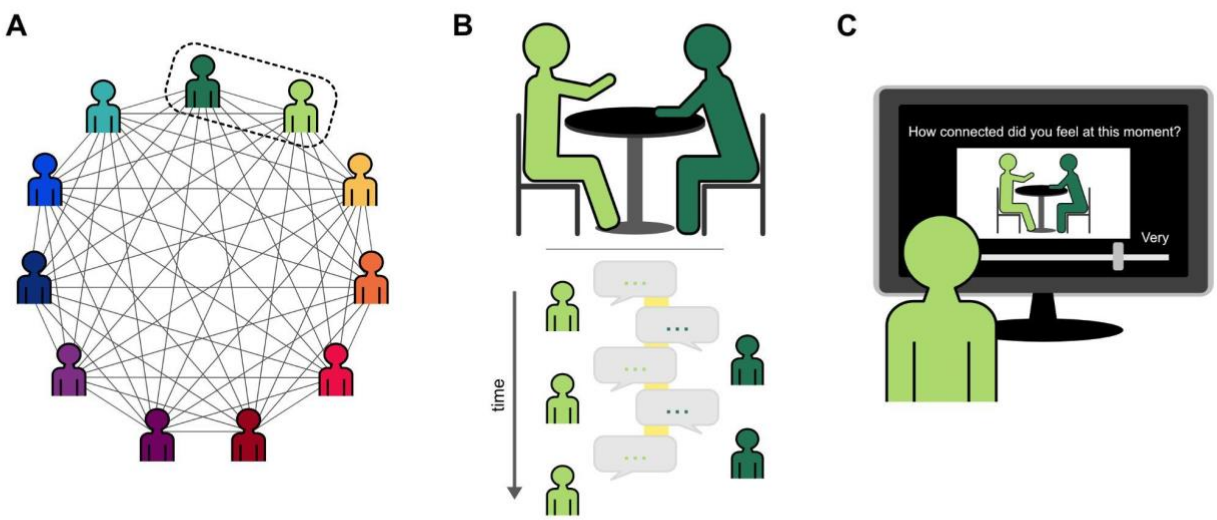


Figure 4. Study design. (A) Each participant was part of an 11-person round-robin. They were scheduled to have 10 study sessions, one with every other member of the round-robin. (B) Top: Each study session began with a 10-minute unstructured conversation. Bottom: A representation of how response time was computed. Each speech bubble represents one speech turn. The space in between the speech bubbles, highlighted in yellow, represents the response time. Response times are the amount of time in between the end of one turn and the start of the next turn. (C) After the conversation, in separate rooms, participants completed a survey about their conversation and then watched a recording of their conversation while continuously rating how connected they felt to their study partner.

We first tested the relationship between response time and conversation outcomes by computing the average response time in each conversation. We then related this value to participant's own reports of their enjoyment and connection within that conversation. In line with our hypothesis, we found that faster response times positively predicted reported enjoyment (b = -0.35, SE = 0.05, p < .001, Fig 5A) and social connection (b = -0.28, SE = 0.05, p < .001, Fig 5B).

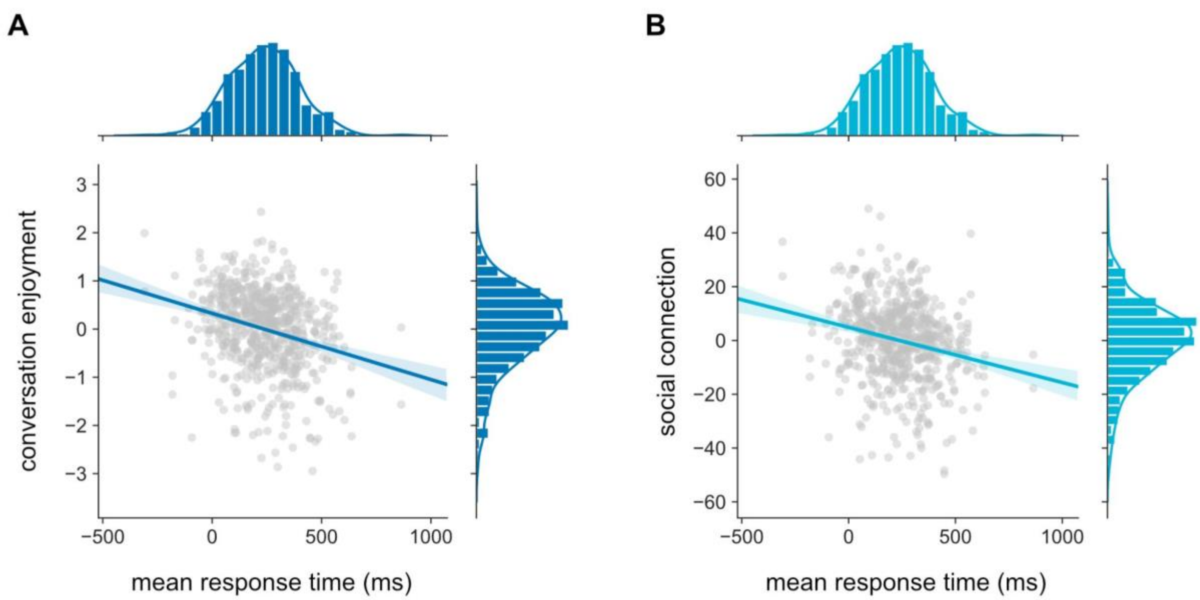


Figure 5. Mean response time predicts (A) conversation enjoyment and (B) social connection. Dependent variables (DVs) of enjoyment and connection are centered within-subject to reflect the random effect structure used in the mixed-effects models. Individual data points are displayed as gray dots. The line represents a regression model relating mean response time and each DV. The distribution of mean response times is plotted above the scatterplots and the distribution of each DV is plotted to the right of the scatterplots.

We also hypothesized that feelings of connection would covary with response time *dynamically, within a conversation*. To test this hypothesis, we divided each 10-minute conversation into twenty 30s bins and within each bin computed the average response time and connection rating for each conversation partner based on their continuous moment-by-moment ratings. We observed a significant effect of time on connection, indicating that participants' reported connection increased over the course of their conversation (b = 0.27, SE = 0.01, p < .001). Controlling for this temporal effect, we also found that response times significantly predicted connection ratings (b = -0.03, SE = .01, p = .002). This effect was invariant to different bin sizes (Fig S3).

We next investigated whether faster responders are better liked by their conversation partners. To test this, we computed each participant's average response time across all of their conversations. Similarly, for each participant, we computed the average amount of conversation enjoyment and connection their *partners* felt after talking with them. We then ran two linear regressions with average response time across all conversations predicting average reports of enjoyment and connection made by each participant's conversation partners. We found that participants with faster average response times evoked more enjoyment (b = -0.64, SE = 0.10, p < .001) and feelings of connection (b = -0.63, SE = 0.10, p < .001; Fig 6) in their partners.

Taken together, we found evidence that faster response times relate to increased social connection across three different levels of analysis—across conversations, within conversations and across individuals.

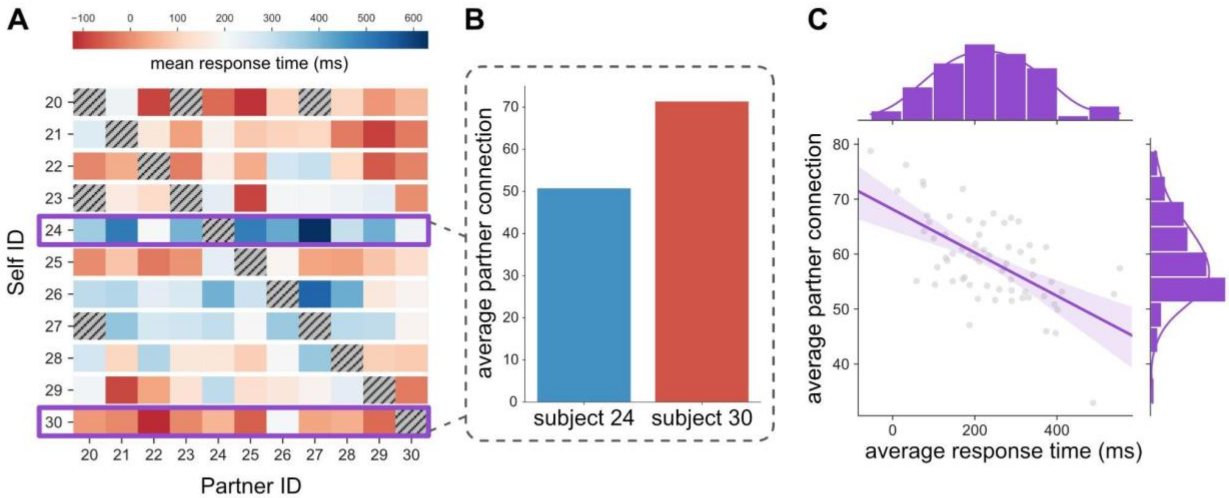


Figure 6. Individual differences in response times predict connection across partners. (A) Real data from one round robin network. The color of each cell indicates the mean response time for a given subject in each of their 10 conversations. The colorbar is centered at 200ms. Gray cells indicate missing data. As highlighted, participant 24 tended to have conversations with relatively slow response times whereas participant 30 had conversations with relatively fast response times. (B) Average partner connection ratings for these two participants. (C) The relationship between average response time and average partner connection across all six round robin participants (dots are individual participants). Distributions at the top and right, depict the probability density functions for average response time and average partner connection, respectively.

Study 2: Generalizing to a Different Context

In Study 1, we found evidence that faster response times in stranger conversations were robustly associated with increased social connection. Next, we were interested in assessing the generalizability of these results to additional conversational contexts. Specifically, we were interested in determining if this relationship was evident in people who were already strongly connected. Thus, in Study 2 we investigated whether response times predicted felt connection for real-world close friends. To test this hypothesis, a subset of participants from Study 1 (N = 22) returned to complete additional conversations with three of their friends (N = 65 conversations).

These conversational partners were someone they: (i) considered to be a close friend, (ii) interacted with regularly, and (iii) were not romantically involved with.

As a manipulation check, we confirmed that close friends in Study 2 rated their conversations more favorably overall than strangers in Study 1 (conversation enjoyment: $M_{friends}$ = 87.95 (SD = 14.52), $M_{strangers}$ = 72.55 (SD = 20.95), $t(251.28) = 10.15$, $p < .001$; see Table S1 for all comparisons). Indeed, reports of overall enjoyment and average connection between close friends were so uniformly high and invariant across dyads that they precluded across-conversation analysis. However, we were able to run the within-conversation analysis and leverage the *dynamics* of the continuous reported connection ratings to test whether time points with faster response times corresponded to relatively higher social connection. We observed that faster response times in conversations between close friends significantly predicted greater feelings of social connection ($b = -0.07$, $SE = .02$, $p < .001$) above and beyond a general increase in reported connection over the course of conversations ($b = 0.25$, $SE = 0.03$, $p < .001$). These results confirm that our findings from Study 1 appear to be robust to conversational context and are present not only in conversations with strangers, but also when interacting with close friends.

Self and Partner Effects

In the analyses reported in Studies 1 and 2, we treated response time as a metric shared by conversation partners. However, this approach obscures *whose* response time is driving the effect. Are my feelings of connection predicted by how quickly I respond to you (self response time)? Are my feelings of connection predicted by how quickly you respond to me (partner response time)? Or are both response times equally important to connection? (Fig 7A).

We first explored this idea using conversations between strangers from Study 1. Using a mixed-effects regression, we found that both self ($b = -0.11$, $SE = .05$, $p = .048$) and partner ($b = -0.27$, $SE = .05$, $p < .001$) response times independently and significantly explained variance in ratings of self enjoyment. In addition, we found that partner (b = -0.22, SE = .05, p < .001), but not self (b = -0.08, SE = .05, p = .075) response times significantly explained variance in ratings of self connection. We compared the magnitude of the self and partner effects using a contrast analysis and found that partner response times were consistently a better explanation of both enjoyment (t(65) = -14.48, p < .001) and connection (t(65) = -8.63, p < .001), compared to self response times (Fig 7B).

Next, we explored whether these partner effects were also present in the connection dynamics within conversations. We tested this hypothesis for both stranger and friend conversations. Using a mixed effects regression, we found that relative changes in connection ratings with strangers were significantly predicted by partner response times (b = -0.03, SE = .00, p < .001) but not self response times (b = -0.01, SE = .01, p = .279) controlling for linear trends. A contrast test revealed that the magnitude of the partner response time effects were consistently stronger than self response times across participants (t(65) = -5.53, p < .001, Fig 7C). We observed a similar pattern of results in the friend conversations. Social connection ratings were significantly independently explained by both partner (b = -0.06, SE = .01, p < .001) and self (b = -0.04, SE = .01, p = .006) response times. However, the magnitude of the partner response effect was consistently greater than the self response time effect across conversations (t(86) = -7.77, p < .001, Fig 7C).

Together, these findings indicate that how much a person enjoys a conversation and feels connected to their partner is predicted more by how quickly their partner responds to them rather than by how quickly they respond to their partner.

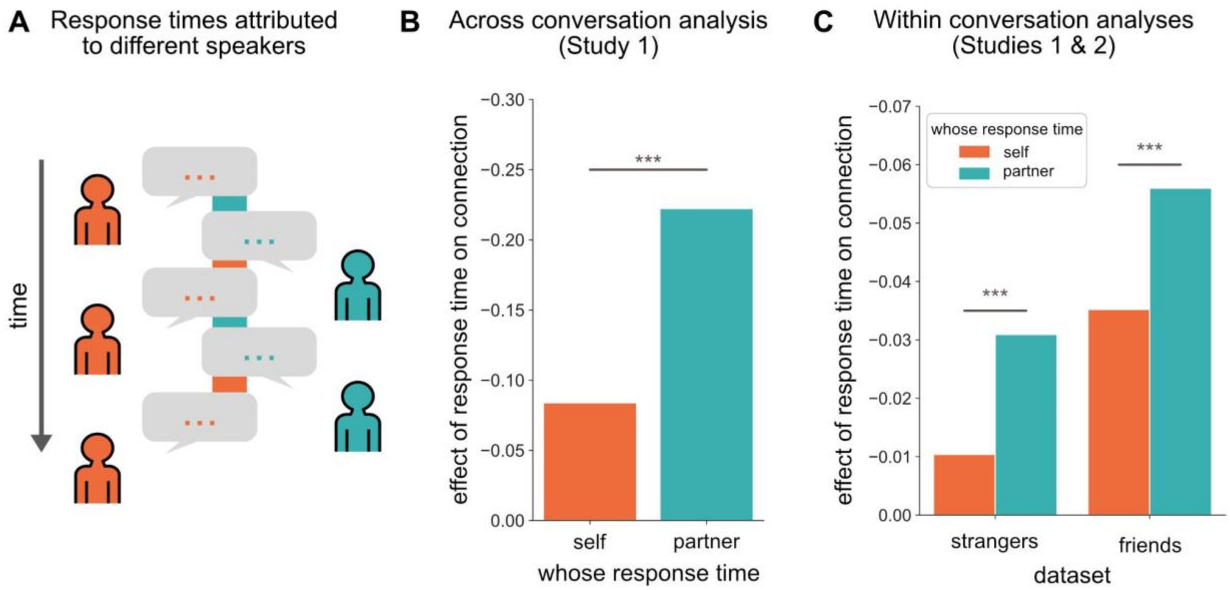


Figure 7. Partner responsivity had a greater influence on connection than self responsivity. (A) Each response time (depicted by rectangles between speech bubbles) was attributed to the speaker who ended the preceding silence. Response times are colored to match the person to whom they are attributed. (B) Beta coefficients for the effect of self and partner response times on self reports of connection in the across conversation analysis (Study 1). (C) Beta coefficients for the effect of self and partner response times on self reports of connection in the within conversation analysis, for strangers (Study 1) and friends (Study 2). Note that the y-axis labels have been flipped for readability, as more negative values indicate a stronger relationship between response time and connection. * $p < .05$, ** $p < .01$, *** $p < .001$

Study 3: Manipulating Response Time

The previous analyses demonstrate that conversational moments with faster response times are robustly associated with increased feelings of enjoyment and connection compared to moments with slower response times across multiple levels of analyses and conversational contexts. Given this relationship, we wondered whether faster response times are, themselves, a *sufficient signal* of enjoyment and connection to outside observers.

In Study 3, we tested whether response time alone signals enjoyment and connection to third party listeners. We selected short audio clips (~10 turns) from the beginning of six conversations recorded in Study 1 and manipulated the length of the response times between speech turns (Fig 8). Response times were shortened to one-fifth the original length in the Fast condition, and lengthened to twice the original length in the Slow condition. The Control condition maintained the original response times. Participants (N = 450) recruited on Amazon's Mechanical Turk listened to all six conversation segments, with each segment randomly assigned to one condition (i.e., Control, Fast, Slow). Participants judged the overall conversation enjoyment and connection between the conversation partners after listening to each segment.

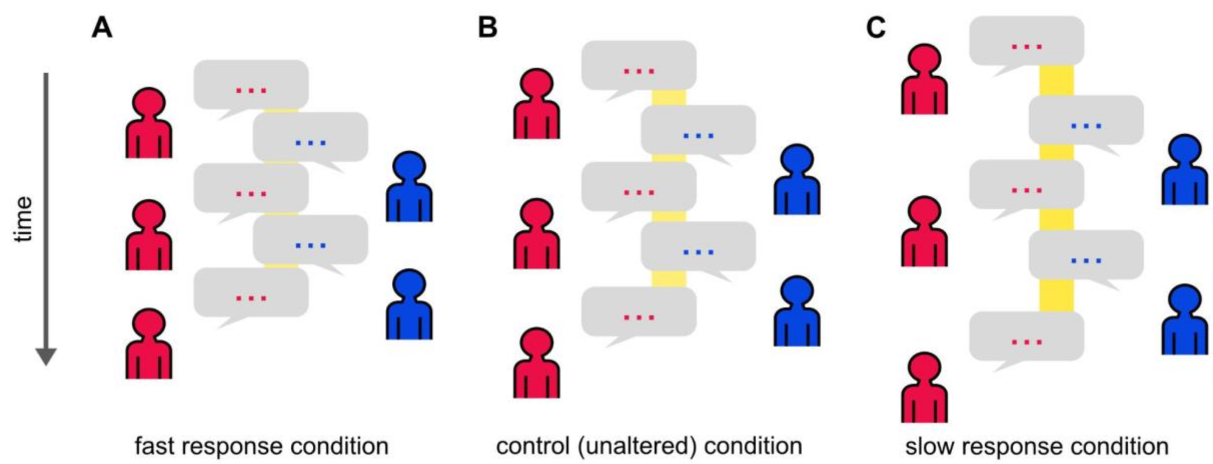


Figure 8. Manipulation of response times. The length of the yellow rectangles indicates the length of each response time. (A) In the fast response condition, each response time was decreased to one-fifth its original length. (B) In the control condition, we used the original (unaltered) response times. (C) In the slow response condition, each response time was double its original length.

We ran two linear mixed effects models with condition (Control, Fast, Slow) predicting each of our two DVs: perceived enjoyment and perceived connection. An ANOVA on these mixed effects models yielded a significant effect of condition such that response time inversely predicted perceived enjoyment ($F(2, 2351.9) = 49.44$, $p < .001$) and connection ($F(2,2344.3) =$

28.51, $p < .001$, Fig 9) by third party listeners. That is, the same conversation was perceived as more enjoyable and connected when response times were decreased and less enjoyable and connected when response times were increased. The same conversation with unaltered response times was rated midway between the two altered versions. Specifically, ratings of enjoyment and connection were significantly lower for the unaltered version compared to when response times were decreased (Fast condition). Ratings of enjoyment (but not connection) were significantly higher for the unaltered version compared to when response times were increased (Slow condition). These findings were replicated in a second sample (Fig S7).

Unlike Studies 1 and 2, the relationship between response time and enjoyment/connection could not be explained by any other feature of the conversation that could conceivably covary with response time (e.g., conversation topic, vocal prosody, etc). This is because only response times varied between versions; everything else about the conversation was held constant. Therefore, Study 3 provides strong evidence that fast response times not only covary with enjoyment and connection, they are a *sufficient* signal of enjoyment and connection to third party listeners.

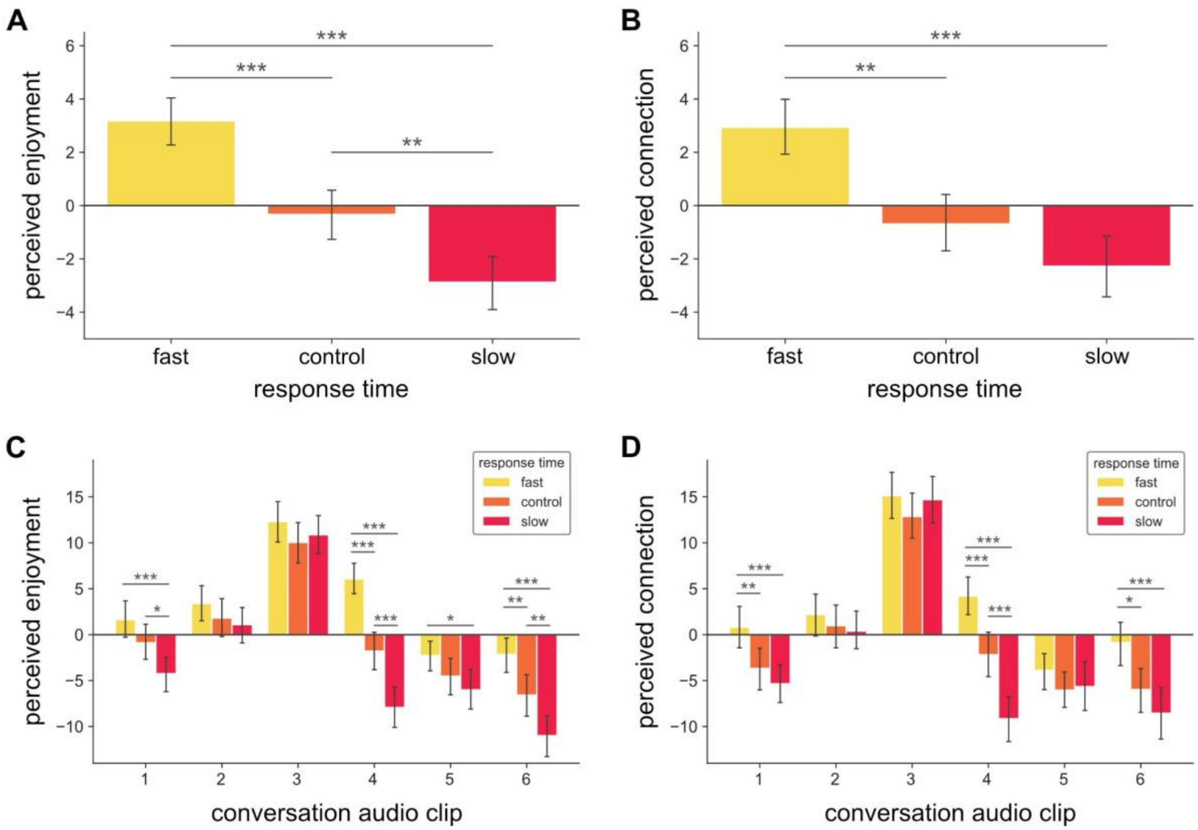


Figure 9. Main effects of response time condition on average ratings of (A) enjoyment and (B) connection across conversations. Effect of response time condition on average ratings of (C) enjoyment and (D) connection for each of the six conversations, separately. All values are centered within-subject to reflect the random effect structure used in the mixed-effects model. P-values have been adjusted for multiple comparisons using Holm's method. Error bars depict 95% confidence intervals. * $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

Conversation is an incredible feat of coordination (Bögels et al., 2015; H. H. Clark, 1996b; Levinson, 2016; Magyari et al., 2014; Riest et al., 2015; Sacks et al., 1974). We must pass the conversational baton within a split second and, as with professional athletes, a few milliseconds can make a striking difference. Here we show that how well people pass this conversational baton is a robust marker of how connected they feel. Across two studies of unstructured natural conversation, we found that faster response times were associated with increased social connection in conversations—both between strangers and friends. Reduced response times likely reinforce feelings of connection. At the same time, because the ability to respond quickly in conversation relies on accurately predicting what your partner is going to say and noticing when their turn is likely to end, we suspect that fast response times may be facilitated by feelings of connection. Natural conversation is likely marked by these mutually reinforcing dynamics.

Conversation enjoyment and connection were better explained by partner (vs self) response times. This suggests that when someone responds quickly it signals to their partner that they had been actively listening. This finding dovetails with the existing literature highlighting the importance of "feeling heard and understood" in conversation (Gramling et al., 2016).

We further demonstrated that response time in conversation is, in and of itself, a sufficient signal of connection to outside observers. Holding everything else about the conversation constant, a split second difference in response time was enough for outside observers to infer connection or a lack thereof. Importantly, listeners were never instructed to pay attention to the timing of turns. Observers may have implicitly learned that response time and connection covary. This finding extends prior work demonstrating that outside observers infer another's preferences (Gates et al., 2021), sincerity (Ziano & Wang, 2021) and certainty (Van de Calseyde et al., 2014) by how many seconds they take to make a decision between available options: faster decisions appear to express stronger or "truer" preferences. Response time in natural conversation, on the order of milliseconds, may similarly be taken as a true signal of connection. That this signal is available to observers further suggests that response time may be used to determine who clicks with whom around us. This is consistent with previous research demonstrating that third party observers are highly attuned to how others connect in their social network (Parkinson et al., 2017). The fact that response times evoke perceptions of connection when holding all else constant, further suggests this heuristic traverses language barriers and may be available to preverbal infants (Seyfarth et al., 2005).

It is important to acknowledge several limitations of these studies. First, the stranger conversations that we recorded consisted of undergraduate students engaging in polite, get-to-know-you talk. Participants knew their conversations would only last for 10 minutes and there was no expectation that they would need to interact with their conversation partners again. These types of interactions happen frequently in our daily lives, especially when we move to a new place or start a new job, and are the locus where most relationships begin. Conversations between close friends offered some generalizability beyond this domain, but there are many other types of conversation contexts that remain unexplored. For example, we might expect that response times relate differently to enjoyment and connection in conversations where there is a clear goal (e.g., negotiation, interview) or in conversations that are antagonistic (e.g., an argument). Conversations with conflict are characterized by people speaking on top of each other and jumping in quickly (Grezes et al., 2013; Trimboli & Walker, 1984). In this context, fast response times might actually signal that two people are *not* listening to each other (Bögels et al., 2018). However, it may also be the case that rapid turn-taking is still signaling psychological investment, either in the partner or the topic being discussed. More research is needed to better understand the role of response times in different conversational contexts. Second, our sample was from a WEIRD (Western, Educated, Industrialized, Rich, Democratic) population (Henrich et al., 2010). Although the average response time between strangers is remarkably consistent cross-culturally (Stivers et al., 2009), that average may obscure interesting cultural variations that may likewise differ across conversational contexts. Finally, our sample does not allow us to investigate the myriad ways that particular dyadic compositions can influence conversational dynamics. More research is needed to explore how turn-taking behavior changes as a function of two or more people sharing or not sharing the same backgrounds, demographics (e.g., gender, race, age), social status, or other aspects of identity.

Although we reliably found stronger effects for partner (vs. self) responsivity, our results cannot adjudicate what determined any specific response time. Speeded responses are likely facilitated by a number of self and partner factors including, but not limited to: partner attention, communicative clarity (e.g., signposting when a turn is ending), emotional salience, and topic expertise. The present finding—that response time indexes connection—opens up future research to investigate the (likely many) mechanisms by which this is achieved.

In summary, across three studies, we showed that response time in conversation has important social consequences. Response times in everyday conversation are remarkably short (Heldner & Edlund, 2010; Levinson & Torreira, 2015; Stivers et al., 2009). They are simply too fast to be under conscious control (Indefrey, 2011; Indefrey & Levelt, 2004) and thus cannot be faked. This brevity is a feat of coordination that provides a natural, “honest” heuristic about how well the conversation is going (Guilford & Dawkins, 1991). Moreover, by virtue of being a feature of conversation itself rather than requiring post-hoc self-report and by virtue of being a signal readily accessible to outside observers, response times may provide a useful metric for future research investigating the conditions that diminish and enhance connection. Conversation is typically discussed in terms of what people talk about. The present findings reveal that the *when* of conversation—how fast one partner responds to the other—is also important, providing a robust, efficient and honest signal of social connection.

Materials and Methods

Study 1

Participants. Sixty-six Dartmouth undergraduate students (33 female) participated in exchange for course credit. We used a round-robin design (Fig 4A), with every round consisting of 11 same-gender participants. We chose to limit this dataset to same-gender dyads given that there may be additional dynamics at play in mixed vs same-gender interactions. All participants were scheduled to complete 10 conversation sessions, one with every other member of the round-robin. We collected six round-robin groups with a goal of recording 330 conversation sessions. We were unable to complete eight sessions, due to medical, scheduling, or technical issues. We collected a total of 322 conversation sessions. All reported studies were approved by the Dartmouth Committee for the Protection of Human Subjects and all participants provided informed consent prior to participation.

The majority of participants had never met each other prior to their conversation. In response to the question “How well did you know your study partner before today?” (0 = Not well at all, 50 = Moderately well, 100 = Extremely well) the mean response was 8.98 (*SD* = 20.55). We therefore refer to participants in this study as “strangers”. Note that all reported effects hold after removing dyads who knew each other before the study.

Study design. In each study session, two participants entered the lab and had an unstructured, 10-minute conversation with each other that was video and audio recorded. Participants were told that they were free to talk about whatever they wanted. After the 10-minute conversation, participants were separated into private rooms where they completed a Qualtrics survey about the conversation they just had and about the conversation partner they just met (see Appendix A for the full list of survey items). Participants then completed a second task that required them to watch the video recording of their conversation. As they watched, participants continuously rated how connected they remembered feeling to their conversation partner at each moment in time. Participants made these ratings by using a computer mouse to move a slider bar on the screen (Fig 4C). Each session took about 30-minutes to complete. Participants never had more than three conversation sessions on any given day.

Defining primary DVs. We conducted a factor analysis on the post-conversation survey items after ensuring our items passed both Bartlett's test of sphericity and the Kaiser-Meyer-Olkin test. The factor that accounted for the most variance (34%) loaded onto questions related to conversation enjoyment. We therefore used this factor as our dependent variable of *conversation enjoyment*. Questions included, "How much did you enjoy the conversation you had with your study partner?" and "How well did this conversation 'flow'?" (see Fig S1 for all factor loadings).

Our second dependent variable was *social connection*. To calculate this measure, we took the mean of the continuous connection ratings that participants made as they watched their conversation recording.

Defining response time. The recorded conversations were transcribed by an external transcription company. Each speech turn in each transcript included the speaker's identity, the timestamp indicating when the speaker started talking, the timestamp when the speaker finished talking, and the transcription of what they said. All of the timestamps included millisecond-precision. This level of fidelity was especially important for our research question, as we expected the average response time to be ~200ms.

Response time was calculated by taking the start timestamp of a given turn and subtracting the end timestamp of the previous turn. Response times with negative values indicate instances when speakers overlap. See Supplement for more details about our transcriptions.

Statistical models. For all reported analyses we used lme4 (Bates et al., 2015) implemented in R (R Core Team, 2018) to perform linear mixed effects regressions. Degrees of freedom and p-values were approximated using Satterthwaite's method and we report standardized regression coefficients to increase interpretability.

Across conversation analysis. We predicted each of our two DVs (i.e., conversation enjoyment and social connection) using average response time in a given conversation. Because subjects participated in multiple conversations, we included Subject ID as a random intercept. Because the relationship between response time and each of our two DVs could vary between different subjects, we also included response time as a random slope.

Within conversation analysis. We ran a mixed linear effects model predicting the temporal dynamics of social connection based on fluctuations in average response time controlling for linear effects of time. To account for variations in average response time between dyads, we included Dyad ID as a random intercept and additionally modeled Subject ID as a random intercept because subjects participated in multiple conversations. We modeled response time as a random slope for Subject ID to account for the fact that the relationship between response time and connection may vary between subjects. We also modeled the linear effect of time as a random slope for Dyad ID to account for the fact that the relationship between time and connection may vary between dyads.

To investigate the robustness of this effect, we generated surrogate data by randomly permuting the order of response times within each conversation using a circle-shifting procedure and re-fitting the model predicting social connection 100 times (Lancaster et al., 2018). This non-parametric analysis generates an empirical null distribution of random shuffles of our data while maintaining the structure of any inherent temporal autocorrelation. Importantly, this demonstrates that our results cannot be explained by any offsets in lag between changes in response time and connection ratings (Fig S3). Moreover, these results appear to be robust to bin size as we observed similar effects across a range of different bin sizes (Fig S3).

Study 2

We invited all 66 participants from Study 1 to participate in this follow-up study. Twenty-two of those participants were willing and able to participate. Participants completed the same study procedure as outlined in Study 1, with three of their close friends as their study partners. Dyads could be same or mixed-gender for this study (Female/Female = 32, Male/Male = 20, Female/Male = 13). Given the small sample sizes within each of these categories, we did not analyze differences between these groups. We collected a total of 65 conversation sessions, transcribing the friend conversations in the same manner as described in Study 1 and similarly computing the response time between each speech turn. We used the same within conversation analysis as described in Study 1 and these analyses also passed the same robustness checks (Fig S4).

Self vs partner effects

Across conversation analysis (Study 1). For the across conversation version, we ran two different linear mixed effects models that included average response time for self and partner as fixed effects predicting each of our two DVs (i.e., conversation enjoyment and social connection). Because subjects participated in multiple conversations, we included Subject ID as a random intercept. Because the relationship between response time and each of our two DVs could vary between different subjects, we also included self response time and partner response time as random slopes.

Within conversation analysis (Studies 1 & 2). For the within conversation version, we ran a mixed linear effects model with average response time for self, average response time for partner, and bin number as fixed effects predicting self connection ratings in each bin. To account for variations in average response time between dyads, we included Dyad ID as a random intercept and additionally modeled Subject ID as a random intercept because subjects participated in multiple conversations. We modeled self and partner response times as random slopes for Subject ID to account for the fact that the relationship between response time and connection may vary between subjects. We also modeled bin number as a random slope for Dyad ID to account for the fact that the relationship between time and connection may vary between dyads.

To run the contrast that determined whether the effect of partner response time was stronger than the effect of self response time, we extracted the beta coefficients for each individual subject and contrasted the betas for the effect of self response time with the betas for the effect of partner response time. We used a one-sample t-test with 0 as the reference point to perform a hypothesis test.

Study 3

In Study 3 we tested the hypothesis that third party perceptions of social connection would be causally influenced by speaker response times. We identified six conversations from Study 1 (three male and three female) that had minimal overlapping speech, where both participants had signed a video release permitting us to use their recording in subsequent studies. For each video, we selected a short audio clip from the start of their conversations that comprised about 10 turns back and forth (min number of turns = 9, max number of turns = 13, mean clip length = 23.33 seconds). We used these stimuli to create three separate conditions by manipulating the response times for each speaker. In the Control condition, the response times between speech turns were the length they were in the original audio file (M = 278.55ms). In the Fast condition, each response time was manipulated to be one-fifth the original length (M = 55.68ms). In the Slow condition, each response time was manipulated to be twice the original length (M = 557.14ms). See Supplement for further details of how we manipulated these audio files. The methods of this study, as well as our hypotheses, were preregistered prior to collecting data (osf.io/u2brn).

Four hundred fifty participants recruited on Amazon's Mechanical Turk listened to one version of each of the six conversation segments, presented in a random order. All participants heard each conversation segment only once and the version (Control, Short, Long) of that conversation segment was randomly assigned. This random assignment was blocked such that, over all participants, each conversation segment was presented an equal number of times across all three conditions.

After listening to each conversation segment, participants responded to two questions: 1) *How much do you think these people enjoyed their conversation*? and 2) *How connected do you think these people felt toward each other*? Participants responded using a slider bar anchored by "Not at all" (0) and "Very much" (100).

To access the study, participants were first asked to complete a simple task (correctly typing the word spoken in the audio file) to ensure that only participants who were able to listen and respond to audio instructions were included in data analysis.

We ran two linear mixed effects models with condition (Control, Short gap, Long gap) predicting each of our two DVs: perceived enjoyment and perceived connection. We included Subject ID and Conversation ID (e.g., which of the 6 conversations was being judged) as random intercepts.

Chapter 1b: Long gaps between turns are awkward for strangers but not for friends

Emma M. Templeton, Luke J. Chang, Elizabeth A. Reynolds, Marie D. Cone LeBeaumont, & Thalia Wheatley. Published in *Philosophical Transactions of the Royal Society B* (2023), Vol. 378 (1875), 20210471.

E. Templeton: Conceptualization, Data curation, Formal analysis, Investigation, Project administration, Software, Visualization, Writing - original draft, Writing - review & editing. **L. Chang**: Conceptualization, Resources, Supervision, Writing - review & editing. **E. Reynolds**: Investigation. **M. Cone LeBeaumont**: Investigation. **T. Wheatley**: Conceptualization, Funding acquisition, Resources, Supervision, Writing - original draft, Writing - review & editing.

Introduction

Conversation is a feat of coordination, often characterized by rapid turn-taking. Indeed, the gaps between speech turns tend to be so short (~200ms (Heldner & Edlund, 2010; Levinson & Torreira, 2015)) that they can only be achieved by predicting what your partner is going to say next (Gisladottir et al., 2018; Magyari et al., 2014; Riest et al., 2015) and planning your response in advance (Barthel et al., 2016; Barthel & Sauppe, 2019; Bögels et al., 2015). More accurate predictions can facilitate faster response times and shorter gaps between turns. These response times have social consequences (Templeton et al., 2022). People in conversations with shorter gaps report enjoying their conversations more and feeling more connected to their conversation partners. When people listen to conversations where the gaps have been manipulated to be shorter, they perceive greater connection than people listening to the same conversation where the gaps have been manipulated to be longer. Given that short gaps are an indication that conversation is going well, do long gaps imply that something has gone wrong?

Existing literature strongly suggests that long gaps should be avoided. Long gaps in conversations between strangers are often attributed to poor social skills (McLaughlin & Cody, 1982). Qualitative research asserts that long gaps signal disagreement and sow discord (Jefferson, 1989; Pomerantz, 1984; Stivers & Robinson, 2006). Participants asked to read or listen to conversations that include long gaps report feeling uncomfortable and tend to assume that the people in those conversations feel uncomfortable as well (Koudenburg et al., 2011; Newman, 1982). Even watching interactions between a human and a robot that contain long gaps can make people feel awkward (Ohshima et al., 2015). Experimentally lengthening the gap between a request and a response has also been shown to create negative impressions (e.g., reluctance to comply, disagreement; (Kendrick & Torreira, 2015; Roberts et al., 2006, 2011; Roberts & Francis, 2013). Fears of awkward silences may be one reason why people avoid talking to strangers even though doing so is most likely to be enjoyable (Sandstrom & Boothby, 2021).

In contrast to these findings, a few studies have found that long gaps may not always be problematic. For example, therapists report strategically using silence to encourage reflection and convey empathy (Hill et al., 2003). Similarly, long gaps in doctor-patient communication can promote connection and increase patients' feelings of being heard and understood (Bartels et al., 2016). These findings suggest that, under certain circumstances, long gaps may convey care and reflection rather than awkwardness. Are unproblematic long gaps limited to therapeutic contexts or are they a feature in close relationships more generally?

The goal of the present research is to examine the social implications of long gaps in conversation for both strangers and friends. If long gaps uniformly signal discomfort and awkwardness, then friends may have fewer of them in their conversations compared to strangers.

On the other hand, friends may have different types of conversations than strangers (Planalp & Benson, 1992), many of which may benefit from pauses that promote deep reflection or savoring of inside jokes. This would suggest that long gaps may also be *experienced* differently by friends compared to strangers, which may also be detected by third-party observers. To investigate these questions we examined gaps within unstructured natural conversations between strangers and friends. In Study 1, we tested whether long gaps differ between strangers and friends in terms of frequency and experienced connection. In Study 2, we explored whether the long gaps of strangers and friends are perceived similarly or differently by outside observers.

Study 1

Participants

We examined dyadic conversations between strangers and between friends to investigate how long gaps are experienced differently across these two relationship types.

Stranger Dataset. Participants in the "stranger" dataset participated in exchange for extra credit in their Psychology or Neuroscience courses. Conversation partners were assigned by an experimenter. To ensure that participants did not know each other we asked them "How well did you know your study partner before today?" (0 = Not well at all, 50 = Moderately well, and 100 = Extremely well). In order to limit our analyses to true strangers who do not know each other, we excluded 61 dyads where both dyad members indicated a response greater than 0 on this question. The analyses reported in this paper come from 261 stranger dyads. However, note that results are similar with all dyads included.

Friend Dataset. All participants in the stranger dataset were invited to participate in the friend dataset. Those who were interested were asked to nominate their close friends to participate with them. Participants in this study had the option of receiving either cash compensation or extra credit in eligible courses. We recorded 65 conversations between dyads of friends.

Methods

Every conversation session began with two participants having a 10-minute unstructured conversation. Participants were seated across from each other at a cafe table. A webcam attached to a desktop computer across the room captured both participants in profile. After the recording was started, the experimenter turned off the Desktop screen so that participants would not be distracted by the recording during their conversation. Participants were told that they could talk about whatever they wanted. After 10 minutes, the experimenter re-entered the room, ending the conversation.

After their conversation, participants were moved to two separate rooms where they privately completed two tasks. They first rated their overall impressions of the conversation via a survey (see Supplement for all items). They then watched a video recording of their conversation while continuously rating how connected they remembered feeling to their conversation partner at each moment in time. Participants made these ratings by using a computer mouse to move an on-screen slider bar (from 0 = None to 100 = Very). The position of the mouse was recorded every 100 milliseconds.

The video recordings of each conversation were transcribed by an external transcription company. Each speech turn in each transcript included the timestamp (in milliseconds) indicating when the speaker started talking and the timestamp when the speaker finished talking. Gap lengths were calculated by subtracting the timestamp at the beginning of a given speech turn from the timestamp at the end of the previous speech turn.

Defining a long gap. Although the average gap length in conversation has been well established (~200ms; (Heldner & Edlund, 2010; Levinson & Torreira, 2015; Stivers et al., 2009), there is no agreed upon minimum threshold that defines a "long" gap. Here, we considered gaps to be "long" when they lasted more than 2 seconds (roughly 3 standard deviations from the mean of the distribution; M = 248 ms, SD = 598 ms). Note that gaps here are simply the absence of verbal speech between speakers. Gaps could therefore contain other non-verbal vocalizations or actions.

Results

Friends have more long gaps than strangers. We first turned to the question of whether long gaps were more prevalent in conversations between friends or strangers. Poisson regression is typically used to model count data. However, we found that our count data were more variable than could accurately be described by a traditional Poisson distribution (i.e., overdispersed (dispersion = 2.95, overdispersion test: $z = 6.49$, $p < .001$)) and also contained more instances of zeros as a consequence of our long-gap threshold (i.e., zero-inflated (ratio of predicted:observed zeros = 0.76)). Therefore, we used a mixed-effects zero-inflated negative binomial regression to predict the number of long gaps based on relationship type (friend or stranger) including subject ID as a random intercept (Brooks et al., 2017). Because each conversation had a different number of turns, we included the total number of gaps for each conversation as an offset parameter. Results reveal that friends have more long gaps than strangers ($b = -1.51$, $SE = 0.15$, $p < .001$, Fig 10). This finding was robust to varying the threshold for what constitutes a "long" gap (see Table S2) and also to the type of statistical model (similar results were found using a negative binomial regression, Poisson regression, and chi-square tests).

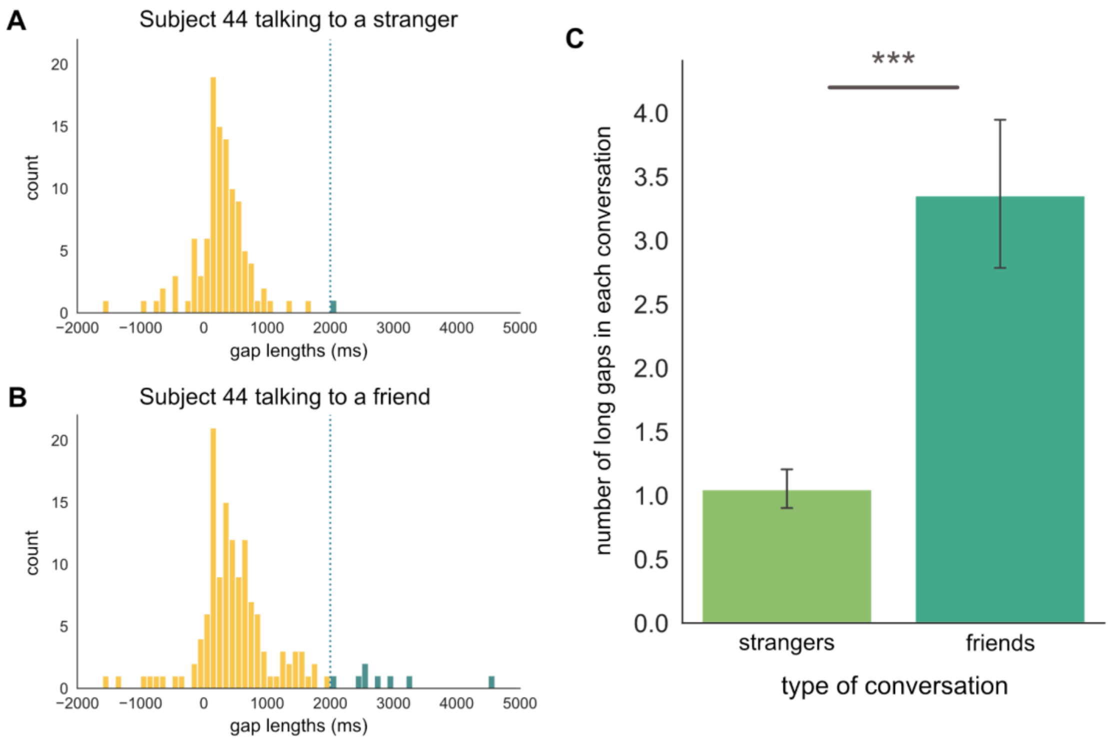


Figure 10. (A) Distributions of gap lengths from one stranger conversation. (B) Distributions of gap lengths from one friend conversation. All gap lengths over 2000ms are highlighted in green. Note there are more long gaps when subject 44 talks to their friend compared to a stranger. (C) Difference in counts of long gaps for each conversation, split by relationship type. Error bars depict 95% confidence intervals. *** p < .001

Strangers enjoy conversations less when they have more long gaps. We investigated the social consequences of these long gaps by relating counts of long gaps in stranger conversations to participants' own reports of conversation enjoyment. A linear mixed-effect model predicted each participant's rating of how much they enjoyed their conversation based on the number of long gaps in that conversation. We included the total number of gaps for each conversation as a fixed effect covariate and subject and dyad ID as a random intercept. Conversations between strangers were rated as more enjoyable when they contained fewer long gaps (b = -1.76, SE = 0.62, p = .005). Although we were not able to run this analysis in the friend dataset due to their uniformly high and invariant enjoyment ratings, we were able to leverage the continuous connection ratings to examine how connection fluctuated around long gaps between friends as well as strangers.

Changes in connection when entering and exiting long gaps. As expected, friends reported significantly higher average connection in their conversations compared to strangers ($M_{friends}$ = 75.55 (*SD* = 13.99), $M_{strangers}$ = 56.62 (*SD* = 19.54), $t(269.16) = 12.66$, $p < .001$, d = 1.11). But did feelings of connection, for either group, change when entering and exiting long gaps? Because long gaps varied in length, we temporally aligned the data by averaging connection ratings for each long gap into a single time interval. We then computed the average connection ratings at time points before and after long gaps in two second intervals. Mixed effects linear regressions modeled the temporal derivative of ratings entering and exiting long gaps, treating participants as a random effect. We found that connection ratings for friends and strangers differed when *entering* a long gap ($b = 1.03$, $SE = 0.35$, $p = 0.004$). Specifically, friends' feelings of connection increased going into a long gap ($b = 0.49$, $SE = 0.23$, $p = 0.043$), whereas strangers' ratings decreased ($b = -0.58$, $SE = 0.27$, $p = 0.038$). When *exiting* a long gap, ratings decreased significantly for strangers ($b = -0.67$, $SE = 0.22$, $p = 0.004$) with no significant difference emerging for friends ($b = -0.48$, $SE = 0.45$, $p = 0.290$). Figure 11 shows how connection ratings change over time from the first time point (i.e., 6 seconds before the long gap), for both relationship types. These findings are robust to varying the threshold for what constitutes a "long" gap (Fig S9) and to varying the length of the intervals surrounding the long gap (Fig S10). In conversations between strangers, long gaps mark moments of diminishing connection: feelings of connection markedly dip entering the long gap and remain low afterwards. For friends, long gaps mark moments of heighted connection: feelings of connection start to build, reaching a crescendo at the long gap.

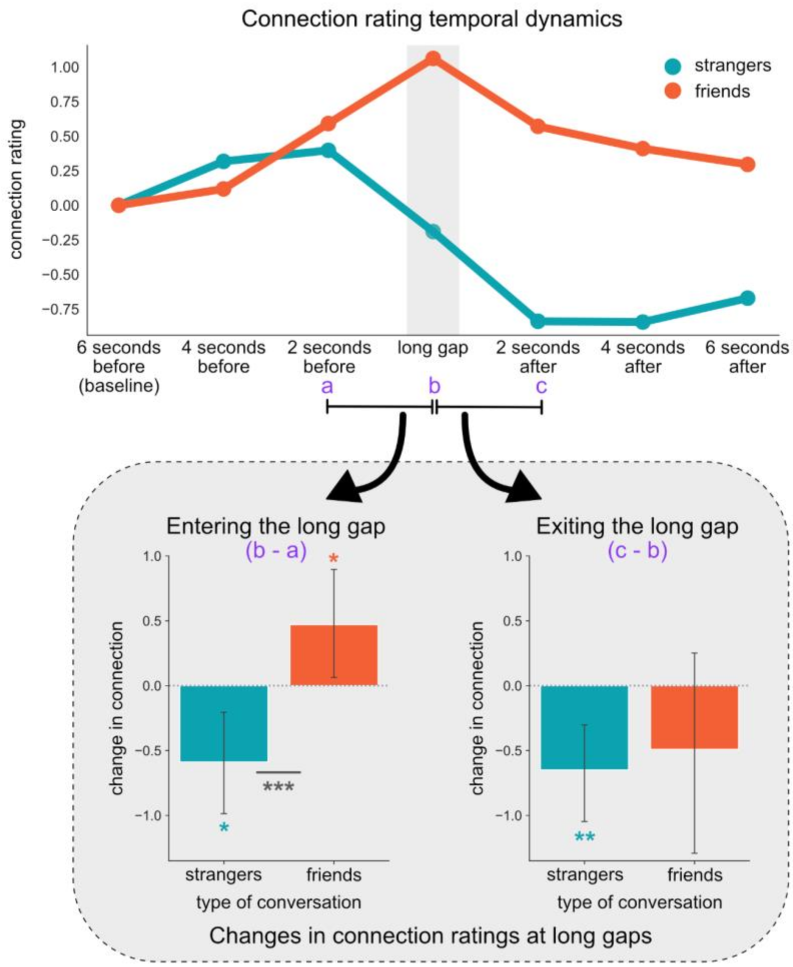


Figure 11. (Top) Depicts the average temporal dynamics of subjective feelings of connection when entering and exiting long gaps starting at an initial baseline 6 seconds prior to the gap. Trajectories are plotted separately for strangers and friends. (Bottom) Depicts the changes in connection ratings entering and exiting the long gap separately for friends and strangers. Error bars depict 95% confidence intervals. * p < .05, ** p < .01, *** p < .001

Changes in connection become stronger as gap length gets longer. We next explored whether the changes in connection ratings at long gaps we observed in the previous set of analyses might be moderated by the duration of the long gap. To test this, we re-ran the same set of analyses described above using gap length to predict the change in connection ratings entering and exiting long gaps. Gap length was log-transformed to account for the exponential distribution of the long gap data (long gaps are defined as being longer than 2s). For friends, the increase in connection when entering into a long gap was stronger for longer gaps (b = 1.66, SE = 0.77, p = 0.031, Fig S9). For strangers, increasing gap length was associated with a greater decrease in feelings of connection when exiting the long gap (b = -3.83, SE = 0.71, p < 0.001, Fig S9). These findings indicate that gap length amplifies the changes in connection ratings observed in friends and strangers in Figure 11.

Study 2

In Study 1, we found evidence that long gaps were more prevalent in friend conversations compared to stranger conversations and that long gaps diminish feelings of connection between strangers while enhancing feelings of connection between friends. In Study 2, we examined whether these differences in felt connection were apparent to outside third-party observers as well. Raters who were blind to the relationship of the conversation partners watched video clips in which long gaps occurred and rated them on a variety of dimensions (e.g., awkwardness, connection, and nonverbal communication).

Method

Independent raters viewed video clips taken from moments in the conversations that had long gap lengths (i.e., >2 seconds). After each video clip, the raters rated their impressions of the gap including: dyadic comfort (*How awkward did the gap seem? How connected did the two people seem during the gap?*), nonverbal communication (*Did any laughter occur during the gap? During the gap, did either participant seem to use any gestures with the intent of communicating something?*), and topic switches (*How closely related were the two turns* *surrounding the gap?*). See Appendix C for the complete list of questions. Raters viewed 100 video clips: 50 from stranger conversations and 50 from friend conversations. Each condition included 10 video clips in each of 5 gap length intervals: 2-2.5 seconds, 2.5-3 seconds, 3-3.5 seconds, 3.5-4 seconds, and greater than 4 seconds. Raters viewed the video clips in a random order and were not informed that clips came from two different relationship types. Detailed procedures and analysis plan for this rating task were preregistered at osf.io/ksnyj.

Video clip selection. The final stimulus set consisted of 100 video clips: 50 from each conversation type (stranger and friend). Within each conversation type, we selected 10 video clips with each of 5 gap length intervals. We used a randomization procedure designed to find clips for each conversation type and interval while also maximizing the number of unique conversations represented in the final stimulus set. This procedure thus minimized the influence on the results from any one conversation.

Each conversation clip contained the full long gap as well as 15 seconds before the start of the long gap and 15 seconds after the end of the long gap. These surrounding epochs were included so that raters could consider the context of the long gap. Raters knew that the gap began 15 seconds into the video clip. The video clips were presented in a Qualtrics survey. Each page of the survey displayed the video clip on top and the set of questions about that clip below. Raters could play the video clip as many times as they wanted to answer the questions about that particular clip. The presentation order of the video clips was randomized for each rater.

Information about raters. Three independent raters viewed and rated all 100 video clips. All of the raters were research assistants approved to be members of the research team by the Dartmouth Committee for the Protection of Human Subjects. None of these research assistants were involved in any of the original studies for which the recordings were made and all were blind to the study hypotheses. The use of research assistants allowed all video-recorded conversations to be rated as opposed to only those with video releases (minimizing potential selection effects). Before completing the rating task, raters viewed and discussed a training set of 24 clips that were not part of the final stimulus set.

Inter-rater reliability scores were computed using Cohen's Kappa for categorical questions (e.g., "Did any laughter occur during the gap?") and Intraclass correlation coefficients for continuous questions (e.g., "How awkward did the gap seem?"). The majority of questions achieved above moderate inter-rater reliability (see Table S3 for inter-rater reliability scores for each of the coded variables).

Models. We used two different approaches to investigate whether raters perceived long gaps differently based on relationship type. For continuous questions (e.g., "How awkward did the gap seem?") we ran separate linear mixed-effects models predicting each rating based on relationship type (friend or stranger), treating Rater ID as a random intercept. We report standardized regression coefficients to increase interpretability. For categorical questions (e.g., "Did any laughter occur during the gap?") a "consensus response" was established by taking the modal response across all raters. A chi-square test examined differences in responses by relationship type.

Results

Long gaps in friend conversations are perceived as qualitatively different from long gaps in stranger conversations. We found that long gaps were rated as less awkward in friend conversations compared to stranger conversations (b = 0.59, SE = 0.11, p < .001, Fig 12A) and friends were perceived to be more connected during long gaps relative to strangers (b = -0.75, SE = 0.11, p < .001, Fig 12B). This finding appeared to be amplified as a function of the gap length (as indexed by the 5 interval bins). We found a significant interaction between relationship type and gap length on ratings of awkwardness (b = 0.47, SE = 0.10, p < .001, Fig 12C), indicating that perceptions of awkwardness *increased* with gap length more for strangers compared to friends. Similarly, a significant interaction between relationship type and gap length on ratings of connection indicate that perceptions of connection *decreased* with gap length more for strangers compared to friends (b = -0.25, SE = 0.10, p = .017, Fig 12D).

We also found evidence that long gaps serve different purposes depending on relationship type. For example, strangers were more likely to switch topics after a long gap compared to friends (b = -0.29, SE = 0.11, p = .011). Whereas a long gap between strangers may create awkwardness and an impetus to change topic, a long gap between friends may serve as a moment to reflect on what was just said. Friends' long gaps were more likely to contain laughter than strangers' long gaps, X^2 (1, N = 100) = 6.05, p = .014) and when laughter did occur, it was perceived as being more genuine (b = -0.48, SE = 0.19, p = .011) compared to strangers. This suggests that the laughter of friends is a genuine response to conversational content whereas the laughter of strangers may be an act of politeness to fill time. (See Table S4 for the effect of condition on every variable measured.)

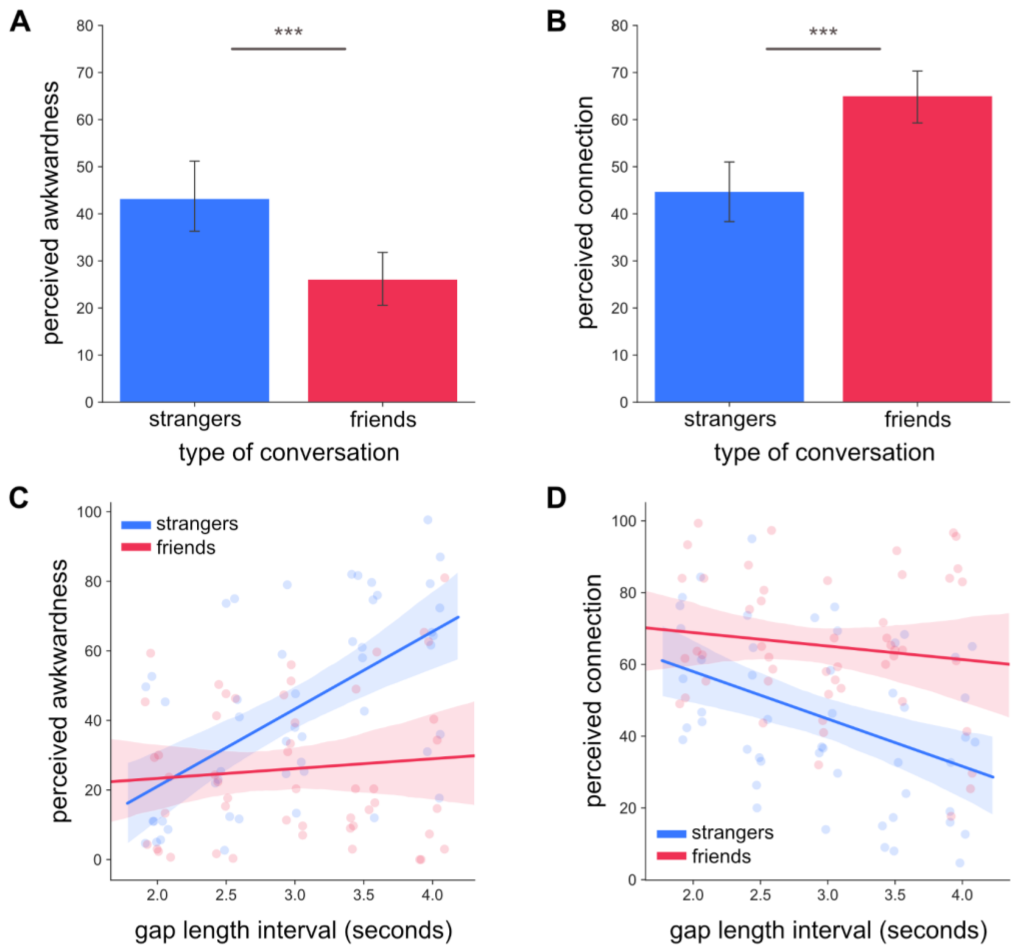


Figure 12. (A) Difference in ratings of awkwardness during moments of long gaps in stranger conversations and friend conversations. (B) Difference in ratings of connection during moments of long gaps in stranger conversations and friend conversations. (C) Effect of gap length interval on ratings of awkwardness split by relationship type (stranger vs friend). (D) Effect of gap length interval on ratings of connection split by relationship type. Lines show linear regression model fit. Jitter was applied to show individual data points; however, each data point belongs to one of the five interval bins. All error bars depict 95% confidence intervals. *** $p < .001$

Discussion

For friends and strangers alike, short gaps are a heuristic for connection: the shorter the gap, the more connected people feel to their relationship partner (Templeton et al., 2022). Here we show that the inverse—the longer the gap, the less connected people feel—is only true for strangers. We found that friends had more instances of long gaps compared to strangers and that long gaps were the site of *increased* connection.

We defined gaps simply in terms of the length of time between verbal speech turns. This definition benefits from being easily computable from an audio file or transcript and therefore facilitates reproducibility and portability to a variety of contexts. It is important to note that the absence of speech in these gaps does not imply the absence of communication. On the contrary, these gaps could contain non-verbal vocalizations, gestures, or postural changes. Further, it is likely that what happens *within* these gaps is illustrative of the particular meaning or context of that gap and that there may be several meaningful subtypes. For example, previous research has defined a "lapse" as a moment when all participants forgo their turn to speak (Hoey, 2015; Sacks et al., 1974). It is possible that "lapses", as so defined, are gaps that are particularly detrimental for connection. Other gaps may be marked by genuine laughter, with positive consequences for connection. Still other gaps may contain postural changes indicating reflection, and so on.

Findings from Study 2 provide some hints as to how long gaps function differently between friends and strangers. Long gaps between strangers contained less laughter overall and less authentic laughter than long gaps between friends. Long gaps between strangers were also much more likely to be followed by a change in topic, compared to friends (see also Fig S11). These findings suggest that long gaps prompted strangers to cast around for something new to say. In contrast, long gaps between friends provided spaces for reacting and reflecting on what was just said. Outside observers also perceived the long gaps of friends as less awkward and more connected compared to the long gaps of strangers—a finding that mirrored the connection ratings of the conversation partners themselves. These findings add critical nuance to previous assertions that long gaps in conversation uniformly signpost trouble (Jefferson, 1989; Kendrick

& Torreira, 2015; Roberts et al., 2006, 2011; Smith & Clark, 1993). Our results indicate that this is only true in conversations where people are getting to know each other.

The present study further illustrates the importance of expanding interaction research beyond the context of strangers. For most of human history, people have lived in communities in which familiar others are their modal conversation partners (R. Dunbar, 1998). Even in modern, WEIRD (Henrich et al., 2010) cultures in which stranger conversations are not infrequent, people prefer to spend the majority of their social lives with friends and family (Kahneman et al., 2004). In contrast, the modal interaction in communications research is that of strangers. This reliance on stranger dyads may lead to an incomplete, if not distorted, understanding of conversational dynamics. As one example, we show that long gaps are associated with markedly different feelings of connection for friends compared to strangers. Long gaps between strangers are a sign of disconnection, and increasingly so the longer they endure. Long gaps between friends signal heightened social connection, regardless of their duration. It is important to understand how conversational dynamics differ between contexts, how they evolve as relationships grow, and how they may signal relationship health (Hadley et al., 2022). A fuller understanding of these dynamics will help paint a more accurate picture of what intimacy looks and sounds like.

For people with a shared history, such as close friends, long gaps may simply be times when communication travels “inside the head” as when reflecting on what was just said or mutually savoring past experiences. This can be triggered by a simple word or phrase (“remember Paris?”). In these instances, a loss of words does not mean a loss of connection or even communication. Because of this, long gaps may not be experienced as gaps at all. The long gaps we remember are instead likely to be ones we enter clumsily and fail to exit gracefully. Such a bias in memory may explain the intuitive yet mistaken assumption that long gaps are uniformly negative. We hope that this work will spur future research that looks more carefully at how features of friends' conversations differ from strangers and how these differences contribute to their social consequences.

Collectively, these studies suggest that long gaps function differently between strangers and friends. For strangers, long gaps are moments of dead air—awkward silences followed by swift changes in topic. For friends, long gaps may not be accurately described or experienced as “gaps” at all. Though devoid of words, the long gaps of friends appear to be full of meaning, providing natural moments for reflection and expression. These differences between the long gaps of strangers and friends are apparent to outside observers: while the long gaps of strangers are hard to watch, the long gaps of friends telegraph connection. These studies add to a growing literature showing that features of conversation change based on shared history and social context (Brennan & Clark, 1996; Garrod & Doherty, 1994; R. D. Hawkins et al., 2021; Holler & Wilkin, 2009; Stolk et al., 2014). Gaps between turns carry meaningful social consequences, and those consequences change with friendship.

Chapter 2: Conversational launch pads help strangers start their conversations

Introduction

My grandmother spent most of her life in Connecticut and she would always say, "If you don't like the weather in New England, just wait a few minutes". Apparently Mark Twain said it first (Twain, 1876), but my grandmother definitely said it more often. Having lived in New Hampshire for the past seven years, I often find myself telling people about this saying and how I learned it from my grandma. I do this so much because a lot of people start their conversations by talking about the weather (myself included)!

Talking about the weather gets a bad rap. People seem to think if they are talking about the weather, they have failed some conversational test. In fact, it seems like a reasonable route to conversation success. I find that I can talk to anyone about the weather. It works whether I'm passing someone in a hallway or sitting down for an hour-long interview. It connects people over Zoom calls ("How's the weather over there?") and generational gaps ("It never used to be this warm"). Not only is talking about the weather an easy and accessible opener, I find that it often leads to much more interesting places. When I tell someone that the 40-degree Fahrenheit weather feels freezing to me, they might respond by explaining that they grew up in Wisconsin and this is actually their ideal weather. Now that I know they are from Wisconsin, I can tell them all about my lab mate Chris who also grew up there. We're off to the races! I can tell a completely different person the same thing (40-degrees, poor me) and they might confess that they've never lived in a cold place before and they are terrified about what winter will bring. Now we start talking about the importance of layering and how to take advantage of all the snow activities that Dartmouth offers. The exact same opening line led to two completely different conversations. Starting with a “boring” topic, like the weather, can give people an easy way to build rapport and find common ground.

As much as I’m an advocate for talking about the weather, I relish the times when I don’t have to rely on it as a conversational crutch. I can walk across the hallway into the Chang Lab, skip all formalities, and ask a graduate student, “What did you all think of that talk?”. I can meet a friend for coffee and immediately upon seeing them say, “You’ll never guess who I ran into at the grocery store”. When I visit my mom at her home in Tennessee, the first thing I might say to her is “oh my god, the couch looks great there!”. These are wildly different ways to start conversations, tailored for the specific person I am talking to and the relationship I have with them. There are certain things that I talk about with my lab, other things that I talk about with friends from college, and a different set of things I talk about with family members.

Every conversation needs to start somewhere, and that choice depends on who you are talking to (Haas & Sherman, 1982). People are more likely to talk to their friends and family about their relationships and more likely to talk to acquaintances about the news (Bearman & Parigi, 2004). When people talk to someone new, they build rapport by searching for common ground (Cassell et al., 2007; Jucker & Smith, 2022; Tickle-Degnen & Rosenthal, 1990). How do strangers best position themselves to start this process?

In this paper, we investigate the intuition that people start their conversations similarly when they talk to strangers, compared to friends. To do this, we leverage a previously collected dataset of face-to-face conversations between dyads of strangers and dyads of friends. We first investigate how the semantic similarity of these conversations varies between groups and over time. We next investigate which topics strangers use to start their conversations. All participants in our dataset are undergraduate students on the same campus, so while they may not rely on the weather, we expect to find a few distinct sets of topics that jump-start their conversations. We then examine whether those starting topics are particularly well-suited to transition into other topics by creating a topic network based on the topic transition structure of the group. We introduce the term "conversational launch pad" to define topics that have the tendency to branch into many different topics, just as the weather did in my example above. It may be no accident that people choose conversation launch pad topics to start their conversations. They can propel people into more interesting places, increasing the likelihood of finding common ground.

Methods and Results

Datasets

Stranger Dataset. Participants in the stranger dataset participated in exchange for extra credit in their Psychology or Neuroscience courses. Conversation partners were assigned by an experimenter. To ensure that participants did not know each other we asked them "How well did you know your study partner before today?" (0 = Not well at all, 50 = Moderately well, and 100 = Extremely well) in a survey following their conversation. In order to limit our analyses to true strangers who do not know each other, we excluded 61 dyads where both dyad members indicated a response greater than 0 on this question. The analyses reported in this paper come from 261 stranger dyads.

Friend Dataset. All participants in the stranger dataset were invited to participate in the friend dataset. Those who were interested were asked to nominate their close friends to participate with them. Participants in this study had the option of receiving either cash compensation or extra credit in eligible courses. We recorded 65 conversations between dyads of friends.

Conversation Sessions. Every conversation session began with two participants having a 10-minute unstructured conversation. Participants were seated across from each other at a cafe table. A webcam attached to a desktop computer across the room captured both participants in profile. After the recording was started, the experimenter turned off the desktop screen so that participants would not be distracted by the recording during their conversation. Participants were told that they could talk about whatever they wanted. After 10 minutes, the experimenter re-entered the room, ending the conversation.

After their conversation, participants were separated into different rooms where they privately completed two tasks. They first rated their overall impressions of the conversation via a survey. They then watched a video recording of their conversation while continuously rating how connected they remembered feeling to their conversation partner at each moment in time. Participants made these ratings by using a computer mouse to move an on-screen slider bar (from 0 = None to 100 = Very). The position of the mouse was recorded every 100 milliseconds.

The video recordings of each conversation were transcribed by an external transcription company. Each transcript contained a start and end timestamp for each speech turn as well as the text of what was said.

Do strangers start their conversations more similarly to each other, compared to friends?

We first tested our hypothesis that strangers would start their conversations more similarly to each other, compared to friends. The general intuition is that conversations between any two sets of stranger conversations are likely to seem much more similar compared to any two sets of friend conversations. We examined this in our datasets by computing all possible pairwise semantic similarity between each group (friend and stranger), over time.

Methods. Although all conversations in our datasets were exactly 10-minutes long, conversations move at different paces with different turn-taking dynamics. To facilitate comparisons across conversations, conversation transcripts were binned into 30-second increments (20 bins per conversation). Each bin contained the text of the speech turns that occurred in that 30-second window. For example, bin 1 contained all the turns that occurred in the first 30 seconds, bin 2 contained all the turns that were said in the second 30 seconds, and so on. This approach ignores speaker identity; if both participants spoke in a 30-second window, both of their turns would be included in that bin. For the purposes of this project, the *conversation* is the unit of analysis, not individual speakers.

We next used the Universal Sentence Encoder to represent the semantic meaning of the text in each bin by transforming it into a 512-dimensional language embedding (Cer et al., 2018). For each bin, within each group (stranger and friend), we computed all pairwise cosine similarity values between the language embeddings. To summarize how similarly each group's conversation was in each bin, we computed the median of the pairwise values. Confidence intervals were computed using subject-wise bootstrapping with 5,000 samples (Chen et al., 2016). Finally, we compared differences in the semantic similarity between each group at each bin using subject-wise permutation (Chen et al., 2016).

Results. The semantic similarity for strangers was significantly higher than friends for 17 out of 20 bins (all ps $< .05$). This difference was highest in the first bin (difference = 0.14, $p < .001$), and second highest in the second bin (difference = 0.08, $p < .001$). Put simply, semantic similarity for strangers is higher than for friends, and that difference is highest at the start of conversations (Fig 13). This effect is robust to bin size and different language models (Fig S12). The findings are consistent with our thinking that strangers start their conversations in a more limited set of ways, compared to friends who can jump into conversations in ways that are idiosyncratic to their particular relationship. Over time, however, strangers start to differentiate their conversations from each other.

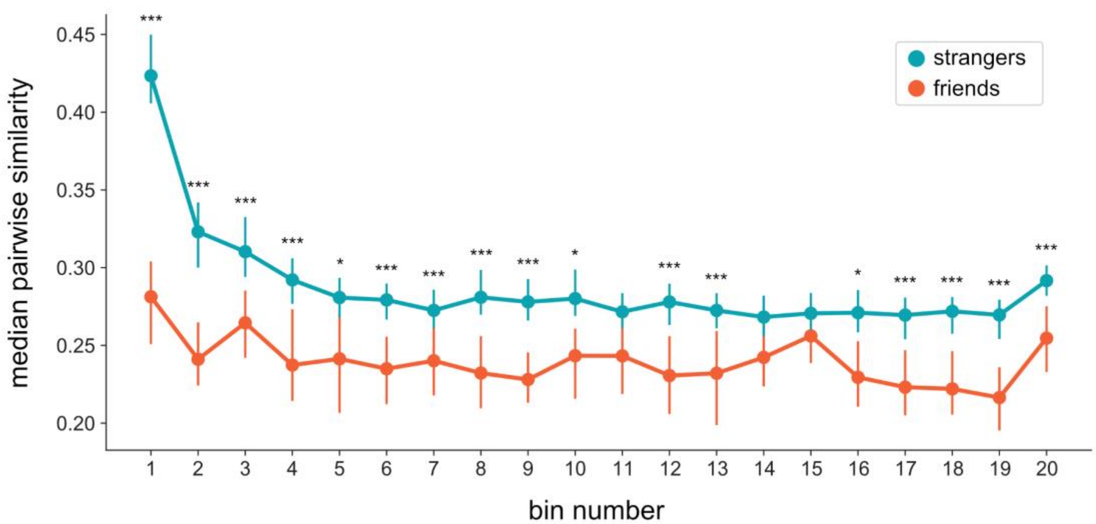


Figure 13. Median pairwise cosine similarity between Universal Sentence Encoder embeddings for each bin of text in stranger and friend conversation transcripts. Each bin represents all the text from the turns spoken in each 30-second interval of a 10-minute conversation. 95% confidence intervals were computed using subject-wise bootstrapping with 5,000 samples. * $p < 0.05$, *** $p < 0.001$

What topics do strangers use to start their conversations?

The previous analysis demonstrated that strangers start their conversations more similarly to each other, and that similarity decreases over time. Language models like the Universal Sentence Encoder are well-suited to describe semantic similarity between text, but the meaning of individual features in the embedding space are not interpretable on their own. Our next step was to cluster the embeddings space to create interpretable topics that we can use to explore *what* strangers talk about.

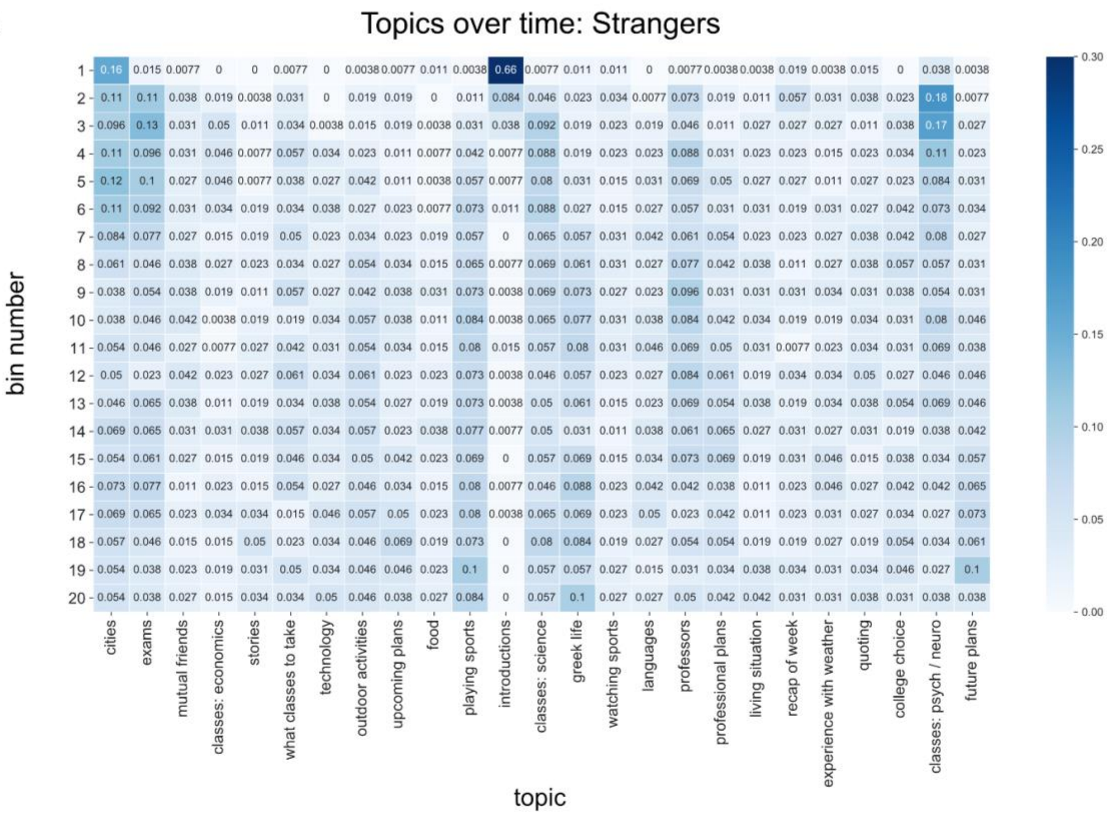
Methods.

Defining topics. We first reduced the dimensionality of the embedding space using uniform manifold approximation and projection (McInnes et al., 2018), a dimensionality reduction technique that aims to preserve distances between observations. We then used-- k-means clustering to divide this reduced space into meaningful clusters. We do not know the "true" number of topics in our datasets (assuming such a thing exists) but nonetheless the clustering analysis requires that we choose a number of topics up-front. Here, we report a clustering solution that first reduces the language embedding feature space to 10-dimensions (Assent, 2012) and then applies a clustering solution with 25 clusters. See the Supplement for a detailed explanation about how each of these choices were made.

We next assigned each bin of text in each conversation to a single topic. So that we can talk about different topics in terms of their *content* (as opposed to an arbitrary cluster number) we labeled each topic to reflect the themes that emerged by inspecting the text of the bins assigned to each topic. See Table S5 for more detailed descriptions of each topic as well as examples of text assigned to each topic.

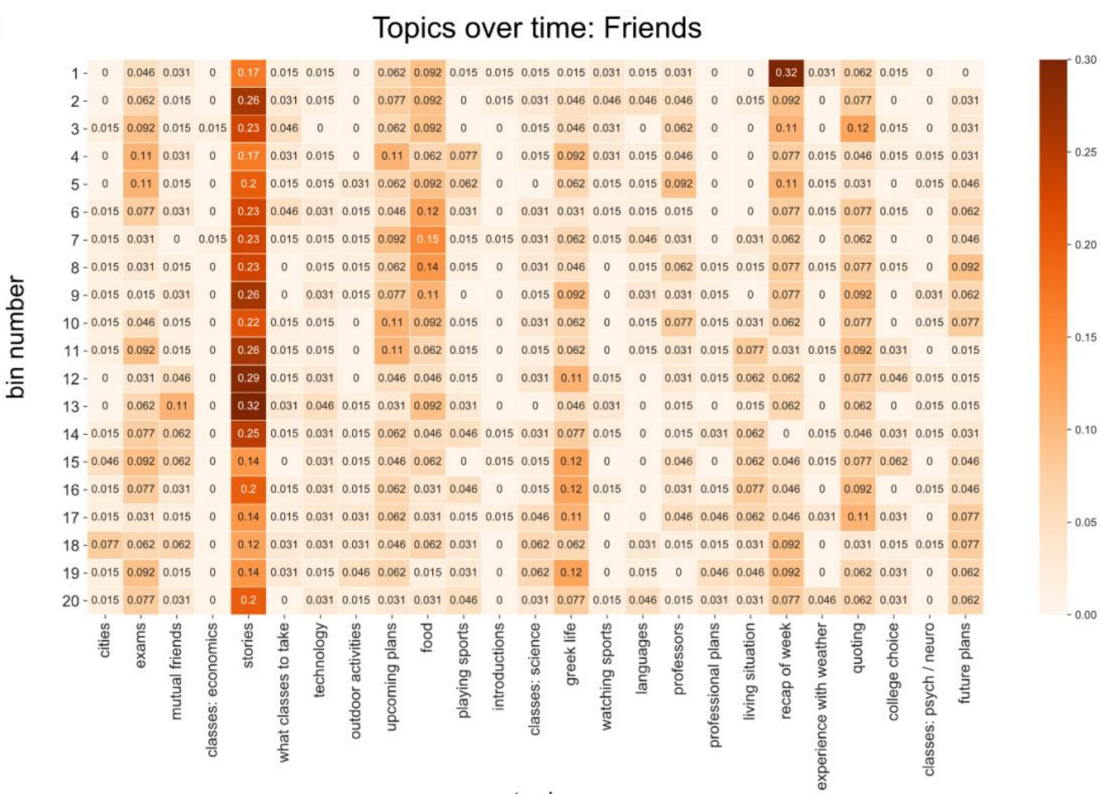
We applied clustering to all bins of text at once; the clustering procedure was agnostic to which group the text came from (friend or stranger) and which time bin it came from. This allows us to compare strangers and friends along the same set of topics. To examine how the topics that people talk about change over time, we computed the proportion of dyads in each topic for each bin, separately for strangers and friends (Fig 14).

A



Topics over time: Strangers
bin number
topic
cities
exams
mutual friends
classes: economics
stories
what classes to take
technology
outdoor activities
upcoming plans
food
playing sports
introductions
classes: science
greek life
watching sports
languages
professors
professional plans
living situation
recap of week
experience with weather
quoting
college choice
classes: psych / neuro
future plans
0.00
0.05
0.10
0.15
0.20
0.25
0.30

B



Topics over time: Friends
bin number
topic
cities
exams
mutual friends
classes: economics
stories
what classes to take
technology
outdoor activities
upcoming plans
food
playing sports
introductions
classes: science
greek life
watching sports
languages
professors
professional plans
living situation
recap of week
experience with weather
quoting
college choice
classes: psych / neuro
future plans
0.00
0.05
0.10
0.15
0.20
0.25
0.30

Figure 14. Proportion of dyads in each topic for each time bin for conversations between strangers (A) and conversations between friends (B). The proportions are annotated in each cell. Colormaps are capped at 0.3 to visually preserve differences between cells by reducing the influence of the bin 1 'introductions' topic in the stranger dataset

Identifying topics that appear early vs late. Now that we can represent each bin as a topic, we investigated the topics that strangers are more likely to *start* their conversations with. To do this, we averaged the proportions for each topic for the first half of the conversation (bins 1-10, Fig 14) and the second half of the conversation (bins 11-20, Fig 14). We then computed a difference score by subtracting the second half values from the first half values. This gives an indication of how likely that topic was to be utilized at the start of a conversation compared to the end (Fig 15). We were specifically interested in the topics that strangers choose to *start* their conversations with; topics that were used at a consistent rate throughout the course of a conversation would not come out in an analysis like this.

Results. As seen in Figure 15, the topics most likely to appear in the first half of a stranger conversation compared to the last half were: introductions, classes: psych / neuro, cities, and exams. These topics are sensible given our population of Dartmouth students who are all participating for extra credit in their psychology and neuroscience courses. They know that they all have psychology and neuroscience courses in common and, even if they are not in the exact same class, they all have exams. In the 'cities' topic, people are generally talking about where they grew up and how their hometown compares to life at Dartmouth. As seen in Figure 14A, the 'introductions' topic is unique in appearing strongly for the first 30-second bin and essentially disappearing afterwards, perhaps suggesting it is fundamentally different from the other early topics.

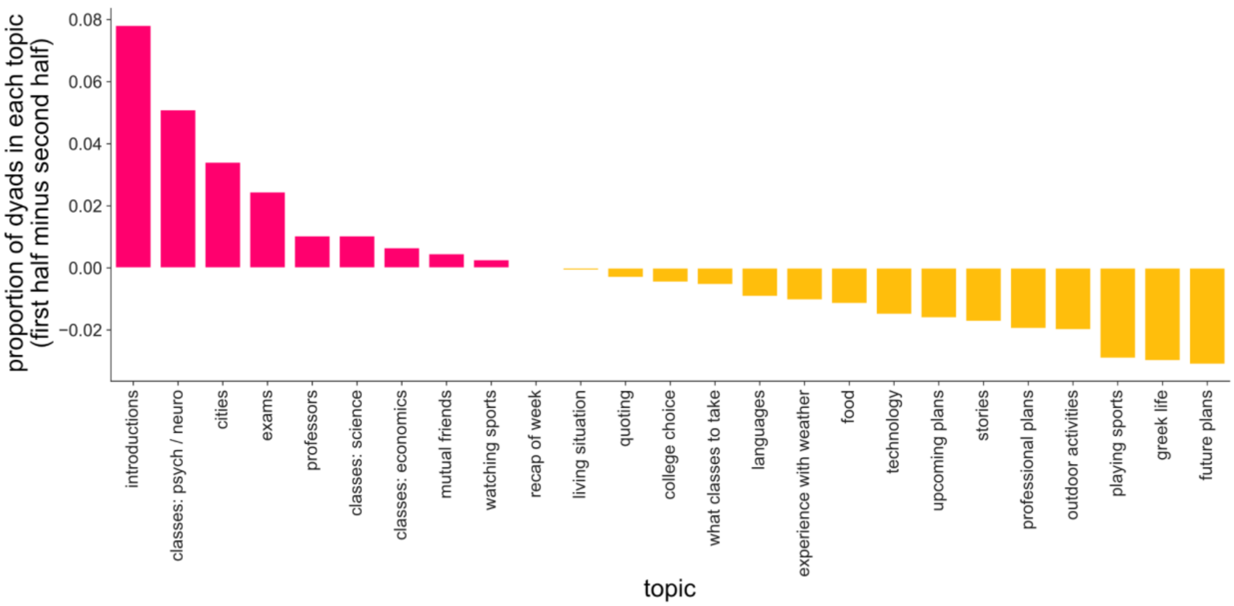


Figure 15. Average proportion of stranger dyads in each topic for the first half (bins 1-10) minus the second half (bins 11-20). Topics in pink were more likely to be used in the first half, compared to the second half. Topics in yellow were more likely to be used in the second half, compared to the first half.

Do starting topics of strangers confer any advantages?

Is there a reason that strangers in this dataset tend to start with these particular topics? We investigated whether certain topics have any structural advantages, in terms of connecting to other topics. We expected that some topics would be better suited to transition people into different sets of topics, perhaps increasing the likelihood of two strangers finding common ground.

Methods. For each stranger conversation, we counted the number of times that people transitioned from one topic to another across their 20 bins. These counts were represented by a topic transition matrix (Fig 16A). This approach is agnostic to the timing of different bins (e.g., early vs late in a conversation), it only records topic transitions between consecutive bins. We then represented this matrix as a weighted, directed graph with each topic acting as a node in the network (Fig 16B). We computed six different node metrics for each topic in the network: in-degree, out-degree, out-degree minus in-degree, eigenvector centrality, betweenness centrality, clustering coefficient. We then clustered these vectors of node features using k-means clustering to discover topics that have similar network properties. The Elbow Method, which is used to select the number of clusters that best minimizes the within-cluster sum of squares (Cui, 2020), revealed that these data are well described by 4 clusters (Fig S14).

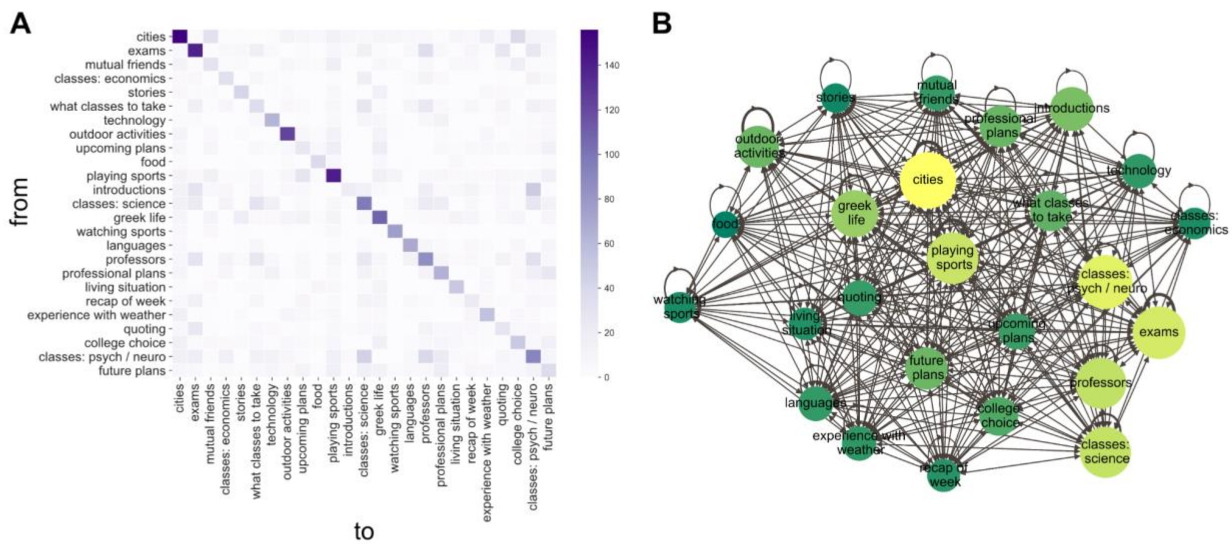


Figure 16. (A) Topic transition matrix for stranger conversation. Each cell represents the number of times there was a transition from the topic on the y-axis to the topic on the x-axis. Darker purple indicates a higher count. (B) Network graph based on the topic transition matrix. Node size and color is based on weighted out-degree. Some topic labels have been shortened for readability.

Results. The topics that were grouped together in each cluster are detailed in Figure 17. As hinted at earlier, the 'introduction' topic is in a cluster by itself because there is no other topic like it. That topic appeared strongly in bin 1 and essentially never appears again (see Fig 14A). Most interestingly, Cluster 1 lists the rest of the topics that are more likely to appear early vs late (see Fig 15). Two additional topics, 'playing sports' and 'greek life' also get included, suggesting that those topics have similar network functions despite tending to appear later in conversations between strangers.

To explore how these four clusters differed from each other, we ran regressions with cluster number predicting each of the six node metrics, separately (Fig 17B). Results show that Cluster 1 is particularly high on out-degree ($F(3, 21) = 86.77$, $p < .001$, $R^2_{\text{Adjusted}} = 0.914$), in-degree ($F(3, 21) = 143.86$, $p < .001$, $R^2_{\text{Adjusted}} = 0.947$), and eigenvector centrality ($F(3, 21) = 58.42$, $p < .001$, $R^2_{\text{Adjusted}} = 0.947$), compared to the other clusters.

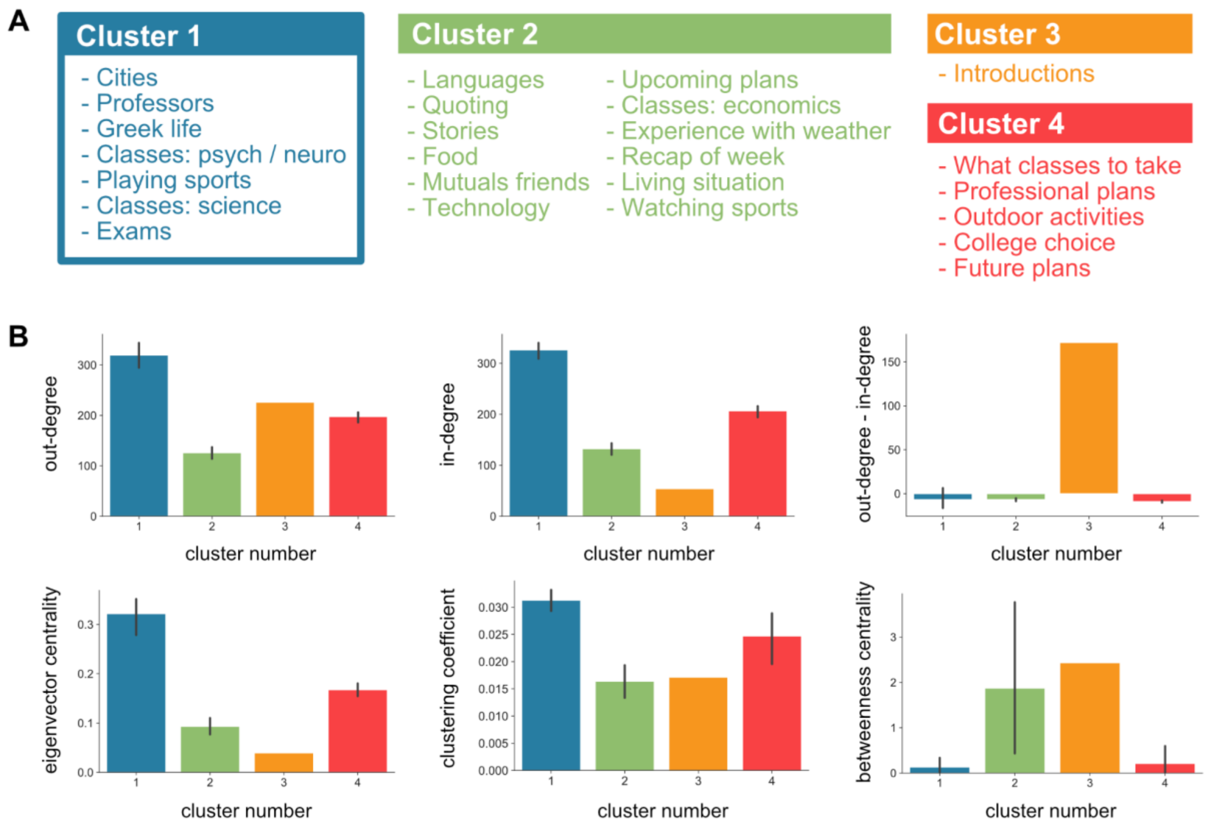


Figure 17. (A) Topics are clustered based on six different node metrics. Here we list the topics that get assigned to each cluster. Cluster 1 contains the candidate "launch pad" topics as most of them appear early (vs late) in stranger conversations. (B) How topics in different clusters vary across the six node metrics. Launch pad topics (Cluster 1) have relatively high in-degree, out-degree, and eigenvector centrality.

Are there any social consequences to using launch pads?

We believe that launch pad topics are used to propel people into more interesting conversational spaces. It follows that something about their use should relate to participants' own reports of how much they enjoyed their conversation and felt connected to their partner.

Methods. We used the topics in Cluster 1 as our list of launch pad topics. For each stranger conversation, we counted the number of bins that contained a launch pad topic. We ran linear mixed-effect regressions with number of launch pad topics predicting enjoyment and connection. We included Subject ID and Conversation ID as random intercepts. We hypothesized that strangers who used launch pad topics *less* would have more enjoyable and connected conversations.

We also computed the number of unique topics that strangers used across their 20 bins. We expected that strangers who were able to find a topic they both enjoyed would stay there, reducing the number of topics they needed to explore. We ran linear mixed-effect regressions with number of unique topics predicting enjoyment and connection. We included Subject ID and Conversation ID as random intercepts. We hypothesized that conversations with fewer unique topics would be rated as more enjoyable and connected.

Results. We did not find a significant relationship between number of bins with launch pad topics and enjoyment (b = 0.04, SE = 0.05, p = .406) or connection (b = 0.05, SE = 0.04, p = .295). We did find that strangers who used fewer topics felt more enjoyment (b = -0.25, SE = 0.05, p < .001) and connection (b = -0.18, SE = 0.04, p < .001, Fig 18). A successful conversation between strangers may be one where people find something they both want to talk about. This is what friends are able to do quickly (Planalp & Benson, 1992). Using conversational launch pads might be an efficient way for strangers to discover and transition into mutually preferred topics.

An exploratory analysis relating these two variables found that conversations with fewer unique topics also had more bins with launch pad topics (b = -0.09, SE = 0.04, p = .037). This raises the possibility that launch pads *themselves* might be appealing destinations.

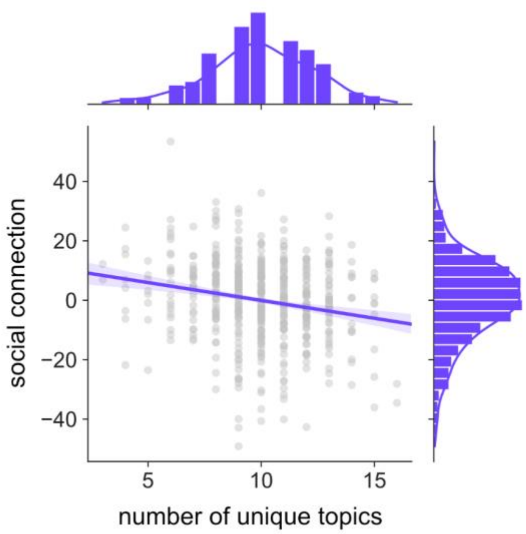


Figure 18. Conversations between strangers have higher connection ratings when they have fewer unique topics. Connection ratings are centered within-subject to reflect the random effect structure used in the mixed-effects model. Individual data points are displayed as gray dots. The line represents a regression model relating the number of unique topics and felt connection. The distribution of number of unique topics is plotted above the scatterplots, and the distribution of connection ratings is plotted to the right of the scatterplots.

Discussion

This project explores how people start brand new conversations with people they have never met before. We found that strangers start their conversations more similarly to each other, compared to friends. In our population of undergraduate students, strangers start their conversations with topics related to school (exams and classes) and about each other (names and hometowns). As conversations progress over time, the semantic similarity between strangers drops and the number and type of topics becomes more varied. Indeed, typical starting topics might have useful transition properties that allow this to happen. We introduce the term "conversational launch pads" to describe topics that serve to launch conversation in many different directions, perhaps increasing the likelihood of finding common ground. The fact that these topics also have high in-degree suggests that they may serve as safe havens for people to return back to. Of course, launch pad topics can sometimes be fruitful places to stay put; when dyads strike upon common ground right away they can grow roots there. Launch pad topics offer many different routes to conversational success, making them a great place to start.

We expect that different populations will utilize different topics as conversational launch pads. In our dataset of undergraduate students, many of those topics center around school, something they all have in common. To push this program of research further, we plan to perform similar analyses in different populations to identify other candidate launch pad topics. We think of good launch pads as being the "lowest common denominator" of topics in a given population. If people can assume they all have something specific in common, it makes sense to start conversations there. We predict that the less information people have about each other, the more general their starting topic should be. Certain topics are always available (e.g., weather) and others rise and fall based on world events, like talking about a virus during a global pandemic (Reece et al., 2023).

We think this preliminary work on launch pads can speak to a mystery in social psychology: If people feel better engaging in deep talk, why do they persist in small talk?

(Kardas et al., 2022; Mehl et al., 2010). It can be risky, and potentially off-putting, to start a conversation too deep (Collins & Miller, 1994). In the absence of an experimental manipulation, *how* do people go deep? We think conversational launch pads offer a safe way for people to (i) narrow down the scope of possible deep topics they might want to get to and (ii) decide if this is someone they *want* to have deep talk with. Starting broad affords people options and ample opportunity to find something they both want to talk about.

Initial analyses suggest that people feel more connected when they persist in a single topic rather than jump around many different topics. Presumably, connected dyads land in a conversation topic both people enjoy and can therefore happily stay put. Though topic analyses get at the *content* of what is being said, other linguistic analyses can get at the *style*. It is possible that as dyads persist in a single topic, the way they talk about it changes. For example, people may start off talking superficially about a single topic and progressively get deeper as time goes on. For dyads that tend to stick to one topic, it would be interesting to compare the language in bins assigned to the same topic based on whether they appear early vs late in a conversation.

Throughout this paper, we have suggested that conversational launch pads are used to transition people into topics where they might be able to find common ground. This is a hypothesis that we have not yet explicitly tested, but it is one that we plan to test soon. For example, we could leverage the moment-by-moment connection ratings in these datasets to test whether connection ratings increase in the bin following a launch pad topic. To better examine how different starting topics influence the trajectory of a conversation, we could run simulations on our transition networks to quantify how different starting topics result in more or less topic exploration. These simulations may reveal conversational "black holes," or topics that are nearly impossible to get out of. Finally, we hope to run follow-up experimental studies where participants are instructed to start their conversations with established launch pad topics vs topics that do not act as launch pads. In these studies, we would measure how dyads transition into different topics based on where they are instructed to start as well as how that decision impacts their enjoyment and connection.

Many of our analyses focused on conversation *topics* that were the result of specific clustering parameters. It will be important to show that our findings are robust to these clustering decisions. However, different clustering decisions will necessarily result in different sets of topics. We expect different clustering solutions to show that strangers tend to start their conversations by talking about school, but those exact topics may be more or less granular than the topics presented here. Of course, the number of clusters is not the only decision that impacts clustering results. Here, we made the decision to cluster stranger and friend data *together*. This allowed us to compare them to each other directly, but did result in some interesting patterns. Most notably, the bulk of friend bins got categorized as the 'stories' topic (Fig 14B). If we performed clustering for friends separately from strangers we would likely see that topic split apart into smaller sub-topics and others, like 'classes: economics,' disappear completely. In the future, we plan to use our transcripts to fine-tune these language models, which are largely trained on written (not spoken) text. We hope this process will result in even cleaner topic clusters, tailored to our specific sample. In sum, the approach of using unsupervised clustering to generate topics requires setting parameters in advance and those decisions need to be made thoughtfully and carefully. Still, we believe our approach to be an improvement upon older work where researchers decided lists of topics on their own, as these decisions can bias the kinds of claims that get made (Bischoping, 1993; R. I. Dunbar et al., 1997; Landis & Burtt, 1924).

Another big decision we made was to bin the conversation transcripts. Every conversation was divided into 20, 30-second bins. This was an important step that allowed us to compare conversations to each other over time. However, in the future, we would like to complement this approach by looking at conversations on a turn-by-turn basis. This would allow us to search for participant-driven topic changes, rather than imposing topic changes at 30-second bins. It would also have the advantage of allowing us to examine the effect of different speakers, rather than collapsing across speakers as we do now.

Every conversation needs to start *somewhere*. Here, we find evidence for the fact that strangers start their conversations in similar ways. Further, the topics they tend to start with may be particularly well-suited to act as conversational launch pads, with the ability to propel the conversation forward into a myriad of different directions. Talking about the weather is not something to mock, but rather a sensible place to start. What matters is where we go from there.

Chapter 3: The spontaneous use of insider language in conversation

Introduction

Inside jokes are an efficient way to use language. By saying a word or phrase that references a shared experience with someone else, it is possible to unfurl a whole story in another person's head without ever needing to explain it. However, these jokes can only land if the person hearing it understands the necessary context. An inside joke told to someone on the "outside" would immediately flop. We call this particular use of language—when words carry specific meaning between some people but not others—*insider language*.

Insider language is an efficient shorthand where a handful of words communicate a complex idea. This shorthand involves a collaborative process of referencing (H. H. Clark & Wilkes-Gibbs, 1986). When one person uses a reference, they look for evidence that the person they are talking to understood what they meant (H. H. Clark & Brennan, 1991). For example, if my sister said to me, "Remember that time you had a meltdown as a kid?" I might seek clarification by replying, "You mean about the ketchup?" which my sister would then confirm with, "Yes, that was such an overreaction." Once an agreement is reached, this reference can be used repeatedly and often gets simplified even further (Brennan & Clark, 1996). Now, when my sister says "ketchup meltdown" to me, I know exactly to which childhood memory she is referring.

Laboratory studies demonstrate that dyads develop insider language over repeated interactions (H. H. Clark & Wilkes-Gibbs, 1986; R. X. D. Hawkins et al., 2017; Krauss & Weinheimer, 1964). In repeated reference games, one person directs their partner to complete a task over and over again. For example, in the Tangram Task, a director needs to describe a set of novel shapes so that their partner can arrange the shapes in a pre-specified order, that changes from round to round. Importantly, the director and partner cannot see each other. In the first iteration of the game, participants must use a lot of words to communicate effectively enough to complete the task (e.g., “upside-down martini glass in a wire stand”). However, as they repeat the task over and over again, they are able to use fewer words (e.g., “martini”) because they have collaboratively established a set of insider references —a kind of shared shorthand—for different aspects of the game (examples from (R. X. D. Hawkins et al., 2017)). Crucially, these references are often peculiar to the dyad who created them, making them relatively unintelligible to new conversation partners. People understand this and have a remarkable ability to keep track of which insider references go with which partners (R. D. Hawkins et al., 2021; Metzing & Brennan, 2003; Wilkes-Gibbs & Clark, 1992).

There are several potential explanations for how dyads can to coordinate with each other to develop insider references over the course of repeated reference games. Follow-up experimental studies demonstrate that directly interacting with someone else in conversation builds common ground, which allows for the development of shared perspective that can become the basis for establishing successful shorthands (H. H. Clark, 1996a; Wilkes-Gibbs & Clark, 1992). Neuroscience work demonstrates that the creation of this common ground can even synchronize brain activity between pairs of people, particularly during times when they are using novel signals (Stolk et al., 2014). Findings from computational modeling suggest that pair-specific references are likely to develop for concepts that are both arbitrary (can be described in many different, idiosyncratic ways) and stable (descriptions tending to persist over longer timescales) (R. X. D. Hawkins et al., 2017). Finally, studies show that people with Autism have difficulty developing shared references in communicative games, compared to matched neurotypical controls (Wadge et al., 2019), suggesting that conceptual alignment with another person (a prerequisite to insider language use) is critical for maintaining smooth social interactions.

In this paper, we are interested in insider language that naturally develops over the course of a relationship (e.g., friendship). In a sense, friends naturally engage in repeated reference tasks, with every conversation providing an opportunity to build common ground and establish references peculiar to that relationship. To take our earlier example, if my advisor overheard my sister say "ketchup meltdown" to me, I would need to provide this additional context to help my advisor understand what my sister meant to communicate:

> *When we were little, my family flew to Connecticut to surprise my grandparents for their 50th wedding anniversary. While we were setting up, our mom ran out to grab us a quick lunch. She came back with burgers and fries but had forgotten to ask for ketchup. I was so upset about the prospect of eating french fries without ketchup that I threw a total fit and made everything really unpleasant for the rest of our family. It was definitely an overreaction.*

This additional context took 82 words of explanation. When my sister and I are alone, she can use the succinct 2-word phrase, "ketchup meltdown" to convey the same information; a net savings of 80 words. Insider language is thus a compression that prompts the other mind to supply the missing information. In this example, "ketchup meltdown" is the *insider language* that references a larger story my sister and I are both familiar with. The additional context I provided to my advisor is the *explanation of insider language*, a proxy for the implicit information shared between my sister and me.

In Study 1, we ask people to write down these explanations of insider language each time they notice an instance of insider language in their conversations. We expect that friends will notice more instances of insider language in their conversations, and consequently, will need to use more words to explain them compared to dyads who have never met. Then, in Study 2, we explore whether (and how) insider language predicts friendship quality and feelings of connection.

Results

Study 1

We first tested the hypothesis that friends use more insider language in their conversations compared to strangers. Dyads of strangers and friends had 10-minute conversations over Zoom. These conversations were unstructured; participants were free to talk about whatever they wanted. Later, each participant watched a recording of their conversation with the goal of explaining all of the insider language that occurred during the conversation. Their task was to fully "unpack" each instance of insider language by writing down all the missing information an outsider would need to understand that moment (Fig 19A).

Friends spontaneously use more insider language compared to strangers. The number of words that participants typed during this task was our measurement of how much insider language needed to be explained. Consistent with our hypothesis, participants in the friend condition typed significantly more words than participants in the stranger condition (b = -137.45, SE = 38.73, p = 0.001, Fig 19B).

To guard against the explanation that friends are simply more verbose in their descriptions of insider language, we also quantified the number of *instances* of insider language that each participant unpacked. Again, results revealed that friends identified significantly more instances of insider language compared to strangers (b = -7.41, SE = 1.25, $p < 0.001$, Fig 19C). As expected, there was also a significant and positive relationship between the number of instances of insider language a participant unpacks and the number of words that they typed in the task overall (b = 23.52, SE = 1.62, $p < 0.001$)

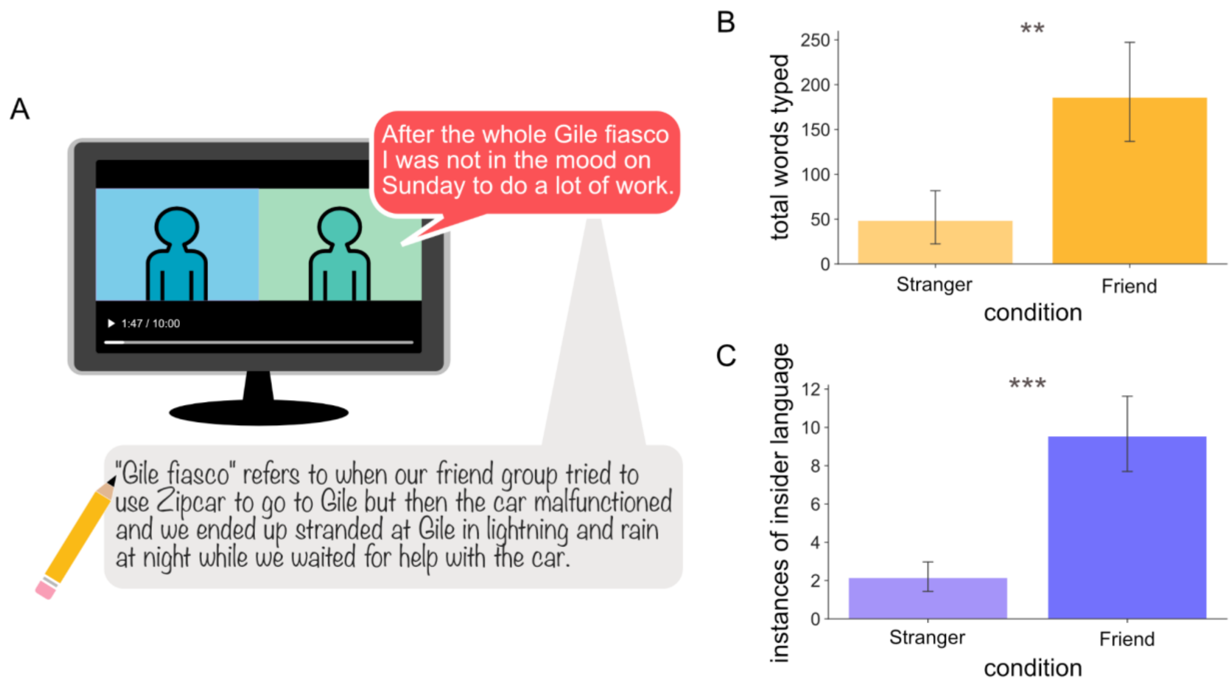


Figure 19. (A) Task design. Participants watched a recording of their conversation (top) and paused the video whenever they came across an instance of insider language. They then used a text box to explain each instance of insider language such that an "outsider" could fully understand what they were communicating (bottom). (B) Participants who had conversations with their friends typed significantly more words than participants who had conversations with a stranger. (C) Participants who had conversations with their friends identified significantly more instances of insider language compared to participants who had conversations with a stranger. ** $p < .01$, *** $p < .001$

Differences in the type of insider language used between strangers and friends. After establishing that friends used more insider language compared to strangers, we next investigated how the *content* of that insider language differed between groups. In other words, when strangers used insider language, did they tend to refer to different types of information compared to friends? To explore this question, research assistants generated a list of potential insider language categories and assigned each instance of insider language to one of those categories. The research assistants did not know that instances of insider language came from two different types of conversations. Inspecting the frequency of these categories in each group revealed that strangers tended to use insider language to refer to expertise (e.g., niche information about sports, classes, or hobbies) and friends tended to use insider information to refer to other people and experiences (Fig 20). See Supplement for more details about these categories.

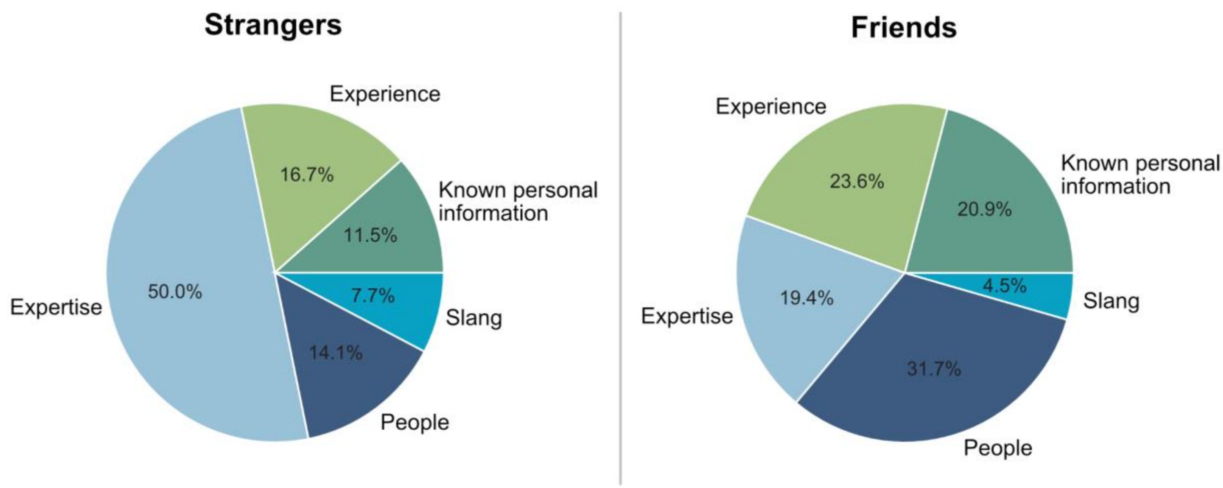


Figure 20. Percentage of insider language assigned to each category for strangers (left) and friends (right). Note that strangers unpack significantly fewer instances of insider language compared to friends; the frequency information in this visualization obscures that fact.

Study 2

We reasoned that friends would use more insider language compared to strangers because friends have had more opportunities to create common ground, and can therefore reference that common ground, every time they talk. It follows that friends who talk to each other *more often* should use insider language *even more*. To test this, we leveraged a previously collected dataset of face-to-face conversations between pairs of friends who reported on how often they talked to each other (Templeton et al., 2022). Independent raters identified instances of insider language use during these conversations on a turn-by-turn basis.

Friends who talk more frequently use more insider language. Participants' responses to the question "How frequently do you talk to this friend?" (0=Monthly, 50=Weekly, 100=Daily) positively related to the amount of insider language identified by independent raters in their conversation (b = 0.411, SE = 0.107, p < 0.001, Fig 21). Friends who talk more often used insider language on a higher percentage of turns compared to friends who talked to each other less often.

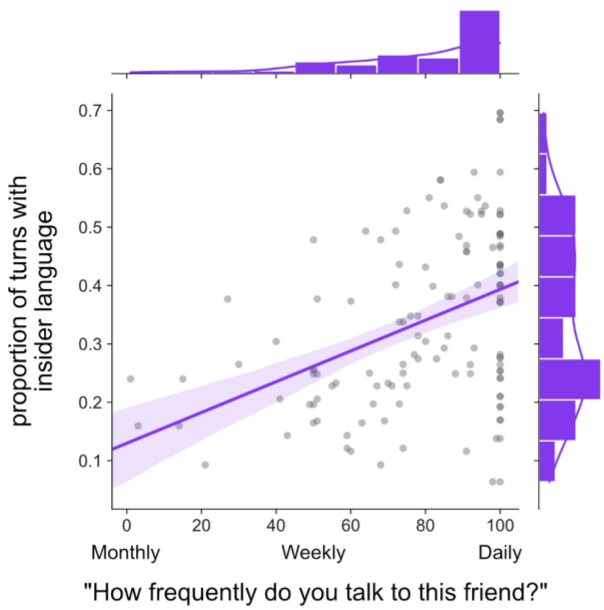


Figure 21. Friends who talk more frequently have conversations determined to contain more insider language. Individual data points are displayed as gray dots. The line represents a regression model relating talk frequency and insider language. The distribution of talk frequency responses is plotted above the scatterplot, and the distribution of proportion of turns with insider language (averaged by each rater for each conversation) is plotted to the right of the scatterplot.

After establishing that *overall* insider language use is related to friendship quality, we next investigated whether insider language use impacts how friends feel about each other *in-the-moment*. In addition to reporting on talk frequency, participants also retrospectively reported on their felt social connection throughout the course of their conversation. We leveraged this information to test whether participants felt more connected to each other when they used insider language.

Insider language use predicts greater feelings of connection. We found a significant positive relationship between the number of independent raters who thought a turn contained insider language and participants' own reports of connection during that turn (b = 0.046, SE = 0.005, p < 0.001, Fig 22). Turns with insider language felt more connected than turns without.

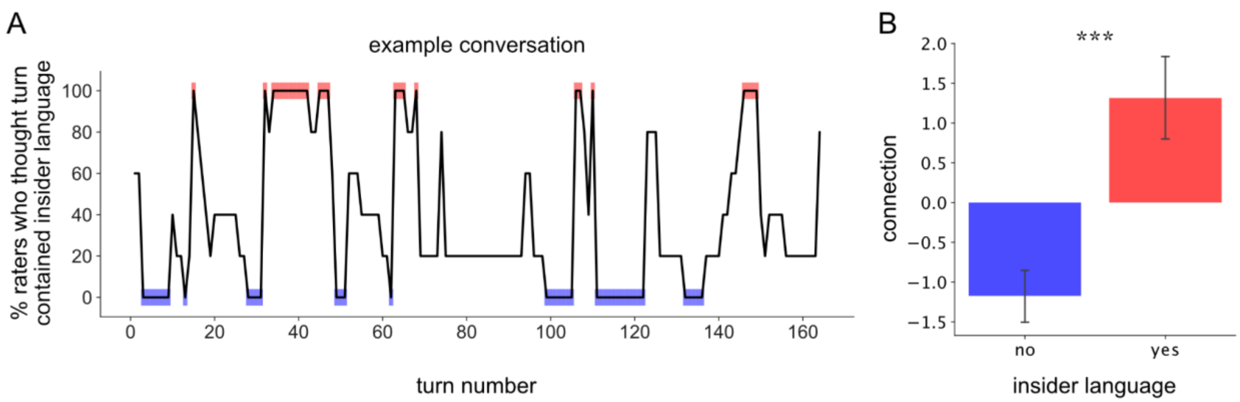


Figure 22. (A) The percentage of raters who identified insider language for each turn in one conversation. Highlighted in blue are turns that no raters identified as having insider language (i.e. rater consensus: no insider language). Highlighted in red are turns that all raters identified as having insider language (i.e. rater consensus: yes insider language). (B) Connection ratings for turns in which insider language was considered absent (blue) and present (red), across all conversations. Connection values are residualized to account for random effects of participant and dyad as well as linear effects of time. Note that the model reported in the main text incorporates insider language consensus score and connection rating for every turn. *** p < .001

Discussion

In Study 1, we found evidence that friends use more insider language in their conversations compared to strangers. In Study 2, we show that friends feel more connected to each other when they use insider language. Participants were unaware that insider language was a focus of this research. Therefore, these studies demonstrate how insider language occurs spontaneously in casual conversation and that it can influence feelings of connection even when participants are not explicitly thinking about it.

We also found that friends who talk to each other more frequently use more insider language in their conversations. This is consistent with previous work showing that dyads who repeatedly play a communication game develop a shorthand way of communicating, allowing them to complete the game faster each round. Insider language in conversation likely affords a similar efficiency. It is easier to use fewer words to reference a story that both conversation partners know than to provide a laborious retelling each time. Our results suggest that there is also a social benefit to using insider language; people feel more connected when they use it.

An exploratory investigation into how categories of insider language differ between friends and strangers revealed that strangers tended to use insider language to communicate niche expertise whereas friends used insider language across a wider range of categories, including talking about people, experiences, and referring to known personal information. Insider language use may reinforce the strength of a relationship by reminding two friends that they have a shared history and by proving that two strangers have common interests.

The phrase "insider language" implies spoken words, but insider information can also be conveyed nonverbally, for example with specific gestures, exaggerated facial expressions, well-timed eye-contact, or shifts in tone of voice. In fact, a recent paper found that gesture duration becomes shorter as the same gestures get repeated (Holler et al., 2022) in the same way that language becomes more concise. Raters in Study 2 were encouraged to use all of the information conveyed in the video recordings when making determinations of inside language use. It would be interesting to examine the different forms that insider language can take, and which sorts of conversational contexts encourage certain modalities over others.

The present work investigated insider language between dyads. By definition, any insider language would need to be understood by both dyad members. However, the same is not true in group settings. There, it is possible to use insider language that is only understood by a subset of the group. Future work should explore how insider language occurs spontaneously in groups and how different levels of inclusion / exclusion can differentially impact feelings of connection between group members. For example, using insider language that only one other person in the group understands could increase feelings of connection between those two people at the expense of alienating the other group members.

Finally, this work has implications for the success of artificial intelligence. Here, we highlight how the words that people use in conversation are not necessarily the whole story. Information living in the heads of two participants does not need to be stated explicitly in dialogue. However, Natural Language Processing techniques can only use the words that are explicitly spoken or written. Computer systems that seek to truly understand what people are saying will need to find ways to notice when insider language is being used and try and determine what words are being left out, but nonetheless communicated. This work joins recent calls for AI to move toward shared concepts vs. literal word meanings (Stolk et al., 2016).

Insider language presents a challenge in studying conversation because important information is going *unsaid*. These studies demonstrate that insider language use is a sign of relationship closeness and a source of connection. Being aware of insider language use in conversation can inspire new study designs and analytical approaches to specifically investigate when and how it is used. People find joy in being able to communicate *more* by saying *less*.

Methods

Study 1

Participants. Study 1 consisted of 40 conversation dyads (20 stranger dyads and 20 friend dyads) comprised of 80 unique participants (57 Female, 23 Male). All reported studies were approved by the Dartmouth Committee for the Protection of Human Subjects, and all participants provided informed consent prior to participation.

Procedure. Participants were randomly assigned to be in either the Friend or Stranger condition. Participants assigned to the Friend condition were asked to nominate someone who they (i) considered to be a close friend, (ii) talked to regularly, and (iii) were not romantically involved with. Participants assigned to the Stranger condition were paired with each other by an experimenter. We confirmed whether Stranger dyads were true strangers by their answer to the question, "Did you know your conversation partner BEFORE your conversation with them?". Dyads assigned to the Stranger condition who knew each other were excluded from analyses.

All participants knew that the study would involve either a recorded zoom conversation with a friend that they nominated themselves or with someone assigned to them by the experimenters (i.e., a stranger). Because participants volunteered prior to knowing which conversation condition they would be in (friend or stranger) and were then randomly assigned to the condition, their assignment could not be attributed to selection effects.

Conversation Session. Each dyad was instructed to have a 10-minute unstructured conversation over the video conferencing platform, Zoom. Participants were free to talk about anything they wanted. A researcher facilitated each conversation session. After explaining the task and answering any questions, the researcher started the recording and then left the zoom room. After 10 minutes had passed, the researcher re-entered the zoom room to pause the recording and to explain the subsequent tasks.

Insider Language Task. A few days after the conversation session, participants received a link to a personalized Qualtrics survey containing the insider language task (Appendix D). Participants were instructed to watch the video recording of their 10-minute zoom conversation and to pause the video every time they encountered an instance of insider language use. Once the video was paused, participants were instructed to type an explanation of the instance of insider language such that an “outsider” would be able to understand. They were told to imagine this outsider was a “typical Dartmouth student”. Participants in the Friend condition paused the video significantly more times than participants in the Stranger condition (b = -4.57, SE = 2.09, p = 0.031) and number of pauses was significantly positively related to number of instances of insider language (b = 0.46, SE = 0.06, p < 0.001) and total words typed in the task (b = 12.80, SE = 1.49, p < 0.001). The methods of this study, as well as the hypotheses, were preregistered prior to collecting data at https://osf.io/hs64n/.

Measures of insider language. Our primary measure of insider language in Study 1 was the word count of the explanations each participant typed over the course of reviewing their 10-minute video. We also used a count of the number of instances of insider language that each participant unpacked.

Text Preprocessing. Participants were instructed to type their explanations of insider language into a text box. This task was purposefully open-ended to allow participants to use whatever words were necessary to fully explain their conversations. However, when inspecting what participants typed, it became clear that they occasionally used the text box for unintended reasons. For example, participants sometimes wrote in the text box about how there was no insider language in their conversation (e.g., “I honestly really don't think anything needed to be explained. We didn't talk about any names or places during the conversation”). Because our primary dependent variable is the number of words typed, we wanted to ensure that this word count reflected explanations of insider language only. The example above would have added 21 words to the total word count, artificially inflating that score. Therefore, we preprocessed the text to remove words that were off-task before computing word counts (see Supplement for more details). The reported results hold when computing word counts without any preprocessing (b = -128.27, SE = 38.83, p = 0.002), suggesting that decisions made with text preprocessing cannot explain the differences between conditions. Similarly, these off-task explanations were removed from the counts of instances of insider language, and the effect of condition still holds when including all instances (b = -6.59, SE = 1.30, p < 0.001).

Statistical Models. For all reported analyses we used lme4 (Bates et al., 2015) implemented in R (R Core Team, 2018) to perform linear mixed effects regressions. Degrees of freedom and p-values were approximated using Satterthwaite’s method and we report standardized regression coefficients to increase interpretability.

A linear mixed-effects model with condition predicting total word count was used to examine whether total word count was higher for friend conversations compared to stranger conversations. Because two participants participated in and made ratings about each conversation, conversation ID was included as a random intercept.

To examine whether the number of instances of insider language was higher for friend conversations compared to stranger conversations, we used a linear mixed-effects model with condition (friend vs. stranger) predicting instances. Conversation ID was included as a random intercept.

Study 2

Participants. In Study 2, 65 friend dyads (female/female = 32, male/male = 20, and female/male = 13) had conversations. Twenty-two participants nominated 3 of their close friends as partners. These conversational partners were someone they (i) considered to be a close friend, (ii) interacted with regularly, and (iii) were not romantically involved with. Twenty-two participants participated in multiple conversations (up to 3) and 65 participants participated in a single conversation.

Conversation Session. In each study session, two participants entered the laboratory and had a 10-minute conversation that was video and audio recorded. Participants were free to talk about whatever they wanted. After their conversation, participants were separated into private rooms where they completed a Qualtrics survey about their relationship with their partner (Appendix B). Participants then completed a second task that required them to watch the video recording of their conversation. As they watched, participants continuously rated how connected they remembered feeling to their conversation partner at each moment in time. Participants made these ratings by using a computer mouse to move a slider bar on the screen. Each conversation session took about 30 min to complete.

Insider language annotations.

Independent raters. Five undergraduate research assistants were tasked with identifying moments of insider language in each conversation. Using research assistants as raters had two advantages. First, unlike workers recruited from Amazon's Mechanical Turk (or similar platforms), research assistants who are covered by the study's IRB were permitted to view the complete set of friend conversation videos, rather than only the subset of videos with signed video releases. Second, just as in the first study, we were interested in moments in a conversation that would need to be explained to a "typical Dartmouth student." By virtue of being Dartmouth students themselves, these raters were well-positioned to differentiate between typical Dartmouth lingo and words used to describe something unique to a particular dyad. None of the raters were informed of the goals and hypotheses of the study beforehand.

Rating Task. Transcripts of each video were randomly ordered and shown to raters one at a time via an online survey. Raters located the corresponding video of each transcript and watched the video with the goal of noticing moments of insider language. Watching the original videos allowed raters to pick up on instances of insider language conveyed via tone of voice or gestures that may not necessarily be apparent in the text of the transcripts alone. When raters noticed a moment of insider language, they found the corresponding speech turn in the transcript and selected it by checking a box (Appendix E). Raters checked every turn that corresponded to a moment of insider language in the video. If the moment persisted over several turns, they selected all of the relevant turns. Raters continued doing this until they completed annotations for all 65 conversations. Raters were given full control over the videos (pause, rewind, etc.) and could return to transcripts as many times as needed.

Statistical Models.

Insider language and talk frequency. To quantify how much insider language was used in each conversation, we computed the average proportion of turns in each transcript that each rater identified as containing insider language (Figure S15). This score was used to predict participants' own reports of how frequently they talked to their partner. Because each dyad member reported on their talk frequency with each other, Dyad ID was included as a random intercept in the linear mixed-effects model. The Supplement describes alternative ways to investigate this same question and every approach shows the same effect.

Insider language and connection. For each turn in each conversation, we computed an average connection rating and an insider language consensus score. The insider language consensus score reflected the percent of raters who identified that turn as containing insider language. This score could go from 0 (no raters thought the turn contained insider language) to 100 (all raters thought the turn contained insider language).

A linear mixed-effects model with the insider language consensus score as a fixed effect was used to predict connection ratings at each turn. Due to a known positive relationship between turn number and connection ratings, turn number was also included as a fixed effect. Because different transcripts have different numbers of turns, we normalized the turn numbers by dividing each turn number by the total number of turns in that transcript. An interaction term was added between both fixed effects. All continuous variables were scaled to interpret the beta coefficients. Conversation ID (the name of the transcript that was rated) and subject ID (the identity of the participant in the conversation who provided continuous connection ratings) were included as random intercepts (Figure S16).

General Discussion

The question that motivated this dissertation was: *What makes conversation good?* Specifically, we were interested in capturing features of conversation that predicted whether people enjoyed them and felt connected to each other. We took the approach of simply recording many different people in many different conversations, hoping to capitalize on this natural variance; some people would walk away feeling really connected with their partner and others would not. We also recorded conversations between close friends to contrast with conversations between strangers, under the assumption that conversations between friends were more likely to contain examples of *good* conversations.

I started this project in my first year of graduate school by collecting one round-robin of conversations between strangers in order to develop my analysis pipeline. My goal was to quantify *every aspect* of the conversation videos by the end of summer. Seven years later, I have something to say about *three aspects* of conversation. (And much more to say about the difficulties and challenges of studying natural conversation in the first place!)

Re-cap and reflections

In Chapter 1 we focused on turn-taking in conversation. What can the length of the gap in between turns tell us about how connected people feel? In the first paper, we found that shorter gaps predict more feelings of connection for friends and strangers alike (Templeton et al., 2022). In the second paper, we turned our focus to extremely *long* gaps in conversation (Templeton et al., 2023). We found that, for strangers, long gaps marked moments of disconnection and awkwardness that became worse as the gap got longer. However we also found that conversations between friends contained *more* instances of long gaps (greater than 2 seconds) and that those long gaps marked moments of increased connection. At first blush, these two findings might seem in opposition to each other. How can shorter gaps *and* long gaps mean more connection for friends? The explanation is in the details of how we approached the analyses in each paper. For the first paper on short gaps, we looked at *averages* (either the average gap length across an entire conversation or the average gap length within different time bins in the conversation). There, a single long gap would simply be absorbed into the average. In the second paper, we looked specifically at those long gaps as *events*. When a long gap occurred, we zoomed in and examined how feelings and impressions of connection changed in the moments immediately surrounding the long gap. By definition, these long gaps were all outlier events (greater than 2 standard deviations from the group mean). While we would not expect them to meaningfully alter average values, the fact that we observe consistent changes in connection around those moments signal the power they have to broadcast how things are going.

Chapter 1's focus on gap lengths is a direct result of the data-driven, exploratory nature of this project. When we designed this study, we thought through all the features in conversation that might impact connection and never considered *gap lengths*. Yet the data were clear—the simple pattern of speaking and silence is a remarkably strong predictor of how well a conversation is going.

In Chapter 2 we focused on how people *start* their conversations. We first transformed the text of what people said at each 30-second window of their conversations into a language embedding and found that strangers start their conversations more similarly to each other compared to friends. We next clustered the language embeddings to identify the topics that strangers tend to use to start their conversations (i.e., their hometowns and classes). By applying network analyses to topic transitions, we found evidence that many of these same topics have high in-degree and out-degree, perhaps making them a strategically smart place to start. These topics offer many pathways *out* but also act as safe topics to *return to*. We call topics with these properties *conversational launch pads*.

Our language analyses and questions have changed dramatically over the years. This was partially due to new developments in language models and methods and partially due to gaining more experience and insight into how these models interact with our particular datasets. Most of these language models are trained on written text and struggle when applied to natural, spoken conversation. Like in many other areas of science, the tools constrain the types of questions we can ask. Language embeddings are built to compute semantic similarity, and we applied this fact to reveal differences in how friends and strangers speak similarly, over time. But the types of questions we will be able to explore using Natural Language Processing will expand over time. Understanding the scope and limitations of these methods is important for designing analyses and interpreting the results.

In Chapter 3 we focused on the words that go *unsaid* in conversation. We used the term *insider language* to describe words in conversation that stand in for a larger concept, understood by both conversation partners. This creates efficiency in the conversation between those two people, but also makes the larger meaning inaccessible to an outside observer. In Study 1, we asked participants to identify and unpack their spontaneous uses of insider language. We showed that insider language happens more often between friends, compared to strangers. In Study 2, we asked outside raters to identify moments of insider language in conversations between friends. We then related those moments to the participants' own feelings of connection to show that friends report feeling more connected during turns determined to contain insider language. Insider language does not only aid in communication efficiency, there are *social benefits* to using it as well.

Our focus on insider language originated from becoming increasingly aware of the limits of Natural Language Processing to represent meaning in our datasets. We noticed moments in the videos where participants were clearly communicating something to each other that we did not understand. We knew we were missing some crucial backstory or shared history. If *we* couldn't understand the context, how could we possibly expect a computer to? This necessitated collecting an entirely new dataset where we explicitly asked participants to fill in that missing information *for us*. The results were amazing. I found myself scrolling through the data with a huge smile on my face as I read participants' explanations of shared information and experiences. When my research assistants worked together to assign each insider language instance to a category, they would often read an explanation aloud that would cause the entire lab to burst out laughing in glee. Outside of the lab, I enjoy noticing these moments between other people as they communicate what sorts of things they prioritize in their relationships. What do they find so salient and important that they can confidently assume their partner will know what they mean when they say a single word or give a knowing glance? Insider language is a joy to experience, even from the outside.

Future directions

Combining features. In this dissertation, we examined three features of conversation—gap length, topics, and insider language—in isolation. Each chapter investigated how a different feature impacted feelings of enjoyment and connection and differed between strangers and friends. Of course, all of these features are unfolding simultaneously over the course of a conversation and it is likely that they influence *each other* (Hadley et al., 2022). In future work, I plan to investigate relationships between these features. For example, are transitions out of launch pad topics followed by shorter gap lengths? Does insider language increase if and when people exit launch pad topics? Do shorter gap lengths correspond to more insider language?

Additional conversation features. Of course, there is more to conversation than gap length, topics, and insider language. There are many features that we have not analyzed here, including body posture and movement, voice and audio properties, and laughter. There are likely many more features that we do not even know to look for yet. The value in the approach we took in this dissertation is that we can continually return to the conversation videos and quantify additional features. The ultimate goal will be to examine how all these features impact each other over the course of a conversation and how they work together to impact feelings of connection and enjoyment.

Better leveraging the round-robin structure. We collected our first conversation dataset as a round-robin: every member of the round-robin had a conversation with everyone else in the round-robin. This was intentionally done to examine different people's preferences and styles in conversation. It allows us to observe a single person in 10 different interactions. It also allows us to observe consistencies in impressions 10 different people have for a single person. Some people are universally loved, others not so much. We utilized this structure in one analysis in Chapter 1, finding that people who have a tendency to use shorter gaps in their conversations are better liked. In the future, I plan to leverage this interesting structure much more to establish different conversational preferences that people have and investigate how feelings of connection may be a result of different preferences aligning or clashing. Connecting with people likely is not a game of being the "best" person (though some people may have an easier time than others) but rather a game of finding the right conversation *partner*.

Other conversation goals. In this dissertation, the outcomes we were most interested in was how much people enjoyed their conversation and how connected people felt to their conversation partners. Of course, people may have different goals in conversation (Yeomans et al., 2022). They may primarily care about coming across as intelligent or respected, they may be hoping to get helpful advice, they may be interested in resolving a conflict, the list goes on. The same patterns of behavior that predict feelings of connection in conversation may not be the ones that achieve other goals. Considering the goals that the people have in conversation is important for situating behavior and for appropriately measuring people's perceptions of how well they succeeded in those goals.

Other conversation contexts. The conversations that we have collected and analyzed for this dissertation were all 10-minute, dyadic conversations between (mostly) undergraduate students at Dartmouth. The generalizability of our findings must always be considered with a grain of salt. However, the generalizability of the *methods* is what is most exciting to me. Regardless of group size, conversation goal, or identity of the participants, conversations all have a similar *structure*. Someone talks and then someone else talks. The techniques used to analyze one conversation can be easily ported to another conversation and I am so excited analyze other conversational contexts. Further, I am prepared to do that work *because of* the insights and skills I gained working with my dissertation datasets. The approaches and methods put forward in this dissertation can be applied to all different kinds of conversations, helping us better understand how minds connect.

Why does this work matter?

I began this dissertation with an overview of different theoretical accounts that have guided prior work on conversation. How do my dissertation studies and findings fit into those existing frameworks? To be clear, we did not collect our datasets with the explicit goal of finding evidence for or against any particular theory. We were instead motivated to record conversations in a way that best resembled how they actually unfold outside of the lab. Still, I think many of our findings lend credence to influential ideas baked into many of these theories. For example, in Chapter 1 we find that people use faster response times when they feel more connected. We believe that faster response times are by-products of mental alignment but we don't know if they are being achieved by a subconscious level of priming (Pickering & Garrod, 2004) or a more concerted effort to find common ground (H. H. Clark & Brennan, 1991). It may be possible to pit these two accounts against each other by using our existing data or by devising a targeted follow-up study focused on gap lengths. It is possible to view our findings from Chapter 2 with the lens of information theory (Shannon, 1948). Good conversational launch pad topics might be ones that can be easily encoded and decoded by many different senders and receivers. They may therefore be particularly useful in reducing uncertainty in conversation. Chapter 3 highlights the spontaneous use of insider language in conversations. We show that friends use insider language more than strangers and that friends who talk to each other more often use it even more than friends who talk to each other less often. These findings are solidly in-line with theoretical frameworks that describe conversation as a process of building and maintaining conceptual spaces (Stolk et al., 2020).

When thinking about how our findings fit into existing theories about conversation, it becomes clear that certain theories rely on particular goals or task structures. For example, the conversation outcome that we cared most about in these studies was social connection. This means that many aspects of Shared Reality Theory are extremely relevant (Echterhoff et al., 2009; Hardin & Higgins, 1996). Indeed, the continuous connection ratings in our datasets could be used to tease apart *objective* shared reality (moments in the conversation when both participants report high connection) from *perceived* shared reality (moments in the conversation when one participant reports high connection while the other participant reports low connection). These ratings could also be used to explore whether the valence of an experience impacts shared reality (e.g., moments in the conversation when both participants report high connection vs moments in the conversation when both participants report low connection). Other accounts, like the Rational Speech Act (Goodman & Frank, 2016), cannot be easily applied to our datasets because our participants do not perform tasks with clear outcomes (e.g., successfully directing a partner to order a set of novel shapes correctly). However, it may be possible to use a similar framework to model different aspects of conversation behavior, like what participants will say next or how quickly they will respond.

More generally, this work helps push us beyond atomistic scientific methods that consider single variables and isolated individuals (Schilbach et al., 2013; Wheatley et al., 2019). In reality, we live messy, complex, and unconstrained lives. Most importantly, we live *social* lives. Understanding what happens when people interact with *each other* is crucial to understanding who we are as a species and how our brains work (Hasson et al., 2012). This dissertation work attempts to better represent the ways that people behave outside of the lab by recording unstructured conversations to examine how behavior and feelings of connection fluctuate together over time and between people (Butler & Randall, 2013).

We spend our lives talking to each other. Conversation is the foundation of our close relationships and friendships, some of the most precious things we have (R. I. M. Dunbar, 2018). Understanding what makes a good conversation has important implications for our mental and physical health (Holt-Lunstad et al., 2015) and can help reveal the nature of who we are and what is important to us.

Ethical considerations

Having a better understanding of what makes conversation good has many promising applications. Being able to spot reliable signals of connection in conversation could enable people to identify relationships that are thriving and flag relationships that might need more attention and care. Results could also lay the foundation for new interventions aimed at reducing conflict during difficult discussions or promoting feelings of connection in close relationships. Of course, any tool that can be used for good can also be harnessed for dubious reasons. For example, technology companies are motivated to create conversational AI systems that will increase user rapport, though the route to getting there will be steep (Cassell et al., 2007; L. Clark et al., 2019). Although we hope that many aspects of connection cannot be "faked" in natural conversation (Templeton et al., 2022), computer systems designed to follow rules may be able to fake them in a way that inappropriately builds trust between human and machine (and the company that operates that machine).

The way conversation research is conducted also needs to adhere to ethical standards. Obtaining conversation data should be done with participant knowledge and consent with expectations of confidentiality upheld. Also, it will be important to more diversely sample conversation behavior from a large swath of cultures, contexts, and populations before any sorts of conversation interventions are attempted (Sanchez et al., 2022). These concerns mirror well-documented demonstrations of how bias in training data can result in models that exhibit racist or sexist behavior (Zou & Schiebinger, 2018). Conversation behavior that promotes connection in one population may be harmful to another population.

Thoughts on working with naturalistic data

Working with naturalistic data is both rewarding and deceptively hard. No two conversations are alike which means we cannot directly compare them. Instead, we have to compare them with respect to some *feature* they both have. Any given feature of interest may be happening at different times and with different frequencies across conversations.

In this dissertation, we illustrate many different ways of utilizing naturalistic data. In Chapter 1, we designed a follow-up experiment to specifically test the relationship between gap length and connection that we observed in our conversation data. Later in Chapter 1, we turned our focus to long gaps and treated each long gap as an *event* that we could isolate and examine further. In Chapter 2, we used binning to put all conversations on a comparable time scale to examine how semantic similarly changed over the course of a conversation. In Chapter 3, we got annotations of insider language separately for each conversation and compared connection for moments with vs without insider language. All of the chapters also demonstrate different ways of creating appropriate null distributions and comparison groups within the data itself.

Naturalistic data is challenging to analyze, requiring new techniques and analytical methods. But the benefits are considerable. By allowing people to talk with each other naturally, we can analyze these features as they occur *in nature*, without creating an artificial experience. The ecological validity of this approach means that we can learn how, when, and why real conversation is such a powerful determinant of human connection.

Conclusion

I recorded unstructured conversations between pairs of strangers and friends to examine how people connect in conversation. Specifically, how gaps between turns signal connection (Chapter 1), how people start their conversations (Chapter 2), and how language moves from spoken words to shared thoughts (Chapter 3). For me, one of the best feelings in the world is suddenly clicking with someone in conversation. It is magical. I hope my dissertation can start to explain what happens in these magical conversations so that we can all have more of them.

Supplementary Materials: Chapter 1a

Studies 1-2

Transcription Details

Many of our analyses relied on the timestamps in each conversation transcript. Here we detail exactly what the transcripts contained and our decisions about what constituted a speech 'turn.'

Format. Below is a screenshot of the first few turns in one transcript, to illustrate the format of the transcripts. Each turn included (i) speaker information (S1 or S2), (ii) a START timestamp, (iii) an END timestamp, and (iv) text of the words spoken.

To compute response time, we subtracted the END timestamp of the previous turn from the START timestamp of a given turn.

```
S1: 00:00:00.000 How's it going? END 00:01.335 END

S2: 00:00:02.236 I'm okay. How are you? END 00:03.215 END

S1: 00:00:03.123 Yeah, I'm good. It's been a busy... END 00:05.599 END

S2: 00:00:05.636 It's been a week. END 00:06.043 END

S1: 00:00:06.397 Yeah, honestly. With rush. Did you rush? END 00:08.880 END

S2: 00:00:08.761 I dropped on Monday. END 00:09.686 END

S1: 00:00:09.582 You dropped? Oh, I'm sorry. END 00:10.731 END

S2: 00:00:10.388 Yeah, I didn't get callbacks. END 00:11.686 END

S1: 00:00:12.067 Oh, okay. END 00:12.432 END
```

Defining a speech 'turn'. The primary unit of analysis is a *speech turn* which we defined in the following ways. For non-overlapping speech, a turn was simply all the words one speaker said before their partner began. For overlapping speech (e.g., backchannels, interjections, interruptions, and false starts), we kept the "coherent thoughts" together (Version A below).

Another way to handle overlapping speech would be to initiate a new turn every time another speaker talks (Version B below). Consider how this real moment of overlapping speech could be transcribed in two different ways:

Version A

Turn 1, S1: It was a beautiful day. That's true. Things could be much worse, but...

Turn 2, S2: How can you be hanging in there today with weather like this?

Version B

Turn 1, S1: It was a beautiful day. That's true. Things could be much--

Turn 2, S2: How can you be--

Turn 3, S1: --worse--

Turn 4, S2: --hanging in there today--

Turn 5, S1: --but.

Turn 6, S2: --with weather like this?

During moments of overlapping speech, Version B results in many more turns with sentences "broken apart" across those turns. We felt that Version A better captured our own experiences engaging in conversation as well as listening to these recorded conversations -- where we are able to effortlessly integrate the words one speaker is saying even if another speaker is simultaneously speaking. Note that in the conversations we recorded, speech doesn't overlap by much time (median negative latency is 297 ms). This is in line with what has been found in previous work on overlaps and so the task of assigning words to the correct speaker is easier than if we were to instruct people to talk over each other in a way that does not tend to happen naturally.

Although we removed the timestamps from this example to increase readability, note that for Version A, the START timestamp for Turn 2 would occur earlier in time than the END timestamp for Turn 1. This would allow us to compute how far into Turn 1 Speaker 2 initiated Turn 2.

It is also important to note that Version A and Version B are both correct representations of what happened in the conversation. However, the different versions emphasize different aspects of turn-taking. While we believed Version A was better suited for our particular question, there may be other projects that benefit from defining turn taking like Version B. What we want to emphasize is that it is important for researchers interested in similar questions to think carefully about how to define a “turn” and to ensure that definition is (i) disclosed and (ii) applied consistently across all transcripts in a given dataset.

Transcription Company. The transcriptions (and therefore the timestamps) that we used to run the response time analyses presented in the paper were all done by one company -- Scribie (https://scribie.com/). More details about the transcription that was completed for each individual conversation video can be found in the Supplement folder of this project's Github repository (ConversationDatasetDetails.doc).

Robustness checks for within-conversation analysis

We examined the relationship between response time and social connection over the course of a conversation by dividing the 10-minute conversation into twenty 30s bins. Our first robustness check was to examine this same relationship over a variety of bin sizes. We examined these bin sizes: 5, 10, 15, 25, 30, 40, 50, 60, 100, 120, 150, 200, and 300 seconds. For each bin, we computed the average response time for turns that occurred in that bin as well as the average connection rating for each speaker in that conversation. We ran a mixed linear effects model predicting the temporal dynamics of social connection based on fluctuations in average response time controlling for linear effects of time. To account for variations in average response time between dyads, we included Dyad ID as a random intercept and additionally modeled Subject ID as a random intercept because subjects participated in multiple conversations. For stranger conversations (Study 1) we observed a significant negative ($p<.05$) effect of response time on connection for all of these bin sizes. For friend conversations (Study 2) we observed a significant negative effect of response time on connection of all of these bin sizes except for the largest bin size of 300 seconds ($p=.06$). This is consistent with the across conversation finding that at these larger timescales friend reports connection become uniformly high and do not sufficiently vary across the conversations to assess the relationship with average response time.

For our second robustness check, we created an empirical null distribution to ensure that the effect we were observing could not be explained by any offsets in lag between changes in response time and connection ratings. To do this, we generated surrogate data by randomly permuting the order of response times within each conversation using a circle-shifting procedure and re-fitting the model predicting social connection 100 times. We performed this procedure for each bin size listed above. As you can see in Figs S3 and S4, our results cannot be explained by any offsets in lag between changes in response time and connection ratings.

Study 3

There are two ways to manipulate the size of response times in a recorded conversation: *proportional* (changing each original response time to be longer or shorter by a specified proportion) and *distributed* (replacing each original response time with one pulled from a specified distribution: short or long). For the pre-registered Study 3 in the main text, we used proportional manipulation which had the benefit of changing the size of each response time while maintaining the natural variance of response times across the conversation. However, we also ran a version of this study that used the alternative method in which the original response times were replaced with response times taken from one of two *distributions* (fast or slow). We are including the methods and results of this additional manipulation study here for two reasons. First, the results demonstrate how our effect replicated in a different set of participants, using a different method of manipulating response time. Second, the methods provide helpful context to explain the changes we made before running the pre-registered version of this study.

Study 3 Replication

Methods. For six of the conversations from Study 1 (3 male and 3 female), we selected a short segment (mean length = 23.33 seconds) to use as stimuli. We picked conversation segments that had minimal overlapping speech, where both participants had signed a video release. These were the same 6 conversation segments that we used in the preregistered version as well.

We used these 6 stimuli to create two different conditions: Slow response times and Fast response time. Specifically, for each segment, we produced two versions by inserting response times with the length drawn from two different distributions: Slow (mean response time = 500 ms, std = 10 ms) and Fast (mean response time = 50 ms, std = 10 ms).

300 participants on Amazon's Mechanical Turk listened to each of 6 conversation segments, presented in a random order. All participants heard each conversational segment only once and the version (Long, Short) of each conversation they were presented was randomly assigned. This random assignment was blocked such that, over all participants, each conversation segment was presented an equal number of times in both conditions.

After listening to each conversation segment, participants responded to two questions -- *1) How much do you think these people enjoyed their conversation?* and *2) How connected do you think these people felt toward each other?*. Participants responded using a slider bar anchored by "Not at all" (0) and "Very much" (100).

To access the study, participants were first required to successfully complete a task delivered via audio instructions. This ensured that we only included those participants who were able to listen and respond to audio instructions, a requirement for the study.

To check for participant compliance, we also collected timing information on each page of the survey. This allowed us to determine whether or not participants submitted their response before the audio file stopped playing. When we filtered the dataset based on this inclusion criteria, our number of observations dropped from 1,800 (300 participants x 6 items) to 1,584 (88% retention). We ran the same set of analyses on the full dataset as well as on this subset and the pattern of results do not change between the two. We present results on the full dataset here.

Results. We ran two mixed linear effects models with condition (Fast, Slow) predicting each of our two DVs: perceived enjoyment and perceived connection. We included subject and item (e.g. which of the 6 conversations was being judged) as random intercepts. Results showed a significant effect of condition such that the same conversation with fast response times was perceived as more enjoyable ($b=6.50$, $SE=0.77$, $p<.001$) and connected ($b=7.52$, $SE=0.88$, $p<.001$; Fig S7) than the version with slow response times.

Additional methods for reported Study 3

Preregistration and materials. We preregistered Study 3 before collecting data. Our full preregistration plan is here: osf.io/u2brn. We've highlighted details from the preregistration plan in this section.

Improvements over the original study. We were encouraged by the results from the first manipulation study, but wanted to be sure our result couldn't be explained by the decision to manipulate the response times by a distributed vs proportional method or acoustic properties in the stimuli. We made three changes. First, we moved from a distributed to a proportional method of manipulating the response times. Second, we recorded the natural background noise of the testing room and used that (rather than pure silence) as the audio for the response times. Third, we included a baseline condition where the response times were not manipulated.

Audio processing. Each conversation segment consisted of audio clips for every speech turn and for every response time in between each speech turn (e.g. silences). The speech turns were spliced out of the original conversation. The silences were taken from a recording of the empty testing room (to re-create the ambient noise) and were trimmed to be the length of the silences in the original conversation (or trimmed to be the length of the new, manipulated response time).

For the Control condition, these audio clips were stitched together, in the order that they appeared in the original conversation segment. We persisted in stitching together the clips in the Control condition (rather than simply playing the entire segment) to maintain continuity across the conditions. The mean response time in the Control condition was 278.88ms.

For the Slow condition, we manipulated the length of the audio clips that consist of silences to be twice as long as they are in the Control condition. The mean response time in the Long condition was 557ms (278 * 2).

For the Fast condition, we manipulated the length of the audio clips that consist of silences to be a fifth as short as they are in the Control condition. The mean response time in the Short condition was 55.7ms (278 / 5).

We picked these multipliers to best match the distributions that we used in our original manipulation study (where the Short condition had a mean response time of 50ms and the Long condition had a mean response time of 500ms).

We took two additional steps to improve the quality of the final conversation segment. For all conditions, the speech turn audio clips have a fade in of length 100ms and a fade out of 300ms. The silence audio clips have a fade in and fade out of length 20ms. This is done to make the audio file sound "smoother" and more natural. For consistency, we decided to keep these lengths constant across all conditions and all files rather than 'tailoring' them to each clip. Finally, because participants in the original study complained that the audio clips were hard to hear, we increased the volume of all audio clips by 6 dB.

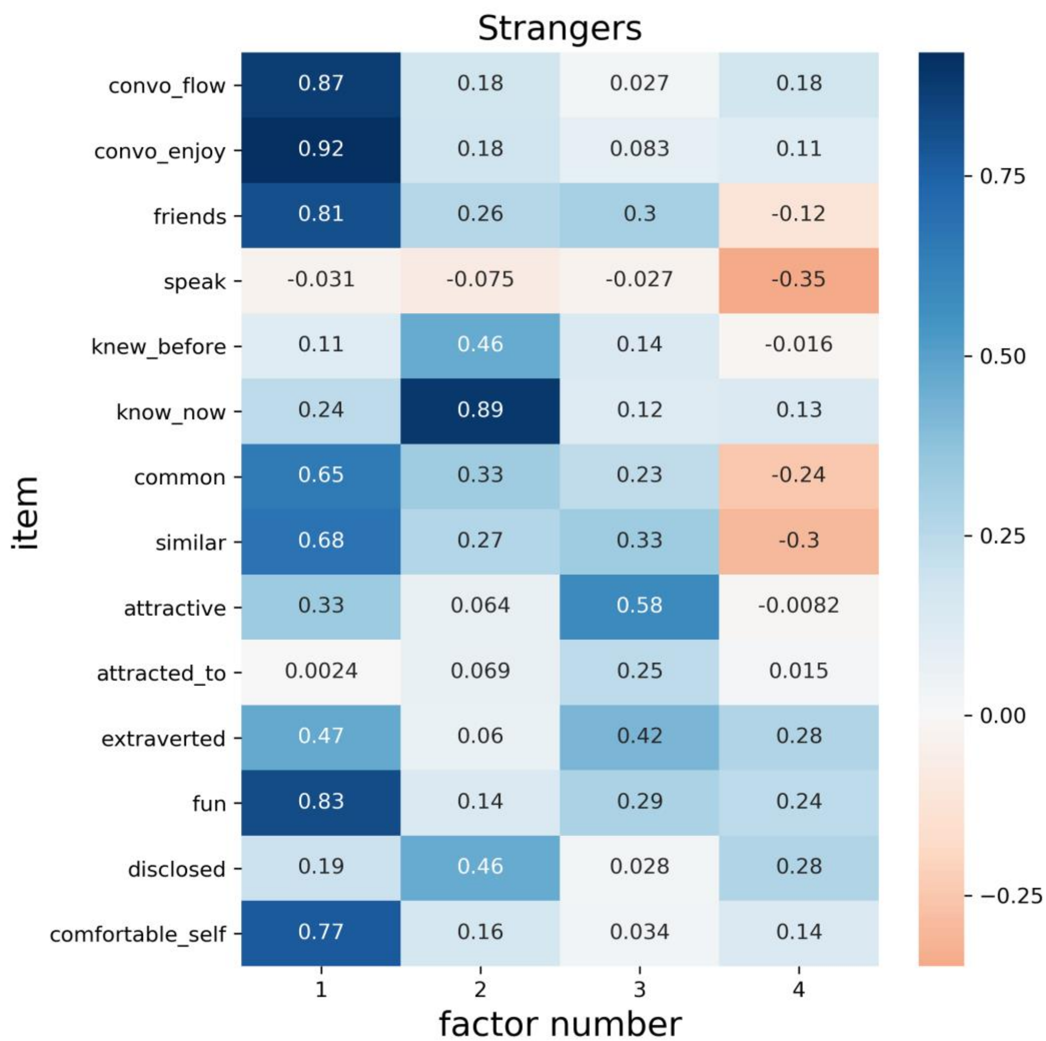


Figure S1: Factor loadings for Study 1. Factor loadings for questions asked across all round robins. Factor 1 was our measure of "conversation enjoyment". The full questionnaire is in Appendix A.

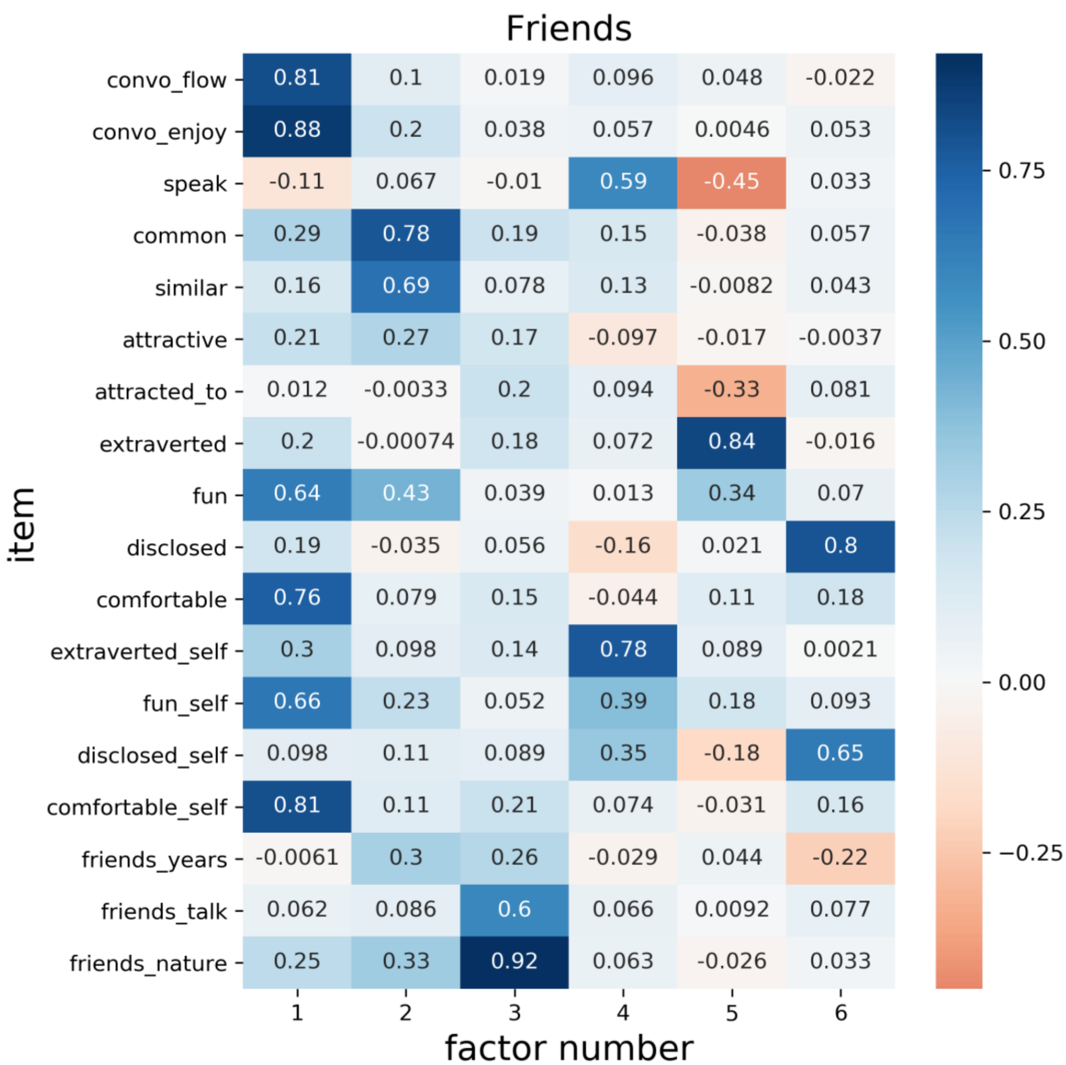


Figure S2: Factor loadings for Study 2. Factor loadings for questions answered after friend conversations. As in Study 1, the first factor maps onto “conversation enjoyment”. The full questionnaire is in Appendix B.

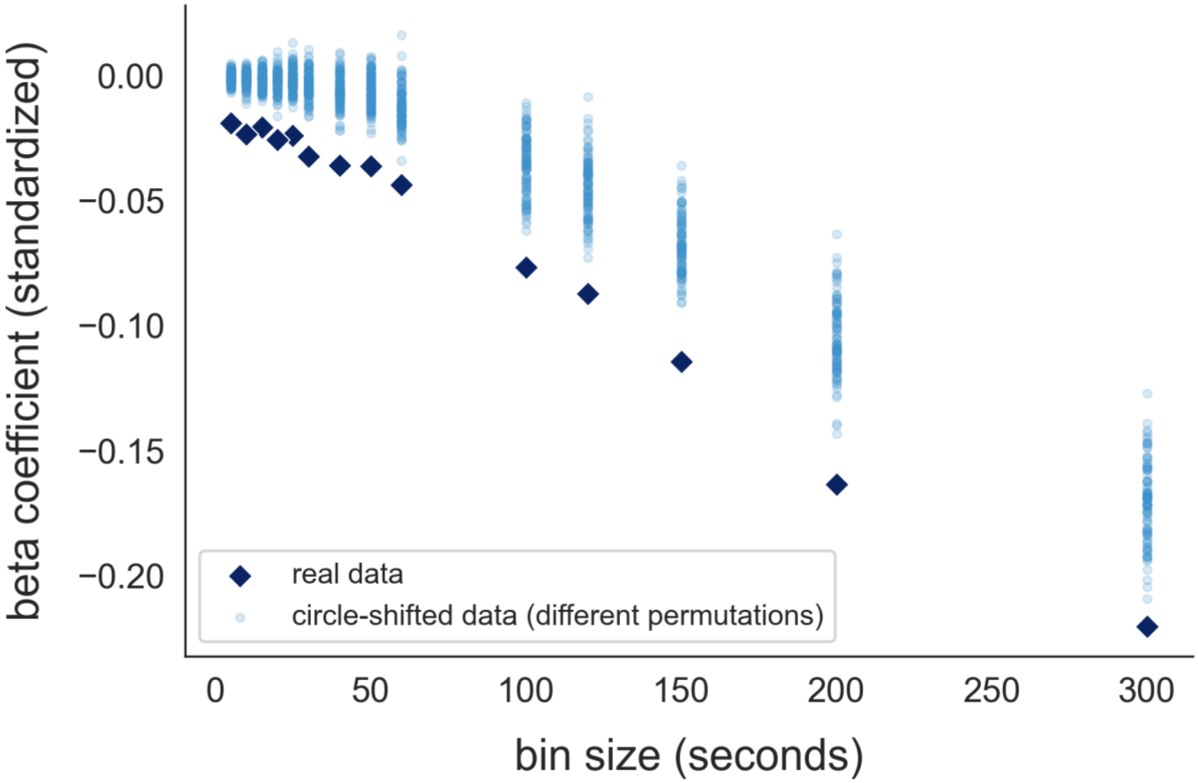


Figure S3: Robustness checks for within-conversation analysis (Study 1). First, how does the relationship between response time and connection ratings within a conversation change by how time is binned? We plot the beta coefficients for this effect across a range of different bin sizes ("real data"). Every bin size yielded a significant ($p < .05$) effect of response time on connection. Second, do these estimates outperform an empirical null distribution? For each bin size, we generated surrogate data by randomly permuting the order of response times within each conversation using a circle-shifting procedure and re-fitting the model predicting social connection 100 times. We plot the beta coefficients for these effects at each bin size, for each permutation ("circle-shifted data"). In general, the smaller the bin size the smaller the magnitude of the effect. One reason for this is that as bin size decreases, the likelihood of missing data also increases. This is because average response time for a given bin can only be computed if a turn occurred in that bin. Additionally, larger bin sizes reflect effects occurring at longer timescales. Because we observed slow and gradual increases in connection ratings over the course of the gap lengths from connection ratings at larger bin sizes.

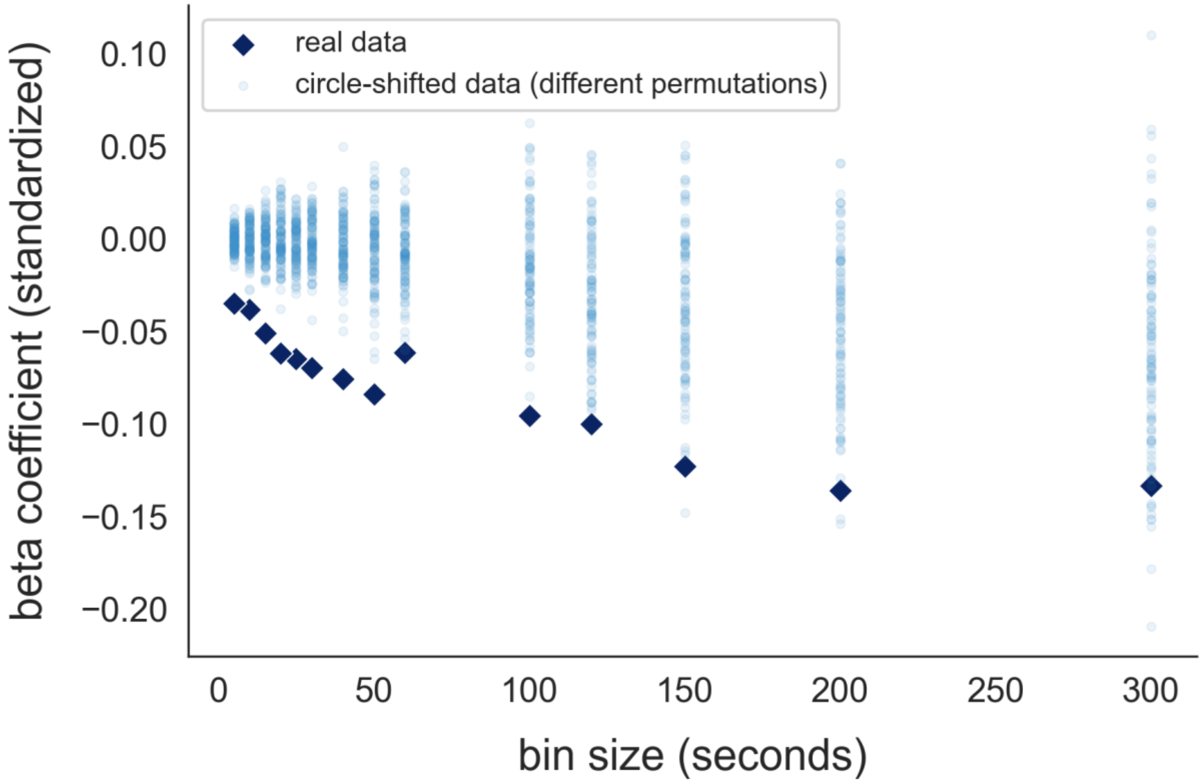


Figure S4: Robustness checks for within-conversation analysis (Study 2). First, how does the relationship between response time and connection ratings within a conversation change by how time is binned? We plot the beta coefficients for this effect across a range of different bin sizes ("real data"). Every bin size except the largest one (300 seconds) yielded a significant ($p < .05$) effect of response time on connection. Second, do these estimates outperform an empirical null distribution? For each bin size, we generated surrogate data by randomly permuting the order of response times within each conversation using a circle-shifting procedure and re-fitting the model predicting social connection 100 times. We plot the beta coefficients for these effects at each bin size, for each permutation ("circle-shifted data"). In general, the smaller the bin size the smaller the magnitude of the effect. One reason for this is that as bin size decreases, the likelihood of missing data also increases. This is because average response time for a given bin can only be computed if a turn occurred in that bin. Additionally, larger bin sizes reflect effects occurring at longer timescales. Because we observed slow and gradual increases in connection ratings over the course of the conversation, there is more variance to model the gap lengths from connection ratings at larger bin sizes. However, unlike the stranger conversations in Study1, friends tend to exhibit less of a gradual change as the connection ratings are already quite high for each conversation.

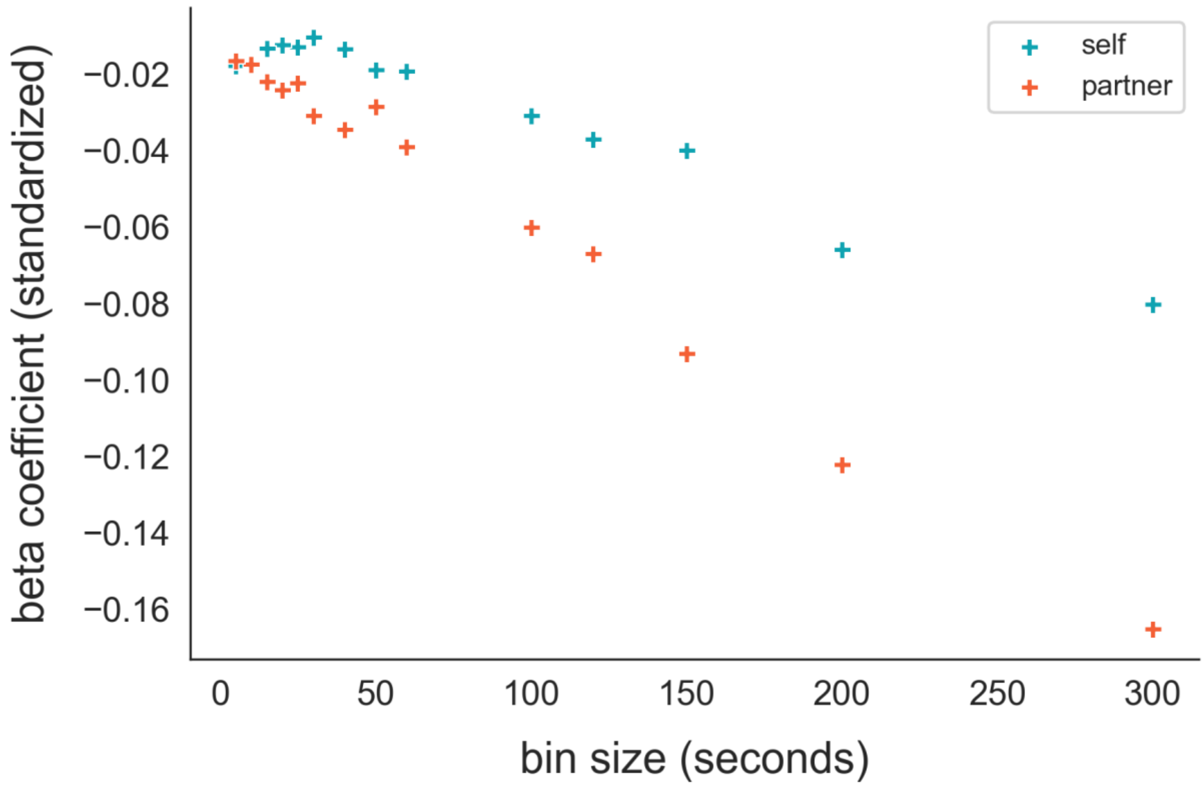


Figure S5: Self / partner effects across different bin sizes (Study 1). How does the relationship between self and partner response time and connection ratings within a conversation change by how time is binned? We plot the beta coefficients for both self and partner effects across a range of different bin sizes. For every bin size, the partner effect is stronger than the self effect. The partner effect yielded a significant ($p < .05$) effect of on connection for every bin size.

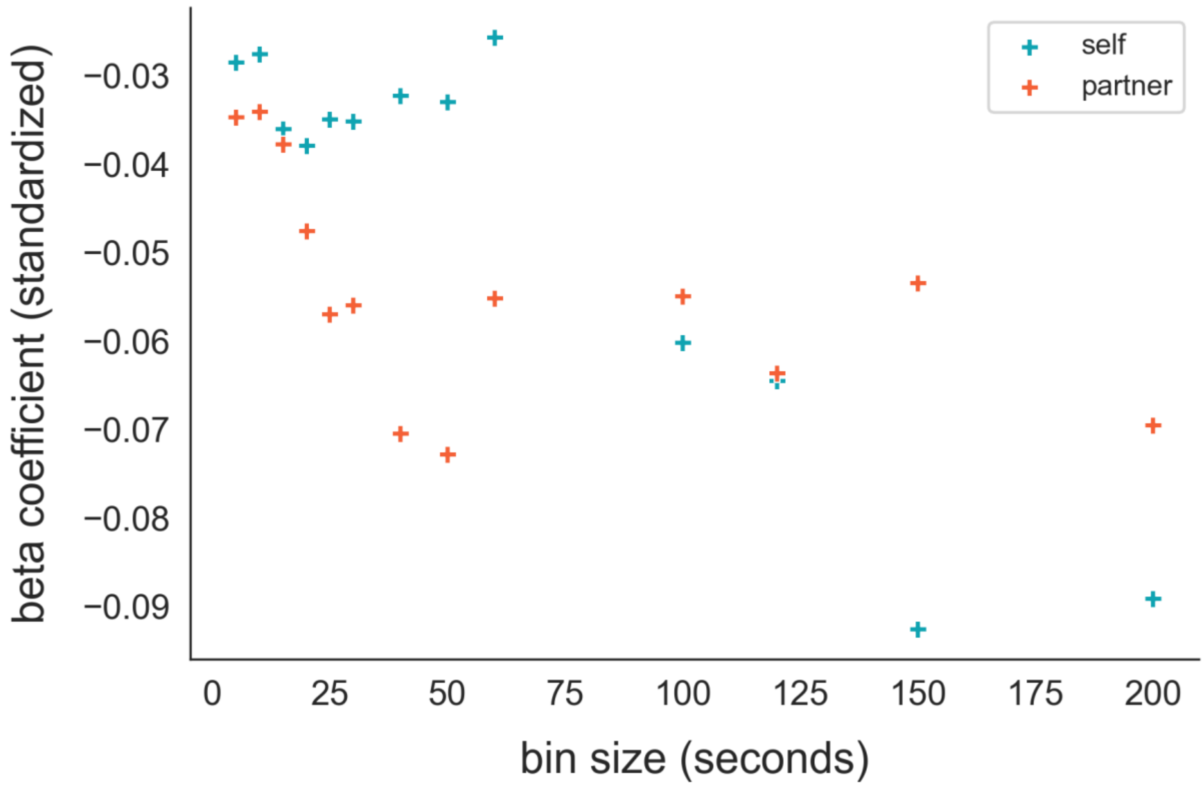


Figure S6: Self / partner effects across different bin sizes (Study 2). How does the relationship between self and partner response time and connection ratings within a conversation change by how time is binned? We plot the beta coefficients for both self and partner effects across a range of different bin sizes. The partner effect is consistently stronger than the self effect for all bin sizes 60 seconds and less. The partner effect consistently yielded a significant ($p < .05$) effect on connection ratings for bin sizes 60 and less. The self effect consistently yielded a significant ($p < .05$) effect on connection ratings for bin sizes 40 and less. It is not possible to detect a relationship between response time and connection at larger bin sizes because connection ratings for close friends are overall higher. Therefore, the variance in connection greatly decreases at larger bin sizes.

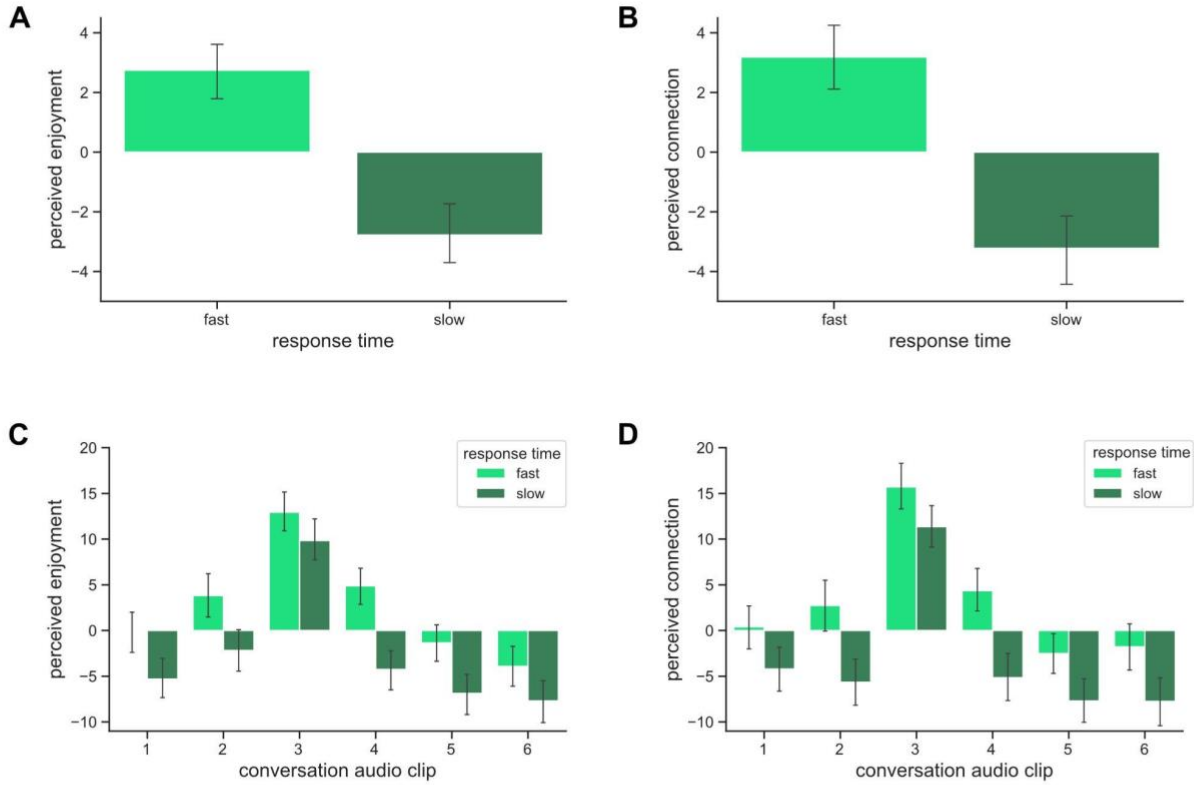


Figure S7: Replication of Study 3 results in a different sample. Main effect of condition (Short, Long) on (A) perceived conversation enjoyment and (B) connection. Effect of response time condition on perceived (C) enjoyment and (D) connection broken down by conversation audio file. All values are centered within-subject to reflect the random effect structure used in the mixed-effects model. Error bars depict 95% confidence intervals.

Table S1: Participants rated conversations with friends more positively than conversations with strangers. Differences in conversation ratings between stranger dyads and friend dyads. As expected, participants rate conversations with their friends more positively than conversations with strangers.

Variable	Mean Friends	Mean Stranger	Welch's t test
convo_flow	88.02	73.25	t(278.12)=10.22, p<.001 ***
convo_enjoy	87.95	72.55	t(251.28)=10.15, p<.001 ***
speak	54.22	53.51	t(168.03)=0.45, p=.66
common	78.64	53.18	t(202.48)=13.33, p<.001 ***
similar	66.65	53.12	t(169.83)=5.73, p<.001 ***
attractive	78.57	55.24	t(220.96)=13.09, p<.001 ***
attracted_to	24.02	8.62	t(149.10)=6.09, p<.001 ***
extraverted	70.25	59.45	t(190.47)=5.02, p<.001 ***
fun	82.57	68.20	t(245.52)=9.35, p<.001 ***
disclosed	55.78	44.16	t(162.19)=4.62, p<.001 ***
comfortable_self	88.86	73.86	t(232.85)=9.87, p<.001 ***

Appendix A

Post-conversation survey items for round-robin dataset (Strangers)

In this short survey, you will make ratings about the conversation you just had. Please answer the following questions about your experience as honestly and completely as possible. Your responses to these questions will be kept confidential and only identified by a numeric identifier, not your name.

1. How well did this conversation "flow"? *(0=Not at all, 100=Very)* **[**variable name = **convo_flow]**
2. How much did you enjoy the conversation you had with your study partner? *(0=Not at all, 100=Very much)* **[convo_enjoy]**
3. How much would you like to be friends with your study partner? *(0=Not at all, 100=Very much)* **[friends]**
4. Think about how much you and your study partner each talked during your conversation and indicate your relative contributions on the scale below *(0=My partner spoke much more than I did, 50=My study partner and I spoke the same amount, 100=I spoke much more than my study partner did)* **[speak]**

5. How well did you know your study partner before today? *(0=Not well at all, 50=Moderately well, 100=Extremely well)* **[knew_before]**
6. If you knew your study partner before today, in what capacity did you know them? *(free response)* **[knew_before_text]**
7. How well did you think you know your study partner now? *(0=Not well at all, 50=Moderately well, 100=Extremely well)* **[know_now]**

8. My study partner and I seemed to have a lot in common. *(0=Strongly disagree, 100=Strongly agree)* **[common]**
9. My study partner and I seemed to have similar personalities. *(0=Strongly disagree, 100=Strongly agree)* **[similar]**
10. My study partner is an attractive person. *(0=Strongly disagree, 100=Strongly agree)* **[attractive]**
11. I am physically attracted to my study partner. *(0=Strongly disagree, 100=Strongly agree)* **[attracted_to]**

12. My study partner seemed to be an extroverted person. *(0=Strongly disagree, 100=Strongly agree)* **[extraverted]**
13. My study partner was a fun person to talk to. *(0=Strongly disagree, 100=Strongly agree)* **[fun]**
14. My study partner disclosed a lot of personal information during our interaction. *(0=Strongly disagree, 100=Strongly agree)* **[disclosed]**
15. My study partner felt comfortable having a conversation with me. *(0=Strongly disagree, 100=Strongly agree)* **[comfortable]**

Please rate your agreement with the following statements, as they relate to the conversation you JUST HAD.

16. I was extroverted in that conversation. *(0=Strongly disagree, 100=Strongly agree)* **[extraverted_self]**
17. I was a fun person to talk to in that conversation. *(0=Strongly disagree, 100=Strongly agree)* **[fun_self]**
18. I disclosed a lot of personal information during that conversation. *(0=Strongly disagree, 100=Strongly agree)* **[disclosed_self]**
19. I felt comfortable having a conversation with my study partner. *(0=Strongly disagree, 100=Strongly agree)* **[comfortable_self]**

The conversation you just had was about 10 minutes long. Sometimes people feel ready for a conversation to end before it actually ends. Sometimes people don't feel that way.

Think back to your conversation. Was there a point in the conversation when you felt ready for it to end? Or do you wish it had gone on longer?

20. How do you feel about the length of the conversation you just had? *(0=I wish it had been much shorter, 50=It was exactly the right length, 100=I wish it had been much longer)* **[length_self]**
21. How do you think YOUR PARTNER felt about the length of the conversation you just had? *(0=They wish it had been much shorter, 50=They thought it was exactly the right length, 100=They wish it had been much longer)* **[length_partner]**

Notes about these survey items.

- Questions **1-14** and **19** were asked across *all* round robins and were therefore the questions that we entered into the factor analysis
- Questions **20-21** were included for a collaborator and were not analyzed by us
- Round Robin **1** answered questions: **1-14**, **19**
- Round Robins **2** & **3** answered questions: **1-19**
- Round Robins **4**, **5**, & **6** answered questions: **1-21**

Appendix B

Post-conversation survey items for Friend dataset

In this short survey, you will make ratings about the conversation you just had. Please answer the following questions about your experience as honestly and completely as possible. Your responses to these questions will be kept confidential and only identified by a numeric identifier, not your name.

1. How well did this conversation "flow"? *(0=Not at all, 100=Very)* **[**variable name = **convo_flow]**
2. How much did you enjoy the conversation you had with your friend? *(0=Not at all, 100=Very much)* **[convo_enjoy]**
3. Think about how much you and your friend each talked during your conversation and indicate your relative contributions on the scale below *(0=My partner spoke much more than I did, 50=My study partner and I spoke the same amount, 100=I spoke much more than my study partner did)* **[speak]**

Please rate your agreement with the following statements:

4. My friend and I have a lot in common. *(0=Strongly disagree, 100=Strongly agree)* **[common]**
5. My friend and I have similar personalities. *(0=Strongly disagree, 100=Strongly agree)* **[similar]**
6. My friend is an attractive person. *(0=Strongly disagree, 100=Strongly agree)* **[attractive]**
7. I am physically attracted to my friend. *(0=Strongly disagree, 100=Strongly agree)* **[attracted_to]**

Please rate your agreement with the following statements:

8. My friend seemed extroverted in that conversation. *(0=Strongly disagree, 100=Strongly agree)* **[extraverted]**
9. My friend was a fun person to talk to in that conversation. *(0=Strongly disagree, 100=Strongly agree)* **[fun]**
10. My friend disclosed a lot of personal information during our interaction. *(0=Strongly disagree, 100=Strongly agree)* **[disclosed]**
11. My friend felt comfortable having a conversation with me. *(0=Strongly disagree, 100=Strongly agree)* **[comfortable]**

Please rate your agreement with the following statements, as they relate to the conversation you JUST HAD:

12. I was extroverted in that conversation. *(0=Strongly disagree, 100=Strongly agree)* **[extraverted_self]**
13. I was a fun person to talk to in that conversation. *(0=Strongly disagree, 100=Strongly*

agree) **[fun_self]**

14. I disclosed a lot of personal information during that conversation. *(0=Strongly disagree, 100=Strongly agree)* **[disclosed_self]**
15. I felt comfortable having a conversation with my friend. *(0=Strongly disagree, 100=Strongly agree)* **[comfortable_self]**

Please answer the following questions about the friend you just talked to.

16. How long have you been friends with them? *(0, 1, 2, 3, 4, 5yrs)* **[friends_years]**
    a. You indicated that you've known your friend for at least 5 years. If you've known them for LONGER than 5 years, please indicate that here: *(open response)* **[friends_years_extended]**
17. How frequently do you talk to this friend? *(0=Monthly, 50=Weekly, 100=Daily)* **[friends_talk]**
18. How would you characterize the nature of your friendship with this person? *(0=acquaintances, 25=friend, 75=close friend, 100=best friend)* **[friends_nature]**
19. Pick the gender that you most identify with: *(Female, Male, Other, Prefer not to answer)* **[gender]**

Supplementary Materials: Chapter 1b

Study 1

Transcription Details

We defined gap lengths based on the timestamps in each conversation transcript. Here we detail exactly what the transcripts contained and our decisions about what constituted a speech 'turn.'

Format. Below is a screenshot of the first few turns in one transcript, to illustrate the format of the transcripts. Each turn included (i) speaker information (S1 or S2), (ii) a START timestamp, (iii) an END timestamp, and (iv) text of the words spoken.

To compute gap length, we subtracted the END timestamp of the previous turn from the START timestamp of a given turn.

```
S1: 00:00:00.000 How's it going? END 00:01.335 END

S2: 00:00:02.236 I'm okay. How are you? END 00:03.215 END

S1: 00:00:03.123 Yeah, I'm good. It's been a busy... END 00:05.599 END

S2: 00:00:05.636 It's been a week. END 00:06.043 END

S1: 00:00:06.397 Yeah, honestly. With rush. Did you rush? END 00:08.880 END

S2: 00:00:08.761 I dropped on Monday. END 00:09.686 END

S1: 00:00:09.582 You dropped? Oh, I'm sorry. END 00:10.731 END

S2: 00:00:10.388 Yeah, I didn't get callbacks. END 00:11.686 END

S1: 00:00:12.067 Oh, okay. END 00:12.432 END
```

Transcription Company. The transcriptions (and therefore the timestamps) that were for the long gap length analyses were performed by one company -- Scribie (https://scribie.com/). More details about the transcription that was completed for each individual conversation video can be found in the Supplement folder of this project's Github repository

(ConversationDatasetDetails.doc).

Speaker-switches vs speaker-stays

In two-person conversations, people typically take turns back and forth. After one person stops speaking, the other person begins speaking. Occasionally, the speaker who last spoke is the one who decides to speak again next. This would be an example of a 'speaker-stay' (vs a 'speaker-switch'). Some researchers may consider this gap in between speaker-stay turns as a silence within a turn rather than an inter-turn gap. Because the focus of this paper is on gaps that are more than 2 seconds long (10x longer than the modal gap length in conversation), we believe it is likely that these long gaps are still *experienced* as gaps by the participants.

We examined the frequency of speaker-switches vs speaker-stays for the turns surrounding the long gaps in our stranger and friend datasets. Strangers had 14 instances of speaker-stays vs 248 instances of speaker-switches. Friends had 19 instances of speaker-stays vs 183 instances of speaker stays. Given the rarity of these events (~7% of long gaps), we did not run any analyses comparing turns with speaker-switches vs speaker-stays. However, we have included this speaker-switch information for each long gap in the Github repository for this project.

Because instances of speaker-stays were rare and because long gaps are likely to be experienced as gaps even when there is a speaker-stay, we included all instances of long gaps in our Study 1 analyses. Note that for Study 2, none of the clips used had gaps that were speaker-stays (all were speaker-switches).

Additional laughter analyses

We investigated whether laughter during the long gap might mediate the effect between relationship type (friend vs stranger) and change in connection. We examined the video footage of all of the long gaps and annotated whether participants laughed during the gap. We first tested for an effect of relationship type on change in connection and found significant effects when entering a long gap (b = 1.03, SE = 0.35, p = 0.004). We next tested for the effect of relationship type on laughter and found a significant effect, such that friends tend to laugh more during long gaps than strangers (b = -0.47, SE = 0.22, p = 0.032). Finally, we tested for the effects of relationship type and laughter on change in connection. When accounting for the effect of laughter, the effect of relationship type still significantly predicted changes in connection when entering a long gap (b = 0.95, SE = 0.35, p = 0.007). This suggests that laughter does not mediate the effect between relationship type and change in connection. Causal mediation analyses confirmed no evidence of a mediation effect (indirect effect = -0.06, p = 0.080).

Exploratory semantic analyses

Previous research using stranger dyads has demonstrated that "minimal responses" (responses that fail to advance the topic) tend to happen *before* a long gap and that question asking tends to *follow* long gaps (Dindia, 1986; McLaughlin & Cody, 1982). These findings align with our own observations of long gaps between strangers.

As a preliminary investigation into the content of what participants tend to say near a long gap, we examined the text of the turns immediately preceding and immediately following each gap. Gaps were categorized based on their lengths (>2000ms (indicating a long gap) and ≤ 2000ms (all other gaps)). We coded for (i) the number of words (as a proxy for minimal responses) and (ii) the presence of a question mark (as a proxy for question asking).

We found that strangers asked more questions and spoke more words after a long gap (> 2000ms, Fig S11) compared to friends. We take this as preliminary evidence that strangers may be "overcompensating" for a long gap by speaking more (as evidenced by the increased word count) and changing the topic (as evidenced by the increased number of questions). However, more comprehensive semantic analyses are needed to better contextualize these initial results.

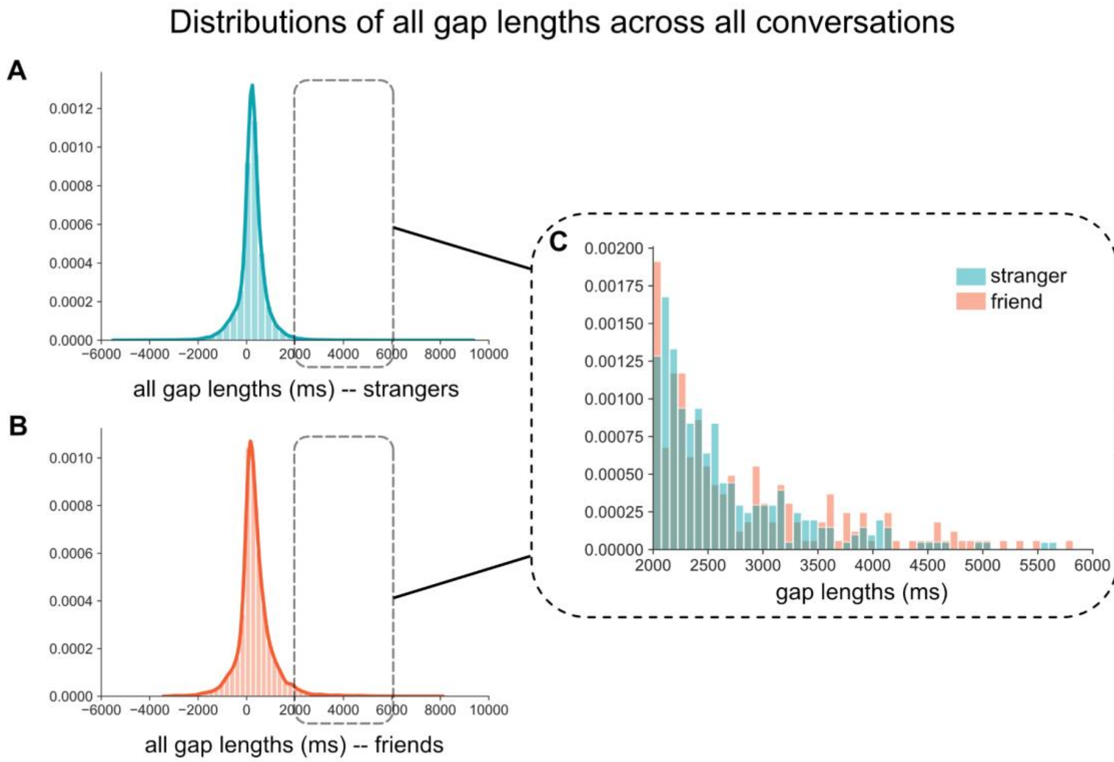


Figure S8. (A) Distributions of all gap lengths across all stranger conversations. (B) Distributions of all gap lengths across all friend conversations. (C) A zoomed-in version of the group distributions highlights the fact that friend conversations contain a greater proportion of long gaps. Note that there are many more stranger conversations compared to friend conversations—261 vs 65, respectively.

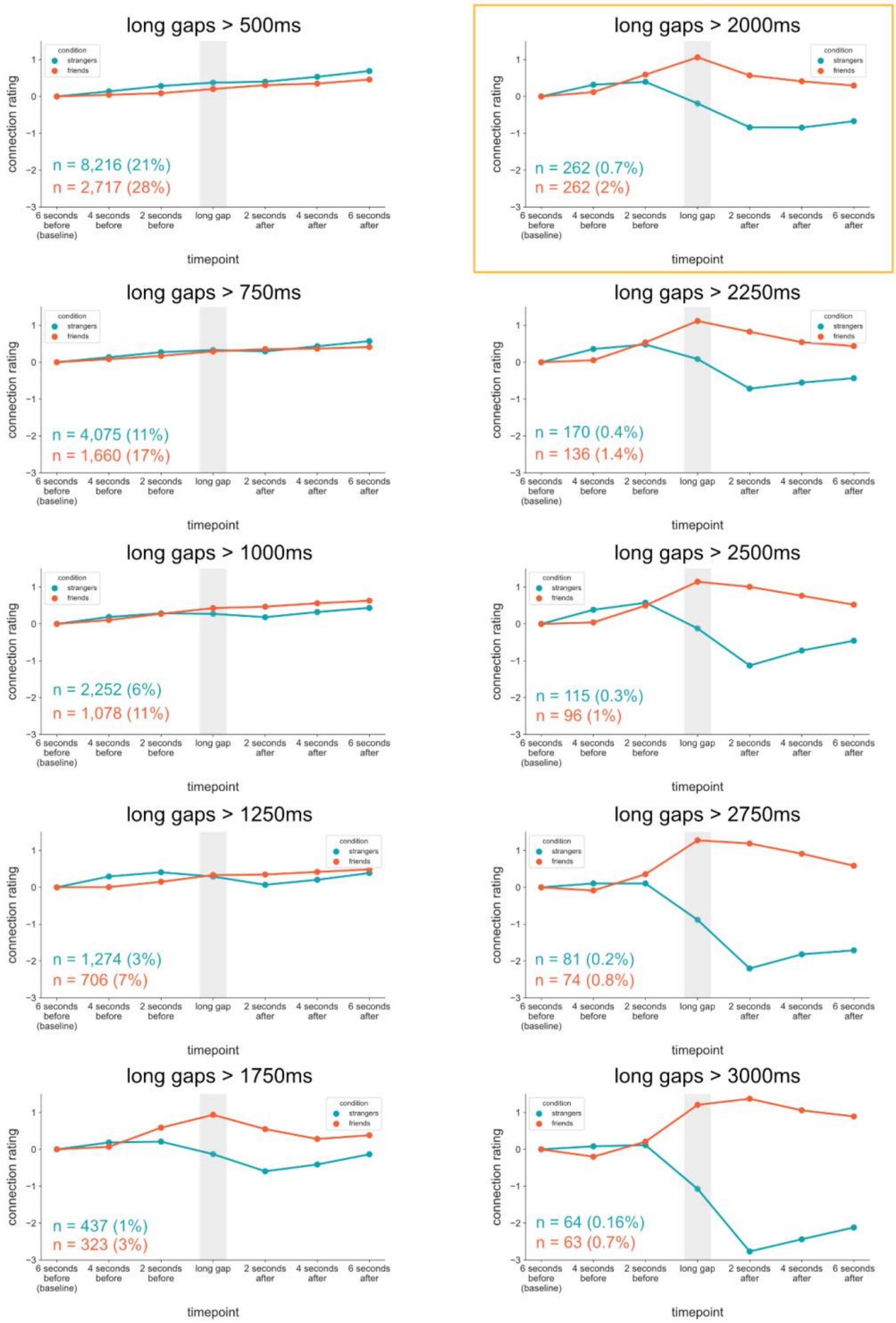
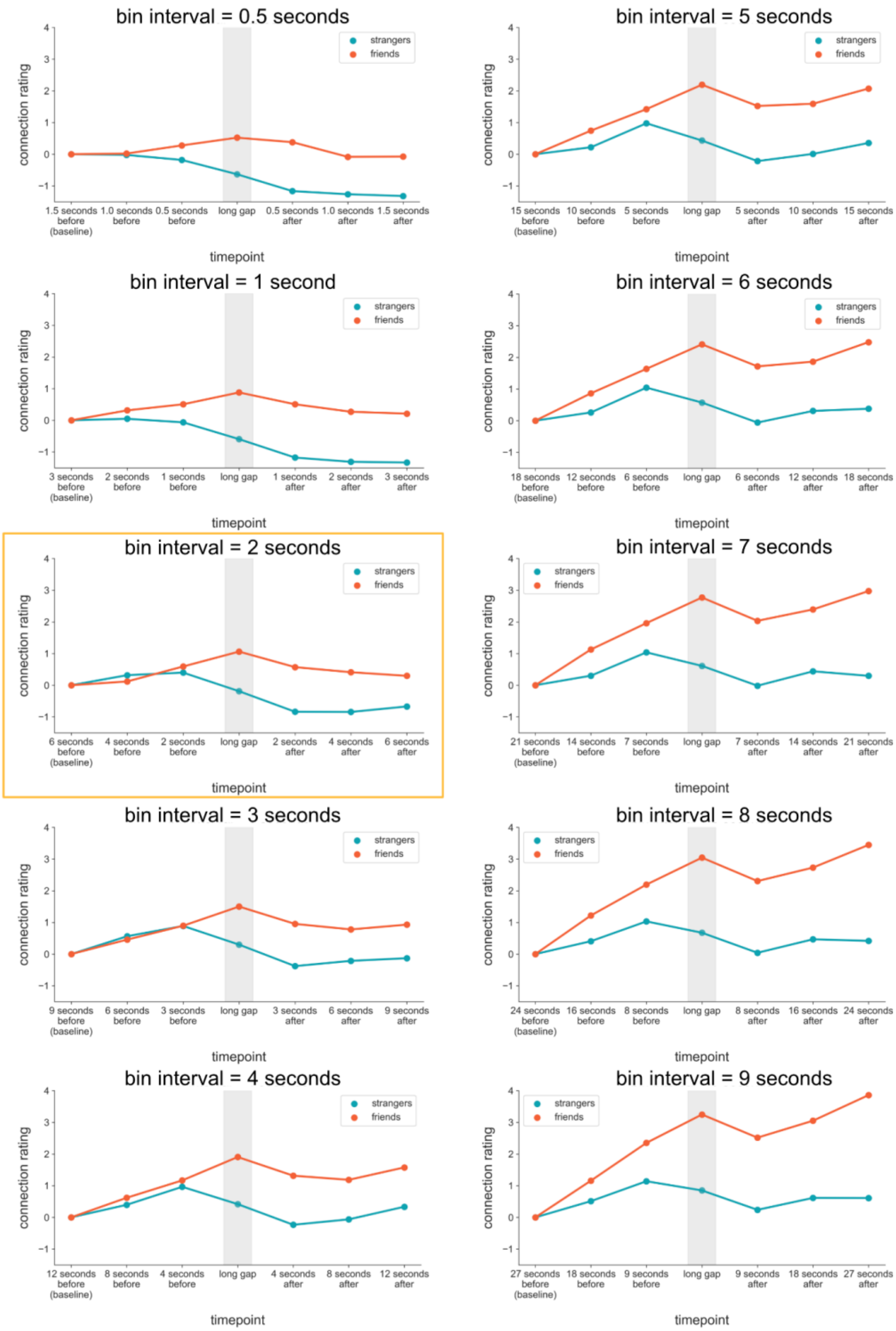

long gaps > 500ms
n = 8,216 (21%)
n = 2,717 (28%)
long gaps > 2000ms
n = 262 (0.7%)
n = 262 (2%)
long gaps > 750ms
n = 4,075 (11%)
n = 1,660 (17%)
long gaps > 2250ms
n = 170 (0.4%)
n = 136 (1.4%)
long gaps > 1000ms
n = 2,252 (6%)
n = 1,078 (11%)
long gaps > 2500ms
n = 115 (0.3%)
n = 96 (1%)
long gaps > 1250ms
n = 1,274 (3%)
n = 706 (7%)
long gaps > 2750ms
n = 81 (0.2%)
n = 74 (0.8%)
long gaps > 1750ms
n = 437 (1%)
n = 323 (3%)
long gaps > 3000ms
n = 64 (0.16%)
n = 63 (0.7%)
condition
strangers
friends
connection rating
timepoint
6 seconds before (baseline)
4 seconds before
2 seconds before
long gap
2 seconds after
4 seconds after
6 seconds after

Figure S9. Each plot depicts the average temporal dynamics of subjective feelings of connection when entering and exiting long gaps starting at an initial baseline 6 seconds prior to the gap. We plot the trajectories separately for strangers and friends. Each plot has a different threshold for defining a "long" gap. The text in blue details the number of long gaps included for strangers with each definition (and what percentage of total gaps that number represents). The text in orange details the number of long gaps included for friends with each definition (and what percentage of total gaps that number represents). Note that these values differ slightly from Table S2 because turns included here can not occur earlier than 6 seconds from the start of the conversation or later that 6 seconds from the end of the conversation. The yellow box indicates the threshold used in the main text. In general, the pattern of results gets stronger as the threshold increases, but note a tradeoff between threshold and number of observations.



bin interval = 0.5 seconds
strangers
friends
connection rating
timepoint
1.5 seconds before (baseline)
1.0 seconds before
0.5 seconds before
long gap
0.5 seconds after
1.0 seconds after
1.5 seconds after
bin interval = 5 seconds
15 seconds before (baseline)
10 seconds before
5 seconds before
5 seconds after
10 seconds after
15 seconds after
bin interval = 1 second
3 seconds before (baseline)
2 seconds before
1 seconds before
1 seconds after
2 seconds after
3 seconds after
bin interval = 6 seconds
18 seconds before (baseline)
12 seconds before
6 seconds before
6 seconds after
12 seconds after
18 seconds after
bin interval = 2 seconds
6 seconds before (baseline)
4 seconds before
2 seconds before
2 seconds after
4 seconds after
6 seconds after
bin interval = 7 seconds
21 seconds before (baseline)
14 seconds before
7 seconds before
7 seconds after
14 seconds after
21 seconds after
bin interval = 3 seconds
9 seconds before (baseline)
6 seconds before
3 seconds before
3 seconds after
6 seconds after
9 seconds after
bin interval = 8 seconds
24 seconds before (baseline)
16 seconds before
8 seconds before
8 seconds after
16 seconds after
24 seconds after
bin interval = 4 seconds
12 seconds before (baseline)
8 seconds before
4 seconds before
4 seconds after
8 seconds after
12 seconds after
bin interval = 9 seconds
27 seconds before (baseline)
18 seconds before
9 seconds before
9 seconds after
18 seconds after
27 seconds after

Figure S10. How do different bin sizes surrounding the long gaps impact the change in connection scores? Each subplots considers long gaps to be all gaps greater than 2 seconds. The subplots differ on how much time is included in each of the 6 timepoints surrounding the long gap (3 before and 3 after). The pattern of results in these subplots are quite similar to the interval used in the main text (2 seconds, highlighted with a yellow box). This demonstrates that our main effects are quite robust to the choice of interval size.

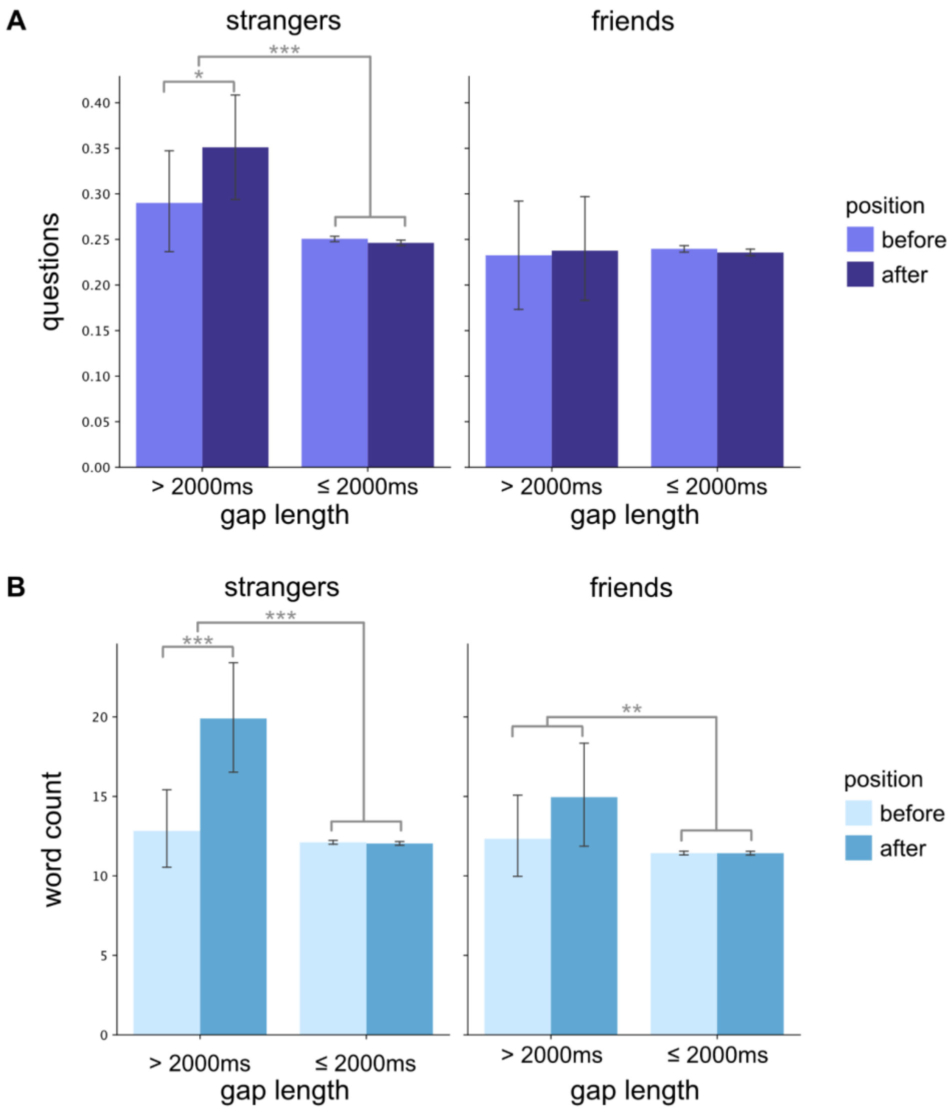


Figure S11. (A) Percentage of turns containing a question mark split by relationship type (stranger vs friend), gap length (> 2000ms (indicating a long gap) vs ≤ 2000ms) and position (before the gap vs after the gap). (B) Word count of turns containing a question mark split by relationship type (stranger vs friend), gap length (> 2000ms (indicating a long gap) vs ≤ 2000ms) and position (before the gap vs after the gap). Strangers are more likely to use turns with question marks and higher word counts immediately after a long gap. Error bars depict 95% confidence intervals. * p < .05, ** p < .01, *** p < .001

Table S2: Friends have more instances of long gaps across a variety of thresholds. In the main text, we use a threshold of 2 seconds to define a "long" gap. We selected our use of 2 seconds because it is 3 SD from the mean gap length across our datasets. However, our threshold is just one way to define a long gap. Here we test whether the frequency of long gaps differs between friends and strangers across a wide range of different thresholds. The results are consistent: Friends have more long gaps than strangers. We used a zero-inflated negative binomial regression to predict the number of long gaps based on relationship type (friend or stranger). Because different conversations had different numbers of turns, we included the total number of gaps for each conversation as an offset parameter. Because subjects could participate in multiple conversations, subject ID was included as a random intercept.

Threshold (ms)	Frequency: Strangers	Frequency: Friends	Condition Effect
500	8,508 (22%)	2,812 (30%)	$b = -0.36$, $SE = 0.04$, $p < .001$
750	4,197 (11%)	1,717 (18%)	$b = -0.62$, $SE = 0.05$, $p < .001$
1000	2,316 (6%)	1,122 (12%)	$b = -0.83$, $SE = 0.07$, $p < .001$
1250	1,315 (3%)	741 (8%)	$b = -1.00$, $SE = 0.08$, $p < .001$
1500	755 (2%)	476 (5%)	$b = -1.10$, $SE = 0.10$, $p < .001$
1750	455 (1%)	343 (4%)	$b = -1.32$, $SE = 0.11$, $p < .001$
2000	274 (0.7%)	218 (2%)	$b = -1.51$, $SE = 0.15$, $p < .001$
2250	178 (0.5%)	150 (1.6%)	$b = -1.47$, $SE = 0.16$, $p < .001$
2500	122 (0.3%)	108 (1.2%)	$b = -1.53$, $SE = 0.18$, $p < .001$
2750	84 (0.2%)	85 (0.9%)	$b = -1.65$, $SE = 0.19$, $p < .001$
3000	66 (0.17%)	70 (0.7%)	$b = -1.68$, $SE = 0.21$, $p < .001$

Table S3: Inter-Rater Reliability (IRR) scores for each variable in Study 2.

Variable name	**IRR Method***	**IRR Score**	**Typical Interpretation****
awkward	Intraclass correlation coefficient (ICC3k)	0.872	Good reliability
connected	Intraclass correlation coefficient (ICC3k)	0.838	Good reliability
topics	Intraclass correlation coefficient (ICC3k)	0.931	Excellent reliability
laughter	Cohen's Kappa (average)	0.845	Almost perfect agreement
laughter_who	Cohen's Kappa (average)	0.615	Substantial agreement
laughter_genuine	Intraclass correlation coefficient (ICC3k)	0.736	Moderate reliability
gestures	Cohen's Kappa (average)	0.366	Fair agreement

*https://scikit-learn.org/stable/modules/generated/sklearn.metrics.cohen_kappa_score.html
*https://pingouin-stats.org/generated/pingouin.intraclass_corr.html
**Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam med*, *37*(5), 360-363.
**Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.

Table S4: Effect of condition (stranger vs friend) on each variable in Study 2. For categorical variables, we used a chi-square test to examine differences in ratings by condition. We took the modal response from the independent raters for each video clip as the "consensus response". For continuous variables, we used this model: scale(variable) ~ condition + (1 | rater ID) which allowed us to use all the ratings, while still accounting for the fact that different raters might have used the continuous scales differently from each other.

Variable name	**Scale**	**Strangers**	**Friends**	**Effect**
awkward	0 = Not at all awkward, 100 = Extremely awkward	M = 43.29	M = 26.15	b = 0.59, SE = 0.11, p < .001 *******
connected	0 = Not at all connected, 100 = Extremely connected	M = 44.80	M = 65.10	b = -0.75, SE = 0.11, p < .001 *******
topics	0 = The turns were on completely different topics, 100 = The turns were on the same topic	M = 59.35	M = 70.50	b = -0.29, SE = 0.11, p = .011 *
laughter	Yes / No	13 / 37	26 / 24	X^2 (1, N = 100) = 6.05, p = .014 *
laughter_who	One person / both people	8 / 18	5 / 8	X^2 (1, N = 39) = 0.01, p = .904
laughter_genuine	1 = Not at all genuine, 9 = Extremely genuine	M = 5.73	M = 6.66	b = -0.48, SE=0.19, p = .011 *
gestures	Yes / No	16 / 34	16 / 34	X^2 (1, N = 100) = 0, p = 1.00

Note: All results hold even when accounting for whether or not the individual raters personally knew someone in the video clip. This model was used to account for this information: variable ~ condition + rater_know + (1 | rater ID)

Appendix C

Questions asked after each video clip in Study 2

1. How awkward did the gap seem? *(0=Not at all awkward, 100=Extremely awkward)* **[**variable name = **awkward]**
2. How connected did the two people seem during the gap? *(0=Not at all connected, 100=Extremely connected)* **[connected]**
3. How closely related were the two turns surrounding the gap (e.g., the turn immediately before and the turn immediately after)? *(0=The turns were on completely different topics, 100=The turns were on the same topic)* **[topics]**
4. Did any laughter occur during the gap? *(Yes / No)* **[laughter]**
5. Who laughed? *(The person on the left / The person on the right / Both people)* **[laughter_who]**
6. How genuine did the laughter seem? *(1=not at all genuine, 9=extremely genuine)* **[laughter_genuine]**
7. During the gap, did either participant seem to use any gestures with the intent of communicating something? (e.g., an exaggerated facial expression, a 'thumbs up', nodding their head, etc.) *(Yes / No)* **[gestures]**
8. Please describe the gesture *(free response)* **[gestures_describe]**
9. Do you personally know either of the people in this video? *(Yes / No)* **[rater_know]**
10. How do you know them? *(free response)* **[rater_know_how]**
11. If you have anything else you want to share about this video clip, or your rating for this video clip, please do so here: *(free response)* **[notes]**

Note: Questions #5,6 only appeared if the rater selected 'yes' for question #4. Question #8 only appeared if the rater selected 'yes' for question #7. Question #10 only appeared if the rater selected 'yes' for question #9.

Supplementary Materials: Chapter 2

Clustering Universal Sentence Encoder embeddings

Any unsupervised clustering approach requires setting parameters in advance. In this section, we document how we thought through those decisions.

Our first step was to reduce the dimensionality of the 512-dimensional feature space of the Universal Sentence Encoder embeddings. This is to deal with the "curse of dimensionality" and the fact that many dimensions are likely correlated with each other (Assent, 2012). We used uniform manifold approximation and projection (UMAP), a dimensionality reduction technique that aims to preserve distances between observations (McInnes et al., 2018). UMAP requires that we specify the number of dimensions of the reduced feature space. To help make this decision, we visualized patterns of pairwise cosine similarity between the embeddings in the original feature space. Our goal was to choose a reduced feature space that preserved the between-dyad similarity structure. We visually inspected how this pattern changed with different UMAP components (e.g., 200, 100, 50, 10, 2). We noticed that once the embeddings were reduced at all, the pattern of pairwise cosine similarity values was quite consistent (e.g., the similarity structure using a UMAP with 200 components looked quite similar to the similarity structure using a UMAP with 10 components). When the number of UMAP components dropped a lot (i.e., to 2) the pattern of similarity values because much more coarse. We opted to use a UMAP solution with 10 components to take advantage of this more granular representation, without including too many components that might adversely impact the clustering algorithm.

Performing k-means clustering requires selecting a *k*, or the number of clusters the algorithm will find. One method of doing this is the "Elbow Method", where k-means clustering is performed over a range of k values and the within-cluster sum of square values are computed for each k. Plotting all this information together should reveal an "elbow" where an increase in k does not dramatically reduce the within-cluster sum of square value. This method did not reveal a clear elbow for our data (Fig S13), though it suggested that a reasonable k might fall in the range of 10-30 clusters. To inspect this range, we performed k-means clustering for 10, 15, 20, 25, and 30 clusters. For each of those clustering solutions, we computed Silhouette scores and visualized Silhouette plots (Fig S13). The Silhouette scores for all clustering solutions were quite similar (~0.3). Finally, for each clustering solution, we also visualized a word cloud based on the word frequency of the text assigned to each cluster. We were ultimately interested in a clustering solution that results in topics that seemed (i) interpretable and (ii) varied, without being redundant. All of this led us to choose a clustering solution with 25 topics.

Although a lot of thought went into each of these decisions, we do recognize that these decisions are still quite arbitrary. It will be important in future work to show that any results are robust to these clustering decisions.

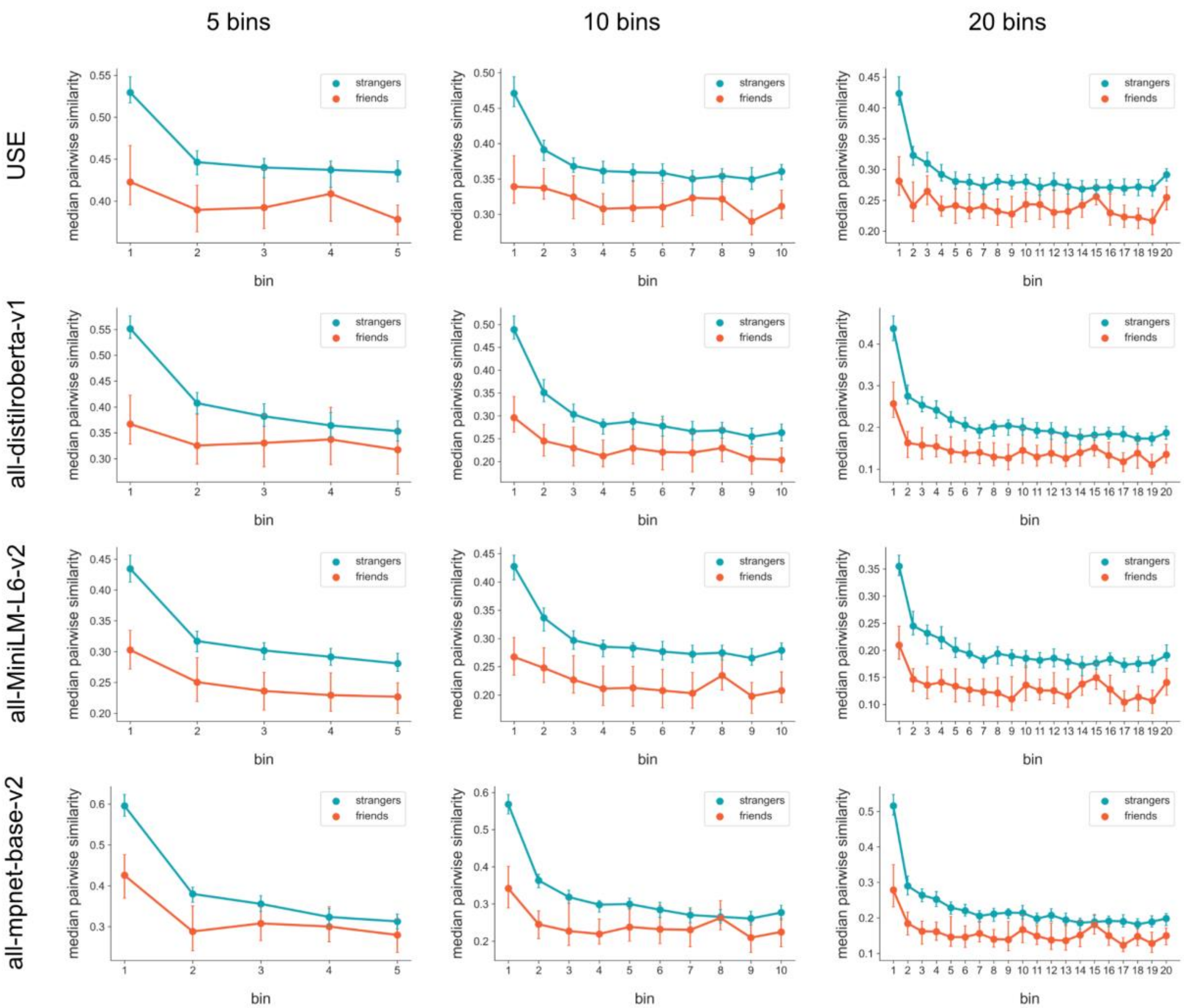


Figure S12. The result presented in Figure 13 is robust to bin size (columns) and language models (rows). Transcripts were divided into 5 bins (2 minute windows), 10 bins (1 minute windows), and 20 bins (30 second windows). Text in each bin was transformed by 4 different language models. Universal Sentence Encoded (USE, top row) is what we report in the main text. Each result is consistent: Semantic similarity for strangers is higher than for friends, and that difference is highest at the start of conversations. Confidence intervals were computed using subject-wise bootstrapping with 5,000 samples.

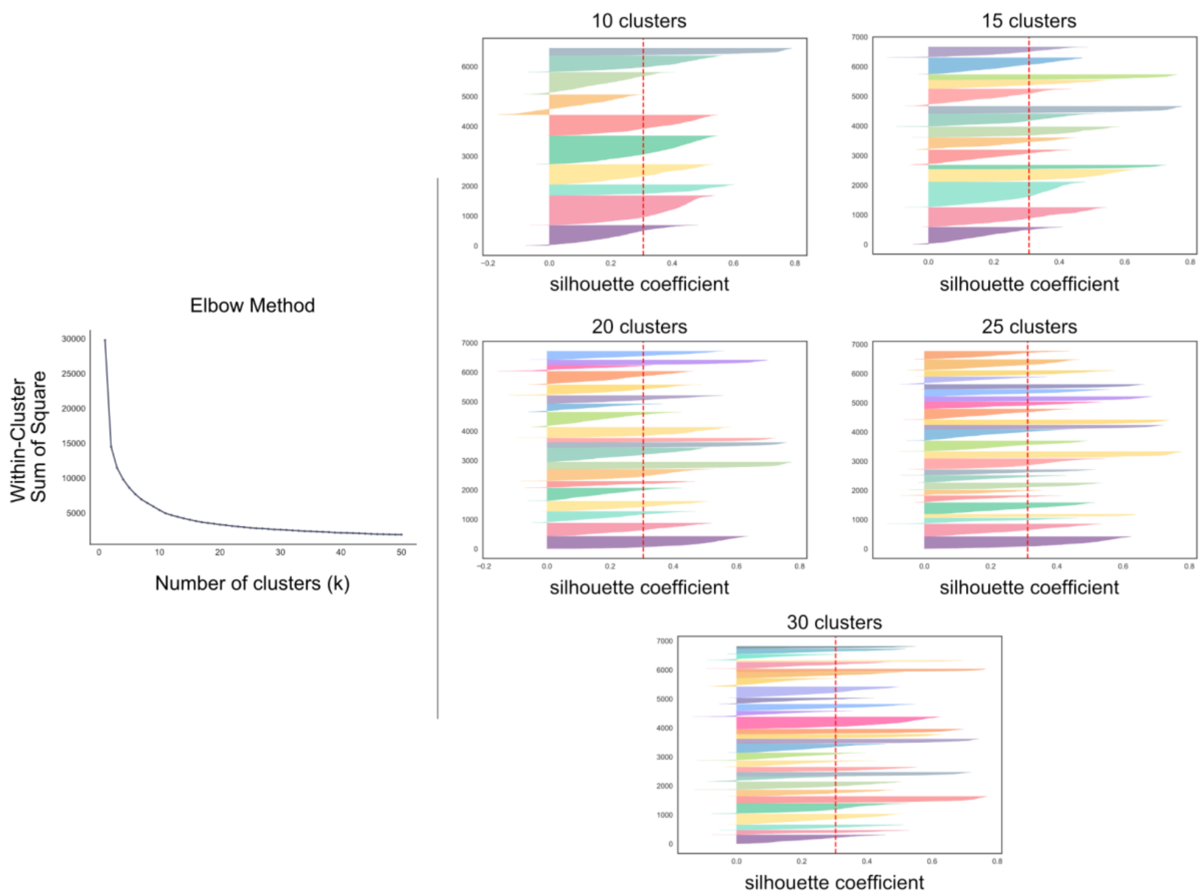


Figure S13. (Right) Elbow method for clustering Universal Sentence Encoder embeddings, for 1 - 50 clusters. (Left) Silhouette plots for 10, 15, 20, 25, and 30 clusters.

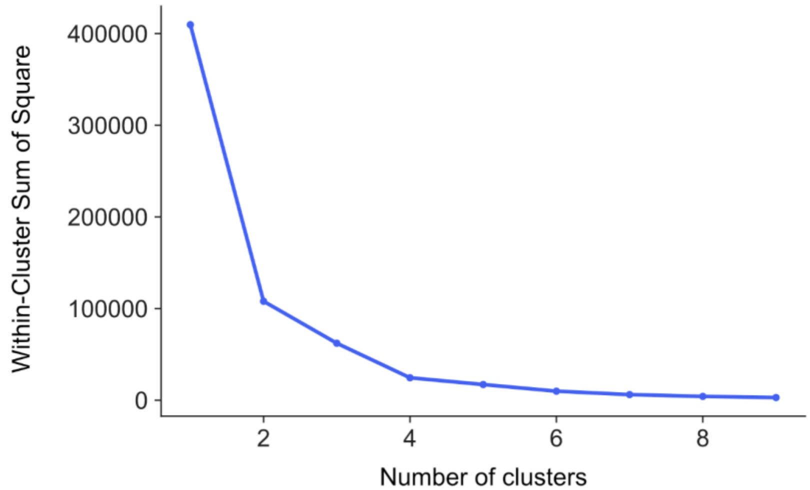


Figure S14. Elbow method based on clustering the node metrics for each topic. Nodes are part of a graph made from the stranger topic transition matrix (Figure 16). This figure suggests that the data are well described by four clusters.

Table S5. More details about the type of text assigned to each topic cluster. Example text is often composed of different turns from different speakers; those differentiations are not made here. This is the exact bin of text that gets transformed into language embeddings. Note that names have been redacted from this table to protect participant identity.

Cluster Number	**Cluster Label**	**Description**	**Example Text**
0	cities	People tend to describe their hometown and how it compares to life at Dartmouth.	When did you move to New Hampshire? Like where did you move from? Texas. What? When I was really little. I was gonna say, I feel like you've been here forever. I was like a year an a half when we went to California, and then... You were... Oh, but you didn't mention that, I feel like. It's in the intro. And then... Professor I wouldn't know. And then, moved to Texas when I was like three, and then moved to New Hampshire when I was like five. It was a really rogue move, like Cali to Texas is like fine... I know. It would've been... Texas to New Hampshire is like, "What?"
1	exams	People talking about the specifics of exams in their different classes.	He usually has 30 people in the class. This term there's 120 'cause he forgot to put a cap on it. EARS 9? What do you learn? About rocks. And what happened? It was just a horrible midterm. It was an hour for an open note test, but 10 short answer questions that were supposed to be a full... No. And 10 multiple choice. In an hour? Dude, no. That's not good. And open notes so you're going like... Trying to find everything?
2	mutual friends	People describe other people that they might both know.	We were friends in elementary school. Woah. That's really crazy. Yeah, I haven't really heard from any of my friends and that, from my Iowa, though... Yeah. It's been really long. Yeah, yeah, I feel it's, like it's pretty hard to keep in touch from just elementary school.
3	classes: economics	People talking about classes in the economics department.	I don't know. I think you have to propose a modified major for... Yeah, that's what I'm going to have to do for Econ if I want to modify it. I have to write up all the classes I'm going to take, Econ and Neuro, and then map my Neuro classes and explain how they all relate to why I want to modify the major... Yeah, that sounds really hard. So it's going to take a while. But if it saves me like six classes, then... Yeah. Do it. Might as well. That's cool. So I haven't taken a single Econ class. Are they super cut-throat like everyone says?
4	stories	Telling stories, often involving one person and	How... Stop. How was your night last night? Oh, were you just inside or... I just... Oh, yeah. I stayed inside. I basically just painted and it started off... Yeah. Yeah, it

		someone else not present in the conversation.	started being pretty cool. I was just abstracting and stuff and it blurred out. And then I tried to draw this person and then he was mad. He looked mad. I was like, "I don't know what to do about this." If it makes you feel better, I think NAME picked up your painting this morning in the common room and said like, "Hey, who did this? This is really cool." And I was like…
5	what classes to take	People describe their thoughts about different classes and give advice about what classes to take.	I'm like trying to figure out kind of what I'm most interested in. So yeah, I was like a little interested in it, so I thought it'd be a good first class to take. It's really interesting. I mean it's not what I wanna do with my life, but it's very interesting to learn about. Yeah, it is. I'd say the first... What are we? We're on almost week eight now? The first like three weeks were kind of boring, but I think that the last four weeks have been pretty cool.
6	technology	People talk about their computer programing classes or about technology in general (e.g., jobs in technology, advances in technology).	Oh my gosh, no, I found NAME. She was... Oh, you did find a coding fairy. Yeah, she helps me. I talked through the code with her. We pseudo-coded. So now I know CS words. Pseudo Code. I was so... Yeah, that's right. Yeah, so we pseudo-coded and I was like, "Ah, I see." And I did it and it worked.
7	outdoor activities	People talk about outdoor activities (e.g., skiing, hiking, biking, etc).	Just downhill? Yeah, just downhill, not nordic, just downhill. You only need three guys to score from every race, and he has four reoccurring spots a year. So that's way more than enough? So he doesn't need a super deep, huge team. I mean, obviously it's good to have a lot of good guys, but he really just wants a couple of great guys. How many downhill... So there are six to eight downhill skiers or more?
8	upcoming plans	Talking about something that will happen in the near-future.	I love it. I mean it's a lot of work right now, but it's just like a great community, you know. I'm excited for the frat bans and stuff 'cause then the... You guys are lucky 'cause like homecoming's early this year. 'Cause last year... Yeah. But we still have six weeks. You still have six weeks, but for last year for us it was eight weeks, I think. I know, that hurt, that sucks. It's crazy. That's crazy long but I mean there's so many fun like with dry parties and stuff like this week. There's one at KDE tonight. Are you gonna go? There's like three different ones tonight. I know. Social calendar. Oh my gosh. My first time hopping around. Perfect. You gotta love it. First of many.

9	food	Talking about food and on campus dining.	There was a bunch of spicy chick peas and like... I don't know. Some of them looked like they were healthy alternatives. Some of them just looked like, I'd rather not kind of things, you know? But there's some poster for it in the hall. Yeah, that's probably where I saw it. Yeah, I don't know. I appreciate you guys trying. But I think I'm mostly gonna be cooking next year. Yeah, I'm so over DBS at this point. I know. I literally walked into Novack today, and just the smell of the bagels was like... I know you pay for them. What did you think of last night? That was super fun. I thought it was really fun.
10	playing sports	People talk about sports that they've played or currently play.	Are you in the field hockey team? Yeah, I am. Do you do... How's that? It's good, I like it. It's definitely a big commitment, but I like it 'cause you get really close with your teammates, and like it kinda like... Sorry, it kinda like keeps your day structured. Like you can't sleep the day away, because you have to go to training and stuff, but... But yeah, I like it. Do you do any clubs or anything?
11	introductio ns	People introduce themselves and exchange names.	What's your name? NAME. NAME, I'm NAME. I'm a '20. Nice to meet you. I'm also a '20. What class are you in late? Did you... Um, PSYC 1. Me too! How do you like it so far? I love it. It's... I think the professors are great and... I agree. How come you took it?
12	classes: science	People talking about STEM classes.	That's interesting. Have you taken CHEM 5 and all that? Yeah, I took CHEM 5 last term. I'm taking CHEM 6 next term. Chem sounds hard. One of my friends is in it and she like dies. But yeah.
13	greek life	People talking about fraternities and sororities on campus.	Yeah, or I don't know, he's fine, but like he was the only one I knew. And then the upperclassmen and the rest of the upperclassmen, I don't know them? I don't know, just like, really nice. Though I think the whole thing with TDX is it's not necessarily like the brothers that are like assholes or whatever that gives them that reputation. But it's a team. It's literally just like, exactly like TDX exists because people want it to exist like the fact that people are like, "Okay, I want there to be a space that I can go to at 2 AM and probably hook up with someone." You know?
14	watching sports	People talking about professional sports teams and players.	Yeah, they're like good... Really, they pressured the Rockets... They beat the Rockets by 40. That was just terrible, oh my god. No, that's pretty wild though, it's definitely cool that at least you have the sports teams that have won championships. If you are from DC, I don't think, a DC team's won a sports championship in any of

			the four major leagues since 1992, so my entire life... You know the Cleveland curse, right though? Yeah, but you guys are getting better, you know. We just... Baker Mayfield, man, he's the future.
15	languages	People talk about studying languages or speaking other languages.	Oh, interesting. So it was grammar that got you not vocab? I mean, vocab I feel like you can solve, it's like memorization, right? But I feel like grammar is something you have to know. And I feel like especially 'cause we're not living in an environment with native speakers, grammar is probably the toughest thing to learn. That's why they have Nihongo Table. Nihongo Table? Have you gone to it? No, I don't even know what that is. What? They send emails out like every week. Is that Japanese society or is it something different?
16	professors	People talk about their professors.	Yeah, I have a friend who's going on that next year and she's an art history minor. That's so cool. Yeah, do you have classes with NAME. NAME? He's the head of the department and I'm taking a couple of classes from him. 'Cause I haven't... And they're easy. They're like... When I need a really good layup, I'm like "Hey, NAME, are you teaching a class this term? Please?"
17	professional plans	People talking about professional plans (e.g., Medical school, law school, etc).	I think that makes a lot of sense that over the summer I interned at a research center at the University of Washington that focuses on health metrics. Because I guess I was, yeah, tied in to my Global Health type interest. And so anyway, there were, that center is run pretty much by academics or doctors, people who have MDs or PhDs, but then they usually either teach or do medical work for some terms then they're researching and working on either making advances in their field during other terms. And that seems a great balance to have, I really liked, you know, the idea of being able to do that.
18	living situation	People talk about their living situations on and off campus.	It's like the Ripley, Woodward, Smith Complex over there. I was gonna say. That's cool. It's not bad. It's like so it's my housing community, that's like why I'm in there but... I don't know. I really hated it at first 'cause I was like, "Oh, it's like this dumpy like... Like I'm gonna be stuck in the same building my entire time here." But it hasn't been that bad. It worked out? That's good. Very cool. For sure. How about you? I'm in Topliff. So not bad. Pretty good location. And I have a single, which is nice. A lot of my friends ended up in the Lodge.
19	recap of week	People recap their weeks.	How was your day, NAME? It's really good. Real good. I didn't get any work done. Actually, that's a lie. I got work

			done... You're working on your problem set? Yeah, I got a problem set done, but then I definitely should have done some work for my English class but I didn't do that. What did you do instead? I... I designed the semi-Invite. You will be receiving that soon. It goes out, I think, tonight.
20	experience with weather	People talk about their feelings and experiences with the weather at Dartmouth and how it compared to other places they've lived.	How much does it rain? Like all the time? Like all the time, I think several years ago, there was a month of straight rain! That's crazy! Every single day! That's great. I wonder, if there's like, higher rates of depression or something... 'Cause that's super like... There are. Yeah, they get, what is it called, like seasonal... Yeah. Affective disorder? They get sad. Umm, yeah, so like... People started turning on lights and using good mood lighting and stuff like that.
21	quoting	This was a challenging topic to label. People have meta conversations about the study they are participating in. They also tell stories that involve directly quoting someone else.	Wait, how are you going to get to class on time? What do you mean? Doesn't your class start at like 12:15... Wait, I'm a dumbass. No... No, I have have a meeting with NAME like... Oh, NAME? Wait, did I show you my tattoo? You were like, "Oh, it's a real tattoo." No, it's not. I showed my mom, and I FaceTimed her yesterday, I was like "Mom, look what I got." And she was just like, "No." And I was like, "Yeah." And she's like, "No."
22	college choice	People discuss why they chose to attend Dartmouth and compare Dartmouth to other colleges.	So I like Dartmouth because it's a small school so it's easier for me to get to know the professors and my peers in class, I think, and so I thought that would be really good for my learning environment. It's also very rural. I wasn't really inclined to go to a school in the city like Columbia or Harvard, so... I felt it was the right choice. I feel that. I feel that. I wanted a campus when I... Not just like... Yeah, not just a city. I feel that. Why did you come to Dartmouth over Grinnell? I have... So I don't know. There were a lot of reasons but for me, I wasn't just like... I ended up making my decision at the very last minute.
23	classes: psych / neuro	People talk about their psychology and neuroscience courses.	In PSYC 1? I mean, I don't know. I'm trying to think if there's a good example in PSYC 1 that I can apply. One of the things I guess maybe I could talk about a little bit is like the sort of evolutionary focus of it. I think a lot of that could probably be carried over into sort of like how people interact with their surroundings and objects and things, which probably applies, but not a ton yet. But I

			think, I'm honestly mostly taking PSYC 1 so that I can take some of those other classes. What about you? What are you thinking major-wise, are you not sure yet or... Yeah, so I'm still figuring it out.
24	future plans	People talk about plans they have on the horizon (e.g., traveling, summer internships, post-college plans).	No, I kinda just found the job through like DartBoard which is Dartmouth's like career site. And I went down to visit and I like really liked it and so... Yeah, I kinda decided that. I'm kind of indecisive, so it was like a big deal. Like I'm like, "Okay, I'm gonna do this." That's good. It's exciting. Yeah, so will you like try to do some like medical stuff on the side or... Yeah, so the job is like in the evenings, so it's like 1:00 to 9:00. And so I think in the mornings/I don't work on Fridays. So I think like those days I'll try to like volunteer at a clinic or something. Or at the hospital or something like that. That's great. That's good.

Supplementary Materials: Chapter 3

Study 1

Post-conversation tasks

Participants were sent a personalized survey link that included survey questions and the insider language task (Appendix D). Due to file size constraints on the Qualtrics survey platform, each 10-minute conversation recording was split into 3 different video segments.

Text preprocessing for word counts

When participants used the text box to convey information that did not directly relate to insider language, it was assigned to one of four categories. First, were words that described the absence of insider language (e.g., "seemed like everything was pretty clear" or "it was all straight forward" or "none"). Second, were words that described information a typical Dartmouth student would know (e.g., "week 9 is the second to the last week in the term" or "rushing is the process of getting into sorority or fraternity or other Greek organization" or "[Hanover] is the town in which we go to school"). Third, was text that included timestamp information. Some participants chose to include this as a way of organizing their explanations of insider language (e.g., "0:16-" or "0:50 –" or "1:00 –"). Fourth, was text that signposted some information. This was similar to the timing category in that it was a way that some participants described what they were about to unpack (e.g., "The guys =" or "given the circumstances-" or "1.").

The word count scores from the four categories were subtracted from the original word count score to yield a final word count score purely reflecting the number of words used to unpack insider language. Text could only appear in one of the four categories to ensure that words were not double-counted and inadvertently removed more than once.

Categorizing instances of insider language use

Participants could write their explanations of insider language in any format they wanted. Research assistants separated this text into discrete instances of insider language. For the most part, this was easy to do: participants tended to put different explanations in new lines or separated them using commas. Afterward, research assistants agreed upon a single category for each instance (Table S6). Total counts of instances of insider language did not include instances assigned to the "Not" or "Incorrect" categories.

Each instance was further assigned to a sub-category (Table S6) that was not analyzed for the present work. However, we believe differences in distributions between these sub-categories may further elucidate how insider language is used in different contexts. In the future, we also plan to transform each instance of insider language into language embeddings and cluster them to reveal discreet topics. It will be interesting to compare the set of topics generated through unsupervised clustering with those generated by human raters.

Study 2

Additional talk frequency analyses

In the main text, we computed the amount of insider language in each conversation by averaging the proportion of turns determined to contain insider language by each rater. An alternative approach would be to consider how many raters thought each *turn* in a conversation contained insider language. We could then apply a threshold (e.g., all turns where 3 or more raters thought there was insider language) to determine the proportion of turns that contained insider language. Doing this for every possible threshold, revealed the same result: Friends who talk more often used insider language on a higher percentage of turns compared to friends who talked to each other less often. (20% threshold: $b = 0.384$, $SE = 0.109$, $p < 0.001$; 40% threshold:

b = 0.400, *SE* = 0.108, *p* < 0.001; 60% threshold: *b* = 0.389, *SE* = 0.108, *p* < 0.001; 80% threshold: *b* = 0.372, *SE* = 0.109, *p* = 0.001; 100% threshold: *b* = 0.367, *SE* = 0.110, *p* = 0.001).

To account for the fact that the distribution of talk frequency values is skewed, we can use Spearman's rank correlation to compare talk frequency and insider language. The results still hold: friends who talk more frequently use more insider language in their conversations (*r* = 0.39, *p* < .001).

Change in connection analysis

In the main text, we related insider language on a given turn to connection on that same turn. We can also examine how insider language on a given turn relates to *change in connection* on that turn (i.e., connection rating on the given turn minus the connection rating on the previous turn). A linear mixed-effects model with the insider language consensus score as a fixed effect was used to predict change in connection ratings at each turn. Conversation ID (the name of the transcript that was rated) and subject ID (the identity of the participant in the conversation who provided continuous connection ratings) were included as random intercepts. We found a significant positive relationship between the number of independent raters who thought a turn contained insider language and change in connection (b = 0.03, SE = 0.007, p < 0.001, Fig S17). Feelings of connection increased when insider language was used.

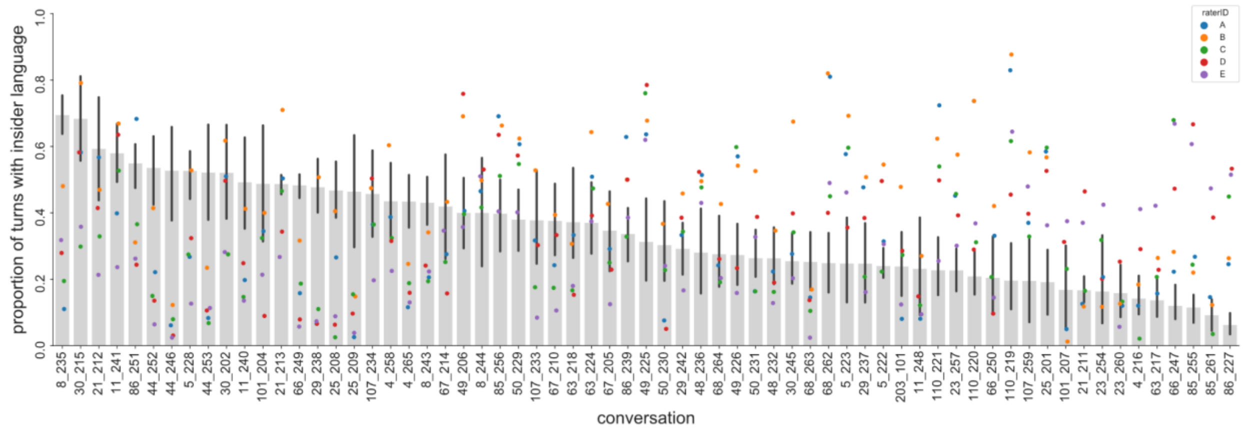


Figure S15. Proportion of turns determined to contain insider language. Individual conversations are listed on the x-axis. For each rater, we computed the proportion of turns they determined contained insider language for each conversation. These values are indicated by the individual data points, with different raters assigned to a different color. The bar graphs show the average proportion of turns with insider language, determined by taking the average of each point for each conversation. Conversations are ordered from highest to lowest average value.

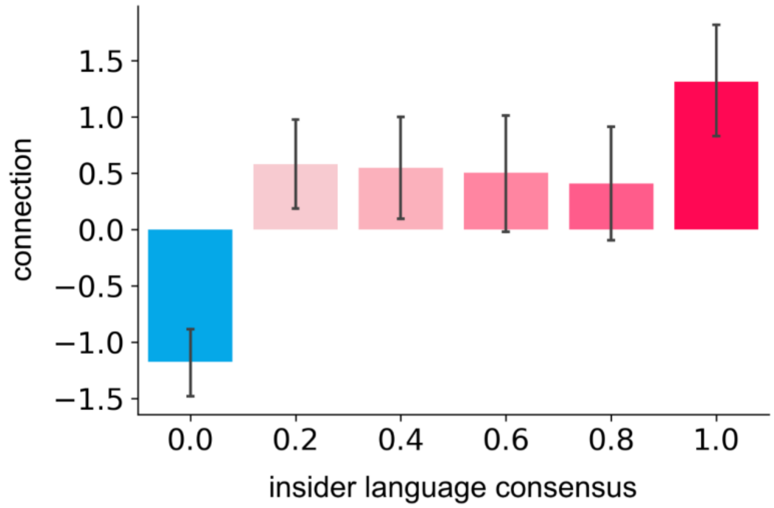


Figure S16. Connection scores by insider language consensus. Insider language consensus is the proportion of raters who thought each turn contained insider language. For example, 0 indicates turns where 0/5 raters thought they contained insider language and 1 indicates the turns where 5/5 raters thought they contained insider language. Average connection values for each turn type is depicted on the y-axis. Connection values are residualized to account for random effects of participant and dyad as well as linear effects of time. Results indicate that turns determined to contain insider language have higher connection ratings.

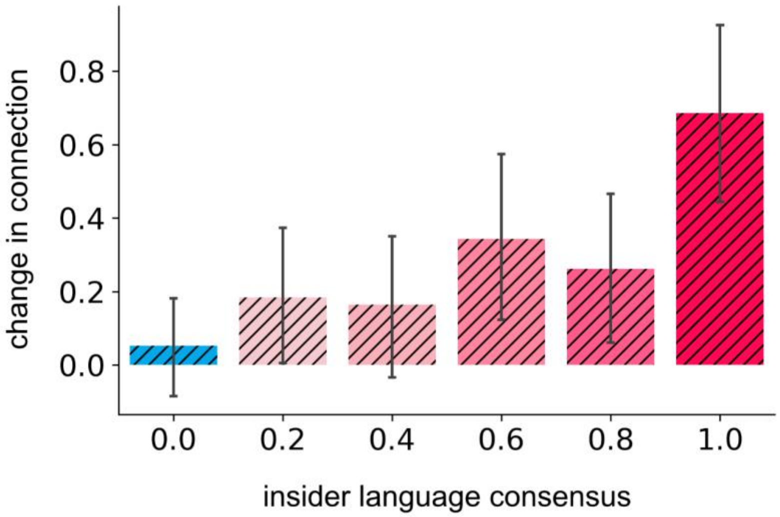


Figure S17. Change in connection scores by insider language consensus. Insider language consensus is the proportion of raters who thought each turn contained insider language. For example, 0 indicates turns where 0/5 raters thought they contained insider language and 1 indicates the turns where 5/5 raters thought they contained insider language. Average change in connection for each turn type is depicted on the y-axis. Change in connection is the connection rating for the current turn minus the connection rating for the previous turn. Results indicate that turns determined to contain insider language have higher increases in connection, compared to the previous turn.

Table S6. Insider language categories and sub-categories, with example text. Each instance of insider language was assigned to a single category and could be assigned to none, one, or multiple sub-categories. Names have been redacted to protect participant privacy.

Category	**Sub-Category**	**Description**	**Example**
Not	General	Not insider language, information a general audience would understand	Spotify is a music streaming service.
	Dartmouth	Not insider language, something a typical Dartmouth student would understand	Hanover: the town in which we go to school
Incorrect		Incorrect response type, typically asserting that there was no insider language	It was all straight forward.
Slang		Explanation of slang or lingo	When NAME said that I'm a good baker, she meant that I take good bong rips.
Expertise	Sports	Niche sports information	This Sunday there is an unusual amount really of good soccer games, meaning good teams are playing each other, so we're super excited to watch since we love to watch soccer. We talk about the timings of the various games, el clásico refers to the Real Madrid vs Barcelona game
	Academics	Niche academic information	In Psych 1, we read an article about a man who suffered from object aphasia. In the article, there were photos of the man's attempts to draw objects or copy lines of text which demonstrated that while he could see lines, he could not processes shapes as a whole.
	Extracurricular	Niche extracurricular information	Appointed position is a position that is decided by the executive committee in the sorority house, rather than being elected on by the people.
	Location	Nice information about a specific location	Topliff is known as a dorm that has a lot of parties this term.
People	Romantic	Talking about someone	NAME is my girlfriend, whose apartment

		who is romantically involved with one of the participants	is near campus.
	Authority	Talking about an authority figure (e.g., coaches, professors)	Prof. Pfister comes to class full of energy and jokes every class!
	Mutual	Talking about someone both participants know personally	NAME is a mutual friend of my partner and I who lives in our building. He has a handheld vacuum that our friend group uses frequently
	One-sided	Talking about someone one of the participants knows personally; the other participant knows *of* the person but there is no evidence they have a personal relationship with them	NAME is a friend of hers, but I do not know her.
Experiences	Past	Talking about an experience that already happened	Gile fiasco refers to when our friend group tried to use Zipcar to go to Gile but then the car malfunctioned and we ended up stranded at Gile in lightning and rain at night while we waited for help with the car.
	Future	Talking about an experience that has not happened yet	There are two small turkeys being cooked for thanksgiving this year, one in the oven and one in the smoker.
	Shared	Talking about an experience both participants had or will have together	We are going to go live in Florida with each other in a week and we are figuring out our plans.
	One-sided	Talking about an experience that only one participant had or will have; the other participants knows about the experience but there is no evidence they were personally involved	Before I left to come to Dartmouth, I had a falling out with my high school friend group.
	Secondhand	Talking about an experience that both participants heard about, but neither of	The emails were sent to people who posted something on their instagram that appeared to violate COVID rules.

		them were personally involved	
Known personal information	Groups	Established knowledge about a group one or both participants are involved in	My partner plays both lacrosse and hockey at Dartmouth.
	Behavior	Established knowledge about behavior one or both participants engage in	NAME says she's disappointed that I bought food because I buy A LOT of things pretty often. I might have a compulsive shopping problem.
	Hometown	Established knowledge about the hometown of one or both participants	I used to live in Quebec and a lot of my family still lives there.
	Housing	Established knowledge about the housing situation of one or both participants	We lived on the 4th floor of our dorm building. The comment about it being a hike was reminiscent of the pain of walking up and down four flights of stairs to get to our rooms.
	Health	Established knowledge about mental or physical health information about one or both participants	NAME says "good enough" in response to being asked how she is doing. NAME struggles with depression, and due to the nature of their relationship, I know that this likely means that NAME is not doing that well with her depression.

Appendix D

Post-conversation survey items for Study 1

Instructions

Welcome to the survey!

Remember that you should only advance forward if you will be able to focus for the next 60 minutes. If that does not describe you right now, please exit out of the survey and come back later. Thank you!

In Part 1 of this study you had a 10-minute Zoom conversation with another person. We will refer to this person as your '**conversation partner**'.

Think back to that conversation and answer the following questions as accurately and honestly as you can. Your conversation partner will NOT view your responses.

Conversation Questions

1. How well did this conversation "flow"? *(0=Not at all, 100=Very)*
2. How much did you enjoy the conversation you had with your conversation partner? *(0=Not at all, 100=Very much)*
3. How connected did you feel to your conversation partner? *(0=Not at all, 100=Very much)*
4. Did you know your conversation partner BEFORE your conversation with them? *(no, yes)*
5. Given that you've met your conversation partner before, how would you characterize the nature of your relationship? *(0=acquaintances, 25=friend, 75=close friend, 100=best friend)* [This question was only asked if the answer to Question #4 was ‘yes’]

Please rate your agreement with the following statements:

6. My conversation partner is an attractive person. *(0=Strongly disagree, 100=Strongly agree)*
7. I was physically attracted to my conversation partner. *(0=Strongly disagree, 100=Strongly agree)*

Shared Reality Questions

Please rate your agreement with the following statements about you and your conversation partner. (1=Strongly disagree, 2=Disagree, 3=Somewhat disagree, 4=Neither agree nor disagree, 5=Somewhat agree, 6=Agree, 7=Strongly agree)

8. During our interaction we thought of things at the exact same time.
9. During our interaction we developed a joint perspective.

10. During our interaction we shared the same thoughts and feelings about things.
11. During our interaction our conversation felt very real.
12. During our interaction the way we thought became more similar.
13. During our interaction we often anticipated what the other was about to say.
14. During our interaction we became more certain of the way we perceived things.
15. During our interaction we saw the world in the same way.

Demographics

16. What is your age? *(open response)*
17. What is your gender? *(choose one: Female, Male, Other w/ optional text box, Prefer not to answer)*
18. What is your ethnicity? *(can select multiple: White, Hispanic or Latino, Black or African American, Native American or American Indian, Asian, Other w/ optional text box, Prefer not to answer)*

Insider language task instructions

Thank you for answering all of those questions.

Now, it is time to complete the main task.

You learned a bit about this task in Part 1 of this study when a researcher gave you an overview. We will now review those instructions with you. Please carefully read these instructions to make sure you fully understand the task before you start.

You will watch a recording of the 10-minute Zoom conversation you had with your conversation partner.

As you watch, **we want you to imagine that another Dartmouth student, who you've never met, is listening in on your conversation**. To be clear, this will never happen.

Whenever you reach a moment in your conversation that this outsider might not understand, we would like you to **pause the video**.

For example, if in the video one of you were to say, "I'm in CS1" this would NOT need to be explained because you can assume that a typical Dartmouth student has knowledge of the courses offered.

However, if one of you were to say, "Did you hear that X and Y are dating?" you would want to unpack that. Who is X? Who is Y? What is their relationship to you and to your conversation partner? Why would their relationship be surprising?

Often when we talk to other people, we use words that carry a backstory that has meaning to us and to the person we are talking to. But if someone new joins your conversation, you would need to explain more for them to understand. This is what we are trying to capture. What are the words that **were NOT said** that would **need to be said** for the new person to understand?

Each time you hear something in your conversation that would need to be explained, pause the video and type that explanation. It doesn't matter if it was something that you said or something that your partner said.

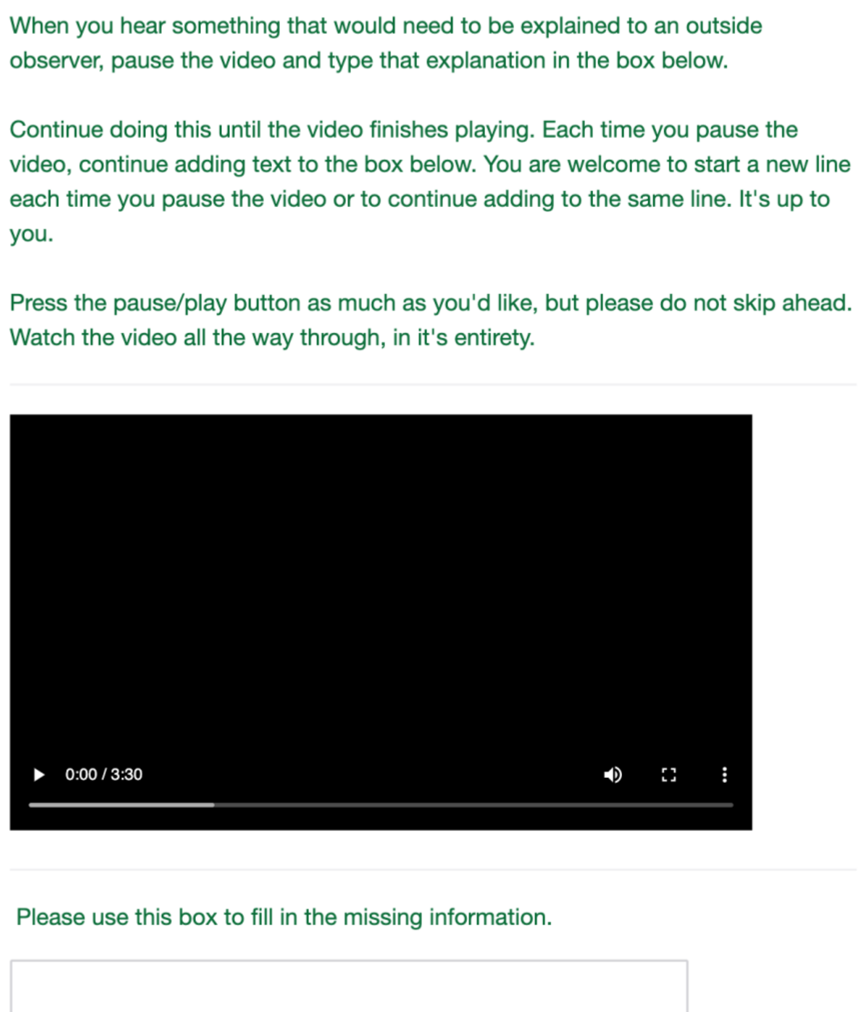
Explain everything that someone would need to know to understand the conversation that you had, including your or your partner's reactions.

Continue doing this until you have finished watching the entire conversation.

If you are in doubt about what needs explaining, just pause the video and go ahead and explain it.

For space constraints, we've divided the recording into two parts. You'll complete this task for the first half of your conversation. Then, you will advance forward in the survey to continue with the second half of your conversation.

Screenshot of what the task looked like



Appendix E

Insider language annotation survey for independent raters in Study 2

Instructions

Welcome! As a refresher, please carefully read through these instructions before beginning the task.

Often when we talk to other people, we use words that carry a backstory that has meaning to us and the person we are talking to. If someone new were to join our conversation, we would need to explain more for them to understand.

This is the experience we are trying to capture in this task.

You will watch video recordings of conversations between two different people. As you watch, imagine you are the "**outsider**" trying to fully understand what the two people are talking about.

The two people involved in the conversation may share "**insider language**." This happens when someone uses language that clearly carries a special meaning shared by both people in the conversation; that special meaning would need to be explained to an outsider.

As you watch these videos, we would like you to notice and identify these instances of insider language.

Your task will be to identify turns that contain insider language. These might be instances where you would ask a clarifying question or where the speakers may have paused their conversation to give you more context.

When you notice an instance of insider language, put a check mark next to those turns on the corresponding transcript in this survey. It's okay if the transcript has some errors. Go by the words that you heard in the conversation video and select the turns in the transcript that best match the moment you found (remember you can use the timestamps to help you).

Different conversations may have different amounts of insider language. Some may have a lot and others may have a little.

Good luck!

A screenshot of what the annotation task looks like. Each turn begins with a timestamp to help raters locate the correct turn from the video. When raters identified turns with insider language, they selected the checkbox next to those turns.

This is conversation **85_261**. Locate the corresponding video file and open it up. Carefully watch the conversation and pause the video each time you come across an instance of insider language. Indicate those moments on the transcript below. Select the turns that best match the moments you've identified (it's okay if the transcripts have errors, always go by what the words you heard in the video and use the timestamps to help). You can pause and re-watch the video as many times as needed.

- [ ] 00:05.360 I woke up 10 minutes ago and then came.
- [ ] 00:07.260 You woke up 10... I mean, I woke up at 9:30, it was fine.
- [x] 00:10.640 Yeah, I woke up at I think, at 9:45. And Brett was still cooped up.
- [x] 00:16.338 Well, he's always passed out, isn't he?
- [ ] 00:17.519 He's always passed out, yeah, and I just drag myself...
- [ ] 00:19.762 He has... Does he have his tough test today? Or is that... Was that yesterday?

References

Antaki, C. (2011). Six Kinds of Applied Conversation Analysis. In C. Antaki (Ed.), *Applied Conversation Analysis: Intervention and Change in Institutional Talk* (pp. 1–14). Palgrave Macmillan UK.

Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to Stop signal response inhibition: role of the subthalamic nucleus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(9), 2424–2433.

Assent, I. (2012). Clustering high dimensional data. *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery*, *2*(4), 340–350.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: quantifying influence on twitter. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 65–74.

Barr, D. J., & Keysar, B. (2006). Chapter 23 - Perspective Taking and the Coordination of Meaning in Language Use. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics (Second Edition)* (pp. 901–938). Academic Press.

Bartels, J., Rodenbach, R., Ciesinski, K., Gramling, R., Fiscella, K., & Epstein, R. (2016). Eloquent silences: A musical and lexical analysis of conversation between oncologists and their patients. *Patient Education and Counseling*, *99*(10), 1584–1594.

Barthel, M., & Sauppe, S. (2019). Speech Planning at Turn Transitions in Dialog Is Associated With Increased Processing Load. *Cognitive Science*, *43*(7), e12768.

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The Timing of Utterance Planning in Task-Oriented Dialogue: Evidence from a Novel List-Completion Paradigm.

Frontiers in Psychology, *7*, 1858.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bearman, P., & Parigi, P. (2004). Cloning headless frogs and other important matters: Conversation topics and network structure. *Social Forces; a Scientific Medium of Social Study and Interpretation*, *83*(2), 535–557.

Bischoping, K. (1993). Gender differences in conversation topics, 1922–1990. *Sex Roles*, *28*(1), 1–18.

Bögels, S., Casillas, M., & Levinson, S. C. (2018). Planning versus comprehension in turn-taking: Fast responders show reduced anticipatory processing of the question. *Neuropsychologia*, *109*, 295–310.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, *5*, 12881.

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, *52*, 46–57.

Boncz, A. (2019). *Communication as Joint Action: The role of cognitive alignment and coupling* [Central European University]. https://cognitivescience.ceu.edu/sites/cognitivescience.ceu.hu/files/attachment/basicpage/3173/ercbonczthesis.pdf

Boothby, E. J., Cooney, G., Sandstrom, G. M., & Clark, M. S. (2018). The Liking Gap in Conversations: Do People Like Us More Than We Think? *Psychological Science*, *29*(11), 1742–1756.

Brambilla, M., Sabet, A. J., Kharmale, K., & Sulistiawati, A. E. (2022). Graph-Based

Conversation Analysis in Social Media. *Big Data and Cognitive Computing*, *6*(4), 113.

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

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

Brooks, Kristensen, & Van Benthem. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R*. https://www.research-collection.ethz.ch/handle/20.500.11850/239870

Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of Child Language*, *2*(1), 1–19.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217.

Butler, E. A., & Randall, A. K. (2013). Emotional Coregulation in Close Relationships. *Emotion Review: Journal of the International Society for Research on Emotion*, *5*(2), 202–210.

Cassell, J., Gill, A., & Tepper, P. (2007). Coordination in Conversation and Rapport. *Proceedings of the Workshop on Embodied Language Processing*, 41–50.

Cer, D., Yang, Y., Kong, S.-Y., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder. In *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/1803.11175

Chen, G., Shin, Y.-W., Taylor, P. A., Glen, D. R., Reynolds, R. C., Israel, R. B., & Cox, R. W. (2016). Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. *NeuroImage*, *142*, 248–259.

Chow, C. P., Mitchell, J. F., & Miller, C. T. (2015). Vocal turn-taking in a non-human primate is learned during ontogeny. *Proceedings. Biological Sciences / The Royal Society*, *282*(1807), 20150069.

Clark, H. H. (1992). *Arenas of Language Use*. University of Chicago Press.

Clark, H. H. (1996a). *Using language*. https://books.google.ca/books?hl=en&lr=&id=b8bLCgAAQBAJ&oi=fnd&pg=PP1&ots=iQn3Sc83PC&sig=9biFA4g0tbq04hhUxzpK2bk_iXM

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

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick (Ed.), *Perspectives on socially shared cognition , (pp* (Vol. 429, pp. 127–149). American Psychological Association, xiii.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*(2), 259–294.

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

Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., Munteanu, C., Wade, V., & Cowan, B. R. (2019). What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Article Paper 475.

Cohen Priva, U. (2017). Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition*, *160*, 27–34.

Cohen Priva, U., & Sanker, C. (2020). Natural Leaders: Some Interlocutors Elicit Greater Convergence Across Conversations and Across Characteristics. *Cognitive Science*, *44*(10), e12897.

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: a meta-analytic review. *Psychological Bulletin*, *116*(3), 457–475.

Cui, M. (2020). Introduction to the K-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, *1*(1), 5–8.

Cuperman, R., & Ickes, W. (2009). Big Five predictors of behavior and perceptions in initial dyadic interactions: Personality similarity helps extraverts and introverts, but hurts "disagreeables." *Journal of Personality and Social Psychology*, *97*(4), 667–684.

Curhan, J. R., & Pentland, A. (2007). Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *The Journal of Applied Psychology*, *92*(3), 802–811.

daSilva, A. W., Huckins, J. F., Wang, W., Wang, R., Campbell, A. T., & Meyer, M. L. (2021). Daily perceived stress predicts less next day social interaction: Evidence from a naturalistic mobile sensing study. *Emotion* , *21*(8), 1760–1770.

Degen, J. (2023). The Rational Speech Act Framework. *Annual Review of Linguistics*, *9*(1), 519–540.

de Vos, C., Torreira, F., & Levinson, S. C. (2015). Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries. *Frontiers in Psychology*, *6*, 268.

Dindia, K. (1986). Antecedents and consequences of awkward silence A replication using revised lag sequential analysis. *Human Communication Research*, *13*(1), 108–125.

Drijvers, L., & Holler, J. (2022). Face-to-face spatial orientation fine-tunes the brain for neurocognitive processing in conversation. *iScience*, *25*(11), 105413.

Dunbar, R. (1998). *Grooming, gossip, and the evolution of language*. https://books.google.ca/books?hl=en&lr=&id=nN5DFNT-

6ToC&oi=fnd&pg=PA11&ots=71H8_XcRRb&sig=S4M6D-CsmBX133DkInZcUQsecdA

Dunbar, R. I. M. (2018). The Anatomy of Friendship. *Trends in Cognitive Sciences*, *22*(1), 32–51.

Dunbar, R. I., Marriott, A., & Duncan, N. D. (1997). Human conversational behavior. *Human Nature* , *8*(3), 231–246.

Eagle, N., & (Sandy) Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, *10*(4), 255–268.

Echterhoff, G., Higgins, E. T., & Levine, J. M. (2009). Shared Reality: Experiencing Commonality With Others' Inner States About the World. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *4*(5), 496–521.

Emler, N. (1990). A Social Psychology of Reputation. *European Review of Social Psychology*, *1*(1), 171–193.

Epley, N., Kardas, M., Zhao, X., Atir, S., & Schroeder, J. (2022). Undersociality: miscalibrated social cognition can inhibit social connection. *Trends in Cognitive Sciences*, *26*(5), 406–418.

Epley, N., & Schroeder, J. (2014). Mistakenly seeking solitude. *Journal of Experimental Psychology. General*, *143*(5), 1980–1999.

Fan, J., Han, F., & Liu, H. (2014). Challenges of Big Data Analysis. *National Science Review*, *1*(2), 293–314.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: quantifying the benefits of linguistic coordination. *Psychological Science*, *23*(8), 931–939.

Fusaroli, R., & Tylén, K. (2016). Investigating Conversational Dynamics: Interactive Alignment, Interpersonal Synergy, and Collective Task Performance. *Cognitive Science*, *40*(1), 145–171.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, *29*(5), 737–767.

Galantucci, B., & Garrod, S. (2011). Experimental semiotics: a review. *Frontiers in Human Neuroscience*, *5*, 11.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*(3), 181–215.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11.

Gates, V., Callaway, F., Ho, M. K., & Griffiths, T. L. (2021). A rational model of people’s inferences about others' preferences based on response times. *Cognition*, *217*, 104885.

Geissmann, T., & Orgeldinger, M. (2000). The relationship between duet songs and pair bonds in siamangs, Hylobates syndactylus. *Animal Behaviour*, *60*(6), 805–809.

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*, 34.

Godfrey, J. J., & Holliman, E. (1997). Switchboard-1 Release 2. *Linguistic Data Consortium, Philadelphia*.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Goodwin, C., & Heritage, J. (1990). Conversation Analysis. *Annual Review of Anthropology*, *19*(1), 283–307.

Gramling, R., Stanek, S., Ladwig, S., Gajary-Coots, E., Cimino, J., Anderson, W., Norton, S. A., AAHPM Research Committee Writing Group, Aslakson, R. A., Ast, K., Elk, R., Garner, K. K., Gramling, R., Grudzen, C., Kamal, A. H., Lamba, S., LeBlanc, T. W., Rhodes, R. L., Roeland, E., … Unroe, K. T. (2016). Feeling Heard and Understood: A Patient-Reported Quality Measure for the Inpatient Palliative Care Setting. *Journal of Pain and Symptom Management*, *51*(2), 150–154.

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, *6*, 1167.

Grezes, F., Richards, J., & Rosenberg, A. (2013). Let me finish: automatic conflict detection using speaker overlap. *Interspeech*, 200–204.

Grice, H. P. (1957). Meaning”: Philosophical Review. *Volume*, *66*, N03.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.

Guilford, T., & Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behaviour*, *42*(1), 1–14.

Haas, A., & Sherman, M. A. (1982). Conversational Topic as a Function of Role and Gender. *Psychological Reports*, *51*(2), 453–454.

Hadley, L. V., Naylor, G., & Hamilton, A. F. de C. (2022). A review of theories and methods in the science of face-to-face social interaction. *Nature Reviews Psychology*, *1*(1), 42–54.

Hall, J. A., Holmstrom, A. J., Pennington, N., Perrault, E. K., & Totzkay, D. (2023). Quality

Conversation Can Increase Daily Well-Being. *Communication Research*, 00936502221139363.

Harari, G. M., Müller, S. R., & Aung, M. S. H. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in*. https://www.sciencedirect.com/science/article/pii/S2352154617300505

Harari, G. M., Müller, S. R., Stachl, C., & Wang, R. (2020). Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of Personality*. https://psycnet.apa.org/record/2019-27734-001

Hardin, C. D., & Higgins, E. T. (1996). Shared reality: How social verification makes the subjective objective. In R. M. Sorrentino (Ed.), *Handbook of motivation and cognition, Vol* (Vol. 3, pp. 28–84). The Guilford Press, xxvi.

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

Hawkins, R. D., Liu, I., Goldberg, A. E., & Griffiths, T. L. (2021). Respect the code: Speakers expect novel conventions to generalize within but not across social group boundaries. *Proceedings of the Annual Conference of the Cognitive Science Society*. https://par.nsf.gov/biblio/10285683

Hawkins, R. X. D., Frank, M. C., & Goodman, N. D. (2017). *Convention-formation in iterated reference games*. cogsci.mindmodeling.org. https://cogsci.mindmodeling.org/2017/papers/0098/paper0098.pdf

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics, 38*(4), 555–568.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2-3), 61–83; discussion 83–135.

Hepburn, A., & Bolden, G. B. (2012). The conversation analytic approach to transcription. *The Handbook of Conversation Analysis*, 57–76.

Hill, C. E., Thompson, B. J., & Ladany, N. (2003). Therapist use of silence in therapy: a survey. *Journal of Clinical Psychology*, *59*(4), 513–524.

Hirschi, Q., Wilson, T. D., & Gilbert, D. T. (2022). Speak Up! Mistaken Beliefs About How Much to Talk in Conversations. *Personality & Social Psychology Bulletin*, 1461672221104927.

Hoey, E. M. (2015). Lapses: How People Arrive at, and Deal With, Discontinuities in Talk. *Research on Language and Social Interaction*, *48*(4), 430–453.

Holler, J., Bavelas, J., Woods, J., Geiger, M., & Simons, L. (2022). Given-New Effects on the Duration of Gestures and of Words in Face-to-Face Dialogue. *Discourse Processes*, *59*(8), 619–645.

Holler, J., & Levinson, S. C. (2019). Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences*, *23*(8), 639–652.

Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes*. https://www.tandfonline.com/doi/abs/10.1080/01690960802095545

Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: a meta-analytic review. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *10*(2), 227–237.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117.

Huang, K., Yeomans, M., Brooks, A. W., Minson, J., & Gino, F. (2017). It doesn't hurt to ask: Question-asking increases liking. *Journal of Personality and Social Psychology*, *113*(3), 430–452.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology*, *2*, 255.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101–144.

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. D. (2001). Rhythms of dialogue in infancy: coordinated timing in development. *Monographs of the Society for Research in Child Development*, *66*(2), i – viii, 1–132.

Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a "standard maximum" silence of approximately one second in conversation. *Conversation: An Interdisciplinary Perspective*, 166–196.

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(45), 16064–16069.

Jolly, E., & Chang, L. J. (2019). The Flatland fallacy: Moving beyond low-dimensional thinking. *Topics in Cognitive Science*, *11*(2), 433–454.

Jucker, A. H., & Smith, S. W. (2022). Explicit and implicit ways of enhancing common ground in conversations. *Pragmatics*, *6*(1), 1–18.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: the day reconstruction method. *Science*, *306*(5702), 1776–1780.

Kang, O., & Wheatley, T. (2017). Pupil dilation patterns spontaneously synchronize across individuals during shared attention. *Journal of Experimental Psychology. General*, *146*(4), 569–576.

Kardas, M., Kumar, A., & Epley, N. (2022). Overly shallow?: Miscalibrated expectations create a barrier to deeper conversation. *Journal of Personality and Social Psychology*, *122*(3), 367–398.

Kendrick, K. H., & Torreira, F. (2015). The Timing and Construction of Preference: A Quantitative Study. *Discourse Processes*, *52*(4), 255–289.

Kenny, D. A., & La Voie, L. (1984). The Social Relations Model. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 18, pp. 141–182). Academic Press.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The Egocentric Basis of Language Use: Insights From a Processing Approach. *Current Directions in Psychological Science*, *7*(2), 46–49.

Koudenburg, N., Postmes, T., & Gordijn, E. H. (2011). Disrupting the flow: How brief silences in group conversations affect social needs. *Journal of Experimental Social Psychology*. https://www.sciencedirect.com/science/article/pii/S002210311000274X

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: a preliminary study. *Psychonomic Science*, *1*(1), 113–114.

Kull, K. (2020). Codes: Necessary, but not Sufficient for Meaning-Making. *Constructivist Foundations*, *15*(2), 137–139.

Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V., & Stefanovska, A. (2018). Surrogate data for hypothesis testing of physical systems. *Physics Reports*, *748*, 1–60.

Landis, M. H., & Burtt, H. E. (1924). A study of conversations. *Journal of Comparative Psychology*, *4*(1), 81.

Levinson, S. C. (2006). *On the human "interactional engine."* https://repository.ubn.ru.nl/bitstream/handle/2066/41724/269151.pdf

Levinson, S. C. (2016). Turn-taking in human communication--origins and implications for language processing. *Trends in Cognitive Sciences*, *20*(1), 6–14.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*, 731.

Levitt, S. D., & List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *The Journal of Economic Perspectives: A Journal of the American Economic Association*, *21*(2), 153–174.

Lewin, K. (1939). Field Theory and Experiment in Social Psychology: Concepts and Methods. *The American Journal of Sociology*, *44*(6), 868–896.

Magyari, L., Bastiaansen, M. C. M., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience*, *26*(11), 2530–2539.

Mastroianni, A. M., Gilbert, D. T., Cooney, G., & Wilson, T. D. (2021). Do conversations end when people want them to? *Proceedings of the National Academy of Sciences of the United States of America*, *118*(10). https://doi.org/10.1073/pnas.2011809118

McFarland, D. A., Jurafsky, D., & Rawlings, C. (2013). Making the Connection: Social Bonding in Courtship Situations. *The American Journal of Sociology*, *118*(6), 1596–1649.

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and

Projection for Dimension Reduction. In *arXiv [stat.ML]*. arXiv. http://arxiv.org/abs/1802.03426

McLaughlin, M. L., & Cody, M. J. (1982). Awkward silences: Behavioral antecedents and consequences of the conversational lapse. *Human Communication Research*, *8*(4), 299–316.

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): a device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *33*(4), 517–523.

Mehl, M. R., Vazire, S., Holleran, S. E., & Clark, C. S. (2010). Eavesdropping on happiness: well-being is related to having less small talk and more substantive conversations. *Psychological Science*, *21*(4), 539–541.

Méndez-Cárdenas, M. G., & Zimmermann, E. (2009). Duetting--a mechanism to strengthen pair bonds in a dispersed pair-living primate (Lepilemur edwardsi)? *American Journal of Physical Anthropology*, *139*(4), 523–532.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213.

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*. https://onlinelibrary.wiley.com/doi/abs/10.1111/josl.1_12177

Moore, H. T. (1922). Further data concerning sex differences. *Journal of Abnormal and Social Psychology*, *17*(2), 210–214.

Mulwa, K. W., & Kucker, S. C. (2022). Coding social interactions in naturalistic settings: The taxonomy of dyadic conversation. *Behavior Research Methods*, 1–15.

Newman, H. M. (1982). The sounds of silence in communicative encounters. *Communication Quarterly*, *30*(2), 142–149.

Ohshima, N., Kimijima, K., Yamato, J., & Mukawa, N. (2015). A conversational robot with vocal and bodily fillers for recovering from awkward silence at turn-takings. *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 325–330.

Onnela, J.-P., Waber, B. N., Pentland, A., Schnorf, S., & Lazer, D. (2014). Using sociometers to quantify social interaction patterns. *Scientific Reports*, *4*(1), 5604.

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2017). Spontaneous neural encoding of social network position. *Nature Human Behaviour*, *1*(5), 1–7.

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

Pierce, C. S. (1931). *Collected Papers, eds. C. Hartshorn, P. Weiss*. Cambridge: Harvard University Press.

Planalp, S., & Benson, A. (1992). Friends' and Acquaintances' Conversations I: Perceived Differences. *Journal of Social and Personal Relationships*, *9*(4), 483–506.

Pomerantz, A. (1984). *Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shaped*. https://scholarsarchive.library.albany.edu/cas_communication_scholar/3/

Pouw, W., & Holler, J. (2022). Timing in conversation is dynamically adjusted turn by turn in dyadic telephone conversations. *Cognition*, *222*, 105015.

Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and*

Clinical Psychology, *79*(3), 284–295.

R Core Team. (2018). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*.

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

Riest, C., Jorschick, A. B., & de Ruiter, J. P. (2015). Anticipation in turn-taking: mechanisms and information sources. *Frontiers in Psychology*, *6*, 89.

Roberts, F., & Francis, A. L. (2013). Identifying a temporal threshold of tolerance for silent gaps after requests. *The Journal of the Acoustical Society of America*, *133*(6), EL471–EL477.

Roberts, F., Francis, A. L., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication*, *48*(9), 1079–1093.

Roberts, F., Margutti, P., & Takano, S. (2011). Judgments Concerning the Valence of Inter-Turn Silence Across Speakers of American English, Italian, and Japanese. *Discourse Processes*, *48*(5), 331–354.

Rossignac-Milon, M., Bolger, N., Zee, K. S., Boothby, E. J., & Higgins, E. T. (2020). Merged minds: Generalized shared reality in dyadic relationships. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000266

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

Sanchez, K. L., Kalkstein, D. A., & Walton, G. M. (2022). A threatening opportunity: The prospect of conversations about race-related experiences between Black and White friends.

Journal of Personality and Social Psychology, *122*(5), 853–872.

Sandstrom, G. M., & Boothby, E. J. (2021). Why do people avoid talking to strangers? A mini meta-analysis of predicted fears and actual experiences talking to a stranger. *Self and Identity: The Journal of the International Society for Self and Identity*, *20*(1), 47–71.

Schegloff, E. A. (2006). Interaction: The infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted. *Roots of Human Sociality: Culture, Cognition and Interaction*, 70–96.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382.

Schegloff, E. A., & Sacks, H. (1973). *Opening up Closings*. *8*(4), 289–327.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience1. *The Behavioral and Brain Sciences*, *36*(4), 393–414.

Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. S. (2009). Signalling signalhood and the emergence of communication. *Cognition*, *113*(2), 226–233.

Serban, I. V., Lowe, R., Henderson, P., Charlin, L., & Pineau, J. (2015). A Survey of Available Corpora for Building Data-Driven Dialogue Systems. In *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/1512.05742

Seyfarth, R. M., Cheney, D. L., & Bergman, T. J. (2005). Primate social cognition and the origins of language. *Trends in Cognitive Sciences*, *9*(6), 264–266.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Shriberg, E., & Stolcke, A. (1996). Word predictability after hesitations: a corpus-based study.

Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96, *3*, 1868–1871 vol.3.

Smith, V. L., & Clark, H. H. (1993). On the Course of Answering Questions. *Journal of Memory and Language*, *32*(1), 25–38.

Snowdon, C. T., & Cleveland, J. (1984). "Conversations" among pygmy marmosets. *American Journal of Primatology*, *7*(1), 15–20.

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*(32), 14425–14430.

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 of the United States of America*, *106*(26), 10587–10592.

Stivers, T., & Robinson, J. D. (2006). A Preference for Progressivity in Interaction. *Language In Society*, *35*(3), 367–392.

Stolk, A., Bašnáková, J., & Toni, I. (2020). Joint epistemic engineering: The neglected process in human communication. In *PsyArXiv*. https://doi.org/10.31234/osf.io/rwfe6

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

Stolk, A., Verhagen, L., & Toni, I. (2016). Conceptual Alignment: How Brains Achieve Mutual Understanding. *Trends in Cognitive Sciences*, *20*(3), 180–191.

Takahashi, D. Y., Narayanan, D. Z., & Ghazanfar, A. A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Current Biology: CB*, *23*(21), 2162–2168.

Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(4). https://doi.org/10.1073/pnas.2116915119

Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2023). Long gaps between turns are awkward for strangers but not for friends. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *378*(1875), 20210471.

Tickle-Degnen, L., & Rosenthal, R. (1990). The Nature of Rapport and Its Nonverbal Correlates. *Psychological Inquiry*, *1*(4), 285–293.

Trimboli, C., & Walker, M. B. (1984). Switching pauses in cooperative and competitive conversations. *Journal of Experimental Social Psychology*, *20*(4), 297–311.

Tsoi, L., Burns, S. M., Falk, E. B., & Tamir, D. I. (2022). The promises and pitfalls of functional magnetic resonance imaging hyperscanning for social interaction research. *Social and Personality Psychology Compass*, *16*(10), e12707.

Twain, M. (1876, December 22). *Speech on the Weather*. New England Society's seventy-first annual dinner, New York City.

Van Bavel, J. J., Rathje, S., Harris, E., Robertson, C., & Sternisko, A. (2021). How social media shapes polarization. *Trends in Cognitive Sciences*, *25*(11), 913–916.

Van de Calseyde, P. P. F. M., Keren, G., & Zeelenberg, M. (2014). Decision time as information in judgment and choice. *Organizational Behavior and Human Decision Processes*, *125*(2), 113–122.

Wadge, H., Brewer, R., Bird, G., Toni, I., & Stolk, A. (2019). Communicative misalignment in Autism Spectrum Disorder. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *115*, 15–26.

Wang, R., Chen, F., Chen, Z., Li, T., & Harari, G. (2014). StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. *Proceedings of the*. https://dl.acm.org/doi/abs/10.1145/2632048.2632054

Wheatley, T., Boncz, A., Toni, I., & Stolk, A. (2019). Beyond the Isolated Brain: The Promise and Challenge of Interacting Minds. *Neuron*, *103*(2), 186–188.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*(2), 183–194.

Wohltjen, S., & Wheatley, T. (2021). Eye contact marks the rise and fall of shared attention in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(37). https://doi.org/10.1073/pnas.2106645118

Wood, A., Templeton, E., Morrel, J., Schubert, F., & Wheatley, T. (2022). Tendency to laugh is a stable trait: findings from a round-robin conversation study. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *377*(1863), 20210187.

Yeomans, M., Schweitzer, M. E., & Brooks, A. W. (2022). The Conversational Circumplex: Identifying, prioritizing, and pursuing informational and relational motives in conversation. *Current Opinion in Psychology*, *44*, 293–302.

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., & Hasson, U. (2017). How We Transmit Memories to Other Brains: Constructing Shared Neural Representations Via Communication. *Cerebral Cortex* , *27*(10), 4988–5000.

Ziano, I., & Wang, D. (2021). Slow lies: Response delays promote perceptions of insincerity. *Journal of Personality and Social Psychology*, *120*(6), 1457–1479.

Zou, J., & Schiebinger, L. (2018, July 18). *AI can be sexist and racist — it's time to make it fair*. Nature Publishing Group UK; nature.com. https://doi.org/10.1038/d41586-018-05707-8