CONVERSATIONAL CUES, INTERPERSONAL JUDGEMENTS, AND THE DECISION TO SPEAK

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by Department of Business Studies

have examined a dissertation entitled

Conversational Cues, Interpersonal Judgements, and the Decision to Speak

Presented by Nicole Abi-Esber

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature _____

Alison Wood Brooks, Chair

Signature _____

Francesca Gino

Signature _____

Lindred Greer

Signature _____

Ethan Burris

Date: April 13, 2023

CONVERSATIONAL CUES, INTERPERSONAL JUDGEMENTS, AND THE DECISION TO SPEAK

How perceptions of others' nonverbal and paralinguistic cues shape decisions to speak in group conversations.

Nicole Abi-Esber

Harvard Business School

Doctoral Dissertation in Organizational Behavior

April 2023

Dissertation Advisor: Alison Wood Brooks Nicole Abi-Esber

Conversational Cues, Interpersonal Judgments, and the Decision to Speak

ABSTRACT

What matters more: what a person says, or how they say it? Words (verbal content) are the carriers of meaning and define conversations. But how someone says something, including facial expressions and body language (nonverbal content) and tone of voice (paralinguistic content) shape how this meaning is interpreted. My work suggests that "What matters more?" may not be the right question: it is the *interplay* of what people say and how they say it that defines the social world. This dissertation aims to explore the critical role of nonverbal and paralinguistic content—alongside verbal content—in shaping how we create interpersonal perceptions about others. This work begins to uncover how perceptions of others' nonverbal and paralinguistic cues shape decisions to speak up, or not, in dyadic conversations, group meetings, and classrooms.

TABLE OF CONTENTS

DEDICATION

I come from a family of scholars and teachers: all four of my grandparents were teachers in my village, Kfarhata, in Northern Lebanon. My PhD dissertation is dedicated to them, and especially to Albert Ayoub, my maternal grandfather. My grandfather was a scholar, a traveler, an adventurer, and most of all: a teacher. He was a teacher and a principal of Kfarhata Public High School for over 40 years. He taught both of my parents and countless others. The school is now named after him. Thank you, Jiddou Albert, for being my inspiration for how to live a life of dignity.

ACKNOWLEDGEMENTS

I am forever grateful for having the most amazing parents in the world: Sana Ayoub and Fares Esber. They left their homes, their parents, and their country in their early twenties to move across the world to raise me and my brother in a country where we would have opportunities to grow and shine. They sacrificed quite literally everything to give us a better life. Thank you for always being positive, supportive, and happy, and endowing me with good looks. My third dissertation chapter is dedicated to you.

To my husband Mateus: you joined mid-way through my PhD journey, but I feel like you've been with me the whole time. Thank you for pushing me to make this dissertation great for myself. Thank you for eschewing traditional gender roles and conventions. You are the kindest, most supportive, warmest, and funniest, and the best Kindle and cuddle partner in the world. I'm so happy I bamboozled you into marrying me. Your video-making skills made my second chapter possible, and my second chapter is dedicated to you.

To my brother Sam, thank you for keeping it real, and for keeping me real. Sorry for always embarrassing you. And sorry, but I will never stop. To my in-laws, the Guzzo family, thank you all for welcoming me with open arms. Love, the eternal paseadora.

To my doctoral advisor and dissertation committee chair, Alison Wood Brooks: you inspire me every day to do rigorous, interesting, difficult work, and to be kind while doing it. Thank you for believing in me and giving me the room to explore my own ideas, while simultaneously giving me early experience with publishing and the review process. Five years is a long time to work closely with one person, and I know that I am one of the luckiest PhD students ever to have had such an amazing, supportive advisor. You never let me feel how busy you must have been—through pregnancies, three kids, teaching a new class during the pandemic, and writing a book.

You were always there, and always generous with your time. Thank you for your support, care, and generosity. I hope to be a scholar, professor, speaker, and mother like you one day. My first chapter is dedicated to you: thank you for supporting my ideas and helping me bring them to fruition.

I am extremely grateful to my amazing dissertation committee members, Lindy Greer, Francesca Gino and Ethan Burris. To Lindy Greer, you are the reason I am here! Once upon a time, in 2016, you responded to a cold email from me and agreed to meet with me. You saw something in me that I didn't see in myself, helped hire me as a research assistant, and continue to inspire me and support me. You helped me refine my research interests, and then introduced me to Alison to study voice and conversations, and in doing so set me up for success in my PhD. Thank you for always pushing me to make sure my work is relevant, interesting, and theoretically rich. Thank you also for opening doors and reaching out to your friends and colleagues on my behalf, both when I was the PhD market and when I was the faculty job market. You're the first person I call when something goes wrong, and you are always there for me, as a colleague and a friend. I am incredibly lucky to have you in my corner.

Francesca Gino, you are a gem as a mentor and friend. I have cherished every conversation with you, and your boundless enthusiasm and intellectual curiosity have been a constant source of inspiration and motivation to me. You have an amazing ability to make people feel like they are the only other person in the room. I often leave our conversations thinking— I love my job! How cool is it that I get to talk to interesting people and explore interesting things? I will miss our long walks and our funny chats. I feel so lucky to have you as a friend, co-author, and committee member; thank you for everything.

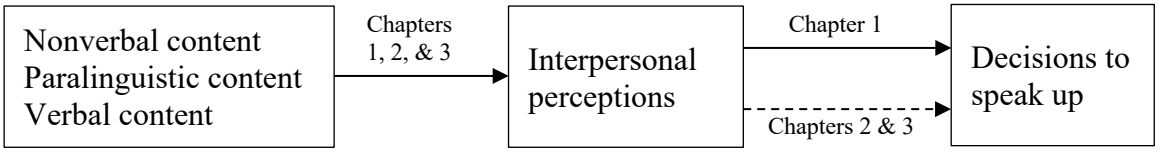
Ethan Burris, thank you for believing in me and joining me as a mentor on this journey. You took a cold email from me in the depths of the COVID pandemic, when I was struggling with my PhD. You listened to my pitch, agreed to join my project, and immediately helped to elevate

the paper, making it more organizationally-relevant and theoretically rich. You believed in what I am trying to do with my research: move from gestalt leadership styles to concrete leader behaviors (and also taught me the word “gestalt”). This project became my job market paper and the first chapter of my dissertation, and I couldn’t have done it without you. I hope to be co-authors and collaborators for a long time.

To the best cohort mates ever, Grace Cormier and Aurora Turek, it’s been a long ride but I’m glad we did it together. Thank you to my co-authors, collaborators, and friends in the academic community: NERD lab, OB lab, Ariella Kristal, Taqua Elleithy, Hanne Collins, Jennifer Abel, Juliana Schroeder, Adam Mastroianni, Mike Yeomans, Katie Boland, Charles Chu, Annebel De Hoogh, Preeti Vani, Hayley Blunden, Dafna Goor, Grant Donnelly, Beth Anne Hagelson, Chistina Bradley, Emily Prinsloo, Ting Zhang, Julian Zlatev, and Jon Jachimowicz. Thank you to the labs that made my experimental research possible: the HBS Be-Lab, Karin Parodi, Chris Nave, and Piper Oren, and the Harvard Decision Sciences Lab and Alki Iliopoulou. To the Doctoral Office: Marais Young, Jen Mucciarone, Maryna MacDonald, Darlene Le, and Finley Foster—I know I’m not the easiest student to work with, and I appreciate your patience, flexibility, and support over the years. Thank you for making everything possible.

INTRODUCTION

What matters more: what a person says, or how they say it? Words (verbal content) are the carriers of meaning and define conversations. But how someone says something, including facial expressions and body language (nonverbal content) and tone of voice (paralinguistic content) shape how this meaning is interpreted. My work suggests that "What matters more?" may not be the right question: it is the *interplay* of what people say and how they say it that defines the social world. This dissertation aims to explore the critical role of nonverbal and paralinguistic content—alongside verbal content—in shaping how we create interpersonal perceptions about others. This work begins to uncover how perceptions of others' nonverbal and paralinguistic cues shape decisions to speak up, or not, in dyadic conversations, group meetings, and classrooms.



Research has only scratched the surface of the interplay between verbal, nonverbal, and paralinguistic cues, usually by comparing cues present or absent in different communication modalities (i.e., examining conversations on the phone versus via text message). In my work, I examine the interplay of verbal, nonverbal, and paralinguistic information by isolating different types of cues at the turn-by-turn level of conversational analysis. I demonstrate how nonverbal cues can increase psychological safety and employee voice (Chapter 1), how nonverbal and paralinguistic content leads to more accurate judgments of others' interpersonal skills (Chapter 2), and how paralinguistic cues can mislead, leading us to evaluate people with foreign accents more negatively than native English speakers, although the verbal content of non-native speakers may

be of higher quality (Chapter 3). I also demonstrate how inferences gleaned from nonverbal and paralinguistic cues shape decisions to speak in group conversations, classrooms, and meetings.

Verbal content defines conversations and meetings. You can have a meeting over the phone (without nonverbal cues), or a conversation via text message (without paralanguage or nonverbals), but can't have either without verbal content. As such, we tend to focus on verbal content. In addition to communicating meaning and information, words also motivate, signal trust, communicate warmth, communicate competence, and have important effects on recipients' feelings (i.e., Aliyev, Karakus, & Ulus, 2013; Imbir, 2017; Stein 1975; Stam, Van Knippenberg, Wisse, & Nederveen, 2018; Swann & Pittman, 1977). However, nonverbal and paralinguistic cues, when available, offer tremendous insight into a person's personality and emotional state (i.e., Ambady et al., 2002; McAleer, Todorov, & Belin, 2014; Hughes, Mogilski, & Harrison, 2014; Scherer, 2003; LaPlante & Ambady, 2003). Seeing somebody's facial expressions, body language, and, as well as hearing their pitch, cadence, volume, and intonation, can help discern a person's true motivations behind language, such as whether a comment is genuine, sarcastic, friendly, or authoritative (Jeong et al., 2019). Nonverbal and paralinguistic cues convey a wealth of information, allowing people to make inferences about someone's personality, emotional state, race, gender, and physical size (McAleer et al., 2014; Belin et al., 2011; Scherer, 2003). Moreover, unlike verbal content, nonverbal and paralinguistic cues are universally recognized—regardless of language and culture (Ekman, 2016). Indeed, nonverbal and paralinguistic behaviors may have an outsized influence on interpersonal dynamics because they may be less easily faked compared to some verbal behaviors (Gottman & Porterfield, 1981; Templeton, Chang, Reynolds, LeBeaumont, & Wheatley, 2022). Contributing to this hypothesis, nonverbal and paralinguistic cues often operate beyond conscious awareness or control (Jones, Feinberg, DeBruine, Little, & Vukovic,

2010; Scherer, 2003; LaPlante & Ambady, 2003), although people can consciously manipulate their voices to seem more intelligent, confident, intelligent, or dominant (Hughes, Mogilski, & Harrison, 2014).

Paralinguistic Content

Having access to paralinguistic content impacts our interpersonal interactions. Several studies have looked at the effect of receiving the same information with someone via text (verbal content only) versus listening to them say the same thing (verbal content and paralinguistic content). These studies have found that paralanguage humanizes the interaction partner, making them seem more mentally capable and possessing more human traits (Schroeder, Kardas, & Epley, 2017), and making them seem more emotional (Berger, Rocklage, & Packard 2022), which in turn, leads to them being more persuasive. Adding paralinguistic cues such as volume (speaking louder) makes people seem more confident, and thus more persuasive, compared to just reading the text of what they say (Van Zant & Berger, 2020). Relatedly, people are more likely to follow recommendations they hear, compared to the same recommendations they read (Mariadassou, Bechler, & Levav, 2022). And, although people largely prefer interacting with a stranger or a friend via writing because they think it will be less awkward (Schroeder, *working*; Kumar & Epley, 2021), several studies find that speaking is a more humanizing conflict-reducing form of discourse, and leads to greater understanding and connection (Schroeder, *working*; Kumar & Epley, 2021).

Paralinguistic information also increases accuracy when it comes to making judgments about other people, their motives, and their relationships. People are more effective at judging when others are being dishonest if they have access to paralinguistic information; sounding uncertain, sounding insecure, and having longer latencies before responding to the self-prediction

question are all signs of dishonesty that others can accurately pick up on (Rogers, Ten Brinke, & Carney, 2016). Callers to potential voters can accurately determine whether someone is actually going to vote—regardless of what they say, verbally—based on the tone of their voice (Rogers, Ten Brinke, & Carney, 2016). Indeed, humans can detect uncertainty in others' voices within milliseconds: it is generally associated with a quiet voice and a rise in pitch at the end of a statement (Jiang & Pell, 2015). Having access to auditory cues leads to an increased ability to identify what kind of relationship two other people have (Dunbar, Robledo, Tamarit, Cross, & Smith, 2022).

Nonverbal Content

Having access to nonverbal content is similarly impactful, and can help increase predictive accuracy when making interpersonal judgments. Having access to someone's nonverbal information leads to more accurate evaluations of if people are being dishonest or honest (Clarke, Krom, & King, 2015), if they are feeling sad or angry (Sternglanz & DePaulo, 2004), whether they are truly remorseful, or faking it (Ten Brinke, MacDonald, Porter, & O'Connor, 2012), and whether they are attracted to another person (Abbey & Melby, 1986). It is easier and more fluent to comprehend a conversation partner when their verbal and nonverbals align, and when they do not align, or there is nonverbal "leakage", nonverbal cues are trusted more (Ten Brinke & Weisbuch, 2020). Some people are particularly good at preventing leakage; individuals with psychopathic traits are better able to suppress emotional expressions and exhibit less leakage of inconsistent emotions when they are lying (Porter, Ten Brinke, Baker, & Wallace, 2011). Nonverbal behaviors are also contagious we can "catch" racial bias from others by merely observing subtle nonverbal cues, even when the content is the same (Willard, Issac, & Carney, 2015). We are also less worried

about monitoring our nonverbal behaviors; we'd rather show our overconfidence via nonverbal cues, compared to verbal statements; nonverbal overconfidence reaps reputational benefits because of its plausible deniability (Tenney, Meikle, Hunsaker, Moore, & Anderson, 2019). Nonverbal cues can even help us judge musical performances; both naïve observers and musical experts were more accurate in predicting how well an orchestra performed if they had access to a video recording of the performance (without sound) compared to an audio clip of the performance (Tsay 2013; 2014). Nonverbal information can also bias information in an undesirable way; musical auditions that have a screen between the auditioner and the judge, preventing the judge from seeing the auditioner, lead to more women being hired and less sex-based discrimination (Goldin & Rouse, 2000). In an fMRI study, people reacted similarly to Diet Coke and Diet Pepsi when they were presented in cups, without any logos or visual cues, but having visual access to the brand logos led to dramatically different brain activation when drinking Diet Coke versus Diet Pepsi (the direction depended on the subject's personal preference; McClure, Tomlin, Cypert, Montague, & Montague, 2004).

Nonverbal cues are so important in communication, that as verbal communication has become more prominent (i.e., emailing and texting), people have begun creating stand-ins for nonverbal cues in written communication. For example, emojis play a significant role in clarifying the intention behind a message, which partially compensates for the absence of nonverbal cues in written communication. For example, winks are found to help people better interpret sarcasm, and ellipses (…) tend to indicate criticism (Thompson & Filik, 2016). Indeed, many elements of 'netspeak' written communication can be understood as efforts to replicate the role played by nonverbals in communication, by conveying context cues such as emotion and the intent behind saying something (Pavalanathan & Eisenstein, 2016). Studies have found that restoring nonverbal

cues on dating apps, by introducing avatars, can help make interactions more pleasant and more likely to lead to real-life relationships (Kotlyar & Ariely, 2013). Nonverbal and paralinguistic cues are indispensable elements of communication, and impact the way we feel and the way we evaluate others.

Summary of Chapters

In my first dissertation chapter, co-authored with Alison Wood Brooks and Ethan Burris, entitled *Feeling Seen: Leader Eye Gaze Promotes Psychological Safety, Participation, and Voice*, I show how eye gaze, a non-verbal cue, affects interpersonal perceptions of psychological safety and ostracism, which affects decisions to speak up. In this paper, we address an emerging argument that existing antecedents to psychological safety, including leadership styles such as transformational leadership and humble leadership, are not specific enough. The multidimensional nature of these leadership constructs obscures mechanistic precision: it is unclear which components lead to safety (Ashford & Sitkin, 2019; Van Knippenberg & Sitkin, 2013; Carton, 2022). And due to survey-based measurements, these constructs also risk having conceptual overlap with safety (Fischer, 2018; Van Knippenberg & Sitkin, 2013). According to a recent review, "we know that leadership matters, but we know less about what leaders must say and do to create psychological safety in concrete terms" (Edmondson & Bransby, 2023). My work increases mechanistic precision by showing that a specific leader's behavior, in this case, leader eye gaze, drives psychological safety and subsequent participation and voice in teams. This is because eye gaze signals attention, respect, dignity, and inclusion (irrespective of what is said). In doing so, I shift scholarly focus from gestalt judgments of leadership styles associated with psychological safety to moment-to-moment behaviors that cause employees to feel safe.

We predict and find that receiving eye gaze causes employees to feel less ostracized and more psychologically safe, which in turn increases participation and voice. We triangulate these effects using mixed methods across three studies. First, using time-lagged analyses within natural conversations in the field, we find that eye gaze received in the first half of a conversation predicts how much employees tend to speak in the second half of the conversation (but not vice versa). Then, in follow-up experiments, we manipulated gaze in two different ways: first with a trained human confederate, in live conversations in the lab and on Zoom, and second with an online conversation simulation using an open-access conversation paradigm I designed. We thus find causal evidence that receiving gaze increases psychological safety, participation, and voice.

We identify several important moderators that define the link between leader eye gaze and employee participation. First, the effect of eye gaze on safety and speaking is strongest when it comes from the group leader (as opposed to another team member). Gaze is also more powerful if it is received while an employee is silent (compared to while they are speaking), more impactful for recipients who are more introverted (as opposed to extraverted), and more impactful for recipients who are racial minorities (as opposed to non-minorities). For eye gaze to be most effective, it depends on whom they receive it from, when they receive it, and what kind of person they are. In summary, Chapter 1 shows how gaze, a non-verbal cue, affects interpersonal perceptions of psychological safety and ostracism, which affects the choice to speak up or not.

In my second dissertation chapter, co-authored with Adam Mastroianni and Alison Wood Brooks, entitled *How Verbal and Nonverbal Cues Mislead Interpersonal Inferences,* we compare the predictive power of verbal, nonverbal, and paralinguistic cues during job interviews. We challenge traditional findings that interviews do not help improve predictive accuracy for GPAs or objective performance measures (Dana et al., 2013; Dubovsky et al., 2008). We show that

nonverbal and paralinguistic skills gleaned from interviews can help predict performance in domains that have not been studied, including interpersonal skills.

For this project, we conducted and recorded structured job interviews with 105 people and asked them to subsequently perform a panel of 7 common work tasks, ranging from a group problem-solving conversation to private math questions. This established ground-truth links between interview behavior and task performance. Next, we modified the interview recordings in a 2 (verbal cues observable: yes v. no) x 2 (nonverbal cues observable: yes v. no) x 2 (paralinguistic cues observable: yes v. no) full factorial design. For example, distorting an audio clip using a low-pass filter preserves the tone of voice but makes the content incomprehensible. Leaving only paralinguistic content, and removing the audio completely from a video and replacing it with subtitles preserves nonverbal and verbal content, but removes paralanguage. We then asked a separate set of participants (over 4,000) to evaluate these stimuli. This design allowed us to explore whether interpersonal perceptions and predictions about job performance differ based on the type of conversational information available to evaluators.

We found that having access to paralinguistic and nonverbal content allowed evaluators to more accurately predict interviewees' performance on tasks evaluating interpersonal skills, such as conversational likeability (as rated by the interviewees' coworkers) and empathic writing (as rated by third-party raters), while verbal content did not increase predictive accuracy. However, we find that, when given the choice, people overwhelmingly prefer to receive verbal content, and felt more comfortable with verbal content when evaluating interpersonal skills—a clear mistake. No type of information increased predictive accuracy for objective tasks like math questions, replicating prior work that interviews do not predict grades and GPAs, (although, again, people felt more confident with verbal content and preferred to receive verbal content). People may intuit

our findings—that their nonverbals (including their appearance, body language, and facial expressions), and tone of voice (which may communicate information such as country of origin and social class) influence others' perceptions of them, including how likable and interpersonally skilled they are. This may lead people who have less-desirable appearances and paralinguistic cues (people from minority backgrounds, people who sound less confident) to speak less in conversations, knowing that nonverbal and paralinguistic cues factor in just as much to peoples' interpersonal perceptions as their verbal content. In this way, this demonstration of how nonverbal and paralinguistic elements may factor more into interpersonal perceptions compared to verbal elements may shed some light on why certain people may choose to remain silent in workplaces, social settings, and classrooms.

In my third dissertation chapter, solo-authored, entitled *Interpersonal Perceptions of Non-Native English Speakers,* I show how paralinguistic cues can, in certain contexts, influence our interpersonal evaluations in negative ways, and lead us astray. I hypothesize that non-native English speakers may be evaluated more negatively when paralinguistic cues are available (i.e., when people can hear their accents). However, I demonstrate that these negative evaluations are misguided because when the verbal content is considered in isolation (i.e., as text only), the evaluation differences between native and non-native speakers disappear, and in some cases even reverse (non-native speakers are evaluated *more* positively). This has to do with the amount of effort it takes to speak in a native language; when speaking a non-native language, people are constantly deliberating, translating, and rehearsing words before speaking, relying overall more on System 2 processing as opposed to System 1 processing. This leads them to be less susceptible to cognitive biases such as framing effects and loss aversion (Costa, Foucart, Arnon, Aparici, & Apesteguia, 2014; Keysar, Hayakawa, & An, 2012). I hypothesize, and find, that although non-

native speakers are judged more poorly in the presence of paralinguistic cues, when evaluating their verbal content in isolation, the words they say are equally high quality or even higher quality compared to native speakers. Similarly to Chapter 2, this project may shed light on non-native speakers' decisions to speak up. Non-native speakers have likely developed a lay intuition that they are being judged based on their paralinguistic information, perhaps even more so than the quality of their verbal content (when paralinguistic cues are available). This may lead non-native speakers to speak less. Thus, my demonstration of how paralinguistic elements obscure judgment of verbal content helps us understand a reason non-native speakers may choose to remain silent in workplaces, social settings, and classrooms.

Overall, my dissertation shows how nonverbal cues can be strategically employed to increase feelings of psychological safety, and consequently the likelihood of speaking up in groups (Chapter 1), and how nonverbal and paralinguistic cues can lead to increased accuracy when evaluating a person's interpersonal skills (Chapter 2). But paying increased attention to nonverbal and paralinguistic cues is necessarily a panacea; I also show how paying attention to paralinguistic cues can lead us astray, negatively evaluating people with accents although their verbal content is just as good—and maybe even better—than those without accents (Chapter 3). Therefore, I also show how nonverbal and paralinguistic cues can mislead our perceptions of others. The latter two chapters, which show how we rely on nonverbal and paralinguistic cues for interpersonal evaluations (Chapter 2) and how paralinguistic cues mislead judgments of verbal content (Chapter 3) help us understand the reasons certain people, who perhaps have less-desirable appearances or paralinguistic cues, may choose to remain silent in workplaces, social settings, and classrooms.

CHAPTER 1: FEELING SEEN: LEADER EYE GAZE PROMOTES PSYCHOLOGICAL SAFETY, PARTICIPATION, AND VOICE

Summary: Psychological safety is a hallmark of effective team functioning. Although prior work shows that leadership styles are associated with employees' feelings of psychological safety and subsequent decisions to speak up, we know very little about "the specific behaviors leaders employ to lead employees to assess an interaction as safe to speak" (Morrison, 2011). In this paper, we shift scholarly focus from gestalt employee evaluations of leadership styles to observing leader behavior directly. We examine how a specific leader nonverbal behavior—eye gaze—affects group members' feelings of psychological safety and subsequent participation and voice behaviors. In three live group conversations, on Zoom and face-to-face (N= 569), and one computer-simulated conversation experiment (N= 578), receiving more eye gaze from the group leader predicted more participation and voice (correlationally and causally), and this relationship was mediated by increased feelings of psychological safety and decreased feelings of ostracism. These relationships were moderated by individual characteristics of gaze recipients, such that the effects of receiving eye gaze were stronger for racial minorities and more introverted individuals. When people feel seen, they feel safer and more included and are more likely to speak up.

Can leaders create a context where employees feel safe to voice their views? Prior work in organizational behavior suggests the answer is yes. But *how*? An extensive body of work has highlighted the benefits of employee participation (Woolley et al., 2010) and voice, a specific type of participation that challenges the status quo and seeks to improve it (Edmondson, 2003; Grant & Ashford, 2008; Morrison, 2011; Van Dyne & LePine, 1998; Barry & Stewart, 1997; Hyatt & Ruddy, 1997; Mesmer-Magnus & Dechurch, 2009; Detert, Burris, Harrison & Martin, 2013), and the important role that leaders play in establishing the psychological safety required to encourage participation and voice (Edmondson, 1999; Nembhard & Edmondson, 2006). However, relatively few studies have examined the behavioral underpinnings that leaders use to establish a workplace environment that employees perceive as safe to speak up (for a recent exception, see Coutifaris & Grant, 2021).

The benefits of team psychological safety are well documented: when employees believe they can take risks without being punished, they participate more often, especially with challenging or dissenting views that are likely to improve their groups and organizations (e.g., Edmondson, 1999; Detert & Burris 2007; Sherf, Parke & Isaakyan, 2021; Frazier, Fainshmidt, Klinger, Pezeshkan, Vracheva, 2017). The fact that leaders play a key role in establishing psychological safety is similarly well established (Edmondson, 1999; Nembhard & Edmondson, 2006). In particular, prior work suggests that certain leadership styles are associated with more psychological safety (e.g., authentic leadership; Hsiung, 2012, leader openness; Detert & Burris, 2007, moral leadership; Chan, 2014, leader trust and empowering leadership; Gao, Janssen, & Shi, 2011) while other leader traits are associated with less psychological safety (e.g., low managerial self-efficacy beliefs; Fast, Burris, & Bartel, 2014).

However, prior work showing leadership styles are associated with increased psychological safety has at least two main drawbacks. First, this work largely relies on gestalt judgments made by employees about their leader's style, as opposed to directly observing leaders' behaviors. Indeed, out of 185 papers in the last decade linking leadership styles with psychological safety and related outcomes, 97% use correlational, survey-based methods (Edmondson & Bransby, 2023). An emerging critique has argued that relying on others' evaluations of leaders (as opposed to directly studying leaders) may lead to confounded or tautological links between leadership styles and outcomes (Van Knippenberg & Sitkin, 2013; Carton, 2022; Fischer, 2018; Fischer & Sitkin, 2023). For instance, an employee that likes their boss and feels psychologically safe and able to speak up at work may also rate her boss as having a positive leadership style, regardless of whether or not her boss exhibits the specific behaviors associated with this style (c.f. Banks, Woznyj, & Mansfield, 2021).

A second drawback comes from the multi-dimensional nature of previously studied leadership constructs. According to a recent review, leadership styles are multi-dimensional to the point of being overly generic: ethical leadership includes 38 behaviors, humble leadership involves 9 behaviors, servant leadership is composed of 28 behaviors, and so on (Carton, 2022). Such multi-dimensionality obscures precise causality (Ashford & Sitkin, 2019; Van Knippenberg & Sitkin, 2013; Fischer & Sitkin, 2023). With so many sub-components, it's hard to know which components of these leadership styles lead employees to feel more psychologically safe, and which do not.

Lastly, although gestalt evaluations of leaders may be positively associated with ongoing, stable climates of psychological safety and related outcomes, they tell us little about how leaders can foster psychological safety. Understanding episodic psychological safety, or the extent to

which employees feel safe during a particular interaction, is important in the context of non-permanent team structures, such as project-based consulting teams, academic ad-hoc committees, or creating a safe environment during one-off interactions (e.g., seminars, workshops, or conferences). With workplace teams becoming increasingly project-based and contract work increasing, understanding how leaders can foster psychological safety quickly is increasingly important. Moreover, understanding how to use behaviors to create episodic psychological safety is important if leaders want to specifically solicit input from a certain employee or a group of employees, whose input may be desired.

This leaves us to ask: What exactly can leaders *do* to foster psychological safety, in the moment? Are there identifiable causal links between leader behaviors and psychological safety? And can other leaders learn to enact those behaviors?

In this paper, we move away from multidimensional leadership styles as antecedents, and instead investigate leader behavior directly. Specifically, we look at a link between discrete nonverbal behavior, leader eye gaze, and employee psychological safety. While the content of a leader's verbal communication is crucial, nonverbal cues can be quite influential, perhaps because people believe they are less easily faked (Hall, Bernieri, & Carney, 2005). Leader nonverbal behaviors can be particularly influential, with recent evidence that leader nonverbal behavior affects perceptions of social hierarchy in a group (Shim et al., 2020) as well as team members' speaking time (Locke & Anderson, 2015).

In this project, we focus on the eye gaze group members receive from the leader of a group. We focus on leader eye gaze for three reasons: first, because it is one of the most important and pervasive forms of nonverbal behavior (people must decide where to look at all times, whether consciously or subconsciously). Second, eye gaze is a profound signal of attention, focus, and

respect, and as such has a profound impact on the feelings of others (Kampe, Frith, Dolan, & Frith, 2001; Conty, George, & Hietanen, 2016; Wirth, Sacco, Hugenberg, & Williams, 2010). Third, we suspect that eye gaze may be particularly influential in group interactions, where more people must share airtime, with less airtime available for verbal contributions per person on average, compared to dyadic interactions (Cooney et al., 2020). Yet, it is also unclear how behavior like eye gaze, which can only be directed to one group member at a time, can influence the psychological safety of all group members who may not receive as much leader attention.

In exploring whether receiving eye gaze from leaders increases psychological safety and decreases feelings of ostracism, which in turn, may influence employee voice and participation, our methodological approach follows an emerging emphasis on analyzing people's discrete behaviors (captured in unfolding interaction data such as transcripts and video recordings) *alongside* their individual characteristics as captured in survey data (e.g., Gallus & Bhatia, 2020; Li, Packard & Berger, 2020; Brown et al., 2020; Hart et al., 2021; Huang et al., 2017; Mastroianni et al., 2021; Sun & Slepian, 2020; Swaab et al., 2021; Truong et al., 2020; Voigt et al., 2017; Yeomans & Brooks, *working*; Yeomans et al., 2021). This approach allows us to capture the interplay between (and disentangle the unique contributions of) a discrete behavior, leader eye gaze, and person-level traits (e.g., employee introversion, hierarchical status, race) on people's feelings (psychological safety and ostracism) and subsequent behavior in response (participation and voice).

Eye Gaze

Together with smiling, close physical distance, direct orientation, and hand gesticulation, eye gaze belongs to a group of nonverbal behaviors that connote "immediacy" (Prisbell, 1985). Direct

gaze from others activates the ventral striatum, a brain region associated with reward processing (Kampe, Frith, Dolan, & Frith, 2001), where the reward, in this case, is interpersonal: interest and attention from one's interaction partner (Conty, George, & Hietanen, 2016; Frischen, Bayliss, & Tipper, 2007; Friedman, 1967; Jarick & Kingstone, 2015; Jones & Cooper, 1971; Khalid, Deska, & Hugenberg, 2016; Wohltjen & Wheatley, 2021).

Humans pay significant attention to where other people are looking and spend a disproportionate amount of time visually fixating on the eyes of others in photographs compared to other faces and body parts (Birmingham, Bischof, & Kingstone, 2008). The amount spent gazing at the eyes of others increases as the number of people in the photo increases, especially when photos are more social in content (i.e., group conversations as opposed to sitting in a waiting room), demonstrating that eye gaze takes on increased importance in group conversations (Birmingham, Bischof, & Kingstone, 2008). Fixation on others' gaze starts from a young age. Infants begin to monitor and follow the direction of adults' visual attention around three months of age (Frischen, Bayliss, & Tipper, 2007), and research on "gaze cueing" suggests that people learn early on to detect stimuli that are placed in line with where a cartoon face is "looking" faster and more readily than stimuli placed away from where the face is "looking" (Friesen & Kingstone, 1998). Eye gaze indicates the attention of others, and when we are looking for social support in group conversations, we look at where other group members are directing their gaze.

Importantly, group conversation is a categorically different task than dyadic interaction—group interaction does not demand continuous, reciprocally sustained engagement, and in groups, people often use eye gaze as a turn-taking mechanism, to signal who should speak next (Cooney et al., 2020). Group members can remain silent, and the conversation can carry on. However, group

members can contribute to the group interaction in meaningful ways even while they aren't speaking by using nonverbal cues like eye gaze.

Eye gaze not only signals the gazer's attention and grabs the perceiver's attention, but it also signals positive evaluation. Watching where others direct their eye gaze imbues the target of their gaze with more positivity and value (Bayliss, Paul, Cannon, & Tipper, 2006; Corneille, Mauduit, Holland, & Strick, 2009), and eye gaze functions as a signal of respect, approval, receptiveness, and recognition toward a receiver. Those who receive it tend to feel valued, accepted, and safe (Friedman, 1967). Even receiving brief eye contact from a passerby can increase feelings of connectedness (Wesselmann et al., 2012), and receiving eye contact from another person increases the perceived value of the relationship (Wirth et al., 2010). Relatedly, withholding eye contact can indicate a lack of involvement and can suggest non-responsiveness toward another (Young, Slepian, Wilson, & Hugenberg, 2014), making the targets feel excluded and ostracized (Wirth et al., 2010).

However, eye gaze is not unequivocally positive. Gaze communicates attention, and sometimes attention is unwanted. In certain contexts, including in the animal kingdom, eye gaze indicates a struggle for dominance (Emery, 2000). Research shows that in contentious interpersonal contexts, the maintenance of eye gaze indicates a struggle for dominance (Schmeichel & Tang, 2015). Maintaining direct gaze with others, especially for a long time, is a characteristic of dominant behavior exhibited by high-power individuals (Carney, Hall, & LeBeau, 2005; Kleck & Nuessle, 1968). In classic research on hierarchies, dominance hierarchies were found to be positively related to more eye contact and direct gaze (Strongman & Champness, 1968). More recent research shows that receiving more eye gaze causes people to make more dominant choices in a gaze ultimatum game (Schmeichel & Tang, 2015), and behave more

competitively in negotiations (Stuhlmacher & Citera, 2005). Thus, the effect of eye gaze is more nuanced than it may first appear.

Leader Eye Gaze

We are interested in understanding the effect of leader gaze specifically and among teams in the context of work. According to evolutionary scientists, monitoring and following the eye gaze of high-status group members is an adaptive process—paying attention to high-status members brings survival advantages (e.g., Van Vugt, Hogan, & Kaiser, 2008). This finding extends to primates; macaque monkeys follow the gaze of high-status monkeys more than the gaze of low-status monkeys (Shepherd, Deaner, & Platt, 2006). Indeed, one of the primary functions of nonverbal behaviors is that they reveal information about a person's personality, intentions, and attitudes (Ambady, Bernieri, & Richeson, 2000). Goffman (1959) noted that we are constantly "giving off signals" via nonverbal behavior that are interpreted by those around us (either accurately or inaccurately) as expressive of our underlying attributes. In workplace settings, where leaders have a large influence over followers' experiences and outcomes (i.e., high power over resources), it makes sense that followers attend to the leader's nonverbal communications to determine appropriate behaviors to display and the extent to which they contribute divergent ideas (Bonaccio et al., 2016).

Because people pay outsized attention to leader gaze, where leaders choose to direct their gaze is likely to impact individual and group outcomes. For example, when leader gaze distribution is more balanced across a team, group members perceive their team as more democratic, egalitarian, and fair—perceptions associated with higher workplace satisfaction and feelings of authenticity (Koch et al., 2010). Experienced leaders may intuit this; eye tracking goggles reveal that, in

classrooms, expert teachers are more egalitarian in their gaze distribution compared to novice teachers (Cortina et al., 2015), and receiving teacher gaze is associated with positive teacher evaluations from students (McCroskey, Richmond, Sallinen, Fayer, & Barraclough, 1995). Recent work suggests that leaders who increased their eye gaze towards racial minorities reduced collective perceptions of status disparities in the group, leading to better group performance in collective decision-making tasks (Shim, Livingston, Phillips, & Lam, 2020). This work also suggests that leaders are able to intentionally direct their gaze, and doing so can change group dynamics.

In sum, eye gaze signals attention, respect, and a visible signal of positive evaluation, increases perceptions of egalitarianism, and decreases status disparities. Early literature on psychological safety underscores the importance of leader respect in establishing climates of psychological safety (Kahn, 1990), while more recent work suggests the importance of employees feeling like their leaders are paying attention to, and listening to them, in establishing climates of safety (Castro et al., 2018). Nembhard & Edmondson (2006) find that leader inclusiveness and decreased status disparities between team members increase psychological safety. Leaders attending to and eliciting participation from employees also predicts psychological safety (Chen, Wadei, Bai, & Liu, 2020). As such, we expect that receiving leader eye gaze will increase feelings of psychological safety.

Hypothesis 1a: Receiving leader eye gaze increases psychological safety.

On the other hand, *not* receiving gaze inhibits the inclusion of individuals in conversation, partitioning the conversation into participants who are included and those who are not (Stivers, 2021; Valsesia et al., 2021). Averted eye gaze is the most frequently used nonverbal cue to enact ostracism (Wirth et al., 2010). Compared to the safe, inclusive feelings associated with receiving

gaze, an averted gaze conveys a negative interpersonal evaluation—receiving less gaze reduces self-efficacy and decreases feelings of social connection with an interaction partner (Wirth et al., 2010), even among strangers (Wesselmann et al., 2012).

An extensive body of research shows that ostracism—feeling excluded or ignored by individuals or groups—is a powerful interpersonal emotion, which threatens people's sense of belonging, control, self-esteem, and meaning, and harms their ability to make sound decisions (Jones & Kelly, 2010; Williams, 2009). Even minimal forms of exclusion can lead to negative psychological consequences. For example, in the popular Cyberball paradigm, used in over 200 published papers with over 19,500 participants, group members can be made to feel ostracized if they are not passed a pretend ball after an initial phase of being included in a simulated game of catch (Hartgerink, Van Beest, Wicherts, & Williams, 2015). The emotional anguish associated with feeling ignored, even within the Cyberball paradigm, is similar to physical pain (Riva, Wirth, & Williams, 2011), and the same neural mechanisms underly the experience of physical pain and social pain (Eisenberger & Lieberman, 2005). The pain of ostracism can be triggered by minimal cues, such as when people feel "out of the loop" during a conversation (e.g., when they are on the outside of inside talk, or under-informed about a topic of shared knowledge among others in the group; Iannone, Kelly, & Williams, 2018).

Because people pay particular attention to leaders' eye gaze, and because receiving gaze can reduce feelings of disrespect and negative evaluation, we predict that receiving leader gaze during a group conversation will decrease feelings of ostracism:

Hypothesis 1b: Receiving leader eye gaze decreases feelings of ostracism.

Psychological Safety, Ostracism, and Speaking Up

We suspect that the attention, respect, and fairness that eye gaze conveys (Bayliss, Paul, Cannon, & Tipper, 2006; Corneille, Mauduit, Holland, & Strick, 2009; Friedman, 1967) not only leads to feelings of connectedness and psychological safety (Wesselmann et al., 2012) but may also impact the behavioral responses of those who receive it. When people feel as though others are paying attention to them and value what they have to say, they are more likely to participate in the group discussion (Liang & Yeh, 2020). A hallmark of classic theories like expectation states theory (Balkwell, 1991) and status characteristics theory (Berger, Cohen & Zelditch, 1972) rest on the assumption that people come to understand the expectations that other group members have for them, which then influences the extent to which they participate in the group's discussion. Nonverbal signals like eye gaze likely send signals that will influence subsequent participation in the group's discussion (Ridgeway, Berger, & Smith, 1985). We expect that eye gaze, especially from a leader, will invite more individuals to share their ideas.

Compared to verbal triggers such as cold calls ("What's on your mind?"), eye gaze may be especially effective in making people feel welcome to participate if and when they are ready to speak (Li & Tian, 2016; Gkorezis, Panagiotou, & Theodorou, 2016; Twenge et al., 2007; Wu et al., 2019), without putting them on the spot. Further, because employee voice is a specific type of communication that is not only improvement-oriented but also challenging in nature (Van Dyne & LePine, 1998), employees are likely attuned to such subtle, nonverbal signals to determine whether their challenging voice is welcome (Dutton et al., 1997). If leaders send signals that convey less receptiveness to an individual or that the relationship is not strong, employees tend to withhold their voice (Burris, Detert & Chiaburu, 2008).

There is a strong and well-documented link between psychological safety and speaking up (Detert & Burris, 2007). Many studies, spanning a wide range of contexts, have found that

psychological safety often mediates the relationship between antecedent variables and speaking up (e.g., Ashford et al., 1998, Miceli & Near 1992, Walumbwa & Schaubroeck, 2009). As such, we expect receiving eye gaze to not only lead to increased feelings of psychological safety, but also to influence subsequent behavior such as participation and voice in response.

Hypothesis 2a: Receiving leader eye gaze increases participation and voice.

Hypothesis 2b: The relationship between leader eye gaze and participation and voice is mediated by increased psychological safety and decreased feelings of ostracism.

Group Member Traits as Moderators

We expect that the effects of eye gaze on psychological safety, ostracism, and speaking up will be moderated by certain employee characteristics, given that individual characteristics interact with situational factors (such as gaze received) to contribute to the likelihood of employees engaging in proactive behavior, such as speaking up in the workplace (Grant & Ashford, 2008). Existing research tells us that speaking may not be highly correlated with expertise (Coffman, 2014; 2014; Anderson & Kilduff, 2009). Instead, speaking is distributed by subjective status in the group, which is strongly influenced by demographic characteristics—for example, underrepresented minorities speak less, introverts speak less, non-native English speakers speak less, and women speak less (Harrison, Harrison, & Shaffer, 2019; Morrison & Milliken, 2000; Rashidi, Yamini, & Shafiei, 2011; Lu, Nisbett, & Morris, 2022; Brescoll, 2011).

In this project, we propose that leader eye gaze will have an outsized impact on employees who typically feel less "visible" and speak less in organizations, and who accordingly are more perceptive to signals of attention, respect, and an expectation for contribution to the group's

conversation. We specifically hypothesize about the effects of race, a demographic cue of power (Talaifar et al., 2021; Ridgeway & Walker, 1995), and employee introversion (Tuovinen, Tang, & Salmela-Aro, 2020; Williams, 2009).

Prior work suggests that non-majority employees face greater pressure to remain silent in the workplace (Morrison & Milliken, 2000) and are less inclined to speak up at work compared to majority members (Harrison, Harrison, & Shaffer, 2019). Even when they do participate, their attempts to speak up are recognized less by managers (Howell, Harrison, Burris, & Detert, 2015). As a result, compared to racial majority members, racial minorities are highly attuned to social cues such as eye gaze, especially from leaders (Shim, Livingston, Phillips, & Lam, 2020). In contrast, the majority of members are likely to feel more psychologically safe and speak up more overall, although their ideas may be less creative because their perspective is shared by many (Harrison, Harrison, & Shaffer, 2019). High-status group members and racial majority members may be less sensitive to the social context when deciding whether to participate and/or voice their views, making the impact of eye gaze on psychological safety and voice less pronounced. Taken together, the effect of eye gaze may be stronger for minority group members.

Similarly, because introverts are already less inclined to participate in groups (Rashidi, Yamini, & Shafiei, 2011) they may be more sensitive to the eye gaze of other team members, including leaders. Introverts are more sensitive to signals that others are socially engaged with them; when introverts perceive others are engaging with and including them, they feel better in social interactions (compared to introverts who do not perceive that others are engaging with them), an effect that is less pronounced for extraverts, who are more vocal overall, and less sensitive to engagement from others (Tuovinen, Tang, & Salmela-Aro, 2020). Compared to introverts, extraverts also feel less ostracized after being told that no other participants in an

experiment wanted to work with them because they require less social validation in social interactions (Williams, 2009). Extraverts tend to feel more psychologically safe and hence participate and engage in voice more frequently overall (LePine & Van Dyne, 2001), irrespective of the level of eye gaze they receive. Thus, the effect of eye gaze may be stronger for introverts than extraverts.

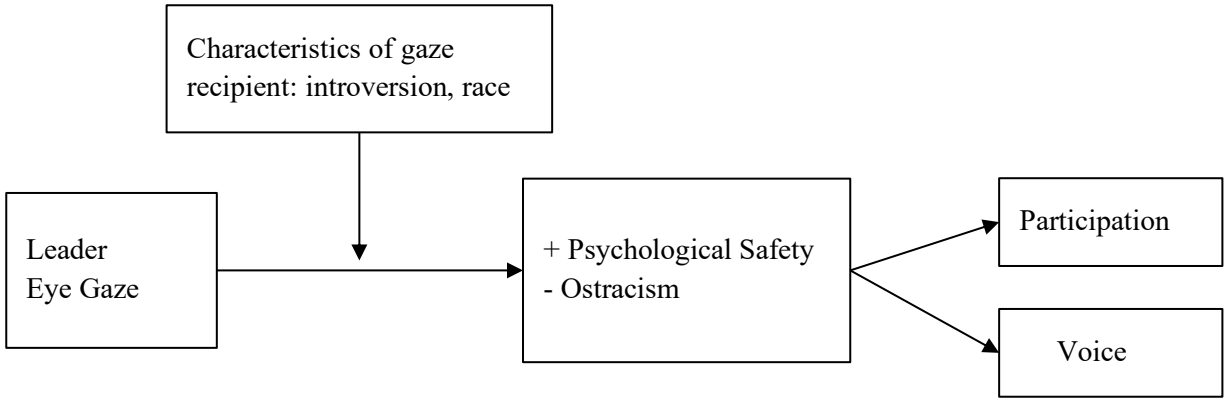
Hypothesis 3a: The relationship between leader eye gaze on psychological safety, ostracism, and speaking up will be stronger for racial minorities (compared to those in the racial majority).

Hypothesis 3b: The relationship between leader eye gaze on psychological safety, ostracism, and speaking up will be stronger for introverts (compared to extraverts).

Our theory centers on how eye gaze signals attention, fairness, and respect, which then fosters psychological safety, which in turn influences participation in group decision-making and voicing challenging opinions. However, because eye gaze is necessarily zero-sum (a leader cannot simultaneously gaze at two people), our theory suggests a potential tension between individual- and group-level outcomes. If an individual receives eye gaze at the expense of other group members, it is possible that the recipient of gaze would feel more psychologically safety while other group members would feel correspondingly less psychological safety. That is, one interpretation of our theoretical model is that feelings of psychological safety are determined by the amount of eye gaze each individual receives, and that as any one group member receives more leader eye gaze, other group members might feel correspondingly less psychological safety.

Yet, because our theory is also tied to conversations where contributions, participation, and ideas should advance the group's decision-making (Berger, Cohen & Zelditch, 1972; Ridgeway, Berger, & Smith, 1985), another possibility may be that eye gaze should increase psychological safety even when directed to others who are speaking (and advancing the group's decision-making). Giving attention and respect to individuals who are contributing to the group's success is likely to be seen as fair and appropriate, and something that groups well recognize (He et al., 2017). By a similar vein, directing gaze to one person while ignoring all others creates the very status disparities that prior work has shown to undermine psychological safety for multiple group members (Nembhard & Edmondson, 2006). Thus, while leader eye gaze might seem like a zero-sum dynamic (a leader can only look at one person at a time), if eye gaze is appropriately distributed, we predict that it can increase psychological safety for all group members. To understand the effects of leader gaze on group-level outcomes, we will look at the effects of leader gaze equity (the extent to which the leader's total gaze time was distributed equitably across group members) in our studies. We will also evaluate the effects of leader attentiveness to speakers (the extent to which the leader's total gaze time was directed toward the person speaking throughout the conversation), and the effects of leader overall attentiveness to others (the extent to which the leader's total gaze time was directed toward any other humans in the conversations versus looking away, or looking at something else).

Figure 1.1: This figure depicts our theoretical model and all hypothesized relationships therein.



Overview of Studies

Across three studies, we explore the relationships between leader eye gaze, and group member feelings (psychological safety and ostracism) and speaking up (participation and voice) in response. In Study 1, we examine the correlational relationship between eye gaze and speaking on a turn-by-turn basis in unconstrained face-to-face conversations between participants.[1] In Study 2A, we experimentally manipulate eye gaze by training an experimenter, positioned as the group leader, to direct high or low levels of eye gaze toward participants in Zoom meetings. In Study 2B, we again use a trained experimenter to manipulate leader eye gaze, but in the context of in-person meetings. In Study 3, we varied the amount of gaze participants received from the group leader in a computer-simulated conversation that allowed us to control verbal content across the experimental conditions.

Open Science Disclosure:

[1] The correlational relationship between eye gaze and airtime is replicated in a supplemental dyadic study in Appendix 1A.

We report how we determined our sample sizes, all data exclusions, all manipulations, and all measures. Data, analysis code, stimuli, and preregistrations for all studies are available on OSF: https://osf.io/68d5j/?view_only=9a0077ed40524257adce0811b439e460.

Study 1: Eye Gaze in Live Conversations

In our exploratory (not pre-registered) Study 1, we observed eye gaze and speaking time during unconstrained group conversations between participants in groups of three. We captured participation (airtime, based on timestamps), eye gaze (based on third-party coding of the conversation videos), and voice behaviors (via post-hoc content coding) at every turn of the conversation. Participants self-rated their feelings of psychological safety after the conversation. This study allowed us to explore correlational relationships between leader gaze and participation, voice, and psychological safety in a naturalistic group setting. Because the groups had a leader and non-leaders, the design also allowed us to benchmark the effects of leader gaze against non-leader gaze, testing our assumption that leader gaze is especially impactful. Lastly, this study allowed us to explore group-level effects of leader gaze, including how the equity of the leader's gaze (the extent to which the leader's total gaze time was distributed equitably across group members) affected group psychological safety. We also explored how leaders' attentiveness to the person speaking, and leaders' overall attentiveness to employees (versus looking away or at something else) affected group psychological safety.

Method

Participants: We recruited 75 groups of three people to participate in live conversations in a behavioral lab (N= 225). One group encountered a technical issue with their video, leaving 74

triads (N= 222) for analysis (70% female, M_{age}= 23.6, SD= 6.27). These 74 triads produced a total of 7,304 conversational turns (M = 98.7 turns per conversation, SD = 44.1).

Design: Participants consented to participate in a three-person, face-to-face group meeting in a behavioral lab. One participant in each group was assigned to be the group leader at the beginning of the session. This person had asymmetric power over the other two participants, because they were assigned to have the final say in who the group chose to hire in the group task, even if the other two participants disagreed with them.

The groups worked on a hidden profile task, a group task in which all participants had asymmetric information about three job candidates, and had to decide who to hire. Some of the clues were shared between all the members, and some were unique to only one participant. The hidden profile task was structured such that one candidate was objectively more qualified than the other two candidates (more positive and less negative points compared to the other two candidates), but the superior candidate's clues were dispersed across group members, such that participants could only discover that they were the superior candidate if all three participants shared their asymmetric information with the group. The leader was given authority on behalf of the group, and the group was told that the leader would make the final decision. Each participant received clues about the three job candidates.

Measures

Eye Gaze: After the study had ended, we hired independent raters to watch the videos and code eye gaze at every conversational turn. Two research assistants coded the 15-minute videos of each conversation, reading alongside the transcripts. They coded the eye gaze of all three participants at each conversational turn (each row of the transcript). Their options for coding gaze were: "At Person A", "At Person B", "At Person C", "Down at their sheet", and "Away." Two

raters coded the first three conversations which consisted of 1,083 codes (361 conversational turns, and three targets of coding for each turn: Person A, Person B, and Person C). When the two raters disagreed, the ratings were flagged and the raters were asked to return to the video to recode. After this training, the intraclass correlation (ICC) for the first 3 videos was between .72 and .81 for the three people coded ($\kappa^{personA}$=.79, $\kappa^{personB}$=.81, and $\kappa^{personC}$=.72). After this, the coders separated the rest of the dataset and each coded half of the conversations. We created a measure of *leader gaze received* for each participant by summing the total number of turns in which they received gaze from the leader, and a measure of *other gaze received* by summing the number of total turns in which they received gaze from the other, non-leader participant. A sample coding guide is included in Appendix 1B.

Psychological Safety: We measured psychological safety using a same seven-item scale adapted from Edmondson (1999). Our measure yielded a Cronbach's alpha of .78. We include all survey measures in our Appendix 1C.

Participation: Because the meetings were video recorded and transcribed, we created a measure of individual participation by aggregating the total number of seconds that each participant spoke during the conversation (airtime).

Voice: The raters also provided ratings of voice (1 for voice, else 0) at each turn of the manuscript, indicating a "1" any time a participant made a comment that was intended to change the status quo or trajectory of the conversation. Voice is distinct from simply sharing unique information, because voice takes the conversational context into account—it aims to change the status quo of the conversation.[2] For example, contributing unique information that only the

[2] We also asked our raters to code for unique information sharing: anytime a participant brought up information that was exclusively known by them and not common knowledge amongst team members. We found that information sharing was distinct from voice, although information sharing and voice were, as expected, significantly correlated; *r*

participant had on their sheet would not necessarily be coded as voice. However, if the participant contributed unique information that went counter to the group's reasoning, the coders would code that as voice (i.e., the group is coalescing around Candidate B as the best option, and the participant mentions disconfirming evidence about Candidate B that only they have on their sheet). Each instance of voice was accompanied by an explanation in our data file. See sample coding guide in Appendix 1B. The intraclass correlation of the two raters was κ=.833, (95% CI = [.822, .843]), p <.001.

Results

Because we were interested in evaluating the effects of leader gaze on others, for the following analyses, we considered the responses of team members only, not leaders (reducing our sample size from 222 to 148).

Main Effects: See Table 1.1a below for correlations between variables. In a mixed effects regression model, with a random effect controlling for group (lme4 package, lmer function, in R), receiving more eye gaze from the leader significantly and positively predicted self-rated psychological safety (b= .16, t= 2.06, p=.041), seconds participated (b= 57.6, t= 8.09, p<.001), and expressions of voice (b= 1.08, t= 8.55, p<.001), supporting Hypotheses 1a and 2a. These results are summarized in Table 1.1c below. For context, each leader gazed at each non-leader an average of 24.77 times (*SD*: 16.89). Each additional instance of leader gaze above this average increased non-leader speaking time by 3.41 seconds. A leader increasing their gaze by one standard deviation (approximately 17 more gazes) led participants to speak approximately 58 seconds longer, assert one additional instance of voice (speaking up to challenge the status quo), and led to a 6% increase in ratings of psychological safety.

= -.20, p = .014. When group members share information, sometimes it challenges the status quo and sometimes it doesn't.

Table 1.1a: Study 1 Correlation Table

	Leader Gaze Received	**Psychological Safety**	Participation	**Voice**
Leader Gaze received	**1**			
Psychological Safety	**.17***	**1**		
Participation	**.81*****	**.28*****	**1**	
Voice	**.58*****	**.21****	**.57*****	**1**

Although we operationalized participation by focusing on the number of seconds spoken by each group member, two other potential measures of participation—total word count spoken and number of turns spoken—were highly correlated with the proportion of time spoken (*r*s > .71, *p*s <.001). Word count and number of turns spoken were also significantly predicted by leader eye gaze (word count: *b*= 135.86, t= 8.12, *p*<.001; number of turns spoken: *b*= 10.54, t= 11.94, *p*<.001). See Table 1.1b for a correlation table with all measures of participation.

Table 1.1b: Study 1 Alternative Operationalizations of Participation Correlation Table

	Seconds Spoken	**Word Count**	Number of turns
Seconds Spoken	**1**		
Word Count	**.71*****	**1**	
Number of turns	**.77*****	**.76*****	**1**

Table 1.1c: Effect of Leader Gaze on Participation, Voice, and Psychological Safety Results

	Dependent variable:		
	Participation (1)	Voice (2)	Psychological Safety (3)
Leader Gaze Received	57.622***	1.076***	0.169*
	(7.119)	(0.126)	(0.082)
Constant	206.170***	1.945***	-0.014
	(6.652)	(0.125)	(0.082)
Table 1.1c (Continued)			
Observations	142	146	146
Log Likelihood	-799.867	-269.082	-207.354
Akaike Inf. Crit.	1,607.734	546.164	422.708
Bayesian Inf. Crit.	1,619.558	558.098	434.643
Note:			+p< .10, *p< .05, **p <. -1, ***p<0.001

Gaze from leader versus employee: Because there were three-person groups, each non-leader participant could receive gaze from either the leader, or the other non-leader participant. As a check for our assumption that gaze is more impactful when it comes from a leader compared to non-leader, we evaluated the effects of receiving gaze from the leader or the other team member.

We created two separate and mutually exclusive measures: the number of times the participant received gaze from the leader (the predictor variable for all the main effects reported above), and the number of times the participant received gaze from the other group member (non-leader). To evaluate which was most impactful, we ran a regression including two simultaneous predictor variables in the same regression: gaze received from the leader and gaze received from the

employee.[3] With psychological safety as the outcome variable, we found that receiving gaze from a leader marginally predicted psychological safety (b= .17, t=1.82, p = .071), while receiving gaze from a non-leader did not significantly predict psychological safety (b= -.02, t=-.23, p = .82). With participation as an outcome variable, we found that receiving gaze from a leader significantly predicted seconds of participation (b= 42.68, t=5.29, p < .001), while receiving gaze from a non-leader also significantly predicted participation, but to a lesser extent (b= 21.38, t=3.02, p = .003)[4]. Lastly, with voice as an outcome variable, we found that receiving gaze from a leader significantly predicted expressions of voice (b= .97, t=6.36, p < .001), while receiving gaze from a non-leader did not significantly predict voice (b= .19, t=1.25, p = .21). See Table 1.2 for a summary of these results. This pattern of results supports our assumption that the effect of eye gaze is stronger when it comes from leaders compared to non-leaders.

Table 1.2: Effect of Leader vs. Non-Leader Gaze Results

	Dependent variable:		
	Participation	Voice	Psychological Safety
	(1)	(2)	(3)
Leader Gaze Received	42.678***	0.968***	0.181+
	(8.074)	(0.152)	(0.100)
Other gaze received	1.256**	0.190	-0.001
	(0.416)	(0.152)	(0.006)
Constant	177.729***	1.945***	0.016

[3] Note: we chose to run a regression with two simultaneous predictors (gaze received from leader, and gaze received from non-leader), allowing us to compare the relative size of the betas. We chose not to run a moderation analysis to avoid a model where one predictor (leader gaze received) is a subset of the other (total gaze received).

[4] Using the alternative operationalizations of participation, we found that receiving gaze from a leader and non-leader significantly predicted word count (b= 91.47, t=5.17, p < .001; b= 70.84, t=4.00, p < .001), and that receiving gaze from a leader and non-leader significantly predicted number of turns spoken (b= 8.94, t=10.46, p < .001; b= 6.47, t=7.57, p < .001)

	(10.997)	(0.125)	(0.154)
Observations	142	146	146
R^2	0.357	0.344	0.029
Adjusted R^2	0.348	0.335	0.015
Residual Std. Error	69.599 (df = 139)	1.512 (df = 143)	0.990 (df = 143)
F Statistic	38.584^{***} (df = 2; 139)	37.492^{***} (df = 2; 143)	2.129 (df = 2; 143)
Note:			$^{+}$p< .10, *p< .05, **p <. -1, ***p<0.001

Group-level effects: We investigated group-level effects of leader gaze—relative gaze—in two distinct ways: 1) the equity with which the leader distributes their gaze across all group members as a proportion of their total gaze, and 2) effect of the leader's consistency in looking at whomever is speaking, and 2).

Equitable Gaze Distribution. In an exploratory analysis, we evaluated whether the equity of gaze, or the extent to which the leader's total gaze time was distributed equitably across group members, affected participants' psychological safety. If the leader gazes proportionally more at one participant, does that participant feel better? To answer this, we calculated the proportion of leader gaze each individual received by dividing the gaze received by each participant, over the total leader gaze given to all participants. In a mixed effects regression model, with a random effect controlling for group, we found that the proportion of leader gaze received by each participant relative to total leader gaze did not significantly affect individual ratings of psychological safety in a simple regression (b= -.01, t=-.77, p = .44).

Aggregating this to the group level, we tested whether the equity of leader gaze affected the group's psychological safety. If the leader looks at everybody equally, does that make the group feel better? We calculated the absolute difference in proportional gaze given to each person (for example if person A received 25% of the leader's gaze, and person B received 75% of the leader's

gaze, the group score would be |.25- .75|= .5). The average leader gaze inequality score was .24 (*M*= 0.24, *SD* = .21). We then created a team-level psychological safety score by averaging the two non-leaders' ratings of psychological safety. Regressing leader gaze balance on team ratings of psychological safety, we found that leader gaze balance did not significantly predict team psychological safety (*b*= -.002, *t*=-.31, *p* = .76).

Leader Attentiveness to Speakers. Our data also allowed us to examine another question: does the leader's attentiveness to the speaker, or the extent to which the leader's total gaze time was directed toward the person speaking, indicating attention and respect to the speaker, increase group level psychological safety? For each conversational turn, we created a new column indicating whether the leader was gazing at the person speaking, or not (0/1 variable). We aggregated this variable to the group-level, to determine the number of turns the leader gazed at the speaker for each group. We also created a group-level psychological safety score by averaging the psychological safety scores of the non-leader group members. We then ran mixed-effects regression model with the leader directing gaze to the speaker as the predictor variable, group psychological safety as the outcome variable, and a random effect controlling for group (using the lme4 package, lmer function, in R). We found that the leader directing gaze to the speaker did significantly predict the group's psychological safety (*b*= .19, *t*=-2.19, *p* = .032).

Overall Leader Attentiveness. Lastly, we looked at the total overall amount of leader gaze towards either employee, as opposed to looking away, checking their phone, or looking down at their instruction sheet or their study materials. We found that overall leader total gaze towards employees significantly predicted team psychological safety (*b*=.11, t= 2.38, *p*= .020).

Exploring Causality with Lagged Analysis: While the data in this study are correlational, the turn-by-turn nature of conversational transcript data allow us to explore causality using a

lagged analysis design to explore the direction of the relationship between receiving gaze and participation (Martin et al., 2022; Voigt et al., 2017), by evaluating instances of gaze receipt and participation in the first half of the conversation (T1) versus the second half of the conversation (T2).

We ran a regression with T1 leader gaze received as a predictor, speaking time in T1 as a control, and speaking time in T2 as an outcome variable, and found that leader gaze received in T1 significantly predicted participation in T2 (b =.15, p <.001). This suggests that receiving more gaze earlier in a conversation may predict speaking later in the conversation.

In another model with speaking time in T1 as a predictor, gaze received in T1 as a control, and gaze received in T2 as an outcome variable, and found that speaking time in T1 does *not* predict gaze received in T2 (p = .48). Thus, speaking early in the conversation does not predict receiving gaze later in the conversation (see Appendix 1B for full results). Taken together, these lagged analyses provide suggestive evidence that receiving eye gaze from the leader in the first part of a conversation may influence speaking time later in the conversation, but not vice versa.

Discussion

In Study 1, we analyzed annotated transcript data from unconstrained face-to-face group conversations. We found that receiving more eye gaze predicted psychological safety, participation, and voice[5]. The equity of the leader's gaze, or, the extent to which the leader's total gaze time was distributed equitably across group members did not influence individual- or group-level psychological safety. However, the leaders' overall attentiveness to employees (versus

[5] We replicated this main effect in a separate sample of dyads, which we include as a Supplemental Study in the Appendix 1A. In dyads, like groups of three people, people who receive more gaze are more likely to speak.

looking away or looking at their materials) and the leader's attentiveness to the person speaking both positively predicted group psychological safety.

This study establishes an initial relationship between gaze, psychological safety, and participation and voice. However, due to the correlational design of this study, it does not preclude the possibility of reverse causality: feeling psychologically safe may lead to receiving more gaze from others, and voicing and participating may also lead to receiving more gaze. We explore these causal relationships by experimentally manipulating gaze in Studies 2A and 2B.

Study 2A: Experimentally Manipulating Gaze in Zoom Conversations

In Study 2A, we experimentally manipulated leader eye gaze during live Zoom meetings, to examine the causal relationship between leader eye gaze and psychological safety, ostracism (exclusion), participation, and voice. We additionally test Hypothesis 3b, which proposes that group members' level of introversion may moderate the relationship between leader gaze and group member response.

We recruited groups of four people, comprised of three participants and one confederate to participate in a study on Zoom. In virtual contexts, gaze cannot be dyadic as it is during face-to-face conversation. Online, an individual cannot give gaze or avert gaze from one specific person. In order to simulate dyadic gaze and the accordant attention it communicates, the confederate was trained to deliver continuous or averted gaze whenever a target participant was speaking. This is both a limitation of our study (and of virtual interaction more broadly), but also offers a conservative test of our hypothesis that receiving more gaze increases feelings of safety and speaking, as continuous or averted gaze may be less impactful in virtual contexts, where group members may feel less personally attended to or ignored overall.

Method

We preregistered hypotheses and analyses (https://aspredicted.org/blind.php?x=ep44hc).

Participants: We aimed to recruit 80 groups (each comprised of 3 participants and one confederate) to participate in live Zoom conversations in exchange for a $13 gift card, with 40 groups per experimental condition. A total of 87 groups (*N*= 261 people) completed the meeting, the post-conversation survey, and were included in our analysis (65% female, 2% non-binary, 33% male, average age= 29.47, SD= 12.31).[6] [7]

Our analyses focus on the target participant, assigned to the role of "Delivery Manager." This person was the target of the confederate's manipulation to provide either continuous gaze (high-gaze condition) or averted gaze (low-gaze condition) while they spoke. A total of 87 participants were assigned to the Delivery Manager role (62% female, average age= 30.68, SD= 13.91).

Participants were recruited via the SONA email listserv from two Behavioral Labs, the online platform Prolific, and other Slack channels and listservs. Participants recruited online were directed to register to take online studies with our behavioral lab recruiting platform (SONA), and study groups included a mix of participants recruited from online samples and from behavioral lab listservs. Groups were randomly assigned to be in the high-gaze or low-gaze condition.

Design and Procedure: Participants agreed to complete a conversation study about interpersonal relationships. We told them that they would have a short conversation with three other people on Zoom, which would be audio and video recorded, and that they would answer survey questions afterward.

[6] One participant did not complete their post-meeting survey, and was not included in our analysis, however we included the other two participants in their group in our analysis.

[7] Three groups were not video-recorded due to a technical issue, so for these three groups we have all self-report (survey) measures, but they are not included in our analysis of transcript measures.

They were assigned to complete a group decision-making task, assuming the roles of employees of a retail store, and discussing different delivery strategies and reaching an agreement by the end of the conversation. We randomly assigned participants to one of three roles: Delivery Manager, Retail Manager, and Warehouse Manager (See Appendix 1C for full participant instructions). The trained confederate always assumed the role of the group leader, Chief Operations Officer. The group was told that the group leader had authority over the group, and would make the final decision on behalf of the group. Importantly, three participants, including the confederate, were assigned to advocate for the same delivery strategy (Large Block Order strategy), while the participant assigned to the role of Delivery Manager was required to advocate for a different, opposing strategy (Just-In-Time strategy), positioning the Delivery Manager as an oppositional, dissenting party. The participant assigned to the Delivery Manager role was always the target participant for the confederate's gaze manipulation.

Groups were assigned to one of two between-subjects conditions: a *high-gaze condition*, and a *low-gaze condition.* In the high-gaze condition, the confederate was trained to maintain a continuous gaze with the target participant (the Delivery Manager). In the low-gaze condition, the confederate was trained to avert their gaze whenever target participant was contributing to the conversation. Specifically, whenever the target participant began speaking, the confederate was instructed to avert their gaze from the camera by either looking elsewhere on their screen, at their phone, or moving their head away from the camera. The confederate was instructed to clearly convey that they were listening to the target participants, and continued listening to, responding to, and asking follow-up questions to the target participant, even while averting their gaze away from the screen. Thus, we manipulated only the confederate's gaze. We focus our analyses on the target participant, who received the gaze manipulation from the confederate.

Measures

Psychological Safety: We measured psychological safety using a seven-item scale adapted from Edmondson (1999). Our measure yielded a Cronbach's alpha of .67. We include all survey measures in Appendix 1C.

Ostracism: We adapted Gerber et al.'s (2017) measure of ostracism, comprised of seven items, such as "During this conversation I felt poorly accepted by the other people I had a conversation with." and "During this conversation I felt good about myself." (1) *Strongly Disagree* to (7) *Strongly Agree*. Our measure yielded a Cronbach's alpha of .86.

Participation: As in Study 1, we created a measure of participation (airtime) for each participant by aggregating the total number of seconds that each participant spoke during the conversation (airtime).

Voice: In contrast to other-rated voice behavior in Study 1, in this study we measured voice using four self-reported items, adapted from Van Dyne and LePine (1998), including "During this conversation I felt I was able to express my views and feelings with the other people in the conversation" and "During this conversation, I suggested new ideas which were beneficial to the team." (1) *Strongly Disagree* to (7) *Strongly Agree*. Our measure yielded a Cronbach's alpha of .87.

Introversion/ Extraversion: We measured introversion and extraversion using four items adapted from the Multidimensional Introversion-Extraversion Scale (MIES scale, Open Psychometrics). This measure included items such as "I don't mind being the center of attention." and "I don't like to draw attention to myself" (R). Response categories ranged from (1) *Strongly Disagree* to (7) *Strongly Agree*. Responses below the scale mean indicated that a participant was

more introverted, while responses above the mean indicated the participant was more extraverted. Our measure yielded a Cronbach's alpha of .70.

Results

Table 1.3: Study 2A Correlation Table

	Psychological Safety	**Ostracism**	Voice	**Participation**
Psychological Safety	**1**			
Ostracism	**-.65*** **	**1**		
Voice	**.76*** **	**-.67*** **	**1**	
Participation	**.12**	**-.05**	**.21***	**1**

Psychological Safety: Analyzing the behavior of the target participant (the Delivery Manager), we found support for Hypothesis 1a: target participants in the high-gaze condition felt significantly more psychologically safe ($M = 4.81$, $SD = 1.10$) compared to participants in the low-gaze condition ($M = 4.30$, $SD = 1.02$), $t(84) = 2.26$, $p = .026$, $d = .49$.

Ostracism: We found evidence for our Hypothesis 1b: target participants in the high-gaze condition indicated that they felt significantly less ostracized ($M = 2.91$, $SD = 1.09$) compared to participants in the low-gaze condition, who felt more ostracized ($M = 3.49$, $SD = 1.46$), $t(84) = -2.07$, $p = .041$, $d = -.45$.

Voice: Providing suggestive evidence for Hypothesis 2a, we found that target participants who received less gaze felt directionally, although not significantly, less able to voice, as indicated in the post-conversation survey. In a t-test, target participants in the high-gaze condition indicated that they felt directionally more able to voice ($M = 5.14$, $SD = 1.11$) compared to participants in the low-gaze condition ($M = 4.72$, $SD = 1.40$), $t(84) = 1.55$, $p = .13$, $d = .33$.

Participation: We found evidence for Hypothesis 2a: target participants in the high-gaze condition participated for significantly more seconds (M = 244.06, SD = 92.78) compared to participants in the low-gaze condition (M = 198.14, SD = 94.38), $t(82) = 2.24$, $p = .028$, $d = .49$.[8]

Mediation: We tested whether ostracism or psychological safety mediated the effect of condition on participation (transcript measure) and voice (survey measure) among target participants. Because we had a sample of N < 100, we performed a permutation analysis recommended by Koopman and colleagues for mediation analyses with small sample sizes (Koopman, Howe, Hollenbeck, & Sin, 2015), using the supplemental code from their paper in R. In a permutation analysis, we found that psychological safety significantly mediated the effect of condition on voice [.03, .85]. In a follow-up bootstrap mediation model (10,000 iterations, lavaan package, R) a significant indirect effect emerged on psychological safety (*indirect effect* = .43, 95% CI [.33, .92], $p = .041$) as a mediator of condition on voice.

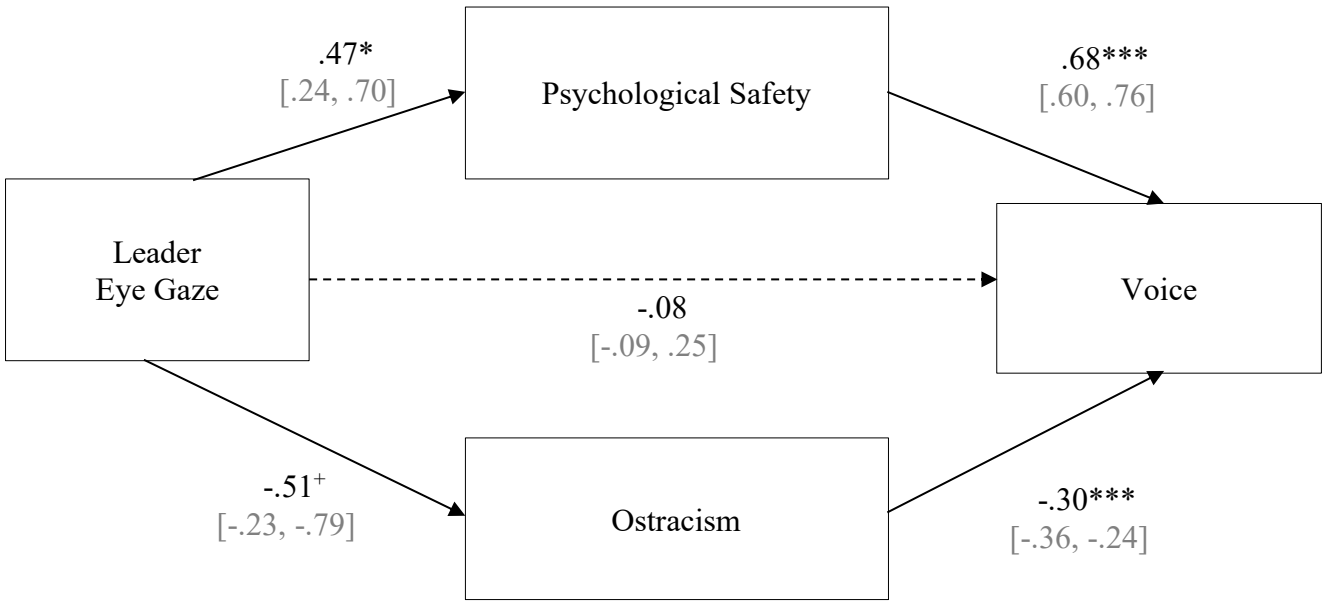
Turning to ostracism, in a permutation analysis, we found that ostracism significantly mediated the effect of condition on voice [.014, .74]. In a follow-up bootstrap mediation model (10,000 iterations, lavaan package, R), a marginal indirect effect emerged for ostracism (*indirect effect* = .31 [.003, .74], p=.060). We did not find evidence for ostracism and psychological safety as mediators on our other outcome variable, participation (p's = .78 and .81).

Results from a parallel mediation analysis (using the lavaan R package) including psychological safety and ostracism as parallel mediators indicated that psychological safety is a significant mediator while ostracism marginally mediates the effect of condition on voice. As depicted in Figure 1.2a, experimental condition predicts psychological safety (a1 = .47, $p = .039$),

[8] Variations in degrees of freedom reflect technical difficulties whereby three groups were not video-recorded, so we have a smaller sample for the participation outcome

which is related to more voice (b1 = .68, $p < .001$). Experimental condition marginally predicts ostracism (a2 = -.51, p = .069), which is significantly related to more voice (b2 = -.30, $p < .001$). Moreover, when considering both mediators, there is no direct effect of condition on voice (c= -.08, $p = .64$).

Figure 1.2a: Study 2A Parallel Mediation Model



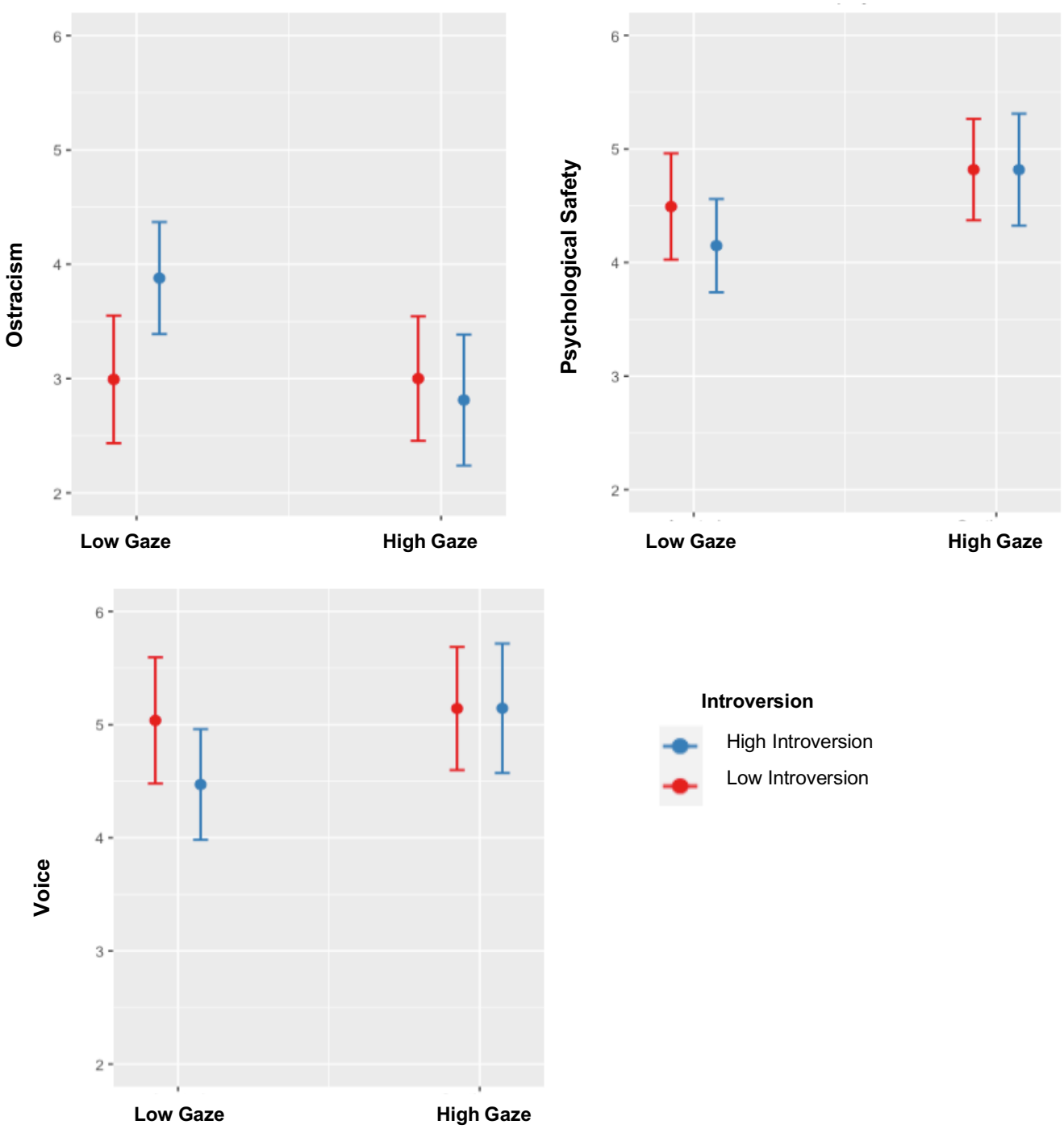
Moderation Analyses: To test Hypothesis 3b, that the effect of gaze on psychological safety is stronger for introverts compared to extraverts, we ran a (non-pre-registered) series of analyses. We ran a regression that predicted ostracism, with condition (high- vs. low-gaze), introversion, and the interaction between condition and introversion as predictors. We found a marginally significant interaction term, $b = 1.07$, $t = 1.94$, $p = .056$. Decomposing this analysis using a median split revealed that for more introverted target participants, there was a significant difference between ratings of ostracism among participants in the low gaze ($M = 3.88$, $SD = 1.58$) and high

gaze conditions (M = 2.81, SD =1.08), t(43)= 2.53, p = .015, d = .77, whereas for more extraverted participants, there was a non-significant difference between ratings of ostracism among participants in the low gaze (M = 2.99, SD =1.14) versus high gaze conditions (M = 3.00, SD = 1.11), t(39)= -.02, p = .98. As shown in Figure 1.3 below.

Turning to psychological safety, we did not find a significant interaction term, b = -.14, t = -.71, p = .46. However, in an exploratory t-test, we found that among more introverted target participants, participants in the high gaze condition felt marginally more psychologically safe (M = 4.82, SD =1.26) compared to participants in the low gaze conditions (M = 4.15, SD =1.13), t(42)= 1.84, p = .07, d = .48, whereas for more extraverted participants there was a smaller and non-significant difference in ratings of psychological safety between participants in the low gaze (M = 4.49, SD =0.84) versus high gaze conditions (M = 4.82, SD = 1.00), t(40)= 1.15, p = .26. This provides some initial, suggestive evidence that gaze may be particularly effective at increasing feelings of psychological safety for more introverted individuals (See Figure 1.3).

Turning to our outcome variables, we found that introversion significantly moderated the effect of condition on participation (interaction term: b = 35.25, t = -2.24, p = .028), and marginally moderated the effect of condition on voice (interaction term: b = .39, t = 1.78, p = .078), such that the manipulation was more impactful for more introverted participants. See Figure 1.3 for a graphical representation of these effects. Taken together these patterns provide suggestive evidence that gaze may be more impactful for more introverted people.

Figure 1.3: Study 2a Moderation by introversion.



Group Analyses: Thus far, we have considered the effects of being in the high-gaze or low-gaze condition for the target participant, who was the subject of our manipulation. However,

there were two other people in this group, which allowed us to explore additional questions about how the eye gaze manipulation affected group-level outcomes. To do so, we analyze the responses from all non-leader participants in the group, not just the target participants' responses.

In an exploratory analysis, we analyzed the effects of the gaze manipulation on the group, using a mixed-effects regression model with a random effect controlling for group (using the lme4 package, lmer function, in R). We found that participants in the low-gaze condition felt significantly more ostracized compared to participants in the high-gaze condition (b = .37, t = 2.45, p = .015). We found that participants in the low-gaze condition felt significantly less psychologically safe compared to participants in the high-gaze condition (b = -.29, t = -2.01, p = .047). We also found that participants in the low-gaze condition felt marginally less able to voice compared to participants in the high-gaze condition (b = -.29, t = -1.76, p = .078). Lastly, participants in the low-gaze condition felt spoke directionally less (M= 183 seconds, SD= 91) than participants in the high-gaze condition (M= 179 seconds, SD= 87), although this difference was not statistically significant (p =.78). This pattern suggests that the confederate's manipulation of providing high-gaze or low-gaze to the target participant had some externalities in terms of affecting the other two non-target participants.

Discussion

Study 2A demonstrates a causal relationship between leader eye gaze, psychological safety, and participation, and provides suggestive evidence for an effect of leader eye gaze on voice. The relationship between leader gaze and voice was mediated by (increased) psychological safety and (decreased) ostracism.

Study 2A also provides evidence that gaze may be more impactful for different people. Specifically, the extent to which the gaze recipient was extraverted significantly moderated the effect of condition on participation, and marginally moderated the effect of condition on ostracism and voice, suggesting that introverted individuals may benefit more from receiving leader eye gaze in conversations. We return to this in Study 2B with a pre-registered analysis.

Furthermore, a group analysis revealed that the high-gaze or low-gaze manipulation, although aimed at one target participant in each group, also impacted the other two participants who were not the target of the gaze manipulations. This implies that in virtual contexts, averting gaze away from the screen may have implications for all group members, even if gaze is only averted when one specific person in talking.

An important limitation of this study is that it was conducted over Zoom, where gaze cannot be directed at a specific person. We addressed this by instructing the confederate to avert their gaze whenever the target person was speaking, thus directing their gaze manipulation at the target participant. However, for a more direct manipulation of eye gaze and averted gaze we replicated this study in a face-to-face context in Study 2B.

Study 2B: Experimentally Manipulating Gaze in Face-to-Face Conversations

Study 2B is an in-person replication of Study 2A (which was conducted over Zoom). We again used a confederate design to manipulate leader eye gaze during live meetings. In this study, we recruited groups of 3 or 4 people to engage in a group meeting in-person, and each group included a confederate, who was always assigned to be the group leader. The confederate either gave continuous gaze to the target individual (high-gaze condition) or averted their gaze away from, the target individual (low-gaze condition).

In this study, we again examine the causal relationship between eye gaze and psychological safety, ostracism (exclusion), participation, and voice. We additionally test Hypotheses 3a and 3b, that eye gaze may be particularly empowering for racial minorities and introverts. Lastly, replicating the same design in a face-to-face context allowed us to explore potential differences in how eye gaze may impact recipients in virtual contexts versus in-person contexts.

Method

We preregistered hypotheses and analyses (https://aspredicted.org/249_CKY).

Participants: We pre-registered recruiting 80 groups (each comprised of 2-3 participants and one confederate) to participate in a live, in-person meeting in exchange for a $15.00 payment. We aimed to collect 40 groups per experimental condition. A total of 81 groups (N= 186) completed the meeting and the post-meeting survey, and were included in our analysis (58% female, 39% male, 3% non-binary, average age= 33.86, SD= 15.29). Participants were recruited via a student and community sample in a Behavioral Lab in an urban center, affiliated with a private university in the Northeastern United States.

As in Study 2A, our analyses largely focused on the target participant, which in each group was the participant assigned to the role of "Delivery Manager". This person was the target of the confederate's manipulation to provide either continuous gaze (high-gaze condition) or avert their gaze (low-gaze condition) towards them. A total of 81 participants were assigned to the Delivery Manager role (64% female, average age= 33.01, SD= 14.68).

Design and Procedure: Participants agreed to participate in a simulated workplace meeting. We told them that they would discuss a case study with other participants in the lab and that they would answer survey questions afterwards.

As in Study 2A, participants were instructed to roleplay a group decision-making task as employees of a retail store, discussing different delivery strategies and reaching an agreement by the end of the conversation. We randomly assigned participants to one of three roles: Delivery Manager, Retail Manager, and Warehouse Manager (See Appendix 1C for full participant instructions). The trained confederate always assumed the role of the group leader, Chief Operations Officer. The leader had authority over the group, and the group was told that the group leader would make the final decision.

Groups were again assigned to one of two between-subjects conditions: a *high-gaze condition*, and a *low-gaze condition.* In the high-gaze condition, the confederate was trained to maintain a continuous gaze with the target participant (the Delivery Manager). In the low-gaze condition, the confederate was trained to avert their gaze from the target participant. The confederate was instructed to clearly convey that they were listening to the target participant, and as such and continued listening to, responding to, and asking follow-up questions to the target participant, even while averting their gaze away from the screen. Thus, only the gaze differed between conditions. The target participant, who was the recipient of the gaze manipulation, is the focus of our analyses.

Measures

We measured Psychological Safety (α= .69), Ostracism (α= .87), Voice (α= .89), and Introversion/ Extraversion (α= .72) using the same scales as Study 2A. We created a measure of participation (airtime) for each participant by aggregating the total number of seconds that each participant spoke during the conversation.

Results

See Table 1.4 for correlations between variables.

Table 1.4: Study 2B Correlation Table

	Psychological Safety	**Ostracism**	Voice	**Participation**
Psychological Safety	**1**			
Ostracism	**-.72*** **	**1**		
Voice	**.66*** **	**-.69*** **	**1**	
Participation	**.20***	**-.24***	**.33****	**1**

Psychological Safety: Analyzing the behavior of the target participant (the Delivery Manager), we found support for Hypothesis 1a: in a t-test, target participants in the high-gaze condition felt more psychologically safe (M = 5.22, SD = .81) compared to participants in the low-gaze condition (M = 4.75, SD = .82), $t(78) = 2.55$, $p = .013$, $d = .57$.

Ostracism: We found evidence for our Hypothesis 1b: in a t-test, target participants in the high-gaze condition indicated that they felt less ostracized (M = 2.56, SD = 1.14) compared to participants in the low-gaze condition, who felt more ostracized (M = 3.17, SD = 1.17), $t(78) = -2.39$, $p = .019$, $d = -.53$.

Participation: We found also evidence for Hypothesis 2a: in a t-test, target participants in the high-gaze condition participated for significantly more overall seconds (M = 375.92, SD = 83.47) compared to participants in the low-gaze condition (M = 311.17, SD = 108.21), $t(78) = 2.24$, $p = .004$, $d = .67$.

Voice: Providing evidence for Hypothesis 2a, we found that target participants who received less gaze felt significantly less able to voice, as indicated in the post-conversation survey. In a t-test, target participants in the high-gaze condition felt significantly more able to voice (M = 5.78, SD = 1.11) compared to participants in the low-gaze condition (M = 5.01, SD = 1.31), $t(78) = 2.81$, $p = .006$, $d = .63$.

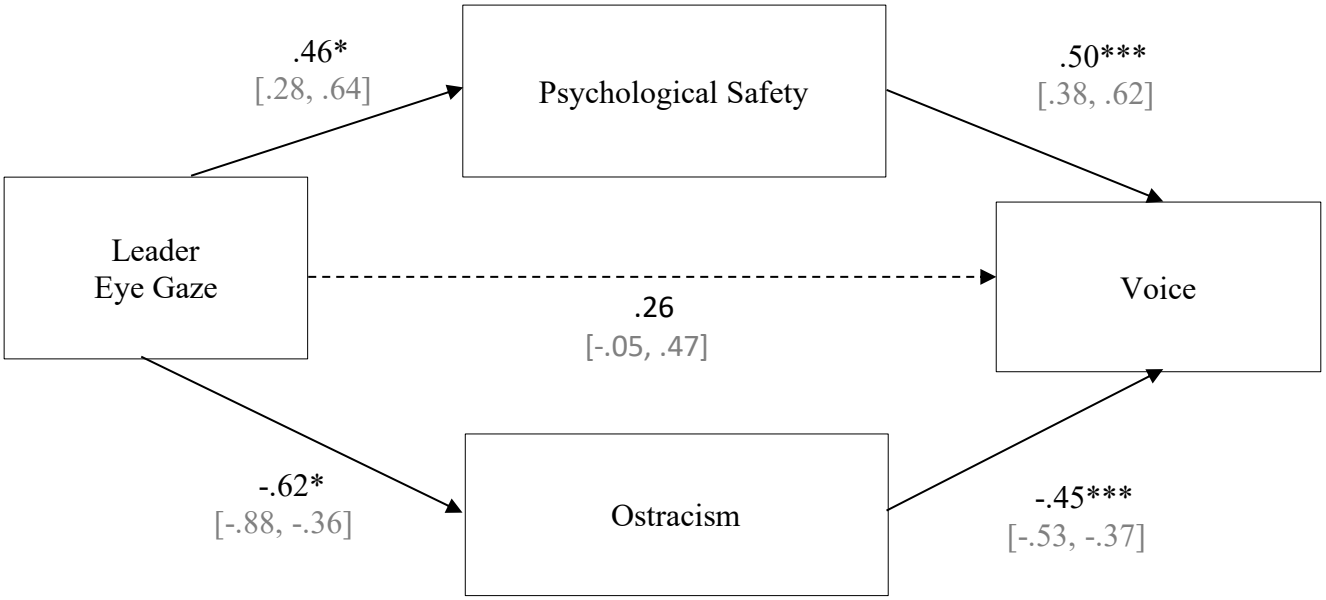
Mediation: We tested whether ostracism or psychological safety mediated the effect of condition on participation (transcript measure) and voice (survey measure) among target participants. Because we had a sample of N < 100, we performed a permutation analysis recommended by Koopman and colleagues for mediation analyses with small sample sizes (Koopman, Howe, Hollenbeck, & Sin, 2015), using their supplemental R code. First, examining voice as a dependent variable, we found that psychological safety significantly mediated the effect of condition on voice [.10, .81]. In a follow-up bootstrap mediation model (10,000 iterations, lavaan package, R) a significant indirect effect emerged for psychological safety (*indirect effect* = .43, 95% CI [.12, .81], p = .009) as a mediator of condition on voice. Turning to ostracism, we found in a permutation analysis that ostracism significantly mediated the effect of condition on voice [-.83, -.08]. In a follow-up bootstrap mediation model (10,000 iterations, using the lavaan package in R), a significant indirect effect emerged for ostracism (*indirect effect* = .43 [.08, .81], p =.019) as a mediator of condition on voice. Thus, we found support for Hypothesis 2b.

Turning to participation as a dependent variable, and psychological safety as a mediator, we found in a permutation analysis that psychological safety did not mediate the effect of condition on participation [-5.26, 24.02], and in a follow-up bootstrap mediation model (10,000 iterations), no significant effect emerged for psychological safety as a mediator on participation (p = .23). We found in a permutation analysis that ostracism did not significantly mediate the effect of condition on participation [-26.07, 1.80]. In a follow-up bootstrap mediation model (10,000 iterations), a marginal indirect effect emerged for ostracism (*indirect effect* = .008 [-.002, .22], p =. 080).

Results from a parallel mediation analysis using the lavaan package in R, including psychological safety and ostracism as parallel mediators and voice as an outcome variable indicated that both psychological safety and ostracism significantly mediate the effect of condition

on voice. As depicted in Figure 1.2b, the experimental condition predicts psychological safety (a1 = .46, p = .010), which is related to increased voice (b1 = .50, p < .001). Experimental condition significantly predicts ostracism (a2 = -.62, p = .016), which is significantly and negatively related to voice (b2 = -.45, p < .001). Moreover, when considering both mediators, there is no direct effect of condition on voice (c= .25, p = .22).

Figure 1.2b: Study 2B Parallel Mediation Model



Moderation Analyses: Race. In order to test Hypothesis 3a (that the impact of eye gaze on our predicted outcome is stronger for racial minorities), we ran a moderation analysis. We created a group of all participants who identified as "Black or African American", "Latino", "American Indian or Alaska Native", "Native Hawaiian or Pacific Islander," or "Other" (non-included groups were "White" and "Asian"). We ran a model with psychological safety as an outcome, and condition (high- vs. low-gaze), race, and the interaction between condition and race as predictors, and found a significant interaction term, $b = 1.10$, $t = 3.01$, $p = .004$. Decomposing this interaction revealed that racial minorities felt significantly more psychologically safe in the high gaze condition ($M = 5.57$, $SD = 0.65$) compared to the low gaze condition ($M = 4.74$, $SD = .88$), $t(23) = 3.84$, $p < .001$, $d = 1.05$, whereas for non-minority participants there was no difference between ratings of ostracism in the low gaze versus high gaze conditions ($p = .36$).

Turning to ostracism, we ran a pre-registered moderation analysis that predicted ostracism as an outcome variable, with condition (high- vs. low-gaze), race, and the interaction between condition and race as predictors, and found a significant interaction term, $b = -1.60$, $t = -3.00$, $p = .004$. Decomposing this interaction revealed that racial minorities felt significantly more ostracized in the low gaze condition ($M = 3.27$, $SD = 1.23$) compared to the high gaze condition ($M = 2.15$, $SD = .82$), $t(23) = 3.86$, $p < .001$, $d = 1.06$, whereas for non-minority participants there was no difference between ratings of ostracism in the low gaze versus high gaze conditions ($p = .93$).

Turning to our outcome variables, voice and participation, we found that there was a non-significant interaction between condition and race as predictors on voice, $b = .88$, $t = 1.53$, $p = .13$. Further analysis revealed that racial minorities felt significantly more able to voice in the high gaze condition ($M = 6.11$, $SD = 0.68$) compared to the low gaze condition ($M = 5.09$, $SD = 1.28$), $t(23) = 3.58$, $p < .001$, $d = .98$, whereas for non-minority participants there was no difference between

ratings of ostracism in the low gaze versus high gaze conditions ($p = .73$). There was no significant moderation for introversion on seconds of participation, $b = 17.93$, $t = .37$, $p = .71$.

Taken together, this pattern of results suggests that gaze may be particularly impactful for more racial minorities, supporting Hypothesis 3a.

Moderation Analyses: Introversion/ Extraversion. To test Hypothesis 3b, that the effect of gaze on psychological safety is stronger for introverts compared to extraverts, we ran a pre-registered analysis with ostracism as an outcome variable, and condition (high- vs. low-gaze), introversion, and the interaction between condition and introversion as predictors. We found a significant interaction term, $b = 1.21$, $t = 2.44$, $p = .017$. Decomposing this using a median split analysis revealed that more introverted target participants felt significantly more ostracized in the low gaze condition ($M = 3.73$, $SD = 1.13$) compared to the high gaze condition ($M = 2.55$, $SD = 1.31$), $t(41) = 3.16$, $p = .003$, d = .96, whereas for more extraverted target participants there no difference between ratings of ostracism among participants in the low gaze versus high gaze conditions ($p = .93$).

We also ran a pre-registered regression with psychological safety as an outcome variable, and condition (high- vs. low-gaze), introversion, and the interaction between condition and introversion as predictors. We found a marginally significant interaction term, $b = .66$, $t = 1.85$, $p = .069$. Decomposing this using a median split analysis revealed that more introverted target participants felt significantly more able to voice in the high-gaze condition ($M = 5.26$, $SD = 0.93$) compared to the low-gaze conditions ($M = 4.48$, $SD = 0.73$), $t(41) = 3.05$, $p = .004$, d = .93, whereas for more extraverted target participants there no difference between ratings of ostracism among participants in the low gaze versus high gaze conditions ($p = .68$).

Turning to our outcome variables, we found that introversion significantly moderated the effect of condition on voice, $b = 1.12$, $t = 2.09$, $p = .040$. Decomposing this using a median split analysis revealed that more introverted target participants felt significantly more psychologically safe in the high-gaze condition ($M = 5.82$, $SD = 1.17$) compared to the low-gaze conditions ($M = 4.55$, $SD = 1.37$), $t(41) = 3.27$, $p = .002$, d = 1.00, whereas for more extraverted target participants there no difference between ratings of ostracism among participants in the low gaze versus high gaze conditions ($p = .66$). There was no significant moderation for introversion on seconds of participation, $b = 14.21$, $t = .32$, $p = .75$.

Taken together, this provides additional evidence that gaze may be particularly impactful for more introverted individuals: leading to decreased ostracism, increased safety, and increased ability to voice, whereas the effects of gaze are less impactful for more extraverted participants. These results support Hypothesis 3b.

Group Analyses: Thus far, we have considered the effects of being in the high-gaze or low-gaze condition for the target participant, who was the subject of our manipulation. However, there were two other people in this group, which allowed us to explore additional questions about how the eye gaze manipulation affects the other people in the group.

We analyzed the effects of the gaze manipulation on all non-leader group members. In a mixed-effects regression model with a random effect controlling for group (lme4 package, lmer function, R), since this set of analyses includes multiple participants per group, we found that there was no significant effect of the manipulation on the non-target members of the group. Participants in the low-gaze condition did not feel more ostracized compared to participants in the high-gaze condition ($b = .19$, $t = 1.22$, $p = .23$), did not feel less psychologically safe compared to participants in the high-gaze condition ($b = -.11$, $t = -.76$, $p = .45$), and did not feel less able to voice compared

to participants in the high-gaze condition (b = -.30, t = -1.56, p = .12). Furthermore, there was no significant difference in speaking time (participation) between non-target participants in the low-gaze condition and the high gaze condition (b = -2.32, t = -.14, p = .88).

These findings contrast with our findings from Study 2A, where we found that the gaze manipulation on Zoom had a negative effect on the non-target participants, who felt more ostracized and less psychologically safe in the low-gaze condition compared to the high-gaze condition.

Discussion

Study 2B replicated the findings of Study 2A, lending further evidence of the causal relationship between leader eye gaze and non-leader feelings (psychological safety, ostracism) and behavior (participation, voice) in response. The relationship between leader gaze and voice was mediated by greater feelings of psychological safety and lower feelings of ostracism.

Study 2B provides further evidence that gaze may be more impactful for different people. Replicating patterns in Study 2A, we find here that the gaze manipulation was more impactful for introverted participants, making them feel more psychologically safe, less ostracized, and more able to voice, while the manipulation had a lesser or no effect for more extraverted participants. We also found that gaze was more impactful for racial minorities; the manipulation increased psychological safety, decreased their feelings of ostracism, and increased the ability to voice for racial minorities, while the manipulation had a lesser or no effect on racial majority group members.

Although the manipulation of the confederate averting their gaze from the target participant negatively impacted the target participant, this manipulation did not impact the other members of the group. This contrasts with our findings in Study 2A, where the averted gaze manipulation

negatively impacted the target participant as well as the non-target members of the group. This suggests an important difference between online and in-person meetings: in online meetings, averted gaze seems to have a dispersed effect, impacting everybody in the meeting, even if averted gaze occurs systematically when one specific group member is speaking. However, it seems that in a face-to-face context, averting gaze away from someone does not negatively affect the rest of the group. This suggests that visual attention from leaders may be even more important in online meetings, given that the impact is likely to be felt by everyone in the meeting.

Importantly, Study 2B had a number of limitations. First, as in Study 2A, gaze was only manipulated for the target participant in each group, limiting our sample size when evaluating individual-level outcomes (N = 81). Although we found evidence that psychological safety and ostracism mediated the effect of condition on voice, we did not find evidence of a mediation on participation (i.e., airtime measured from the transcript), likely due to the limited sample size. Secondly, both an asset and drawback of our three in-person studies (Study 1, Study 2A, and Study 2B) were that they allowed for naturalistic, open-ended conversation. However, these naturalistic designs did not control for or consider the substantive verbal content of the conversation—the words and meaning of what was said during the conversation, nor did it consider other nonverbal cues like facial expressions, trunk lean, hand gesticulation, and on. There may have been aspects of verbal content during these conversations that influenced participants' assessments of psychological safety, ostracism, and speaking up. In Study 3, we use a more tightly controlled experimental design to isolate the effects of leader gaze, where many confounding variables (e.g., verbal, paralinguistic, and other nonverbal content) are stripped away.

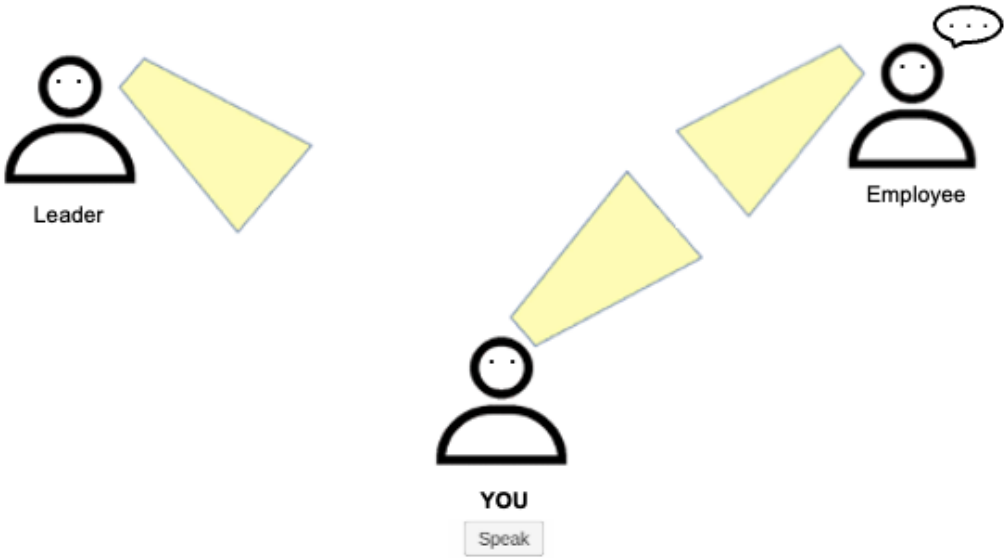
Study 3: Computer-Mediated Gaze Experiment

In Study 3, to control the content of the conversation and isolate the impact of leader eye gaze on group member responses, we developed a simulated conversation environment. This design was inspired by Cyberball, a popular research paradigm in which participants throw a virtual ball to each other. Prior work shows that participants who receive fewer ball tosses compared to the other players feel ostracized (Williams, Cheung, & Choi, 2000; for a meta-analysis, see Hartgerink, Van Beest, Wicherts, & Williams, 2015). We developed an original online gaze simulation in which participants received more or less gaze during a simulated conversation. There was a visual representation of gaze (yellow "vision fields"), which was directed at participants from other players, while participants could choose their own gaze behavior (as in a real conversation). No verbal content was exchanged, but an animated text bubble appeared above the current speaker's head to indicate speaker turns. Whenever participants chose to "speak," the animated speech bubble would appear over their avatar. See Figure 1.4 for screenshots.[9]

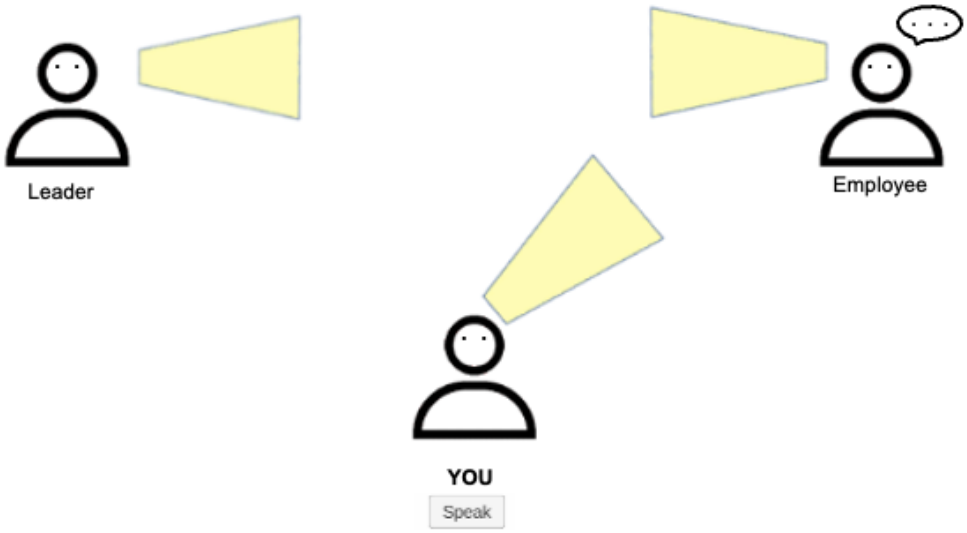
Figure 1.4: Screenshots from the conversation simulation paradigm (Study 3). The yellow shapes represent fields of vision, and the participant can choose to speak at any time by pressing the speak button. The animated speech bubbles indicate who is currently "speaking".

Screenshot from high gaze condition:

[9] Our Gaze Paradigm is online, open-source, and freely available for use by researchers. The paradigm offers a variety of features that are not used in the current paper but may be of interest to other researchers, including varying the group size, gender, race, names, or roles of the group members, or fully customizing who looks where and who speaks when. To customize the paradigm researchers can visit this link: http://gazeparadigm.com/Start.html, where they can click "design your own." After customization, they will receive a custom code, which they can give to participants along with the link to begin the paradigm: http://gazeparadigm.com/Streets3/index.html.



Screenshot from low gaze condition:



Method

We preregistered our hypotheses and analyses (https://aspredicted.org/GG7_HDZ]).

Participants: We pre-registered recruiting 600 participants from Amazon Mechanical Turk, with 200 participants per condition. In total, 578 people passed our attention checks and took our survey (44% female, 2% prefer to self-identify, 54% male; average age= 37.14, *SD*= 12.30).

Design and Procedure: Participants were recruited to participate in a conversation game, purportedly with two other participants (to create groups of 3 people). Participants were told that they would participate in two rounds of a conversation simulation with two other players. In reality, the two other players were computerized agents we had programmed to gaze more or less at the participant based on the experimental condition. The participant was always assigned to be an "employee," while one of the other players was labeled as an "employee," and the third player was labeled as the "leader." Participants were told that after two conversations they would return to a survey and answer questions independently. They were given the game instructions, including that they could choose to "speak" at any time, that it was a game, and as such no words would be exchanged, but speaking would be represented by animated speech bubbles, and that the other players' eye gaze (where they are looking) would be indicated by yellow vision fields (see Figure 1.4 above for screenshots).

We increased realism by keeping participants in a waiting room while they were ostensibly "paired" with other participants. In reality, the other two players in the group were a pre-programmed set of responses built into the game.

Conditions: The "leader" agent was pre-programmed to give more or less gaze to the participant, depending on the experimental condition. There were three experimental conditions, mirroring the Cyberball paradigm (Williams, Cheung, & Choi, 2000). In the *high-gaze condition*, participants received gaze from the leader 75% of the time, intermittently throughout the conversation. In the *medium-gaze condition*, participants received gaze from the leader 50% of the time. Finally, in the *low-gaze condition*, participants received gaze from the leader 25% of the time. When the leader was not looking at the participant, the leader was programmed to split their time between looking at the other employee and looking away (i.e., gazing at neither employee).

That is to say, when the leader was not looking at the target participant, they were not looking at the other employee by default (and excluding the participant); half the time they were not gazing at anybody. The participant could speak at any time, regardless of when or if they received gaze. Lastly, the gaze manipulation came only from the leader; the other employee avatar was programmed to gaze equally at the leader and the participant in all conditions.

Measures

Manipulation Check. We asked a manipulation check question: "During this game, I felt seen by the leader of this group" (1= Strongly Disagree, 7= Strongly Agree).

Behavioral Dependent Variable: After the first conversation finished, we asked participants if they wanted to stay with their same group, or switch to a new group: "You will now move on to Round 2 of the game. Would you like to play with the same three people again, or would you like to be matched with a new group? (Yes, I would like the same group/ No, please match me with a new group)."

Survey Measures: Finally, participants completed post-conversation survey measures to indicate their psychological safety (α= .74), ostracism (α= .92), voice (α= .87), and introversion/ extraversion (α= .78). We used the same measures as in Studies 1 & 2 (See Appendix 2B for items).

Results

See Table 1.5 for correlations between variables.

Table 1.5: Study 3 Correlation Table

	Psychological Safety	**Ostracism**	Voice
Psychological Safety	**1**		
Ostracism	**-.73*** **	**1**	
Voice	**.68*** **	**-.70*** **	**1**

Manipulation Check: A one-way analysis of variance (ANOVA) showed a main effect of the experimental condition on our manipulation check, $F(2, 362) = 22.85$, $p < .001$, $\eta^2 = .11$. Participants in the high-gaze condition felt significantly more seen by the leader than participants in the low-gaze condition, and participants in the medium gaze condition felt significantly more seen by the leader than participants in the low-gaze condition (t's > 5.38, p's < 0.001, d's > .56). However, contrary to our theorizing, there was no significant difference in ratings of feeling seen by the leader between participants in the high gaze and medium gaze conditions, although the measure was directionally higher in the high-gaze condition (M= 5.06, SD =1.46) compared to the medium-gaze condition (M= 4.87, SD =1.46, t(375)=1.31, p =.19, d= .13).

Psychological Safety: A one-way analysis of variance (ANOVA) revealed a main effect of eye gaze on psychological safety, F (2, 363) = 13.15, $p < .001$, $\eta^2 = .06$. In line with our pre-registered hypotheses, participants in the high-gaze condition felt significantly more psychologically safe than participants in the low-gaze condition, and participants in the medium gaze condition felt significantly more psychologically safe than participants in the low-gaze condition (t's > 3.94, p's < 0.001, d's > .41). However, contrary to our pre-registered hypothesis, there was no significant difference in ratings of psychological safety between participants in the high gaze and medium gaze conditions, although psychological safety ratings from participants in the high-gaze condition were directionally higher (M= 4.77, SD =1.00) compared to participants in the medium-gaze condition (M= 4.65, SD =0.92, t(375) =1.16, p =.25, d= .12). See Figure 1.5 below.

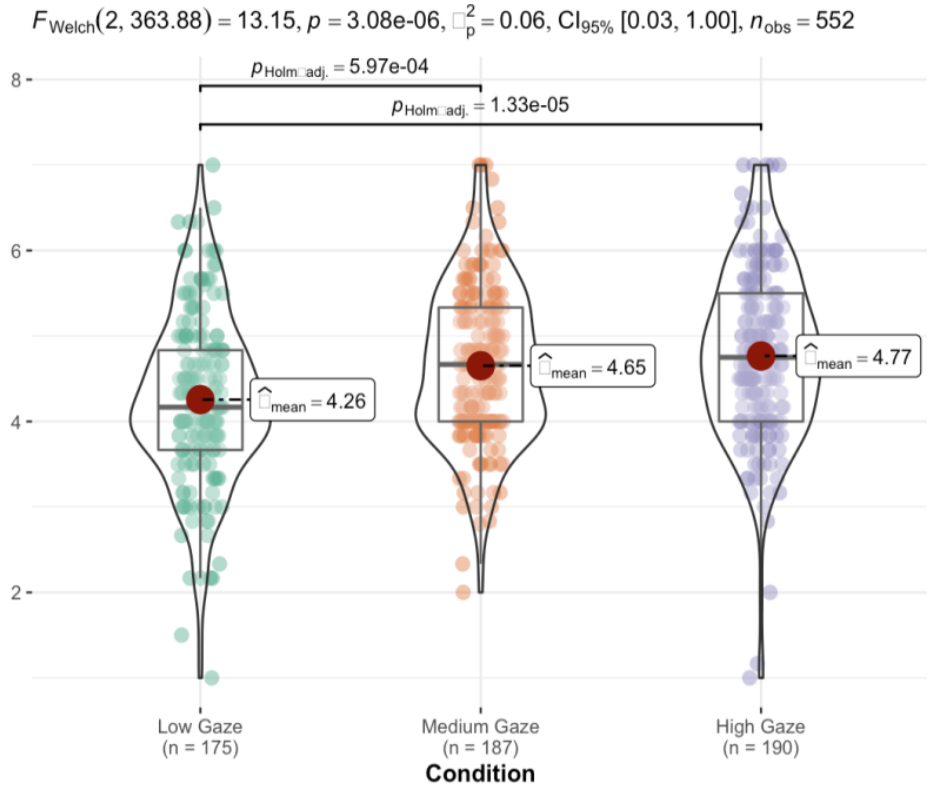
Ostracism: A one-way analysis of variance (ANOVA) showed a main effect of the experimental condition on ostracism, $F(2, 364) = 18.18$, $p < .001$, $\eta^2 = .09$. In line with our pre-registered hypotheses, participants in the high-gaze condition reported lower levels of ostracism

compared to participants in the medium-gaze condition, who in turn had lower levels of ostracism compared to participants in the low-gaze condition (t's < -2.43, p's < .016, d's < -.25). See Figure 1.5 below.

Voice: A one-way analysis of variance (ANOVA) showed a main effect of experimental condition on voice, $F(2, 363) = 9.12$, $p < .001$, $\eta^2 = .04$. In line with our pre-registered hypotheses, participants in the high-gaze condition felt they had significantly more voice than participants in the low-gaze condition, and participants in the medium gaze condition felt they had significantly more voice than participants in the low-gaze condition (t's > 3.06, p's < 0.002, d's > .32). However, contrary to our pre-registered hypothesis, there was no significant difference in ratings of voice between participants in the high gaze and medium gaze conditions, although voice ratings from participants in the high-gaze condition were directionally higher (M= 5.09, SD =1.28) compared to participants in the medium-gaze condition (M= 4.94, SD =1.31, t(375) =1.17, p =.24, d= .12). See Figure 1.5 below.

Figure 1.5: The effect of eye gaze condition on self-reported psychological safety, ostracism, and ability to voice.

Psychological Safety



Ostracism

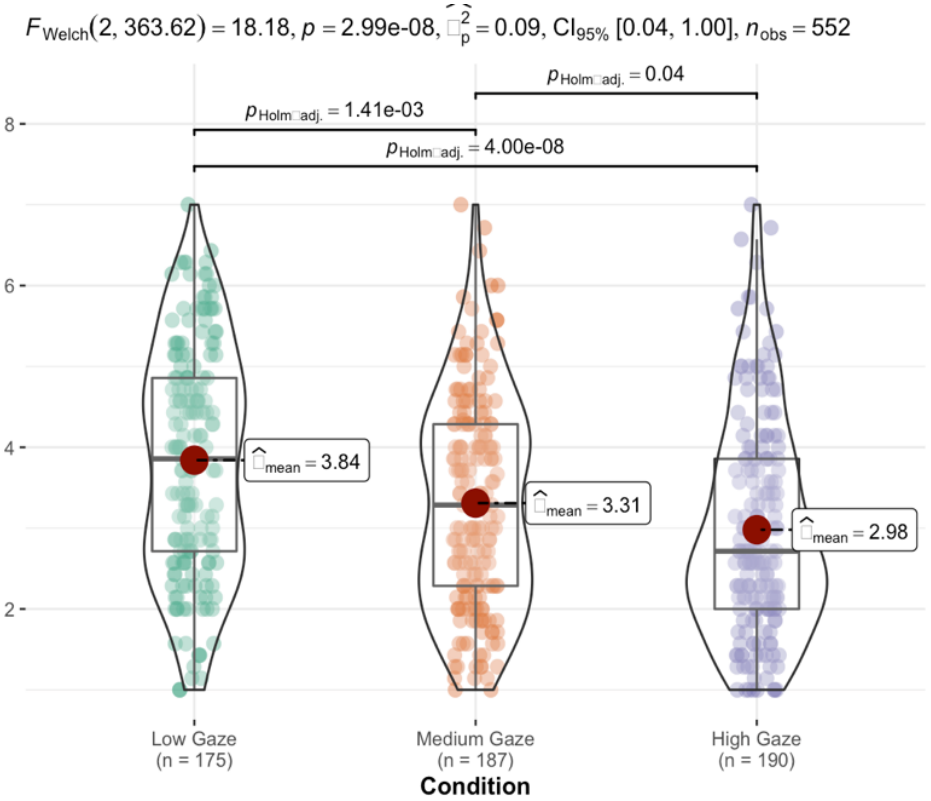
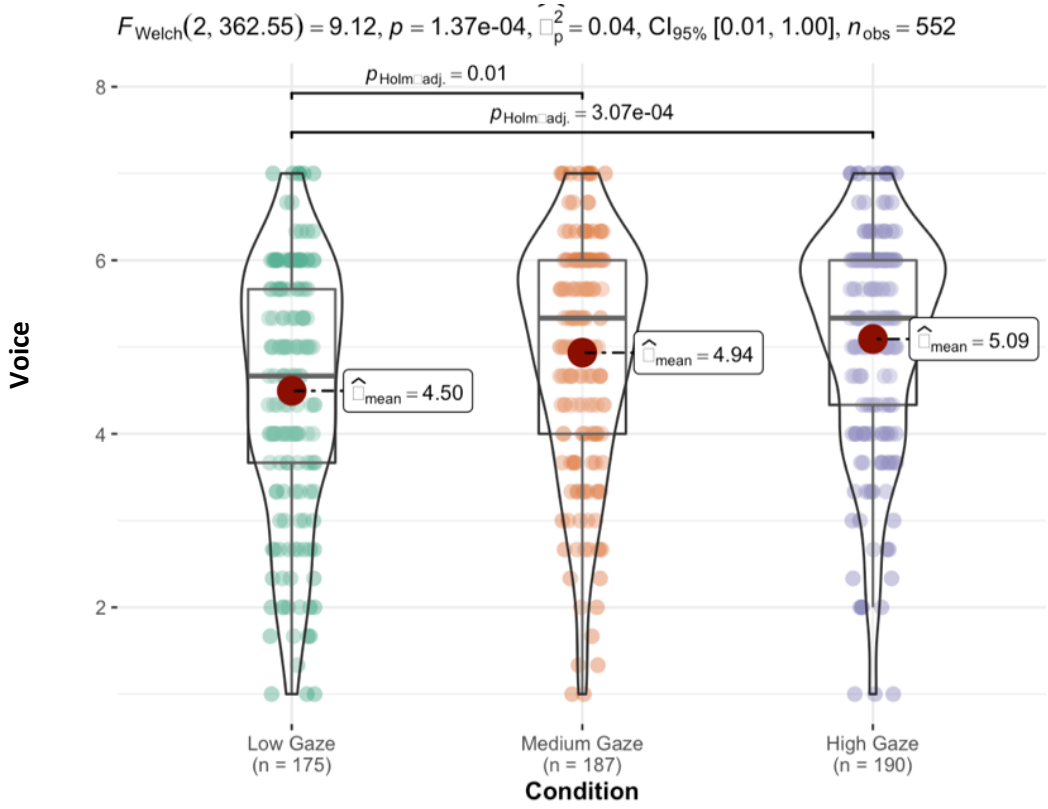


Figure 1.5 (Continued):



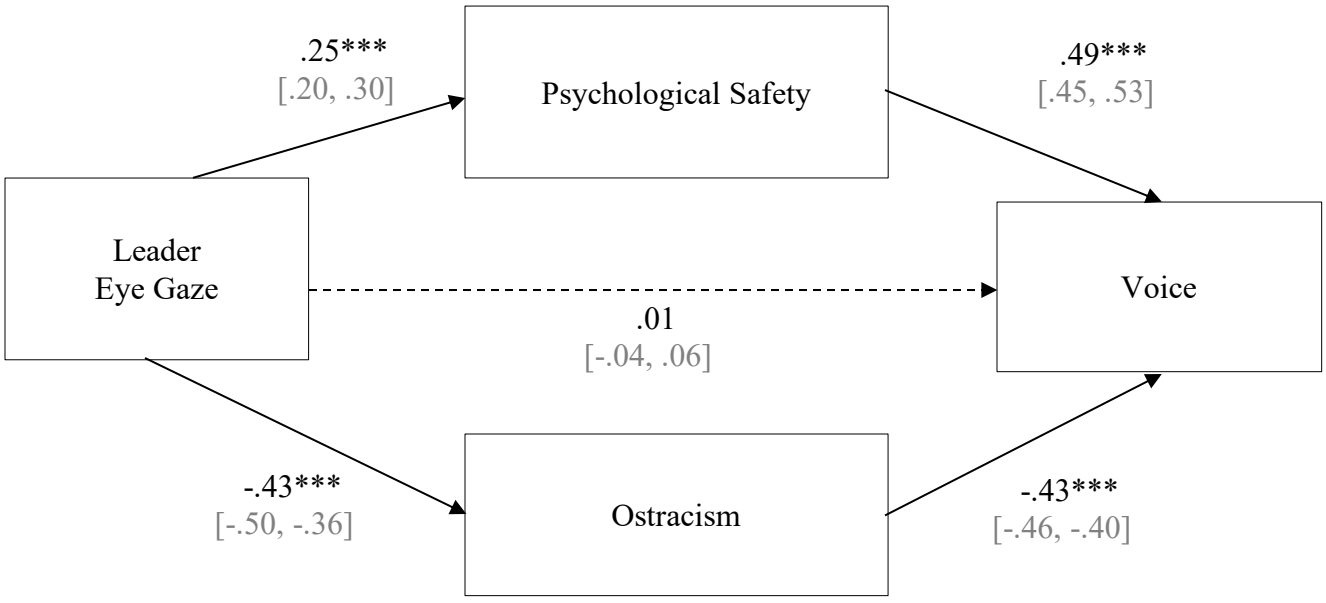
Behavioral Choice: Stay or Switch Groups: More participants in the high-gaze condition chose to stay in the same group for round two (63%), compared to participants in the medium-gaze condition (60%) and participants in the low-gaze condition (49%), $X^2(2, 574) = 7.86$, $p = 0.020$, suggesting that participants in the high-gaze condition and medium-gaze condition felt more included in their groups compared to participants in the low-gaze condition.

Mediation: We tested whether ostracism or psychological safety significantly mediated the effect of condition on voice. Because N>100, we used a bootstrap mediation model (10,000 iterations) that included gaze condition as the independent variable, self-reported measures of ostracism and psychological safety as mediators, and self-rated ability to voice as a dependent variable, a significant indirect effect emerged for both psychological safety (*indirect effect* = .36,

95% CI [.17, .55], $p < .001$) and ostracism (*indirect effect* = -.36, 95% CI [.17, .56], $p < .001$) as mediators of condition on voice.

Results from a parallel mediation analysis using the lavaan R package, including psychological safety and ostracism as parallel mediators, indicated that psychological safety and ostracism in parallel significantly mediate the effect of condition on voice. As depicted in Figure 1.6 below, experimental condition significantly predicts psychological safety (a1 = .25, $p < .001$), which significantly predicts more voice (b1 = .49, $p < .001$). Experimental condition significantly and negatively predicts ostracism (a2 = -.43, $p < .001$), which significantly predicts more voice (b2 = -.43, $p < .001$). Moreover, when considering both mediators, there is no direct effect of condition on voice (c= .01, $p = .79$).

Figure 1.6: Study 3 Parallel Mediation



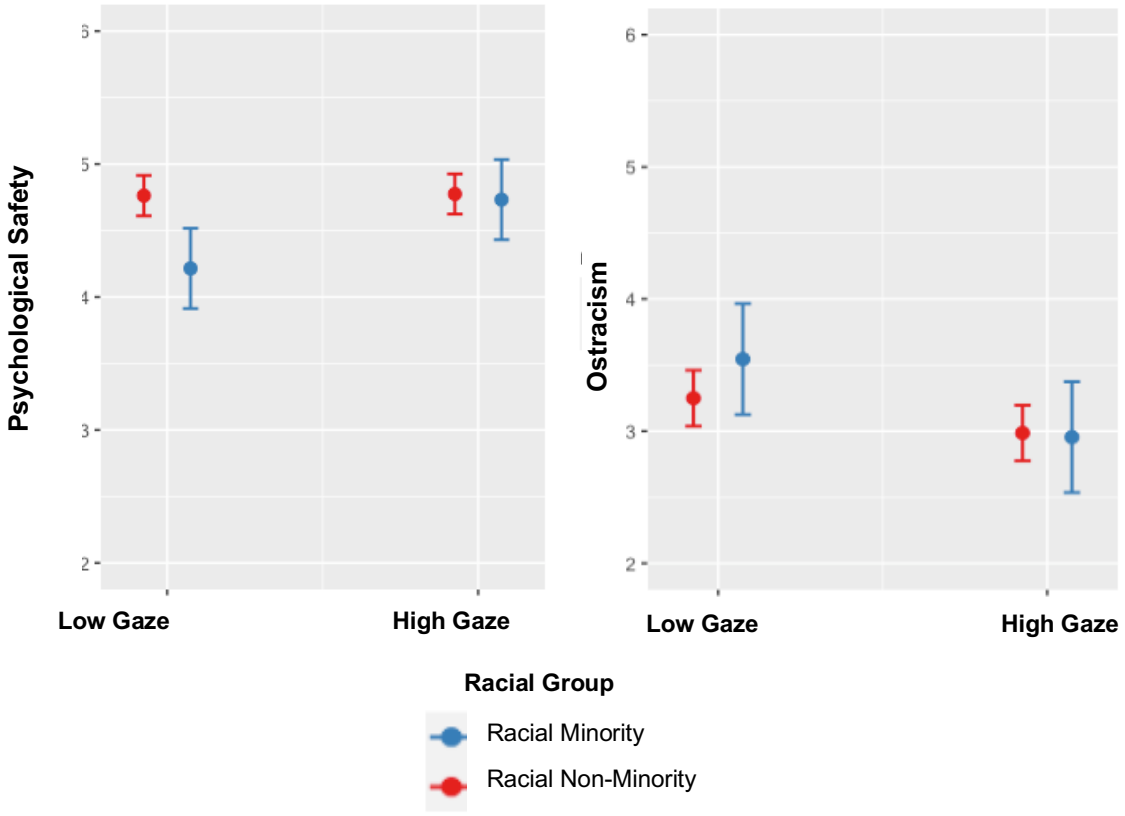
Moderation Analyses: Race. In order to test Hypothesis 3a (that the impact of eye gaze on our predicted outcome is stronger for racial minorities), we ran a moderation analysis. We created a group of all participants who identified as "Black or African American", "Latino", "American Indian or Alaska Native", "Native Hawaiian or Pacific Islander," or "Other" (non-included groups were "White" and "Asian"). There were 117 total participants (out of 576 total) included in the racial minority grouping.

We ran a pre-registered regression analysis that predicted psychological safety, with condition (high- vs. medium-, vs. low-gaze), racial group, and the interaction between condition and racial group as predictors. We found a main effect for condition, $b = .40$, $t = 4.95$, $p < .001$, and a significant interaction term, $b = .51$, $t = 2.92$, p = .004. Decomposing the effect found that racial minorities felt significantly less psychologically safe in the low-gaze condition ($M = 4.21$, $SD = 0.86$) compared to non-minorities ($M = 4.76$, $SD = 0.90$), $t(185) = -3.38$, $p < .001$, $d = -.61$, and this difference was less pronounced in the high-gaze condition, where minorities ($M = 4.73$, $SD = 1.09$) and non-minorities ($M = 4.77$, $SD = 0.98$) felt more similar levels of psychological safety $t(188) = -.23$, $p = .82$, $d = -.04$, (see Figure 1.7a below).

We ran another pre-registered regression analysis that predicted ostracism, with condition (high- vs. medium-, vs. low-gaze), racial group, and the interaction between condition and racial group as predictors. We found a main effect for condition, $b = -.63$, $t = 5.60$, $p < .001$, but did not find a significant interaction term, $b = -.32$, $t = -1.32$, $p = .19$. Decomposing the effect found that racial minorities felt directionally, but not significantly more ostracized in the low-gaze condition ($M = 3.55$, $SD = 1.31$) compared to non-minorities ($M = 3.25$, $SD = 1.31$), $t(185) = 1.24$, $p = .22$, $d = .23$, and this difference was less pronounced in the high-gaze condition, where minorities (M

= 2.95, *SD* = 1.36) and non-minorities (*M* = 2.99, *SD* = 1.32) felt similar levels of ostracism *t*(188) = -.12, *p* = .90, *d* = -.02 (see Figure 1.7a).

Figure 1.7a: Study 3, moderation by race

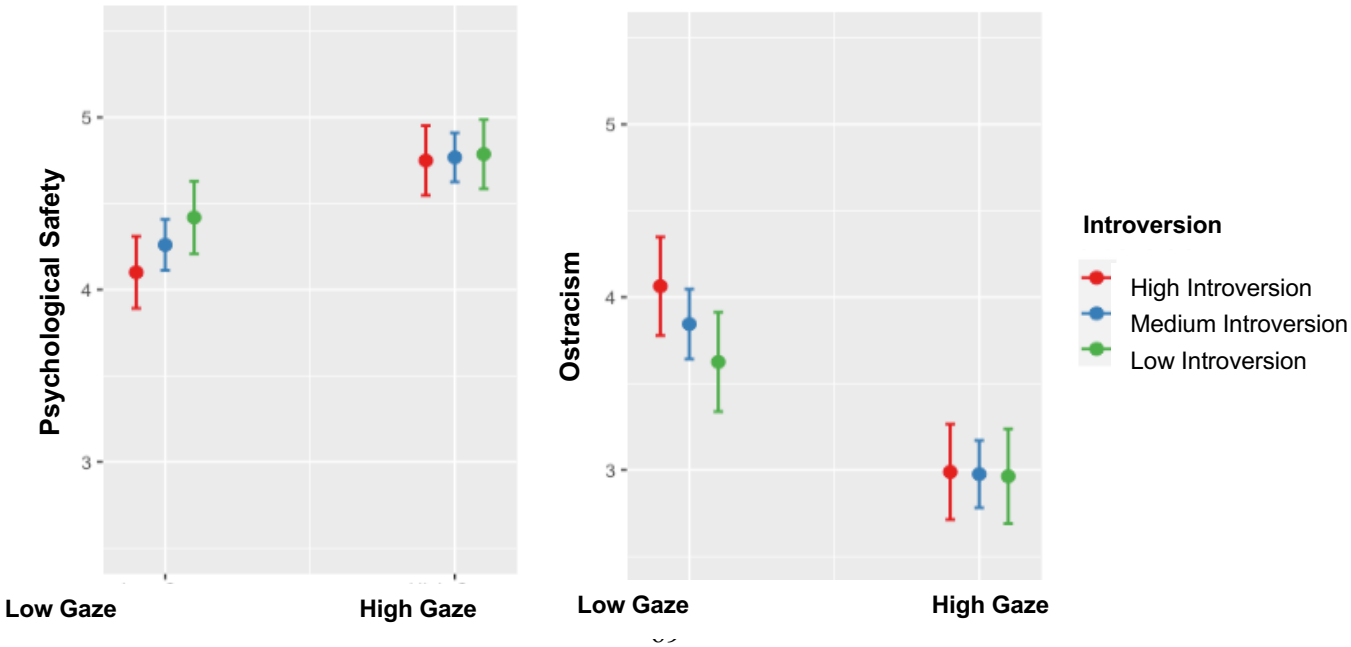


Moderation Analyses: Introversion/ Extraversion. In order to test Hypothesis 3b (that the impact of eye gaze on our predicted outcome is stronger for introverted individuals), we ran a pre-registered regression analysis that predicted psychological safety, with condition (high- vs. medium-, vs. low-gaze), introversion, and the interaction between condition and introversion as predictors. We found a main effect for condition, *b* = .60, *t* = 3.19, p = .002, a marginal main effect for introversion, *b* = .07, *t* = 1.69, p = .092, but did not find a significant interaction term, *b* = .08, *t* = 1.34, *p* = .18. However, decomposing the effect with a median split found that introverted

participants felt directionally but not less psychologically safe in the low-gaze condition ($M = 4.19$, $SD = 1.02$) compared to extraverted participants ($M = 4.35$, $SD = 0.97$), $t(172) = -1.04$, $p = .30$, $d = -.16$, and this difference was less pronounced in the high-gaze condition, where introverts ($M = 4.76$, $SD = 1.09$) and extraverts ($M = 4.78$, $SD = 0.98$) felt more similar levels of psychological safety $t(186) = -.13$, $p = .90$, $d = -.02$ (see Figure 1.7b).

Lastly, we ran a pre-registered regression analysis that predicted ostracism, with condition (high- vs. medium-, vs. low-gaze), introversion, and the interaction between condition and introversion as predictors. We found a main effect for condition, $b = .96$, $t = 3.66$, $p < .001$, but did not find a significant interaction term, $b = .12$, $t = 1.44$, $p = .15$. Decomposing the effect with a median split found that introverted participants felt significantly more ostracized in the low-gaze condition ($M = 4.04$, $SD = 1.37$) compared to extraverted participants ($M = 3.58$, $SD = 1.41$), $t(172) = 2.13$, $p = .035$, $d = .33$, and this difference was less pronounced in the high-gaze condition, where introverts ($M = 3.04$, $SD = 1.42$) and extraverts ($M = 2.90$, $SD = 1.20$) felt more similar levels of ostracism $t(186) = .73$, $p = .47$, $d = .10$ (see Figure 1.7b).

Figure 1.7b: Study 3, moderation by introversion



Discussion

In Study 3, we developed a computer-mediated conversation paradigm to experimentally test the effects of eye gaze in a simulated conversation, controlling for many aspects of conversational content—there was no verbal (words) or paralinguistic (acoustic) content, and very little nonverbal content (only eye gaze). Our findings replicate findings from Studies 2A and 2B that receiving more eye gaze decreases feelings of ostracism, increases psychological safety, and increases the likelihood of voice. We found mediational evidence that psychological safety and ostracism, in parallel, mediate the effect of condition on voice. And, behaviorally, receiving less gaze caused participants to be (directionally) more likely to want to switch to a new group for a subsequent conversation, suggesting that they may have felt less belonging to their initial group.

We also tested Hypothesis 3a and 3b. We found evidence that the effect of gaze on psychological safety was more pronounced for racial minorities, but only directional evidence that the effect of gaze on ostracism was more pronounced for racial minorities. We also found directional evidence that the effect of gaze on psychological safety and ostracism was more pronounced for introverts, compared to extraverts, but these effects did not reach significance.

In this study, we were also able to test the difference between three conditions: high, medium, and low gaze conditions. This allowed us to evaluate: what has a stronger effect—more gaze (relative to a baseline) or less gaze (relative to a baseline). We found the biggest differences between the low- and medium-gaze condition, and often found no differences between the medium- and high-gaze conditions. This shows that the biggest difference comes from giving at least some gaze to someone and moving from a low gaze to medium gaze. Thus, it seems important to establish a baseline of giving any gaze at all, to convey respect and attention to another person.

Overall, we suspect that this study was a conservative test of the effects of eye gaze on conversational dynamics because the conversations were simulated, conversation partners were represented as avatars, and no actual words were exchanged. It is notable that minimal social exclusion in the form of receiving less gaze in a computer game with simulated strangers still led to feelings of ostracism, decreased psychological safety, and less self-rated ability to voice.

General Discussion

Psychological safety is an important ingredient for effective group functioning, especially in creating environments where employees can freely speak up and voice ideas that challenge the status quo (Edmondson, 1999; Detert & Burris, 2007). While the importance of psychological safety is rarely questioned, the conventional advice for leaders largely focuses on creating general impressions (e.g., be more open, more transformational), with little attention paid to the behavioral underpinnings that lead followers to form those impressions. Our studies focused on a specific, non-verbal behavior that signals attention, respect, and expectations of valued contributions, thereby increasing psychological safety: eye gaze. Results across four studies with multiple methodologies, multiple measurement strategies, and different participant populations show consistent effects: receiving more eye gaze from leaders increases psychological safety, decreased ostracism, and increased speaking (both participating generally and engaging in voice specifically).

Importantly, our moderation analyses suggest that increasing visual attention may empower speaking up more for those who may tend to feel more invisible in the workplace and receive less attention and respect at baseline. We find significant evidence that race moderates the effect of gaze on psychological safety, such that the effects of gaze are stronger for

underrepresented minorities (Studies 2B and 3), as well as significant (Study 2B) and suggestive (Study 3) evidence that is particularly effective in reducing ostracism for underrepresented minorities. Turning to introversion, we found significant (Study 2B) and suggestive (Studies 2A and 3) evidence that gaze is especially impactful at increasing psychological safety and decreasing ostracism for introverted participants (whereas extraverted participants generally feel higher psychological safety and less ostracism, and are less impacted by gaze manipulations). Taken as a whole, these patterns seem to suggest that the positive ramifications of eye gaze may be more impactful for those who may feel more invisible and be less likely to speak in the workplace.

Our findings point to interesting group-level externalities. In Study 1, which evaluated in-person conversations with no gaze manipulation, the extent to which the leader's gaze was equitable did not impact individual or group-level psychological safety. However, leaders' overall attentiveness to others (as opposed to looking at their papers, laptops, and so on) did impact group-level psychological safety, as did the leader's attentiveness to the person speaking. When it comes to our studies with a gaze manipulation, in which the leader avoided giving any eye contact at all to one group member, this had a negative impact on the group as a whole in a Zoom meeting (Study 2A), but did not have a significant effect during in-person conversations (Study 2B). Thus, we find that the leader giving attention and respect to somebody else does not negatively impact individual or group psychological safety. However, in online settings, seeing a leader visually ignore another colleague does detract from group psychological safety.

This leads us to another point: our findings demonstrate a difference between Zoom conversations and in-person conversations. In Study 2A, which was on Zoom, although the leader only looked away whenever the target participant spoke, the other two participants felt less psychologically safe and more ostracized. However, in Study 2B, an in-person replication of Study

2A, the leader averting their gaze from one participant did not negatively impact the other people in the group. We suspect that this is because gaze is not dyadic on Zoom: even if you try to avert your gaze away from one person only, everybody in the meeting is able to notice, and may be negatively affected. It may thus be especially important for leaders to maintain visual attention during online meetings, as it has the potential to have a more dispersed effect on everyone in the meeting.

Contributions to Theory

These findings make several fundamental contributions. First, this work identifies eye gaze as an important behavioral strategy for leaders to establish psychological safety. Although the outcomes for teams with high levels of psychological safety is well established (Edmondson, 1999; Baer & Frese, 2003; Nembhard & Edmondson, 2006), relatively few studies have examined how employees and teams come to view a context as psychologically safe in the first place (a recent exception is Coutifaris & Grant, 2021). We offer a new theoretical perspective that explains the specific drivers of psychological safety and subsequent participation and voice behaviors in teams. Rather than focus on what leaders can say to compellingly create a culture of safety (Coutifaris & Grant, 2021), we outline how a non-verbal behavior, irrespective of what is said, can signal the attention, respect, and value that drives feelings of safety and associated behavioral outcomes.

Second, this research shifts scholarly focus from overall gestalt judgments of leader traits associated with psychological safety and speaking up to the moment-to-moment behaviors that cause employees to assess a climate as safe to speak up. Previous research has focused on what types of leadership styles lead to psychologically safe environments, such as authentic leadership (Hsiung, 2012); leader openness (Detert & Burris, 2007); moral leadership (Chan et al., 2013); and empowering leadership (Gao, Janssen, & Shi, 2011). However, this work relies on judgments made

by employees about their perceptions of their leader's style, as opposed to directly observing leaders' actual behaviors, and relying on others' evaluations (as opposed to directly studying leaders) may lead to confounded or tautological links between leadership styles and outcomes (Van Knippenberg & Sitkin, 2013; Carton, 2022; Fischer & Sitkin, 2023). Also, these leadership constructs are multidimensional, which obscures precise causality: with so many sub-components, it's hard to know what exactly is leading employees to feel more psychologically safe (Ashford & Sitkin, 2019; Van Knippenberg & Sitkin, 2013). Rather than focus on who leaders *are* in terms of their traits and styles, we offer a new theoretical perspective by demonstrating what leaders *do* in terms of nonverbal behaviors to signal that an environment is safe, insights that can emerge by examining turn-level conversation data, which allows for more granularity and precision compared to trait-level survey measures.

Third, we find that the effects of eye gaze were moderated by important features of the group members, building on the literature on how individual factors interact with situational factors (i.e., Ten Berge & De Raad, 2001; 2002; Parrigon, Woo, Tay, & Wang, 2017). Specifically, we find suggestive evidence that receiving eye gaze was particularly empowering for introverted individuals and racial minorities, suggesting that eye gaze may be most empowering for those who may feel more invisible at work. This sheds new light on best practices for managing diverse teams by demonstrating a behavioral tactic for the inclusion of racial minorities, and more introverted team members. Past theory suggests that people with lower status may be more responsive to situational factors that increase voice (LePine & Van Dyne, 1998), and our studies lend evidence to that hypothesis.

Fourth, our theory rests at the intersection of a fundamentally dyadic behavior (leader eye gaze can only be directed to one other person at a time) that has broad implications for group-level

feelings of psychological safety. Although we theorize that leader gaze conveys attention and respect, we find that gaze is also tied to contributions that advance group decision-making and hence gaze does not detract from the psychological safety of other group members. Rather, if appropriately directed, gaze has the potential to enhance the safety of those group members speaking and those who are observing those contributions. Such a finding challenges other theories noting that leaders can sometimes create differentiation through their positive attention directed only to some employees (He et al., 2017). This finding moreover brings to the forefront the need to formally examine and theorize the conditions under which leader behavior directed toward one employee can positively versus negatively impact the feelings of other group members.

Limitations and Avenues for Future Research

Our work has limitations that may offer fruitful avenues for future research. First, although we compared high gaze versus low gaze, there is a question of where the leader's gaze is directed if not at the other humans in the group. Is there a difference if an individual's attention is directed toward something else (e.g., a phone, a laptop), versus looking into space as if in thought, or out the window? Many questions remain about the visual information people attend to amidst conversation, and how conversation partners perceive their partners' gaze. Research shows that the negative effects of someone averting their gaze and looking at a mobile phone are pronounced (Nakamura, 2013; 2015), which suggests that averted gaze may be particularly pernicious when attention is averted towards a mobile device or a computer—a cue that does not correlate strongly with actual listening, Collins et al., *working*. Relatedly, if a person gives an attribution or justification for their reduced gaze ("I'm taking notes"; "my camera is on my other monitor"), could that attenuate the negative effects of averted gaze?

Further, feigning interest is a common phenomenon. Future research could examine the object of attentional gaze, whether others are able to tell whether interest is feigned, and how and whether feigned versus sincere attention affects psychological safety and ostracism in online versus face-to-face settings. Emerging research shows that people cannot easily discern between feigned and genuine attention from others, suggesting that merely feigning attention may be enough to lead others to feel included in online settings (Collins et al., *working*).

Third, future work could look into eye gaze, psychological safety, and speaking up in more naturalistic settings with real work teams who have worked together and who know each other, as group familiarity could alter the effects that we observed. Prior research on group familiarity suggests that groups who know each other well are more productive (Goodman & Leyden, 1991) and more likely to share information (Gruenfeld, Mannix, Williams, & Neale, 1996). Though our experiments suggest that leader eye gaze offers a highly practical tool for increasing group engagement and participation, more data is needed to understand how this phenomenon would play out in established groups.

Fourth, we hypothesize that leader gaze impacts group members because it conveys attention and respect. Our theorizing would predict that other verbal, nonverbal, or paralinguistic cues that convey attention and respect would similarly increase safety and speaking. There is a vast array of behaviors people can enact amidst conversation, from nonverbal cues like nodding and smiling, to paralinguistic cues like laughing and back-channeling (providing verbal affirmations such as "yeah" and "uh-huh"), to verbal cues, such as acknowledging and summarizing what somebody has just said (i.e., paraphrasing or demonstrating conversational uptake, Demszky et al., 2021). Backchannels in particular function as a signal of engagement and attention in a conversation without adding conversational content (Holroyd, Rich, Sidner, & Ponsler, 2011), and may thus

perform a similar role, conveying attention and respect. Moreover, different cues may become more important as group conversations and meetings increase in size, and each individual spends less time speaking, on average (Cooney et al., 2020).

Relatedly, because giving direct eye contact to another person is impossible in online settings, further research could examine other ways to convey attention and respect to one specific person in an online group meeting. Conversational behaviors that are possible online include sending a private chat message to someone, sending emojis, displaying exaggerated facial expressions, or nodding while someone is speaking. Emerging research shows that people who receive emojis feel more confident after the conversation (Fadhil et al., 2018) indicating that emojis, like actual eye gaze, may also signal attention and respect.

Lastly, the finding that gaze is particularly impactful for minorities demonstrates a behavioral tactic for the inclusion of minorities and their voice. Future research could build upon this finding, to test if eye gaze, and other leader behaviors that convey attention and respect, could be a way to unleash some of the benefits that diversity promises.

Practical Implications

Our findings are of high practical relevance. Our research calls on leaders to modulate their visual attention (eye gaze) during the course of group discussions.

In terms of prescriptions for leader behavior, our studies suggest that it is important for leaders to gaze at employees, instead of looking down at their laptops, papers, phones, and so on. In terms of who leaders should look at, our studies suggest that leaders should aim to maintain a consistent gaze with speakers during a group conversation, and also make sure to make eye contact (even briefly) with group members who aren't speaking, to signal attention and respect.

Moreover, leaders should also make a concerted effort to direct their visual attention to group members who may feel "invisible" at work, especially introverts and racial minorities. We demonstrate that merely making a conscious effort to make eye contact with employees who tend to be more silent in meetings may be particularly empowering to those employees.

Our research also has important implications for remote teams. Study 2A showed that receiving gaze of the group leader (a confederate) during a Zoom meeting made participants feel more psychologically safe and less ostracized, and led to them speaking more. Although people are not able to tell which individual face each person is looking at in a Zoom meeting, our study demonstrated that people can tell if someone is paying attention while people are speaking, or if they are looking away or at something off-screen.

Conclusion

Although the benefits of increased psychological safety and speaking up in the workplace are well-documented, relatively little research has examined how to increase psychological safety. The studies in this paper show that simply changing patterns of visual attention—a costless, time-neutral nonverbal cue—can convey attention and respect, and thus increase feelings of psychological safety and decrease feelings of ostracism, which leads to more participation and voice. In order for employees to feel seen and empowered, leaders should make sure to look at them.

CHAPTER 2. HOW VERBAL, NONVERBAL, AND PARALINGUISTIC CUES MISLEAD INTERPERSONAL INFERENCES

Summary: In interpersonal interactions, people convey three types of information: verbal (their words), nonverbal (what they look like), and paralinguistic (how they sound). Research on interpersonal perception in the domain of job interviews suggests that interactional interviews do not help observers predict subsequent job performance. The current results both replicate and challenge these findings. Across a factorial exploration of verbal, nonverbal, and paralinguistic information (N = 4,143), although observers believed that observing verbal content (words) from interviews would help them make the most accurate predictions about interviewees' subsequent performance, observing verbal information from the interviews did not improve predictive accuracy on any subsequent performance tasks. Further, no type of interview information (verbal, nonverbal, or paralinguistic) led observers to make more accurate predictions of performance on objectively-evaluated tasks (e.g., math problems). However, observing the nonverbal and paralinguistic cues conveyed by interviewees improved observers' predictive accuracy on subjectively-evaluated tasks (e.g., the interviewee's likability and empathy, as rated by other people). This work replicates prior findings that interviews do not help predict GPAs or test scores, but suggests that nonverbal and paralinguistic cues may convey important information about people's interpersonal skills.

Keywords: interviews, conversation, verbal, nonverbal, prosody, paralanguage, prediction, estimation, hiring, job performance

How Verbal, Nonverbal, and Paralinguistic Cues Shape Interpersonal Inferences

In 1979, fifty University of Houston Medical School applicants were interviewed and rejected—but were later admitted when the Texas legislature mandated a larger class size. Years later, those who had been initially rejected based on their interviews performed just as well in school as their classmates who "passed" the interviews (Devaul et al., 1987), leading to the conclusion that interviews do not provide useful information. This natural experiment has been replicated in research showing that interviews do not predict subsequent performance on objectively-evaluated tasks such as standardized test scores and GPAs (Dana et al., 2013; Dubovsky et al., 2008; Kausel et al., 2016; Conrad, 1988). Although even small samples of information allow people to make quick judgments about the person's gender, the size of the person's social network, and how much they like someone (e.g., Ambady et al., 2000; Mobasseri et al., 2022), this information may not be sufficient to predict task-specific abilities.

Why don't interviews provide insight into future behavior? One possibility is that interviews do not provide any diagnostic information. Any conversation can only convey a small sample of someone's personality, abilities, habits, and skills, and interviewees can easily hide their weaknesses and emphasize their strengths (Yeomans et al., 2022). When predicting how well people might do at a job, then, talking to them may not help at all.

Another possibility is that interpersonal interactions do contain useful information, but observers miss it. In conversation, speakers convey three distinct types of information: verbal content (their words), nonverbal content (what they look like), and paralinguistic content (how they sound). These three types of conversational content exert separable influence on interpersonal perception, inference, and behavior (e.g., Berger, Rocklage, & Packard 2022; Camp, Voight, & Jurafsky, & Eberhardt, 2021; Carney et al., 2015; Jeong et al., 2019, Packard & Berger, 2021;

Locke & Anderson, 2015, Ten Brinke et al., 2012; Tsay, 2013; 2014; Yeomans et al., 2021). It is possible that when forming interpersonal perceptions of others, observers overweight certain content which does not predict later performance (e.g., the words people say), while underweighting other content that does (e.g., their nonverbal and paralinguistic content). People could attend to the wrong information altogether, or could attend to useful information, but use it to predict performance on the wrong tasks.

Existing work has documented the inability to predict objective outcomes like test scores based on interpersonal interviews (Dana et al., 2013; Dubovsky et al., 2008; Kausel et al., 2016; Conrad, 1988). However, interviews are interpersonal exchanges, and a person's success on an interview is subjectively evaluated by another person. As such, interviews may be better suited to predict performance on tasks that are similarly subjectively-evaluated (e.g., how empathetic someone is) as opposed to objective outcomes (e.g., their performance on a standardized test).

We tested these possibilities by conducting conversational interviews with participants and asking them to subsequently complete a variety of subjectively-evaluated tasks (e.g., conversing with a group) and objectively-evaluated tasks (e.g., standardized test questions) to establish a ground-truth link between interview behavior and subsequent performance on a broad range of tasks (See Table 2.1 below for a summary of the performance tasks in Study 1A). We then recruited a separate sample of participants to view recordings of the interviews and predict the interviewees' subsequent performance, manipulating the type(s) of conversational content could see and hear (see Table 2.2 below for a summary of experimental conditions) in Study 1B.

Table 2.1: Performance Tasks and Evaluations

	Performance Task	Mode of Evaluation
1	Group conversation	Subjective (likeability rated by other participants)
2	Case summary	Subjective (quality rated by third-party observers)
3	Termination letter	Subjective (quality rated by third-party observers)
4	LSAT logic questions	Objective
5	GRE math questions	Objective
6	Bonus distribution	Objective
7	Slider persistence	Objective

Table 2.2: Experimental Conditions: A Factorial Treatment of Verbal, Nonverbal, and Paralinguistic Conversational Information

	Condition	Type(s) of Conversational Interview Content Observed
0	Average Guess	None (predict average performance)
1	Video with full audio	Verbal, Nonverbal, Paralinguistic
2	Video with filtered audio	Nonverbal, Paralinguistic
3	Video muted	Nonverbal
4	Video muted, with subtitles	Verbal, Nonverbal
5	Full audio	Verbal, Paralinguistic
6	Filtered audio	Paralinguistic
7	Transcript text	Verbal

If conversations do provide useful information that people miss or misuse, we expect to see differences in prediction accuracy based on what content types are available to observers (verbal v. nonverbal v. paralinguistic). If conversations don't provide useful information for

predicting future behavior, we should see that none of these conditions leads to increased predictive accuracy.

This experiment is, to our knowledge, the first to isolate and compare all permutations of verbal, nonverbal, and paralinguistic content, not just in an interview context, but in any context. Although prior work has examined naturally-occurring cue differences, often by comparing different modalities of communication (e.g. video with sound vs. without sound, Tsay 2013; 2014; speaking on the phone vs. in person, Schroeder, Lyons, & Epley, 2021; speaking on Zoom vs. in person, Boland et al., 2021; reading text vs. listening to audio, Berger, Rocklage, & Packard, 2022; Schroeder, Kardas, & Epley, 2017; media richness theory, Daft & Lengel, 1986), this is the first empirical exploration of verbal, nonverbal, and paralinguistic content in a simultaneous factorial design.

Open Science Disclosure: We report how we determined our sample sizes, all data exclusions, all manipulations, and all measures. Data, analysis code, stimuli, and preregistrations for all studies are available on OSF: https://osf.io/3qtyp/.

Study 1A: Recording Conversations and Measuring Performance

We recruited participants to complete a conversational job interview and seven subsequent performance tasks. This approach allowed us to examine potential ground-truth links between conversation content (verbal, nonverbal, and paralinguistic information) and subsequent task performance (on subjective-evaluated v. objective-evaluated tasks).

Method:

Participants: To create naturalistic interview stimuli, we recruited 105 participants, in 35 groups of three, from across the United States (M_{Age} = 33.10, *SD* = 14.32; 65% female) to participate in an online experiment in exchange for a $12 electronic gift card.

Part 1: Conversational Interview. As soon as three participants had joined the online meeting, the lead researcher sent one participant and one researcher to each of the three Zoom breakout rooms. The participants were told that they would be interviewing for a project manager position, which involved "coordinating between many different teams, communicating deadlines and priorities, and executing tasks along with a team." Researchers asked participants a series of 10 common interview questions, including "Tell me about a time that you tried something and failed" (see Appendix 2A). These questions were selected for being common interview questions, and this strongly resembled an interview; people shared interesting and relevant anecdotes, were very engaged in the process, and often thanked us for providing them an opportunity to practice interviewing. At the end of 10 minutes, the breakout rooms closed and all participants and researchers returned to the main room for Part 2.

Part 2: Subjectively-Evaluated Performance Tasks. Once all participants and researchers had re-entered the main meeting room, two of the researchers left. The lead researcher remained and instructed participants that they would be engaging in a 10-minute conversation with the two other participants, and they could talk about whatever they liked, with the goal to get to know each other. These 3 participants were strangers prior to the experiment. The lead researcher then turned off their video and instructed all participants to enable the "Hide non-video participants" option on Zoom so that they would only see each other's faces. At the end of the group conversation, participants moved on to complete other tasks and survey measures on their own.

Group conversation: Likeability. We asked participants to report how much they liked the other two participants in their group conversation using 7-point Likert scales with endpoints labeled *1 (not very much)* and *7 (very much)*.

Writing Task 1: Case Summary. Next, participants completed two writing tasks. The first writing task required them to read a case study and prepare an executive summary. Participants were told to "to highlight the major points and issues included, as well as detail the potential solutions that are under consideration." Participants were given 10 minutes to read the case and write their summary.

Writing Task 2: Termination Letter. The second writing task required participants to read a workplace scenario and prepare a termination letter. Participants were told that "the letter should confirm the employee's end of employment. It could also contain the reasons for termination and any necessary information that the individual should know to move forward." Participants were given 10 minutes for this task.

Three research assistants later evaluated participants' case summaries and termination letters, rating them from 1 (*very poor quality*) to 7 (*excellent quality*). The intraclass correlation of these ratings was .82, (95% CI = [.70, .90]) for the termination letters and .81 (95% CI = [.69, .89]) for the case summaries.

Part 3: Objectively-Evaluated Performance Tasks. Last, participants completed four tasks that were evaluated objectively.

LSAT Task. Participants had 10 minutes to complete 10 LSAT logic questions. All questions were taken from LSAT practice questions (see Appendix 2A for an example).

GRE Math Task. Participants had five minutes to complete five GRE math questions. All questions were taken from GRE practice questions (see Appendix 2A for an example).

Bonus Distribution Task. Participants were then told they would be entered into a lottery for a bonus payment of $12. They had the option of distributing this bonus between themselves and the other participants in their conversation group; they could keep the entire bonus for themselves, give it away, or choose any option in between.

Slider Persistence Task. Lastly, participants were asked to move 50 sliders to the exact number displayed on the left of each slider. Participants were asked to complete as many sliders as possible but were allowed to advance from the page at any point.

Study 1B: Observers Predict Performance

To test human prediction accuracy, we asked a separate sample of participants to observe the interviews and forecast subsequent performance. In a full factorial experimental design, we created different versions of each interview recording, which contained all possible combinations of verbal, nonverbal, and paralinguistic content (see Table 2.1 above), and randomly assigned participants to view one kind of content. For example, participants in Condition 4 saw a muted video of the interview with subtitles, which conveyed verbal and nonverbal, but not paralinguistic, information. After observing one stimulus, participants predicted how the interviewee had performed on each of the subsequent tasks and activities, with full information about how their performance on each task had been evaluated (subjectively or objectively). This approach allowed us to examine which types of conversational input information led people to make more (and less) accurate performance predictions. We also included an average-guess condition in which participants merely guessed the average performance on each task without seeing an interview at all.

Method

We preregistered our recruitment and analysis plan: https://aspredicted.org/G11_1FF.

Participants: We planned to recruit 3,940 participants. One participant's video from Study 1A was unusable due to a computer error, so this video was not included in Study 1B. Taken together, we had 104 interview recordings, and we created seven versions of each recording, resulting in 728 total stimuli. Each stimulus was to be viewed at least 5 times by 5 unique participants, requiring 3,640 participants. We recruited 300 additional participants for the average-guess condition, which presented no interview stimuli. In order to participate, participants were required to pass an audio, video, or reading attention check (presented before the interview stimuli were provided). In total, 4,143 adults from Prolific agreed to participate in exchange for $2.10 ($M_{age}$ = 27.14 years, SD = 9.33, 74% female).

Design: We randomly assigned participants to one of eight conditions (seven stimuli conditions and one average-guess no-information condition). Based on their condition, participants either read a transcript of an interview, listened to an audio clip of an interview (with verbal content included or obscured), watched a video of an interview (with verbal content included or obscured), or made predictions with no interview information (see Table 2.1 above). After viewing or listening to the interview, participants were told that the interviewee completed several tasks after the interview. Participants read about each task, including an example question and a description of how each task was evaluated, and then predicted the interviewee's performance.

Liking Task. Participants predicted how much each interviewee was liked by the two other participants using a 7-point Likert scale with endpoints labeled *1* (*not very much*) and *7* (*very much*).

Writing Task 1: Case. Participants predicted the grade that each participant received on the case task using a 7-point Likert scale with endpoints labeled *1* (*very low grade*) and *7* (*very high grade*).

Writing Task 2: Termination letter. Participants predicted the grade that each participant received on the termination letter task using a 7-point Likert scale with endpoints labeled *1* (*very low grade*) and *7* (*very high grade*).

LSAT Task. Participants predicted how many LSAT questions the interviewee had answered correctly by entering a number between 1 and 17, inclusive.

GRE Math Task. Participants predicted how many GRE questions the interviewee had answered correctly by entering a number between 1 and 5, inclusive.

Bonus Distribution Task. Participants predicted how much of the bonus the interviewee had kept for themselves by entering a number between $1 and $12, inclusive.

Slider Task. Participants predicted how many sliders the interviewee dragged to the correct location by entering a number between 1 and 50, inclusive.

Confidence: For each task, participants reported their confidence in each prediction using a Likert scale with endpoints labeled *1 (not very confident)* and *7 (very confident).*

Results

Analysis plan: We analyzed these data to answer our main (pre-registered) research question: does access to verbal versus nonverbal versus paralinguistic cues differentially influence observers' predictive accuracy? If the interviews did provide useful information, participants may be more accurate in conditions that provide access to more-predictive information. For example, if observing interviews is more helpful for evaluating subjective tasks, participants may be more accurate when using that information to predict performance on subjectively- versus objectively-

evaluated tasks. In a pre-registered linear mixed effects model, we exclude the average-guess condition, and create 3 predictors by dummy coding the remaining conditions based on whether they include verbal, nonverbal, or paralinguistic information. This model included fixed effects for condition and random intercepts for each interviewee. We compared conditions to each other using planned post-estimation contrasts and a Holm-Bonferroni correction for multiple comparisons.

We evaluated an additional research question: does observing conversational content improve human prediction accuracy above and beyond seeing nothing and trying to guess an average? We calculated human prediction accuracy between the predictions of participants who evaluated interview stimuli in Study 1B and participants' actual task performance in Study 1A[10]. We compared the accuracy of those who saw (or heard) stimuli to the accuracy of those in the condition of guessing an average. This model included fixed effects for condition and random intercepts for each interviewee. We compared conditions to each other using planned post-estimation contrasts and a Holm-Bonferroni correction for multiple comparisons.

Question 1: What type of content led to the most prediction accuracy (verbal, nonverbal, or paralinguistic)? We investigated the extent to which observing each type of conversational content influenced participants' prediction accuracy—and their confidence in their predictions. We coded each of the seven conditions based on the type of information they contained: verbal, nonverbal, and/or paralinguistic. Conditions 1-4 received a code of 1 for nonverbal content because they contained visual information, conditions 1, 2, 5, and 6 received a code of 1 for paralinguistic content because they contained acoustic elements, and conditions 1, 3,

[10] We were primarily concerned with the level of accuracy, rather than the direction of potential inaccuracies (i.e. whether participants under- or over-predicted interviewee performance). We report directional findings in Appendix 2C.

5, and 7 received a code of 1 for verbal content because they contained the words the interviewee said. All other codes were set to zero. As pre-registered, Condition 0 (the average-guess condition) was omitted from this analysis.

Further, we separated task types in our analyses, in order to evaluate whether different content types led to more accurate predictions for different tasks (subjectively-evaluated tasks v. objectively-evaluated tasks). This categorization of tasks was not pre-registered, so we additionally present the results separately for each task (see Table 2.3).

Table 2.3: Task Type and Content Categories: Which Content Types Increase Predictive Accuracy? Nonverbal and paralinguistic content significantly predict predictive accuracy in subjectively-evaluated tasks, while they lead to increased inaccuracy in some objectively-evaluated tasks.

	Task	Task Type	Verbal content	Nonverbal content	Paralinguistic content
Subjectively-evaluated tasks	Group Conversation	Peer-rated likeability (1-7)	.05	.12***	.24***
	Termination letter	Third-party-rated quality (1-7)	.04	.14***	.18***
	Case summary	Third-party-rated quality (1-7_	.01	$.06^{+}$	.10***
Objectively-evaluated tasks	LSAT Logic	Number correct out of 14	-.08	-.35***	-.08
	GRE Math	Number correct out of 5	.05	-.10***	.04
	Slider Persistence	Number completed out of 50	$.68^{+}$	-.78*	-.86*
	Bonus Donation	Dollars allocated to self v. others	-.02	.02	-.005

Note: $^{+}p< .10$, $^{*}p< .05$, $^{**}p < .01$, $^{***}p<0.001$

Subjectively-Evaluated Tasks. We scaled and collapsed the accuracy measures across the three subjectively-evaluated tasks (subjective ratings of performance in the group conversation, termination letter, and case summary). In a model with verbal, nonverbal, and paralinguistic cues as predictors, a random effect for interviewee (since interviewees were evaluated multiple times) and a random effect for task (since we combined accuracy scores for several tasks), we found that observing paralinguistic information *increased* human predictive accuracy for subjectively-evaluated tasks (b = .17, t = 7.77, $p < .001$) as did observing nonverbal information (b = .11, t = 4.86, $p < .001$), while observing verbal content did not significantly affect prediction accuracy (b = .03, t = 1.49, p = .14). These results were consistent when each task was evaluated separately as well (another pre-registered analysis, represented in Table 2.3 and Appendix 2B).

Though it did not improve prediction accuracy, observing verbal information significantly increased participants' *confidence* in their predictions on subjectively-evaluated tasks (b = .20, t = 7.79, $p < .001$), while observing nonverbal or paralinguistic information did not affect their confidence.

Objectively-Evaluated Tasks. We scaled and collapsed the accuracy measures across the four objectively-evaluated tasks (LSAT, GRE math, slider, and bonus distribution tasks). We found that, overall, none of the three information types (paralinguistic, nonverbal, or verbal) significantly improved human prediction accuracy on these tasks, and suggestive evidence that having access to these cues significantly *decreased* predictive accuracy (nonverbals: b = -.30, t = -2.44, p = .015; paralanguage: b = -.25, t = -1.98, p = .048; verbals: b = -.21, t = -1.65, p = .10). These results are consistent when each task was evaluated separately as well (another pre-registered analysis, represented in Table 2.3 and Appendix 2B).

Although verbal information did not improve accuracy in predicting any performance task, verbal information significantly increased participants' feelings of *confidence* in their ability to predict the objectively-evaluated tasks ($b = .19$, $t = 7.65$, $p < .001$).

Question 2: Did observing conversational content improve human prediction accuracy above and beyond a guess about average performance? To answer this question, we compared the average-guess condition (in which participants guessed average task performance without observing interviews at all) to all other conditions. All other comparisons are reported in Appendix 2B.

Group Conversation: Liking. As shown in Figure 1, there were no significant differences between participants in the average-guess and other conditions when predicting subjective ratings of liking ($ps > .42$), except that participants in the average-guess condition were significantly better at predicting subjective ratings of liking compared to participants in the transcript condition ($b = .56$, $p < .001$). Participants in the transcript condition were also significantly less accurate compared to participants in all other conditions (all $bs < -0.37$, all $ps < .001$). Although participants in the average-guess condition were just as accurate as those who observed interview content, those with no information were significantly less confident than participants in all other conditions ($bs < -.69$, $ps < .001$).

Writing Task 1: Termination letter. There were no significant differences between participants in the average-guess and other conditions when predicting subjectively-evaluated performance in the termination letter task ($ps > .30$), except that participants in the average-guess condition were significantly better at predicting letter-writing performance compared to participants in the transcript condition ($b = .48$, $p < .001$). Participants in the transcript condition were again significantly less accurate compared to participants in all other conditions (b's $< -.23$,

p's < .044). Participants were less confident in the average-guess condition compared to several other conditions: full video (*b* = -.42, *p* <.001), full audio (*b* = -.42, *p* <.001), and transcript conditions (*b* = -.34, *p* =.003).

Writing Task 2: Case. There were no significant differences between participants' accuracy in predicting subjectively-evaluated performance on the case writing task (*p*s > .14). Participants were less confident in the average-guess condition compared to the full video condition (*b* = -.29, *p* =.024).

LSAT Task. There were no significant differences between conditions in participants' accuracy in predicting performance on the LSAT task (*p*s > .17). There were no significant differences in confidence between average-guess and other conditions.

GRE Math Task. Participants in the average-guess condition were significantly more accurate compared to participants in the full video condition (*b* = .39, *p* <.001). There were no significant differences in confidence between average-guess and other conditions.

Bonus Distribution Task. There were no significant differences between conditions in participants' accuracy in predicting bonus distributions (*p*s = 1). Participants were significantly less confident in predicting the bonus distribution in the average-guess condition, compared to all other conditions (*b*s < -.33, *p*'s <.02).

Slider Task. There were no significant differences between participants in the average-guess and other conditions when predicting performance on the slider task (*p*s > .09), except that participants in the average-guess condition were significantly more accurate compared to participants in the full video condition (*b* = 3.27, *p* =.040). There were no significant differences in confidence between average-guess and other conditions.

General Discussion

While our results replicate previous findings that observing interviews doesn't help people predict task performance on objective tasks (Devaul et al., 1987; Dana et al., 2013; Dubovsky et al., 2008; Elfenbein, Barsade, & Eisenkraft, 2015), they offer a window into how there may be useful information contained in conversational interviews. Participants were more accurate in predicting performance on subjectively-evaluated tasks when they received paralinguistic and nonverbal content (compared to verbal content). Consistent with prior work, our findings suggest that some paralinguistic and nonverbal cues, such as vocal pitch, cadence, volume, intonation, pausing, and eye contact, which often operate beyond conscious awareness or control (i.e., they are difficult to "fake") may offer valuable insight into people's personality and emotional states (Ambady et al., 2002; McAleer et al., 2014; Belin et al., 2011; Collins et al., *working*; Gottman & Porterfield, 1981; Hughes, Mogilski, & Harrison, 2014; Jones, Feinberg, DeBruine, Little, & Vukovic, 2010; Scherer et al., 2003; Laplante & Ambady, 2003; Templeton et al., 2022), and can help discern motivations behind language, such as whether a comment is genuine or sarcastic (Culpeper, 2011) or friendly or authoritative (Jeong et al., 2019). While the visual and acoustic properties of people's conversational behavior may be the most consistent aspects of their interpersonal behavior, observers do not intuit this, preferring instead to rely on verbal information.

Nonverbal and paralinguistic information were the most helpful for predicting performance on subjectively-evaluated tasks, such as writing an empathetic termination letter, compared to objectively-evaluated tasks such as solving math problems. This suggests that if conversational interviews are conducted at all, they may be best suited to predict performance on subjectively-evaluated tasks (which more closely resemble the interpersonal experience of a conversational interview).

When comparing simple effects across conditions, we found that participants in the full video condition (who had access to all three types of content) had poorer predictive accuracy than those in the conditions with nonverbal content and paralinguistic content presented independently. This suggests that access to verbal content may distract observers from perceiving and interpreting nonverbal and paralinguistic cues when presented together. This is a counterintuitive effect: In a separate pilot study (see Appendix 2D), when given the power to choose which interview stimuli they preferred, participants overwhelmingly chose to view the full video to predict subsequent job performance. Similarly, in Study 1B, participants indicated that they felt more confident when they received more information, but it didn't make them any better at predicting performance, and sometimes it made them worse.

Cognitively, these errors may be triggered because interviewers are overloaded with information, drowning out predictive signals with noise (Nisbett, Zukier, & Lemley, 1981; Yeomans et al., 2021). Lending support to this idea, in our data, participants were more accurate when hearing paralinguistic information alone (in a distorted audio clip) than when hearing paralanguage together with verbal content (in a full audio clip).

Although interpersonal interactions do not lead to as much predictive accuracy as people expect, they may have benefits outside the scope of this study. First, interacting directly with someone may be useful for forging a direct personal connection. Second, conversations serve goals in addition to interpersonal evaluation and prediction of future behavior. Interviews, for example, may benefit applicants by giving them a chance to gather information about their potential job or employer. And third, conversations may be more predictive of future behavior over longer time horizons. For example, many jobs require on-the-job learning—perhaps interviews provide predictive evidence about an interviewee's ability to learn or adapt over time.

Conclusion

These results show that observers often miss or misuse the information they glean from interpersonal interactions, overlooking predictive signals, such as paralinguistic cues and nonverbal cues, in favor of noise, such as verbal content. People additionally apply predictive signals to predict the wrong outcomes: paying attention to someone's nonverbals and paralanguage was helpful for predicting how likable and empathetic they were rated by others, while verbal information didn't help predict performance on any tasks (and, still, people would rather have access to verbal information). Taken together, these findings provide a more nuanced understanding of what interviews can be used for: you may not be able to predict how well someone can do on objective tasks like math problems and LSAT questions, but paying attention to their nonverbal presence and the sound of their voice (compared to the words they say) provides useful information about their transferrable interpersonal skills.

CHAPTER 3. THE ROLE OF PARALINGUISTIC CUES IN JUDGING THE COMPETENCE AND INTELLIGENCE OF NON-NATIVE SPEAKERS

Summary: Speaking in a non-native language is difficult. It is more cognitively demanding and less comfortable than speaking in a native language, and having an accent increases stereotyping and negative evaluations. But could speaking in a non-native language also have positive outcomes? Research demonstrates that people rely more on System 2 processing when speaking a non-native language (deliberating, translating, and rehearsing words before speaking). I hypothesize that due to non-native speakers utilizing System 2 processing more relative to native speakers, non-native speakers will produce higher-quality verbal content. In Study 1, I interview participants for a job, experimentally manipulating the presence of paralinguistic cues (including accents) by creating two conditions: an audio clip of the interview (with paralinguistic content and verbal content) and an interview transcript condition (with verbal content only). I replicate prior work, finding that in the audio clip condition, with accents audible, non-native speakers are evaluated more negatively compared to native speakers. However, in the transcript condition, I find suggestive evidence that non-native speakers are evaluated *more* favorably than native speakers. In Study 2, I examine recordings of Q&A sessions during a university speaker series. I replicate prior work demonstrating that non-native speakers speak up less (speaking for fewer seconds overall), and also replicate my finding in Study 1 that non-native speakers' comments are judged as higher quality relative to native speakers' comments when presented without paralinguistic information. Overall, I replicate past findings that non-native speakers speak up less, and are evaluated more negatively when they do speak, and introduce a novel upside to speaking in a non-native language: producing higher-quality verbal content.

Millions of people communicate in their non-native language in the workplace and in their daily lives, with English swiftly becoming the global lingua franca (Neeley, 2017). In the United States alone, 1 in 5 people speak a native language other than English (compared to 1 in 10 people in 1980; US Census, 2019). People are usually less proficient in their second language compared to their native language, and their second language is often learned in the classroom, in the workplace, and/ or due to immigration.

Speaking in a non-native language is difficult. It is cognitively demanding (Cook, 1997; Favreau and Segalowitz, 1983; Van Heuven and Dijkstra, 2010) and anxiety-inducing (Horwitz, Horwitz & Cope, 1986; MacIntyre, 1999). Speaking in a non-native language, and specifically speaking English with an accent, can lead to discrimination. For example, non-native speakers are perceived as less truthful (Lev-Ari & Keysar, 2010), are less likely to receive promotions (Huang, Frideger & Pearce, 2013), and are less likely to receive startup funding (Huang, Frideger & Pearce, 2013).

However, there are some benefits to speaking in a non-native language. Research has demonstrated both enhanced behavioral outcomes and positive brain changes in bilingual and multilingual individuals (Li & Jeong, 2020). In terms of behavioral outcomes, multilingual exposure promotes effective communication by enhancing perspective taking in children (Fan et al., 2015), and even infants (Liberman, Woodward, Keysar, & Kinzler, 2017). Converging evidence also shows that bilinguals have an advantage in theory of mind, or the ability to predict the mental states of other people and to predict and explain their behavior based on these predictions (Schroeder, 2018). In terms of brain development, a recent body of research provides increasing evidence that bilingualism contributes to neuroplasticity and brain growth in the executive control region, because individuals must resist interference from their native language

when speaking in a non-native language, and because "bilingualism is a long-term cognitively challenging experience" (e.g., Pliatsikas, Pereira Soares, Voits, DeLuca, & Rothman, 2021). The long-term effect of this is that bilingual and multilingual individuals are more resistant to cognitive decline and the onset of Alzheimer's disease (for a meta-analysis, see: Anderson, Hawrylewicz, & Grundy, 2020).

System 1 & System 2 in Non-Native Language

One benefit with particular relevance to this inquiry is that people speaking in a non-native language rely more on System 2 processing: a more deliberate, controlled, and analytical cognitive pathway (compared to System 1 processing that is faster, more intuitive, and more affective, e.g., Kahneman, 2003; Sloman, 1996; Stanovich & West, 2000). A number of studies demonstrate that when people use a foreign language, they systematically make different choices than when they use their native tongue (e.g., Costa et al., 2014; Keysar, Hayakawa, & An, 2012). In particular, it appears that choices made when problems are presented in a foreign language are less subject to intuitive biases (Costa et al., 2014). Speaking in a non-native language leads to a reduction in decision-making biases; when choices are presented in a non-native (versus native) language the framing effect disappears, and the effect of loss aversion is reduced (Keysar, Hayakawa, & An, 2012). Overall, people tend to choose utilitarian options much more often when using a foreign language than their native language (e.g., Cipolletti, McFarlane, & Weissglass, 2016; Corey et al., 2017; Costa et al., 2014; Geipel, Hadjichristidis, & Surian, 2015)

System 2 could be activated more in non-native language processing due to either a decrease in effective System 1 processing, an increase in deliberative System 2 processing, or both happening simultaneously. In terms of decreasing System 1 processing, scholars argue that speaking in a foreign language creates a distancing effect that moves people away from immediate

intuitive and emotional responses to a more analytical mode of thinking (e.g., Pavlenko, 2005). This reduction in emotional reaction may decrease the impact of emotional processes, allowing individuals to rely more on deliberative and analytical processes when making decisions. Consequently, making decisions in a foreign language could reduce reliance on System 1 and consequent emotional bias (Keysar, Hayakawa, & An, 2012).

Another possibility is that foreign language contexts may prompt more analytic thinking relative to native language contexts because of the increased deliberation and processing difficulty required when speaking in a non-native language (Alter, Oppenheimer, Epley, & Eyre, 2007; Costa et al., 2014). According to scholars, the level of cognitive fluency is a crucial factor that determines how much influence intuitive processes have on decision-making (i.e., Kahneman, 2011; Reber, Schwarz, & Winkielman, 2004). Specifically, in situations where cognitive fluency is high and little cognitive effort is required (such as speaking in a native language), the intuitive System 1 has a greater impact on decision-making compared to situations where cognitive fluency is disrupted. In the latter situations, people tend to be more cautious when responding, raising their attention levels and allowing System 2 to have a greater influence on their responses (Costa et al., 2014). For example, presenting problems in a difficult-to-read print reduces intuitive biases in decision-making problems, even in contexts where there is no emotional component (Alter et al., 2007). Based on such findings, Kahneman (2011) argues that any factor that increases cognitive tension and reduces cognitive fluency, regardless of its origin, would activate System 2 and decrease the impact of the quick and automatic response provided by System 1. Thus, according to this increased deliberation account, because a foreign language is typically processed less automatically and is less familiar than a native tongue, this reduced fluency and corresponding increased cognitive tension promotes more System 2 processing (Favreau & Segalowitz, 1983;

Keysar, Hayakawa, & An, 2012). According to this argument, speaking in a non-native language should promote System 2 processing regardless of the reduction in emotional distance that is proposed to reduce System 1 processing (Costa et al., 2014).

Because speaking in a non-native language promotes increased System 2 processing, and System 2 processing increases reflection, deliberation, and thoughtfulness, and makes people less susceptible to intuitive biases, we hypothesize that when non-native speakers speak, their content will be of a higher quality compared to native speakers.

Hypothesis 1: Non-native speakers will produce higher-quality verbal comments compared to native speakers.

A second consequence of foreign-language processing being more costly and disrupting cognitive fluency is that people are more cautious of their responses (Costa et al., 2014). Speaking in a non-native language is notoriously anxiety-inducing, and often debilitatingly so (Horwitz, Horwitz & Cope, 1986; MacIntyre, 1999). Research shows that non-native speakers speak less than native speakers, in laboratory studies (Freiermuth, 2001), in workplaces (Morrison, 2011; Aichorn & Puck, 2017; Behtoui, Boréus, Neergaard, & Yazdanpanah, 2017), in school classrooms (Riasati, 2012; Effiong, 2016), and even in MBA classrooms (Lu, Nisbett, & Morris, 2022. We expect to find a similar effect.

Hypothesis 2: When speaking their non-native language, people will speak less.

Biases Against Non-Native Speakers

In general, listeners exhibit bias against non-native speakers, and have a preference for native accents over non-native accents (Livingston, Schilpzand & Erez, 2017). This manifests in a variety of ways in the workplace and broader society. Non-native speakers are less likely than

native speakers to be recommended to receive promotions at work, including in management positions (Huang, Frideger & Pearce, 2013). Having an accent influences credibility; people perceive statements as less truthful when spoken by non-native speakers and view non-native speakers overall as less credible (Lev-Ari & Keysar, 2010). Non-native speakers also were found to have a significantly lower likelihood of receiving new-venture funding (Huang, Frideger & Pearce, 2013). On a broader scale, asymmetries in language fluency contribute to an us-vs-them dynamic in global teams; language proficiency is a fault-line along which groups can fragment (Neeley, Hinds, & Cramton 2012). Thus, we expect that in our studies, people will similarly evaluate non-native speakers more negatively, but only if their non-native status is perceptible (i.e., hearing an audio clip of someone speaking versus reading a transcript of them speaking).

> *Hypothesis 3: If it is perceptible that someone is speaking in a non-native language (i.e., because they have an audible accent), people will evaluate them more negatively than native speakers. This effect will disappear if the accent is imperceptible (i.e., if they are evaluating a transcript).*

Study 1A: Stimuli Creation

We recruited participants to complete a job interview. These job interviews were video-recorded, and we used the recordings to develop two conditions of stimuli, an audio clip condition (audio + verbal content present) and a transcript condition (only verbal content present). This allowed us to control for verbal content while manipulating whether audio cues, which allow for the identification of non-native speakers, were present.

Method.

Participants: We recruited 105 participants, in 35 groups of three, from a behavioral lab listserv (M_{Age} = 26.35, *SD* = 8.64; 59% female) to participate in a laboratory experiment in exchange for a $12 payment.

Of these 105 participants, 31 were non-native speakers of English, the language in which the job interview was conducted (M_{Age} = 25.19, *SD* = 6.61; 71% female), and the remaining 74 were native speakers of English (M_{Age} = 26.85, *SD* = 9.38; 54% female).

Design: Each participant entered a room with a researcher. The participants were told that they would be interviewing for a project manager position, which involved "coordinating between many different teams, communicating deadlines and priorities, and executing tasks along with a team." The interviews were video-recorded. Researchers asked participants a series of 10 common interview questions, including "Tell me about a time that you tried something and failed" (see Appendix 2A). Researchers were trained to remain neutral between questions, responding to each answer by simply saying "*Thank you for your answer/ thank you for sharing that. My next question is…*". At the end of 10 minutes, the interview concluded.

Stimuli Creation: We modified the video recordings of the interviews to create two conditions. In the audio condition, an audio file of the interview was created (preserving verbal content and auditory content). In the transcript condition, a transcript was created of the interview text, preserving verbal content alone (no audio content).

Study 1B: Evaluating Native and Non-Native Interviewees

In Study 1B, we recruited a new set of participants to evaluate the stimuli created in Study 1A. Participants either listened to the audio clip of the interview (audio condition), or read a transcript of the interview (transcript condition), and then judged the interviewee on how

competent they seemed, how warm they seemed, and the quality of their responses to the interviewer's questions.

Method.

Participants: We recruited 1,089 participants on Prolific academic (M_{Age} = 26.06, *SD* = 7.80; 49% female, 50% male, 1% prefer to self-identify) to participate in an experiment in exchange for a $1.20 payment.

Design: Participants were randomly assigned to be in either the audio condition (audio + verbal content available), or the transcript condition (only verbal content available). In the audio condition, participants listened to an audio clip of 2 minutes of the interview (seconds 0:30- 2:30). In the transcript condition, participants read a transcript corresponding to the same 2-minute section of the interview. The verbal content was identical in both conditions. There were a series of attention checks in both conditions (audio attention checks in the audio condition, and reading attention checks in the transcript condition).

Measures: To measure the perceived *quality* of the interviewees' responses, we asked participants to answer the following question: "Overall the **quality** of the interviewees' responses were of..." (1) *Very Low Quality* to (7) *Very High Quality.*

To measure the *task competence* of the interviewee, we asked participants to rate the candidates' task competence. This measure included statements such as "The interviewee was capable" and "The interviewee was efficient"; responses ranged from (1) *Strongly Disagree* to (7) *Strongly agree.* See Appendix 3A for a full list of items. This measure yielded a Cronbach's alpha of .95.

To measure the *trait competence* of the interviewee, we asked participants to rate several questions about the candidate, adapted from Cuddy, Fiske, & Glick (2008). These included

questions such as “How competent was the candidate?” and “How confident was the candidate?”; responses ranged from (1) *Not at all* to (7) *Very much.* See Appendix 3A for a full list of items. This measure yielded a Cronbach’s alpha of .94.

To measure the *trait warmth* of the interviewee, we asked participants to rate several questions about the candidate, adapted from Cuddy, Fiske, & Glick (2008). These included questions such as “How friendly was the candidate?” and “How good-natured was the candidate?”; responses ranged from (1) *Not at all* to (7) *Very much.* See Appendix 3A for a full list of items. This measure yielded a Cronbach’s alpha of .92.

Results

Quality of Responses. In the audio condition, where non-native accents were perceptible, participants judged native English speakers as giving higher quality answers to the interviewer’s questions ($M = 4.83$, $SD = 1.57$) compared to non-native speakers ($M = 4.48$, $SD = 1.45$), $t(493) = 2.24$, $p = .026$, $d = .22$. In a Kruskal-Wallis test to correct for uneven samples among native and non-native speakers, this difference persisted ($X^2 = 6.02$, $df = 1$, $p = .014$). However, in the transcript condition, with only verbal content available, this effect was directionally (although not significantly) reversed. In the transcript condition, participants judged native English speakers as giving directionally lower-quality answers to the interviewer’s questions ($M = 4.22$, $SD = 1.77$) compared to non-native speakers ($M = 4.06$ $SD = 1.78$), $t(514) = -.90$, $p = .37$, $d = -.09$ (Kruskal-Wallis test: $X^2 = .97$, $df = 1$, $p = .32$).

Task Competence. Turning to the task competence scale, in the audio condition, participants judged native English speakers as being higher in task competence ($M = 4.85$, $SD = 1.58$) compared to non-native speakers ($M = 4.30$, $SD = 1.53$), $t(493) = 2.89$, $p = .004$, $d = .29$ (Kruskall-Wallis: $X^2 = 9.58$, $df = 1$, $p = .002$). However, in the transcript condition, with only verbal content available,

this effect was marginally (although not significantly) reversed: participants judged native English speakers as being marginally lower in task competence (M = 4.11, SD = 1.81) compared to non-native speakers (M = 4.39 SD = 1.92), $t(514)$ =-1.61, p = .11, d = -.29 (Kruskal-Wallis test: X^2 = 2.84, df = 1, p = .09).

Trait Competence. Turning to trait competence, in the audio condition, participants judged native English speakers as being higher in trait competence (M = 5.26, SD = 1.44) compared to non-native speakers (M = 4.82, SD = 1.51), $t(493)$ = 3.12, p = .002, d = .31 (Kruskall-Wallis: X^2 = 8.37, df = 1, p = .004). However, in the transcript condition, this effect was marginally (although not significantly) reversed: participants judged native English speakers as being marginally lower in trait competence (M = 4.65, SD = 1.67) compared to non-native speakers (M = 4.95 SD = 1.57), $t(514)$ =- 1.83, p = .067, d = -.20 (Kruskal-Wallis test: X^2 = 2.87, df = 1, p = .09).

Warmth. Lastly, turning to warmth, in the audio condition, participants judged native English speakers as warmer (M = 5.74, SD = 1.25) compared to non-native speakers (M = 5.47, SD = 1.26), $t(493)$ = 2.22, p = .027, d = .22 (Kruskall-Wallis: X^2 = 6.04, df = 1, p = .014). However, in the transcript condition, this effect was significantly reversed: participants judged native English speakers as being significantly lower in warmth (M = 5.38, SD = 1.36) compared to non-native speakers (M = 5.66 SD = 1.27), $t(514)$ =- 2.18, p = .030, d = -.22 (Kruskal-Wallis test: X^2 = 6.04, df = 1, p = .014).

Discussion

In this study, we replicated prior work showing that non-native English speakers are evaluated more negatively compared to native English speakers. We find that non-native English speakers are judged as less warm and competent, and that their responses to interview questions are judged as lower-quality compared to native speakers.

However, by creating a transcript condition that strips away the audio content and preserves only the verbal content, we can compare the verbal content in isolation, without cues that allow for the identification of native or non-native speakers. We find that the difference between native speakers and non-native speakers goes away in the transcript condition, and moreover find suggestive evidence that this difference may even be reversed; in the transcript conditions, non-native speakers are evaluated as directionally *more* competent and their responses as directionally *higher* quality compared to native English speakers. Non-native speakers are also evaluated as significantly warmer compared to native English speakers in the transcript condition. Together, this indicates that although non-native English speakers are evaluated more negatively when their accents are perceptible, they may be producing overall better verbal content when they speak, as evidenced by evaluations of verbal content evaluated in isolation.

Study 2: University Public Lecture Series

In Study 2, we tested our hypotheses using a series of publicly available lectures online. After each lecture, students were able to ask questions to the speaker. We transcribed these questions, and asked third parties to evaluate the transcribed questions for the perceived intelligence of the speaker and the quality of the question. We used natural variation in whether the question-askers were native English speakers or non-native English speakers, and thus were able to test whether there was a difference in question quality based on native versus non-native speakers. We moreover looked at the duration of the questions asked, to test whether there was a difference in question length based on whether the question-asker was a native versus non-native speaker.

Method.

Question Dataset: We created a dataset of a university lecture series, based on publicly available videos. We chose a lecture series at a public policy school, where there were relatively more attendees that were international, to ensure that there were enough non-native speakers in the dataset. We used every video posted in a two-year period (2021 and 2022), which gave us a dataset of 33 lecture videos. Within these 33 lecture videos, there were 176 audience questions. Of these, 149 question-askers were native English speakers, and 27 were non-native English speakers.

All questions were transcribed, and the time stamps of the question-start time and question-end time were noted. Two research assistants coded the question-askers as being native-English speakers or non-native English speakers (0/1). They coded the first 2 lectures (13 questions) independently and had 100% agreement on their classifications. The coders then each coded half of the remaining dataset. For the purposes of this project, no further distinctions were considered (i.e., native English speakers who were American versus British, Australian, and on). Any names or identifying characteristics provided by the question-askers during their questions were redacted from the transcript.

Participants: We aimed to recruit 400 participants, and eventually recruited 407 participants from an online sample. 4 participants did not pass attention checks, and a total of 403 participants completed the survey (M_{Age} = 40.19, SD = 14.43; 48% female, 51% male, 1% non-binary or other gender) which took a total of 5-minutes in exchange for a $1.00 payment.

Design: Before each question, the participant read a few sentences of context about the lecture and the speaker who was invited (there was one passage about context/ about the lecturer for each of the 33 lectures). All participants then viewed a randomly selected sample of 5 transcribed questions (text only), from the 176 possible questions.

Measures: To measure the perceived *intelligence* of the question-asker, we asked participants to answer: "How intelligent do you think the person asking this question is?", with responses ranging from (1) *Not at All* to (10) *Very Much*). To measure the perceived *warmth* of the question-asker, we asked participants to answer: "How friendly/ warm do you think the person asking this question is?", with responses ranging from (1) *Not at All* to (10) *Very Much*). To measure the perceived *competence* of the question-asker, we asked participants to answer: "How competent do you think the person asking this question is?", with responses ranging from (1) *Not at All* to (10) *Very Much*). Lastly, to measure the perceived *quality* of the question, we asked participants to answer: "Overall, how would you rank the quality of this question?", with responses ranging from (1) *Very Low Quality* to (10) *Very High Quality*.

To measure the *duration* of the interviewees' questions in seconds, we subtracted the question ending timestamp from the question beginning timestamp.

Results

Duration. Native English speakers spoke for about 10 seconds longer when asking questions ($M = 44.60$, $SD = 18.73$) compared to non-native speakers ($M = 34.56$, $SD = 24.90$), $t(174) = 2.43$, $p = .016$, $d = .51$, supporting Hypothesis 2. In a Kruskal-Wallis test to correct for uneven samples among native and non-native speakers, this difference persisted: $X^2 = 10.00$, $df = 1$, $p = .002$. In a regression controlling for the lecture number, the year of the lecture, and the format of the speech (virtual versus hybrid), being a non-native speaker continued to emerge as a negative predictor of length of comment ($b = -6.90$, $t = -5.28$, $p < .001$). In this model, the format of the speech (virtual versus hybrid) also emerged as a marginally significant predictor of length ($b = 2.40$, $t = 1.79$, $p = .073$); suggesting that people may speak for longer when asking questions in virtual lectures

versus in-person lectures. (Lecture format had no significant interaction effect with native versus non-native English speakers).

Perceived Intelligence. Participants rated the transcribed comments and questions given by non-native English speakers to be significantly more intelligent (M = 4.98, SD = 1.43) compared to those given by native speakers (M =4.61, SD = 1.50), $t(2008) = 3.90$, $p < .001$, $d = .25$. In a Kruskal-Wallis test to correct for uneven samples among native and non-native speakers, this difference persisted: $X^2 = 16.15$, $df = 1$, $p < .001$. In a mixed-level regression controlling for the lecture number, the year of the lecture, the format of the speech (virtual versus hybrid), and with a random effect controlling for the respondent (since each respondent rated 5 transcribed comments), being a non-native speaker continued to emerge as a positive predictor of third-party rated intelligence of the comment (b= .41, t= 4.62, p <.001).

Perceived Competence. Participants rated the questions and comments from non-native English speakers to be significantly more competent (M = 4.98, SD = 1.52) compared to those given by native speakers (M =4.57, SD = 1.53), $t(2008) = 4.12$, $p < .001$, $d = .26$. In a Kruskal-Wallis test to correct for uneven samples among native and non-native speakers, this difference persisted: $X^2 = 19.09$, $df = 1$, $p < .001$. In a mixed-level regression controlling for the lecture number, the year of the lecture, the format of the speech, and with a random effect controlling for the respondent, being a non-native speaker continued to emerge as a positive predictor of third-party rated competence of the comment (b= .44, t= 4.78, p <.001).

Perceived Quality. Participants rated the questions and comments from non-native English speakers to be of significantly higher quality (M = 4.75, SD = 1.67) compared to those from native speakers (M =4.30, SD = 1.69), $t(2008) = 4.20$, $p < .001$, $d = .27$. In a Kruskal-Wallis test to correct for uneven samples among native and non-native speakers, this difference persisted: $X^2 = 18.82$,

df = 1, $p < .001$. In a mixed-level regression controlling for the lecture number, the year of the lecture, the format of the speech, and with a random effect controlling for the respondent, being a non-native speaker continued to emerge as a positive predictor of perceived quality of the comment (*b*= .51, *t*= 4.87, $p < .001$).

Perceived Warmth and Friendliness. Lastly, participants rated the questions and comments from non-native English speakers to be significantly warmer and friendlier (*M* = 4.77, *SD* = 1.44) compared to those from native speakers (*M* =4.47, *SD* = 1.39), *t*(2008) = 3.45, $p < .001$, *d* = .22. In a Kruskal-Wallis test to correct for uneven samples among native and non-native speakers, this difference persisted: $X^2 = 12.32$, *df* = 1, $p < .001$. In a mixed-level regression controlling for the lecture number, the year of the lecture, the format of the speech, and with a random effect controlling for the respondent, being a non-native speaker continued to emerge as a positive predictor of perceived warmth and friendliness of the comment (*b*= .39, *t*= 4.90, $p < .001$).

Discussion

In Study 2, I assembled a dataset of students asking questions and giving comments after a guest speaker's lecture. The dataset included 2 years' worth of lectures, and a total of 33 videos, from YouTube, which allowed me to test my hypotheses in a naturalistic environment. The comments were transcribed, and research assistants noted how long each person spoke. A separate population of third parties was recruited to rate each comment based on its quality, intelligence, competence, warmth and friendliness.

I found significant evidence that third-party raters rated transcribed comments of non-native English speakers as more intelligent, more competent, higher quality, and warmer and friendlier compared to the comments of native English speakers, supporting Hypothesis 1. I also found

significant evidence that non-native speakers spoke for shorter durations of time, supporting Hypothesis 2.

General Discussion

The studies in this paper replicate prior work which finds that non-native speakers speak less than native speakers, and are evaluated more negatively than native speakers when they do speak. However, these studies present a potential upside to speaking in a non-native language: producing higher quality verbal content, which I hypothesize is due to increased reliance on System 2 processing in non-native speakers, relative to native speakers.

In Study 1, we replicate prior findings that non-native speakers are evaluated more negatively compared to native speakers (Lev-Ari & Keysar, 2010; Huang, Frideger & Pearce, 2013), in a job interview setting. We similarly find that non-native English speakers are judged as less competent, and less warm, and that their responses to interview questions are judged as lower quality compared to native speakers, supporting Hypothesis 3. However, we introduce a boundary to this finding—this is only the case if paralinguistic cues are available, leading to their identification as non-native speakers. In the transcript condition, without the presence of paralinguistic cues, obscuring their status as non-native speakers, these differences go away. Moreover, we find suggestive evidence that this difference may even be reversed; in the transcript conditions, non-native speakers are evaluated as significantly *more* warm, directionally *more* competent, and their responses as directionally *higher* quality compared to native English speakers. This provides initial support for Hypothesis 1: that non-native speakers may produce higher-quality verbal content compared to native speakers.

In Study 2, we use a publicly available database of a university lecture series to replicate prior findings that non-native speakers speak less than native speakers (Freiermuth, 2001;

Morrison, 2011; Aichorn & Puck, 2017; Behtoui, Boréus, Neergaard, & Yazdanpanah, 2017; Riasati, 2012; Effiong, 2016; Lu, Nisbett, & Morris, 2022, finding that non-native speakers speak for shorter durations of time when they do speak. We asked third parties to evaluate the transcribed questions and comments raised by students, and find additional support for Hypothesis 1: the questions and comments raised by non-native speakers are rated as more intelligent, competent, warm, and of higher quality compared to those raised by native speakers.

These findings make three major contributions to the research on communication, interpersonal perception formation, and multilingualism. First, this work contributes to the evidence suggesting that System 2 is activated more in non-native language processing. These findings further contribute to the scholarly debate about whether non-natives rely on System 2 more because due to an increase in deliberative System 2 processing when speaking a non-native language, or due to a decrease in effective System 1 processing. In the latter case, scholars argue that speaking in a foreign language creates a distancing effect that moves people away from immediate intuitive and emotional responses to a more analytical mode of thinking (e.g., Pavlenko, 2005), and that this reduction in the emotional reaction may decrease the impact of emotional processes, allowing individuals to rely more on deliberative and analytical processes when making decisions. However, this work suggests that non-native speakers are not less emotional, finding that non-native speakers are evaluated as warmer compared to native speakers (Studies 1&2). This work suggests that it is more likely that non-native speakers rely more on System 2 because foreign language contexts prompt more analytic thinking, increased deliberation, and more caution and attention compared to speaking in a native language (Alter, Oppenheimer, Epley, & Eyre, 2007; Costa et al., 2014). Indeed, I find that non-native speakers' verbal content is evaluated as more intelligent and competent (suggestive evidence in Study 1, confirmatory evidence in Study 2).

Thus, my work lends additional evidence to the increased deliberation account of reliance on System 2 in non-native language speakers.

A second contribution of this work is an examination of the interplay between paralinguistic cues and verbal cues. Prior work finding that non-native speakers are evaluated more negatively has examined them in contexts in which paralinguistic cues are available, but has not evaluated verbal cues in isolation. Because paralinguistic cues have a significant impact on the interpretation of verbal cues (Clark and Clark, 1977; Mehrabian, 1972), evaluating verbal content without paralinguistic cues can yield important insights.

Lastly, although research has highlighted the challenges and difficulties associated with speaking in a non-native language, such as increased anxiety (Horwitz, Horwitz & Cope, 1986; MacIntyre, 1999), and negative evaluations from others (Lev-Ari & Keysar, 2010 Huang, Frideger & Pearce, 2013), there may be potential benefits to this experience that have not yet been fully explored. I present a novel and significant upside to speaking in a non-native language: non-native speakers may actually produce higher-quality verbal content. This is a promising result with important implications for the millions of people speaking in their non-native language on a daily basis, as well as those who manage them, supervise them, mentor them, and support them. Additionally, this research may encourage a reevaluation of the traditional bias toward native speakers present in classrooms, workplaces, and social settings. This finding furthermore presents a compelling case for the value of diversity along the dimensions of country-of-origin in organizations in classrooms.

Limitations and Avenues for Future Research

Firstly, although two findings in this paper are replications of prior work (that non-native speakers speak less, and that they are evaluated more negatively when they are heard), future

studies are needed to replicate the novel finding that non-native speakers may produce higher-quality verbal content. Nonetheless, this preliminary evidence offers a promising avenue for future research.

Secondly, I hypothesize that non-native speakers may produce higher-quality verbal content than native speakers, due to an increased reliance on System 2 processing. However, I have not tested and confirmed a relatively greater reliance on System 2 as a mechanism. Future studies could evaluate the extent to which native and non-native speakers rely on System 1 versus System 2 processing, directly surveying native and non-native speakers during or after job interviews or classroom discussions. Alternatively, other behavioral indicators, such as reaction times, or other markers of reliance on System 2 versus System 1 processing could be utilized to assess cognitive load and test this hypothesized mechanism.

Thirdly, the present work examines the interplay between verbal and paralinguistic cues; however, a third major category of interpersonal cues - nonverbal (visual) cues - has been excluded. These include a person's appearance, body language, and facial expressions. It is possible that the presence of certain nonverbal cues may lead to an unconscious inference that someone is a non-native speaker, which could similarly lead to negative evaluations of verbal content, even in the absence of paralinguistic content. For instance, it is plausible that verbal content associated with an individual who possesses physical features typical of immigrant groups (such as those of Asian, Middle Eastern, Latinx, and Southeast Asian heritage) might be evaluated less positively. Future research should examine the extent to which nonverbal cues typical of non-native speakers influence third-party evaluations of verbal content, and whether this effect is mediated by the perception that the speaker is a non-native speaker.

Fourthly, future work could also investigate the impact of the format of the classroom, meeting, or conversational setting. As workplaces and classrooms are increasingly moving to incorporate virtual elements, it is important to understand the impact on the behaviors and perceptions of non-native speakers in these formats. Prior research indicates that non-American speakers in US classrooms tend to speak more in virtual formats (Lu, Nisbett, & Morris, 2022). While our study revealed a main effect indicating that students tend to speak longer in virtual formats, we did not find a significant interaction effect demonstrating that this was especially true for non-native speakers. However, given the relatively small sample size of the total number of lecture videos in our study (33 videos, with 5 virtual and 28 in-person), further research with a larger sample size could assess format as a potential moderator.

Conclusion

This work replicates prior findings that non-native speakers of English speak less and are evaluated more negatively when they do speak, but presents a potential novel upside to speaking in a non-native language: non-native speakers may produce higher-quality verbal content, which is likely due to their increased reliance on System 2 processing when speaking. This work speaks to the importance of examining the interplay between verbal and paralinguistic cues when studying interpersonal interactions and interpersonal perception formation.

REFERENCES

Abbey, A., & Melby, C. K. (1986). The effects of nonverbal cues on gender differences in perceptions of sexual intent. *Sex Roles*, 15(5–6), 283–298.

Aichhorn, N., & Puck, J. (2017). "I just don't feel comfortable speaking English": Foreign language anxiety as a catalyst for spoken-language barriers in MNCs. *International Business Review*, *26*(4), 749-763.

Aliyev, R., Karakus, M., & Ulus, H. (2013). Teachers' verbal cues that cause students to feel various emotions. *The Anthropologist*, *16*(1-2), 263-272.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of experimental psychology: General*, *136*(4), 569.

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in experimental social psychology* (Vol. 32, pp. 201-271). Academic Press.

Ambady, N., LaPlante, D., Nguyen, T., Rosenthal, R., Chaumeton, N., & Levinson, W. (2002). Surgeons' tone of voice: a clue to malpractice history. *Surgery*, 132(1), 5-9.

Anderson, J. A., Hawrylewicz, K., & Grundy, J. G. (2020). Does bilingualism protect against dementia? A meta-analysis. *Psychonomic Bulletin & Review*, 27, 952-965.

Anderson, C., & Kilduff, G. J. (2009). Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance. *Journal of personality and social psychology*, *96*(2), 491.

Ashford, S. J., Rothbard, N. P., Piderit, S. K., & Dutton, J. E. (1998). Out on a limb: The role of context and impression management in selling gender-equity issues. *Administrative Science Quarterly*, 23-57.

Ashford, S. J., & Sitkin, S. B. (2019). From problems to progress: A dialogue on prevailing issues in leadership research. *The Leadership Quarterly*, *30*(4), 454-460.

Baer, M., & Frese, M. (2003). Innovation is not enough: Climates for initiative and psychological safety, process innovations, and firm performance. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *24*(1), 45-68.

Balkwell, J. W. (1991). From expectations to behavior: An improved postulate for expectation states theory. *American Sociological Review*, 355-369.

Banks, G. C., Woznyj, H. M., & Mansfield, C. A. (2021). Where is "behavior" in organizational behavior? A call for a revolution in leadership research and beyond. *The Leadership Quarterly*, 101581.

Barry, B., & Stewart, G. L. (1997). Composition, process, and performance in self-managed groups: The role of personality. *Journal of Applied psychology*, *82*(1), 62.

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic bulletin & review*, 13(6), 1061-1066.

Behtoui, A., Boréus, K., Neergaard, A., & Yazdanpanah, S. (2017). Speaking up, leaving or keeping silent: Racialized employees in the Swedish elderly care sector. *Work, employment and society*, *31*(6), 954-971.

Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*(4), 711-725.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior research methods*, *40*(2), 531-539.

Berger, J., Cohen, B. P., & Zelditch Jr, M. (1972). Status characteristics and social interaction. *American Sociological Review*, 241-255.

Berger, J., Rocklage, M. D., & Packard, G. (2022). Expression modalities: How speaking versus writing shapes word of mouth. *Journal of Consumer Research*, *49*(3), 389-408.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Social attention and real-world scenes: The roles of action, competition and social content. *The Quarterly Journal of Experimental Psychology*, *61*(7), 986-998.

Boland, J.E., Fonseca, P., Mermelstein, I., & Williamson, M. (2021). Zoom disrupts the rhythm of conversation. *Journal of Experimental Psychology: General.*

Bonaccio, S., O'Reilly, J., O'Sullivan, S. L., & Chiocchio, F. (2016). Nonverbal behavior and communication in the workplace: A review and an agenda for research. *Journal of Management*, *42*(5), 1044-1074.

Brescoll, V. L. (2011). Who takes the floor and why: Gender, power, and volubility in organizations. *Administrative Science Quarterly*, *56*(4), 622-641.

Brown, M., Westrich, B., Bates, F., Twibell, A., & McGrath, R. E. (2020). Preliminary evidence for virtue as a cue to long-term mate value. *Personality and Individual Differences*, *167*, 110249.

Burris, E. R., Detert, J. R., & Chiaburu, D. S. (2008). Quitting before leaving: the mediating effects of psychological attachment and detachment on voice. *Journal of applied psychology*, *93*(4), 912.

Camp, N. P., Voigt, R., Jurafsky, D., & Eberhardt, J. L. (2021). The thin blue waveform: Racial disparities in officer prosody undermine institutional trust in the police. *Journal of Personality and Social Psychology*.

Carney, D. R., Cuddy, A. J., & Yap, A. J. (2015). Review and summary of research on the embodied effects of expansive (vs. contractive) nonverbal displays. *Psychological science*, *26*(5), 657-663.

Carney, D. R., Hall, J. A., & LeBeau, L. S. (2005). Beliefs about the nonverbal expression of social power. *Journal of Nonverbal Behavior*, *29*, 105-123.

Carton, A. M. (2022). The science of leadership: A theoretical model and research agenda. *Annual Review of Organizational Psychology and Organizational Behavior*, *9*, 61-93.

Castro, D. R., Anseel, F., Kluger, A. N., Lloyd, K. J., & Turjeman-Levi, Y. (2018). Mere listening effect on creativity and the mediating role of psychological safety. *Psychology of Aesthetics, Creativity, and the Arts,* 12(4), 489.

Chan, S. C., Huang, X., Snape, E., & Lam, C. K. (2013). The Janus face of paternalistic leaders: Authoritarianism, benevolence, subordinates' organization-based self-esteem, and performance. *Journal of Organizational Behavior,* 34(1), 108-128.

Chan, S. C. (2014). Paternalistic leadership and employee voice: Does information sharing matter? *Human Relations*, *67*(6), 667-693.

Chen, L., Wadei, K. A., Bai, S., & Liu, J. (2020). Participative leadership and employee creativity: a sequential mediation model of psychological safety and creative process engagement. *Leadership & Organization Development Journal.*

Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral foreign-language effect. *Philosophical Psychology*, *29*(1), 23-40.

Clark, H. H., & Clark, E. V. (1977). *Psychology and Language.*

Clarke, S., Krom, C., & King, R. M. (2015). Identification of Dishonesty Through the Observation of Non Verbal Cues. *Psyking.net.*

Coffman, K. B. (2014). Gender differences in willingness to guess. *Management Science*, *60*(2), 434-448.

Collins, H., Minson, J., Kristal, A., & Brooks, A.W. (*Working)* Listening detection: Amidst live conversation, people can't tell when others are listening or not.

Conrad, P. (1988). Learning to doctor: Reflections on recent accounts of the medical school years. *Journal of health and social behavior*, 323-332.

Conty, L., George, N., & Hietanen, J. K. (2016). Watching eyes effects: When others meet the self. *Consciousness and cognition*, *45*, 184-197.

Cook, G. 1997. 'Language play, language learning'. *English Language Teaching Journal* 51/3: 224-231.

Cooney, G., Mastroianni, A. M., Abi-Esber, N., & Brooks, A. W. (2020). The many minds problem: disclosure in dyadic versus group conversation. *Current Opinion in Psychology*, *31*, 22-27.

Corneille, O., Mauduit, S., Holland, R. W., & Strick, M. (2009). Liking products by the head of a dog: Perceived orientation of attention induces valence acquisition. *Journal of Experimental Social Psychology*, *45*(1), 234-237.

Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with expert and novice teachers. *International Journal of Science and Mathematics Education*, *13*(2), 389-403.

Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). "Piensa" twice: On the foreign language effect in decision making. *Cognition*, *130*(2), 236-254.

Coutifaris, C. G., & Grant, A. M. (2021). Taking Your Team Behind the Curtain: The Effects of Leader Feedback-Sharing and Feedback-Seeking on Team Psychological Safety. *Organization Science*.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in experimental social psychology*, *40*, 61-149.

Culpeper, J. (2011). *Impoliteness: Using language to cause offence* (Vol. 28). Cambridge University Press.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management science*, *32*(5), 554-571.

Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment and Decision making*, *8*(5), 512.

Demszky, D., Liu, J., Mancenido, Z., Cohen, J., Hill, H., Jurafsky, D., & Hashimoto, T. (2021). Measuring Conversational Uptake: A Case Study on Student-Teacher Interactions. *arXiv preprint arXiv:2106.03873.*

Detert, J. R., & Burris, E. R. (2007). Leadership behavior and employee voice: Is the door really open? *Academy of Management Journal*, *50*(4), 869-884.

Detert, J. R., Burris, E. R., Harrison, D. A., & Martin, S. R. (2013). Voice flows to and around leaders: Understanding when units are helped or hurt by employee voice. *Administrative Science Quarterly*, *58*(4), 624-668.

DeVaul, R. A., Jervey, M. S., Chappell, M. D., Caver, P., & O'Keefe, S. (1987). of Initially Rejected Students. *JAMA*, *257*, 47-51.

Dubovsky, S. L., Gendel, M. H., Dubovsky, A. N., Levin, R., Rosse, J., & House, R. (2008). Can admissions interviews predict performance in residency?. *Academic Psychiatry*, *32*(6), 498-503.

Dunbar, R. I. M., Robledo, J. P., Tamarit, I., Cross, I., & Smith, E. (2022). Nonverbal auditory cues allow relationship quality to be inferred during conversations. *Journal of Nonverbal Behavior*, 1-18.

Dutton, J. E., Ashford, S. J., O'neill, R. M., Hayes, E., & Wierba, E. E. (1997). Reading the wind: How middle managers assess the context for selling issues to top managers. *Strategic Management Journal, 18(5), 407–425.*

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative science quarterly*, *44*(2), 350-383.

Edmondson, A. C., & Bransby, D. P. (2023). Psychological Safety Comes of Age: Observed Themes in an Established Literature. *Annual Review of Organizational Psychology and Organizational Behavior*, *10*, 55-78.

Edmondson, A. C. (2003). Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of management studies,* 40(6), 1419-1452.

Effiong, O. (2016). Getting them speaking: Classroom social factors and foreign language anxiety. *Tesol Journal*, *7*(1), 132-161.

Ekman, P. (2016). What scientists who study emotion agree about. *Perspectives on psychological science*, *11*(1), 31-34.

Eisenberger, N. I., & Lieberman, M. D. (2005). Broken hearts and broken bones: The neurocognitive overlap between social pain and physical pain. *The social outcast Ostracism, social exclusion rejection, and bullying*, 109-127.

Elfenbein, H. A., Barsade, S. G., & Eisenkraft, N. (2015). The social perception of emotional abilities: Expanding what we know about observer ratings of emotional intelligence. *Emotion*, 15(1), 17

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & biobehavioral reviews*, *24*(6), 581-604.

Fadhil, A., Schiavo, G., Wang, Y., & Yilma, B. A. (2018, May). The effect of emojis when interacting with conversational interface assisted health coaching system. In *Proceedings of the 12th EAI international conference on pervasive computing technologies for healthcare* (pp. 378-383).

Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2015). The exposure advantage: Early exposure to a multilingual environment promotes effective communication. *Psychological science*, *26*(7), 1090-1097.

Fast, N. J., Burris, E. R., & Bartel, C. A. (2014). Managing to stay in the dark: Managerial self-efficacy, ego defensiveness, and the aversion to employee voice. *Academy of Management Journal*, *57*(4), 1013-1034.

Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in the first-and second-language reading of fluent bilinguals. *Memory & cognition*, *11*, 565-574.

Fischer, T. (2018). Leadership: Processes and ambiguities (*Doctoral dissertation, Université de Lausanne, Faculté des hautes études commerciales)*.

Fischer, T., & Sitkin, S. B. (2023). Leadership styles: a comprehensive assessment and way forward. *Academy of Management Annals*, *17*(1), 331-372.

Frazier, M. L., Fainshmidt, S., Klinger, R. L., Pezeshkan, A., & Vracheva, V. (2017). Psychological safety: A meta-analytic review and extension. *Personnel Psychology*, *70*(1), 113-165.

Freiermuth, M. R. (2001). Native speakers or non-native speakers: Who has the floor? Online and face-to-face interaction in culturally mixed small groups. *Computer assisted language learning*, *14*(2), 169-199.

Friedman, N. 1967. The social nature of psychological research: The psychological experiment as a social interaction. New York, NY: *Basic Books*

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic bulletin & review*, *5*(3), 490-495.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, *133*(4), 694.

Gallus, J., & Bhatia, S. (2020). Gender, power and emotions in the collaborative production of knowledge: A large-scale analysis of Wikipedia editor conversations. *Organizational Behavior and Human Decision Processes*, *160*, 115-130.

Gao, L., Janssen, O., & Shi, K. (2011). Leader trust and employee voice: The moderating role of empowering leader behaviors. *The Leadership Quarterly*, 22(4), 787-798.

Geipel, J., Hadjichristidis, C., & Surian, L. (2015). How foreign language shapes moral judgment. *Journal of experimental social psychology*, *59*, 8-17.

Gerber, J. P., Chang, S. H., & Reimel, H. (2017). Construct validity of Williams' ostracism needs threat scale. *Personality and Individual Differences*, *115*, 50-53.

Gkorezis, P., Panagiotou, M., & Theodorou, M. (2016). Workplace ostracism and employee silence in nursing: the mediating role of organizational identification. *Journal of advanced nursing*, *72*(10), 2381-2388.

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American economic review*, *90*(4), 715-741.

Goodman, P. S., & Leyden, D. P. (1991). Familiarity and group productivity. *Journal of applied psychology*, 76(4), 578.

Goffman, E. (1959). The moral career of the mental patient. *Psychiatry*, *22*(2), 123-142.

Gottman, J. M., & Porterfield, A. L. (1981). Communicative competence in the nonverbal behavior of married couples. *Journal of Marriage and the Family*, 817-824.

Grant, A. M., & Ashford, S. J. (2008). The dynamics of proactivity at work. *Research in organizational behavior*, *28*, 3-34.

Gruenfeld, D. H., Mannix, E. A., Williams, K. Y., & Neale, M. A. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational behavior and human decision processes*, *67*(1), 1-15.

Hall, J. A., Bernieri, F. J., & Carney, D. R. (2005). Nonverbal behavior and interpersonal sensitivity. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 237–281). Oxford University Press.

Harrison, D. A., Harrison, T., & Shaffer, M. A. (2019). Strangers in strained lands: Learning from workplace experiences of immigrant employee.

Hart, E., VanEpps, E. M., & Schweitzer, M. E. (2021). The (better than expected) consequences of asking sensitive questions. *Organizational Behavior and Human Decision Processes,* 162, 136-154.

Hartgerink, C. H., Van Beest, I., Wicherts, J. M., & Williams, K. D. (2015). The ordinal effects of ostracism: A meta-analysis of 120 Cyberball studies. *PloS one*, *10*(5), e0127002.

He, W., Fehr, R., Yam, K. C., Long, L. R., & Hao, P. (2017). Interactional justice, leader–member exchange, and employee performance: Examining the moderating role of justice differentiation. *Journal of Organizational Behavior*, 38(4), 537-557.

Holroyd, A., Rich, C., Sidner, C. L., & Ponsler, B. (2011, July). Generating connection events for human-robot collaboration. In *2011 RO-MAN* (pp. 241-246). IEEE.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern language journal*, *70*(2), 125-132.

Howell, T. M., Harrison, D. A., Burris, E. R., & Detert, J. R. (2015). Who gets credit for input? Demographic and structural status cues in voice recognition. *Journal of Applied Psychology*, *100*(6), 1765.

Hsiung, H. H. (2012). Authentic leadership and employee voice behavior: A multi-level psychological process. *Journal of business ethics*, *107*(3), 349-361.

Huang, L., Frideger, M., & Pearce, J. L. (2013). Political skill: Explaining the effects of nonnative accent on managerial hiring and entrepreneurial investment decisions. *Journal of Applied Psychology*, *98*(6), 1005.

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.

Hughes, S. M., Mogilski, J. K., & Harrison, M. A. (2014). The perception and parameters of intentional voice manipulation. *Journal of Nonverbal Behavior*, *38*(1), 107-127.

Hyatt, D. E., & Ruddy, T. M. (1997). An examination of the relationship between work group characteristics and performance: Once more into the breech. *Personnel Psychology,* 50(3), 553-585.

Iannone, N. E., Kelly, J. R., & Williams, K. D. (2018). "Who's that?": The negative consequences of being out of the loop on pop culture. *Psychology of Popular Media Culture*, *7*(2), 113.

Imbir, K. K. (2017). Valence and origin of emotional words influence on subsequent perception of ambiguous stimuli in terms of competence versus warmth. *Journal of Psycholinguistic Research*, *46*, 1549-1571.

Jarick, M., & Kingstone, A. (2015). The duality of gaze: eyes extract and signal social information during sustained cooperative and competitive dyadic gaze. *Frontiers in psychology*, *6*, 1423.

Jeong, M., Minson, J., Yeomans, M., & Gino, F. (2019). Communicating with warmth in distributive negotiations is surprisingly counterproductive. *Management Science*, 65(12), 5813-5837.

Jiang, X., & Pell, M. D. (2015). On how the brain decodes vocal cues about speaker confidence. *Cortex*, *66*, 9-34.

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, *79*(1), 57-62.

Jones, E. E., & Kelly, J. R. (2010). "Why am I out of the loop?" Attributions influence responses to information exclusion. *Personality and Social Psychology Bulletin*, *36*(9), 1186-1201.

Jones, R., & Cooper, J. (1971). Mediation of experimenter effects. *Journal of Personality and Social Psychology,* 1971, 20, 70-74.

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal,* 33(4), 692-724.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kampe, K. K., Frith, C. D., Dolan, R. J., & Frith, U. (2001). Reward value of attractiveness and gaze. *Nature*, *413*(6856), 589-589.

Kausel, E. E., Culbertson, S. S., & Madrid, H. P. (2016). Overconfidence in personnel selection: When and why unstructured interview information can hurt hiring decisions. *Organizational Behavior and Human Decision Processes*, *137*, 27-44.

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological science*, *23*(6), 661-668.

Khalid, S., Deska, J. C., & Hugenberg, K. (2016). The eyes are the windows to the mind: Direct eye gaze triggers the ascription of others' minds. *Personality and Social Psychology Bulletin*, *42*(12), 1666-1677.

Kleck, R. E., & Nuessle, W. (1968). Congruence between the indicative and communicative functions of eye contact in interpersonal relations. *British Journal of Social and Clinical Psychology*, *7*(4), 241-246.

Koch, S. C., Baehne, C. G., Kruse, L., Zimmermann, F., & Zumbach, J. (2010). Visual Dominance and Visual Egalitarianism: Individual and Group-Level Influences of Sex and Status in Group Interactions. *Journal of Nonverbal Behavior*, 34(3), 137–153.

Koopman, J., Howe, M., Hollenbeck, J. R., & Sin, H.-P. (2015). Small sample mediation testing: Misplaced confidence in bootstrapped confidence intervals. *Journal of Applied Psychology*, 100(1), 194-202.

Kotlyar, I., & Ariely, D. (2013). The effect of nonverbal cues on relationship formation. *Computers in Human Behavior*, *29*(3), 544-551.

Kumar, A., & Epley, N. (2021). It's surprisingly nice to hear you: Misunderstanding the impact of communication media can lead to suboptimal choices of how to connect with others. *Journal of Experimental Psychology: General*, *150*(3), 595.

Laplante, D., & Ambady, N. (2003). On how things are said: Voice tone, voice intensity, verbal content, and perceptions of politeness. *Journal of language and social psychology*, *22*(4), 434-441.

LePine, J. A., & Van Dyne, L. (1998). Predicting voice behavior in work groups. *Journal of applied psychology*, *83*(6), 853.

LePine, J. A., & Van Dyne, L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: evidence of differential relationships with big five personality characteristics and cognitive ability. *Journal of applied psychology*, *86*(2), 326.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of experimental social psychology*, *46*(6), 1093-1096.

Liang, H. L., & Yeh, T. K. (2020). The effects of employee voice on workplace bullying and job satisfaction: The mediating role of leader–member exchange. *Management Decision*.

Liberman, Z., Woodward, A. L., Keysar, B., & Kinzler, K. D. (2017). Exposure to multiple languages enhances communication skills in infancy. *Developmental science*, *20*(1), e12420.

Li, Y., Packard, G., & Berger, J. (*working*). Conversational Dynamics: When Does Employee Language Matter?

Li P, Jeong H. (2020) The social brain of language: grounding second language learning in social interaction. Nature. doi:10.1038/s41539-020-0068-7.

Li, C. F., & Tian, Y. Z. (2016). Influence of workplace ostracism on employee voice behavior. *American Journal of Mathematical and Management Sciences*, *35*(4), 281-296.

Livingston, B. A., Schilpzand, P., & Erez, A. (2017). Not what you expected to hear: Accented messages and their effect on choice. *Journal of Management*, *43*(3), 804-833.

Locke, C. C., & Anderson, C. (2015). The downside of looking like a leader: Power, nonverbal confidence, and participative decision-making. *Journal of Experimental Social Psychology*, *58*, 42-47.Lu, J. G., Nisbett, R. E., & Morris, M. W. (2022). The surprising underperformance of East Asians in US law and business schools: The liability of low assertiveness and the ameliorative potential of online classrooms. *Proceedings of the National Academy of Sciences*, 119(13), e2118244119.

MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere*, *24*, 41.

Mariadassou, S., Bechler, C. J., & Levav, J. (2022). The Effect of Auditory and Visual Recommendations on Choice. *Psychological Science*, *34*(1), 47-59.

Martin, S.R., Harrison, S.H., Hoopes, C., Schroeder, J., & Belmi, P.R. (2022). Talking shop: An exploration of how talking about work affects our initial interactions. *Organizational Behavior and Human Decision Processes, 168*, 1-16.

Mastroianni, A. M., Cooney, G., Boothby, E. J., & Reece, A. G. (2021). The liking gap in groups and teams. *Organizational Behavior and Human Decision Processes*, *162*, 109-122.

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices. *PloS one*, *9*(3), e90779.

McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004). Neural correlates of behavioral preference for culturally familiar drinks. *Neuron*, *44*(2), 379-387.

McCroskey, J. C., Richmond, V. P., Sallinen, A., Fayer, J. M., & Barraclough, R. A. (1995). A cross-cultural and multi-behavioral analysis of the relationship between nonverbal immediacy and teacher evaluation. *Communication Education*, *44*(4), 281-291.

Mehrabian, A. (1972). *Nonverbal Communication*. Routledge.

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of applied psychology*, *94*(2), 535.

Miceli, M. P., & Near, J. P. (1992). *Blowing the whistle: The organizational and legal implications for companies and employees*. Lexington Books.

Mobasseri, S., Stein, D. H., & Carney, D. R. (2022). The accurate judgment of social network characteristics in the lab and field using thin slices of the behavioral stream. *Organizational Behavior and Human Decision Processes*, 168, 104103.

Morrison, E. W. (2011). Employee voice behavior: Integration and directions for future research. *Academy of Management Annals*, *5*(1), 373-412.

Morrison, E. W., & Milliken, F. J. (2000). Organizational silence: A barrier to change and development in a pluralistic world. *Academy of Management review*, *25*(4), 706-725.

Nakamura, L. (2013). *Cybertypes: Race, ethnicity, and identity on the Internet*. Routledge.

Nakamura, L. (2015). The unwanted labour of social media: Women of colour call out culture as venture community management. *New Formations*, *86*(86), 106-112.

Neeley, T. (2017). The language of global success. In *The Language of Global Success*. Princeton University Press.

Neeley, T. B., Hinds, P. J., & Cramton, C. D. (2012). The (un) hidden turmoil of language in global collaboration. *Organizational Dynamics*, *41*(3), 236-244.

Nembhard, I. M., & Edmondson, A. C. (2006). Making it safe: The effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *27*(7), 941-966.

Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive psychology*, *13*(2), 248-277.

"Open-Source Psychometric Project." The Multidimensional Introversion-Extraversion Scales https://openpsychometrics.org/tests/MIES/

Packard, G., & Berger, J. (2021). How Concrete Language Shapes Customer Satisfaction. *Journal of Consumer Research*, *47*(5), 787-806.

Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of personality and social psychology*, 112(4), 642.

Pavalanathan, U., & Eisenstein, J. (2016). More emojis, less:) The competition for paralinguistic function in microblog writing. *First Monday*.

Pavlenko, A. (2005). Bilingualism and Thought. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 433–453). Oxford University Press.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015.*

Pliatsikas, C., Pereira Soares, S. M., Voits, T., DeLuca, V., & Rothman, J. (2021). Bilingualism is a long-term cognitively challenging experience that modulates metabolite concentrations in the healthy brain. *Scientific Reports*, *11*(1), 7090.

Porter, S., Ten Brinke, L., Baker, A., & Wallace, B. (2011). Would I lie to you? "Leakage" in deceptive facial expressions relates to psychopathy and emotional intelligence. *Personality and Individual Differences*, *51*(2), 133-137.

Prisbell, M. (1985). Assertiveness, shyness and nonverbal communicative behaviors. *Communication Research Reports*, *2*(1).

Rashidi, N., Yamini, M., & Shafiei, E. (2011). Oral communication apprehension and affective factors: Self-esteem and introversion/extroversion. *Journal of English Language Teaching and Learning.*

Riasati, M. J. (2012). EFL learners' perception of factors influencing willingness to speak English in language classrooms: A qualitative study. *World Applied Sciences Journal*, *17*(10), 1287-1297.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and social psychology review*, 8(4), 364-382.

Ridgeway, C. L., Berger, J., & Smith, L. (1985). Nonverbal cues and status: An expectation states approach. *American Journal of Sociology*, 90(5), 955-978.

Ridgeway, C. L., & Walker, H. A. (1995). Status structures. *Sociological perspectives on social psychology*, 281, 310.

Riva, P., Wirth, J. H., & Williams, K. D. (2011). The consequences of pain: The social and physical pain overlap on psychological responses. *European Journal of Social Psychology*, 41(6), 681-687.

Rogers, T., Ten Brinke, L., & Carney, D. R. (2016). Unacquainted callers can predict which citizens will vote over and above citizens' stated self-predictions. *Proceedings of the National Academy of Sciences*, *113*(23), 6449-6453.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech communication*, *40*(1-2), 227-256.

Schmeichel, B. J., & Tang, D. (2015). Individual differences in executive functioning and their relationship to emotional processes and responses. *Current Directions in Psychological Science*, *24*(2), 93-98.

Schroeder, J. (*working*). The Role of Conversation Structure in Conflict: Speaking is More Effective for Resolving Conflict than Writing

Schroeder, J., Kardas, M., & Epley, N. (2017). The humanizing voice: Speech reveals, and text conceals, a more thoughtful mind in the midst of disagreement. *Psychological science*, *28*(12), 1745-1762.

Schroeder, S. R. (2018). Do bilinguals have an advantage in theory of mind? A meta-analysis. *Frontiers in Communication*, *3*, 36.

Schroeder, J., Lyons, D., & Epley, N. (2021). Hello, stranger? Pleasant conversations are preceded by concerns about starting one. *Journal of Experimental Psychology: General.*

Shepherd, S. V., Deaner, R. O., & Platt, M. L. (2006). Social status gates social attention in monkeys. *Trends Cogn. Sci*, *4*, 138-147.

Sherf, E. N., Parke, M. R., & Isaakyan, S. (2021). Distinguishing voice and silence at work: Unique relationships with perceived impact, psychological safety, and burnout. *Academy of Management Journal*, *64*(1), 114-148.

Shim, S. H., Livingston, R., Phillips, K. W., & Lam, S. S. (2020). The Impact of Leader Eye Gaze on Disparity in Member Influence: Implications for Process and Performance in Diverse Groups. *Academy of Management Journal*.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological bulletin*, *119*(1), 3.

Stam, D., Van Knippenberg, D., Wisse, B., & Nederveen Pieterse, A. (2018). Motivation in words: Promotion-and prevention-oriented leader communication in times of crisis. *Journal of management*, *44*(7), 2859-2887.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and brain sciences*, *23*(5), 701-717.

Stein, R. T. (1975). Identifying emergent leaders from verbal and nonverbal communications. *Journal of Personality and Social Psychology*, *32*(1), 125.

Sternglanz, R. W., & DePaulo, B. M. (2004). Reading nonverbal cues to emotions: The advantages and liabilities of relationship closeness. *Journal of Nonverbal Behavior*, *28*, 245-266.

Strongman, K. T., & Champness, B. G. (1968). Dominance hierarchies and conflict in eye contact. *Acta Psychologica*, *28*, 376-386.

Stivers, T. (2021). Is conversation built for two? The partitioning of social interaction. *Research on Language and Social Interaction*, *54*(1), 1-19.

Stuhlmacher, A. F., & Citera, M. (2005). Hostile behavior and profit in virtual negotiation: A meta-analysis. *Journal of Business and Psychology*, *20*, 69-93.

Sun, K. Q., & Slepian, M. L. (2020). The conversations we seek to avoid. *Organizational Behavior and Human Decision Processes*, *160*, 87-105.

Swaab, Roderick I., Robert B. Lount Jr, Seunghoo Chung, and Jeanne M. Brett. (2021)."Setting the stage for negotiations: How superordinate goal dialogues promote trust and joint gain in negotiations between teams." *Organizational Behavior and Human Decision Processes* 167: 157-169.

Swann Jr, W. B., & Pittman, T. S. (1977). Initiating play activity of children: The moderating influence of verbal cues on intrinsic motivation. *Child development*, 1128-1132.

Talaifar, S., Buhrmester, M. D., Ayduk, Ö., & Swann Jr, W. B. (2021). Asymmetries in mutual understanding: People with low status, power, and self-esteem understand better than they are understood. *Perspectives on Psychological Science*, *16*(2), 338-357.

Templeton, E. M., Chang, L. J., Reynolds, E. A., LeBeaumont, M. D. C., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences*, 119(4).

Ten Berge, M. T., & De Raad, B. D. (2001). The construction of a joint taxonomy of traits and situations. *European Journal of Personality*, 15(4), 253-276.

Ten Berge, M. A., & De Raad, B. (2002). The structure of situations from a personality perspective. *European Journal of Personality*, 16(2), 81-102.

Ten Brinke, L., MacDonald, S., Porter, S., & O'connor, B. (2012). Crocodile tears: facial, verbal and body language behaviours associated with genuine and fabricated remorse. *Law and human behavior*, *36*(1), 51.

Ten Brinke, L., & Weisbuch, M. (2020). How verbal-nonverbal consistency shapes the truth. *Journal of Experimental Social Psychology*, *89*, 103978.

Tenney, E. R., Meikle, N. L., Hunsaker, D., Moore, D. A., & Anderson, C. (2019). Is overconfidence a social liability? The effect of verbal versus nonverbal expressions of confidence. *Journal of personality and social psychology*, *116*(3), 396.

Thompson, D., & Filik, R. (2016). Sarcasm in written communication: Emoticons are efficient markers of intention. *Journal of Computer-Mediated Communication*, *21*(2), 105-120.

Truong, M., Fast, N. J., & Kim, J. (2020). It's not what you say, it's how you say it: Conversational flow and networking success. *Organizational Behavior and Human Decision Processes*, 158, 1-10/

Tsay, C. J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences*, *110*(36), 14580-14585.

Tsay, C. J. (2014). The vision heuristic: Judging music ensembles by sight alone. *Organizational Behavior and Human Decision Processes*, 124(1), 24-33.

Tuovinen, S., Tang, X., & Salmela-Aro, K. (2020). Introversion and social engagement: Scale validation, their interaction, and positive association with self-esteem. *Frontiers in Psychology*, *11*, 3241.

Twenge, J. M., Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., & Bartels, J. M. (2007). Social exclusion decreases prosocial behavior. *Journal of personality and social psychology*, *92*(1), 56.

US Census, Language Use in the United States, 2019, https://www.census.gov/library/stories/2022/12/languages-we-speak-in-united-states.html

Valsesia, F., Nunes, J. C., & Ordanini, A. (2021). I am not talking to you: Partitioning an audience in an attempt to solve the self-promotion dilemma. *Organizational Behavior and Human Decision Processes*, *165*, 76-89.

Van Dyne, L., & LePine, J. A. (1998). Helping and voice extra-role behaviors: Evidence of construct and predictive validity. *Academy of Management journal*, *41*(1), 108-119.

Van Heuven, W. J., & Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain research reviews*, *64*(1), 104-122.

Van Knippenberg, D., & Sitkin, S. B. (2013). A critical assessment of charismatic—transformational leadership research: Back to the drawing board?. *Academy of management Annals*, *7*(1), 1-60.

Van Vugt, M., Hogan, R., & Kaiser, R. B. (2008). Leadership, followership, and evolution: some lessons from the past. *American Psychologist*, *63*(3), 182.

Van Zant, Alex B., and Jonah Berger. "How the voice persuades." *Journal of personality and social psychology* 118.4 (2020): 661.

Voigt, R., Camp, N. P., Prabhakaran, V., Hamilton, W. L., Hetey, R. C., Griffiths, C. M., ... & Eberhardt, J. L. (2017). Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences,* 114(25), 6521-6526.

Walumbwa, F. O., & Schaubroeck, J. (2009). Leader personality traits and employee voice behavior: mediating roles of ethical leadership and work group psychological safety. *Journal of applied psychology*, *94*(5), 1275.

Wesselmann, E. D., Cardoso, F. D., Slater, S., & Williams, K. D. (2012). To be looked at as though air: Civil attention matters. *Psychological Science*, *23*(2), 166-168.

Williams, K. D. (2009). Ostracism: A temporal need-threat model. *Advances in experimental social psychology*, *41*, 275-314.

Williams, K. D., Cheung, C. K., & Choi, W. (2000). Cyberostracism: effects of being ignored over the Internet. *Journal of personality and social psychology*, *79*(5), 748.

Willard, G., Isaac, K.-J., & Carney, D. R. (2015). Some evidence for the nonverbal contagion of racial bias. *Organizational Behavior and Human Decision Processes, 128,* 96–107. https://doi.org/10.1016/j.obhdp.2015.04.002.

Wirth, J. H., Sacco, D. F., Hugenberg, K., & Williams, K. D. (2010). Eye gaze as relational evaluation: Averted eye gaze leads to feelings of ostracism and relational devaluation. *Personality and Social Psychology Bulletin*, *36*(7), 869-882.

Wohltjen, S., & Wheatley, T. (2021). Eye contact marks the rise and fall of shared attention in conversation. *Proceedings of the National Academy of Sciences*, *118*(37).

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*(6004), 686-688.

Wu, W., Qu, Y., Zhang, Y., Hao, S., Tang, F., Zhao, N., & Si, H. (2019). Needs frustration makes me silent: Workplace ostracism and newcomers' voice behavior. *Journal of Management & Organization*, *25*(5), 635-652.

Yeomans, M., Kantor, A., & Tingley, D. (2018). The politeness Package: Detecting Politeness in Natural Language. *R Journal*, *10*(2).

Yeomans, M., Schweitzer, M. E., & Brooks, A. W. (2021). The Conversational Circumplex: Identifying, Prioritizing, and Pursuing Informational and Relational Motives in Conversation. *Current Opinion in Psychology*, *44,* 293-302.

Yeomans, M. & Brooks, A.W. (*Working*) Topic preference detection in conversation: A novel approach to understand perspective taking.

Young, S. G., Slepian, M. L., Wilson, J. P., & Hugenberg, K. (2014). Averted eye-gaze disrupts configural face encoding. *Journal of Experimental Social Psychology*, *53*, 94-99.

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.

TABLES, FIGURES & APPENDICES

Chapter 1 Appendices

Appendix 1A: Supplemental Study (Eye Gaze and Airtime in Dyads)

In an exploratory (not pre-registered) supplemental study, we explore the relationship between eye gaze and speaking. Participants came to the lab in pairs, and were video-taped having a live conversation. They completed survey measures after the conversation about perceptions of their partner. After the conversation, third-parties coded the videos for how much eye contact each participant received, and transcripts were assessed for amount of speaking. We investigated potential correlational relationships between eye gaze received from their partner and amount of speaking.

Method

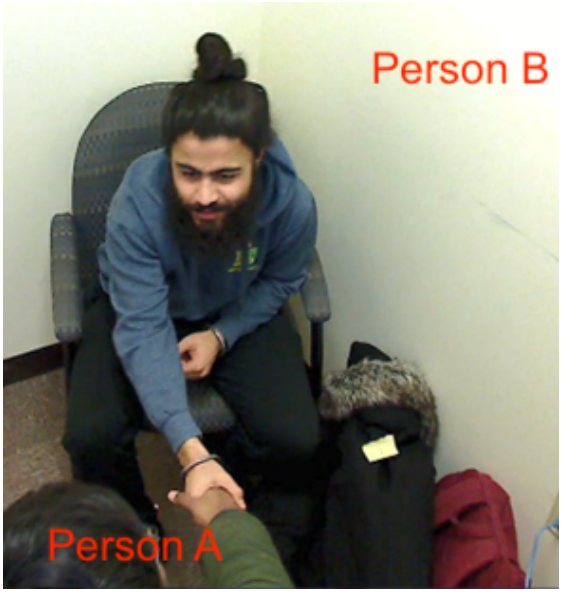
Participants: Lab: We aimed to recruited 160 dyads to participate in live conversations. Ultimately 162 dyads participated and were included in our analysis (N = 324; 55% female; M_{age} = 23 years). All conversation partners were strangers.

Participants: Coders. A total of 419 participants (37% female, M_{age} = 34 years) were recruited through Amazon's Mechanical Turk (mTurk) to complete a 15-minute online survey.

Design: The experiment was originally designed to test if participants could tell who was listening to them. As such, there were three conditions.

Procedure: Lab. When they arrived at the lab, participants were randomly paired with a partner and were told that they would spend 5 minutes in conversation. Specifically, we instructed participants to get to know their partner and determine whether or not they would make good roommates. Within each dyad, one participant sat with their back to a TV which was

in the room, was instructed to ignore the videos playing behind them. This was the "unmanipulated partner" (Person A). There are no videos of this partner, but we know how much gaze they received. The other participant (the "manipulated" partner, Participant B) was seated in full view of the TV screen, and was randomly assigned to pay attention to the TV or not. We have video recordings of Person B, and these video recordings were later coded for eye gaze and other verbal and nonverbal cues. We controlled for condition in our analysis because we were not interested in the effect of condition on outcomes for the purposes of this analysis.



Supplemental Figure 1.1: Screenshot of the video recording of the participants. Person B (manipulated) is being recorded, and their body language and eye gaze was subsequently coded. Person A is the recipient of Person B's eye gaze, and our study evaluates whether their speaking time is associated with their conversation partner's gaze.

Procedure: Coders. After all of the videos were collected, mTurk participants watched and evaluated the videos. In the videos, the camera is positioned facing Person B partner, offering the viewer an unobstructed frontal view of this participant. Participants on mTurk were recruited to act like coders: they watched a video from one of the interactions and were asked to

report the extent to which the participant in the video engaged in several behaviors. To gather fine-grained coding of the videos, participants were asked to watch each video divided into 1-minute segments. After each 1-minute segment, the video would pause and the participant would be asked to report how frequently the individual engaged in various behaviors. Then the video would resume, and pause again after another minute.

Dependent Measures:

Eye Gaze: For each minute-long video segment, participants rated the extent to which Person B made eye contact with Person A on a scale from 1: "Not at all" to 4: "A lot." This process repeated until the video ended.

Exploratory verbal and nonverbal cues: To benchmark eye gaze against other verbal and nonverbal cues, coders also evaluated cues such as: the extent to which Person B interrupted Person A, the extent to which Person B nodded, leaned forward, and smiled at Person A, and verbally affirmed Person A ("yeah", "totally"). All scales were from 1: "Not at all" to 4: "A lot."

Participation: To evaluate how much Participant A spoke, we calculated their word count during the conversation (note that these transcripts did not have time stamps).

Exploratory measures: After the conversation, participants were asked to answer a few questions about their experience during the conversation. This included the extent to which they believed that their partner was tolerant, warm, and sincere. All response scales ranged from 1= *extremely disagree* to 6 = *extremely agree*.

Results

Participation. We found evidence supporting Hypothesis 2a: in a linear regression, the amount of eye gaze Person B gives to Person A significantly and positively predicts Person A's participation, $b = 70.8$, $t = 2.07$, $p = .046$. Controlling for experimental condition shows a similar

although marginal pattern of receiving eye gaze on increased speaking up: $b = 68.4$, $t = 1.95$, $p = .053$.

Interpersonal evaluations of conversation partner. Our results show that the amount of eye gaze Person B gives to Person A significantly and positively predicts Person A's ratings of how tolerant, sincere, and warm Person B is (controlling for condition, b's > .57, t's > 2.34, p's > .02).

Other cues. Benchmarking eye gaze against other verbal and nonverbal cues, we found that interruptions, leaning forward, smiling, and nodding by Person B did not significantly predict Person A's speaking. However, receiving verbal affirmations (such as "yeah", "totally") significantly and positively predict Person A's speaking (controlling for condition, $b = 66.8$, $t = 2.34$, $p = .02$),

Discussion

In our supplemental study, we find that receiving more eye gaze is associated with speaking more during dyadic conversation. As an exploratory analysis, we find that receiving more eye gaze is significantly positively related to subjective perceptions of the conversation partner, such as feeling like they are warm and tolerant, which are conceptually related to psychological safety and ostracism. Moreover, eye gaze predicts speaking above and beyond other cues such as interruptions, leaning forward, smiling, and nodding, which did not significantly predict speaking. Verbal affirmations, which are similarly a conversational cue of attention and respect, did predict speaking.

These findings offer preliminary evidence of the relationship between eye gaze and speaking. Notably we do not measure leader eye gaze, as the dyads are peers, nor do we

manipulate eye gaze to understand the direction of causality. We address these concerns in Studies 2A-B.

Appendix 1B

Study 1 Coding Example:

This table shows sample Rows from a transcript coded for voice. Person A has received gaze 5 times, and Person B has received gaze 3 times, and Person C has received gaze 2 times.

Speaker	Time Stamp	Text	Where is Person A looking?	Where is Person B looking? (Group leader)	Where is Person C looking?	Voice
A	1:07	I guess we can just kind of discuss each candidate and their strengths and weaknesses?	At Person C	At Person A	At Person A	0
C	1:30	Sounds good. So candidate A speaks five languages fluently so that feels important. I	Down at their sheet	Down at their sheet	At Person A	0
A	1:43	Person A speaks five languages? Wait, are you sure? I don't have that	Down at their sheet	Down at their sheet	Down at their sheet	0
C	1:46	Interesting. Do we all have different information?	At Person C	Down at their sheet	At Person A	0
B	…2:52	Yes it seems like Candidate B is the best overall. He can fly a MIG, and has black belt in Taekwondo.	At Person B	At Person A	At Person B	0
C	2:58	I see here that Candidate B has had a burnout in the past- maybe he is not the best candidate.	Down at their sheet	Down at their sheet	At Person B	1
A	3:00	What else do we know about Candidate B?	At Person C	Down at their sheet	At Person A	0

C	2:02	I see that he is born in Cleveland.	Down at their sheet	Down at their sheet	At Person A	0

Study 1 Additional Analyses

In order to understand causality, we also conducted a lagged analysis. We found that the amount of gaze received in the first half of the conversation significantly predicted the number of speaking time in the last half of the conversation, $b = .40$, $p < .001$. The amount of gaze received during the first half of the conversation also predicts the speaking time during the first half, but the co-efficient smaller: $b = .25$, $p < .001$. This pattern demonstrates that gaze in T1 had a stronger effect on speaking time in T2 compared to the effect that gaze in T2 had on speaking in T1. We repeated this analysis varying the percentages of time compared and found the same pattern of a bigger effect size for T1 gaze predicting T2 speaking compared to T2 speaking predicting T1 gaze in the table below:

Supplemental Table 1.1: Study 1 Lagged Analyses

	T1 gaze → T2 speaking	***T2 speaking → T1 gaze***
First 10%/ last 10%	b= .15, p<.001	b=.14, p<.001
First 20%/ last 20%	b= .23, p<.001	b=.17 p<.001
First 30%/ last 30%	b= .31, p<.001	b= .19, p<.001
First 40%/ last 40%	b= .33, p<.001	b= .25, p<.001
First 50%/ last 50%	b= .40, p<.001	b= .25, p<.001

Appendix 1-C

Additional Measures from Study 2A & 2B

Survey Items reported in the main text:

Ostracism survey items (adapted from Gerber et al., 2017)

1. During this conversation I felt poorly accepted by the other members of the group.
2. During this conversation I felt somewhat inadequate.
3. During this conversation I felt like an outsider.
4. During this conversation I felt that other people in the group did not perceive me as worthy and likable.
5. During this conversation I felt somewhat frustrated.
6. During this conversation I felt good about myself. (R)
7. During this conversation I felt in control. (R)

Voice survey items (adapted from LePine & Van Dyne, 1998)

1. During this conversation I felt I was able to express my views and feelings with the other people in the conversation.
2. During this conversation, I suggested new ideas which were beneficial to the team.
3. During this conversation, I felt comfortable speaking up.
4. During this conversation, I communicated my opinion even if my opinion was different and others disagreed with me.

Introversion/ Extraversion survey items (adapted from the MIES scale)

1. I don't mind being the center of attention.
2. I talk to a lot of different people at parties.
3. I prefer to socialize 1 on 1 than with a group. (R)

4. I don't like to draw attention to myself. (R)

Additional methodological detail

Part 1. At the start of the meeting, all participants joined the Zoom meeting. Once at least three participants and the confederate joined the Zoom meeting, the study could begin. The researcher changed the Zoom names to reflect the role that each participant (and the confederate) was assigned. Once assigned their roles, the participants were sent their individual task instructions through private Zoom chat and instructed to read through these instructions separately for 7 minutes. After seven minutes, the researcher then prompted the participants to do a comprehension check, where participants were asked to think about and write down their role and supporting evidence for their role. After two minutes of reflection, the researcher then instructed the three participants and one confederate to begin their 15-minute discussion of the task. After the conversation, participants were directed to Part 2, where they each completed a Qualtrics survey that included individual questions.

At the end of 7 minutes, the researcher then instructed the participants and confederate to return to the survey and complete the comprehension check. At the end of 2 minutes, the researcher then instructed the participants to begin their 15-minute conversation surrounding the task. At the end of 15 minutes, the researcher informed the participants and confederate that their time was up and began Part 2.

These were the instructions that the researcher read for Part 1 of this study:

> Part 1 Task Instructions: “Thank you for participating today. Before getting started, please turn on your videos. On the top of your screen, please make sure that you are in ‘gallery’ view. Great! You will now have 7 minutes to read the task instructions. These will be sent to you via private message in the chat. Then you will have 15 minutes to discuss the task with the three other people.
>
> I will send you a document via Zoom chat in a moment. This has the task instructions on it. Please read the whole task and try to remember it. You can refer to the document again

if you need to or if you forget something, but please try your best to look at and engage with your other group members during the discussion, and try not refer to the document too much during the conversation. Does this make sense? Did everybody receive their private message on Zoom chat? Please give me a thumbs up or say yes.

There is one more step before you can start reading. Please check the Zoom chat and follow the link there. Let me know when you are able to open the link. This has the consent form. Please pull up the form in another tab on your computer and sign it. Let me know if you have any questions.

Please pay close attention to the next few instructions that I will step you through. Your group number is ___. Please input your group number now. Each of you also has a participant role as your Zoom name. Please choose your participant role from the dropdown menu. And the next page also will ask you to input the email address that you would like to receive payment. Does everyone have these fields filled out? You should arrive at a page that tells you to STOP.

Great! Please do not close out of this survey. We ask you to keep this open because we will be asking you to return to it later. For now, please turn to the task instructions that I sent you. You will now have 7 minutes to read through the task."

Part 1 Comprehension Check: "Hi everyone! Your 7 minutes is up! At this time, please go back to the survey and answer the next two questions briefly. I will give you about 2 minutes to do so now."

Part 1 Conversation Instructions: "Once again, please do not close out of this survey. Now, you will have 15 minutes to discuss the task with the three other people. I understand that this is a lot of information to process, but please do your best to get creative and think on the spot in order to keep the conversation flowing. Please unmute yourselves and begin now."

Part 2. Finally, all participants were instructed to complete the rest of the Qualtrics survey that included individual questions. Once all participants signaled they were back to the survey, the researcher instructed participants to complete the survey. The final part of the survey included four demographic questions related to gender, age, degree, and race.

Chapter 2 Appendices

Appendix 2A: Study 1A Additional Methods and Results

Study 1A Additional Methods:

Study 1A Interview questions: The ten questions that were asked of participants were:

1. Tell me a little bit about yourself.
2. What sets you apart from others – what are skills and strengths that you have that others generally do not have?
3. What is your biggest achievement so far?
4. Tell me about a time that you dealt with a difficult social situation.
5. What are you passionate about?
6. Tell me about a time that you took a risk.
7. How do you think other people would describe you?
8. Tell me about a time that you tried something and failed.
9. What is a tough decision you had to make?
10. Where do you see yourself in 5 years?

Study 1A example LSAT question: "A kennel owner is scheduling dog play times for dogs labeled A-F on her chart. They will play in a common fenced area and the schedule must follow these guidelines: - Dogs A and B cannot play together. – Dogs C and D must play first. – Dog E cannot play with dog A. – Dog F must play last. Which of the following scenarios would follow all of the guidelines? 1. Two play times: A, C, E; then B, D, F. 2. If dog E goes home, the remaining dogs could play together. 3. If dogs A, D, and F go home, the remaining dogs could play together. 4. Two play times: B, E, F; then A, C, D."

Study 1A example GRE Math question: "Alan has two more than twice as many chocolates as does Alice, and half as many chocolates as does Nadia. If Alice has x number of chocolates, then in terms of x, how many chocolates do Alan, Alice and Nadia have?"

Appendix 2B: Study 1B Additional Results

Study 1B Additional Results: Did observing conversational content improve human prediction accuracy?

Although we were only interested in comparing the baseline to other conditions in the first analysis of Study 1B other between-condition comparisons emerged:

Liking. Participants were significantly less confident in the prosody audio only condition compared to: the full video condition (b = -.33, p =.001), the muted video with audio condition (b = -.27, p =.020), the full audio condition (b = -.39, p <.001), and the transcript condition (b = -.32, p =.002).

Writing Task 1: Termination letter. Participants in the full video condition were more confident compared to: participants in the muted video (b = .47, p <.001), prosody video (b = .42, p <.001), and prosody audio conditions (b = .44, p <.001). Participants in full audio were more confident compared to participants in prosody video (b = .41, p <.001) and prosody audio conditions (b = .44, p <.001). Lastly, participants in the transcript condition were more confident compared to: participants in prosody video (b = .34, p <.001), muted video (b = .39, p <.001), and prosody audio conditions (b = .37, p <.001).

Writing Task 2: Case. Participants in the full video condition were more confident compared to: participants in the muted video (b = .42, p <.001), prosody video (b = .36, p <.001), and prosody audio conditions (b = .44, p <.001). Participants in full audio were more confident compared to: participants in prosody video (b = .32, p =.002), muted video (b = .39, p <.001), and prosody audio conditions (b = .40, p <.001). Lastly, participants in the transcript condition were more confident compared to: participants in prosody video (b = .26, p =.03), muted video (b = .32, p =.001), and prosody audio conditions (b = .34, p <.001).

LSAT. When predicting LSAT performance, participants were more confident in the full video condition compared to: the muted video condition, (b = .30, p =.016), the prosody audio condition (b = .37, p = .001). Participants in the full audio condition were more confident compared to those compared to the prosody audio condition (b = .35, p = .002). Lastly, those in the transcript condition were more confident compared to those in the prosody audio condition (b =.33, p =.005).

Bonus Distribution Task. Participants were more confident in the full video condition compared to the prosody video condition (b = .30, p = .023) and the prosody audio condition (b = .36, p = .002). Participants in the transcript condition were more confident compared to the prosody audio condition (b = .28, p =.044).

Slider Task. The only significant differences between conditions with regards to confidence in the slider task predictions were that participants were more confident in the full video condition compared to the prosody video and prosody audio condition (b = .35, p < .01; b = .36, p < .01)

Study 1B Additional Results: Why doesn't observing conversational information improve human prediction accuracy?

In the main text we reported subjectively- and objectively-evaluated tasks in aggregate, here we report on each task individually.

Group Conversation: Liking. Observing paralinguistic information (b = .24, p < .001) and nonverbal information (b = .12, p < .001) increased accuracy in predicting the extent to which subsequent conversation group members liked the interviewee, but observing verbal content did not influence prediction accuracy (p = .17). However, only verbal information made participants feel more confident in their predictions (b = .17, p < .001); nonverbal and paralinguistic information did not affect confidence.

Writing Task 1: Termination letter. Observing paralinguistic information increased accuracy in predicting the quality of the termination letter ($b = .18$, $p < .001$) as did observing nonverbal content ($b = .14$, $p < .001$). Verbal information did not significantly affect prediction accuracy. Once again, observing verbal information increased participants' confidence ($b = .22$, $p < .001$); nonverbal and paralinguistic information did not affect confidence.

Writing Task 2: Case summary. Observing paralinguistic information increased accuracy in predicting the quality of the case summary ($b = .10$, $p < .01$). Nonverbal content marginally increased accuracy ($b = .06$, $p = .081$) and observing verbal content did not significantly affect accuracy ($p = .70$). Once again, observing verbal information significantly increased participants' confidence ($b = .21$, $p < .001$); nonverbal and paralinguistic information did not affect confidence.

Objectively-Evaluated Tasks. Observing nonverbal information improved predictive accuracy on the LSAT task ($b = .35$, $p < .001$), while observing verbal and paralinguistic information did not improve predictive accuracy. Observing nonverbal information also improved predictive accuracy on the GRE math task ($b = .10$, $p = .001$), while observing verbal and paralinguistic information did not improve predictive accuracy. Observing paralinguistic, verbal, and nonverbal information did not improve predictive accuracy in slider completion task or the bonus donation task.

Although verbal information did not lead to increased accuracy in any task, verbal information significantly increased participants' *confidence* in their ability to predict the LSAT task ($b = .22$, $p < .001$), the GRE math task ($b = .18$, $p < .001$), the slider task ($b = .18$, $p < .001$), and the bonus donation task ($b = .19$, $p < .001$).

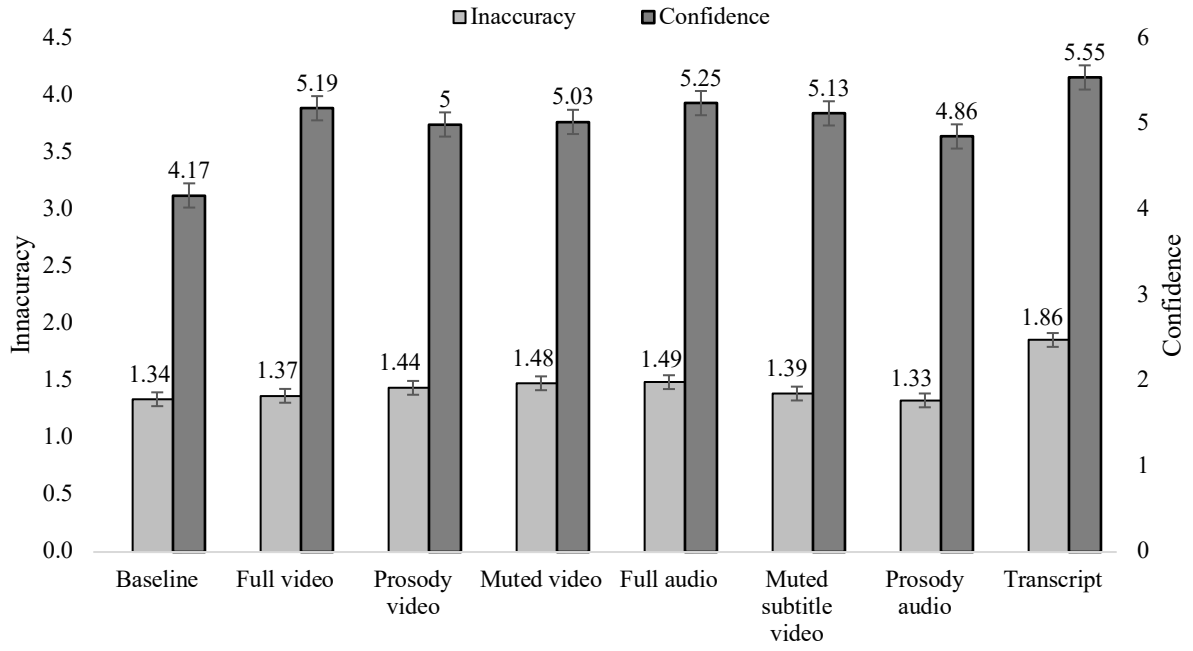
Study 1B Additional Pre-Registered Analysis

Another pre-registered analysis for Study 1B (model #2 in the pre-registration) was a linear mixed effects model with accuracy as the outcome variable, condition as a fixed effect, and random intercepts controlling for interviewee. In this model, we included only conditions three, six, and seven using planned pairwise contrasts between each condition. This was in order to isolate the effects of the condition with *only* nonverbal content (condition 3), *only* paralinguistic content (condition 6), and *only* verbal content (condition 7).

Liking. The results of this model showed that the participants in the paralanguage-only condition were more accurate at predicting liking compared to participants in the nonverbal-only condition ($b = .14$, $p = .035$) and the transcript-only condition ($b = .53$ $p < .001$). Participants in the nonverbal-only condition were also more accurate at predicting liking compared to participants in the transcript-only condition ($b = .38$, $p < .001$).

Writing Tasks. Participants in the paralanguage-only condition and in the nonverbal-only condition were more accurate at predicting performance on the case writing task ($b = .21$, $p = .018$; $b = .20$, $p = .018$) and the termination letter writing task ($b = .39$, $p = .001$; $b = .41$, $p < .001$) compared to participants in the transcript-only condition.

Other tasks. There were no significant differences between conditions in predictive accuracy of the LSAT task, the GRE math task, the slider completion task, or the bonus donation task.



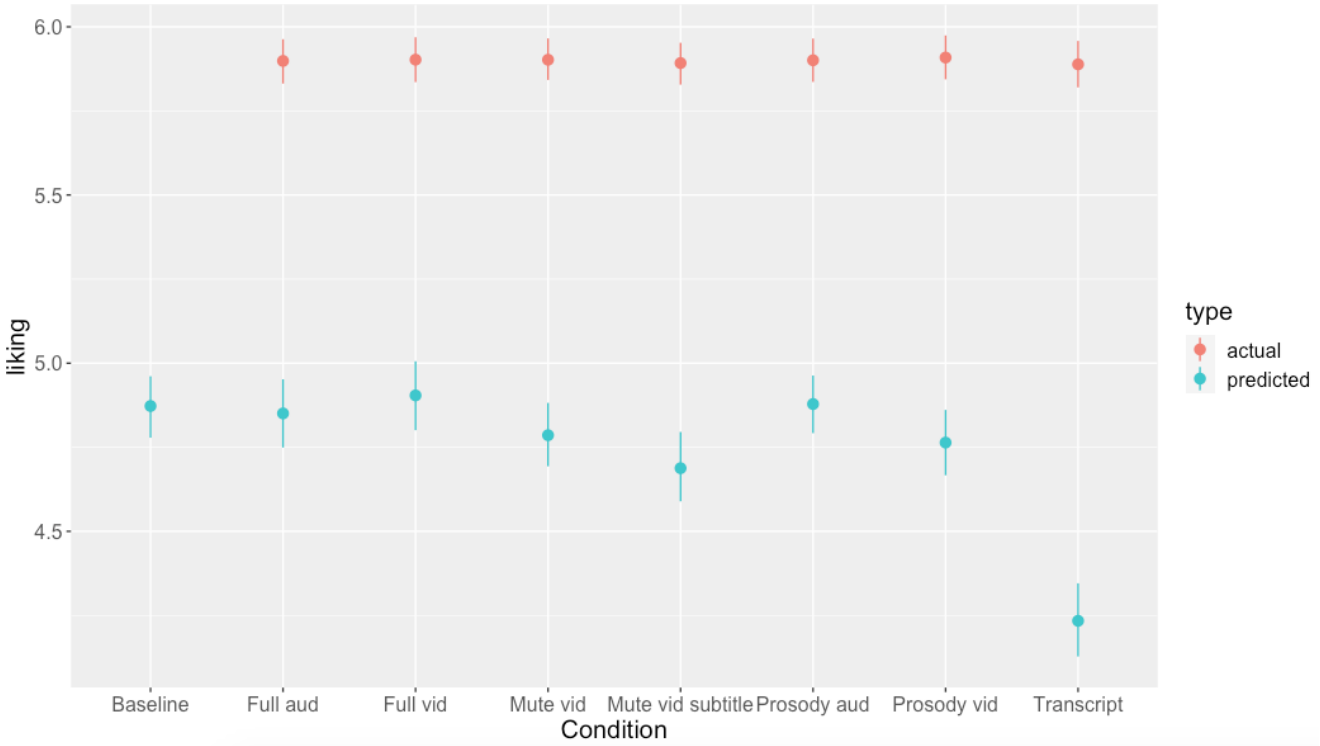
Supplemental Figure 2.1: When observers were asked to predict group members' evaluations of the interviewee's likeability, their predictions were most accurate with no interview information and least accurate when they read only the interview transcript, while their ratings of confidence showed the reverse: observer confidence was lowest with no information and highest after reading the transcript only.

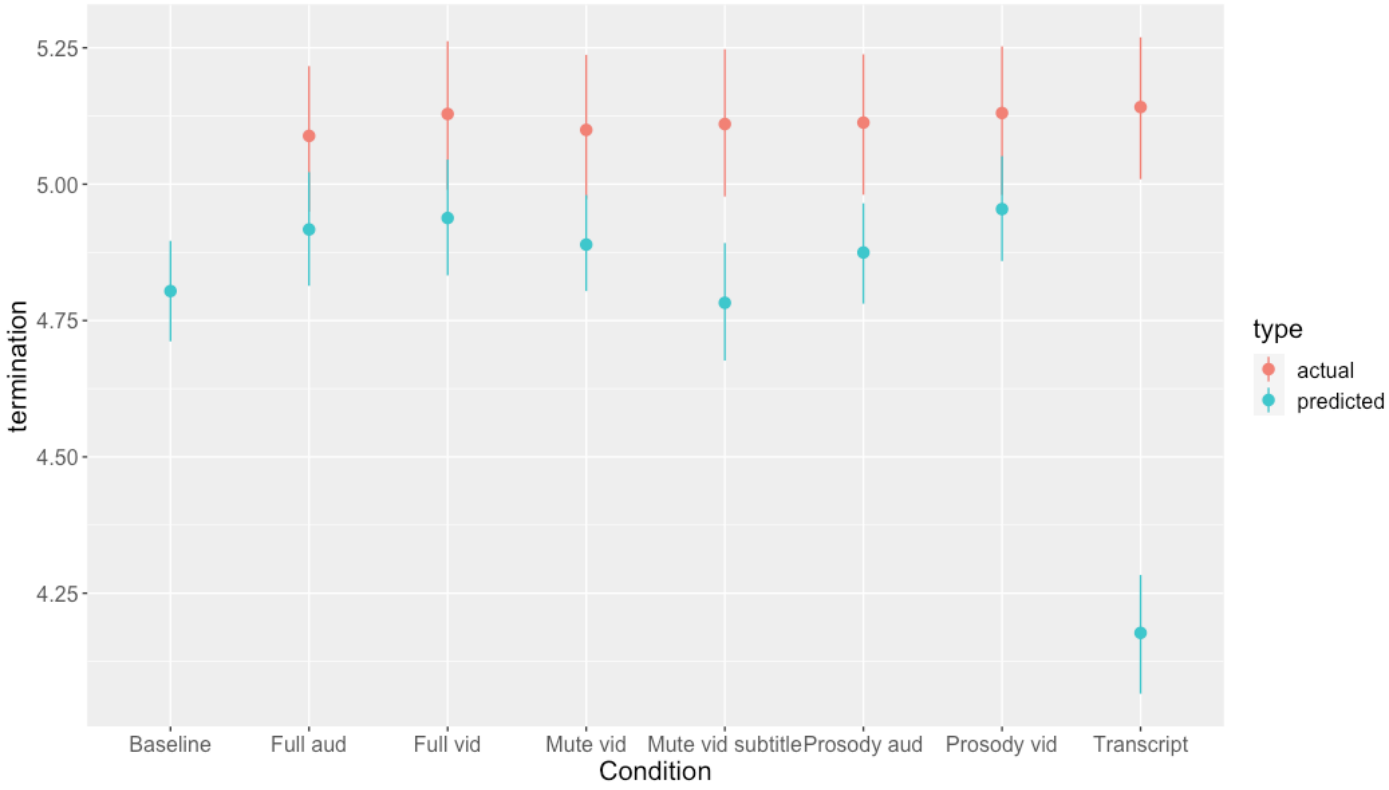
Appendix 2C: The Direction of Bias for Each Task in Study 1B

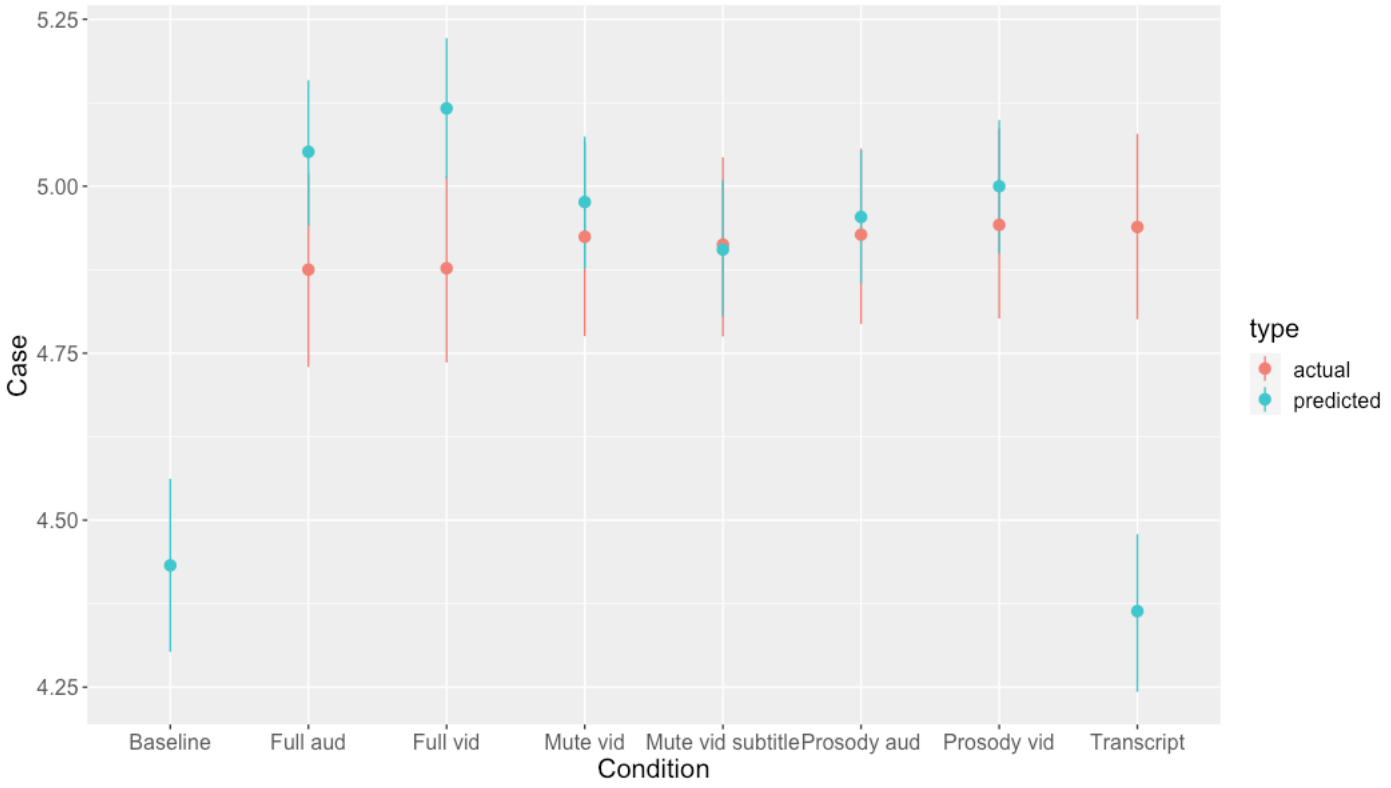
Supplemental Table 3: Summary of the direction of bias for each task. Participants seemed to underestimate performance on subjectively-evaluated tasks and to overestimate performance on objectively-evaluated tasks.

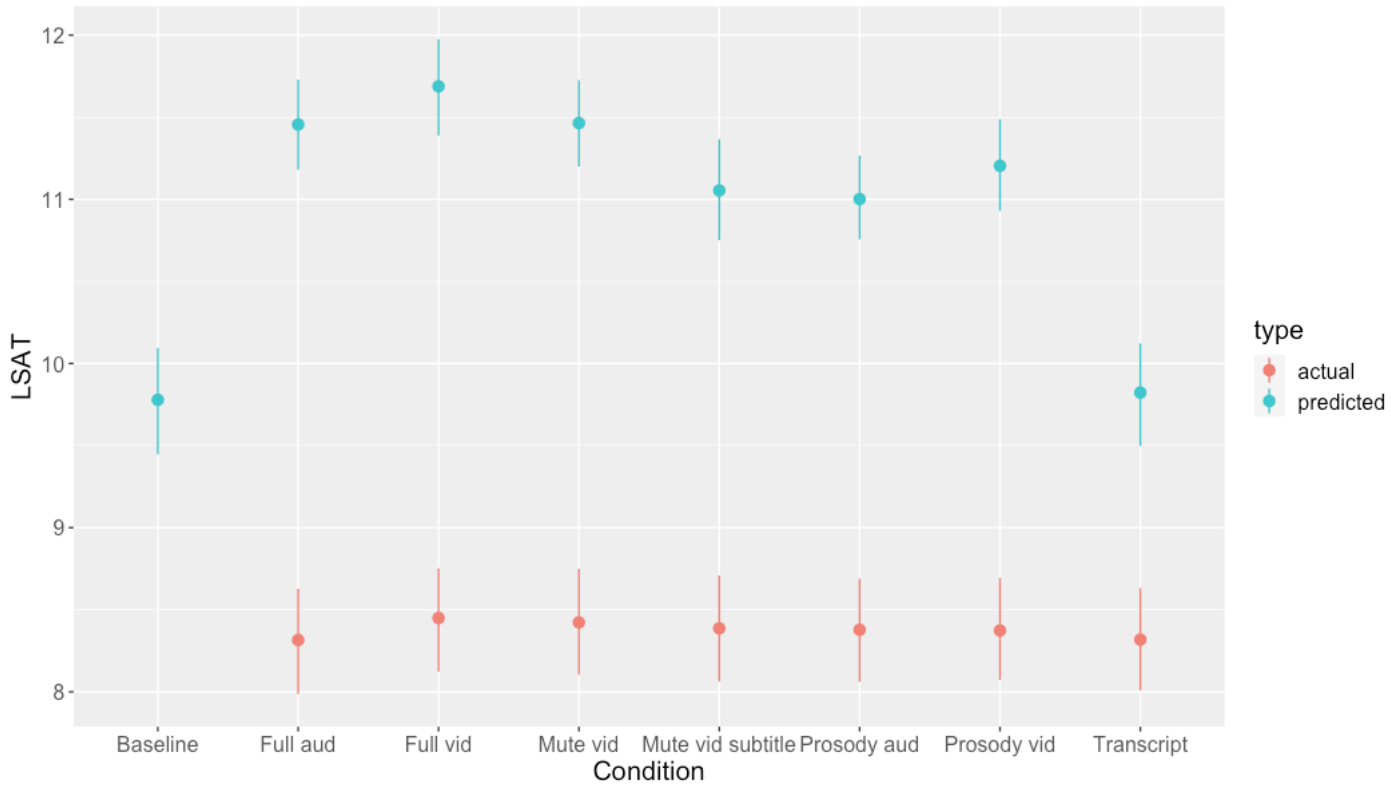
Task	Direction of Bias
Liking	Observers significantly underestimated the interviewee's likeability in group conversation, in every condition
Termination letter	Observers underestimated interviewee's writing score, not always significant
Case	Mixed: Observers sometimes overestimated (based on full video/full audio), sometimes underestimated (transcript) the interviewee's writing score
Bonus distribution	Mixed: Observers sometimes underestimated (full video/ muted subtitles), sometimes overestimated (transcript) the interviewee's generosity. Also, with no info, they thought interviewees would keep more money.
LSAT	Observers significantly overestimated LSAT performance in every condition.
Math GRE	Observers significantly overestimated Math GRE performance in every condition.
Slider Persistence	Observers significantly overestimated persistence in every condition.

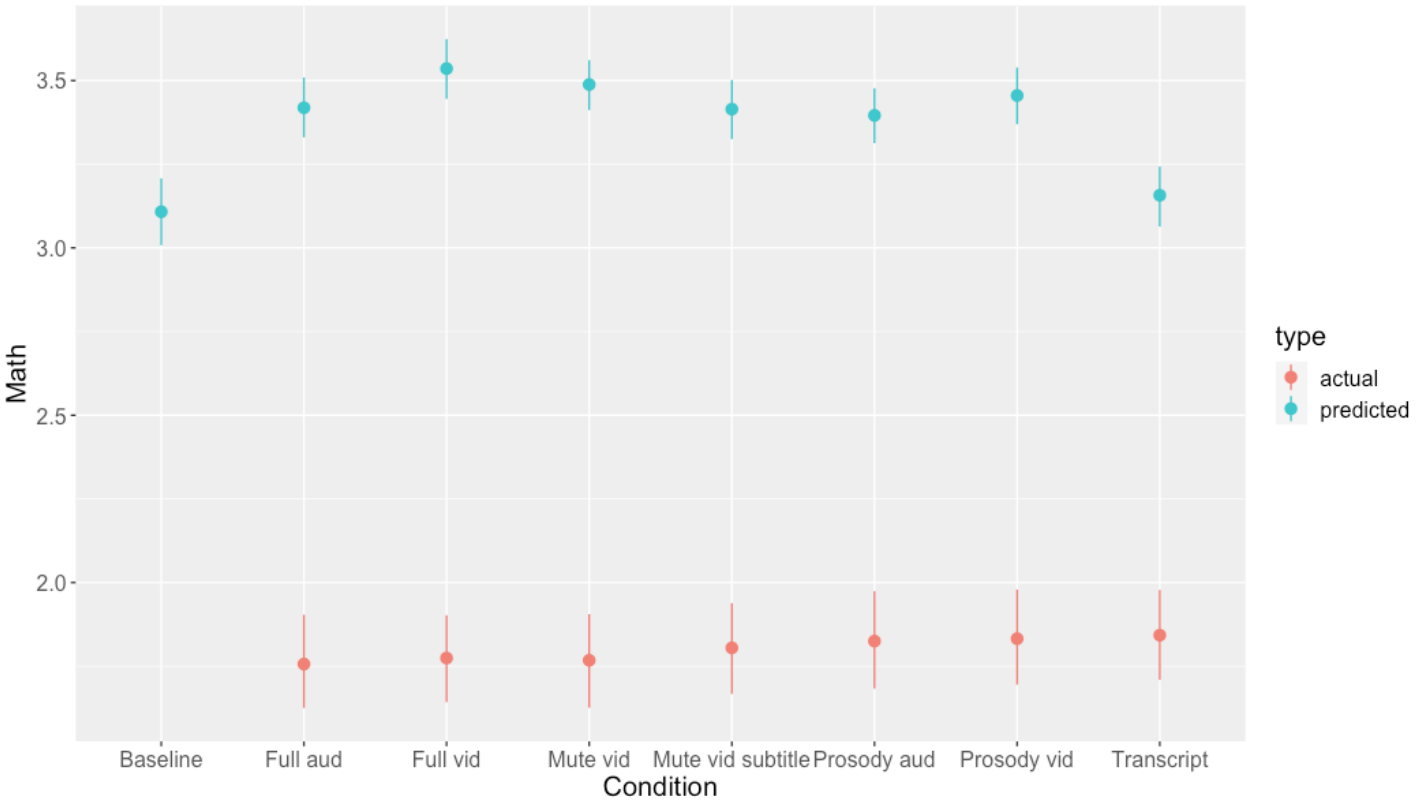
Supplemental Figures 2.2 to 2.8: Predicted and actual average performance for each task across conditions

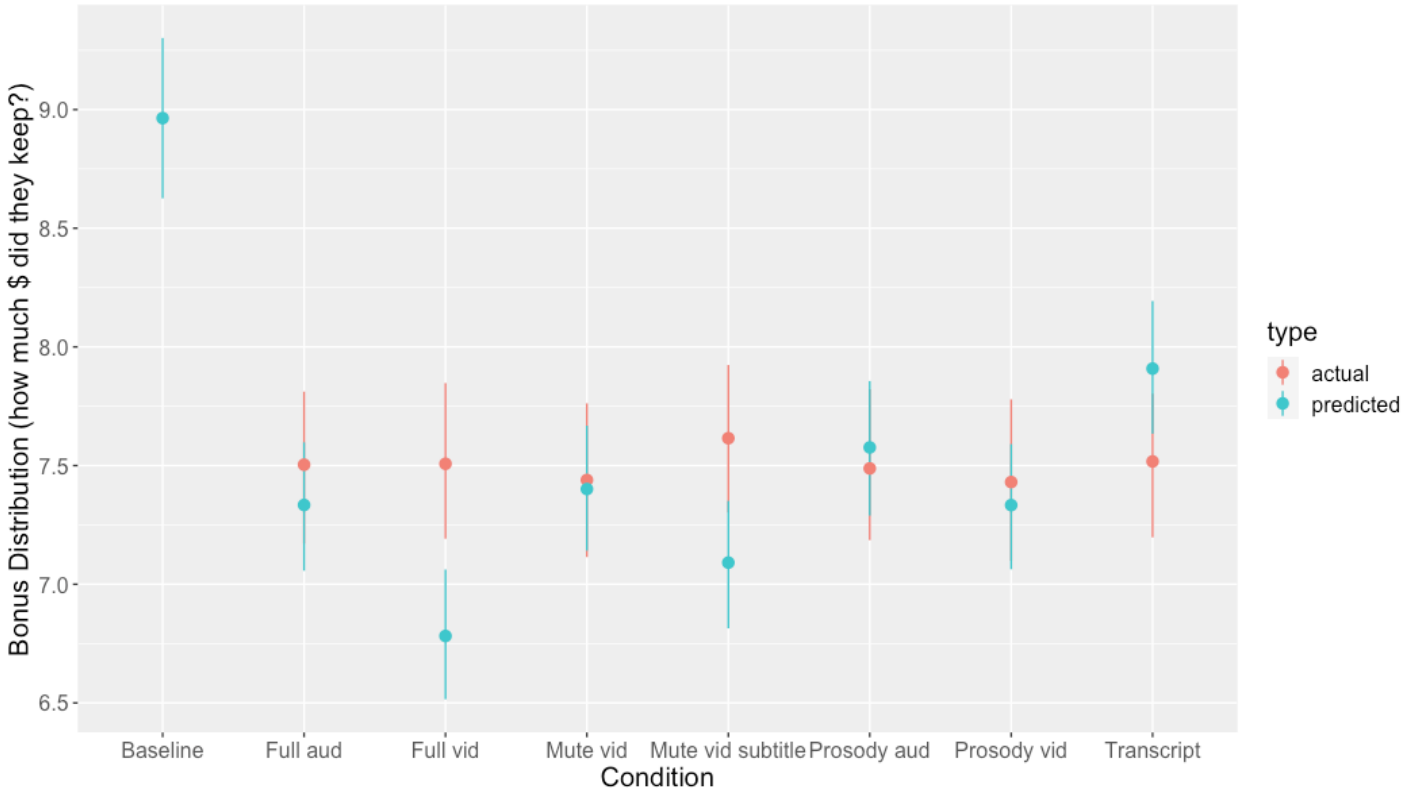


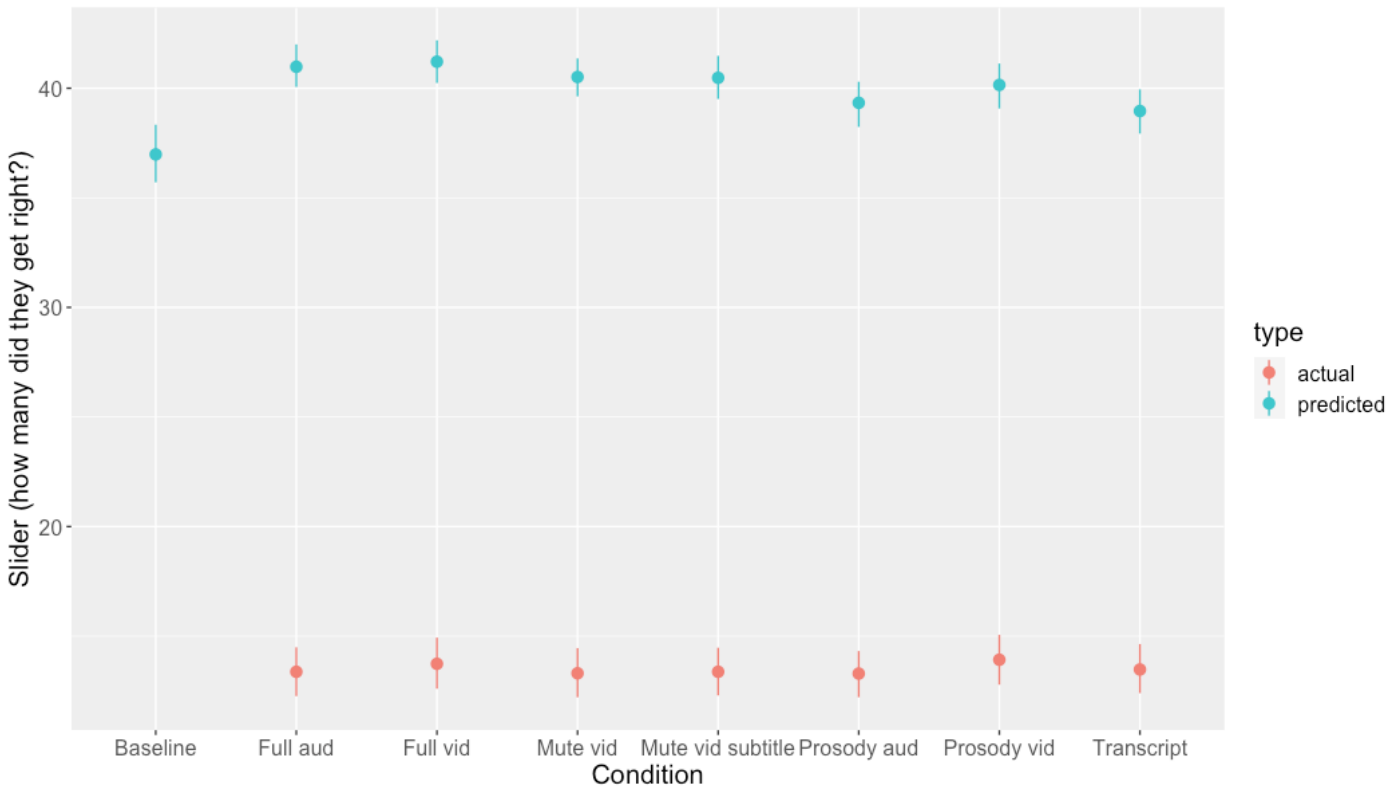



5.25
5.00
4.75
4.50
4.25
Case
Baseline
Full aud
Full vid
Mute vid
Mute vid subtitle
Prosody aud
Prosody vid
Transcript
Condition
type
actual
predicted


12
11
10
9
8
LSAT
Baseline
Full aud
Full vid
Mute vid
Mute vid subtitle
Prosody aud
Prosody vid
Transcript
Condition
type
actual
predicted


Math
3.5
3.0
2.5
2.0
Baseline
Full aud
Full vid
Mute vid
Mute vid subtitle
Prosody aud
Prosody vid
Transcript
Condition
type
actual
predicted


Bonus Distribution (how much $ did they keep?)
9.0
8.5
8.0
7.5
7.0
6.5
Baseline
Full aud
Full vid
Mute vid
Mute vid subtitle
Prosody aud
Prosody vid
Transcript
Condition
type
actual
predicted


Slider (how many did they get right?)
40
30
20
Baseline
Full aud
Full vid
Mute vid
Mute vid subtitle
Prosody aud
Prosody vid
Transcript
Condition
type
actual
predicted

Appendix 2D

Pilot Study: Lay Beliefs about the Predictive Power of Interviews

In a pilot study, we examined people's lay beliefs about the predictive power of interview information: what information do people *believe* will be most informative for forecasting performance, and how confident are they in their beliefs? We include all pre-registrations, materials, data, and code here: https://osf.io/3qtyp/.

Method

Participants. We recruited 204 participants on Amazon Mechanical Turk (M_{Age} = 25.8, *SD* = 7.74; 48% female) to participate in an online survey in exchange for $2.

Information choice. We asked participants what kind of information they would prefer to receive if they needed to predict someone's generalized job performance based on a conversational interview. They chose from the following options, presented in random order: "A written transcript of all the content of the interview"; "An audio clip of the interview"; "A video of the interview without sound"; "A video of the interview without sound, but with subtitles of what they were saying"; "A full video of the interview (with sound)"; "An audio clip of the interview where I could hear their tone of voice, but not the words they are saying"; "A video of the interview where I could hear their tone of voice, but not the words they are saying"; "I would rather have no information and just guess how an average person would do".

Judging subjective ability. Next, we asked participants what kind of information they would prefer to receive if they had to decide "how kind, friendly, likeable, and empathetic" the job applicant was. Participants picked from the same response options as above.

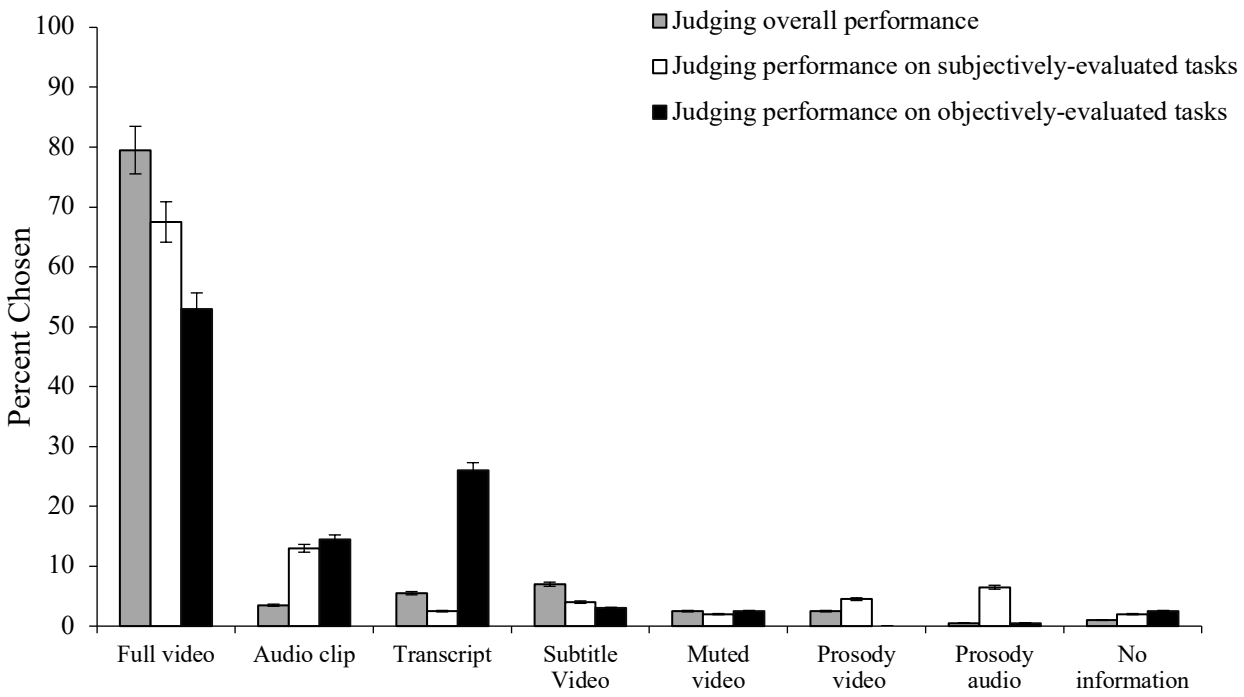
Judging objective ability. Then we asked what kind of information they would prefer to receive if they had to decide "how intelligent the person is, and how good they are at completing tasks well and efficiently." Participants picked from the same response options as above.

Confidence. Last, we asked participants how confident they would feel making the three judgments above if they only had one version of the eight versions of input information (summarized in Table 1). Participants rated their confidence for each judgment and each source of information using a 7-point Likert scale with endpoints labeled *1 (not very confident)* and *7 (very confident)*.

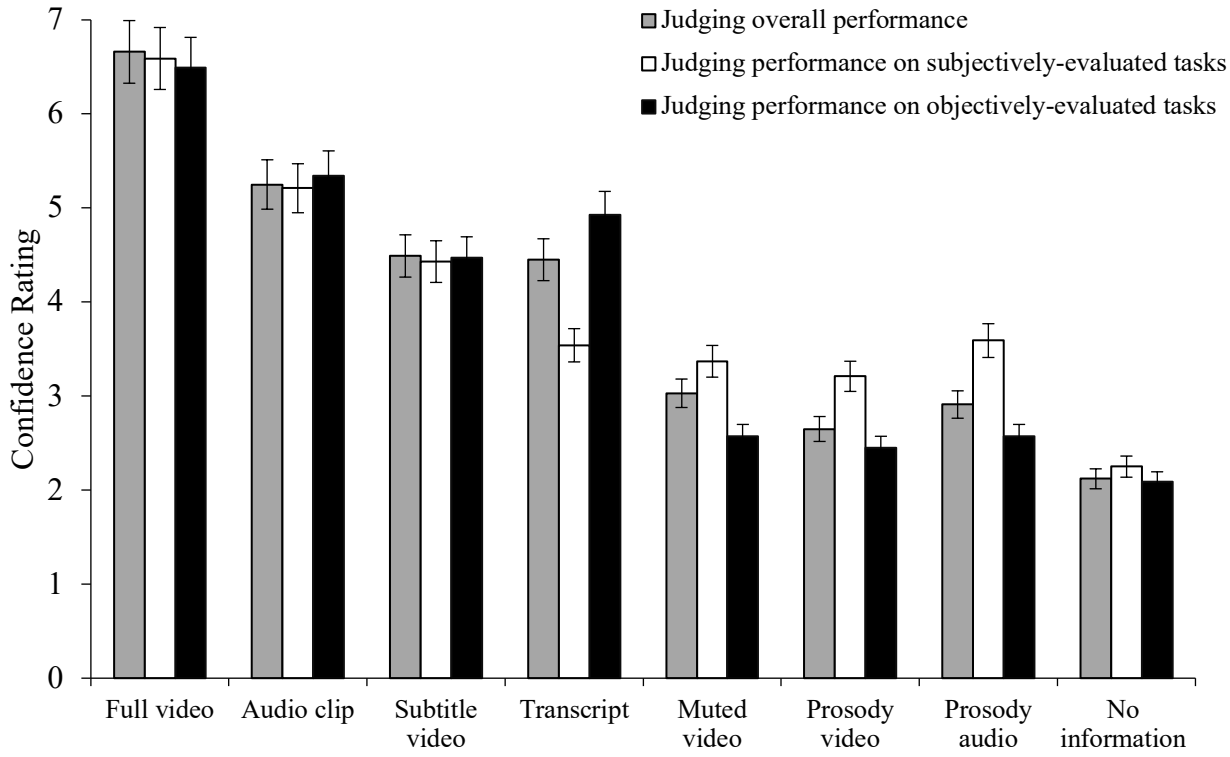
Results

Choosing Information. We summarize these results in Supplemental Figure 9. Participants overwhelmingly preferred the full video: 78% of participants chose it to predict general job performance, 66% chose it for judging subjective skills, and 52% chose it for judging objective skills. Twenty-five percent of participants preferred to observe the transcript to judge competence. As shown in Figure 1, no other options were chosen by more than 15% of participants.

Confidence. We summarize these results in Supplemental Figure 10. Participants reported that they would feel most confident making all three judgments using a full video. Participants were least confident with no information.



Supplemental Figure 2.9. The information participants prefer to observe to predict future performance



Supplemental Figure 2.10. Participants reported their confidence in their judgments, based on information observed

Chapter 3 Appendices

Appendix 3A: Additional Measures from Study 1B

Survey items:

Quality of Interviewee's Responses
Overall, the **quality** of the interviewees' responses were of... (1—Very low quality, 7- Very high quality)

Task Competence (1- Strongly disagree, 7- Strongly agree)
The interviewee was capable.
The interviewee was efficient.
The interviewee was organized.
The interviewee was thorough.
The interviewee met the interviewer's needs.
The interviewee performed as expected.

Trait competence (1- Not at all, 7- Very much)
How competent was the interviewee?
How confident was the interviewee?
How intelligent was the interviewee?
How skillful was the interviewee?
How efficient was the interviewee?

Warmth (1- Not at all, 7- Very much)
How friendly was the interviewee?
How well-intentioned was the interviewee?
How warm was the interviewee?
How good-natured was the interviewee?
How sincere was the interviewee?