Perceptions of conversational listening are inaccurate

Abstract

Across all domains of human social life—from learning to mating to commerce—being perceived as a "good listener" improves well-being, reproductive success, and interpersonal flourishing. But a fundamental question remains: Are perceptions of listening accurate? No prior research has empirically tested whether humans can detect others' auditory processing, interpretation, and comprehension (i.e., listening), particularly during live conversation. Across five studies (total N = 1,450), we experimentally manipulated "listening" during live conversation by instructing participants to divert their attention, or by constraining auditory input. We find that perceivers of all types—naive conversation partners, third-party observers, and even the manipulated listeners themselves watching video recordings of their own behavior—are surprisingly poor at distinguishing between attentive, feigned, and inattentive listening. These results contribute to a growing body of work on the failings of human-to-human mind perception and the surprising human inability to detect disfluencies during conversation. Our findings call for a re-examination of a fundamental and misunderstood social behavior ("listening"), and underscore a stark difference between *being* heard and *feeling* heard during conversation.

Key words: listening; conversation; mind perception; interpersonal relations

Perceptions of conversational listening are inaccurate

Imagine you're sharing a personal story over dinner with friends. While you're talking, your friends seem engaged: they hold your eye gaze, smile at the funny parts, nod in agreement, and laugh warmly. You decide that your friends were interested and listening because they offered every commonly known sign of doing so. But were they *really* listening? If you asked them about the details of the story tomorrow, would they remember? Even better, if you probed their minds during the conversation, what were they actually attending to and thinking about? Were they "good listeners," or just faking it well?

From business to medicine to romance, being seen as a "good listener" is widely advised and highly desired. And for good reason—perceptions of listening are associated with many beneficial outcomes. In the workplace, employees who feel that their supervisor listens to them report lower emotional exhaustion, lower turnover intentions, greater internal motivation (Lloyd, Boer, Keller & Voelpel, 2015), and seek more feedback (Qian, Wang, Song, Li, Wu, & Fang, 2019). In healthcare, patients who feel that their healthcare provider listens to them are not only more satisfied with the care that they receive (Wanzer, Booth-Butterfield & Gruber, 2004), but also show higher levels of medication adherence (Shafran-Tikva, & Kluger, 2016). In romantic relationships, signals of listening are associated with improved dyadic coping and overall relationship satisfaction (Bodenmann, 2005; Kuhn, Bradbury, Nussbeck, & Bodenmann, 2018). Even in voice-to-voice encounters, customer perceptions of listening during a phone call with call center employees drive satisfaction ratings (De Ruyter & Wetzels, 2000). Importantly, although the prior literature has shown that perceptions of listening are consequential, research has not shown whether they are *accurate.*

From marriage counselors to international mediators, conflict resolution professionals are frequently tasked with repairing the relational damage resulting from one party accusing the other of failing to listen, and the frequency of relationship partners not feeling heard has spawned a large self-help literature. This literature aims to teach well-intentioned conversationalists how to "show" they are listening so they get credit for their efforts.

Yet, the opposite error can also occur. Individuals often assume that their partner is listening when they are not, with deleterious consequences for information exchange and productivity. Consider a work meeting with colleagues diligently feigning attention while doodling on their notepads. While the meeting organizer might leave with the impression that she effectively communicated important information, her colleagues may be just as uninformed as they were at the start. Pretending to listen does not enable learning, but may leave the speaker with the mistaken impression that learning occurred.

One reason people misperceive others' listening may be the complexity of conversation, which is a dynamic, relentless, and cognitively demanding decision environment. Participants must constantly multi-task by keeping track of their partner's contributions (verbal, nonverbal, and prosodic expressions) while also deciding what to say and do in turn (Truong, Fast, & Kim, 2020; Yeomans & Brooks, working). Determining whether one's partner is really listening attentively requires perspective taking, a notoriously challenging task (Epley, 2008; Epley & Waytz, 2008). While people manage this deluge of cognitive demands, how accurate can they possibly be at evaluating their partner's listening? How frequently do people make inferential errors when they give credit to their partners for being good listeners or blame for being bad ones? In the current work, we examine whether people know the difference between attentive, feigned, and distracted listening during live conversation.

Previous Research on Listening – A Conundrum

The psychological process of listening is more complex than simply hearing words and sounds (i.e., auditory processing). Listening is experienced cognitively in the privacy of one's mind, expressed through words and behaviors, and perceived by others with three types of cues: verbal cues (e.g., words of affirmation, repetition, follow-up questions), nonverbal cues (e.g., nodding, eye gaze, trunk lean, facial expression), and prosodic cues (e.g., backchannels like "uh huh," silence, laughter) that may (or may not) accurately signal the underlying cognitive process. Janusik (2007) describes this conundrum aptly: "Listening research is a challenge, as listening is performed cognitively and perceived behaviorally, but listening cognitions and behaviors are not always congruent (Witkin, 1990)."

Accordingly, prior psychological research on listening has focused on two broad themes: the intrapersonal (cognitive) experience of listening, versus the interpersonal perception of listening and the consequences of those perceptions. Prior cognitive models of listening focus on two factors: 1) auditory attention, "the ability to direct and sustain attention to sounds," and 2) memory, "the ability to remember or use what has been heard," both of which are integral to the act of listening (Witkin, 1990). Indeed, early listening scholars attempted to measure listening using hearing, comprehension, and recall measures. For example, Nichols (1947) saw information retention as the key feature of good listening, thus setting the tone for listening research and training in the 1950s (Nichols & Lewis, 1954; Nichols & Stevens, 1957). However, recent scientific consensus has concluded that auditory, comprehension, and recall measures of listening cannot meaningfully distinguish listening from memory capacity (Thomas & Levine, 1994). Thus, researchers to date have not uncovered a robust measure that accurately captures listening as a cognitive activity and distinguishes it from other related processes.

Prior interpersonal models of listening describe how people *perceive* others' listening. This work has largely relied on two types of measures: 1) the listener's behavioral cues that lay belief suggests are indicative of whether or not they are or are not listening, and 2) their partners' self-reported perceptions of listening (e.g. "How well do you think your partner listened?") (e.g., Itzchakov, Castro, & Kluger, 2016; Wanzer, Booth-Butterfield & Gruber, 2004).

Much of the attention in this area has focused on "active listening," a style of listening initially suggested by the pioneering humanistic psychologist Carl Rogers and considered effective in many fields, including psychotherapy (see Hafen & Crane, 2003 for a review), education (Cheon & Grant, 2009, McNaughton et al., 2008), and business (Kubota et al., 2004, Mishima et al., 2000, Rautalinko & Lisper, 2004). Active listening embraces the benefits of listening, as well as the importance of communicating that one is listening. For example, books that teach "active listening skills" usually include a section on "behavioral listening," or non-verbal behaviors such as maintaining eye contact, leaning forward, nodding, and smiling when appropriate (e.g., Hoppe, 2006; Rogers & Farson, 1957). According to this literature, by engaging in behaviors that people associate with listening (see Bodie, Cyr, Pence, Rold, & Honeycutt, 2012), the listener can signal to their partner that they are listening, which will improve the interaction and, ultimately, the relationship.

Experimental investigations of active listening strategies suggest benefits for the listener and their conversation partners: active listeners are better liked and people find interacting with them more satisfying (e.g. Weger, Castle, & Emmett, 2010; Weger, Castle Bell, Minei, & Robinson, 2014). However, research on active listening has focused exclusively on *perceptions* of listening, usually by surveying the speaker. Thus, it remains unclear whether the active

listener has actually performed the cognitive work of listening well, or whether they have simply done a better job convincing their partner.

Further, research on active listening and perceptions of listening broadly suggests that being a good listener may not merely require a healthy auditory cortex and a good working memory, but also the ability to engage with the speaker in an empathic manner. For example, psychological research on *responsiveness*, the extent to which a listener expresses understanding, validation, and care, suggests that higher responsiveness to a partner strongly predicts the success of long-term relationships by improving intimacy, attachment, and emotional health (Laurenceau et al., 1998; Capriariello & Reis, 2011; Reis et al., 2004; Reis, 2020). Responsiveness may capture what individuals mean when they report "feeling heard." Like the cognitive study of listening, though, there is no way to directly capture whether a listener is truly and sincerely understanding, validating, and caring for the speaker—or just going through the motions. Indeed, though there are well-developed self-report scales for interpersonal perceptions of responsiveness (Reis et al., 2018), capturing actual responsiveness may pose an even more profound measurement challenge than assessing listening through measures of recall and information processing.

In sum, prior work on listening has broadly struggled to connect the experience and perception of listening. Some philosophers have even gone as far as to suggest that a large portion of human conversation includes very little listening at all. Abraham Kaplan (1994), for example, proposed the idea of the "duologue," a conversation in which neither party listens to the other. A duologue is more than a monologue because people take turns speaking and not speaking, but it is less than a dialogue because the participants aren't really listening to each other—akin to the idea of parallel play among children.

And thus, the question remains: How does the internal experience of listening, which may or may not be expressed behaviorally, influence others' perceptions of listening quality (if at all)? Are perceptions of listening an accurate reflection of actual listening? Do people know when someone is listening to them (or not) in conversation? The current research seeks to address these questions by manipulating (rather than measuring) listening during live conversation.

Mind Perception in Conversation

Despite the frequency with which conversation occurs, it is a difficult task. Conversation exerts a deluge of cognitive demands, and humans are good at only some of them. For example, people are mostly good at understanding what has been said, and they exert effort to maintain a shared understanding (Clark & Schaefer, 1989, Hardin & Higgins, 1996; Pickering & Garrod, 2004; Schegloff, 2007).

Establishing a shared understanding in conversation requires people to consider (and understand) others' points of view—including their wants, needs, beliefs, and intentions (Clark Wilkes-Gibbs, 1986; Fussell & Krauss, 1992; Isaacs & Clark, 1987; Schober, 1993; Wilkes-Gibbs & Clark, 1992; c.f. Epley & Waytz, 2010)—a task that is notoriously difficult for the human mind. Failure to successfully reason about the minds of others often leads to miscommunication and exacerbates conflict (Keysar & Barr, 2002; Ross & Ward, 1995). Indeed, inadequate perspective-taking has been postulated to be the greatest barrier to resolving conflict (see Friend & Malhotra, 2019 for a review).

Although conversations rely heavily on the ability to reason about the minds of others, there is a large literature documenting the limits of this ability (e.g., Epley, 2008; Epley & Waytz, 2010). Without direct access to the content of another's mind, people must rely on

various other inputs to guide their inferences about others' mental states. One strategy is to consult the content of one's own mind (Goldman, 2006). However, people regularly rely too heavily on their own mental state and fail to properly adjust for how another's mental state may differ (Epley, Keysar, Van Boven, & Gilovich, 2004; Tamir & Mitchell, 2013). Another strategy is to rely on people's behavior to help infer the contents of their mind. However, the same behavior can lead to different mental-state inferences based on the believed intention (e.g., Hamlin, 2013), as well as the objective outcome (e.g., Baron & Hershey, 1988).

An emerging literature has begun to investigate how simple failures of mind perception stand in the way of good conversational outcomes. For example, though there are systematic cues in people's speech that signal their topic preferences, even in cooperative interactions, humans struggle to recognize whether a person wants to keep discussing a topic or switch to a different topic (Yeomans & Brooks, working). Likewise, parties in conflict systematically misjudge the emotions experienced by their counterparts, attributing greater fear of being wrong to those who disagree with them than is actually the case (Dorison & Minson, 2020, working). While people often overestimate their expressed receptiveness to a divergent point of view (Yeomans, Minson, Collins, Chen & Gino, 2020) and the intrusiveness of asking personal questions (Hart, VanEpps, & Schweitzer, working), they systematically underestimate the benefits of asking questions (Huang et al., 2017) and the extent to which people like them after a conversation has ended (Boothby, Cooney, Sandstrom, & Clark, 2018; Mastroianni et al., 2020).

Extending this work to the domain of listening, work by Galantucci and colleagues suggests that conversationalists are insensitive to disruptions in conversation because they are often not actually listening (or only partially listening). For example, Galantucci & Roberts (2014) introduced major disruptions to conversational coherence during instant messaging

conversations. When they repeatedly crossed unrelated conversations (i.e., swapped partners from one text conversation to another unbeknownst to the parties)—between 27% and 42% of the participants did not notice that their conversation partners had changed at all, lending supportive evidence to Kaplan's suggestion of "duologue," at least in text-based conversation. Roberts et al. (2016) replicated this finding with different obvious disruptions (e.g., incorrect references to someone's gender). Finally, Galantucci et al. (2018) showed that people also don't notice if nonsensical phrases are said by conversation partners. When a confederate said "Colorless green ideas sleep furiously" during a dyadic interaction in a clear voice, only one third of participants admitted noticing the nonsensical sentence. Taken together, it seems that people spend a significant amount of time not listening (and not noticing) during conversation, despite the social value placed on listening.

We contribute to this growing literature by suggesting that evaluations of others' listening—on which so many important social judgments and decisions rely—are riddled with misperceptions. These misperceptions are rooted in people's fundamental inability to read the minds of others, coupled with the belief that they can. Though prior work has focused almost exclusively on the consequences of *perceptions* of listening—people's beliefs about whether or not their conversation partner is listening to them (e.g., Lloyd, Boer, Keller & Voelpel, 2015; Itzchakov, Castro, & Kluger, 2016; Shafran-Tikva, & Kluger, 2016; Wanzer, Booth-Butterfield & Gruber, 2004)—in the present work, we examine the correspondence between perceptions of listening and *actual* listening by experimentally manipulating listening behavior and/or auditory input during live conversation.

The Current Work

Across five studies, we investigate whether people can accurately detect listening in conversation. We suspect that although speakers tend to *feel* heard (or ignored), they often don't know if they have *actually* been heard or not.

In our pilot study, we document people's lay beliefs about their own and others' listening ability and ability to perceive listening in others (i.e., listening detection). Then, in our main set of experiments (Studies 1-4), dyads engage in live conversation, with one individual later asked to assess the extent to which the other was listening attentively. Using a variety of paradigms, we manipulate actual listening between experimental conditions by introducing distractions, adjusting incentives, or using technology to limit participants' physical ability to hear their partner—and then measure perceptions of listening. This approach allows us to circumvent the challenge of measuring the cognitive experience of listening and instead introduces a "ground truth" metric of how well participants listened. We can compare ground truth listening quality (by experimental condition) to *perceptions* of listening quality and other interpersonal evaluations offered by their partners—methods only possible when listening is studied in real dyadic encounters. All data and materials for each study are available here: https://osf.io/w4nf9/?view_only=db83cc05213c4964951ab88b1f9e5de0.

Pilot Study: Overconfident Beliefs about Listening and Listening-Detection

We conducted a survey of people's lay beliefs about their own and others' listening. Specifically, we were interested in whether people believe themselves to be good listeners and good evaluators of others' listening. We also wanted to know how they evaluate their own skills in this domain relative to those of others. Finally, we wanted to document the behaviors that

people consider to indicate good listening. All procedures and analyses were pre-registered (https://aspredicted.org/blind.php?x=7qc4se).

Method

Participants. We recruited participants through Prolific (N = 829 after exclusions; 177 participants excluded for failing our attention check; 50% male, M_{age} = 32.14) to participate in a 3-minute study.

Procedure. We employed a 3 (conversation partner type) x 2 (self v. other) between-subjects design: Participants were randomly assigned to think about conversations they had with a romantic partner, friend, or co-worker. Then participants were randomly assigned to respond to the questions in reference to *themselves* or the *other* person in the conversation. Participants then answered a series of questions about their lay perceptions of listening.

Listening Behavior. We asked participants to think back to conversations they had with a romantic partner/a friend/or a co-worker (randomly assigned) and estimate the percentage of time in those conversations they themselves ("self" condition) or their romantic partner/friend/co-worker ("other" condition) spent "listening attentively (versus pretending to listen or tuning out altogether)." Participants provided a percentage estimate of the amount of time spent "attentively listening" and "pretending to listen, or tuning our altogether" and were told their responses must sum to 100%.

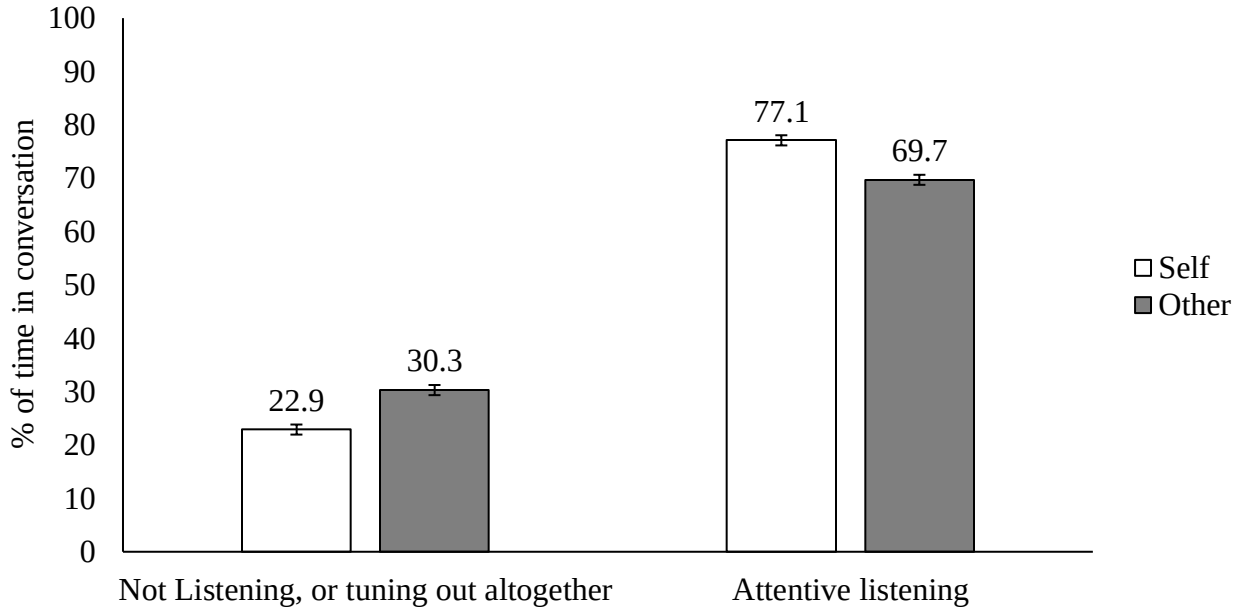
Listening Detection. Next, participants were asked to think back to these same conversations and report how often they ("self" condition) or their romantic partner/friend/co-worker ("other" condition) were able to accurately detect whether their conversation partner was "attentively listening…versus pretending to listen or tuning out altogether" (1: *Never*, 2: *Rarely*, 3: *Sometimes*, 4: *Often*, 5: *Always*).

Listening cues. Finally, we asked participants what they look for when they are trying to detect whether someone is "listening attentively to you, pretending to listen, or tuning out altogether." Participants listed "3 qualities, behaviors, and/or tendencies" using an open-ended response format.

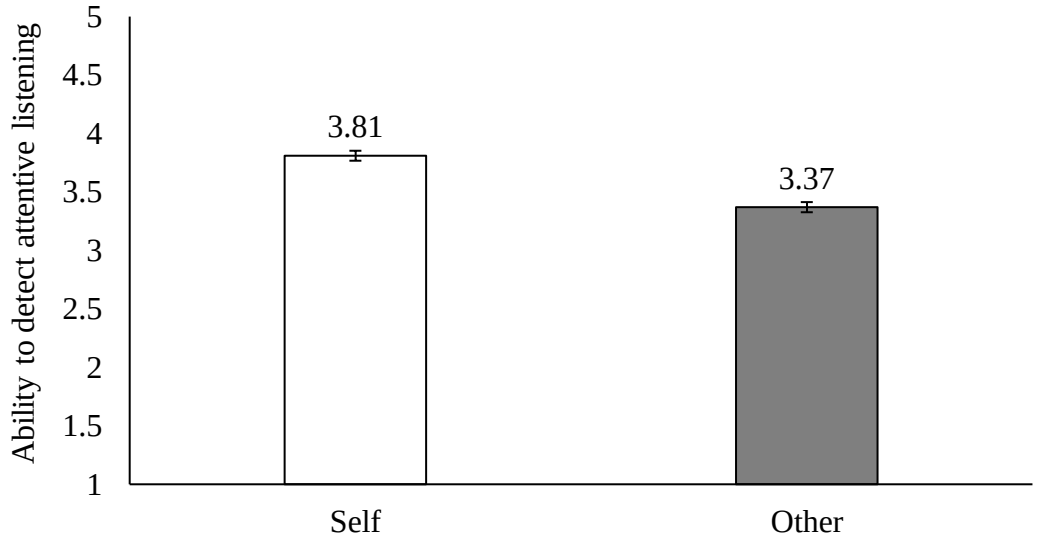
Results

Listening Behavior. First, we present our results collapsed across social categories to investigate the main effects of self- vs. other-assessments. Across all social categories, participants readily reported that they spend about a quarter of their conversations not listening to their partners (M = 23% "pretending to listen, or tuning out altogether", SE = 0.94). However, they believed that other people commit this offense even more frequently (M = 30% "pretending to listen, or tuning out altogether", SE = 0.95), $b = 7.39$, $SE = 1.33$, $p < .001$, Cohen's d = -0.39, 95%CI [-0.52, -0.25] (see Figure 1).

We find no significant interaction between self- ($M_{coworkers} = 26$, SE = 1.65; $M_{friends} = 23$, SE = 1.58; $M_{partner} = 21$, SE = 1.63) vs. other-condition ($M_{coworkers} = 33$, SE = 1.65; $M_{friends} = 30$, SE = 1.67; $M_{partner} = 28$, SE = 1.60) and social category on beliefs about listening behavior, $F(2, 823) = 0.04$, $p = .96$. In other words, irrespective of relationship type, participants believed that a substantial portion of their conversations featured one party not listening to the other, and they were less guilty of this offense themselves than their conversation partners.



Panel A. Participants report spending significantly more time attentively listening, and significantly less time not listening, or tuning out altogether compared to others (romantic partners, friends, and coworkers).



Panel B. Participants report being able to detect attentive listening (vs. not listening, or tuning out altogether) in their conversation partners more often than others (romantic partners, friends, and coworkers) can detect their attentive listening in conversation.

Figure 1. Self-other differences in beliefs about time spent attentive listening and listening detection abilities, across social categories (Pilot Study).

Listening Detection. Importantly, participants believed that they were better than others in detecting attentive listening. Collapsing across social categories, people thought they were able to detect when their conversation partners were attentively listening more accurately (M = 3.81, SE = 0.04) than they thought their conversation partners could (M = 3.37, SE = 0.04), b = -0.44, SE = 0.06, $p < .001$, Cohen's d = 0.50, 95%CI [0.36, 0.64,] (see Figure 1).

We find a significant interaction between self- vs. other-condition and social category on beliefs about listening detection ability, $F(2, 823) = 4.55$, $p = .01$. Conducting Tukey's test for post-hoc comparisons, we find that in the "self" condition, there were no differences between the three social categories in participants' self-reported abilities to detect attentive listening ($M_{coworkers}$ = 3.79, SE = 0.07; $M_{friends}$ = 3.78, SE = 0.07; $M_{partner}$ = 3.87, SE = 0.07; friends vs coworkers: b = -0.009, SE = 0.10, p = 0.99; partners vs coworkers: b = 0.08, SE = 0.10, p = 0.73; partners vs friends: b = 0.09, SE = 0.10, p = 0.67). However, in the "other" condition, participants reported that their romantic partners were much better able to detect their attentive listening ($M_{partner}$ = 3.62, SE = 0.07) than their coworkers ($M_{coworkers}$ = 3.10, SE = 0.07; b = 0.52, SE = 0.10, $p < .001$), and marginally better than their friends ($M_{friends}$ = 3.37, SE = 0.08; b = 0.25, SE = 0.10, p = .05). Additionally, their friends were better able to detect their attentive listening than their co-workers ($M_{coworkers}$ = 3.10, SE = 0.07; b = 0.27, SE = 0.11, p = .03).

Pilot Study Discussion

Our exploratory results suggest that people show a self-serving bias in their beliefs about listening ability: not only do they think they spend more time during their conversations actually listening than others, but they also think they are better than their partners at detecting actual (vs. feigned) listening. Further, we find that feigned listening is a common occurrence: people

reported spending approximately a quarter of their conversation time giving the false appearance of listening.

Study 1: Can Conversationalists Detect Listening?

In Study 1, we pursue our main prediction: that people do not accurately detect when their conversation partner is listening attentively or inattentively during a live conversation. In order to induce variation in listening, we instructed one participant in each dyad to pay careful attention to their partner (listening condition), direct their attention elsewhere (distracted condition), or direct their attention elsewhere while *pretending* to listen to their partner (feigned listening condition).

Study 1 Method

Participants. A total of 162 dyads came into the lab to participate in a study on "Everyday Conversations." Eleven dyads were excluded from analysis (2 dyads knew each other; 2 dyads did not complete the questionnaires; 7 dyads experienced an experimenter error).[1] Our analyses are based on the remaining 151 dyads (N = 302; 55% female; M_{age} = 23 years).

Procedure. When they arrived at the lab, participants were randomly paired with an unfamiliar 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. We also told participants that there would be a series of videos playing on a screen in the conversation room. The instructions regarding these videos varied by condition and are explained below.

Listening Manipulation. Within each dyad, one participant sat with their back to the video screen and was instructed to ignore the videos playing behind them. This was the

[1] Participants in these seven dyads experienced a technical malfunction (4 dyads) or received a questionnaire that did not match their condition assignment (3 dyads).

"unmanipulated partner." The other participant (the "manipulated" partner) was seated in full view of the video screen, and was randomly assigned to one of three conditions: (1) *Listening* condition (n = 50), (2) *Distracted* condition (n = 47), or (3) *Feigned Listening* condition (n = 52). The unmanipulated partner did not know the manipulated partner's private listening instructions. The manipulated participants were video recorded throughout the interaction.

We instructed participants in the *Listening* condition to ignore the videos and listen attentively to their partner. Participants learned that they could earn a bonus based on how well they remembered what their partner said.

Participants in the *Distracted* condition were instructed to pay attention to the video playing on the screen behind their partner. They learned that the video consisted of a series of muted commercials and that they could earn a bonus for each commercial that they correctly recalled.

Participants in the *Feigned Listening* condition were instructed to pay attention to the commercials playing on the screen while pretending to listen attentively to their partner. These participants were offered a bonus for each commercial they correctly recalled *only if* their partner reported thinking they were listening throughout the conversation.

After both participants in a dyad read their instructions independently, they were brought into the conversation room and were seated face-to-face.

Measures. Our primary question was whether the listening manipulation influenced the impressions that the unmanipulated participants formed of their manipulated partners. To this end, we asked the unmanipulated participants to judge the quality of their partner's listening, evaluate the conversation overall, as well as their partner's characteristics.

Perceptions of Listening. Unmanipulated participants reported the extent to which they thought their partner was "a good listener," was "interested in what I had to say," and was "engaged in this conversation" on a scale from 1: "Extremely disagree" to 7: "Extremely agree." Responses to these three items were averaged to create an overall measure of the extent to which participants thought their partner was listening ($\alpha = 0.91$).

Enjoyment. Unmanipulated participants also reported the extent to which they agreed with fives statements assessing their enjoyment of their partner as well as their conversation (e.g., "I liked my partner" and "I found the conversation with my partner interesting"; $\alpha = 0.91$).

Warmth, Competence and Power. Unmanipulated participants also rated their partner's warmth (4-items; e.g., "I think my partner is tolerant"; $\alpha = 0.78$; Fiske, Cuddy, & Glick, 2002), competence (5-items; e.g., "I think my partner is independent"; $\alpha = 0.64$; Fiske, Cuddy, & Glick, 2002), and power (3-items; e.g., "I think my partner is dominant"; $\alpha = 0.78$; Smith, Wigboldus, & Dijksterhuis, 2008). Importantly, both participants were told that their interpersonal ratings would remain private to mitigate potential impression management concerns.

Manipulation Check. To assess whether participants followed our instructions, manipulated participants were asked to recall as many of the commercials as they could. We tallied the number of commercials participants correctly recalled (out of 9). Manipulated participants also predicted the extent to which their partner would report that they were a good listener on a scale from 1: "Extremely disagree" to 6: "Extremely agree."

To measure the extent to which participants learned about their partner, unmanipulated participants also completed the Activities Preferences Questionnaire (APQ; Surra & Longstreth, 1990; Swann & Gill, 1997) prior to the conversation, and manipulated participants predicted their responses to the questionnaire after the conversation. However, in reviewing the

conversation transcripts, it was clear that none of the conversation pairs discussed the activities included in the questionnaire, and thus this measure is not included in our main analyses.

Study 1 Results

Self-Perceptions. Our manipulation successfully shifted participants' attention as evidenced by the number of commercials the manipulated participants recalled, $F(2, 148) = 90.11$, $p < .001$. Participants in the *Distracted* condition correctly recalled the greatest number of commercials ($M = 5.61$, SD = 1.93). Conducting Tukey's test for post-hoc comparisons, participants in the *Feigned Listening* condition recalled significantly fewer commercials than those in the *Distracted* condition ($M = 4.46$, SD = 2.36; $b = -1.15$, 95%CI [-1.88, -0.42], SE = 0.37, $p = .002$), but recalled significantly more than those in the *Listening* condition ($M = 0.84$, SD = 0.93; $b = 3.62$, 95%CI [2.90, 4.35], $SE = 0.37$, $p < .001$; see Figure 2). These results suggest that participants in the *Distracted* and *Feigned Listening* conditions were indeed devoting a substantial amount of attention to a stimulus other than their partner.

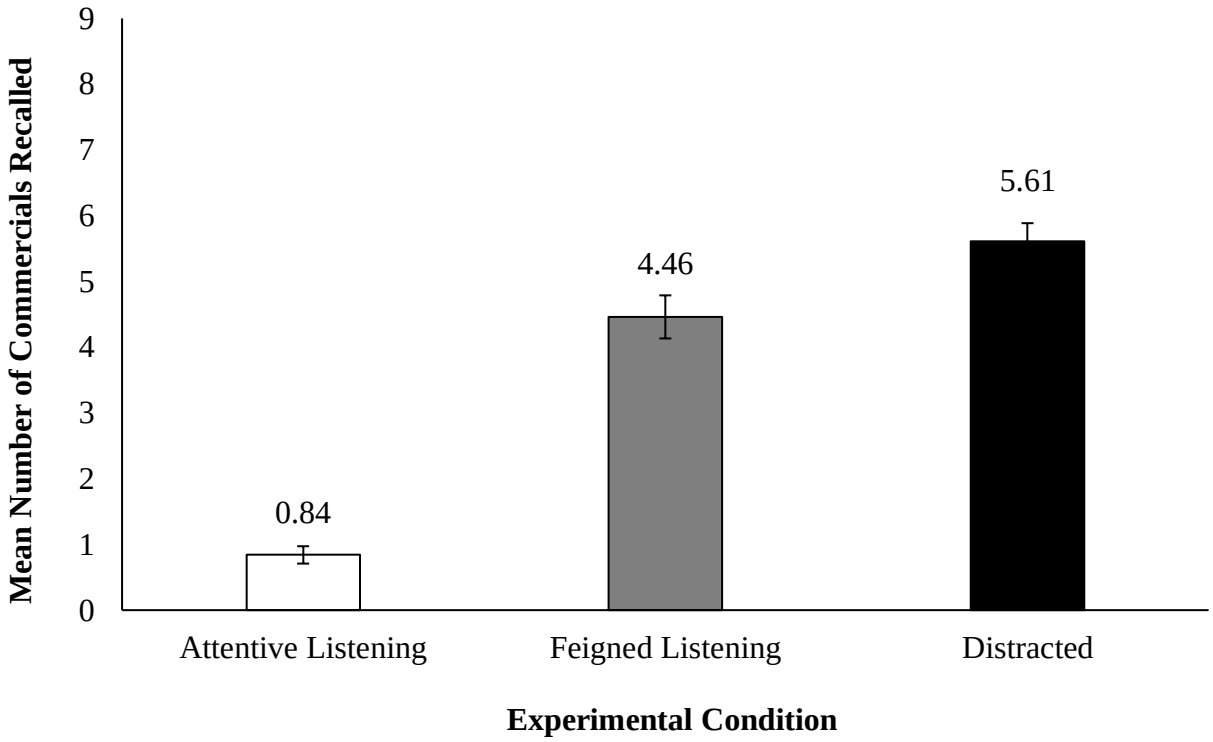


Figure 2. Mean number of commercials correctly recalled by participants across conditions (Study 1), with error bars representing standard errors. Distracted listeners recalled significantly more commercials than feigned listeners, who recalled significantly more than attentive listeners.

Furthermore, manipulated participants believed that the quality of their listening would be easily detected, $F(2, 148) = 5.79$, $p = .004$. Again, using Tukey's test for post-hoc comparisons, participants in the *Distracted* condition thought that their partner would rate them as a significantly worse listener ($M = 4.78$, SD = 1.18) than participants in the *Listening* ($M = 5.32$, SD = 0.77; $b = -0.54$, 95%CI [-0.89, -0.20], $SE = 0.17$, $p = 0.002$) and *Feigned Listening* ($M = 5.25$, SD = 0.56; $b = -0.47$, 95%CI [-0.82, -0.13], SE = 0.17, $p = 0.007$) conditions. The two latter conditions did not differ from each other ($b = -0.07$, 95%CI [-4.09, 2.69], $SE = 0.17$, $p = 0.68$), suggesting that participants in the *Feigned Listening* condition felt confident in their ability to convince their partner that they were listening attentively (Figure 3).

Partner Perceptions. Though participants believed that their differences in attention and listening would be detectable, we find no differences between the three conditions in the partner's evaluations of listening quality ($M_{Listening}$ = 5.36, $SD_{Listening}$ = 0.91; $M_{Distracted}$ = 5.18, $SD_{Distracted}$ = 1.01; $M_{Feigned\ Listening}$= 5.42, $SD_{Feigned\ Listening}$ = 0.82; $F(2, 147) = 0.84$, $p = .43$; Listening vs Distracted: b = -0.18, 95%CI[-0.54, 0.19], SE = 0.19, p = 0.34; Listening vs Feigned Listening: b = 0.06, 95%CI[-0.30, 0.41], SE = 0.18, p = 0.75; Distracted vs Feigned Listening: b = 0.23, 95%CI[-0.13, 0.60], SE = 0.18, p = 0.21). In other words, participants who conversed with a partner who was secretly memorizing commercials rated their listening quality similarly to participants interacting with a fully attentive partner (Figure 3).

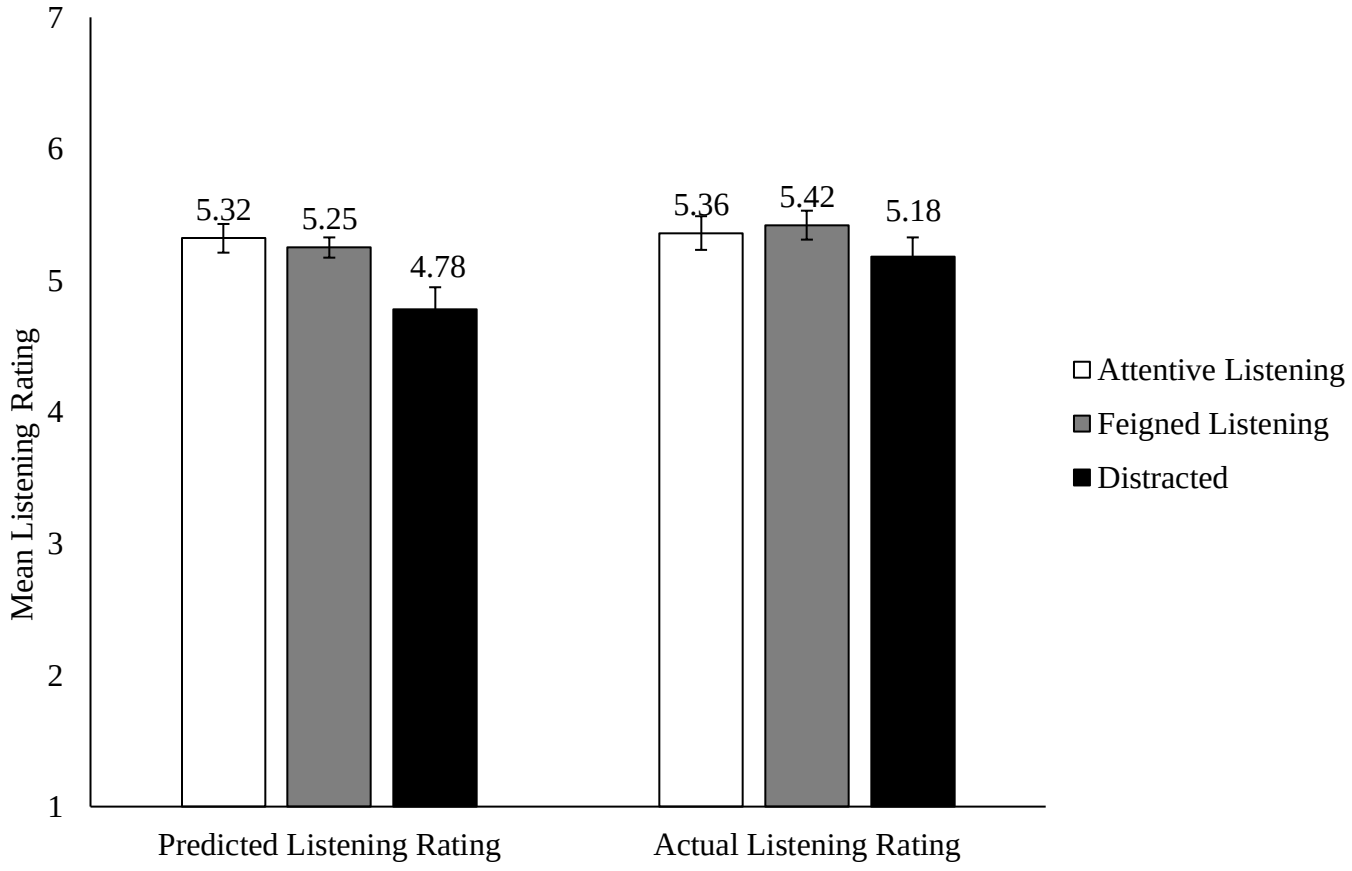


Figure 3. "Actual Listening Ratings" show ratings of listening provided by the unmanipulated partners, while "Predicted Listening Ratings" show how manipulated partners believed their partner would rate them (Study 1). Though participants in the distracted condition predicted that

their partner would rate them as a poorer listener, there were no differences in actual listening ratings across conditions. Error bars represent standard errors.

Similarly, there were no significant differences between the three conditions in ratings of warmth ($M_{Listening}$ = 5.23, $SD_{Listening}$ = 0.69;$M_{Distracted}$ = 5.10, , $SD_{Distracted}$ = 0.82; $M_{Feigned\ Listening}$= 5.20, $SD_{Feigned\ Listening}$ = 0.73; F(2, 147) = 0.37, p = 0.69; Listening vs Distracted: b = -0.12, 95%CI[-0.42, 0.18], SE = 0.15, p = 0.42; Listening vs Feigned Listening: b = -0.02, 95%CI[-0.32, 0.27], SE = 0.15, p = 0.87; Distracted vs Feigned Listening: b = 0.10, 95%CI[-0.20, 0.40], SE = 0.15, p = 0.51), competence ($M_{Listening}$ = 4.83, $SD_{Listening}$ = 0.62; $M_{Distracted}$ = 4.88, $SD_{Distracted}$ = 0.62; $M_{Feigned\ Listening}$= 4.88, $SD_{Feigned\ Listening}$ = 0.63; F(2, 147) = 0.12, p = 0.89; Listening vs Distracted: b = 0.05, 95%CI[-0.20, 0.30], SE = 0.13, p = 0.67; Listening vs Feigned Listening: b = 0.05, 95%CI[-0.20, 0.30], SE = 0.12, p = 0.69; Distracted vs Feigned Listening: b = -0.003, 95%CI[-0.25, 0.24], SE = 0.13, p = 0.98), or power ($M_{Listening}$ = 3.50, $SD_{Listening}$ = 0.77; $M_{Distracted}$ = 3.71, $SD_{Distracted}$ = 0.98; $M_{Feigned\ Listening}$= 3.54, $SD_{Feigned\ Listening}$ = 1.10; F(2, 147) = 0.69, p = 0.50; Listening vs Distracted: b = 0.21, 95%CI[-0.17, 0.60], SE = 0.19, p = 0.27; Listening vs Feigned Listening: b = 0.04, 95%CI[-0.34, 0.42], SE = 0.19, p = 0.82; Distracted vs Feigned Listening: b = -0.17, 95%CI[-0.55, 0.21], SE = 0.1, p = 0.37). Surprisingly, conversation partners in the *Feigned Listening* (M = 5.28, SD = 0.76) condition reported significantly greater enjoyment of the conversation than partners in the *Listening* condition (M = 4.91, SD = 1.03; b = 0.37, 95%CI[0.05, 0.69], SE = 0.16, p = .03), but did not differ from the *Distracted* condition (M = 5.08, SD = 0.63; b = 0.20, 95%CI[-0.13, 0.52], SE = 0.17, p = .24). There were no significant differences in enjoyment between the *Listening* and *Distracted* conditions, b = 0.18, 95%CI [-0.15, 0.50], SE = 0.17, p = .29; F(2, 146) = 2.56, p = .08).

Study 1 Discussion

The results of Study 1 suggest that people's ability to detect others' listening is limited. Though some partners were listening attentively, others were pretending to listen, and others still were largely focusing on something other than the conversation, partners did not discern these differences. Our results may partly stem from the cognitive demands of engaging in live conversation, which is always high, but may be particularly high when interacting with a stranger in an experimental setting. It is possible that these participants were too concerned with their own behavior to accurately judge their partner. We examine this possibility in Study 2.

Study 2: Can Third-Party Observers Detect Listening?

In Study 2, we investigated whether removing the cognitive demands of live conversation might improve peoples' ability to accurately detect listening quality. Specifically, we tested whether third-party observers were able to accurately detect listening between two other people.

Method

Participants. A total of 419 participants (37% Female, Age = 34.82) were recruited through Amazon's Mechanical Turk (mTurk) to complete a 15-minute online survey.

Procedure. In this study, participants watched and evaluated the videos of the conversations we collected in Study 1. When we recorded these videos, we positioned the camera facing the "manipulated" partner in each conversation, 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, we asked participants 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. This process repeated until the video ended.

Measures. Based on prior listening theory (e.g., Bodie et al., 2012) and the open-ended responses of our pilot study participants, we generated a list of verbal and non-verbal behaviors that scientists and laypeople had suggested might represent valid cues of good and bad listening. For each video segment, participants rated the extent to which the manipulated conversation partner engaged in each of those behaviors (detailed below) on a scale from 1: "Not at all" to 4: "A lot."

Participants were asked to report the extent to which the manipulated conversation partner engaged in *verbal interruptions* (cutting the partner off, talking over the partner, etc.) and *verbal affirmations* (defined as 'mhmm', 'uh-huh', 'yes', etc.). Participants were also asked to report the extent to which the manipulated conversation partner engaged in *nodding*, *eye contact*, *looking away* (behind the partner's head, up, down, or to the side), *smiling*, *fidgeting*, *leaning forward*, and *leaning backward*.

Predicting Listener Condition. After watching a full video, participants received an explanation of the lab paradigm that we used to collect the video (including an explanation of the experimental manipulation). We asked participants to guess which condition the manipulated conversation partner had been assigned: attentive, distracted, or feigned listening.

Own and Partner Perceptions. After watching a full video, participants also reported whether they thought the individual was a good listener, and whether they were likeable using the same items as in Study 1. Further, we asked participants to guess the evaluations that the Study 1 conversation partners offered to each of the manipulated participants on the video. Specifically, we were interested in whether mTurk raters thought that each manipulated

participant was seen as a good listener and enjoyable conversation partner using the same scales used in Study 1.

Results

Behavioral Coding. Surprisingly, the mTurk participants did not detect any significant differences across the three conditions in terms of the frequency of any of the verbal or non-verbal behaviors associated with good listening by the prior literature or suggested by our own participants (see Table 1). In other words, conversation partners who were instructed to listen attentively were no more likely to nod, affirm, lean forward, or maintain eye contact than those explicitly assigned to direct their attention elsewhere.

Condition Guess. 36% (95%CI [0.29, 0.42]) of mTurk participants correctly guessed the listener's condition assignment. This differed by video condition ($F(2, 123) = 106.1$, $p < .001$): Significantly more participants were correct when the individual in the video was in the *Listening* condition ($M_{Listening} = 77\%$, $SD_{Listening} = 29\%$) than the *Distracted* ($M_{Distracted} = 13\%$, $SD_{Distracted} = 22\%$; $b = -0.64$, 95%CI[-0.74, -0.538], SE = 0.05, $p < .001$) and *Feigned Listening* conditions ($M_{Feigned\ Listening} = 14\%$, $SD_{Feigned\ Listening} = 17\%$, $b = -0.63$, 95%CI[-0.73, -0.53], SE = 0.05, $p < .001$), with no differences between the *Distracted* and *Feigned Listening* conditions ($b = 0.007$, 95%CI[-0.09, 0.11], SE = 0.05, $p = .90$). This difference was driven by an omnibus tendency: 67% of participants guessed that the individual in the video was in the *Listening* condition (across all conditions). It seems that participants across conditions were quite good at giving the appearance of listening, fooling both their partners and third-party raters.

Own and Partner Perceptions. There were no significant differences across the three conditions in mTurkers' ratings of their own perceptions of listening and liking of the manipulated individual in the video. Similarly, there were no differences in mTurkers'

predictions of how the manipulated conversation participant was perceived by their partner during the live interaction (see Table 1).

	***F*(2, 123)**	***p***	**Attentive Listening *M (SE)***	**Feigned Listening *M (SE)***	**Distracted *M (SE)***
Behavior					
Verbal interruptions	2.54	.08	1.74_{a} (0.07)	1.93_{b} (0.07)	1.75_{ab} (0.07)
Verbal affirmations	0.60	.55	2.89_{a} (0.06)	2.82_{a} (0.07)	2.79_{a} (0.07)
Nodding	1.06	.35	2.95_{a} (0.07)	2.82_{a} (0.07)	2.83_{a} (0.07)
Eye Contact	0.10	.90	3.39_{a} (0.06)	3.38_{a} (0.06)	3.42_{a} (0.06)
Looking Away	0.50	.61	2.01_{a} (0.07)	2.10_{a} (0.08)	2.10_{a} (0.07)
Smiling	1.35	.26	3.02_{a} (0.07)	2.98_{a} (0.08)	3.15_{a} (0.08)
Fidgeting	1.25	.29	2.58_{a} (0.08)	2.69_{a} (0.08)	2.50_{a} (0.08)
Leaning Forward	0.48	.62	1.90_{a} (0.09)	2.01_{a} (0.09)	1.91_{a} (0.09)
Leaning backwards	0.92	.40	1.62_{a} (0.08)	1.76_{a} (0.08)	1.62_{a} (0.08)
Predicted Partner Perceptions					
Listening	0.23	.79	5.06_{a} (0.08)	5.03_{a} (0.08)	5.10_{a} (0.08)
Liking	0.28	.76	4.67_{a} (0.08)	4.73_{a} (0.08)	4.75_{a} (0.08)
Enjoyment	0.07	.93	4.92_{a} (0.08)	4.87_{a} (0.09)	4.88_{a} (0.08)
Own Perceptions					
Listening	0.14	.87	4.69_{a} (0.12)	4.68_{a} (0.12)	4.76_{a} (0.12)
Liking	0.83	.44	4.72_{a} (0.09)	4.78_{a} (0.09)	4.62_{a} (0.09)

*Table 1***.** Means in each row with different subscripts were significantly different at the $p < .05$ level. For example, a mean with subscript 'a' differs from a mean with subscript 'b'.

Study 2 Discussion

Even when people did not bear the cognitive load of active conversation, they were unable to accurately detect when someone was listening or merely pretending to listen to their conversation partner. Indeed, our raters reported no differences between conditions on any verbal or non-verbal behaviors, and were woefully inaccurate when asked to guess the condition assignment of the targets. Further, these results help clarify our understanding of Study 1—it

does not appear that manipulated participants provided obvious behavioral cues of inattention that their partner simply missed, but rather, inattentive listeners were either able to display the same behavioral cues of listening as their attentive counterparts, or their behavioral cues of inattentive listening were sufficiently subtle that even third-party viewers couldn't readily perceive them.

Another possible explanation for this surprising result is that people are particularly poor at detecting the listening of strangers. This finding would still be quite important, given how much of time people spend in important conversations with new people (e.g. doctors, potential employers, first dates, new students). But in Study 3, we examine whether people can accurately detect listening by the one person they are most familiar with: themselves.

Study 3: Can People Perceive Their Own Listening in Retrospect?

In studies 1 and 2, people were unable to accurately detect listening—both those actively engaged in a conversation and dispassionate third-party observers. In Study 3, we test an even more extreme version of our question: Can people detect when they *themselves* are listening?

Method

Participants. 90 participants (M_{age} = 37, 49% female) were recruited to our behavioral lab to participate in a 30-minute study.

Procedure. In this study, we used a within-subjects design. Participants were seated in a room with an experimenter, who read two stories to each participant, with story order counterbalanced, while music was playing in the room. We selected two songs ("I am my own grandpa," and "Big rock candy mountain") that were obscure enough that participants were unlikely to be familiar with them, but used plain language that would be easy to understand and remember. Participants received different instructions about how to direct their attention and

behave while each story was read, according to the experimental condition to which they were assigned. All participants were told to listen attentively to one story, and were randomly assigned to one of two conditions for the second story (described below). Participants were video recorded as they listened.

Attentive Listening Instructions. During one of the stories, participants were asked to "listen as attentively as possible to the story." They were told that they would answer comprehension questions about the story and would receive a $1.00 bonus for each question they answered correctly.

Inattentive-Listening Instructions. During the other story (counterbalanced), all participants were incentivized to experience one of two levels of cognitive distraction, which we manipulated between participants. In the *Inattentive-Listening* condition (n = 45), participants were asked to "listen as attentively as possible to the song playing in the room" and that they would receive a $1.00 for each line of lyrics that they correctly recalled. In the *Semi-Attentive Listening* condition (n = 45), participants were asked to "listen as attentively as possible to the story AND the song playing in the room" and that they would receive a $1.00 bonus for each comprehension question they answered *and* each line of song lyrics they correctly recalled.

Manipulation Check. Participants completed seven comprehension questions about the story their partner read to them, as well as seven fill-in-the-blank questions about the lyrics of the song playing in the room. Further, after completing each listening task, participants reported how "attentively" they listened to the story, and how "distracted" they were while their partner was reading the story (1: "*Not at all"*; 5: "*Extremely"*).

Detecting their Own Listening. After participants finished the listening task, they were shown 10 separate 5-second muted video clips of themselves listening (5 clips from each story),

a methodology borrowed from "thin-slice" research (e.g., Ambady & Rosenthal, 1992; Ambady, Bernieri, & Richeson, 2000). For each video clip, participants were asked to guess which set of instructions they were following in the video. Then participants were asked "How confident are you in your answer?" (1=Not at all confident; 7=Extremely confident). Finally, participants provided a guess of the number of the 10 clips for which they correctly identified their condition assignment.

Results

Manipulation Check. Our listening manipulation was successful: Participants in the *Inattentive Listening* condition reported listening more attentively to the story when they were instructed to do so (M = 5.69, SD = 1.12) than when they were instructed to listen to the song (M = 3.18, SD = 1.85; $t(44) = 9.03$, $p < .001$, Cohen's d = 1.60, 95%CI[1.07, 2.13]). Similarly, participants in the *Semi-Attentive Listening* condition reported listening more attentively to the story when they were instructed to do so (M = 5.91, SD = 1.00) than when they were instructed to listen to both the story and the song (M = 5.16, SD = 1.28, t(44) = 4.19, $p < .001$, Cohen's d = 0.65, 95%CI[0.31, 0.99]). Additionally, participants in the *Inattentive Listening* condition reported feeling more distracted from the story when they were instructed to listen to the song (M = 4.73, SD = 1.74) than when they were instructed to listen to the story (M = 3.67, SD = 1.49, t(44) = 4.28, $p < .001$, Cohen's d = 0.66, 95%CI[0.32, 0.99]). Those in the *Semi-Attentive Listening* condition were similarly more distracted when they were instructed to listen to the song and story (M = 4.87, SD = 1.70) than when they were instructed to only listen to the story (M = 3.91, SD = 1.87, t(44) = 4.17, $p < .001$), Cohen's d = 0.53, 95%CI[0.26, 0.80]. Across both methods of distraction, participants felt more attentive and less distracted when they were instructed to listen only to the story.

Looking at the story comprehension questions, participants in the *Inattentive Listening* condition answered more story comprehension questions correctly when they were asked to listen to the story (M = 3.78, SD = 1.61) than when they were asked to listen to the song (M = 2.60, SD = 1.37, $t(44) = 3.70$, $p < .001$, Cohen's d = 0.79, 95%CI[0.30, 1.27]). However, those in the *Semi-Attentive Listening* condition answered a similar number of story comprehension questions correctly when they were asked to listen to the story (M = 3.56, SD = 1.37) as when they were asked to listen to both the story *and* the song (M = 3.04, SD = 1.43, $t(44) = 1.70$, p = 0.10, Cohen's d = 0.36, 95%CI[-0.08, 0.80]).

Further, participants in the *Inattentive Listening* condition recalled more song lyrics correctly when they were asked to listen to the song (M = 2.11, SD = 1.70) than when they were asked to listen to the story (M = 0.33, SD = 0.56, $t(44) = 6.74$, p < .001, Cohen's d = 1.40, 95%CI[0.82, 1.98]). Additionally, participants in the *Semi-Attentive Listening* condition recalled more song lyrics correctly when they were asked to listen to both the story and the song (M = 1.31, SD = 1.41), than when they were asked to listen only to the story (M = 0.69, SD = 1.00, $t(44) = 2.85$, p = .007, Cohen's d = 0.50, 95%CI[0.13, 0.88]).

In sum, these results suggest that our manipulations were effective. Both self-report and behavioral measures suggest that participants paid less attention to the story when instructed to do so.

Listening Detection. First, we assess participants' self-reported confidence in their guesses. In the *Inattentive Listening* condition, participants felt similarly confident in their guesses for clips in which they were instructed to listen to the story (M = 5.33, SD = 1.39) and when they were instructed to listen to the song only (M = 5.40, SD = 1.36; $t(44) = 0.46$, p = .65, Cohen's d = 0.05, 95%CI[-0.16, 0.26]). Participants in the *Semi-Attentive Listening* condition

were also equally confident in their guesses for clips in which they were instructed to listen to the story (M = 4.95, SD = 1.30) and when they were instructed to listen to the story and the song (M = 4.91, SD = 1.39; $t(44) = 0.37$, $p = .71$, Cohen's d = 0.03, 95%CI[-0.13, 0.19]).

Next, we turn to participants' actual guesses. Participants in the *Inattentive Listening* condition (who attended only to the song) guessed that they were listening more often during clips in which they were instructed to listen to the story (M = 3.78, SD = 1.28) than when they were only instructed to listen to the song (M = 1.96, SD = 1.59; $t(44) = 4.93$, $p < .001$, Cohen's d = 1.27, 95%CI[0.58, 1.95]). Participants in the *Semi-Attentive Listening* condition guessed they were listening more often during clips when they were instructed to listen to the story (M = 3.04, SD = 1.35) compared to clips when they were instructed to listen to the song (M = 2.22, SD = 1.58; $t(44) = 2.71$, $p = .01$, Cohen's d = 0.56, 95%CI[0.12, 1.00]).).

Although statistically significant, the results are not encouraging regarding the extent of participants' discernment. Specifically, participants only guessed their condition assignment correctly in 63% of the video clips that they viewed. In other words, participants could not accurately detect their own state of listening in over one third of the trials. When participants viewed clips in which they were instructed to listen to the story (*Attentive Listening*), they guessed this to be the case 68% of the time (76% in the *Inattentive Listening* condition; 61% in the *Semi-Attentive Listening* condition). When they were instructed to listen to both the story and the song (*Semi- Attentive Listening*), they still guessed that they were instructed to listen to the story 44% of the time. Even when participants were only instructed to listen to the song (*Inattentive Listening*), they guessed they had been asked to listen to the story 39% of the time (see Table 2).

	Attentive Listening ***M***	**Semi-Attentive Listening** ***M***	**Inattentive Listening** ***M***
Percentage of clips identified as attentive listening	68%	**44%**	**39%**
Percentage of clips identified as inattentive- or semi-attentive listening	**32%**	56%	61%

Table 2. Percentage of clips identified by participants as instances of attentive listening and semi- or inattentive listening (depending on condition). Bolded numbers represent participants' inaccurate identifications.

These data suggest that some signal is coming through. Participants who viewed a clip of themselves listening attentively to a story were significantly more likely to guess that they were indeed listening than participants who viewed a video of themselves fully or somewhat distracted by music. Although statistically significant, these results are not particularly encouraging regarding the extent of this ability – participants still thought they were listening attentively more than one third of the time when they actually were not.

Study 3 Discussion

The results of Study 3 suggest that people do not *accurately* detect listening, even when they observe their *own* nonverbal cues immediately after engaging in attentive, inattentive, or semi-attentive listening. In particular, mirroring the results of Studies 1-2, people radically over-estimate (their own) attentive listening. These results are particularly striking because our participants had just completed the listening task minutes before they were asked to guess whether they were listening or not. However, it is possible that people are extraordinarily good at dividing their attention (e.g., between a TV screen and a conversation partner or between simultaneous audio inputs like a story and ambient music). We examine this possibility further in Study 4.

Study 4: How well do people know when others cannot hear them?

In Studies 1 and 3, we guided participants' listening behavior with *instructions* (to listen attentively, inattentively, or semi-attentively). However, it's possible that the human mind is highly capable of dividing its attention between multiple stimuli, and that all participants in Studies 1 and 3—even those whose attention was divided—were able to process their partner's words (even while simultaneously attending to video advertisements or music lyrics).

In Study 4, we tested this explanation by strictly limiting participants' *ability to hear* their conversation partner's words—by garbling portions of the conversation. We incentivized "listeners" to act as if the conversation was proceeding without disruption—their payment depended on maintaining their partner's ignorance about the sporadically garbled content. This design helps us answer an important question: how pervasive can lapses in hearing and listening become before conversation partners begin to notice? As previously noted, recall (which was used in the previous studies) is an imperfect measure of listening (Thomas & Levine, 1994), thus we circumvent this challenge completely in Study 4. All procedures and analyses were pre-registered (https://aspredicted.org/blind.php?x=sf5nz6).

Method

Participants. 229 participants[2] (M_{age} = 32, 52% female) were recruited to our behavioral lab in groups of 4-6 to participate in a 45-minute study.

Protocol. Participants completed 10 min, one-on-one, round-robin video chats with 2-3 different partners. Conducting these conversations over video chat (instead of face-to-face) allowed us to asymmetrically manipulate audio for one participant in each dyad. The number of participants that arrived for each experimental session determined the number of conversations

[2] We were unable to reach our pre-registered sample size due to the COVID-19 pandemic. This study was actively running in the lab when local shelter-in-place restrictions required data collection to stop.

each participant completed (we maximized the number of unique round-robin dyads possible in each session). Though all participants were tasked with getting to know their conversation partners, they also received private instructions. Half the participants in each session were randomly assigned to the role of "Listener," the other half to the role of "Partner."

Listeners. Those assigned as Listeners learned that there may be times during the conversation when they couldn't understand their partner—instead of hearing their partner's words, they would hear garbled sound, and this garbled sound was an *intentional* part of the study, not a technical glitch. By design, we used a computer program to obscure what the Listeners (but not their Partners) could hear during the conversation. Unbeknownst to the Listeners, we randomly assigned them (at the session level) to one of four conditions corresponding to the amount of time the 10 min conversation would be garbled from their perspective only: 0% (0 seconds garbled), 25% (150 seconds garbled), 50% (300 seconds garbled), or 75% (450 seconds garbled). We configured the computer program to intermittently turn on and off a voice filter that obscured what the Listeners heard from their partner at specified intervals (25% garbled condition: 30 seconds filter off, 10 seconds filter on; 50% garbled condition: 10 seconds filter off, 10 seconds filter on; 75% garbled condition: 3 seconds filter off; 10 seconds filter on), which repeated throughout the 10-minute conversation. Random assignment to level of garbling (0%, 25%, 50%, 75%) was performed at the session level.

Our manipulation allowed us to test whether people can give the impression that they are listening, even when they cannot actually hear the words that their partner is saying, thus testing if divided attention could explain our earlier results. Importantly, we incentivized Listeners to act as if everything was normal:

> *"Your primary goal is to be (or at least appear to be) a GREAT LISTENER. After each conversation, your partner will rate how well they think you listened to them. If your*

> *partner reports that you were a good listener (i.e., one of the 40 best listeners in the study)…then you will earn a $20 bonus…In order to earn the "good listener" bonus, your partner should think the conversation has played out naturally and smoothly."*

Partners. Those assigned as Partners were completely ignorant of the Listeners' instructions and the conversational garbling. But for fairness in payment potential, the Partners were also incentivized:

> *"Your primary goal is to be likeable. After each conversation, your partner will rate how much they liked you. If your partner reports that you were highly likeable (i.e., one of the 40 most likeable people in the study)…then you will earn a $20 bonus."*

In this manner, both listeners and partners were financially incentivized to make a positive impression on each other.

After each 10-minute conversation, participants completed a post-conversation survey, which included self-reported items about the conversation and their partner. At the end of the lab session, participants completed a final demographic survey and received payment.

Measures completed by Partners. After each conversation, the Partners (who were blind to the manipulation) reported their perceptions of their (manipulated) counterpart's listening ("My partner was a good listener," "My partner was engaged in the conversation") and responsiveness ("My partner made me feel heard," "My partner made me feel validated," "I felt that my partner cared about me") on a scale from 1: "*Strongly Disagree"* to 7: "*Strongly Agree*." They also reported the extent to which they agreed or disagreed that their partner "worked hard to listen to me," "was attentive to what I was saying," and "understood what I was saying" (1: "*Strongly Disagree"*; 7: "*Strongly Agree*)". These evaluations represent our key dependent variables in this study. Partners also estimated the percentage of the things they said during each conversation that they believe their listener heard (1-100%).

Partners also rated their general assessments of their (manipulated) counterpart, including likeability ("My partner is likable," "I liked my partner," "I would enjoy spending time with my partner," "I disliked my partner (R)"), intelligence ("My partner is smart"), and interestingness ("My partner is interesting") on a scale from 1: "*Strongly Disagree"* to 7: "*Strongly Agree*."

Measures completed by Listeners. Listeners made predictions about their partner's perceptions of them. They predicted their partner's perceptions of their listening ("My partner thought I was a good listener," and "My partner thought I was engaged in the conversation") and responsiveness ("My partner felt heard," "My partner felt validated," and "My partner felt that I cared about them") on a scale from 1: "*Strongly Disagree"* to 7: "*Strongly Agree*." Listeners also reported whether they "worked hard to listen to my partner," "was attentive to what my partner was saying," and "understood what my partner was saying" on a scale from 1: "*Strongly Disagree* "to 7: "*Strongly Agree*." Listeners also reported whether they thought their partner would say that they could hear them (yes/no) and guessed "My partner would say that I could hear ___% of what they said" (1-100%). These measures served as manipulation checks.

Finally, Listeners predicted how likeable their partner would rate them as being ("My partner thinks I'm likable," "My partner liked me," "My partner would enjoy spending time with me," "My partner dislikes me," 1: "*Strongly Disagree"* to 7: "*Strongly Agree"*).

Both participants were aware that their ratings of each other would remain private so as to assuage any impression management concerns.

Results

After excluding participants who expressed confusion about the instructions to the experimenters or disclosed the listening manipulation to their partner (n = 9), we analyzed data from 305 conversations (n_0 = 80; n_{25} = 87; n_{50} = 70; n_{75} = 68). Because each participant engaged

in several conversations, we conducted mixed-model regressions, clustering at the participant level and controlling for order effects to account for repeated measures using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015). We performed separate contrasts between all conditions, using the "eff" method in the 'emmeans' package, which is a Bonferroni-based method, and thus is considered slightly conservative (Lenth, 2020). We report specific results for each dependent variable, specifying the results of each pairwise comparison. All results are presented in aggregate in Table 3.

	Experimental Condition: Percentage of Garbled Conversational Content			
	0% ***M (SE)***	**25%** ***M (SE)***	**50%** ***M (SE)***	**75%** ***M (SE)***
Partner Ratings				
Listening	6.22 (0.17)$_a$	6.17 (0.15)$_a$	6.19 (0.17)$_a$	6.01 (0.17)$_a$
Responsiveness	5.98 (0.19)$_a$	5.72 (0.18)$_{ab}$	5.99 (0.19)$_a$	5.44 (0.20)$_b$
Worked Hard	5.75 (0.19)$_a$	5.96 (0.17)$_a$	6.13 (0.19)$_a$	5.66 (0.20)$_a$
Attentive	6.16 (0.17)$_a$	6.13 (0.16)$_a$	6.23 (0.18)$_a$	5.86 (0.18)$_a$
Understood	6.16 (0.19)$_a$	6.12 (0.17)$_a$	6.12 (0.19)$_a$	5.40 (0.19)$_b$
Guess Percent Heard	88.40 (2.54)$_a$	85.70 (2.35)$_a$	81.90 (2.61)$_a$	74.30 (2.64)$_b$
Liking	6.00 (0.18)$_a$	6.06 (0.16)$_a$	5.99 (0.18)$_a$	5.71 (0.18)$_a$
Intelligence	6.34 (0.18)$_a$	6.16 (0.16)$_{ab}$	6.08 (0.18)$_{ab}$	5.83 (0.18)$_b$
Interestingness	5.98 (0.21)$_a$	5.94 (0.19)$_a$	5.93 (0.21)$_a$	5.54 (0.21)$_a$
Listener Ratings				
Predicted Listening	6.23 (0.16)$_a$	6.23 (0.16)$_a$	5.94 (0.17)$_{ab}$	5.58 (0.17)$_b$
Predicted Responsiveness	6.05 (0.16)$_a$	6.02 (0.15)$_a$	5.88 (0.17)$_{ab}$	5.44 (0.17)$_b$
Predicted Liking	5.57 (0.14)$_a$	6.18 (0.14)$_b$	5.99 (0.15)$_b$	5.58 (0.15)$_a$
Worked Hard	5.96 (0.19)$_a$	6.61 (0.18)$_b$	6.75 (0.20)$_b$	6.62 (0.20)$_b$
Attentive	6.54 (0.15)$_a$	6.64 (0.14)$_a$	6.60 (0.16)$_a$	6.43 (0.16)$_a$
Understood	6.60 (0.24)$_a$	5.52 (0.23)$_b$	4.62 (0.25)$_c$	3.35 (0.25)$_d$
Percent Heard	94.0 (2.66)$_a$	68.90 (2.56)$_b$	50.60 (2.83)$_c$	26.90 (2.85)$_d$
Guess Partner Rating of Percent Heard	78.70 (9.57)$_a$	78.00 (5.66)$_a$	67.20 (6.31)$_{ab}$	50.50 (5.35)$_b$

Table 3. Means in each row with different subscripts were significantly different at the $p < .05$ level. For example, a mean with subscript 'a' differs from a mean with subscript 'b'.

Was the listening manipulation successful? Several results demonstrate that our listening manipulation was successful. First, listeners' reports of the percentage of the conversation that they heard decreased significantly with each increase in garbling ($M_{0\%}$ = 94.20, $M_{25\%}$ = 68.90, $M_{50\%}$ = 50.40, $M_{75\%}$ = 27.00; Figure 4; 0 vs 25%: b = -25.20, 95%CI[-32.50, 18.00], p < .001; 25 vs. 50%: b = -18.60, 95%CI[-26.00, -11.10], p < .001; 50 vs 75%: b = -23.30, 95%CI[-31.20, -15.50], p < .001). Further, listener's self-reported understanding during the conversation was significantly different between all conditions, diminishing with each increase in the percentage of speech that was obscured ($M_{0\%}$ = 6.59, $M_{25\%}$ = 5.52, $M_{50\%}$ = 4.64, $M_{75\%}$ = 3.34; Figure 5; 0 vs 25%: b = -1.07, 95%CI[-1.71, -0.42], p = .002; 25 vs. 50%: b = -0.88, 95%CI[-1.55, -0.21], p = .01; 50 vs 75%: b = -1.30, 95%CI[-2.01, -0.59], p < .001).

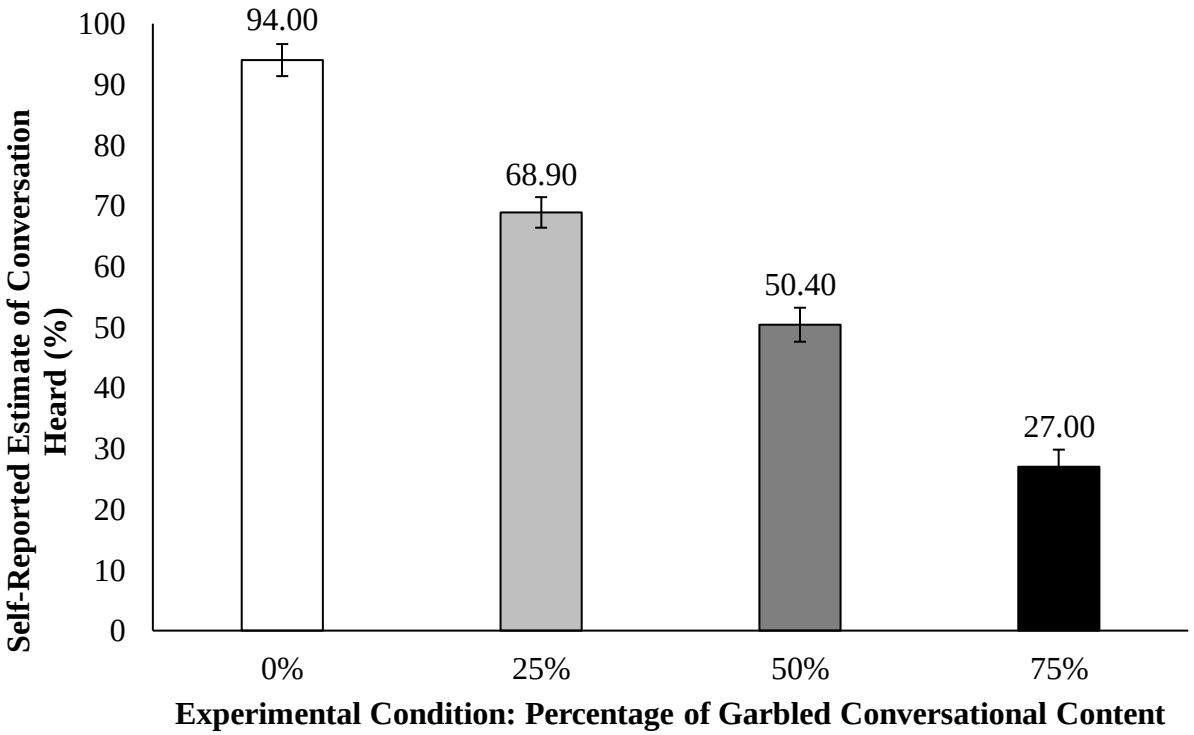


Figure 4. Listeners estimated the percentage of conversational content they heard, with error bars representing standard errors (Study 4). Participants' self-reported estimates were remarkably accurate. Those who heard 100% of the content estimated they heard 94%, those who heard 75%

estimated 69%, those who heard 50% estimated 50%, and those who heard 25% estimated 27%. Each condition significantly differed from all others.

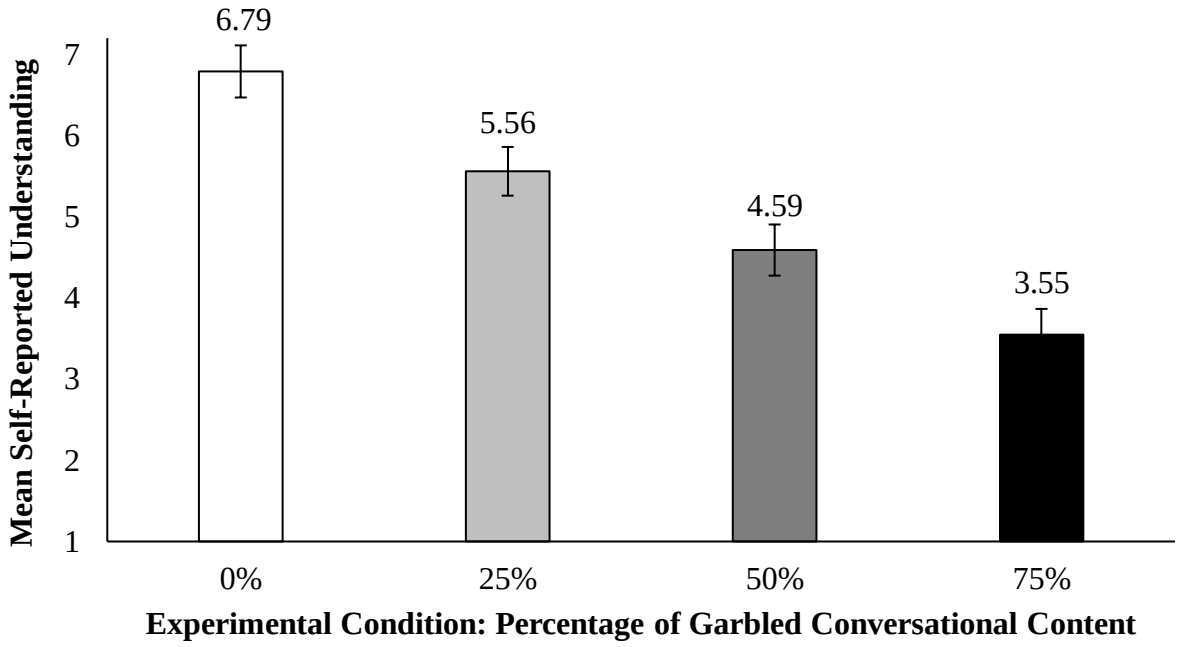


Figure 5. Listeners' mean self-reported understanding during the conversation, with error bars representing standard errors (Study 4). Each condition significantly differed from all others.

Additionally, listeners in the 25%, 50%, and 75% garbled conditions reported working harder to listen to their partner than those in the 0% condition (0 vs 25%: b = 0.67, 95%CI[0.15, 1.19], p = .01; 0 vs. 50%: b = 0.83, 95%CI[0.28, 1.37], p = .003; 0 vs 75%: b = 0.69, 95%CI[0.14, 1.24], p = .01), though there were no significant differences on this measure between the obscured conditions ($M_{0\%}$ = 5.94, $M_{25\%}$ = 6.61, $M_{50\%}$ = 6.77, $M_{75\%}$ = 6.63; 25 vs 50%: b = 0.16, 95%CI[-0.38, 0.69], p = .56; 25 vs. 75%: b = 0.02, 95%CI[-0.52, 0.56], p = .94; 50 vs 75%: b = -0.13, 95%CI[-0.70, 0.42], p = .63).

Importantly, listeners reported no differences in their attentiveness during the conversation across the conditions ($M_{0\%}$ = 6.54, $M_{25\%}$ = 6.64, $M_{50\%}$ = 6.59, $M_{75\%}$ = 6.42; 0 vs

25%: b = 0.11, 95%CI[-0.30, 0.51], p = .60; 25 vs. 50%: b = -0.05, 95%CI[-0.46, 0.37], p = .82; 50 vs 75%: b = -0.17, 95%CI[-0.61, -0.27], p = .44; all other ps > .30).

Did listeners think they could fake it? Comparing listener's predictions of their partner's perceptions of their listening ($M_{0\%}$ = 6.22, $M_{25\%}$ = 6.23, $M_{50\%}$ = 5.94, $M_{75\%}$ = 5.57) and responsiveness ($M_{0\%}$ = 6.04, $M_{25\%}$ = 6.02, $M_{50\%}$ = 5.88, $M_{75\%}$ = 5.42), we find that participants in the 75% condition predicted that their partner would rate them as a poorer listener than those in the 0% (b = -0.65, 95%CI[-1.12, -0.17], p = .008) and 25% conditions (b = -0.66, 95%CI[-1.12, -0.20], p = .006), and as less responsive than in the 0% (b = -0.62, 95%CI[-1.08, -0.16], p = .009) and 25% conditions (b = -0.60, 95%CI[-1.05, -0.15], p = .01), but show no differences from the 50% condition (listening: b = -0.37, 95%CI[-0.85, 0.12], p = .14; responsiveness: b = -0.45, 95%CI[-0.93, 0.02], p = .06). Additionally, when asked to predict how likable their partner would find them, participants in the 0% condition actually predicted their partner would find them less likable than those in the 25% (b = 0.62, 95%CI[0.22, 1.01], p = .002) and 50% (b = 0.42, 95%CI[0.001, 0.83], p = 0.0497) conditions, but would be similarly likable to those in the 75% condition (b = 0.00007, 95%CI[-0.42, 0.42], p > .99; $M_{0\%}$ = 5.56, $M_{25\%}$ = 6.18, $M_{50\%}$ = 5.98, $M_{75\%}$ = 5.56). Further, those in the 75% condition also felt they would be seen as less likable than those in the 25% condition (b = -0.62, 95%CI[-1.02, -0.21], p = .003) and marginally less so than those in the 50% (b = -0.42, 95%CI[-0.84, 0.01], p = .06) condition.

Finally, when comparing listeners' predictions of how much of the conversation their partner thought they heard ($M_{0\%}$ = 78.70, $M_{25\%}$ = 78.00, $M_{50\%}$ = 67.20, $M_{75\%}$ = 50.50), participants in the 75% reported that their partner would think they heard significantly less than in the 0% and 25% conditions (0 vs 75%: b = -28.20, 95%CI[-50.50, -5.92], p = .01; 25 vs 75%: b = -27.52, 95%CI[-43.40, -11.61], p = .001), and marginally less than in the 50% condition (b =

-16.71, 95%CI[-33.50, 0.11], p = .05), with no other between-condition differences (ps > .20). These results suggest that, except for those whose conversations were 75% garbled, most listeners felt they were able to convince their partner of their listening.

Did evaluators notice? As in Studies 1-3, we find no differences across conditions in (unmanipulated) Partners' perceptions of their partner's (manipulated) listening ($M_{0\%}$ = 6.24, $M_{25\%}$ = 6.17, $M_{50\%}$ = 6.19, $M_{75\%}$ = 6.00; all ps > .34; Figure 6; 0 vs 25%: b = -0.07, 95%CI[-0.44, 0.31], p = .77; 0 vs. 50%: b = -0.04, 95%CI[-0.44, 0.35], p = .85; 0 vs. 75%: b = -0.23, 95%CI[-0.63, 0.16], p = .33; all other ps > .43).

Further, when reporting how responsive their partner was in the conversation, partners rated listeners in the 0%, 25% and 50% conditions as similarly responsive ($M_{0\%}$ = 6.00, $M_{25\%}$ = 5.72, $M_{50\%}$ = 5.99, $M_{75\%}$ = 5.44; 0 vs 25%: b = -0.28, 95%CI[-0.78, 0.22], p = .28; 0 vs. 50%: b = -0.005, 95%CI[-0.53, 0.52], p = .99; 25 vs. 50%: b = 0.27, 95%CI[-0.24, 0.79], p = .30), and only when the listener heard 75% of the conversation garbled were they rated as slightly less responsive, although these differences barely attained traditional levels of significance (0 vs 75%: b = -0.56, 95%CI[-1.09, -0.03], p = .04; 25 vs 75%: b = -0.28, 95%CI[-0.80, 0.23], p = .30; 50 vs 75%: b = -0.5, 95%CI[-1.10, -0.01], p = .045; Figure 6).

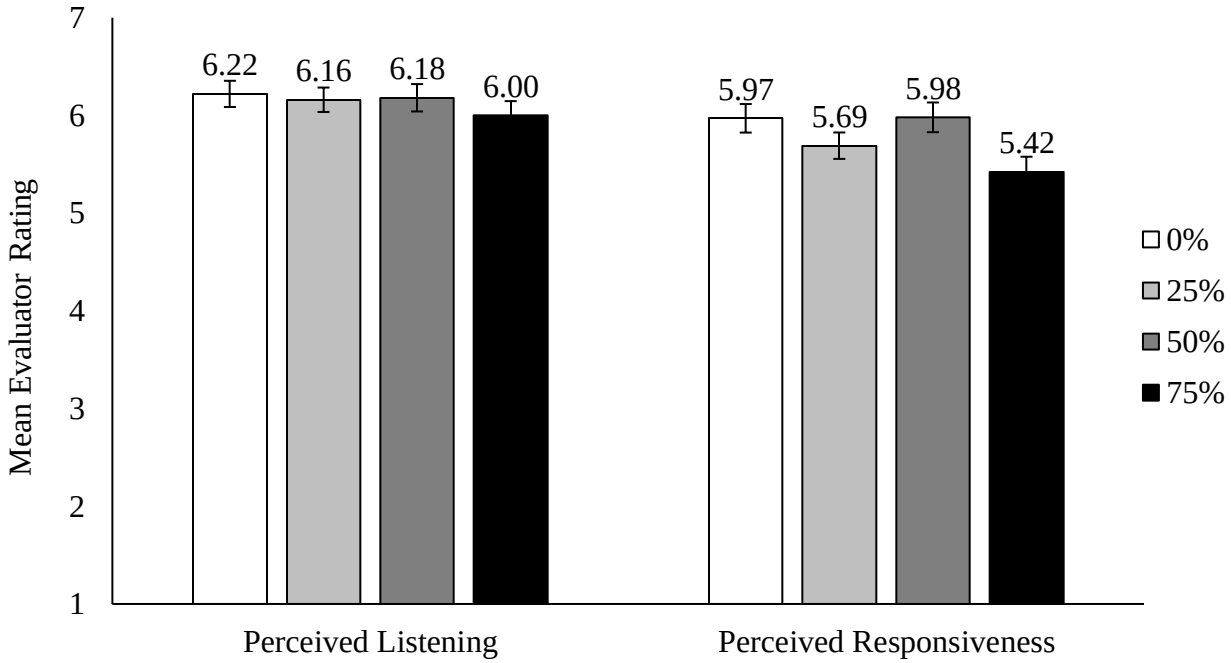


Figure 6. Evaluators' mean ratings of their partner's listening and responsiveness across conditions, with error bars representing standard errors. There were no significant differences in perceived listening across conditions. Perceived responsiveness was significantly lower in the 75% garbled condition, compared to the 0% and 50% (but not 25% garbled) conditions.

Partners' ratings of Listeners' effort ($M_{0\%}$ = 5.79, $M_{25\%}$ = 5.96, $M_{50\%}$ = 6.13, $M_{75\%}$ = 5.66; 0 vs 25%: b = 0.17, 95%CI[-0.33, 0.68], p = .50; 0 vs. 50%: b = 0.35, 95%CI[-0.19, 0.88], p = .20; 0 vs. 75%: b = -0.13, 95%CI[-0.66, 0.41], p = .64; all other ps > .09) and attentiveness ($M_{0\%}$ = 6.18, $M_{25\%}$ = 6.13, $M_{50\%}$ = 6.24, $M_{75\%}$ = 5.85; 0 vs 25%: b = -0.05, 95%CI[-0.51, 0.42], p = .84; 0 vs. 50%: b = 0.06, 95%CI[-0.43, 0.55], p = .80; 0 vs. 75%: b = -0.32, 95%CI[-0.82, 0.17], p = .20; all other ps > .13) during the conversation also did not differ across conditions. Like ratings of responsiveness, we find no differences between the 0%, 25% and 50% conditions on perceptions of the listeners' understanding (0 vs 25%: b = -0.07, 95%CI[-0.56, 0.43], p = .79; 0 vs. 50%: b = -0.06, 95%CI[-0.58, 0.47], p = .83; 25 vs. 50%: b = 0.008, 95%CI[-0.50, 0.52], p = .97). Only when Listeners heard 75% of the conversation garbled did Partners perceive that the Listeners' understanding suffered (0 vs 75%: b = -0.79, 95%CI[-1.32, -0.26], p = .004; 25 vs.

75%: b = -0.72, 95%CI[-1.23, -0.21], p = .006; 50 vs. 75%: b = -0.73, 95%CI[-1.2, -0.19], p = .008; $M_{0\%}$ = 6.18, $M_{25\%}$ = 6.12, $M_{50\%}$ = 6.13, $M_{75\%}$ = 5.39).

When directly asked to estimate the percentage of the conversation their partner was able to hear, (unmanipulated) Partners made statistically equivalent estimates across the 0%, 25% and 50% garbled conditions (0 vs 25%: b = -2.73, 95%CI[-9.52, 4.06], p = .43; 0 vs. 50%: b = -6.49, 95%CI[-13.64, 0.67], p = .08; 25 vs. 50%: b = -3.76, 95%CI[-10.72, 3.20], p = .29). Though Partners guessed that Listeners heard less in the 75% condition, even in that condition, Partners massively underestimated their Listeners' inability to hear, guessing that they could *hear* 75% of the conversation, when they could only hear 25% of it (0 vs 75%: b = -14.18, 95%CI[-21.36, -7.01], p = .002; 25 vs. 75%: b = -11.45, 95%CI[-18.44, -4.47], p = .002; 50 vs. 75%: b = -7.70, 95%CI[-15.03, -0.36], p = .04; $M_{0\%}$ = 88.40%, $M_{25\%}$ = 85.70%, $M_{50\%}$ = 81.90%, $M_{75\%}$ = 74.20%).

Finally, when asked to make person-level judgments about the Listeners, the unmanipulated Partners reported no differences in liking ($M_{0\%}$ = 6.03, $M_{25\%}$ = 6.06, $M_{50\%}$ = 5.99, $M_{75\%}$ = 5.71; 0 vs 25%: b = -0.03, 95%CI[-0.43, 0.50], p = .89; 0 vs. 50%: b = -0.04, 95%CI[-0.53, 0.46], p = .89; 0 vs. 75%: b = -0.31, 95%CI[-0.81, 0.18], p = .21; all other ps > .15) or judgments of interestingness ($M_{0\%}$ = 6.00, $M_{25\%}$ = 5.94, $M_{50\%}$ = 5.93, $M_{75\%}$ = 5.54; 0 vs 25%: b = -0.06, 95%CI[-0.60, 0.49], p = .84; 0 vs. 50%: b = -0.07, 95%CI[-0.65, 0.51], p = .81; 0 vs. 75%: b = -0.47, 95%CI[-1.04, 0.11], p = .11; all other ps > .15). Judgments of intelligence ($M_{0\%}$ = 6.35, $M_{25\%}$ = 6.16, $M_{50\%}$ = 6.08, $M_{75\%}$ = 5.83; 0 vs 25%: b = -0.20, 95%CI[-0.66, 0.27], p = .40; 0 vs. 50%: b = -0.28, 95%CI[-0.77, 0.21], p = .26; 25 vs. 50%: b = -0.08, 95%CI[-0.56, 0.39], p = .74) only suffered for those in the 75% condition, who were rated as less smart than those in the

0% condition (0 vs 75%: b = -0.53, 95%CI[-1.02, -0.04], p = .04; 25 vs. 75%: b = -0.33, 95%CI[-0.81, 0.15], p = .17; 50 vs. 75%: b = -0.25, 95%CI[-0.75, 0.25], p = .33).

Discussion

The results of Study 4 suggest that perceptions of listening are surprisingly impervious to a lister's inability to actually hear their partner's words. It was only when 75% of the conversation was inaudible that participants noticed a difference—they believed that their conversation partners heard less, understood less, and seemed less responsive and intelligent. But even in this extreme case, partner ratings of listening, effort, attentiveness, interestingness, and likeability did not differ (compared to listeners who could hear more—and those who could hear everything). These results reinforce the findings from Studies 1-3 that perceptions of listening are remarkably inaccurate: there seems to be a striking difference between *feeling* heard and actually *being* heard.

General Discussion

Conversational listening is a pervasive and consequential phenomenon. Information transmission, social connection, conflict management, happiness—the key foundations of human flourishing—hinge critically on our ability to hear, understand, and respond to other people (e.g., Shiller, 1996; Yeomans et al., 2020).

A large body of work finds that speakers and listeners alike experience myriad benefits when people express "good listening," or are perceived as "good listeners" (e.g., Bodenmann, 2005; Huang et al., 2017; Lloyd, Boer, Keller & Voelpel, 2015; Kuhn, Bradbury, Nussbeck, & Bodenmann, 2018; Qian, Wang, Son, Wu, & Fang; Shafran-Tikva, & Kluger, 2016; Wanzer, Booth-Butterfield & Gruber, 2004; Yeomans et al., 2020). At the same time, a rich literature on the failings of mind perception (Epley, 2008) and emerging work on people's insensitivity to

conversational coherence (e.g., Galantucci & Roberts, 2014) call into question whether perceptions of listening accurately reflect the actual cognitive experience of listening (i.e., *being heard*) or merely reflect a subjective experience simulated through specific cues during social interaction (i.e., *feeling heard*). Although the subjective experience of feeling heard matters immensely, it may not represent the construct it is understood to represent: actual listening.

Across five studies, we find that people hold self-serving beliefs about their listening abilities, expressions, and perceptions: people believe they spend more time in conversation listening to others than others spend listening, are better able to convey attentive listening when not actually doing so, and are better able to detect variations in other's listening than others are able to detect. When put to the test in live conversation, though, people did not accurately perceive their partner's listening attentiveness. Though people were able to nimbly adjust their listening in line with instructions—by either listening attentively, inattentively, or dividing their attention—their conversation partners were scarcely able to detect differences in instructed listening and ability to hear (auditory input) across our experimental conditions. This failing extends to third-party observers who were not immersed in the conversation as well as listeners who looked back on their own conversational listening.

Theoretical Contributions

Our work makes several contributions that fundamentally advance scientific understanding of perspective-taking (theory of mind), listening, and the psychology of conversation more broadly. First, our findings add to a body of prior work that show how effectively people can simulate good listening behaviors (e.g., eye contact, nodding, leaning forward). This ability seems robust, even when attention is divided or when the necessary auditory input is severely limited. When motivated to do so, it appears that people are skilled at

conveying the impression of good listening, even in the absence of hearing, learning, or processing.

Second, our findings contribute to a large psychological literature about misperceptions of how others direct their attention. For example, consider the seminal spotlight effect (Gilovich, Medvec, & Savitsky, 2000). Just as people misperceive others' attention when they are wearing a funny t-shirt (they overestimate how much people will notice), people seemingly misperceive others' attention during a conversation, when a distractor (e.g., television or music) is present nearby. Our findings support recent work that suggests we are surprisingly poor at mind perception and perspective taking. We are unable to understand what those around us are hearing and thinking, especially during conversation, a fraught environment where inferential errors are rampant (e.g., Boothby et al., 2019; Mastroianni et al., 2020).

Further, these findings contribute to an emerging literature that is beginning to shed light on people's inattentiveness to others during conversation. Recent work has shown that people are surprisingly insensitive to conversational coherence—in both their attention to turns of phrase and to who their conversational partners are (e.g., Galantucci & Roberts, 2014; Galantucci, Roberts & Langstein, 2018; Roberts, Langstein, & Galantucci, 2016). We extend this work to live conversations, and focus more specifically on an important and misunderstood conversational phenomenon – listening. Together with previous work, the findings of the current research add to a growing literature about the pervasiveness of egocentrism in conversation (Yeomans & Brooks, working).

Our findings are particularly striking when juxtaposed with individuals' seeming overconfidence regarding their ability to detect listening and the important social and professional consequences that follow from such evaluations. One important question is *why* do

people have such faith in their judgments of listening? Is the conviction in the accuracy of one's evaluations driven by the artful performance of the listener? Or does a set of intra-psychic processes experienced by the speaker lead to a feeling of being heard? Answering these questions has profound implications for our understanding of many interpersonal judgments.

In either case, individuals' confidence in their ability to accurately evaluate listening is likely difficult to shake. There are few situations when people receive feedback on whether their perceptions of their partners' listening were accurate or not. In particular, when someone falsely believes that their partner is listening attentively, neither party would have the motivation to correct the error. Indeed, overconfident social judgments are particularly difficult to debias (Logg & Minson, working). Yet, as with beliefs in the predictive powers of the zodiac (Lu, Liu, Liao, & Wang, 2020) or bogus personality tests, we should be cautious about basing important judgments on evaluations of someone's listening skills. It turns out that most people are surprisingly poor judges of others' listening.

Limitations and Future Directions

Our methods and findings are qualified by several limitations that offer fruitful avenues for future research. For example, we observed conversations between strangers in a behavioral lab setting. Future work should explore the listening behaviors between people who know each other, and move to more naturalistic settings beyond the lab.

Second, future work should examine how misperceptions of actual listening influence important downstream consequences, such as learning, education, productivity, decision-making, or relationship maintenance. Although such consequences seem highly likely, given prior results, the *extent* to which errors in listening detection lead to errors in downstream behaviors should be documented.

Third, to minimize the number of experimental conditions in our studies, we did not include a "natural" condition wherein we did not provide extra instructions to the listener. It would be helpful to compare attentive, inattentive, and divided attention listening behavior to the behavior of people who are in their most natural state of listening. We are left wondering: which manipulated condition most resembles natural listening?

Lastly, our findings suggest an important distinction between different types of conversational cues or content. In Study 3, when people watched videos of themselves listening, they had not engaged in an active conversation (recall they had listened to a story and/or music) and the videos they watched were muted, so there was no way they could rely on verbal or prosodic cues (their own or the experimenter's). These results suggest that nonverbal cues may provide an impoverished source of information about actual listening—perhaps because they can be "faked" (i.e., misrepresented). Though participants in Studies 1, 2, and 4 had access to their partner's verbal and prosodic expressions of listening (as well nonverbal cues), it is possible that people aren't particularly good (on average) at using verbal content to *convey* listening. Indeed, emerging work suggests that this is the case—and points to benefits for those who can express listening through their verbal and prosodic expressions.

For example, emerging research on the psychology of conversation points to features of speech that accurately and explicitly convey attentive listening through verbal content, such as asking follow-up questions (e.g., Huang et al., 2017; Yeomans et al., 2019), paraphrasing (e.g., Yeomans & Brooks, working), expressing affirmation and agreement (e.g., Yeomans et al., 2020), and call-backs to previous topics (Sezer et al., working) -- verbal cues that respond directly to what has been previously said (or done) by one's partner. On the other hand, there are verbal cues that can serve as solid proof that someone has failed to listen as well, such as

boomerasking (Hauser et al., working) or redundancies (i.e., repeating what someone has already said as if it is novel and original). Taken together, we suspect that the most accurate, meaningful, and noticeable cues of listening may not happen while people are silently nodding and smiling, but while they are speaking.

Further, we hope future work can capture the prevalence of type I and type II errors in this phenomenon. Are we more likely to believe someone is listening when they're not or to believe someone isn't listening when they are? Our data suggest that inattentive listeners often get more credit than they deserve. For example, are people who appear to not be listening, and then show that they really were listening all along better off in some ways than people who appear to be listening all along? Or are nonverbal indicators of listening (that seemingly do not correlate strongly with actual listening) more important than actual listening for outcomes like impression management?

Conclusion

Recent work suggests that people are surprisingly blind to major disruptions in logical coherence during interpersonal encounters (Galantucci & Roberts, 2014; Galantucci, Roberts & Langstein, 2018; Roberts, Langstein, & Galantucci, 2016). On one hand, this is unsurprising: conversation is a complex, overwhelming decision environment that requires relentless thinking, perceiving, monitoring, and deciding—the human mind is bound to make lots of mistakes. On the other hand, the extent of listening blindness that we document here is quite surprising: a great body of scholarly and lay attention has been paid to developing and identifying "good listening" skills, and people believe they are excellent listening detectors. Taken together, our findings suggest that most people frequently misjudge when their partners are listening or not and call for a re-examination of an important and pervasive social behavior.

References

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

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, *111*(2), 256.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of personality and social psychology*, *54*(4), 569.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

Bodenmann, G. (2005). Dyadic coping and its significance for marital functioning. In T. A. Revenson, K. Kayser, & G. Bodenmann (Eds.), Couples coping with stress: Emerging perspectives on dyadic coping (pp. 33–49). Washington, DC, US: American Psychological Association.

Bodie, G., St. Cyr, K., Pence, M., Rold, M., & Honeycutt, J. (2012). Listening Competence in Initial Interactions I: Distinguishing Between What Listening Is and What Listeners Do. International Journal of Listening, 26(1), 1-28.

Boothby, E.J., Cooney, G., Sandstrom, G.M., & Clark, M.S. (2018). The liking gap in conversations: Do people like us more than we think? *Psychological Science, 29(11)*.

Caprariello, P. A., & Reis, H. T. (2011). Perceived partner responsiveness minimizes defensive reactions to failure. *Social Psychological and Personality Science, 2*, 365-372.

Cheon, J., & Grant, M. (2009). Web-Based Assessment Tool for Communication and Active Listening Skill Development. Tech trends, 6(53), 24-34.

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

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

Epley, N. (2008). Solving the (Real) Other Minds Problem. Social and Personality Psychology Compass, 2(3), 1455-1474.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of personality and social psychology*, *87*(3), 327.

Epley, N., & Waytz, A. (2010). Mind perception. *Handbook of social psychology*.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of personality and social psychology*, *82*(6), 878.

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. Journal of Personality and Social Psychology, 62, 378 – 391.

Friend, W., & Malhotra, D. (2019). Psychological barriers to resolving intergroup conflict: An extensive review and consolidation of the literature. *Negotiation Journal*, *35*(4), 407-442.

Galantucci, B., & Roberts, G. (2014). Do We Notice when Communication Goes Awry? An Investigation of People's Sensitivity to Coherence in Spontaneous Conversation.

Galantucci, B., Roberts, G., & Langstein, B. (2018). Content deafness: When coherent talk just doesn't matter. *Language and Communication*, *61*, 29–34.

Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: an egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of personality and social psychology*, 78(2), 211.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press.

Hafen, M. J. and Crane, D. R., 2003. When marital interaction and intervention researchers arrive at different points of view: the active listening controversy. In Journal of Family Therapy, Vol. 25, No. 1, pp 4-14.

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

Hart, E., VanEpps, E., & Schweitzer, M. E. (2019). I Didn't Want to Offend You: The Cost of Avoiding Sensitive Questions. *Available at SSRN 3437468*.

Hamlin, J. K. (2013). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition*, *128*(3), 451-474.

Hoppe, M. (2006). Active listening: improve your ability to listen and lead (Ideas into action guidebooks). Greensboro, N.C.: Center for Creative Leadership.

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.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. Journal of Experimental Psychology: General, 116 , 26 – 37.

Itzchakov, G., Castro, D. R., & Kluger, A. N. (2016). If you want people to listen to you, tell a story. *International Journal of Listening, 30*(3), 120-133. doi:10.1080/10904018.2015.103744

Janusik, L. A. (2007). Building listening theory: The validation of the conversational listening span. Communication Studies, 58(2), 139-156.

Kaplan, A. (1994). The life of dialogue. *The reach of dialogue: Confirmation, voice, and community*, 34-46.

Keysar, B., & Barr, D. J. (2002). Self-anchoring in conversation: Why language users do not do what they "should." In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 150 – 166). New York: Cambridge University Press.

Kubota, S. et al., 2004. A Study of the Effects of Active Listening on Listening Attitudes of Middle Managers. In Journal of Occupational Health, Vol. 46, pp 60-67.

Kuhn, R., Bradbury, T. N., Nussbeck, F. W., & Bodenmann, G. (2018). The power of listening: Lending an ear to the partner during dyadic coping conversations. *Journal of Family Psychology*, *32*(6), 762.

Laurenceau, J. P., Barrett, L. F., & Pietromonaco, P. R. (1998). Intimacy as an interpersonal process: The importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. *Journal of personality and social psychology*, *74*(5), 1238.

Lenth, R (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.4. https://CRAN.R-project.org/package=emmeans

Lloyd, K. J., Boer, D., Keller, J. W., & Voelpel, S. (2015). Is my boss really listening to me? The impact of perceived supervisor listening on emotional exhaustion, turnover intention, and organizational citizenship behavior. *Journal of Business Ethics*, *130*(3), 509-524.

Lu, J. G., Liu, X. L., Liao, H., & Wang, L. (2020). Disentangling stereotypes from social reality: Astrological stereotypes and discrimination in China. *Journal of Personality and Social Psychology*.

Mastroianni, A.M., Cooney, G., Boothby, E.J., Reece, A.G. (2020). The liking gap in groups and teams. *Organizational Behavior and Human Decision Processes.*

McNaughton, D. et al., 2008. Learning to Listen: Teaching an Active Listening Strategy to Preservice Education Professionals. In Topics in Early Childhood Special Education, Vol. 27, No. 223.

Mishima, N. et al., 2000. The Development of a Questionnaire to Assess the Attitude of Active Listening. In Journal of Occupational Health, Vol. 42, pp 111-118.

Nichols, R. G., (1947, February), Listening: Questions and problems. Quarterly Journal of Speech, *33*(1), 83-86

Nichols, R. G., & Lewis, T, R, (1954), Listening and speaking: A guide to effective oral communication. Dubuque, IA: William C, Brown,

Nichols, R. G., & Stevens, L, A, (1957), Are you listening? New York: McGraw-Hill

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169-190.

Qian, J., Wang, B., Song, B., Li, X., Wu, L., & Fang, Y. (2019). It takes two to tango: The impact of leaders' listening behavior on employees' feedback seeking. *Current Psychology*, *38*(3), 803-810.

Rautalinko, E. and Lisper, H.-O., 2004. Effects of Training Reflective Listening in a Corporate Setting. In Journal of Business and Psychology, Vol. 18, No. 3, pp 281-299.

Reis, H. T. (in press). Perceived partner responsiveness as an organizing theme for the study of relationships and well-being. In L. Campbell & T. J. Loving (Eds.), *Close relationships: An interdisciplinary integration*. Washington, DC.: APA Books.

Reis, H. T., Clark, M. S., & Holmes, J. G. (2004). Perceived partner responsiveness as an organizing construct in the study of intimacy and closeness. In D. Mashek & A. Aron (Eds.), *The handbook of closeness and intimacy* (pp. 201-225). Mahwah, NJ: Lawrence Erlbaum Associates.

Reis, H. T., Crasta, D., Rogge, R. D., Maniaci, M. R., & Carmichael, C. L. (2018). Perceived Partner Responsiveness Scale (PPRS). In D. L. Worthington & G. D. Bodie (Eds.), The Sourcebook of Listening Research: Methodology and Measures (pp. 516-521). Hoboken, NJ: Wiley Blackwell.

Roberts, G., Langstein, B., & Galantucci, B. (2016). (In)sensitivity to incoherence in human communication. *Language and Communication*, *47*, 15–22.

Rogers, C. R., & Farson, R. E. (1957). *Active listening*. Industrial Relations Center of the University of Chicago.

Ross, L., & Ward, A. (1995). Psychological barriers to dispute resolution. In *Advances in experimental social psychology* (Vol. 27, pp. 255-304). Academic Press.

De Ruyter, K., & Wetzels, M. G. (2000). The impact of perceived listening behavior in voice-to-voice service encounters. *Journal of Service Research*, *2*(3), 276-284.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, 47, 1 – 24.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis I* (Vol. 1). Cambridge university press.

Shafran-Tikva, S., & Kluger, A. N. (2018). Physician's listening and adherence to medical recommendations among persons with diabetes. *International Journal of Listening*, *32*(3), 140-149.

Smith, P. K., Wigboldus, D. H., & Dijksterhuis, A. P. (2008). Abstract thinking increases one's sense of power. *Journal of Experimental Social Psychology*, *44*(2), 378-385.

Surra, C. A., & Longstreth, M. (1990). Similarity of outcomes, interdependence, and conflict in dating relationships. *Journal of Personality and Social Psychology*, 59, 501–516. http://dx.doi.org/10.1037/0022- 3514.59.3.501

Swann, W. B., Jr., & Gill, M. J. (1997). Confidence and accuracy in person perception: Do we know what we think we know about our relationship partners? *Journal of Personality and Social Psychology*, 73, 747–757. http://dx.doi.org/10.1037/0022-3514.73.4.747

Tamir, D. I., & Mitchell, J. P. (2013). Anchoring and adjustment during social inferences. *Journal of Experimental Psychology: General*, *142*(1), 151.

Thomas, L. T., & Levine, T. R. (1994). Disentangling listening and verbal recall: Related but separate constructs? Human Communication Research, 21, 103–127.

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

Wanzer, M. B., Booth-Butterfield, M., & Gruber, K. (2004). Perceptions of health care providers' communication: relationships between patient-centered communication and satisfaction. *Health communication*, *16*(3), 363-384.

Weger Jr, H., Castle, G. R., & Emmett, M. C. (2010). Active listening in peer interviews: The influence of message paraphrasing on perceptions of listening skill. *The International Journal of Listening*, *24*(1), 34-49.

Weger, H., Castle Bell, G., Minei, E., & Robinson, M. (2014). The Relative Effectiveness of Active Listening in Initial Interactions. International Journal of Listening, 28(1), 13-31.

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

Witkin, B. R. (1990). Listening theory and research: The state of the art. Journal of the International Listening Association, 4, 7–32.

Yeomans, M. & Brooks, A.W. Topic Selection in Conversation. *Manuscript in prep.*

Yeomans, M., Brooks, A.W., Huang, K., Minson, J., & Gino, F. (2019). It helps to ask: The cumulative benefits of asking follow-up questions. *Journal of Personality and Social Psychology, 117 (6),* 1139-1144.

Yeomans, M., Minson, J., Collins, H., Chen, F., & Gino, F. (2020). Conversational receptiveness: Improving engagement with opposing views. *Organizational Behavior and Human Decision Processes, 160,* 131-148.