Supplemental Material

Exploratory Study

Supplemental Method

Post-Conversation Measures. After their 25-minute conversation, participants completed a brief survey in which they reflected on their conversation. Specifically, they responded to measures of listening, liking, and conversational enjoyment.

Listening. Participants provided post-hoc evaluations of their own (in the 'target' condition; $\alpha = 0.89$) or their partner's (in the 'perceiver' condition; $\alpha = 0.91$) listening. They responded to five items: (1) [I/My partner] was a good listener, (2) [I/My partner] was engaged in the conversation, (3) [I/My partner] made [my partner/me] feel heard, (4) [I/My partner] made [my partner/me] feel validated, (5) [My partner/I] felt that [I/my partner] cared about [them/me] (1: *Strongly Disagree*, 4: *Neither agree nor disagree*, 7: *Strongly Agree*).

Liking. Participants reported how much they liked their partner using a 4-item scale (Boothby et al., 2018; 1: *Strongly Disagree*, 4: *Neither agree nor disagree*, 7: *Strongly Agree*). The items included: (1) I generally liked my partner, (2) I would be interested in getting to know my partner better, (3) If given the chance, I would like to interact with the other participant again, (4) I could see myself becoming friends with the other participant ($\alpha = 0.93$).

Enjoyment. Finally, participants reported their enjoyment of the conversation with their partner using three items (Huang et al., 2017; 1: *Strongly Disagree*, 4: *Neither agree nor disagree*, 7: *Strongly Agree*). The items included: (1) I enjoyed this conversation, (2) I thought this conversation was engaging, (3) I had an interesting conversation with this person ($\alpha = 0.95$).

Supplemental Results

In our second pre-registered analysis, we conducted a binary logistic regression to predict target responses (actual listening) from a fixed effect for perceiver guesses (values recoded such that 0=*[My/my partner's] mind was inattentive*, 1= *Yes, [I/my partner] was fully attentive*) and time-point (which of the 5 ratings), and a random effect for dyad to account for repeated observations (Bates et al., 2015). We found a positive association between perceiver ratings and target self-reports, $b = 0.49$, *95% CI* [0.03, 0.96], *SE* = 0.49, *p* = .04, and a negative effect of time, $b = -0.16$, *95% CI* [-0.28, -0.05], *SE* = 0.06, *p* = .006 (see Figure 2). Translating these into an odds ratio, these results suggest that perceivers were 1.64 times more likely to report that the target was listening to them than the target was to report they were indeed listening.

To assess the correlation between perceiver and target reports of listening, we conducted a linear regression predicting target self-reports of listening from a fixed-effect for perceiver ratings, and a random effect for dyad to account for repeated observations (Bates et al., 2015).[1] Results showed a small positive correlation between perceiver and target ratings of listening, $b = 0.09$, *95% CI* [0.02, 0.16], *SE* = 0.04, *p* = .02, but perceiver ratings only accounted for 0.6% of the variance In target ratings (marginal $R^2 = 0.0059$; Vonesh et al., 1996).

Post-Conversation Measures. Though ratings of listening diverged *during* the conversation, differences in post-conversation ratings of listening (a composite of our five items) were not as dramatic, perhaps reflecting the targets' motivation to self-present as a good listener ($M_{target} = 5.93$, $SE_{target} = 0.09$; $M_{perceiver} = 6.03$, $SE_{perceiver} = 0.09$; $b = 0.10$, *95% CI* [-0.14, 0.35], *SE* = 0.12, *p* = 0.40). This remained true when we looked at single-items measures that may more directly measure cognitive attention: "good listener" ($M_{target} = 6.05$, $SE_{target} = 0.10$; $M_{perceiver}$ = 6.27, $SE_{perceiver} = 0.10$; $b = 0.22$, *95% CI* [-0.04, 0.48], *SE* = 0.13, *p* = 0.10) and "engaged in the

---

[1] Recent work has found that linear regression yields unbiased estimates of effects on binary outcomes (Gomila, 2021).

conversation" ($M_{target} = 6.29$, $SE_{target} = 0.10$; $M_{perceiver} = 6.27$, $SE_{perceiver} = 0.10$; b = -0.02, *95% CI* [-0.27, 0.23], *SE* = 0.13, *p* = 0.88). We found no significant interaction with gender, b = -0.24, *95% CI* [-0.78, 0.30], *SE* = 0.28, *p* = 0.38 (women vs. men).

We also found a strong positive relationship between post-conversation ratings of listening (self-reports for *targets*; perceptions for *perceivers*) and liking (b = 0.73, *95% CI* [0.56, 0.91], *SE* = 0.09, *p* < .001)—though this was qualified by a significant interaction with role such that the relationship was stronger between perceived listening and liking of one's partner than between self-reported listening and liking of one's partner (b = 0.42, *95% CI* [0.07, 0.76], *SE* = 0.18, *p* = .02). Similarly, we find a strong positive relationship between post-conversation ratings of listening and enjoyment for targets and perceivers alike (b = 0.98, *95% CI* [0.81, 1.16], *SE* = 0.09, *p* < .001; and a non-significant interaction with role: b = 0.25, *95% CI* [-0.08, 0.58], *SE* = 0.17, *p* = .13).

Gender Differences for Main Effects

On an exploratory basis, we investigated gender differences in our main effects (comparing men and women because the number of participants identifying as "Non-binary/Other" was too small to power our statistical analysis, n = 6). Gender effects could influence our results either at the level of the listener or the perceiver—however, we find no evidence that accuracy of listening perceptions differed for men and women at either level. Indeed, perceiver guesses matched targets' self-reports of listening 70% of the time when listeners were women, and 69% of the time when listeners were men. Additionally, when perceivers were women, their guesses matched targets' self-reports 70% of the time while when the perceiver was a man, their guesses matched 69% of the time.

The misalignment between perceptions and target self-reports of listening were directionally biased in the same direction—there were significantly more Type I errors (perceiver thought target was listening when target reported mind-wandering) than Type II errors (perceiver reported target was not listening when target reported listening attentively) when the listener was a woman (77% Type I error; 14% Type II error) vs. a man (80% Type I error; 18% Type II error), and when the perceiver was a woman (76% Type I error; 17% Type II error) vs. a man (80% Type I error; 13% Type II error).

Experiment 1

Gender Differences for Main Effects

We explored whether there were gender differences in our main effect (again, comparing men and women because we did not have a large enough sample size to compare to individuals who identified as "Non-binary/Other", n = 3). First, we investigated whether perceptions of listening across the three experimental conditions varied as a factor of whether the *perceiver* was a man vs. a woman. We found no significant interaction between condition and the gender of the unmanipulated participant for perceptions of listening (Listening vs. Distracted: $b_{interaction}$ = -0.42, *95% CI* [-1.30, 0.46], *SE* = 0.44, *p* = .34; Listening vs. Feigned Listening: $b_{interaction}$ = -0.65, *95% CI* [-1.45, 0.16], *SE* = 0.41, *p* = .12; Distracted vs. Feigned Listening: $b_{interaction}$ = -0.22, *95% CI* [-1.10, 0.65], *SE* = 0.44, *p* = .61).

Further, we investigated whether perceptions of listening across the three experimental conditions were affected by the gender of the *listener*. Again, we found no significant interaction between condition and the gender of the manipulated participant for perceptions of listening (Listening vs. Distracted: $b_{interaction}$ = 0.46, *95% CI* [-0.38, 1.30], *SE* = 0.43, *p* = .28; Listening vs.

Feigned Listening: $b_{interaction}$ = -0.04, *95% CI* [-0.82, 0.75], *SE* = 0.40, *p* = .93; Distracted vs.

Feigned Listening: $b_{interaction}$ = -0.50, *95% CI* [-1.32, 0.33], *SE* = 0.42, *p* = .24).

Experiment 2

Gender Differences for Main Effects

We explored whether there were differences in the behaviors that manipulated listeners in Experiment 1 enacted during their conversations based on their gender across the three conditions. Interaction analyses revealed differences between men and women across conditions for verbal interruptions (increases in verbal interruptions in the feigned listening vs. listening condition was greater for men than women), looking away (increases in looking away in the feigned listening vs. listening condition was greater for men than women), leaning forward (increases in leaning forward in the distracted vs. listening condition was greater for men than women), and leaning backward (increases in leaning backward in the feigned listening vs listening condition was larger for men than women). There were no differences on any of the other behaviors. All results are included in the table below.

Finally, the results showed no differences in how accurately third-party observers were able to identify listeners' assigned condition for men vs. women: 16% were correct for men and 20% for women, b_{gender} = -0.18, SE = 0.45, *95%CI* [-1.10, 0.71], *p* = .68.

	Attentive Listening *M (SE)*		Feigned Listening *M (SE)*		Distracted *M (SE)*	
Behavior	Men	Women	Men	Women	Men	Women
Verbal interruptions	**1.56 (0.10)**	*1.86 (0.10)*	**2.05 (0.09)**	*1.89 (0.10)*	1.74 (0.13)	1.71 (0.08)
Verbal affirmations	2.98 (0.10)	3.05 (0.08)	2.87 (0.09)	2.84 (0.10)	2.66 (0.13)	2.97 (0.08)
Nodding	3.14 (0.11)	3.04 (0.09)	2.88 (0.10)	2.81 (0.11)	2.72 (0.14)	2.95 (0.09)
Eye Contact	3.46 (0.09)	3.43 (0.08)	3.46 (0.08)	3.49 (0.09)	3.50 (0.12)	3.46 (0.07)
Looking Away	**1.84 (0.11)**	*2.15 (0.10)*	**2.22 (0.10)**	*1.93 (0.12)*	2.19 (0.15)	2.09 (0.10)

Smiling	3.04 (0.12)	3.14 (0.10)	2.92 (0.11)	3.14 (0.12)	3.15 (0.16)	3.25 (0.10)
Fidgeting	2.56 (0.14)	2.61 (0.12)	2.79 (0.12)	2.63 (0.14)	2.50 (0.18)	2.46 (0.11)
Leaning forward	**1.68 (0.14)**	*2.06 (0.12)*	1.97 (0.12)	2.10 (0.14)	**2.03 (0.19)**	*1.76 (0.11)*
Leaning backwards	**1.41 (0.12)**	*1.75 (0.10)*	**1.84 (0.10)**	*1.67 (0.12)*	1.59 (0.16)	1.53 (0.10)

Note: Significant interaction effects are identified with a combination of bold and italics. Differences between bolded means in the same row are significantly different from the differences between bolded and italicized numbers.

Experiment 3

Gender Differences for Main Effects

We explored whether there were differences between men and women (no other gender-identities were present in our demographic data) in their ability to identify their own listening. Results showed that men and women did not differ in their accuracy. Men correctly guessed their listening on 63% of trials (30% when listening attentively to the story; 33% when listening inattentively to the story), while women correctly guessed their listening on 63% of trials (31% when listening attentively to the story; 32% when listening inattentively to the story).

Experiment 4

Gender Differences for Main Effects

We investigated whether gender influenced our main effects (focusing on comparisons between men and women due to sample constraints, $n = 2$ participants identified as "Non-binary/Other"). Across conditions, women were perceived to be better listeners (controlling for condition and round), $b = 0.29$, SE $= 0.13$, *95%CI* [0.03, 0.55], $p = 0.03$, and were perceived to be significantly more responsive, $b = 0.37$, SE $= 0.14$, *95%CI* [0.10, 0.65], $p = .009$.

Additionally, while we find no significant interaction between listener gender and condition for ratings of perceived listening (all p's$>.15$), we do find a significant interaction for ratings of perceived responsiveness. Specifically, we find that the reduction in perceived

responsiveness for participants in the 75% condition compared to the 0% ($b_{interaction}$ = -0.85, SE = 0.40, 95%CI[-1.64, -0.05], p = 0.04) and 25% conditions ($b_{interaction}$ = -0.95, SE = 0.45, 95%CI[-1.84, -0.06], p = 0.04) was greater for women than for men.

We also investigated the effect of perceiver gender on our main results. Results showed no main effect (*p*'s>.87), or significant interaction with condition (*p*'s>.06), for ratings of perceived listening or responsiveness.