



Supporting Online Material for

Are Women Really More Talkative Than Men?

Matthias R. Mehl,* Simine Vazire, Nairán Ramírez-Esparza, Richard B. Slatcher, James W. Pennebaker

*To whom correspondence should be addressed. E-mail: mehl@email.arizona.edu

Published 6 July 2007, *Science* **317**, 82 (2007)

DOI: 10.1126/science.1139940

This PDF file includes:

Materials and Methods

Fig. S1

Table S1

References

Supporting Online Material for

Are Women Really More Talkative Than Men?

Matthias R. Mehl*, Simine Vazire, Nairán Ramírez-Esparza, Richard B. Slatcher, James W. Pennebaker

* To whom correspondence should be addressed. E-mail: mehl@email.arizona.edu

This PDF file includes:

Materials and Methods

Table S1

Figure S1

References

Materials and Methods

Participants and Samples

Participants in samples 1, 2, 4, 5 and 6 were students at the University of Texas at Austin (*S1, S2, S4–S6*). Participants in sample 3 were students at the Autonomous University of Nuevo Leon in Monterrey, Mexico (*S3*). References to the original publications of the data sets along with further information about the samples are included in Table S1.

Sample 1. Participants in sample 1 ($N = 112$) were couples who had been dating for at least six months and were recruited via fliers and online advertisements. Each couple received \$150 in exchange for participation. The age of the participants ranged from 18 to 29 ($M = 20.2$, $SD = 1.8$) and the sample was 50% female. Participants in this sample wore the EAR for approximately 7 days altogether. Specifically, the study had two 3.5-day monitoring sessions (from Thursday afternoon to Sunday night) that were separated by 2 weeks (*S1*).

Sample 2. Participants in sample 2 ($N = 79$) were recruited from introductory psychology classes and received \$50 in exchange for their participation. The age of the participants ranged from 17 to 23 ($M = 18.7$, $SD = 1.4$) and the sample was 53% female. Participants in this sample wore the EAR for approximately four days, from Friday evening to Tuesday evening (*S2*).

Sample 3. Participants in sample 3 ($N = 51$) were recruited from Psychology classes and were paid the equivalent of \$30 (300 Mexican pesos) for their participation. The age of the participants ranged from 17 to 25 ($M = 17.9$, $SD = 1.6$) and the sample was 61% female. Participants in this sample wore the EAR for approximately four days. About half of the participants wore the EAR from Friday at noon to Tuesday at noon, and the other half wore the EAR from Thursday morning to Monday morning (*S3*).

Sample 4. Participants in sample 4 ($N = 96$) were recruited from introductory psychology classes and received partial course credit in exchange for their participation. The age of the participants ranged from 17 to 22 ($M = 18.7$, $SD = 0.9$) and the sample was 49% female (*S4*). Participants in this sample wore the EAR for approximately two days from Monday morning to Wednesday morning ($n = 22$) or from Wednesday afternoon to Friday afternoon ($n = 74$).

Sample 5. Participants in sample 5 ($N = 11$) were recruited from psychology classes and received \$100 in exchange for their participation. The age of the participants ranged from 18 to 26 ($M = 20.9$, $SD = 2.4$) and the sample was 64% female. Participants in this sample wore the EAR for approximately ten days from September 11, 2001 to September 20, 2001 (*S5*).

Sample 6. Participants in sample 6 ($N = 47$) were recruited from introductory psychology classes and received partial course credit in exchange for their participation. The age of the participants ranged from 17 to 23 ($M = 19.0$, $SD = 1.3$) and the sample was 57% female. Participants in this sample wore the EAR for altogether approximately 4 days. Specifically, the study had two two-day monitoring sessions (Monday morning to Wednesday morning and Wednesday afternoon to Friday afternoon) that were separated by 4 weeks (*S6*).

EAR System

The EAR system in samples 1 to 5 consisted of a digital voice recorder (SONY Memory Stick ICD-MS1), an external microphone (OPTIMUS Tie Clip Microphone), and a controller microchip. The EAR system in sample 6 consisted of an analog micro-cassette recorder (OPTIMUS Micro-32), an external microphone (OPTIMUS Tie Clip Microphone), and a controller microchip (*S7*). With both systems, the chip was programmed with a 30-sec on and 12.5-min off cycle. The digital EAR operated with high precision and reliably produced 30-sec

recordings and 4.8 recording cycles per hour. The analog EAR system operated with less precision and produced a net recording time of on average 22 seconds and 4.5 cycles per hour due to a fading in and fading out period at the beginning and the end of each recording period. Thus, the digital EAR recorded 4% of the time and the analog EAR system recorded roughly 2.75% of the time. Participants carried the EAR around in a shock-protected case in one of their pockets or in a purse-like bag. An external microphone was clipped to the collar of their shirts. Due to the covert recording (the EAR devices were wrapped in padding and put in sealed protective cases) it was impossible for participants to sense when the EAR was on or off.

Procedures

Participants were instructed to wear the EAR during their waking hours. Participants were encouraged to wear the device as much as possible and to remove it only when its functioning was in jeopardy (e.g., when taking a shower or playing sports). Participants were informed that the EAR would be sampling less than 5% of the time and were lead to believe that the recordings were randomly spaced; they were only informed about the periodic nature of the EAR recordings at the end of the study as part of the debriefing. Participants also were informed that they would have the opportunity to review their recordings and that they could erase any part they wished before the researchers had access to them. This procedure is part of the set of safeguards we implement in all EAR studies to protect participants' privacy and to ensure the confidentiality of the data. These safeguards are critical for the ethical conduct of EAR research and have proven highly effective at alleviating potential concerns that participants may have about the method.

To increase the reliability of the analyses, participants were also asked to keep an end-of-day diary where they documented major activities of the day and times where they were unable to wear the EAR. The recorder was then activated and handed to participants in a sealed, protective case. At the end of the scheduled monitoring period, participants returned to the lab, dropped off the EAR device, were debriefed, and were provided with an opportunity to review their recordings and erase parts they wished.

EAR Compliance and Obtrusiveness

Detailed information on the obtrusiveness of the EAR and participants' compliance with wearing the device is provided in the original publications of the data sets (S1–S6). A recent systematic empirical analysis based on two samples (S4–S5) found that compliance was high and obtrusiveness was not a major concern for participants. With respect to compliance, the rate of EAR-coded “not wearing the EAR” was 2% in S5 (the 10-day sample) and 9% in S4 (the 2-day sample) and essentially converged with participants' reported compliance. These compliance rates are high and comparable to those reported for other ambulatory assessment methods (S8).

With respect to obtrusiveness, participants generally report the EAR as being minimally obtrusive. Specifically, across these two samples, the average rating on the item “To what extent did the microphone influence your way of talking?” was 1.5 on a scale ranging from 1 = *not at all*, to 5 = *a great deal*. In addition, participants do not talk much about the EAR—arguably a critical behavioral marker of how obtrusive the method is. Specifically, the percentage of recordings in which the EAR was mentioned was 2.6% in S5 and 3.6% in S4. This evidence suggests that the EAR operates fairly unobtrusively in everyday life.

Finally, most participants do not request that any of their files be deleted. For example, in samples S2 and S4 participants deleted less than 0.1% of their sound files. Thus, the participants seemed confident in the researchers' privacy and confidentiality policies. Taken together, these

analyses suggest that the data obtained from EAR sampling is largely representative of participants' daily activities.

Transcription of the EAR recordings

English-speaking research assistants transcribed the EAR recordings for samples 1, 2, and 4-6. Spanish-English bilingual research assistants transcribed the EAR recordings for sample 3. All research assistants received training in dealing with the challenges of transcribing oral language, such as handling of repetitions, filler words, nonfluencies, or slang (*S9*). Partial words were transcribed only if they could be unambiguously completed from the captured word fragment. For ethical considerations and to comply with a mandate by the Institutional Review Board of the University of Texas at Austin, the recording of language from people around the participants was minimized by setting the recording sensitivity to low. In cases where other people's language ended up being audible on the recordings, it was not transcribed or used in analyses (see *S6* for an exception where bystanders' language was transcribed and used in analyses). Research assistants used a sample voice recording that participants provided at the beginning of the study to discriminate among the participants' voices and the voices of other people around the participants. The verbatim transcripts were then submitted to Linguistic Inquiry and Word Count (LIWC), a computerized text analysis program, which provided the word count for each participant (*S10*). Intertranscriber agreement was estimated from 392 training EAR recordings that were independently transcribed by 8 research assistants. Intertranscriber agreement, calculated as the intraclass correlation based on a two-way random effect model (ICC [2, k]), was very high, $r = .99$ (*S4*).

Data Analytic Strategy

The number of words that female and male study participants used was estimated in the following way: First, for each participant, the total number of recorded words was divided by the number of valid and waking recordings to obtain an estimate of the average number of words spoken per EAR recording. In a second step, the computed average number of words per recording was then extrapolated to an estimate of the number of words spoken within a day. For this, the number of words per recording was divided by the recording length (30 seconds for samples 1-5, 22 seconds for sample 6) and multiplied by 3,600 seconds per hour and an average of 17 waking hours per day. The National Sleep Foundation's most recent poll found that adults in America report sleeping on average 6.9 hours a night when considering both sleep on weekdays and weekends (*S11*). An average of 17 waking hours per day was therefore considered the best available estimate for the purpose of this study. We used this number rather than sample-based information in the algorithm to minimize the effects of sampling bias and to maximize the generalizability of the resulting estimate. That way, the estimated average number of words spoken per day across all samples was 15,958 ($SD = 7,949$; $Min = 695$; $Max = 47,016$).

In an additional analysis, we explored the possibility that women and men differed in their use of distinct words. For each participant, LIWC computed the percentage of unique words, also known as type-token ratio (i.e. the number of different words divided by the number of total words multiplied by 100). Across the six samples, female and male study participants did not differ in their use of unique words, $M = 32.02$ versus $M = 32.84$, $t(394) = 0.782$, $p = .434$.

Finally, we explored the possibility that in the absence of mean differences in daily word use between women and men, women might be overrepresented in the positive tail of the distribution. These "outliers", then, could serve as perceptually salient exemplars and contribute

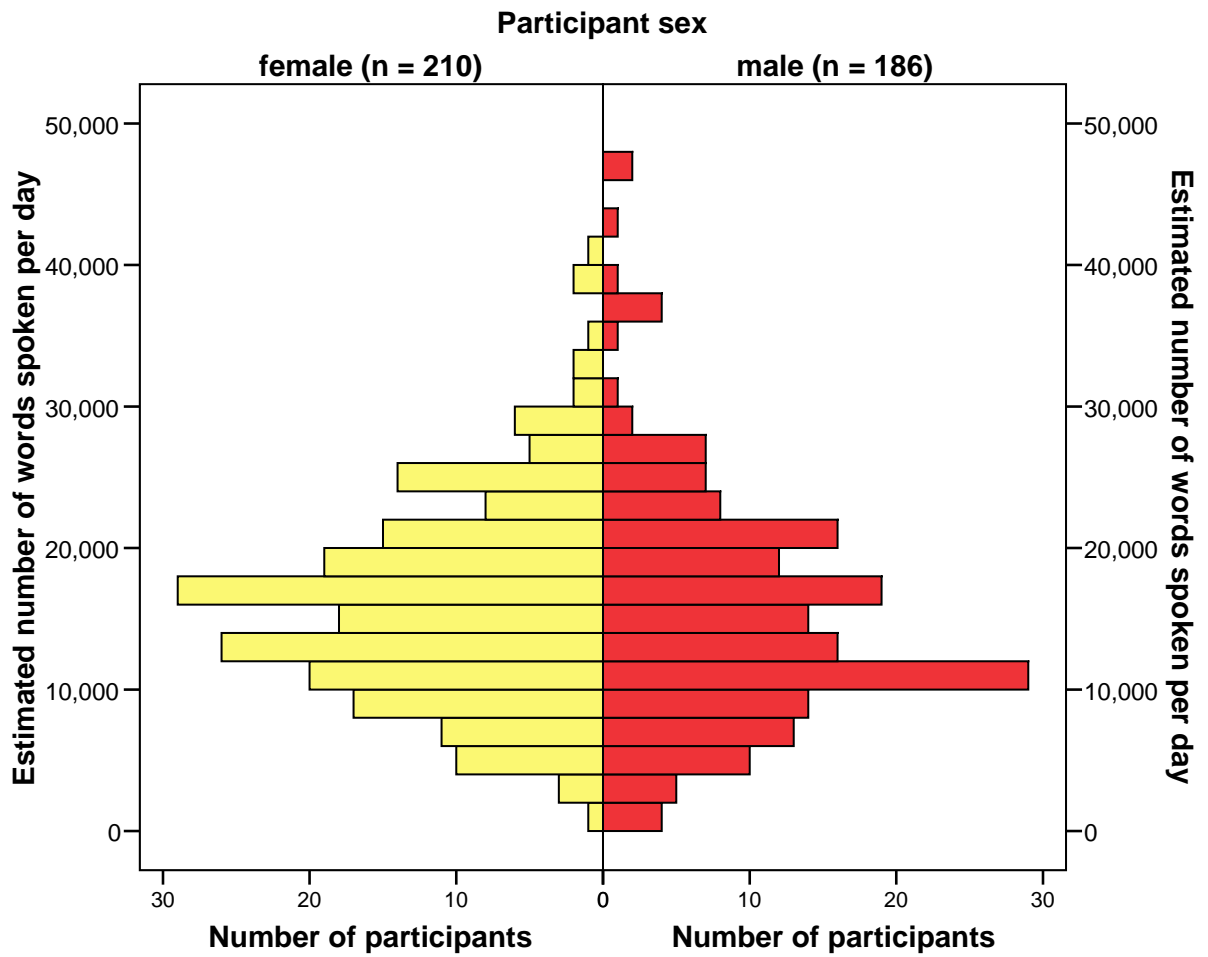
to the creation and maintenance of the stereotype of female talkativeness. To test this idea, we examined whether the participants with the highest daily word use tended to be women. Specifically, we tested this in two ways. First, we compared the skewness of the distributions for female and male participants in each sample. All distributions were positively skewed (range = 0.385 – 1.877), but in none of the six samples did women have a more positive skew than men. This suggests that women were not more likely than men to be among the most talkative participants in the samples. In addition, we also examined whether women were overrepresented in the top 17% of word users in each sample (which corresponds approximately to being more than 1 SD above the mean). Among the top 17% of word users in each sample, 50% were female and 50% male. These numbers are not significantly different from what would be expected by chance (across all six samples 53% of the participants were female and 47% male). Finally, a visual inspection of the two histograms of the distributions of male and female daily word use (see Figure S1) reveals that the three participants with the highest word use across all six samples were in fact men. Taken together, this suggests that it is unlikely that talkative female outliers serve as perceptually salient exemplars on the basis of which the stereotype of female talkativeness may have evolved or can be maintained.

Table S1. Additional information regarding the estimation of male and female study participants' daily word use in the 6 samples.

| Sample | Ref. | Net record. length | Valid waking recordings | | Total words recorded | | Words per valid waking recording | |
|------------------|-----------|--------------------------|-------------------------|------------------------|------------------------|------------------------|----------------------------------|------------------------|
| | | | Women | Men | Women | Men | Women | Men |
| | | | <i>M</i> (<i>SD</i>) | <i>M</i> (<i>SD</i>) | <i>M</i> (<i>SD</i>) | <i>M</i> (<i>SD</i>) | <i>M</i> (<i>SD</i>) | <i>M</i> (<i>SD</i>) |
| 1 | <i>S1</i> | 30 sec | 373 (142) | 410 (150) | 3,490 (2,087) | 3,306 (1,723) | 9.041 (3.657) | 8.126 (3.859) |
| 2 | <i>S2</i> | 30 sec | 296 (102) | 304 (108) | 2,017 (1,110) | 1,983 (1,407) | 7.008 (3.157) | 6.892 (4.444) |
| 3 | <i>S3</i> | 30 sec | 258 (98) | 301 (42) | 1,805 (1,021) | 2,184 (1,262) | 7.208 (3.047) | 7.364 (3.855) |
| 4 | <i>S4</i> | 30 sec | 128 (39) | 131 (29) | 983 (509) | 1,024 (538) | 7.930 (3.686) | 8.122 (4.465) |
| 5 | <i>S5</i> | 30 sec | 660 (106) | 713 (101) | 4,885 (2,733) | 8,098 (2,326) | 7.726 (4.405) | 11.790 (5.005) |
| 6 | <i>S6</i> | 22 sec | 210 (46) | 198 (48) | 1,184 (471) | 902 (637) | 5.930 (2.845) | 4.625 (2.999) |
| Weighted average | | | 274 (151) | 287 (160) | 2,136 (1,736) | 2,166 (1,821) | 7.671 (3.520) | 7.500 (4.229) |

Note: $N = 396$; **Ref.** provides a reference to the original publication of the respective data set; **Net record. length** is the effective recording time per EAR recording; **Valid waking recording** is the average number of valid (i.e. compliant and free of technical problems) waking (i.e. non-sleeping) EAR recordings; **Total words recorded** is the average number of full words that the EAR captured across all recordings; **Words per valid waking recording** is the average number of full words per valid waking recording; the weighted average weighs the respective sample group mean by the sample size of the group. The average number of words spoken per day (Table 1) was estimated by dividing the average number of words per valid waking recording by the net recording length in the sample and multiplying the resulting estimate of average words spoken per second by 3600 seconds per hour and by an average of 17 waking hours per day (*S11*).

Figure S1. Histograms for the distributions of the estimated number of words spoken per day for female and male study participants across the six samples ($N = 396$).



Supporting References

- S1. R. B. Slatcher, J. W. Pennebaker, paper presented at the Summer Research Institute of the Center for Health Promotion and Disease Prevention Research, Austin, TX (21st June, 2006).
- S2. S. Vazire, M. R. Mehl, paper presented at the 7th Annual Meeting of the Society for Personality and Social Psychology, Palm Springs, CA (28th January, 2006).
- S3. N. Ramírez-Esparza, M. R. Mehl, J. W. Pennebaker, paper presented at the 6th Annual Meeting of the Society for Personality and Social Psychology, New Orleans, LA. (21st January, 2005).
- S4. M. R. Mehl, S. D. Gosling, J. W. Pennebaker, *J. Pers. Soc. Psychol.* **90**, 862 (2006).
- S5. M. R. Mehl, J. W. Pennebaker, *Psychol Sci.* **14**, 579 (2003).
- S6. M. R. Mehl, J. W. Pennebaker, *J. Pers. Soc. Psychol.* **84**, 857 (2003).
- S7. M. R. Mehl, J. W. Pennebaker, M. Crow, J. Dabbs, J. Price, *Beh. Res. Meth. Instr. Comp.* **33**, 517 (2001).
- S8. N. Bolger, A. Davis, E. Rafaeli, *Annu. Rev. Psychol.* **54**, 579 (2004).
- S9. H. Sacks, E. A. Schegloff, G. Jefferson, *Language.* **50**, 696–735 (1974).
- S10. J. W. Pennebaker, M. E. Francis, R. J. Booth, *Linguistic Inquiry and Word Count: LIWC2001* (Erlbaum, Mahwah, NJ, www.liwc.net, 2001)
- S11. National Sleep Foundation. 2005 ‘*Sleep in America*’ Poll.
http://www.sleepfoundation.org/site/c.huIXKjM0IxF/b.2419039/k.14E4/2005_Sleep_in_America_Poll.htm (30 March, 2007)