Polite speech emerges from competing social goals – Supplementary Materials

Abstract

Language is a remarkably efficient tool for transmitting information. Yet human speakers make statements that are inefficient, imprecise, or even contrary to their own beliefs, all in the service of being polite. What rational machinery underlies polite language use? Here, we show that polite speech emerges from the competition of three communicative goals: to convey information, to be kind, and to present oneself in a good light. We formalize this goal tradeoff using a probabilistic model of utterance production, which predicts human utterance choices in socially-sensitive situations with high quantitative accuracy, and we show that our full model is superior to its variants with subsets of the three goals. This utility-theoretic approach to speech acts takes a step towards explaining the richness and subtlety of social language use.

Keywords: politeness, computational modeling, communicative goals, pragmatics

Word count: 5500

Polite speech emerges from competing social goals – Supplementary Materials

Supplementary Materials

Model details

The *literal listener* L_0 is a simple Bayesian agent that takes the utterance to be true:

$$P_{L_0}(s|w) \propto \delta_{[\![w]\!](s)} \cdot P(s).$$

where $\delta_{[\![w]\!](s)}$ is the Kronecker delta function which uses to the truth-functional denotation of the utterance $[\![w]\!](s)$ to return a value of 1 if the utterance w is true of the state s and 0 otherwise. The literal meaning is used to update the literal listener's prior beliefs over world states $P(s)$.

The *speaker* S_1 chooses utterances approximately optimally given a utility function, which can be decomposed into two components. First, informational utility (U_{inf}) is the amount of information a literal listener L_0 would still not know about world state s after hearing a speaker's utterance w. Second, social utility (U_{soc}) is the expected subjective utility of the state inferred given the utterance w. The utility of an utterance subtracts the cost $c(w)$ from the weighted combination of the social and epistemic utilities.

$$U(w; s; \phi) = \phi \cdot \ln(P_{L_0}(s \mid w)) + (1 - \phi) \cdot \mathbb{E}_{P_{L_0}(s|w)}[V(s)] - C(w).$$

The speaker then chooses utterances w softmax-optimally given the state s and his goal weight mixture ϕ:

$$P_{S_1}(w \mid s, \phi) \propto \exp(\alpha \cdot \mathbb{E}[U(w; s; \phi)]).$$

Literal semantic task

We probed judgments of literal meanings of the target words assumed by our model and used in our main experiment.

Participants. 51 participants with IP addresses in the United States were recruited on Amazon's Mechanical Turk.

Design and Methods. We used thirteen different context items in which a speaker evaluated a performance of some kind. For example, in one of the contexts, Ann saw a presentation, and Ann's feelings toward the presentation (true state) were shown on a scale from zero to three hearts (e.g., two out of three hearts filled in red color; see Figure **??** for an example of the heart scale). The question of interest was "Do you think Ann thought the presentation was / wasn't X?" and participants responded by choosing either "no" or "yes." The target could be one of four possible words: *terrible*, *bad*, *good*, and *amazing*, giving rise to eight different possible utterances (with negation or no negation). Each participant read 32 scenarios, depicting every possible combination of states and utterances. The order of context items was randomized, and there were a maximum of four repeats of each context item per participant.

Behavioral results. We analyzed the data by collapsing across context items. For each utterance-state pair, we computed the posterior distribution over the semantic weight (i.e., how consistent X utterance is with Y state) assuming a uniform prior over the weight (i.e., a standard Beta-Binomial model). Meanings of the words as judged by participants were as one would expect (Figure 1).

Full statistics on human data

We used Bayesian linear mixed-effects models (`brms` package in R; Bürkner, 2017) using crossed random effects of true state and goal with maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013; Gelman & Hill, 2006). The full statistics are shown in Table 1.

Model fitting and inferred parameters

Other than speaker goal mixture weights explained in the main text (shown in Table **??**), the full model has one global parameter: the speakers' (both S_1 and S_2) soft-max

Table 1

Predictor mean estimates with standard deviation and 95% credible interval information for a Bayesian linear mixed-effects model predicting negation production based on true state and speaker goal (with both-goal as the reference level).

Predictor	Mean	SD	95% CI-Lower	95% CI-Upper
Intercept	0.88	0.13	0.63	1.12
True state	2.18	0.17	1.86	2.53
Goal: Informative	0.47	0.17	0.14	0.80
Goal: Kind	0.97	0.25	0.51	1.49
True state * Informative	-1.33	0.18	-1.69	-0.98
True state * Kind	-0.50	0.22	-0.92	-0.07

Table 2

Inferred negation cost and speaker optimality parameters for all model variants.

Model	Cost of negation	Speaker optimality
ninformational only	1.58	8.58
ninformational, presentational	1.89	2.93
ninformational, social	1.11	3.07
ninformational, social, presentational	2.64	4.47
presentational only	2.58	9.58
social only	1.73	7.23
social, presentational	2.49	5.29

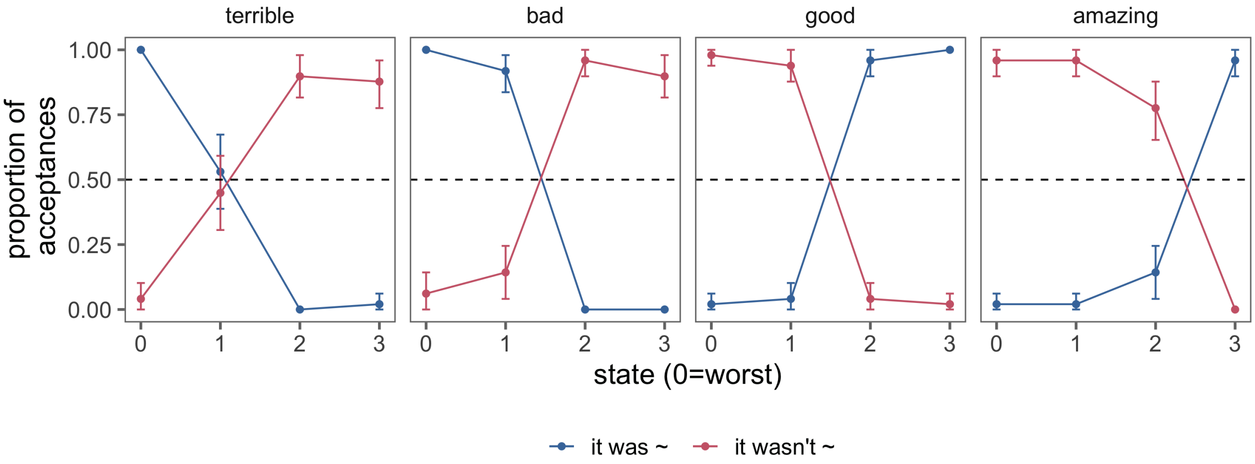

Figure 1. Semantic measurement results. Proportion of acceptances of utterance types (shown in different colors) combined with target words (shown in different facets) given the true state represented on a scale of hearts. Error bars represent 95% confidence intervals.

parameter, which we assume to be the same value α and infer from the data. We put a prior that was consistent with those used for similar models in this model class: $\alpha \sim Uniform(0, 20)$. Finally, we incorporate the literal semantics data into the RSA model by maintaining uncertainty about the semantic weight of utterance w for state s, for each of the states and utterances, and assuming a Beta-Binomial linking function between these weights and the literal semantics data (see *Literal semantics task* above). We infer the posterior distribution over all of the model parameters and generate model predictions based on this posterior distribution using Bayesian data analysis (Lee & Wagenmakers, 2014). We ran 4 MCMC chains for 80,000 iterations, discarding the first 40,000 for burnin. The inferred values of parameters are shown in Table 2.

Data Availability

Our model, preregistration of hypotheses, procedure, data, and analyses are available at https://github.com/ejyoon/polite_speaker.

Supplemental Figures

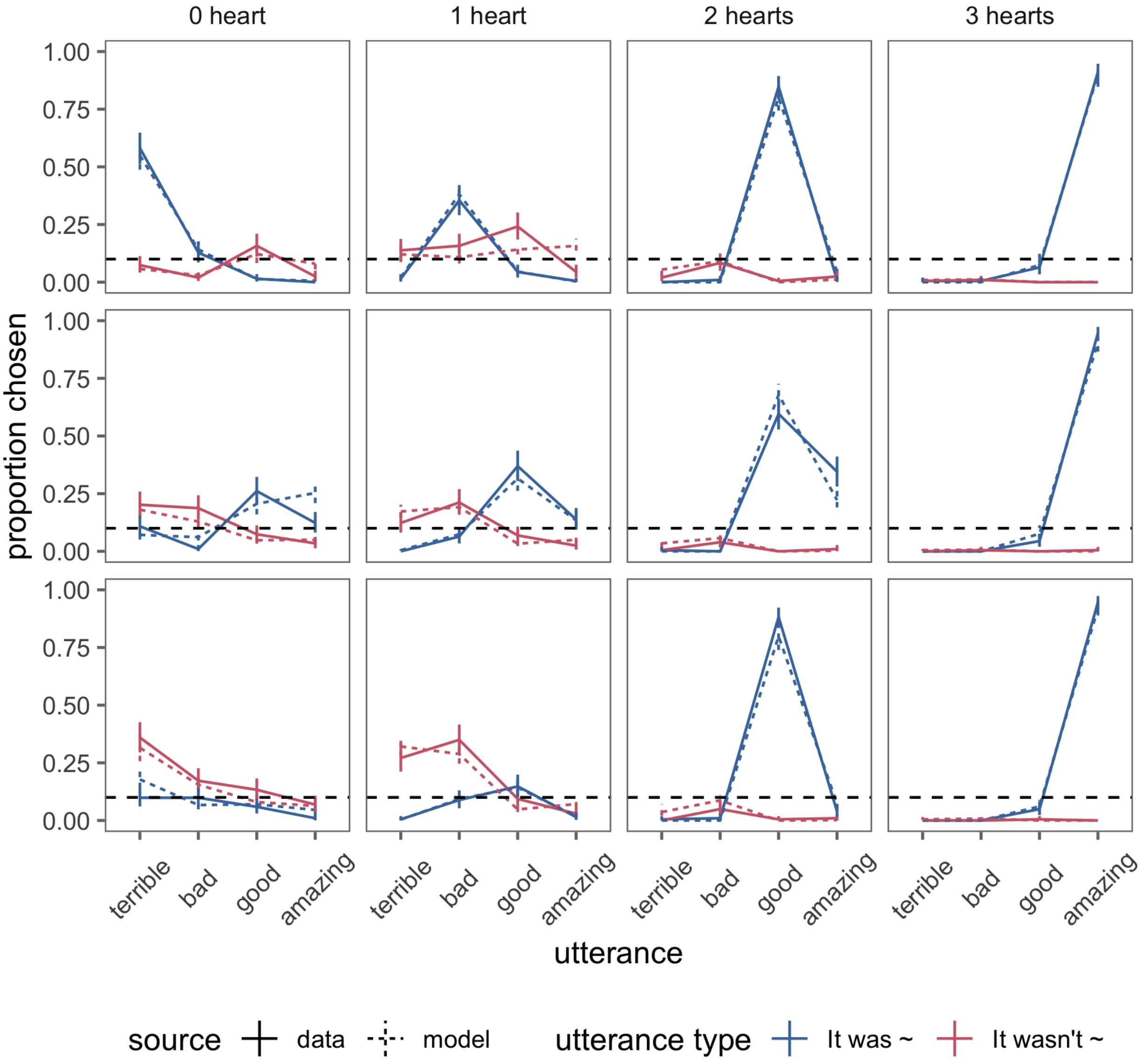

Figure 2. Experimental results (solid lines) and fitted predictions from the full model (dashed lines) for speaker production. Proportion of utterances chosen (utterance type – direct vs. indirect – in different colors and words shown on x-axis) given the true states (columns) and speaker goals (rows). Error bars represent 95% confidence intervals for the data and 95% highest density intervals for the model. Black dotted line represents the chance level.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bürkner, P.-C. (2017). brms: An R package for bayesian multilevel models using Stan.

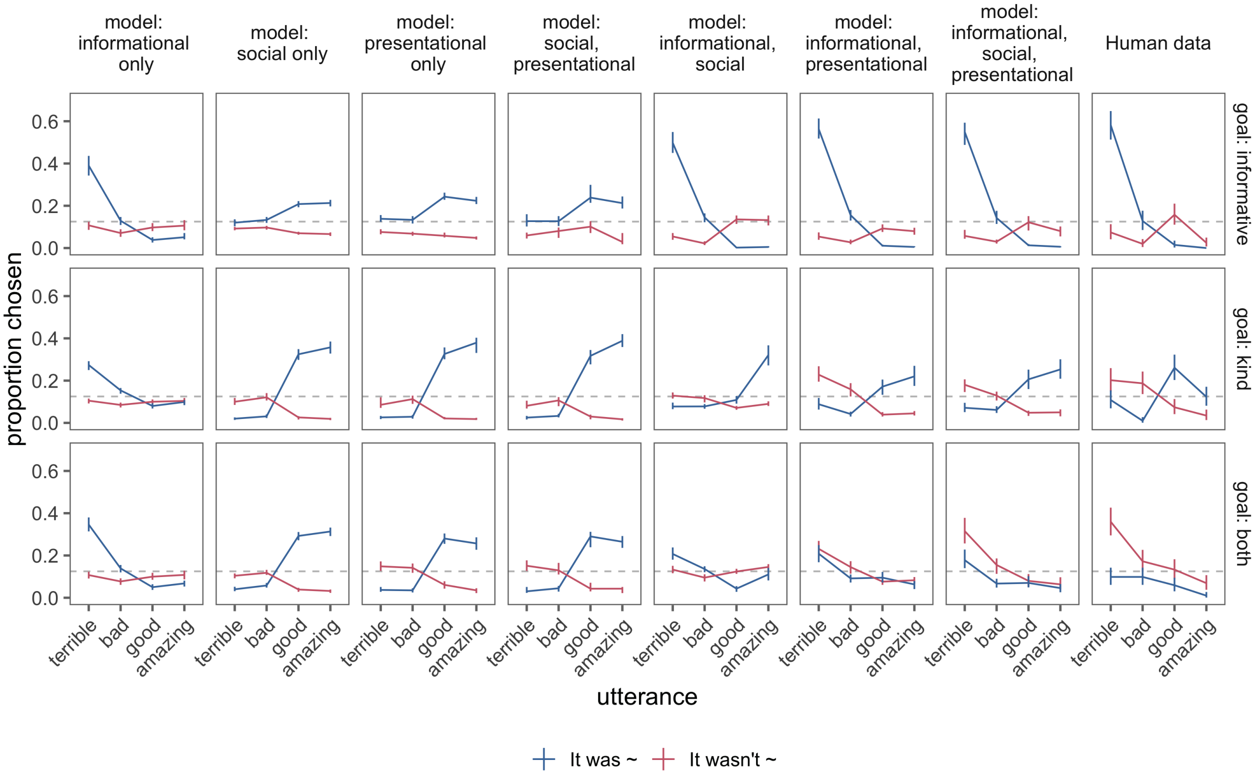

Figure 3. Comparison of predictions for proportion of utterances chosen by pragmatic speaker from possible model variants (left) and human data (rightmost) for average proportion of negation produced among all utterances, given true state of 0 heart and speaker with a goal to be informative (top), kind (middle), or both (bottom). Gray dotted line indicates chance level at 12.5%. Error bars represent 95% confidence intervals for the data (rightmost) and 95% highest density intervals for the models (left).

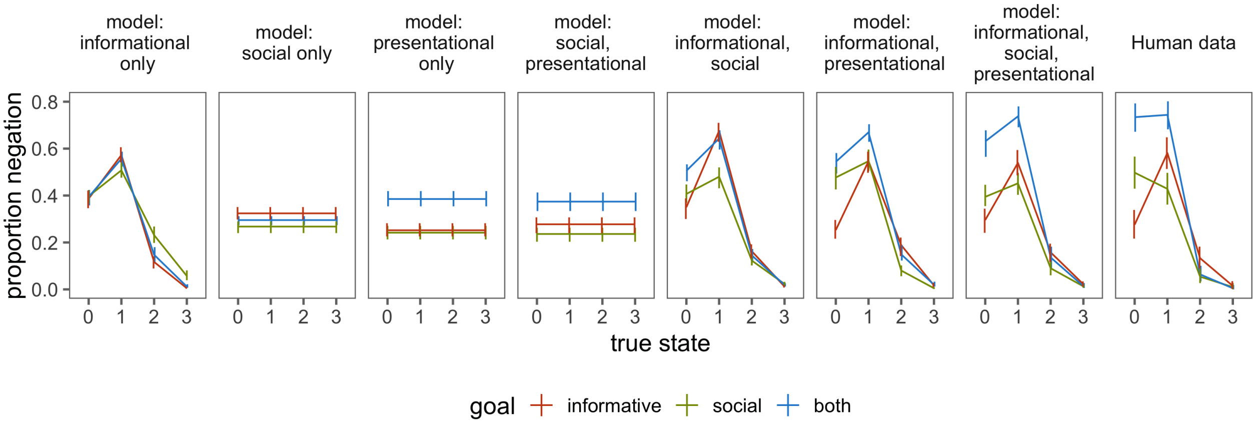

Figure 4. Experimental results (left) and fitted model predictions (right) for average proportion of negation produced among all utterances, given true states (x-axis) and goals (colors).

Journal of Statistical Software, *80*(1), 1–28. doi:10.18637/jss.v080.i01

Gelman, A., & Hill, J. (2006). *Data analysis using regression and* *multilevel/hierarchical models*. Cambridge university press.

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge Univ. Press.