Polite speech emerges from competing pressures to be (and look) informative and kind

Erica J. Yoon[1,F], Michael Henry Tessler[1,F], Noah D. Goodman[1], & Michael C. Frank[1]

[1] Department of Psychology, Stanford University

[F] These authors contributed equally to this work.

Author Note

FIXME.

Correspondence concerning this article should be addressed to Erica J. Yoon, 450 Serra Mall, Bldg. 420, Rm. 290, Stanford, CA 94305. E-mail: ejyoon@stanford.edu

Abstract

Conveying information in a false or indirect manner in consideration of listeners' wants (i.e. being polite) seemingly contradicts an important goal of a cooperative speaker: information transfer. We model production of polite speech in which speakers deviate from being maximally informative for social reasons. In this work, we show that polite speech emerges from a set of competing goals: to be informative, to be kind and provide positive value to others, and to be self-presentational and *appear* helpful. We formalize this tradeoff between speaker's competing goals using a utility-theoretic model, and show the model is able to predict people's polite speech production judgments. Our extension of formal theories of language to account for speakers' social goals represents an advance in understanding of human speech.

Keywords: keywords

Word count: X

Polite speech emerges from competing pressures to be (and look) informative and kind

We don't always say what is on our mind. Although "close the window!" would be sufficient, we say "can you please ... ?" or "would you mind ... ?" And rather than telling the uncomfortable truth, we lie ("Your dress looks great!") and prevaricate ("your poem was so... appropriate to the occasion"). These kinds of utterances are puzzling for standard views of language use, which see communication as the transfer of information from a sender to a receiver (Bühler, 1934; M. C. Frank & Goodman, 2012; Jakobson, 1960; Shannon, 1948). Under information-based views, the transfer ought to be efficient and accurate: The speaker should choose a succinct utterance from which the listener can recover their intended meaning (Grice, 1975; Searle, 1975), and the information transferred should be accurate and truthful to the extent that the speaker knows or believes to be true. Polite speech – like the examples above – then violates basic expectations about the nature of communication: it is typically inefficient and underinformative, and sometimes even outright false. So why are we polite?

Theories of politeness explain deviations from optimal information transfer in language by assuming that speakers take into account social, as well as informational, concerns. These concerns are sometimes expressed as sets of polite maxims (Leech, 1983) or social norms (Ide, 1989), but the most influential account of politeness relies on the notion of "face" to motivate deviations (Brown & Levinson, 1987; Goffman, 1967). On this theory, speakers seek to maintain both their and the listeners' freedom to act ("negative face") as well as desires to be liked, approved, and related to ("positive face"). Both inefficient indirect speech and untruthful lies in communication are then due to speakers' strategic choices relative to possible face threats.

This face-based framework provides an intuitive and appealing explanation of many types of polite speech, but applying it to make quantitative predictions in any individual circumstance can be complicated. It is often not obvious how to quantify a face threat in a given situation (e.g., how much of the listener's positive face will be damaged by hearing "your poem was terrible"), or how social and informational motivations will trade off in the mind of a speaker (given that the poem recital was terrible, should the speaker say that the listener's poem was "okay," "not bad," or "marvelous"?). Concretely, such theorizing does not constrain how an artificial agent like a robot should go about making polite requests, conveying negative evaluations, or delivering bad news. Further, it does not take into account the recursive nature of reasoning about face: Speakers may choose particular strategies not only to preserve face genuinely, but also to be *seen* as doing so, hence appearing to be considerate and socially apt.

To address these challenges, we develop a utility-theoretic model for understanding polite speech, in a unified framework to quantify tradeoffs between different goals that a speaker may have. In our model, speakers attempt to maximize a set of competing utilities: an informational utility, derived via effective information transmission; a social utility, derived by being kind and providing positive affect to others; and a self-presentational utility, derived by appearing in a particular way to other agents. Speakers then can choose between different utterances on the basis of their expected utility (including their cost to utter, approximated by the length of the utterance). The lie that a poem "was good" provides social utility by making its writer feel good, but does not inform about the true state of the world. Further, if the writer suspects that it was in fact terrible, the speaker runs the risk of being seen as uncooperative.

Formally, these utilities are weighed within a rational speech act (RSA) model. RSA models take a probabilistic approach to pragmatic reasoning in language (M. C. Frank & Goodman, 2012; Goodman & Frank, 2016): Speakers are modeled as agents who choose utterances by reasoning about their effects on a listener relative to their cost, while listeners are modeled as choosing interpretations by reasoning about speakers and their goals. This class of models has been effective in understanding a wide variety of complex linguistic behaviors, including vagueness (Lassiter & Goodman, 2017), hyperbole (Kao, Wu, Bergen, & Goodman, 2014), and irony (Kao & Goodman, 2015), among others. More broadly, RSA models provide a instantiation for language of the idea that human social cognition can be approximated via reasoning about others as rational agents who act to maximize their subjective utility (Baker, Saxe, & Tenenbaum, 2009), a hypothesis which has found support in a wide variety of work with both adults and children (e.g., Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016; Liu, Ullman, Tenenbaum, & Spelke, 2017).

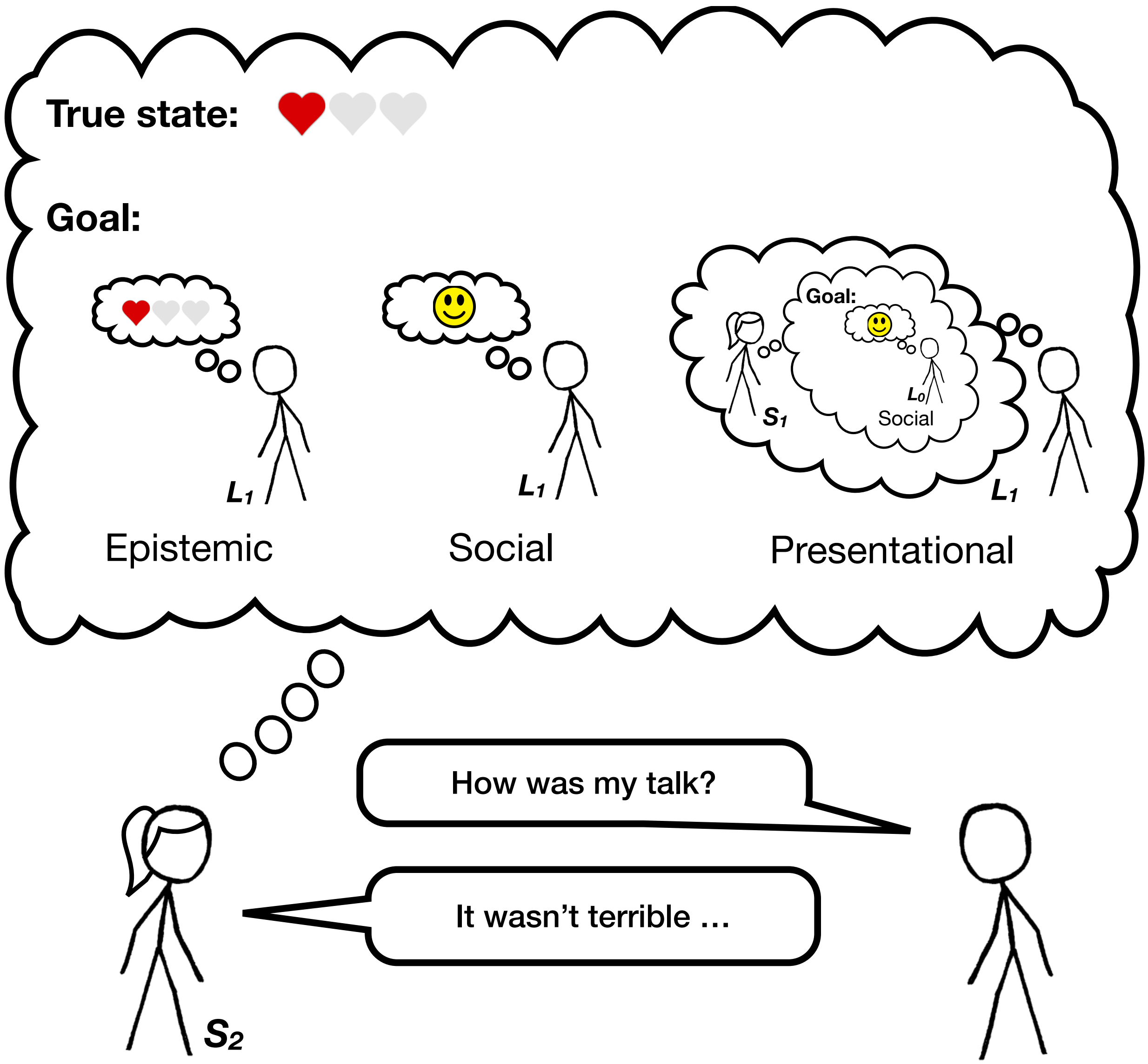

Figure 1. Diagram of the model: The pragmatic speaker observes the true state and determines her goal between three utilities (epistemic, social, and presentational), and produces an utterance.

RSA models are defined recursively such that speakers reason about listeners, and vice versa. By convention the level of this recursion is numbered such that a pragmatic listener L_1 reasons about what intended meaning and goals would have led a speaker S_1 to produce a particular utterance. Then S1 reasons about a "literal listener" L_0, who is modeled as attending only to the literal meanings of words (rather than their pragmatic implications), and hence grounds the recursion. The target of our current work is a model of a polite speaker S_2: S_2 reasons about what utterance to say to L_1 by considering the set of utilities described above: namely, whether an utterance results in L_1 gaining information, feeling positively, or judging S_2 to be either informative or kind (Figure 1).

We evaluate our model by predicting human behavioral data in situations where polite language use is expected. Imagine Bob recited his poem and is ignorant of the quality of his poem recital; he asks Ann how well he did. Ann (the pragmatic speaker S_2) produces an utterance w based on the true state of the world s (i.e., the rating truly deserved by Bob's recital) and a set of goal weights $\hat{\phi}$, each of which determines how much she would like to prioritize a particular goal compared to other possible goals. The speaker chooses utterances depending on their expected utility, specifically as a softmax which interpolates between maximizing and matching (via the parameter λ_{S_2}; Goodman & Stuhlmüller, 2013):

$$P_{S_2}(w|s, \hat{\phi}) \propto \exp(\lambda_{S_2} \cdot \mathbb{E}[U_{total}(w; s; \hat{\phi})])$$

.

What goals must the speaker consider to arrive at a polite utterance? We consider three utilities: informational, social, and presentational. The total utility of an utterance is the weighted combination of the three utilities minus the cost $C(w)$:

$$U_{total}(w; s; \hat{\phi}) = \phi_{inf} \cdot U_{inf}(w; s) + \phi_{soc} \cdot U_{soc}(w; s) + \phi_{pres} \cdot U_{pres}(w; s) - C(w)$$

.

The first utility term is a standard *informational utility* (U_{inf}), which represents the speaker's desire to be epistemically helpful. The informational utility captures the amount of information a literal listener (L_0) would still not know about the world state after hearing the speaker's utterance:

$$U_{inf}(w) = \ln(P_{L_1}(s|w))$$

.

For aspects of the world with affective consequences for the listener (e.g., Bob and his poem recital), we assume speakers produce utterances that make listeners feel like they are in a good state. *Social utility* (U_{soc}) is the value, or expected subjective utility, to the listener of the state inferred given the utterance. This value captures the idea that people want to hear that they are in a good state of the world (e.g., that Bob's poem recital was good). We use a simple linear value function (V) to map states to subjective values: better ratings are more positively valued:

$$U_{soc}(w) = \mathbb{E}_{P_{L_1}(s|w)}[V(s)]$$

.

If listeners are aware that speakers have goals and try to infer what those goals are, speakers may choose utterances in order to convey that they had certain goals in mind. The third component, *presentational utility* (U_{pres}), captures the extent to which the speaker appears to the listener to have a particular goal in mind (e.g. to be kind). The speaker gains presentational utility when her listener believes she has certain goals – that she is trying to be informative or kind. Formally,

$$U_{pres}(w) = \ln(P_{L_1}(\phi_{S_1} \mid w)) = \ln \int_s P_{L_1}(s, \phi_{S_1} \mid w)$$

.

The speaker considers the beliefs of listener L1, who hears an utterance and jointly infers both the speaker's utilities and the true state of the world:

$$P_{L_1}(s, \hat{\phi}|w) \propto P_{S_1}(w|s, \hat{\phi}) \cdot P(s) \cdot p(\hat{\phi})$$

.

This presentational utility, which is the most novel aspect of our model, is higher-order in that it can only be defined for a speaker thinking about a listener who evaluates a speaker. (That is, it can be defined for S_2, but not S_1.)

Finally, utterances that are more complex incur a greater cost, $C(w)$ – capturing the general pressure towards economy in speech. In our work, utterances with negation (e.g., "not terrible") are slightly more costly than their equivalents with no negation (inferred from data; see Supplemental Materials).

Intuitively, when Bob's performance is good, Ann's utilities align to lead her to say something positive. By saying "[Your poem recital] was amazing," Ann is being both truthful and kind, and that is likely to be clear to Bob. But, if Bob's recital is poor, Ann is in a bind: She could be kind and say it was great, but she does so at the cost of conveying the wrong information to Bob, if he mistakenly infers Ann's goal to be truthful and his recital to be actually good. Worse yet, Bob could infer that she is "just being nice," inferring her goal to be social, and discount her comment as uninformative. Alternatively, she could directly say the truth ("It was bad"), but then Bob would think Ann didn't care about him. What is a socially-aware speaker to do? Our model predicts that indirect speech – like "It wasn't bad" – helps navigate Ann's dilemma. It conveys some true information while being sufficiently open-ended to spare Bob's feelings. Further, by incurring the slightly higher cost involved in producing another word suggests that Ann had reasons for not saying a simpler alternative like "It was good," and thus it provides a signal to Bob that Ann takes his feelings into account in her choice.

We made a direct, pre-registered test of our model by instantiating the example above in an online experiment (N = 202). Participants read scenarios in which we provided information on the speaker's (Ann's, in our example) feelings toward some performance or product (e.g., poem recital; *true state*), which were shown on a scale from zero to three hearts (e.g. one out of three hearts). For example, one trial read: "Imagine that Bob gave a poem recital, but he didn't know how good it was. Bob approached Ann, who knows a lot about poems, and asked"How was my poem?" We also manipulated the speaker's *goal* across trials: to be *informative* and "give accurate and informative feedback"; to be *social* and "to make the listener feel good"; or to be *both* informative and social at the same time. We hypothesized that each of the three goals will represent a tradeoff between the three utilities in our model described above (their inferred values are available in the Supplementary Materials). In a single trial, each scenario was followed by a question that asked for the most likely utterance by Ann. Participants selected one of eight possible utterances, by choosing between *It was* vs. *It wasn't* and then among *terrible*, *bad*, *good*, and *amazing.*

Our primary behavioral hypothesis was that speakers who found themselves describing bad states (e.g., Bob's performance was bad) and who had as goals to be both informative and social would produce more indirect, negative utterances ("It wasn't terrible"). These indirect speech acts serve to save the listener's face while also conveying a vague estimate of the true state. This prediction was confirmed: a Bayesian mixed-effects model predicting negation as a function of true state and goal yielded a significant interaction, such that a speaker with both informational and social goals produced more negation in worse states compared to a speaker with only the informational goal (M = -1.33, [-1.69, -0.98]) and social goal (M = -0.50, [-0.92, -0.07]). Rather than eschewing one of their goals to increase utility along a single dimension, participants chose utterances that jointly satisfied their conflicting goals by producing indirect, polite speech.

To connect these behavioral data more directly to our model, we next built a Bayesian data analytic model to integrate out the parameters of the RSA model (e.g., the condition-specific goal-weights for the speaker) and provide a principled way to incorporate

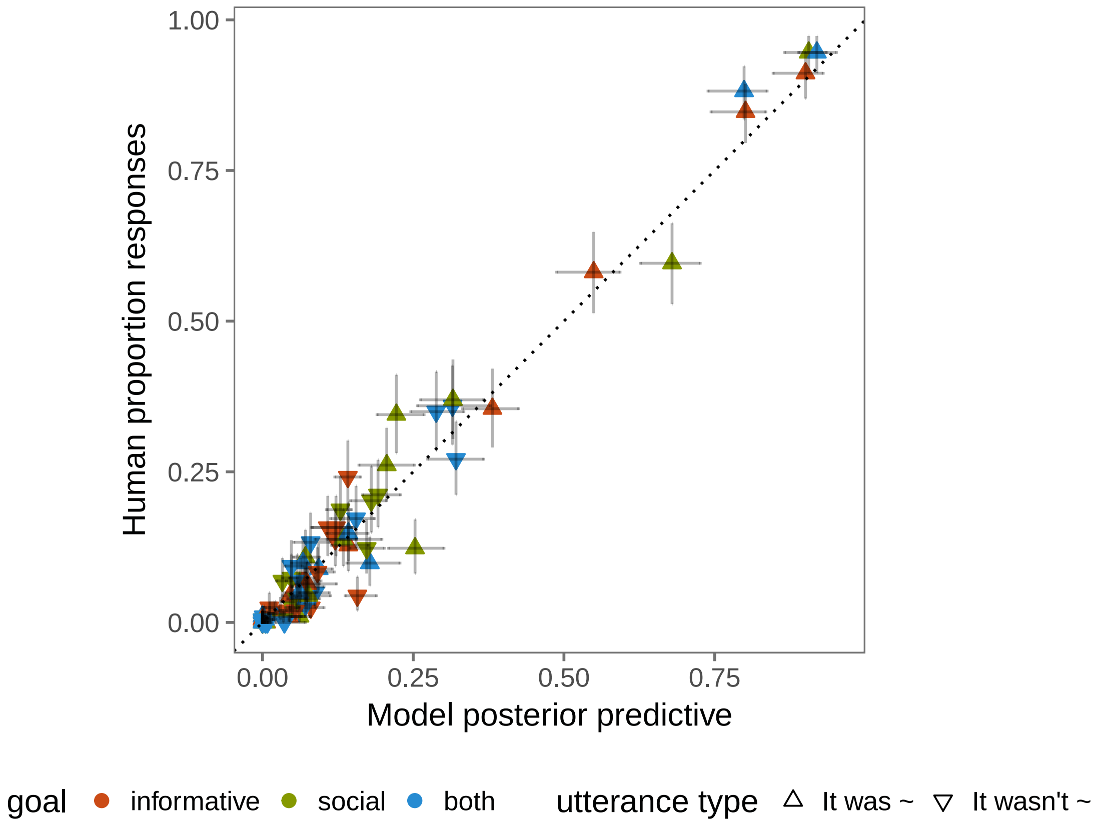

Figure 2. Full distribution of human responses vs. model predictions. Error bars represent 95% confidence intervals for the data (vertical) and 95% highest density intervals for the model (horizontal).

judgments about the literal meanings of the utterances into our model's predictions [Lee and Wagenmakers (2014); see Supplementary Materials]. Using an independent sample of N=51 participants, we measured how participants judged our possible utterances to apply to each of the levels on the heart scale (e.g., to what extent is "terrible" true of 2 out of 3 hearts?). These measurements are used in the Bayesian data analysis to approximate the semantics of the words as interpreted by the literal listener agent L0 (see Supplementary Materials for literal semantic results; see our pre-registered model, hypothesis, and procedure at FIXME).

Predictions from the full polite speaker model showed a strong fit to participants' utterance choices ($r^2(96) = 0.97$; Figure 2). We also compared the predictions of our model with model variants containing different subsets of the three utilities in the full model

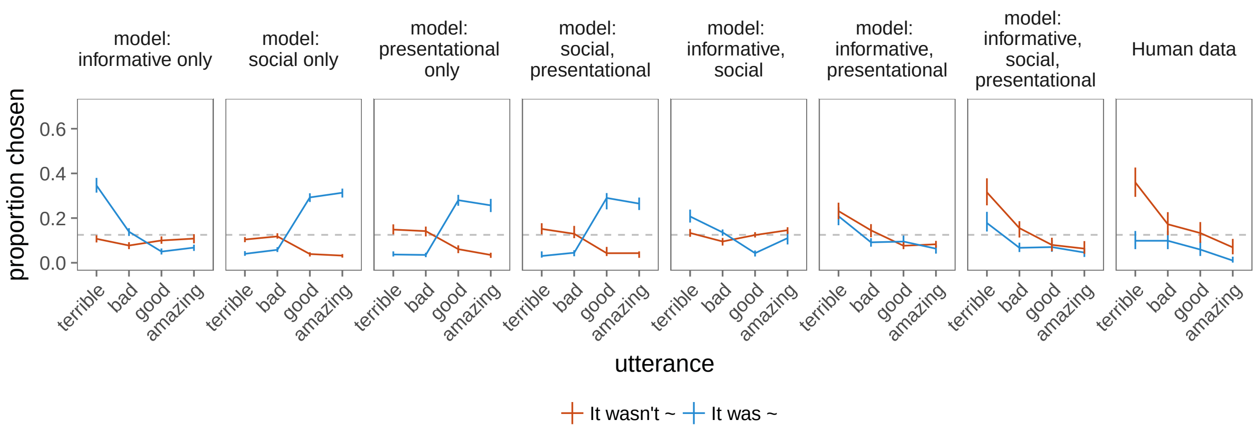

Figure 3. Comparison of predictions for proportion of utterances chosen by pragmatic speaker from possible model variants (left) and human data (rightmost) for average proportion of negation produced among all utterances, given true state of 0 heart (on a scale of 0 to 3) and speaker with both goals. Gray dotted line indicates chance level at 12.5%.

(Figure 3; see Supplemental Materials: Model Comparison). Both the variance explained and the likelihood were the highest for the full model (see Table 1 and Figure 3). In particular, only the full model captured the participants' preference for negation in the condition in which the speaker had both goals to be informative and social about truly bad states, as we hypothesized. The full model was superior to: the model with social and presentational utilities, which predicted outright false statements ("It was good"); the model with informational and social utilities, which predicted truthful statements "It was terrible" and "It wasn't amazing" (that is semantically true when the poem was terrible); and to the model with informative and presentational utilities, which predicted that the speaker would now care about being seen as informative and nice, but still wanting to be as truthful ("It was terrible") as she is presentational ("It wasn't terrible"). Thus, all three – informational, social, and presentational – utilities were required to fully explain participants' choices, correctly predicting that they would prefer to say an indirect speech ("It wasn't terrible") about a bad performance.

Politeness is a puzzle for purely informational accounts of language use. Incorporating

Table 1

Comparison of variance explained for each model variant and log Bayes factors (with the marginal likelihood for the full model as the denominator).

Model	Variance explained	log BF
model: informative only	0.83	274.89
model: social only	0.22	885.52
model: presentational only	0.23	873.83
model: social, presentational	0.23	864.00
model: informative, social	0.92	25.06
model: informative, presentational	0.96	11.14
model: informative, social, presentational	0.97	1.00

social motivations can provide an explanatory framework, but such intuitions have been resistant to formalization or precise testing. To overcome this issue, we created a utility-theoretic model of language use that captured the interplay between competing informational, social, and presentational goals. A preregistered experimental test of the model confirmed its ability to capture human judgments, unlike comparison models that used only a subset of the full utility structure.

To better measure choice behavior, our experiment abstracted away from natural interactions in a number of ways. Real-life Anns will have access to a potentially infinite range of utterances to manage the same tradeoff ("It's so hard to write a good poem," "That metaphor in the second stanza was so relatable!"). Under our framework, each utterance will have strengths and weaknesses relative to the speaker's goals, though computation in an unbounded model presents technical challenges (see Goodman & Frank, 2016).

Managing listeners' inferences is a fundamental task for a socially conscious speaker. Following Brown and Levinson (1987) we hypothesize that cross-cultural differences in politeness are a product of different weightings within the same utility structure. Systematic measurements of these weights could be an approach to understanding the vast range of politeness practices found across languages. Further, politeness is only one of the ways that language use deviates from pure information transfer. When we flirt, insult, boast, and empathize, we balance information transmission with the goal to affect others' feelings or present particular views of ourselves. A similar utility structure to the one we employed here could give insights into these behaviors as well.

The formalization of the presentational utility is especially meaningful in that it begins to precisely define self-oriented motivations behind polite speech and other related behaviors. Brown and Levinson, and other theories of politeness described that other- vs. self-oriented strategies are different (e.g., maximize approval of other, minimize praise of self; Leech, 1983), but did not explain how the motivations of the two are related or how they trade off to inform the speaker's utterance choices. In our current model, the self-oriented concern stems from an other-oriented concern, as the speaker wants to appear to care about the other person's face or access to knowledge. The model then makes precise predictions about how the speaker considering both of these concerns will choose her utterances. This work then can be extended to not only other speech acts, but also a wide range of behaviors that can be modeled as utility-driven inference in a social context (Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017; Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013) where agents need to take into account concerns about both self and others.

In sum, this work takes a concrete step toward quantitative models of the nuances of human speech. And it moves us closer to courteous computation – to computers that communicate with tact.

Acknowledgments

This work was supported by NSERC PGS Doctoral scholarship PGSD3-454094-2014 to EJY, NSF Graduate Research Fellowship DGE-114747 to MHT, ONR grant

N00014-13-1-0788 to NDG, and NSF grant BCS 1456077 to MCF.

Supplemental Materials

Materials and Methods

Literal semantic task. We probed judgments of literal meanings of the target words assumed by our model and used in all our experiments. 51 participants with IP addresses in the United States were recruited on Amazon's Mechanical Turk. We used 13 different context items in which someone 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). 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 five possible words: terrible, bad, good, and amazing, giving rise to ten 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. For this and the subsequent experiment, 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. Meanings of the words as judged by participants were as one would expect (see Figure S1). We used the fraction of participants that endorsed utterance w for state s to set informative priors to infer posterior credible values of the literal meanings from data in the speaker production experiment.

Speaker production task. 202 participants with IP addresses in the United States were recruited on Amazon's Mechanical Turk. As in the literal semantic task above, we used scenarios in which a person (e.g., Bob) gave some performance and asked for another person (e.g., Ann)'s opinion on the performance (see Fig. 2). Additionally, we provided information on the speaker Ann's goal – to make Bob feel good, or to give as accurate and informative feedback as possible, or both – and the true state – how Ann actually felt about Bob's performance (e.g., two out of three hearts, on a scale from zero to three hearts). Each participant read 12 scenarios, depicting every possible combination of goals (3) and states (4). The order of context items was randomized, and there were a maximum of two repeats of each context item per participant. Each scenario was followed by a question that read, "If Ann wanted to make Bob feel good but not necessarily give informative feedback (or to give accurate and informative feedback but not necessarily make Bob feel good, or BOTH make Bob feel good AND give accurate and informative feedback), what would Ann be most likely to say?" Participants indicated their answer by choosing one of the options on the two dropdown menus, side-by-side, one for choosing between It was vs. It wasn't and the other for choosing among terrible, bad, good, and amazing.

Supplementary Text

Data analysis. We used R (Version 3.4.3; R Core Team, 2017) and the R-packages *BayesFactor* (Version 0.9.12.2; Morey & Rouder, 2015), *bindrcpp* (Version 0.2; Müller, 2017a), *binom* (Version 1.1.1; Dorai-Raj, 2014), *brms* (Version 2.0.1; Bürkner, 2017), *coda* (Version 0.19.1; Plummer, Best, Cowles, & Vines, 2006), *directlabels* (Version 2017.3.31; Hocking, 2017), *dplyr* (Version 0.7.4; Wickham, Francois, Henry, & Müller, 2017), *forcats* (Version 0.2.0; Wickham, 2017a), *ggplot2* (Version 2.2.1; Wickham, 2009), *ggthemes* (Version 3.4.0; Arnold, 2017), *gridExtra* (Version 2.3; Auguie, 2017), *here* (Version 0.1; Müller, 2017b), *jsonlite* (Version 1.5; Ooms, 2014), *langcog* (Version 0.1.9001; Braginsky, Yurovsky, & Frank, n.d.), *lme4* (Version 1.1.15; Bates, Mächler, Bolker, & Walker, 2015), *magrittr* (Version 1.5; Bache & Wickham, 2014), *Matrix* (Version 1.2.12; Bates & Maechler, 2017), *papaja* (Version 0.1.0.9655; Aust & Barth, 2017), *purrr* (Version 0.2.4; Henry & Wickham, 2017), *RColorBrewer* (Version 1.1.2; Neuwirth, 2014), *Rcpp* (Eddelbuettel & Balamuta, 2017; Version 0.12.14; Eddelbuettel & François, 2011), *readr* (Version 1.1.1; Wickham, Hester, & Francois, 2017), *rwebppl* (Version 0.1.97; Braginsky, Tessler, & Hawkins, n.d.), *stringr* (Version 1.2.0; Wickham, 2017b), *tibble* (Version 1.3.4; Müller & Wickham, 2017), *tidyr*

Table 2

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: Social	0.97	0.25	0.51	1.49
True state * Informative	-1.33	0.18	-1.69	-0.98
True state * Social	-0.50	0.22	-0.92	-0.07

(Version 0.7.2; Wickham & Henry, 2017), and *tidyverse* (Version 1.2.1; Wickham, 2017c) for all our analyses.

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 (**???**, **???**).

Model fitting and inferred parameters. In the speaker production task, participants were told what speakers' intentions were (e.g., wanted to make Bob feel good). We assume that the intention descriptions conveyed the weight mixtures ϕ_{epi}, ϕ_{soc}, ϕ_{pres}, and ϕ_{S_1} that the speaker was using. We put uninformative priors on each of these mixtures ($\phi \sim Uniform(0, 1)$) and inferred their credible values separately for each goal condition ("wanted to X") using Bayesian data analytic techniques (Lee & Wagenmakers, 2014). We ran 4 MCMC chains for 80,000 iterations, discarding the first 40,000 for burnin. The inferred values of weight mixtures for each model variant (with different ϕ components) and other

Table 3

Inferred phi parameters from all model variants with more than one utility.

Model	goal	ϕ_{inf}	ϕ_{soc}	ϕ_{pres}	ϕ_{S_1}
informative, social, presentational	both	0.36	0.11	0.54	0.36
informative, social, presentational	informative	0.36	0.02	0.62	0.49
informative, social, presentational	social	0.25	0.31	0.44	0.37
informative, presentational	both	0.64	NA	0.36	0.17
informative, presentational	informative	0.77	NA	0.23	0.33
informative, presentational	social	0.66	NA	0.34	0.04
informative, social	both	0.54	0.46	NA	NA
informative, social	informative	0.82	0.18	NA	NA
informative, social	social	0.39	0.61	NA	NA
social, presentational	both	NA	0.38	0.62	0.55
social, presentational	informative	NA	0.35	0.65	0.75
social, presentational	social	NA	0.48	0.52	0.66

parameters are shown in Table 3 and Table 4 respectively.

Table 4

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

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

Supplemental Figures

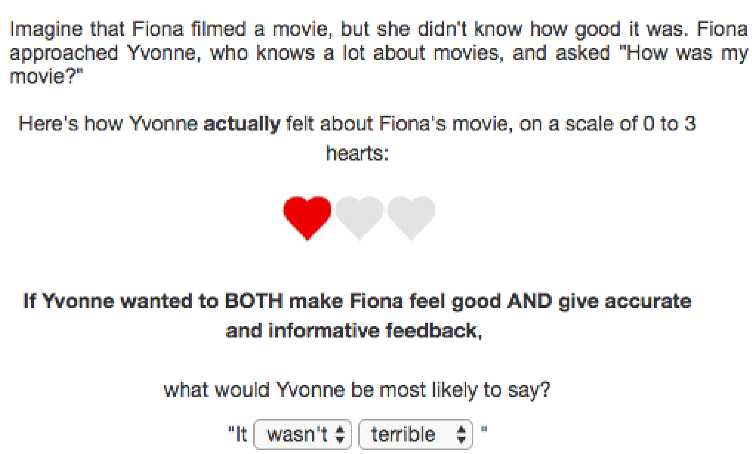

Figure 4. Example of a trial in the speaker production task.

.

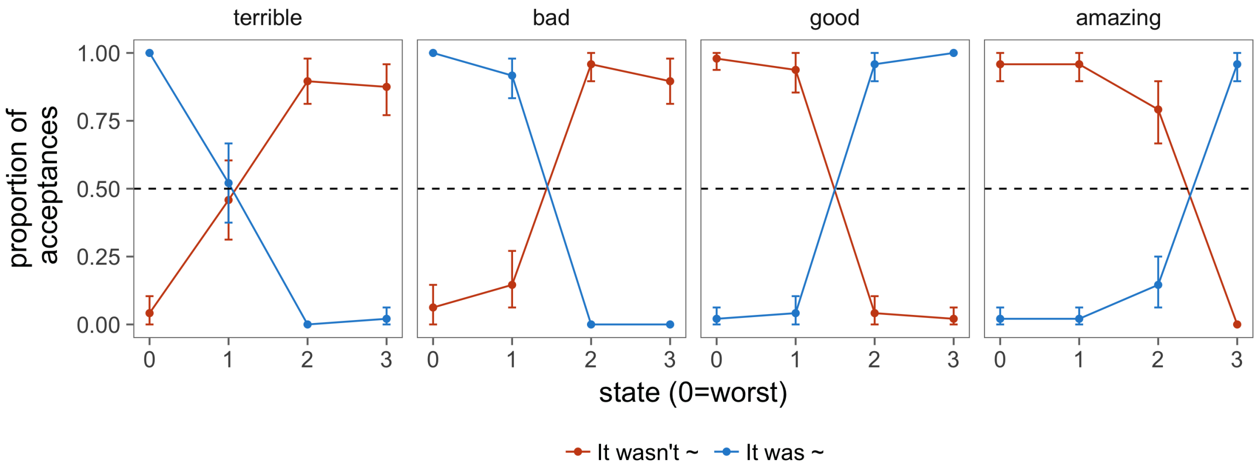

Figure 5. Semantic measurement results. Proportion of acceptances of utterance types (shown in different colors) combined with target words (shown in different facets) given the true state represented on a scale of hearts. Error bars represent 95% confidence intervals.

References

Arnold, J. B. (2017). *Ggthemes: Extra themes, scales and geoms for 'ggplot2'*. Retrieved from https://CRAN.R-project.org/package=ggthemes

Auguie, B. (2017). *GridExtra: Miscellaneous functions for "grid" graphics*. Retrieved from https://CRAN.R-project.org/package=gridExtra

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Bache, S. M., & Wickham, H. (2014). *Magrittr: A forward-pipe operator for r*. Retrieved from https://CRAN.R-project.org/package=magrittr

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 0064.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.

Bates, D., & Maechler, M. (2017). *Matrix: Sparse and dense matrix classes and methods*.

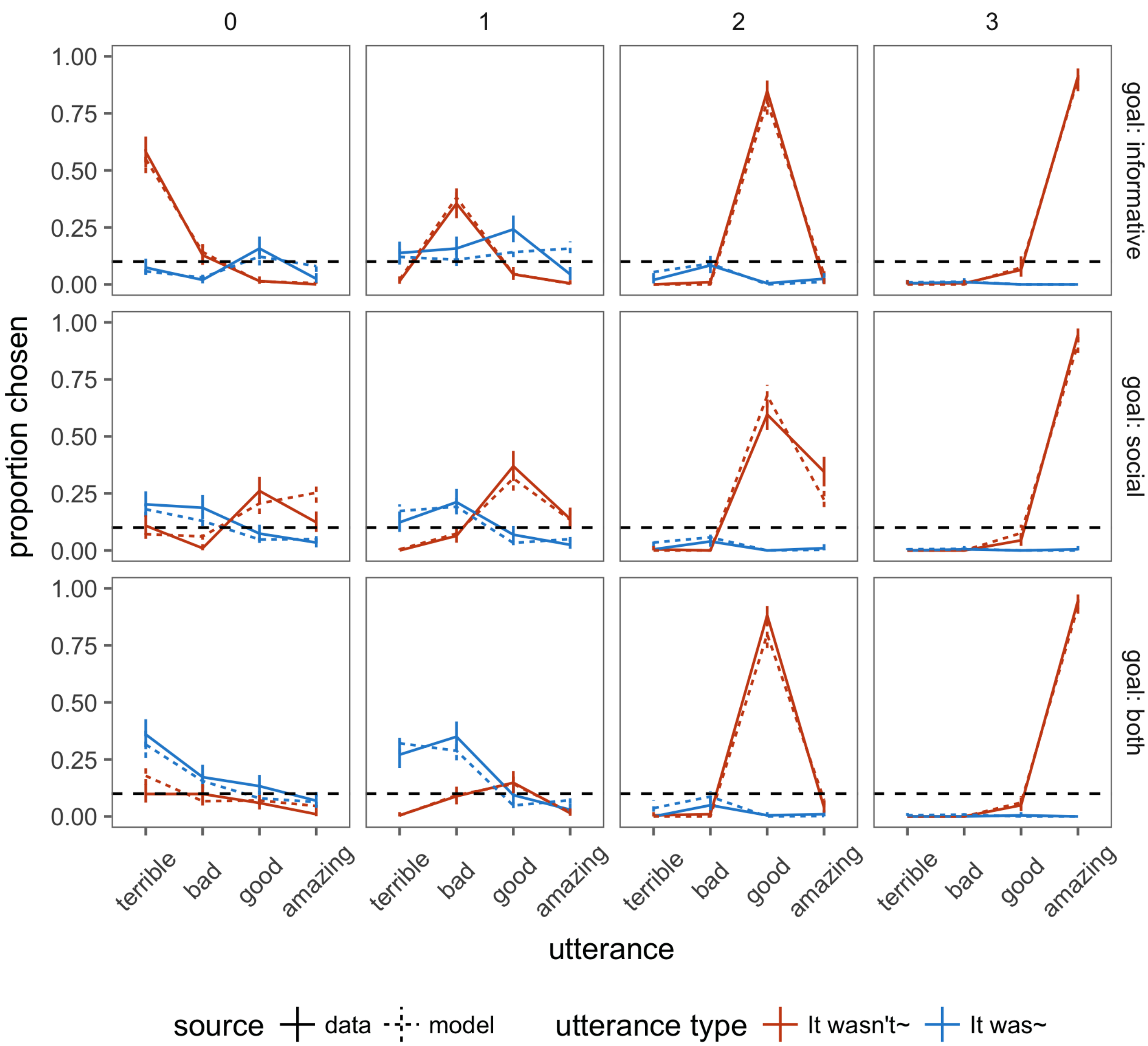

Figure 6. Experimental results (solid lines) and fitted predictions from the full model (dashed lines) for speaker production. Proportion of utterances chosen (utterance type – direct vs. indirect – in different colors and words shown on x-axis) given the true states (columns) and speaker goals (rows). Error bars represent 95% confidence intervals for the data and 95% highest density intervals for the model. Black dotted line represents the chance level.

Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models

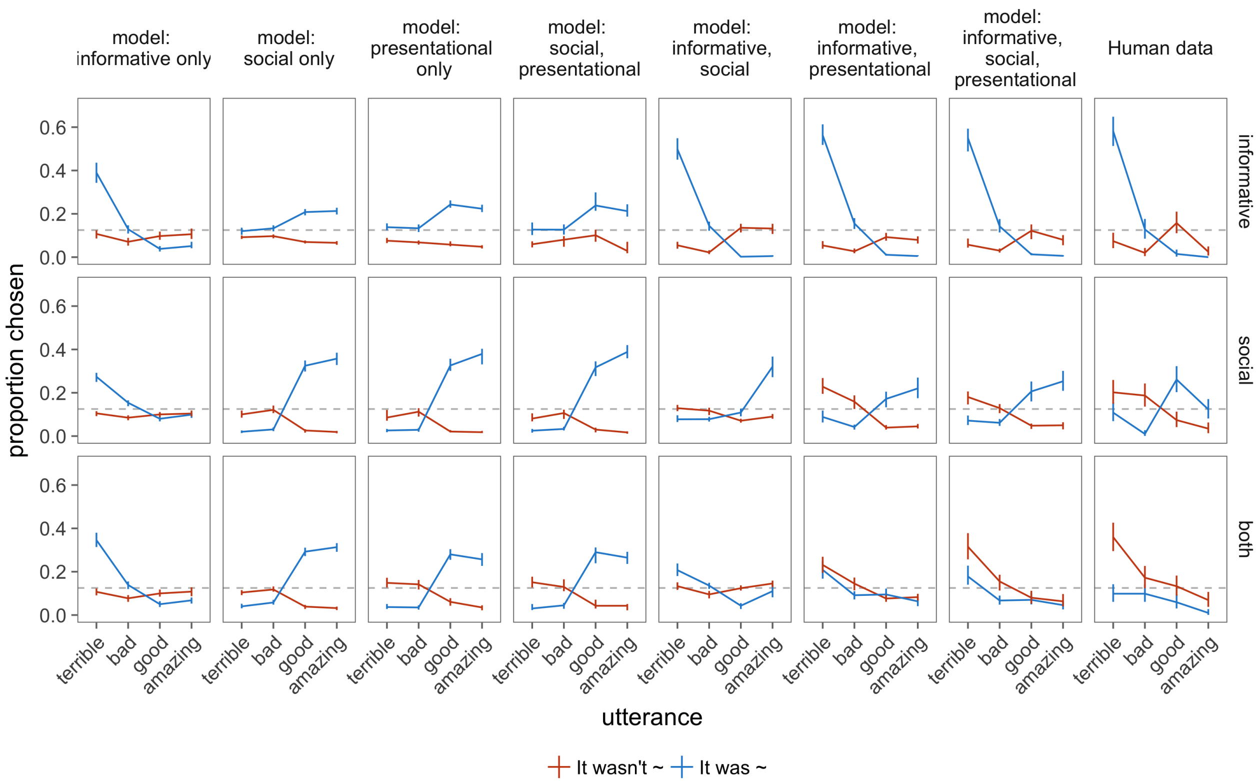

Figure 7. Comparison of predictions for proportion of utterances chosen by pragmatic speaker from possible model variants (left) and human data (rightmost) for average proportion of negation produced among all utterances, given true state of 0 heart and speaker with informative (top), social (middle), and both goals (bottom). Gray dotted line indicates chance level at 12.5%.

using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Braginsky, M., Tessler, M. H., & Hawkins, R. (n.d.). *Rwebppl: R interface to webppl*. Retrieved from https://github.com/mhtess/rwebppl

Braginsky, M., Yurovsky, D., & Frank, M. (n.d.). *Langcog: Language and cognition lab things*. Retrieved from http://github.com/langcog/langcog

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge university press.

Bühler, K. (1934). *Sprachtheorie*. Oxford, England: Fischer.

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

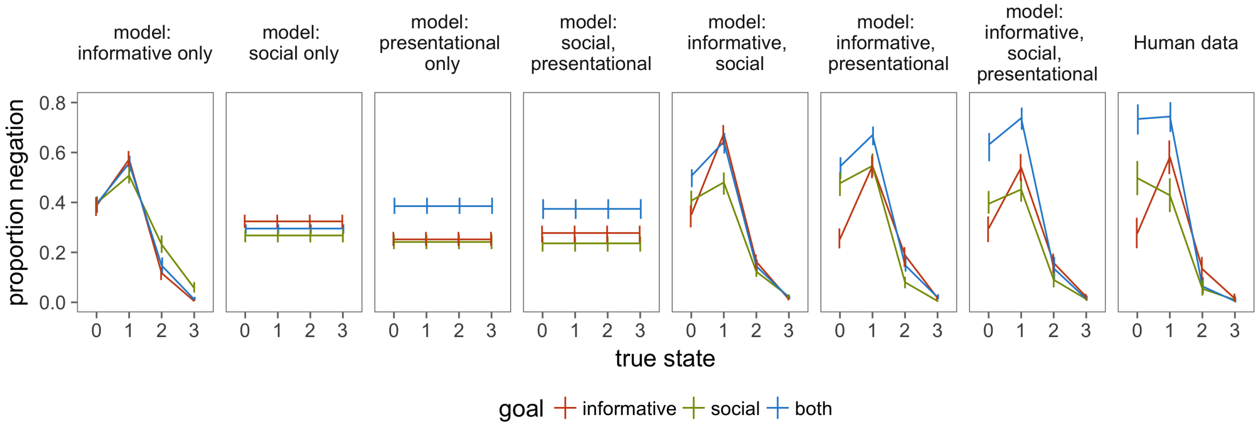

Figure 8. Experimental results (left) and fitted model predictions (right) for average proportion of negation produced among all utterances, given true states (x-axis) and goals (colors).

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

Dorai-Raj, S. (2014). *Binom: Binomial confidence intervals for several parameterizations*. Retrieved from https://CRAN.R-project.org/package=binom

Eddelbuettel, D., & Balamuta, J. J. (2017). Extending extitR with extitC++: A Brief Introduction to extitRcpp. *PeerJ Preprints*, *5*, e3188v1. doi:10.7287/peerj.preprints.3188v1

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. doi:10.18637/jss.v040.i08

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face interaction*. Aldine.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*(1), 173–184.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and* *semantics* (Vol. 3, pp. 41–58). Academic Press.

Hamlin, K. J., Ullman, T. D., Tenenbaum, J. B., Goodman, N. D., & Baker, C. L. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science*, *16*(2), 209–226.

Henry, L., & Wickham, H. (2017). *Purrr: Functional programming tools*. Retrieved from https://CRAN.R-project.org/package=purrr

Hocking, T. D. (2017). *Directlabels: Direct labels for multicolor plots*. Retrieved from https://CRAN.R-project.org/package=directlabels

Ide, S. (1989). Formal forms and discernment: Two neglected aspects of universals of linguistic politeness. *Multilingua-Journal of Cross-Cultural and Interlanguage Communication*, *8*(2-3), 223–248.

Jakobson, R. (1960). Linguistics and poetics. In *Style in language* (pp. 350–377). MA: MIT Press.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589–604.

Kao, J. T., & Goodman, N. D. (2015). Let's talk (ironically) about the weather: Modeling verbal irony. In *Proceedings of the 37th annual conference of the Cognitive Science Society*.

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007.

Lassiter, D., & Goodman, N. D. (2017). Adjectival vagueness in a bayesian model of interpretation. *Synthese*, *194*(10), 3801–3836.

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

Cambridge Univ. Press.

Leech, G. (1983). *Principles of pragmatics*. London, New York: Longman Group Ltd.

Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science*, *358*(6366), 1038–1041.

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of bayes factors for common designs*. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Müller, K. (2017a). *Bindrcpp: An 'rcpp' interface to active bindings*. Retrieved from https://CRAN.R-project.org/package=bindrcpp

Müller, K. (2017b). *Here: A simpler way to find your files*. Retrieved from https://CRAN.R-project.org/package=here

Müller, K., & Wickham, H. (2017). *Tibble: Simple data frames*. Retrieved from https://CRAN.R-project.org/package=tibble

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer palettes*. Retrieved from https://CRAN.R-project.org/package=RColorBrewer

Ooms, J. (2014). The jsonlite package: A practical and consistent mapping between json data and r objects. *arXiv:1403.2805 [Stat.CO]*. Retrieved from https://arxiv.org/abs/1403.2805

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for mcmc. *R News*, *6*(1), 7–11. Retrieved from https://journal.r-project.org/archive/

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Searle, J. (1975). Indirect speech acts. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 59–82). Academic Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, *27*, 623–656.

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from http://ggplot2.org

Wickham, H. (2017a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2017b). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H. (2017c). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from https://CRAN.R-project.org/package=tidyverse

Wickham, H., & Henry, L. (2017). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Francois, R., Henry, L., & Müller, K. (2017). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., Hester, J., & Francois, R. (2017). *Readr: Read rectangular text data*. Retrieved from https://CRAN.R-project.org/package=readr