

Dissertation Proposal: Modeling polite speech: Developmental and cross-cultural investigations

1 Introduction

Language users hear and produce **polite speech** on a daily basis. But being polite conflicts with one important goal of cooperative communication: exchanging information efficiently and accurately (Grice 1975). To be polite, people produce indirect requests that are much longer than simple imperatives (“It would be great if you could close that window” as opposed to “Close that window.”), and tell white lies to make others feel good (“Your new dress is gorgeous!”) Thus, speakers convey information inefficiently and risk losing accurate information (indirect remarks) or even intentionally convey wrong information (lies). If information transfer was the only currency in communication, a cooperative speaker would find polite utterances undesirable because they are potentially misleading.

Following a classic theory that a cooperative speaker has both an epistemic goal to improve the listener’s knowledge state as well as a social goal to minimize any potential damage to the interactants’ self-image (*face*; Brown and Levinson 1987), we conceptualize polite speech as reflecting a tradeoff between information transfer and face-saving (*tradeoff hypothesis*). In our previous work, we developed a novel computational model (described in sec 3.1) that captures this tradeoff, providing some support for this model from the judgments of US adults in simple language comprehension tasks (see sec 3.2).

In the current proposal, we propose to perform empirical research that informs basic questions regarding politeness, and also tests our formal model and explores its predictions across much wider range of phenomena. In particular, we have three aims:

1. Verify tradeoff hypothesis through two case studies of polite speech: **white lies** and **indirect speech**;
2. Examine **children**’s inferences about polite speech and track the development of polite speech understanding, with the hypothesis that children’s inferential patterns can be explained using our formal model; and
3. Capture **cultural variations** in reasoning about polite speech by comparing adult and child participants in US versus India.

In sum, I will attempt to provide an in-depth understanding of a range of polite speech phenomena through a formal model and empirical work probing developmental and cross-cultural differences. These efforts take a concrete step toward quantitative models of the nuances of polite speech, thereby contributing to a richer understanding of pragmatic language processing and production more generally.

2 Backgrounds

Language users hear and produce polite speech on a daily basis. Polite speech ranges from words of apology (“sorry”) or of gratitude (“thanks”) to compliments (“Your new dress is gorgeous!”) and indirect requests (“It would be great if you could pass that salt”). Simple formulaic markers of politeness (e.g. “sorry,” “thanks,” “please”) fit in well with theoretical views that polite speech arises from people’s tendencies to follow social norms for what constructs ‘proper’ social conduct (e.g., Ide 1989).

More complicated polite expressions are difficult to explain under these views, however, as the literal meanings of the utterances do not precisely match the speakers’ intentions or knowledge states. For example, saying “Can you speak a little louder?”, when the actual intended meaning is “Speak louder!”, does not convey the intended meaning in a maximally efficient manner. Also, a compliment such as “Your new dress is gorgeous!” seemingly indicates that the dress is literally and truthfully gorgeous, but it is natural to think about situations where speakers do not feel that way and are trying to hide their intentions (“That dress is really ugly”).

As such, polite utterances seem to often conflict with one important goal of cooperative communication: accurate and efficient information transfer (Grice 1975). If information transfer was the only currency in communication, a cooperative speaker would find polite utterances undesirable because they are potentially

misleading. People do speak politely, however. Adults and even young children spontaneously produce requests in polite forms (Clark and Schunk 1980; Axia and Baroni 1985), and speakers use politeness strategies even while arguing, preventing unnecessary offense to their interactants (Holtgraves 1997). Listeners even attribute ambiguous speech to a polite desire to hide a truth that could hurt another’s self-image (e.g., Bonnefon et al. 2009). In fact, it is difficult to imagine a world in which human speech was used purely as a medium for conveying only the truth. Intuitively, politeness is one prominent characteristic that differentiates human speech from stereotyped robotic communication, which may try to follow rules to say “please” or “thanks” yet still lack genuine politeness.

Do these facts about politeness imply that people are not cooperative communicators in the Gricean sense? Brown and Levinson (1987) recast the notion of a *cooperative speaker* as one who has both an epistemic goal to improve the listener’s knowledge state as well as a social goal to minimize any potential damage to the hearer’s (and the speaker’s own) self-image, which they called *face*. In their analysis, if the speaker’s intended meaning contains no threat to the speaker or listener’s face, then the speaker will choose to convey the meaning in an efficient manner, putting it *on the record*. As the degree of face-threat becomes more severe, however, a speaker will choose to be polite by producing more indirect utterances. Saying “Can you please speak a little louder?” rather than “Speak louder!” is a more indirect form of request instead of order, which gives the listener a sense of autonomy or freedom from imposition and also bestows better reputation upon the speaker herself. Thus, in Brown and Levinson (1987)’s claim, the motivation for politeness is that deviation from truthfulness of informativity leads to face-saving.

One possible proposal based on Brown and Levinson (1987)’s argument is that language users think about polite speech as reflecting a tradeoff between information transfer and face-saving. When a speaker tries to save face, she hides or risks losing information in her intended message by making her utterance false or indirect to some degree. On the other hand, when a speaker prioritizes truthfulness and informativity, she may risk losing the listener’s (or the speaker’s own) face. In a previous work, my collaborators and I developed a novel computational model (described below) that captures the idea that cooperative speakers attempt to balance between the two goals: information transfer and face-saving. This model builds on a recent formal framework for modeling pragmatic language understanding, the “rational speech act” model (Goodman and Frank 2016).

3 Prior work

In this section I describe our current model for understanding polite speech as reflecting tradeoff between information transfer and face-saving, and the current empirical support for this model based on the case study of white lies. These results have been reported to the Cognitive Science Conference in Yoon et al. (2016).

3.1 Current model

Politeness poses a challenge for formal models of pragmatic language understanding, which assume that speakers’ goals are to communicate informatively about some aspect of the world. The Rational Speech Act (RSA) framework (Goodman and Frank 2016) is part of a family of approaches to formalizing Gricean pragmatics, and builds on earlier work in game theoretic models of language understanding (Benz et al. 2006). The RSA models describe language understanding as recursive probabilistic inference between a pragmatic listener and an informative speaker. This framework has been successful at capturing the quantitative details of a number of language understanding tasks, but it neglects the social goals a speaker may pursue. Our model extends standard RSA, which assumes a speaker with only an epistemic goal to improve the listener’s knowledge state, to take into account a speaker with both the usual epistemic goal and a competing social goal: be kind (i.e., save face).

RSA models a listener as reasoning about a speaker, who chooses utterances optimally given a utility function. Goodman and Stuhlmüller (2013) define speaker utility by the amount of information a *literal listener* would still not know about world state s after hearing a speaker’s utterance w (i.e. *surprisal*), what

we call *epistemic utility*: $U_{epistemic}(w; s) = \ln(P_{L_0}(s | w))$, where the literal listener is a simple agent that takes the utterance to be true. We modeled this agent using Bayesian modeling formalism, which makes use of probabilistic information to capture belief representations (Tenenbaum et al. 2011).

In our new model, “Polite RSA,” we proposed there is a second component to the speaker’s utility related to the intrinsic value of the state in the eyes of the listener, what we call *social utility*. We defined the social utility of an utterance to be the expected utility of the state the listener would infer given the utterance w :

$$U_{social}(w; s) = \mathbb{E}_{P_{L_0}(s|w)}[V(s)],$$

where V is a value function that maps states to subjective utility values—this captures the affective consequences for the listener of being in state s . We take the overall speaker utility to be a weighted combination of epistemic and social utilities:

$$U(w; s; \hat{\beta}) = \beta_{epistemic} \cdot U_{epistemic} + \beta_{social} \cdot U_{social}.$$

The speaker chooses utterances w softmax-optimally given the state s and his goal weights $\hat{\beta}$. The pragmatic listener, denoted L_1 , infers the world state based on this speaker model. We assumed the listener does not know exactly how the speaker weights his competing goals, however. We assumed the pragmatic listener jointly infers the state s and the utility weights of the speaker, $\beta_{epistemic}$ and β_{social} (Goodman and Lassiter 2015; Kao et al. 2014):

$$P_{L_1}(s, \hat{\beta} | w) \propto P_{S_1}(w | s, \hat{\beta}) \cdot P(s) \cdot P(\hat{\beta}) \quad (1)$$

We implemented this model using the probabilistic programming language WebPPL Goodman and Stuhlmüller (2014) and a complete implementation can be found at <http://forestdb.org/models/politeness-cogsci2016.html>.

In summary, given assumptions about word meanings and social utilities for a particular speaker and context, Polite RSA allows us to make quantitative predictions for listener inferences about speaker’s goals and true states of the world.

3.2 Empirical support for the Polite RSA

The predictions of Polite RSA were tested in two experiments. Here I provide details on both experiments, as they will be identical in design to proposed experiments in Section 3.2.3. Within our experimental domain, we assumed there are five possible states of the world corresponding to the value placed on a particular referent (e.g. how good is the presentation the speaker is commenting on): $S = \{s_1, \dots, s_5\}$. We further assumed a uniform prior distribution over possible states of the world. The states have subjective numerical values $V(s_i) = \alpha \cdot i$, where α is a scaling parameter (later inferred from data). The set of utterances is $\{\text{terrible}, \text{bad}, \text{okay}, \text{good}, \text{and amazing}\}$.

3.2.1 Experiment P1: True state inference. We examined listeners’ inferences about the likely state of the world s given a speaker’s utterance (e.g. “It was good”) and a description of the speaker’s intentions (e.g. the speaker wanted to be nice). We presented scenarios in which a person (e.g. Ann) asked for another person (e.g. Bob)’s opinion on her performance. We provided information on Bob’s goal (to be *honest*, *nice*, or *mean*) and what Bob actually said to Ann (e.g. “It [your cake] was okay”), where Bob used one of the five possible words: *terrible*, *bad*, *okay*, *good*, or *amazing*. Then we asked participants to infer the true state of the world (e.g. how Bob actually felt about Ann’s cake). Thus, participants read each story (e.g. Ann baked a cake and asked Bob about it) followed by a prompt that said, e.g., “Bob wanted to be nice: “It was okay,” he said. How do you think Bob actually felt about Ann’s cake?” Participants indicated their answer on a scale of five hearts.¹

¹The experiment can be viewed at: http://langcog.stanford.edu/expts/EJY/polgrice/L2_S/polgrice_L2_S.html.

3.2.2 Experiment P2: Goal inference. Experiment P2 probed listeners’ inferences of the speaker’s goals, given an utterance (e.g. “*It was good*”) and a true state. We presented the same context items and utterances as Experiment 1. But instead of goals, we provided information on the true states (i.e. how Bob actually felt towards Ann’s performance). Then we asked participants to infer the likelihood of Bob’s goals to be *honest*, *nice*, and *mean*. Participants read each scenario followed by a question that read, “Based on what Bob said, how likely do you think that Bob’s goal was to be: honest; nice; mean,” with the three goals placed in a random order below three slider bars, on which the participant could indicate each goal’s likelihood.²

3.2.3 Findings. The findings were consistently in favor of the model predictions: participants’ inferences about the true state of the world (e.g. how much Ann actually liked the talk that Bob gave) differed based on what the speaker said (e.g. Ann said “It was okay”) and whether the speaker’s intended goal was to be honest, nice or mean. For example, when the speaker wanted to be nice and said “It was okay,” participants predicted that Ann actually thought it was mediocre (2 out of 5; see Figure 1). Also, participants’ attributions of different social goals to speakers depended on how well the literal utterance meaning matched the actual rating the performance deserved. When Ann actually thought Bob’s talk was bad but said “It was good”, then people attributed more niceness and less honesty to Ann. Overall, our Polite RSA model displayed strong quantitative fits to the inference data in both experiments ($r^2(75) = 0.74$ and $r^2(75) = 0.82$, respectively).

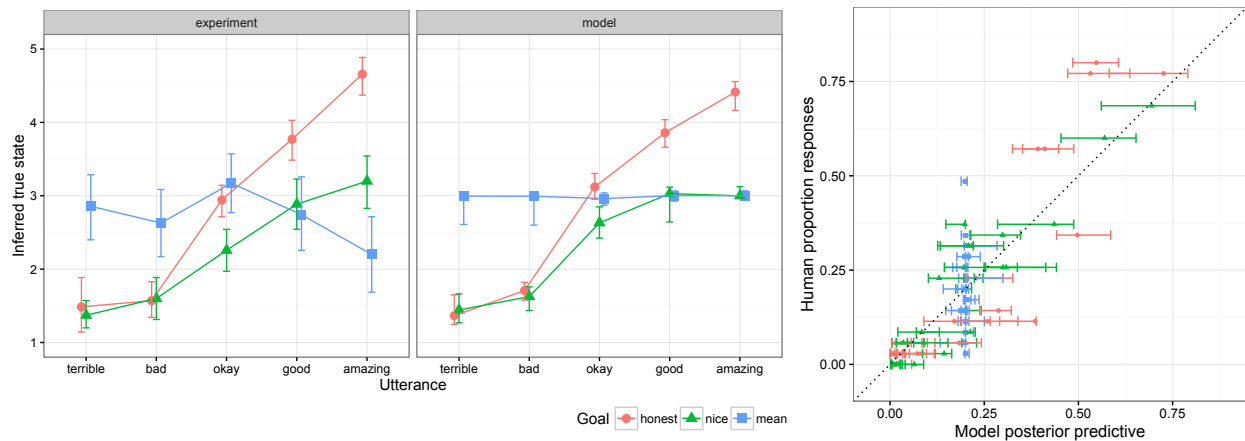


Figure 1: Results from Expt. 2 (left) and model predictions (center) for average states inferred based on a speaker’s goal and utterance. Right: Full distribution of human responses vs. model predictions. Error bars represent 95% confidence intervals for the data and 95% highest density intervals for the model.

3.3 Limitations of previous empirical work

Our prior work provides a promising start for investigating polite speech phenomena, but there are limitations that narrow the scope of our understanding. First, in our prior work we only considered one particular domain of politeness phenomena: white lies. If other domains of politeness phenomena reflect the same tradeoff between face-saving and informativity, then we should see similar inferential patterns for other kinds of polite speech, such as indirect remarks (“I wouldn’t say that her dress is the most wonderful thing I’ve ever seen”).

Second, the empirical support for the model only came from a small, particular population of adults in the US, which incidentally is what comprises almost all of major pragmatic language research as well as most experimental work in psychology (Henrich et al. 2010). But to examine whether our model or any model of pragmatic language understanding applies more broadly, it is crucial to examine other groups that are underrepresented in the literature. Specifically, I propose to look at the same inferences in children as

²The experiment can be viewed at: http://langcog.stanford.edu/expts/EJY/polgrice/L2_G/polgrice_L2_G.html.

well as another cultural group (India). I propose to look at these developmental and cross-cultural differences as well as different instances and factors contributing to polite speech, which will allow us to present a comprehensive theoretical framework to capture the understanding of polite speech.

4 Proposed work

We now take up the challenge of developing the RSA framework to broaden its scope and address the weaknesses described above. Subsections correspond to the broad goals described in sec. 1. Each has a common format: a review of the limitations of the current RSA framework and then descriptions of the theoretical extensions we plan to explore and our proposed empirical tests of the theory.

4.1 Indirect remarks

So far, I have tried to capture one specific area of politeness phenomena: what people infer about white lies given a lay speaker's utterance. However, our model can be extended to make predictions about other kinds of polite speech. Here I focus on one additional case study: indirect remarks. Through indirect remarks, speakers try to convey a particular message in a more nuanced way. This is different from white lies, in that speaker's intentions are no longer hidden, but revealed with suboptimal efficiency. Below I present partially completed work (as reported in Yoon et al. (2017)) as well as proposed research looking at production and comprehension of indirect remarks.

4.1.1 Background and model implications. Why would people speak indirectly? Theoretical accounts of indirect speech in situations of potential conflicts argue that indirect language maintains plausible deniability (Pinker et al. 2008), and suggests higher stakes for the listener than speaker in case the speaker's wants are not fulfilled (Franke and Jäger 2016). For example, a mobster trying to coerce a restaurant owner into paying protection money, who utters, "Your daughter is very sweet. She goes to the school in Willow Road, I believe." avoids the risk of being sued for threat but also suggests to the owner that his stakes for not paying money are high (his daughter would be in danger).

Indirect speech in politeness situations, however, is distinguished from speech aimed upon avoiding legal liability or persuading the listener to defer to the speaker's propositions. Why would it be better to say "I would love another glass of wine, thanks." than "Pour me more wine"? The latter more clearly conveys the guest's intention for the waiter to pour more wine. But the former is less imposing on the waiter, and circumvents an impression that the speaker is in a position to give orders to the listener. Indeed, even when the implied meaning of the requests is the same, people prefer requests whose literal meanings ask for the listener's permission ("Could I ask you where Jordan Hall is?") to those with literal meanings that assume listener's obligation to respond ("Shouldn't you tell me where Jordan Hall is?") (Clark and Schunk 1980). Indirect requests are complicated to manipulate for many reasons, for example due to various possible semantic forms of imperatives, and my proposed work focuses on *indirect remarks* as a simpler case study.

What may lead a speaker to produce indirect remarks? An indirect remark may be motivated by the speaker's goal to convey some face-threatening information, while being seen as a polite person who avoids threatening others' face. In our previous work, we described a pragmatic listener that jointly inferred the true state and the goals of the speaker (Yoon et al. 2016). Building on this model, I describe here a speaker whose goal is to lead this pragmatic listener to infer the true state *and* attribute to the speaker certain goals (e.g., face-saving). For instance, "It wasn't amazing" does not preclude the possibility that the presentation was bad, and may in fact be pragmatically strengthened to mean that it was actually bad. Yet because the speaker does not choose the more direct "It was bad", the listener will infer a face-saving goal. Thus saying "It wasn't amazing" can accomplish the goal of conveying that the presentation was bad while the speaker is seen as not wanting to make the listener feel bad. On the other hand, if the speaker does not care about being seen as face-saving, she will produce less indirect speech. Further, if the presentation was actually good, or even decent, the speaker will prefer to produce a directly positive remark ("It was good") in either case. Thus I predict more indirect speech when the true state is bad, and an interaction with the speaker's

desire to both be informative and be seen as wanting to save face.

There have been some relevant cross-cultural evidence that people do take into account face-informativity tradeoff for polite indirect speech: Hebrew adult speakers rate conventional indirect requests as more polite than direct orders or hints, and reason that both face concern and informativity are important for speech to be considered polite (Blum-Kulka 1987); similarly, English and Korean adult speakers find evasive remarks to be better than direct or irrelevant remarks (Holtgraves and Joong-nam 1990). Finally, Holtgraves and Perdeu (2016) suggested that people think of subtle utterances as reflecting varying degrees of both politeness or uncertainty (related to informativity).

Thus I hypothesize that, similar to white lies, indirect speech reflects speaker’s desires to balance between the goal to be informative (convey information in the most direct manner possible) and the goal to save face, this time concerning the speaker’s own face as well (maintain her reputation for conveying accurate information with intentions to be polite). In the next two subsections, I describe our extended model up to date and its empirical support, as reported in Yoon et al. (2017).

4.1.2 Model extensions and predictions. Here I report extensions to the previous model up to date. We build on pRSA by adding negative utterances and modeling a more sophisticated speaker. First, we extend the utterance alternatives to include negation. Previously we considered five possible utterances: {It was *terrible*, *bad*, *okay*, *good*, and *amazing*}, all direct assertions of specific states (e.g., “It was amazing” would be true for the state of 5 but untrue for the states of 1 or 2). Now the speaker may say, {It *wasn’t* terrible, bad, okay, good, and amazing}. These utterances indirectly address the referent by negating certain state. We assume that it is more costly to say utterances with negation, which makes the utterance morphemically longer and is harder to process (Clark and Chase 1972). In the full data analysis, We put a prior on this negation cost parameters and infer its likely values from the data.

Most importantly, we extended the recursive reasoning in the model. For our experiment, we consider the pragmatic speaker (S_2) who chooses an utterance based on the pragmatic listener model (Eq. 1), thinking about the state as well as goal weights that the pragmatic listener will infer.

$$P_{S_2}(w \mid s, \hat{\beta}) \propto \exp(\lambda_2 \cdot \ln(P_{L_1}(s, \hat{\beta} \mid w)) - C(w))$$

This crucially captures the idea that the speaker both wants to convey the state s , and to be seen as someone with goals $\hat{\beta}$. we simplify from the Yoon et al. (2016) model by including only a single mixture parameter ϕ governing the extent to which the speaker is being informative vs. face saving: $\beta_{epi} = \phi$, $\beta_{soc} = 1 - \phi$.

The S_2 speaker in this proposed model has the goal to convey the state and to be seen as having a particular set of goals. We explored predictions for 3 hypothetical speakers, corresponding to 3 different ϕ mixture parameter weights: (a) an *informative* speaker who wants to convey high epistemic utility (prioritizing information transfer; $\phi = 0.9$) (b) a *social* speaker who wants to convey high social utility (making the listener feel good; $\phi = 0.1$) (c) a *both-goal* speaker who wants to convey a balance between the two utilities ($\phi = 0.5$).³

Figure 2 (left) shows the speaker’s production probabilities associated with producing an indirect speech act (i.e., an utterance with negation) for the three different speakers as the true state of the world is varied. We see, consistent with the intuition, that indirect speech was relatively more preferred in bad states than in good states. As well, we see higher probability of negation production for the speaker who wants to convey both goals (epistemic and social) relative to each goal independently. Indirect speech does not convey that much information and so the informative speaker (a) would disprefer it. The social speaker (b) who wants to convey a face-saving goal would tend to signal a better-than-actual state through direct positive remarks. The

³In addition, the model has a few parameters not of theoretical interest. For the purposes of generating model predictions *a priori*, we assigned values to these parameters consistent with the previous literature with this class of models: the speaker optimality parameter (λ_1 assigned to 2); the pragmatic speaker optimality parameter (λ_2 to 2); the value scale parameter (α to 1) in the utility function; and the parameter governing the cost of producing a negation ($C(u)$ to 2).

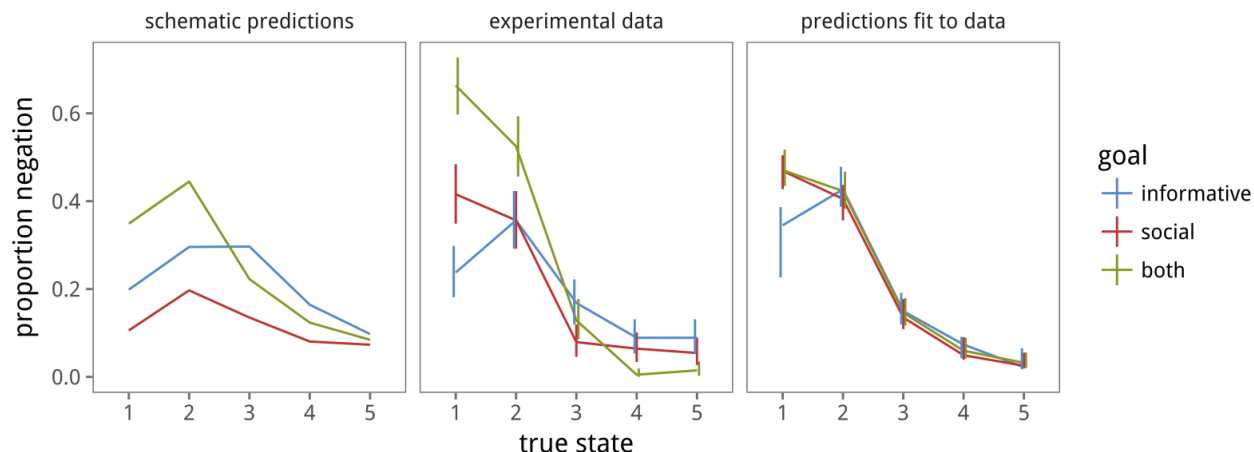


Figure 2: Schematic model predictions (left), experimental results (center) and fitted model predictions (right) for average proportion of negation produced among all utterances, given true states (x-axis) and goals (colors).

both-goal speaker produces indirect remarks to avoid direct remarks that are either true but face-threatening, or face-saving but false.

4.1.3 Empirical test: Experiment 5. To compare against the model predictions, I ran the first experiment to measure participants’ predictions for the most likely utterance (w) produced by the speaker, given a description of the true state. For example, given that Ann wanted to make Bob feel good but felt that his poem deserved 2 out of 5 hearts, what would she say? I hypothesized that when there was no tradeoff between informativity and face-threat avoidance (i.e., when the addressee’s performance was great), speakers would use truthful and face-saving direct remarks (“[Your poem] was amazing”) regardless of their described goals. However, when there was a conflict between the epistemic and social goals (i.e., when the addressee’s performance was poor), a speaker who tried to convey both goals would use vague indirect remarks (“[Your poem] wasn’t terrible”) more often than direct face-threatening remarks (“[Your poem] was bad”; preferred by a speaker who only considered the epistemic goal) or direct face-saving remarks (“[Your poem] was good”; preferred by a speaker who wanted to convey only a social goal).

I 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. Additionally, I 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., 2 out of 5 hearts). Each participant read 15 scenarios, depicting every possible combination of goals and states. 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 *was* vs. *wasn’t* and the other for choosing among *terrible*, *bad*, *okay*, *good*, and *amazing* (see Figure 3).


Our hypotheses for utterance production by speakers with different goals were borne out (see full results in Figure 4).

For good states (4 and 5 hearts), positive direct remarks were judged to be the most likely utterances across all three goal conditions. For less-than-perfect, but still decent states, there was a greater degree of expectation of white lies (e.g., “It was amazing” for 4 hearts) given a social goal. For bad states (1 and 2 hearts),

as predicted, there were more instances of expected indirect remarks overall across all goal conditions given bad states. Critically, speakers with both informative and social goals produced more indirect remarks than were observed in the other two goal conditions (Figure 2, center).

Imagine that Justine wrote a review for a book, but Justine didn't know how good it was. Justine approached Kelly, who knows a lot about writing reviews, and asked "How was my review?"

Here's how Kelly **actually** felt about Justine's review:



If Kelly wanted to make Justine feel good, but not necessarily give informative feedback,

What would Kelly be most likely to say?

"It

Figure 3: Example of a trial in Experiment 1.

4.1.4 Empirical test: Experiment 6. Whereas Experiment 5 looked at the speaker production, I will also look at listener inferences. I will use tasks identical in design to Experiments P1-2, but looking at indirect speech instead of white lies. I will test participants' judgments in contexts in which Bob asks Ann for her opinion on his cookie. I hypothesize that participants will attribute more niceness but less informativity to Ann when she says "It wasn't amazing," compared to "It was terrible." This will show that indirect remarks are similar to white lies and reflect speakers' consideration of face-informativity tradeoff.

4.2 Children's polite speech understanding

Previously we have shown that adults in the US think about polite speech as reflecting a tradeoff between information transfer and face-saving. Will children reason similarly? Here we extend our tradeoff hypothesis to suggest that, starting at a young age, children distinguish between these goals and reason about the optimal degree of tradeoff (i.e. what is the nicest, most helpful thing to say) based on the speaker's intention, listener's need, and cultural expectations. Thus, adults and children should reason about polite language as broadly reflecting a tradeoff between information-saving and face-saving, although such understanding may mature over time.

4.2.1 Background. There is not as much evidence for young children's understanding of polite speech (specifically white lies) as for children's production of polite speech. For example, children as young as 3 years start to tell white lies; they lie to the giver of an undesirable gift and say that the gift is nice (Talwar et al. 2007).

Interestingly, understanding of lies in general and motivations for white lies is displayed in later years than production: 8- and 11-, but not 4-year-olds, demonstrate correct categorization of truths as truths and lies as lies (Bussey 1999); 7- to 11-year-olds rate lies as more positive in politeness than transgression contexts (Heyman et al. 2009). However, no study has yet looked at how children spontaneously reason about motivations behind lies and truths, especially those related to the epistemic-social tradeoff that we are currently interested in.

4.2.2 Empirical test: Experiment 1. To examine our hypotheses about children's polite language understanding, we would like to look at how children reason about speakers who speak truthfully but impolitely, or politely but untruthfully. We propose a procedure that is based on the scenarios used with adults (Experiments P1 and P2), but simplified: participants will be asked to read a story in which, for example, two speakers tasted yucky cookies and were asked how they liked it, either by the baker himself, or someone else. One speaker speaks truthfully ("the cookie was yucky") and the other untruthfully ("the cookie was

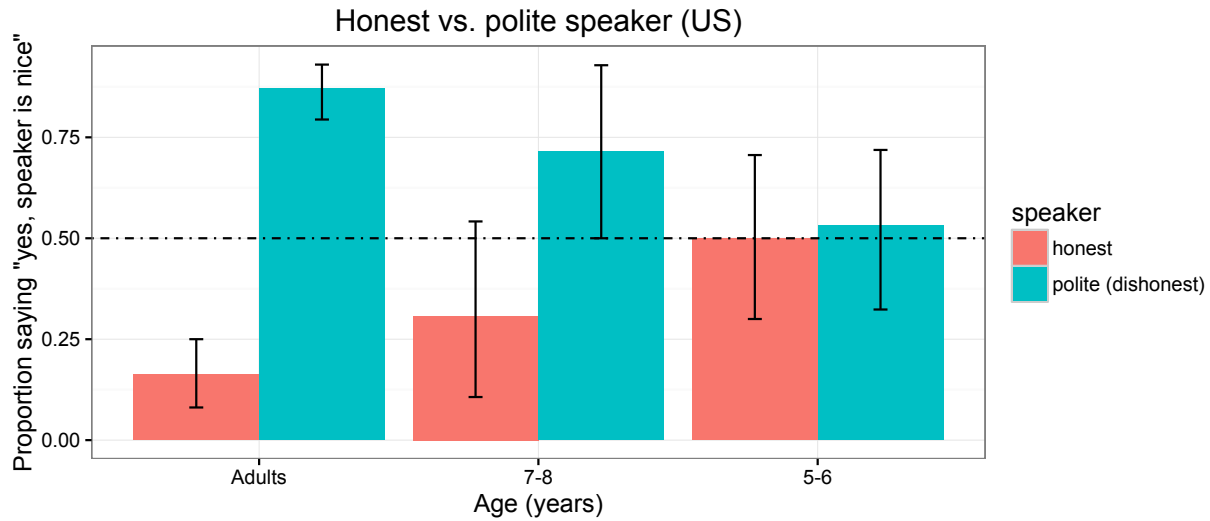


Figure 4: Results from Experiment 1 (pilot), for children’s niceness judgment for honest versus polite (but dishonest) speaker. Error bars represent 95% confidence intervals.

tasty”). Participants will then judge whether each speaker was telling the truth and whether she was nice or mean, and indicate whether they prefer to play with the truthful speaker or the untruthful speaker. This procedure has been piloted and has worked well in India, as well as other two sites of interest (Pilot findings are discussed in the next subsection).

Data collection procedure. Participants will be recruited at children’s museum and nursery/elementary schools in US. Across these three different sites, we will recruit three different age groups (3-4, 5-6, and 7-8-year-olds), with 24 participants per condition. Those who are not exposed to English at least 75% daily will be excluded from data analysis.

Design. There are two independent variables: (1) listener type: whether or not the context was such that the ‘listeners’ in our story asked about their own performance (e.g., cookie that they baked) or another, unknown person’s performance (e.g., a cookie that was lying around for free); and (2) speaker type: whether the ‘speakers’ in our story decide to tell a lie or truth about the performance.

Each participant will be randomly assigned to one ‘listener’ condition (i.e., whether the story is about listener’s performance or someone else’s) and will go through both ‘speaker’ trials (i.e., speaker who tells a lie and speaker who tells the truth). Thus, the ‘listener’ variable is a between-subjects, and the ‘speaker’ variable is a within-subjects factor. There will be two sets of two speaker trials, and the order of the sets will be counterbalanced across subjects.

We will ask the following questions to participants: (1) In each trial, after telling the story of each speaker: “Was [the speaker] in the story nice? Was she mean? Was she telling the truth?” (the order of questions will be counterbalanced); (2) After one set of two trials, comparing two speakers, one who told a lie vs. one who told the truth: “Who do you want to play with more, [polite speaker] or [honest speaker]?”

Analysis plan. We will use a mixed-effects logistics regression model to analyze: (1) participants’ rating for a given speaker’s niceness/meanness/truth-telling; (2) comparisons of polite vs. honest speakers, as indicated by whom they “want to play with.”

4.2.3 Results from pilot work. We have run a pilot study of the proposed work, and Figure 4 is an example analysis conducted on pilot data. We see an interesting developmental trend where 8-year-olds show similar patterns of niceness attribution to adults but do not as strongly differentiate between an honest and polite speaker, whereas 6-year-olds do not seem to differentiate between the honest and polite speaker at all.

4.2.4 Implications for the model. Based on our pilot results, children seem to show similar patterns of inferences for honest versus polite speakers, in that they judge polite speaker as nice and honest speaker as not nice. However, this pattern is less clear in 6-year-olds, who do not differentiate between the two speakers. Interpreting these results, if this pattern holds, is possible in two ways: younger children have the same mechanistic process for evaluating the epistemic-social tradeoff to determine the speaker’s social utility, but this is obscured due to some cause such as task demand (e.g. understanding the question at hand), or younger children do not see the utterance as reflecting the epistemic-social utility at all. The former could be accommodated in our model by describing developmental change via increases in social utility, but further evidence (with greater statistical power) is needed to assess the model fit to the data.

4.3 Cultural variations in polite speech understanding

US adults show understanding of polite speech as reflecting epistemic-social tradeoffs. Now we turn to other populations of interest: different cultural groups. On our model, variability might arise because different cultures may place different emphases on one of these two communicative goals; some cultures may consider it noble to try to make others feel good by lying, whereas some may regard honesty as the most important virtue. We hypothesize that cultural variations in polite communication arise due to these differences in optimal tradeoff, and we test this prediction in the experiment described below.

4.3.1 Background. To the best of our knowledge, there has been no study looking at cross-cultural variations in adults’ polite speech understanding. There have been only a few studies that have looked at cultural variations in children’s ratings of white lies (e.g. Fu and Lee 2007; Ma et al. 2011). For example, Chinese children as young as 7 years rate lie-telling as positive in a public than private situation, and as negative when accurate information would help the listener (Ma et al. 2011). But these studies do not offer comparisons with European American children (cf. Lee et al. 1997). In our proposed work, we hope to directly compare three groups of interest, for both adults and children: US, India, and South Korea. These areas cover a wide range of languages used, religions and socioeconomic statuses, and thus will help provide a comprehensive view over the general trend of pragmatic language understanding across different cultures.

4.3.2 Empirical tests. Experiment 2: Indian and Korean adults’ true state inference. This experiment will be identical in structure as Experiment P1, and participants will reason about the true state given speaker’s utterance and goal. Indian participants will be recruited on Amazon’s Mechanical Turk, and Korean participants by advertising on Facebook. In order to test for linguistic (rather than cultural) differences, we will recruit participants in two batches for each site: one batch that speaks Hindi and the other that speaks English from India; and one batch for Korean and one for English from Korea, and we will compare the data to examine any differences that may arise from language.

Experiment 3: Indian and Korean adults’ goal inference. This experiment will be identical in structure as Experiment P2, and they will reason about the true state given speaker’s utterance and goal. Participants will be recruited in an identical manner to Experiment P2.

Experiment 4: Indian and Korean children’s understanding of polite speech. This experiment will be identical in structure as Experiment 1. Participants will be recruited at elementary schools in India and South Korea. The Korean site has been established by Co-PI Yoon as a testing site in previous work, and the Indian site has been used extensively by PI Frank in unrelated research on mathematics (Frank and Barner 2012; Barner et al. 2016). Across these three different sites, we will recruit three different age groups (3-4, 5-6 and 7-8-year-olds), with a minimum of 10 and maximum of 24 participants per condition, age group and site, allowed by the total testing time at the field sites. They will be tested in English or their native language (Korean, Hindi or Gujarati), with a script that is directly translated from English by an educated native speaker. Those who are not exposed to the language of instruction at least 75% daily will be excluded from data analysis.

4.3.3 Results from pilot work. Figure 5 shows an example analysis conducted on pilot data. Whereas US children attributed more niceness to polite speaker than honest speaker with increasing age, Indian children

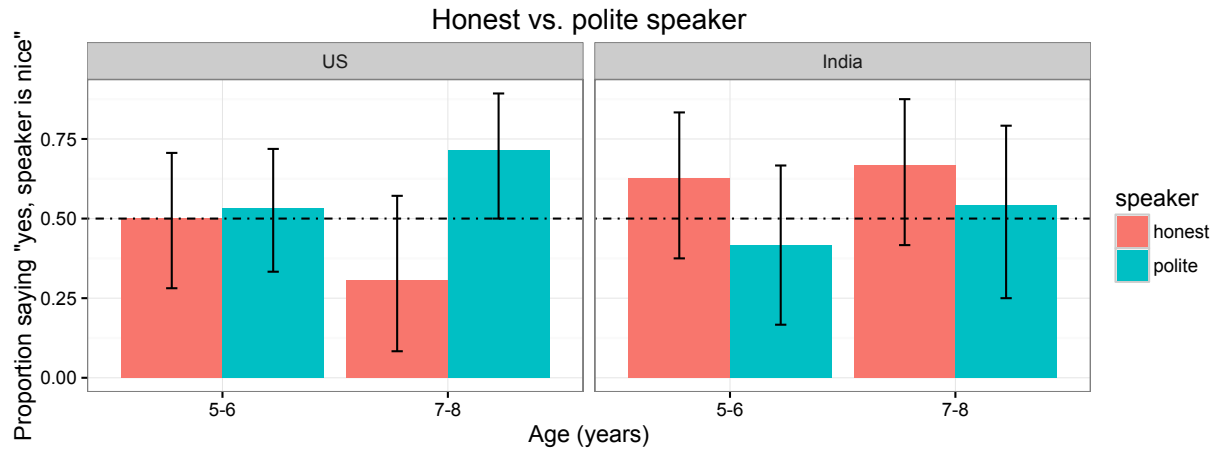


Figure 5: Results from Experiment 4 (pilot), for US versus Indian children’s niceness judgments for honest versus polite (but dishonest) speaker. Error bars represent 95% confidence intervals.

showed the opposite trend of attributing more niceness to honest than polite speaker across both age groups. This is preliminary yet promising evidence that construal of niceness (linked to the notion of optimal tradeoff of truthfulness versus face-saving) may be different across cultures. We will seek to verify this trend with larger sample sizes to account for variability and ensure robustness of judgment patterns.

4.3.4 Implications for the model. Preliminary data from Indian child participants show an interesting reversal in speaker niceness attribution: Indian children tend to say that an honest speaker is nicer than a polite speaker. In light of our model, two interpretations are possible (if the patterns hold up): either Indian children do not reason about polite language based on epistemic-social tradeoff as we hypothesized, or Indian children have a different construal of the word “nice,” to mean “helpful for future performance” (i.e. giving useful feedback) rather than “kind” or “caring”. If the latter is true, then it informs that language of instruction is a critical consideration in conducting these studies, and that Indians may have a different sense of what comprises most helpful communication from the US population. Our proposed work addresses both of these issues, in that we will conduct each Experiment at each site in two languages: the local language and English, to look at any differences in responses that are caused by language rather than cultural differences; and we will inquire what each cultural group thinks of as an optimal, cooperative speaker.

References

- Axia, G. and Baroni, M. R. (1985). Linguistic politeness at different age levels. *Child Development*, pages 918–927.
- Barner, D., Alvarez, G., Sullivan, J., Brooks, N., Srinivasan, M., and Frank, M. C. (2016). Learning mathematics in a visuospatial format: A randomized, controlled trial of mental abacus instruction. *Child development*.
- Benz, A., Jäger, G., and Van Rooij, R. (2006). An introduction to game theory for linguists. In *Game theory and pragmatics*, pages 1–82. Springer.
- Blum-Kulka, S. (1987). Indirectness and politeness in requests: Same or different? *Journal of pragmatics*, 11(2):131–146.
- Bonnefon, J.-F., Feeney, A., and Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, 112(2):249–258.
- Brown, P. and Levinson, S. C. (1987). *Politeness: Some universals in language usage*, volume 4. Cambridge Univ. Press.
- Bussey, K. (1999). Children’s categorization and evaluation of different types of lies and truths. *Child Development*, 70(6):1338–1347.
- Clark, H. H. and Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive psychology*, 3(3):472–517.
- Clark, H. H. and Schunk, D. H. (1980). Polite responses to polite requests. *Cognition*, 8(2):111–143.
- Frank, M. C. and Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, 141(1):134.
- Franke, M. and Jäger, G. (2016). Probabilistic pragmatics, or why bayes? rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, 35(1):3–44.
- Fu, G. and Lee, K. (2007). Social grooming in the kindergarten: the emergence of flattery behavior. *Developmental science*, 10(2):255–265.
- Goodman, N. D. and Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11):818–829.
- Goodman, N. D. and Lassiter, D. (2015). *Probabilistic Semantics and Pragmatics: Uncertainty in Language and Thought*. Wiley-Blackwell.
- Goodman, N. D. and Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, 5(1):173–184.
- Goodman, N. D. and Stuhlmüller, A. (2014). The Design and Implementation of Probabilistic Programming Languages. <http://dippl.org>.
- Grice, H. P. (1975). *Logic and conversation*. Blackwell.
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences*, 33(2-3):61–83.

- Heyman, G. D., Sweet, M. A., and Lee, K. (2009). Children's reasoning about lie-telling and truth-telling in politeness contexts. *Social Development*, 18(3):728–746.
- Holtgraves, T. (1997). Yes, but... positive politeness in conversation arguments. *Journal of Language and Social Psychology*, 16(2):222–239.
- Holtgraves, T. and Joong-nam, Y. (1990). Politeness as universal: Cross-cultural perceptions of request strategies and inferences based on their use. *Journal of personality and social psychology*, 59(4):719.
- Holtgraves, T. and Perdue, A. (2016). Politeness and the communication of uncertainty. *Cognition*, 154:1–10.
- 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.
- Kao, J. T., Wu, J. Y., Bergen, L., and Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, 111(33):12002–12007.
- Lee, K., Cameron, C. A., Xu, F., Fu, G., and Board, J. (1997). Chinese and canadian children's evaluations of lying and truth telling: Similarities and differences in the context of pro-and antisocial behaviors. *Child development*, 68(5):924–934.
- Ma, F., Xu, F., Heyman, G. D., and Lee, K. (2011). Chinese children's evaluations of white lies: Weighing the consequences for recipients. *Journal of experimental child psychology*, 108(2):308–321.
- Pinker, S., Nowak, M. A., and Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of sciences*, 105(3):833–838.
- Talwar, V., Murphy, S. M., and Lee, K. (2007). White lie-telling in children for politeness purposes. *International journal of behavioral development*, 31(1):1–11.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285.
- Yoon, E. J., Tessler, M. H., Goodman, N. D., and Frank, M. C. (2016). Talking with tact: Polite language as a balance between kindness and informativity. In *Proceedings of the Thirty-Eighth Annual Conference of the Cognitive Science Society*.
- Yoon, E. J., Tessler, M. H., Goodman, N. D., and Frank, M. C. (2017). “i won't lie, it wasn't amazing”: Modeling polite indirect speech. Manuscript submitted for publication.