Nonliteral language understanding for number words

Justine T. Kao [*], **Jean Wu,** [*] **Leon Bergen** [†] **and Noah D. Goodman** [*]

[*]Stanford University, and [†]MIT

Submitted to Proceedings of the National Academy of Sciences of the United States of America

One of the most puzzling and important facts about communication is that people do not always mean what they say; speakers often use imprecise, exaggerated, or otherwise literally false descriptions to communicate experiences and opinions. Here we focus on the nonliteral interpretation of number words, in particular hyperbole (interpreting unlikely numbers as exaggerated and conveying affect) and pragmatic halo (interpreting round numbers imprecisely). We provide a computational model of number interpretation as social inference regarding the communicative goal, meaning, and affective subtext of an utterance. We show that our model predicts humans' interpretation of number words with high accuracy. Our model is the first computational model that quantitatively predicts a range of nonliteral effects in number interpretation, and our modeling framework provides a theory of nonliteral language understanding more generally.

Pragmatics | Language understanding | Computational modeling

Introduction

Imagine a friend describing a new restaurant where she recently dined. Your friend says, "It took 30 minutes to get a table." You are likely to interpret this to mean she waited approximately 30 minutes. Suppose she says: "It took 32 minutes to get a table." You are more likely to interpret this to mean exactly 32 minutes. Now, suppose she says: "It took a million years to get a table." You will probably interpret this to mean that the wait was shorter than a million years, but importantly that she thinks it took much too long. One of the most fascinating facts about communication is that people do not always mean what they say—a crucial part of a listener's job is to understand an utterance even when its literal meaning is false. People's ability to interpret nonliteral language poses a critical puzzle for research on language understanding.

A rich body of literature in psychology and linguistics has examined how people use and understand nonliteral language (1-4). However, much of the previous literature has been qualitative, with little focus on analyzing aspects of an utterance that predict the quantitative details of people's figurative interpretations. Here we present a model that formalizes and integrates three general principals of language and communication to explain the computational basis of nonliteral language understanding. First, speakers and listeners communicate with the assumption that their interlocutors are rational and cooperative agents; second, speaker and listener utilize common ground—their shared knowledge of the world—to communicate effectively; third, listeners assume that speakers choose utterances to maximize informativeness and relevance with respect to their communicative goals. The ideas we propose here have connections to Gricean pragmatics and relevance theory. Rather than view nonliteral language as a separate mode of communication that requires specialized language processing strategies, several scholars have argued that basic principles of communication drive the meaning that a listener infers from a figurative utterance (Sperber & Wilson, 2008; Wilson & Carston, 2006). Relevance theory, in particular, posits that listeners interpret utterances with the assumption that speakers produced them because they are maximally relevant. Proponents of the theory argue that this principle of relevance explains how listeners infer the meaning of a metaphor as well as other forms of loose talk where the meaning of an utterance is underspecified (Sperber & Wilson, 1985; Wilson & Sperber, 2002). By formalizing a notion of informativeness and relevance in a computational model and applying it to a case study on number words, we show that nonliteral interpretations can arise from these components and their interactions alone without positing dedicated processing mechanisms for nonliteral language.

A recent body of work has formalized communication as an interaction between rational and cooperative agents. These models view pragmatic language understanding as probabilistic inference over recursive social models and are able to quantitatively explain a range of phenomena in human pragmatic reasoning (7, 8, 9, 10). At the core of these models, a listener and a speaker recursively reason about each other to arrive at pragmatically enriched meanings. Given an intended meaning m, speaker S_n reasons about listener L_{n-1} and chooses utterance u based on the probability that the listener will successfully infer the intended meaning (9):

$$S_n(u|m) \propto L_{n-1}(m|u) \cdot e^{-C(u)}$$

The listener L_n then reasons about S_n and uses Bayes Rule to infer the meaning m given utterance u:

$$L_n(m|u) \propto P(m)S_n(u|m)$$

This framework predicts that it is never optimal for a speaker to choose an utterance whose literal meaning directly contradicts her intended meaning. However, this is precisely the case in nonliteral language. For example, "Juliet is the sun conveys that Juliet is a beautiful woman and not, in fact, the sun, and "It took a million years to get a table" conveys that the wait time was long but not, in fact, a million years. This suggests that the basic model is incomplete and requires additional elements to explain nonliteral communication.

Previous work has examined people's reasons for using figurative language and suggested that certain discourse goals, such as conveying emotion and emphasis, are best satisfied by nonliteral language (1). Here we propose that language understanding in general, and nonliteral language understanding in particular, relies on reasoning about communicative goals during interpretation. We introduce a model in which the listener is uncertain about the speaker's communicative goal

Reserved for Publication Footnotes

and performs joint inference on both the goal and the intended meaning. Importantly, the interpretation space has multiple dimensions, and different communicative goals are satisfied by different aspects of the inferred meaning. A speaker's goal may be to maximize the probability of successfully conveying information along one dimension of meaning but not another, which makes it possible for a literally false utterance to be optimal as long as it is informative along the target dimension. We explore the case where the interpretation space has two dimensions: the state of the world and the speaker's affect. The speaker is now modeled as

$$S_n(u|g) \propto \sum_{s,a} L_n(s,a|u) g(s,a) \cdot e^{-C(u)}$$

where the intended meaning includes s (the state of the world) and a (the speakers affect). g is a function that denotes whether a communicative goal is satisfied by s and a, and C is a function for utterance costs (see Methods for details). The listener then performs joint inference on both the goal and the meaning:

$$L_n(s,a|u) \propto \sum_{g} P_S(s) P_A(a|s) P_G(g|s,a) S_{n-1}(u|g)$$

Based on this equation, the listener's model utilizes prior knowledge such as the probability of a state P_S and the probability of having a particular affect given a state P_A. The affective subtext a that a listener infers from an utterance arises from the common knowledge that certain kinds of affect are usually associated with certain states of the world. Common knowledge of a domain thus allows the speaker to communicate additional dimensions of meaning such as affect without explicitly describing them.

This formulation of language understanding as joint inference of the communicative goal, state of the world, and affective subtext of an utterance provides a computational model of nonliteral language understanding. As a case study, we focus on the nonliteral interpretation of number words. We chose number words because they have precise literal meanings that can be easily modeled, whereas the literal meanings of concepts such as "Juliet" and "the sun" are more difficult to formalize. We aim to capture two particular well-known phenomena regarding number interpretation: hyperbole and pragmatic halo. Hyperbole is a figure of speech that uses exaggeration to convey emphasis and emotion. While hyperbolic utterances are literally false, such indirect communication is readily understood and serves many purposes (1, 11-13). Pragmatic halo refers to people's tendency to interpret round numbers such as 100 imprecisely and complex numbers such as 103 precisely (14). While this effect has been formalized via game theory as a rational choice given different utterance costs and a notion of pragmatic slack (15, 16), other research has shown that speakers' tendency to choose simple number expressions decreases when more precise information is relevant to the listener. This suggests that higher-level pragmatic considerations such as communicative goals directly impact the production and interpretation of round versus sharp numbers (17). As a result, our model uses alternative communicative goals coupled with differential utterance costs to model the pragmatic halo effect. We show that our framework for pragmatic inference makes quantitative predictions for both hyperbole and pragmatic halo in the nonliteral interpretation of number words.

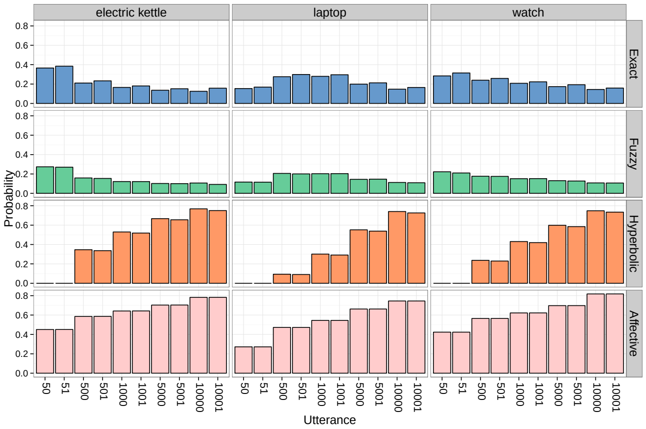

Fig. 1. Each vertical panel column shows the probabilities of different kinds of interpretations given utterances about an item (see text).

Results

We tested our model on number words referring to the prices of three types of everyday items: electric kettles, watches, and laptops. We selected these items because they have distinct price distributions, which we measured empirically by asking participants to rate the probability of various prices for the three items (see Experiment 3a in Methods). We also obtained an affect prior by asking participants to rate the probability of a speaker thinking that an item is too expensive given a price state (see Experiment 3b). A speaker can say, "The electric kettle cost u dollars," for $u \in U$, and a listener can interpret this to mean that the kettle cost s dollars, for $s \in S$. Each utterance is either "round" (divisible by 10 and less costly to utter) or "sharp" (not divisible by 10 and more costly to utter). A formal description of these assumptions is in the Methods section.

Model simulations. Using the price priors and affect priors measured for each of the three items, we obtained the full posterior meaning distribution predicted by the model for each utterance (see Figure S1). Figure 1 summarizes this distribution into different types of interpretations. The first three are model interpretations regarding the price state: exact (e.g., "1000" interpreted as 1000), fuzzy (e.g. "1000" interpreted as 1001 or "1001" interpreted as 1000), and hyperbolic (e.g. "1000" interpreted as 100). Utterances whose literal meanings are less likely given the price prior are more likely to be interpreted hyperbolically (e.g. "1000" is more likely to be interpreted hyperbolically for electric kettles than laptops), which shows the model captures a basic feature of hyperbole. Round utterances such as "500" and "1000" are interpreted less exactly and more fuzzily than their sharp counterparts, which shows the model captures pragmatic halo. On the affect dimension, affective interpretation refers to the probability that an utterance conveys the speaker's opinion that the price is expensive. Utterances whose literal meanings are associated with higher affect priors (such as "10000" and "10001") are more likely to be interpreted as conveying affect, which shows the model predicts the affective subtext of hyperbole.

To build intuition for these predictions, consider a pragmatic listener who recursively reasons about a speaker and analyzes her choice of utterance. The pragmatic listener hears "10,000 dollars" and knows its literal meaning is extremely unlikely. However, given that the speaker reasons about a literal listener who interprets "10,000 dollars" literally and believes that the speaker very likely thinks it is expensive,

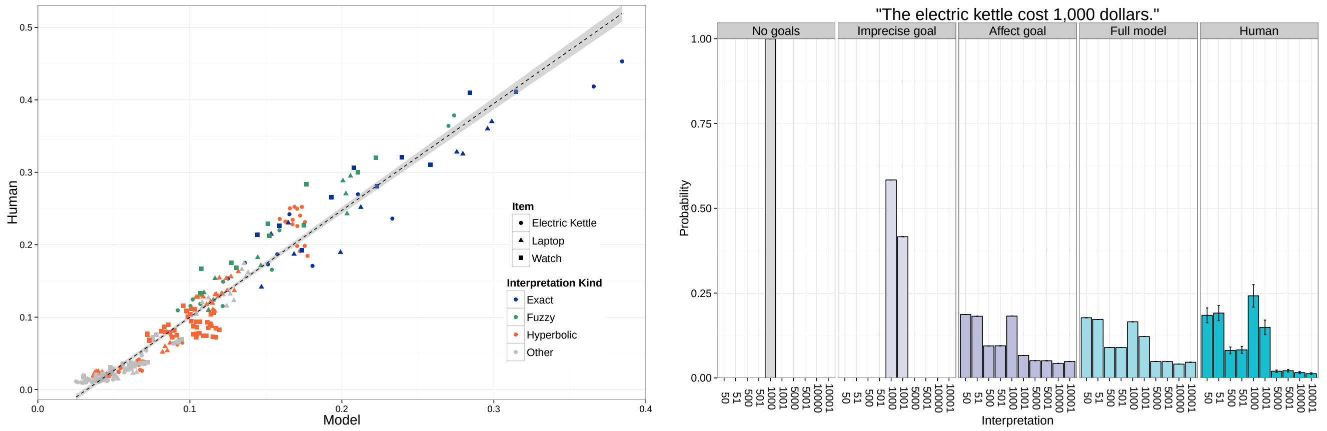

Fig. 2. (A) Model predictions (x-axis) versus average human responses (y-axis) for 300 data points (3 Items × 10 Utterances × 10 Price States) in Experiment 1. (B) Human interpretations of a sample utterance and model predictions given different communicative goals. A model that considers both affect and precision goals closely matches human data.

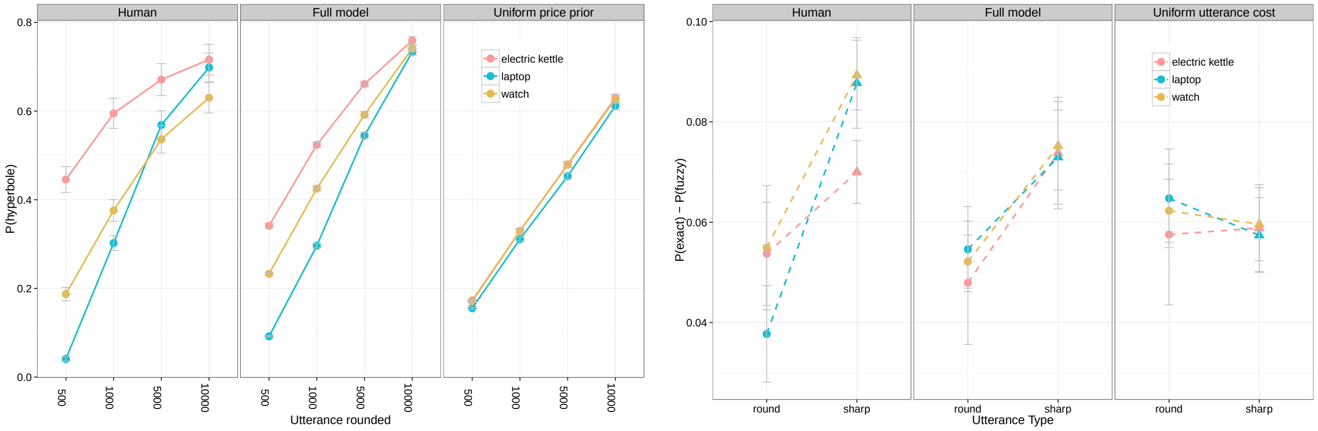

Fig. 3. (A) Probability of hyperbolic interpretation across utterances and items. The leftmost panel shows human data (error bars are standard errors). A full model that uses empirical price priors matches human data; a model that uses uniform price priors does not distinguish among item types and shows weaker hyperbole effects. (B) Bias for exact interpretation for round/sharp utterance types. Humans have a bias for exact interpretations of sharp utterances. A full model that assigns higher costs to sharp numbers matches human data; a model that uses uniform utterance costs does not.

"10,000 dollars" is an optimally informative utterance if the speakers goal is to communicate that the kettle is expensive (without concern for the actual price). Since the pragmatic listener uses this information to perform joint inference on the speaker's communicative goal and the meaning of the utterance, he infers that "10,000 dollars" is likely to mean less than 10,000 dollars but that the speaker thinks it is too expensive (i.e., strong affect).

Behavioral experiments. We conducted Experiment 1 to evaluate the model's predictions for the interpreted price. Participants read scenarios in which a buyer produces an utterance about the price of an item he bought, for example: "The electric kettle cost 1000 dollars." Participants then rate the likelihood that the item actually cost s dollars for $s \in S$ (see Experiment 1 in Methods). Figure S2 shows humans' interpretation distributions across all utterances. Participants were more likely to interpret utterances as hyperbolic when their literal meanings have lower probabilities under the items prior price distribution ($F(1, 10) = 44.06, p < 0.0001$). To examine the halo effect, we computed the difference between the probability of an exact interpretation and the probability of a fuzzy interpretation for each utterance. This difference is significantly smaller for round numbers than for sharp numbers ($F(1, 28) = 18.94, p < 0.001$), which indicates that round numbers tend to be interpreted less precisely than sharp numbers. These results match the model's qualitative predictions for hyperbole and halo. To quantitatively evaluate the model's fit, we compared model and human interpretation probabilities across all utterances and showed that model predictions are highly correlated with human interpretations of number words ($r = 0.974, p < 0.0001$) (Figure 2(A)).

We explore simpler comparison models to show that each component of the proposed model is responsible for capturing effects observed in the human data. Figure 2(B) compares model interpretations of the utterance "The electric kettle cost 1,000 dollars" given considerations of different communicative goals. A model that does not consider alternative communicative goals interprets the utterance entirely literally. A model that considers a speaker whose goal may be to communicate precisely or imprecisely interprets the utterance as meaning either 1000 or 1001. A model that considers a speaker whose goal may be to communicate the precise price state or her affect prefers price states with higher prior probabilities. Finally, a model that considers the full range of goals produces

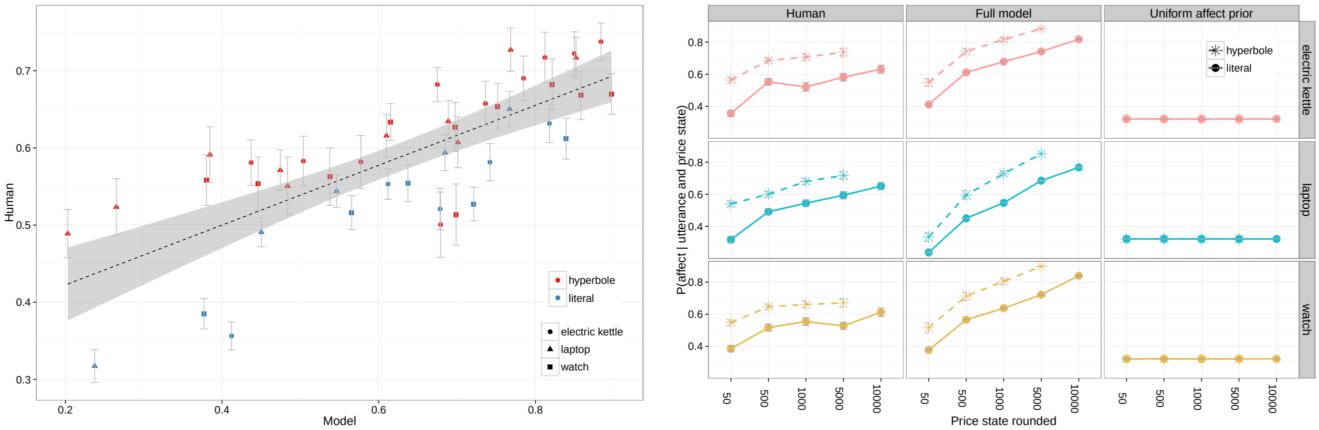

Fig. 4. (A) Model predictions of affect (x-axis) versus human responses (y-axis) for 45 data points (3 Items × 15 Utterance-Price state pairs where $u \geq s$) in Experiment 2. (B) Probability of inferring affect given a price state and a hyperbolic or literal utterance. Humans infer higher probability of affect given higher price states and higher affect given hyperbolic utterances. A full model that uses empirical affect priors matches human data; a model that uses uniform affect priors predicts neither affect across price states or the rhetorical effect of hyperbole.

interpretations that demonstrate hyperbole and halo effects that closely match humans' interpretations. This suggests that reasoning about a speaker's communicative goals is crucial for the nonliteral interpretation of number words. Figure 3(A) shows probabilities of an utterance being interpreted hyperbolically by humans, the full model, and a version of the model that takes a uniform price prior for each item type. The full model faithfully captures the human data, while the "lesioned" model fails to differentiate among hyperbole effects for the three item domains. This confirms that people use their knowledge of a domains prior distribution to infer hyperbolic interpretations instead of the semantics of the number words alone. Figure 3(B) shows the halo effect in humans, the full model that assigns higher utterance costs to sharp numbers, and a version of the model where the costs of utterances are uniform. The full model replicates humans' pragmatic halo effect, while the simpler model does not. This suggests that people consider utterance costs and communicative efficiency when inferring exact versus fuzzy interpretations.

Does the model capture the rhetorical effect of hyperbole? We conducted Experiment 2 to examine humans' inference of affect in hyperbolic versus literal utterances. Participants read scenarios in which a speaker bought an item that cost s dollars and says it cost u dollars, where $u \geq s$. They then rate how likely it is that the buyer thinks the item was too expensive (see Experiment 2 in Supplementary Materials). We focused on the affect of an item being too expensive due to previous findings suggesting that hyperbole is more often used to communicate negative attitudes and emotions (1, 11). Results showed that utterances u where $u > s$ are rated as significantly more likely to convey affect than utterances where $u = s$ ($F(1, 25) = 12.57, p < 0.005$). This confirms the hypothesis that listeners infer affective subtext from hyperbolic utterances. Quantitatively, we compared model and human interpretations of affect for each of the 45 items where $u \geq s$. While there is a significant amount of noise in the human judgments (average split-half correlation is 0.833), the model predicts human interpretations of the utterances affective subtext significantly better than chance ($r = 0.772, p < 0.00001$), capturing most of the reliable variation in these data (Figure 4(A)). Figure 4(B) shows probabilities of inferring affect given a price state and a literal or hyperbolic utterance for humans, the full model, and a version of the model that takes in uniform affect priors. The human data shows that higher actual price states are associated with higher probability of affect. Within the same price state, hyperbolic utterances are interpreted as conveying more affect than literal utterances. Both effects are replicated by the full model, but not by the "lesioned" model. This shows that the rhetorical effect of hyperbole is driven in part by prior knowledge of affect associated with different prices.

Discussion

We have presented the first computational model of nonliteral language understanding that quantitatively predicts people's hyperbolic and imprecise interpretations of number words. Our model and behavioral results show that complex patterns in nonliteral number interpretation depend on common ground between speaker and listener, consideration of communicative efficiency, and reasoning about relevance to the speaker's communicative goal.

The model presented here is intended to give a computational account of how people utilize prior knowledge and pragmatic reasoning to arrive at potentially nonliteral interpretations; it does not serve to predict process-level details, such as whether literal interpretations must be considered before they are rejected in favor of nonliteral interpretations. Instead, our goal was to show that formalizations of basic communicative principles such as informativeness and relevance can explain nonliteral language understanding as well as its rhetorical effects. We were able to examine nonliteral language at a fine-grained level and understand how the quantitative details of an utterance in context predict specific interpretations of a number word. Our model's predictions closely match humans' judgments of hyperbole, a complex phenomenon previously beyond the scope of computational models. By capturing the communicative goals in other figures of speech such as irony and metaphor, we hope to extend our model to explain nonliteral language understanding more broadly. We believe that this framework significantly advances the flexibility and richness of formal models of language understanding, such that some day probabilistic models will explain *everything* (hyperbolically speaking).

Materials and Methods

Model. Here we describe our model in detail. Let u be the utterance a speaker utters. The meaning of the utterance has two dimensions, one concerning the actual price state s, and one concerning the speaker's affect a. We defined the set of price states $S = \{50, 51, 500, 501, 1000, 1001, 5000, 5001, 10000, 10001\}$ and assumed that the set of utterances U is identical to S. We defined the set of affect states $A = \{0, 1\}$ (0 means no affect and 1 means with affect). Given the set of price states S and set of affect states A, the set of possible meanings M is given by $M = S \times A$. We denote each possible meaning as s, a, where $s \in S$ and $a \in A$.

Let g be the communicative goal, which also has two dimensions, one concerning the price state, and the other concerning the speaker's affect. We denote each communicative goal as $g_{\mathbf{s},\mathbf{a}}$ where $\mathbf{s} \in 2^S$ is an equivalence class of price states and $\mathbf{a} \in 2^A$ is an equivalence class of affect states. These equivalence classes represent states of the world that are sufficiently close to the true state of the world, for the purposes of the speaker. Formally, the goal $g_{\mathbf{s},\mathbf{a}}$ is a function $g_{\mathbf{s},\mathbf{a}} : M \rightarrow \{0, 1\}$, such that $g_{\mathbf{s},\mathbf{a}}(m_{s,a}) = 1$ if and only if $s \in \mathbf{s}, a \in \mathbf{a}$. Thus, a meaning satisfies this goal if it belongs to the state and affect equivalence classes of the goal. We assume that there are two types of price-related goals: the speaker either wants to communicate the price state exactly or approximately. Exact goals are represented by subsets that consist of a single price state, i.e. $\mathbf{s} = \{i\}$ (for some $i \in S$), and approximate goals are represented by subsets that consist of the price states within a distance of 1 of some state, i.e. $\mathbf{s} = \{j | j \in S, |j - i| \leq 1\}$. Subsets of price states which do not satisfy either of these conditions are assigned probability 0 by the model.

The prior probability of a price state s is taken from an empirically derived price prior $P_S(s)$, and the probability of an affect a given a price state s is taken from an empirically derived conditional affect prior $P_A(a|s)$ (see Experiments 3a and 3b). The probability distribution $P_G(\cdot|s, a)$ over goals given that the speaker knows meaning s, a is defined to be uniform over goals consistent with s, a, i.e. uniform over goals $g_{\mathbf{s},\mathbf{a}}$ such that $g_{\mathbf{s},\mathbf{a}}(\mathbf{s}, \mathbf{a}) = 1$. This is equivalent to assuming that the speaker either wants to communicate her meaning exactly or approximately.

A literal listener L_0 provides the base case for recursive social reasoning between the speaker and listener. L_0 interprets an utterance u literally without taking into account the speaker's communicative goals:

$$L_0(s, a|u) = \begin{cases} P_A(a|s) & \text{if } s = u \\ 0 & \text{otherwise} \end{cases}$$

The speaker S_n is assumed to be a rational planner who optimizes the probability that the listener will infer a meaning m that satisfies her communicative goal while minimizing the cost of her utterance. S_n chooses utterances according to a softmax decision rule that describes an approximately rational planner [17]:

$$S_n(u|g_{\mathbf{s},\mathbf{a}}) \propto e^{U_n(u|g_{\mathbf{s},\mathbf{a}})} \qquad [1]$$

Optimizing the probability of the speaker's goal being satisfied can be accomplished by minimizing the goal's information-theoretic surprisal. Given an utterance u, the listener L_n will guess that the meaning is s, a with probability $L_n(s, a|u)$. The probability of the speaker's goal being satisfied is therefore the following:

$$\sum_{s,a} L_n(s, a|u) g_{\mathbf{s},\mathbf{a}}(s, a) \qquad [2]$$

The utility function U_n is composed of both the negative surprisal of the goal and the negative of the utterance cost $C(u)$. U_n is therefore defined by:

$$U_n(u|g_{\mathbf{s},\mathbf{a}}) = \log(\sum_{s,a} L_n(s, a|u) g_{\mathbf{s},\mathbf{a}}(s, a)) - C(u) \qquad [3]$$

Combined with equation 2, this leads to:

$$S_n(u|m, g) \propto \sum_{s,a} L_n(s, a|u) g_{\mathbf{s},\mathbf{a}}(s, a) \cdot e^{-c(u)} \qquad [4]$$

We used $C(u) = 1$ when u is a round number and $C(u) = 1.8$ when u is a sharp number for all model simulations reported.

The listener L_n performs Bayesian inference to guess the intended meaning given the prior P and his internal model of the speaker. To determine the speaker's intended meaning, the listener will marginalize over the possible goals under consideration.

$$L_n(s, a|u) \propto \sum_g P_S(s) P_A(a|s) P_G(g|s, a) S_{n-1}(u|g). \qquad [5]$$

We obtained a posterior distribution for all possible meanings s, a given an utterance u. Raw data for model predictions are here: . Figure S1 shows the full posterior distributions for all utterances.

Experiment 1a: Halo and hyperbole. 120 participants were recruited on Amazon's Mechanical Turk. We restricted participants to those with IP addresses in the United States. Each participant read 15 scenarios in which a person (e.g. Bob) buys an item (e.g. a watch) and is asked by a friend whether the item is expensive. We randomized the order of the trials as well as the names of the buyers. Bob responds by saying "It cost u dollars," where $u \in \{50, 50 \pm k, 500, 500 \pm k, 1000, 1000 \pm k, 5000, 5000 \pm k, 10000, 10000 \pm k\}$, where k was randomly selected from the set $\{1, 2, 3\}$ for each trial. We will refer to this set of utterances as U. Numbers devisable by 10 are considered "round" numbers, while numbers not devisable by 10 are "sharp" numbers.

Given an utterance u, participants rated the probability of Bob thinking that the item was expensive. They then rated the probability of the item costing the following amounts of money: $50, 50 \pm k, 500, 500 \pm k, 1000, 1000 \pm k, 5000, 5000 \pm k, 10000, 10000 \pm k$, where k was randomly selected from the set $\{1, 2, 3\}$ for each trial. We will refer to this set of prices as S. Ratings for each price state were on a continuous scale from "impossible" to "extremely likely," represented as real values between 0 and 1. There are a total of 30 possible trial configurations (3 Items × 10 Utterances). The stimuli for Experiment 1 can be found here: .

We normalized participants' ratings across price states for each trial to sum up to 1. There are a total of 300 normalized average ratings (3 Items × 10 Utterances × 10 Price States). The average normalized ratings across participants for each item/utterance pair is shown in Figure S2. The raw ratings can be found here: , and the normalized ratings are here: . To adjust for humans' biases against using the extreme ends of the slider bars, we performed a Luce choice transformation on the models distribution. We multiplied the models predicted probability for each meaning by a free parameter λ and renormalized the probabilities to sum up to 1 for each utterance. Fitting the λ to the behavioral data to optimize correlation, we obtained the best fit with $\lambda = 0.34$, resulting in a correlation between model predictions and participant ratings of $r = 0.974$ (see main text). All figures and analyses that we report in the main text are with this transformation. Without Luce choice transformation and with no free parameters in the model, correlation between model predictions and participant ratings is still very high ($r = 0.907$).

For the analysis reported in Figure 3(A), we computed the probability of a participant interpreting an utterance u as hyperbolic by summing up his or her probability ratings for each interpreted price state s, where $u > s$. Since our analysis of hyperbole does not involve utterance costs, we collapsed across round and sharp versions of utterances and price states. For example, "1001" interpreted as 1000 does not count as hyperbole. Since 50 and 51 are the lowest available price states, the probabilities for hyperbolic interpretation of utterances "50" and "51" are 0. We computed the average probability of a hyperbolic interpretation across subjects for each utterance. We then showed the hyperbole effect by building a linear regression model with prior probabilities for the utterances' literal meanings as predictor and the probabilities for hyperbolic interpretation as response. Results indicated that participants were more likely to interpret utterances as hyperbolic when their literal meanings have lower probabilities under the items prior price distribution ($F(1, 10) = 44.06, p < 0.0001$).

For the analysis reported in Figure 3(B), we analyzed the pragmatic halo effect by computing each subjects bias for interpreting an utterance u exactly ("1000" interpreted as 1000) versus fuzzily ("1000" interpreted as 1001). Bias was measured by subtracting the probability of a fuzzy interpretation from the probability of an exact interpretation. We then obtained the average bias for each utterance across subjects. We showed that the average bias for exact interpretation is significantly higher for sharp utterances than for round utterances ($F(1, 28) = 18.94, p < 0.001$).

Experiment 1b: Affective subtext. 160 participants were recruited on Amazon's Mechanical Turk. We restricted participants to those with IP addresses in the United States. Each participant read 30 scenarios in which a person (e.g. Bob) buys an item that costs s dollars and is asked by a friend whether the item is expensive. We randomized the order of the trials as well as the names of the buyers. Bob responds by saying "It cost u dollars," where $u \in U$ and $u \geq s$. Participants then rated how likely Bob thinks the item was expensive on a continuous scale ranging from "impossible" to "absolutely certain," represented as real values between 0 and 1. There are a total of 180 trial configurations (3 Items × 60 $\{u, s\}$ pairs where $u \geq s$). The stimuli for Experiment 2 can be found here: ; the raw data is here: .

Since our analysis of affective subtext does not involve utterance cost, for the analyses reported in Figure 4(A) and 4(B), we collapsed round and sharp versions of each utterance and price state such that there are a total of 45 combinations of utterances and price states under consideration. Utterances u for which $u = s$ are considered literal; utterances u for which $u > s$ are hyperbolic. For the analysis reported in Figure 4(B), we obtained average ratings of affect for each utterance given that it is literal or hyperbolic. A linear regression model showed that hyperbolic utterances are rated as having significantly higher affect than literal utterances across price states ($F(1, 25) = 12.57, p < 0.005$).

Experiment 2a: Price prior. To obtain peoples prior knowledge of the price distributions for electric kettles, laptops, and watches, 30 participants were recruited from Amazon's Mechanical Turk. We restricted participants to those with IP addresses in the United States. Each participant rated the probability of an electric kettle, laptop, and watch costing s dollars, where $s \in S$. We randomized the order of the trials as well as the names of the buyers. Ratings for each price state were on a continuous scale from "impossible" to "extremely likely," represented as real values between 0 and 1. The stimuli for Experiment 3a can be found here: . We normalized participants' ratings across price points for each trial to sum up to 1. The average normalized ratings across participants for each item were taken as the prior probability distribution of item prices. These price distributions were used in the model to determine the prior probability of each price state. The normalized ratings can be found here:

Experiment 2b: Affect prior. To obtain people's prior knowledge of the affect likelihood given a price state, 30 participants were recruited from Amazon's Mechanical Turk. We restricted participants to those with IP addresses in the United States. Each participant read 15 scenarios where someone had just bought an item that cost s dollars ($s \in S$). We randomized the order of the trials. They then rated how likely the buyer thinks the item was expensive on a continuous scale ranging from "impossible" to "absolutely certain," represented as real values between 0 and 1. The stimuli for Experiment 3b can be found here: . The average ratings for each item/price state pair were taken as the prior probability of an affect given a price state. This was used in the model to determine the prior probability of an affect given each price state. The data can be found here:

Appendix: Appendix title

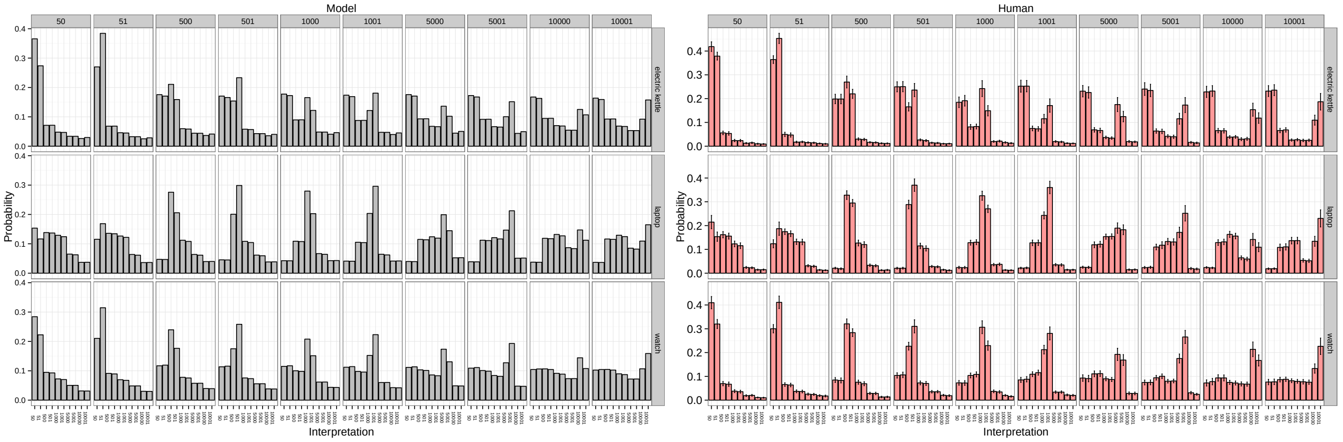

Fig. 5. (A)Full posterior meaning distribution predicted by the model for each utterance. Each column of panels is an utterance, and each row of panels is an item type. Each panel represents the interpretation distribution given an utterance for an item. (B) Full meaning distribution produced by humans for each utterance. Each column of panels is an utterance, and each row is an item type. Each panel represents the interpretation distribution given an utterance for an item. Error bars are standard errors.

ACKNOWLEDGMENTS. Thanks!

1. Grice, H.P., Logic and conversation, (1975), pp. 41–58
2. Clark, H.H., Using language, (1996), Cambridge University Press Cambridge
3. Frank, M.C. and Goodman, N.D., Predicting pragmatic reasoning in language games, Science, 336(6048), (2012), pp. 998
4. Goodman, N.D. and Stuhlmüller, A. Knowledge and implicature: Modeling language understanding as social cognition, Proceedings of CogSci conference, (2012)
5. Bergen, L. and Goodman, G.D. and Levy, R., That's what she (could have) said: How alternative utterances affect language use, Proceedings of CogSci conference, (2012)
6. Jäger, G. and Ebert, C., Pragmatic rationalizability, Proceedings of Sinn und Bedeutung, 13, (2009), pp. 1–15
7. Lasersohn, P., Pragmatic halos, Language, (1999), pp. 522–551
8. Bastiaanse, H., The rationality of round interpretation, Vagueness in communication, (2011), pp. 37–50
9. Krifka, M., Approximate interpretation of number words: A case for strategic communication, Cognitive foundations of interpretation, (2007), pp. 111–126
10. McCarthy, M. and Carter, R, There's millions of them: hyperbole in everyday conversation, Journal of pragmatics, 36(2), (2004), pp. 149–184.
11. Gibbs, R.W., Irony in talk among friends, Metaphor and symbol, (2000), pp. 5–27
12. Cano Mora, L., At the Risk of Exaggerating: How Do Listeners React to Hyperbole?, Anglogermanica online: Revista electrónica periódica de filología alemana e inglesa, (2003), pp. 2–13
13. Kreuz, R.J. and Caucci, G.M., Lexical influences on the perception of sarcasm, Proceedings of the Workshop on Computational Approaches to Figurative Language, (2007), pp. 1–4, Association for Computational Linguistics
14. Davidov, D. and Tsur, O. and Rappoport, A., Semi-supervised recognition of sarcastic sentences in twitter and amazon, Proceedings of the Fourteenth Conference on Computational Natural Language Learning, (2010), pp. 107–116, Association for Computational Linguistics
15. Reyes, A. and Rosso, P., Mining subjective knowledge from customer reviews: a specific case of irony detection, ACL HLT 2011, (2011), pp. 118

16. van Kruijsdijk, R., Algorithm Development in Computerized Detection of Sarcasm using Vocal Cues, (2007)

17. Sutton, R.S. and Barto, A.G., Reinforcement learning: An introduction, 28, (1998)