Nonliteral understanding of 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 hyperbolic and pragmatic halo effects in number interpretation using a unified framework. This modeling framework provides an approach to 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 the 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, 2, 3, 4]. However, much of the work 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, listeners assume that speakers choose utterances to maximize informativeness with respect to their communicative goals; third, speaker and listener utilize common ground—their shared knowledge of the world—to communicate effectively. These ideas have important connections to Gricean pragmatics [5, 6] and relevance theory [7, 8]. For instance, proponents of relevance theory argue that listeners infer the meaning of a metaphor, as well as other forms of loose talk, by assuming that speakers maximize relevance [9, 10, 11]. Here we formalize the notion that speakers have a goal to maximize informativeness about a given topic. By applying this computational approach to a case study on number words, we show that nonliteral interpretations can arise from principles of communication 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 Rational Speech Act (RSA) 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 [12, 13, 14, 15]. 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 [14]:

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

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

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

The RSA framework predicts that it is never optimal for a speaker to choose an utterance whose literal meaning directly contradicts her intended meaning. However, this contradictory use is precisely the case in nonliteral language. For example, "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 RSA model is incomplete and requires additional elements to explain nonliteral communication.

Previous work has examined people's communicative reasons for using figurative language and suggested that certain 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 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 (e.g. the true amount) and the speaker's affect or opinion[1]. The speaker

Reserved for Publication Footnotes

[1]In what follows we describe the subtext dimension as "affect," but it could be other kinds of speaker opinion, *mutatis mutandis*.

is now modeled as:

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

where the intended meaning includes s (the state of the world) and a (the speaker's affect), and $C(u)$ is the utterance's cost. The function g projects the listener's inferred meaning onto relevant dimensions: the speaker's communicative goal is to be informative (only) along this "topic" dimension. A literal listener interprets utterances literally without reasoning about the speaker, while a pragmatic listener performs joint inference on both the speaker's goal and her intended meaning (see Materials):

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

The listener utilizes prior knowledge of the probability of a state (P_S) and the probability of having a particular affect given a state (P_A). The literal meaning of utterance u may convey nothing about the affective subtext a; it conveys information about the state s. However, the common knowledge that affect is usually associated with certain states of the world allows the speaker to use an assertion that the state is s to convey the corresponding information about a. If it is known that the goal is to convey affect, and not the state, then the pragmatic listener will discount the information about s but retain the information about a—a nonliteral interpretation is obtained. Even when the pragmatic listener is not certain of the speaker's goal, a joint inference of goal, state, and affect results in the same nonliteral interpretation. Common knowledge of a domain and joint reasoning about communicative goals thus allows the speaker to communicate additional dimensions of meaning, such as affect, without explicitly describing these dimensions.

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 approach extending the RSA framework to 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, and apply to domains (such as prices) that lend themselves to quantitative measurement. 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 [16]. While hyperbolic utterances are literally false, such indirect communication is readily understood and serves many purposes [1, 16, 17, 18]. Pragmatic halo refers to people's tendency to interpret round numbers such as 100 imprecisely and complex numbers such as 103 precisely [19]. The halo effect has been formalized in game theoretic models as a rational choice given different utterance costs and a possibility of pragmatic slack [20, 21]. Other research has shown that speakers' tendency to choose simple number expressions decreases when more precise information is relevant to the listener [22]. This suggests that higher-level pragmatic considerations such as communicative goals directly impact the production and interpretation of round versus sharp numbers. 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 interpretation of number words.

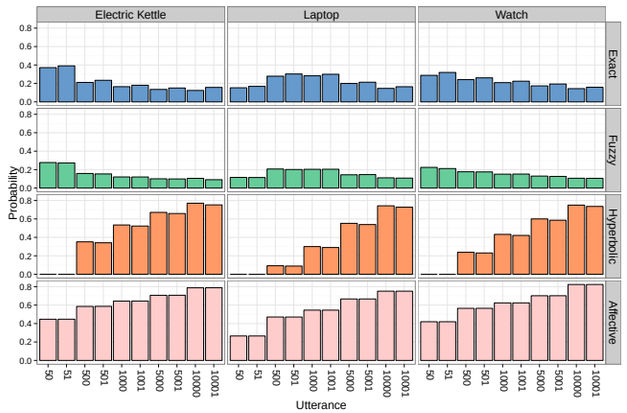

Fig. 1. Each vertical panel column shows the probabilities of different kinds of interpretations given utterances about an item (see text).

Results

We tested our model on number words that refer to the prices of three types of everyday items: electric kettles, watches, and laptops. We selected these items because they have distinct price distributions, the prior P_S, 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, P_A, by asking participants to rate the probability of a speaker thinking that an item is too expensive given a price state (see Experiment 3b). Given these priors, we aimed to model people's interpretations of utterances such as, "The electric kettle cost u dollars," for $u \in U$, assuming that a listener can interpret this utterance to mean that the kettle cost s dollars, for $s \in S$. We selected U such that each number word 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 posterior meaning distribution predicted by the model for each utterance (see Figure 5a). 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), 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 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 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—predicting the affective subtext of hyperbole.

To build intuition for these predictions, consider a pragmatic listener who reasons about a speaker and analyzes her choice of utterance. The pragmatic listener hears "10,000 dollars" and knows that 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

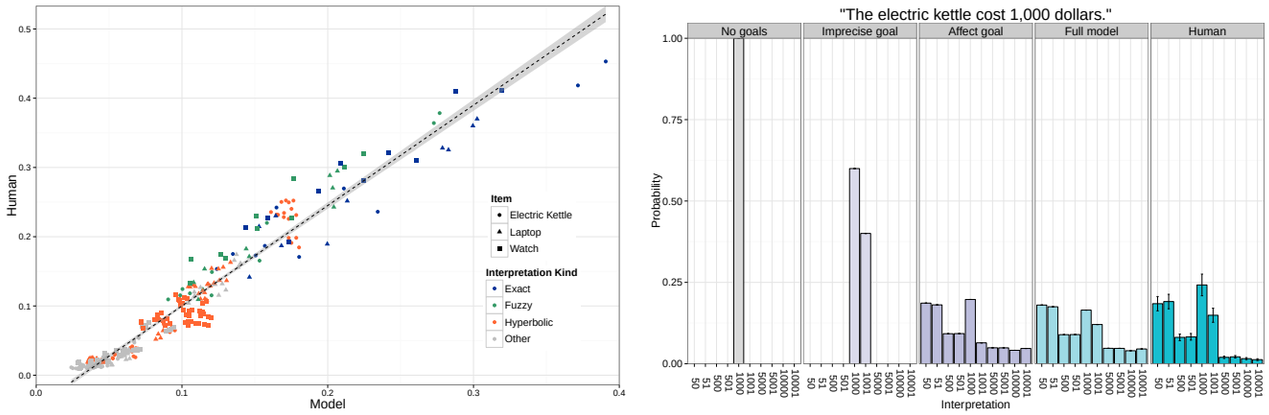

Fig. 2. (a) Model predictions (x-axis) v.s. average human responses (y-axis) for 300 data points (3 Items × 10 Utterances × 10 Price States) in Experiment 1. (b) Human interpretations of an utterance and model predictions given different communicative goals. A model that considers both affect and precision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

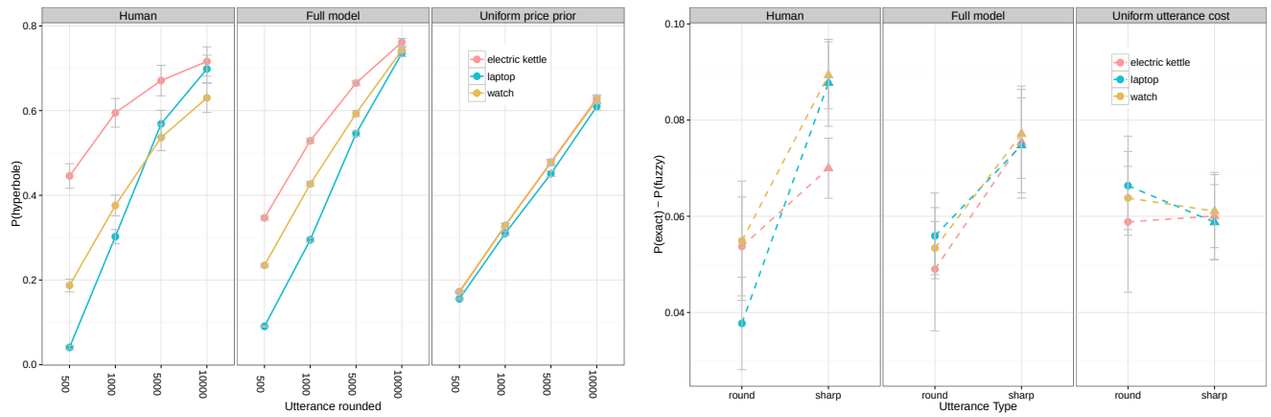

Fig. 3. (a) Probability of hyperbolic interpretation across utterances and items. The leftmost panel shows human data (error bars are standard errors). A full model that uses empirical price priors matches human data; a model that uses uniform price priors does not distinguish among item types and shows 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.

speaker very likely thinks it is expensive, "10,000 dollars" is an informative utterance if the speaker's 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 5b shows humans' interpretation distributions across all utterances. We first test the model's qualitative predictions for hyperbole and halo: Participants were more likely to interpret utterances as hyperbolic when their literal meanings have lower probabilities under the item's 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. 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 ($r = 0.973, p < 0.0001$) (Figure 2a).

To show how each component of the proposed model is necessary for capturing effects observed in the human data, we explore a series of simpler comparison models. For illustration, Figure 2b 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 price state *or* her affect prefers price states with higher prior probabilities. Finally, a model that considers the full range of goals produces interpretations that demonstrate hyperbole

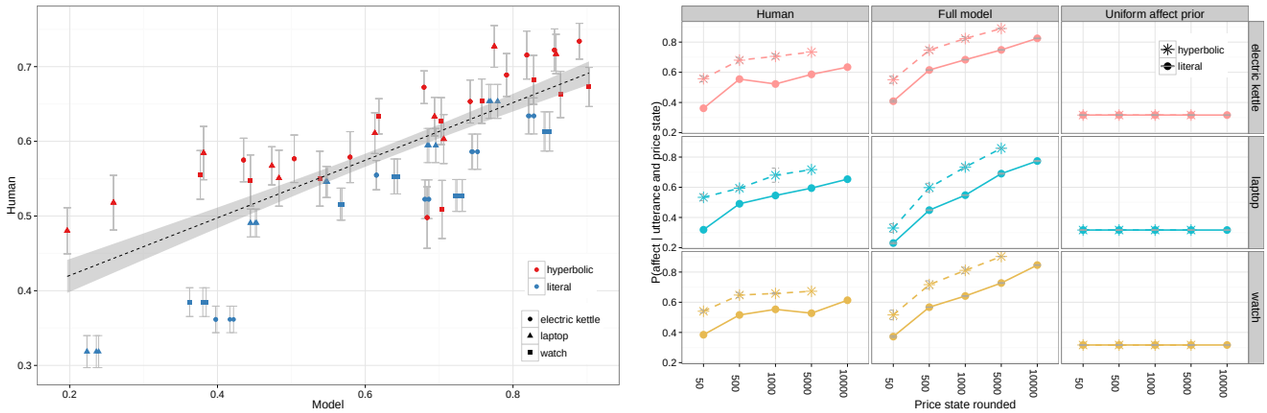

Fig. 4. (a) Model predictions of affect (x-axis) versus human responses (y-axis) for 45 data points (3 Items × 15 Utterance-Price state pairs where $u \geq s$) in Experiment 2. (b) Probability of inferring affect given a price state and a hyperbolic or literal utterance. Humans infer higher probability 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.

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 3a 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 rely on (common) knowledge of a domain's prior distribution to infer hyperbolic interpretations, not the semantics of the number words alone. Figure 3b 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, yielding 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 ($r = 0.771, p < 0.00001$), capturing most of the reliable variation in these data (Figure 4a). Figure 4b 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 uses 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: 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 of language. However, 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—informativity with respect to a goal—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.

Added Our model has important connections to theories of communication and linguistic meaning as well as their relation to nonliteral language. By modeling the speaker as choosing an utterance that maximizes informativeness with respect to her goal, we showed that people can infer nonliteral meaning by assuming that the speaker follows a principle of relevance. By introducing a dimension of meaning that is associated with a state of the world and not encoded into the semantics of an utterance, we formalized an encyclopedic approach to meaning that views meaning as a combination of linguistic and extralinguistic knowledge. While our model is currently limited to two dimensions of meaning and goals, in future work we

hope to capture dimensions of meaning central to other figures of speech such as irony and metaphor, thus extending our model to explain nonliteral language more broadly. We believe that our 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. Let u be the utterance a speaker utters. The meaning of the utterance has two dimensions: the actual price state s and the speaker's affect a. We defined the set of price states $S = \{50, 51, 500, 501, 1000, 1001, 5000, 5001, 10000, 10001\}$. We 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—this binarization is purely for simplicity). Given the price states S and 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 $a \in A$.

The speaker S_n is assumed to be a planner who's goal is to be informative about some relevant topic. We write both the goal and its topic as g. S_n chooses utterances according to a softmax decision rule that describes an approximately rational planner [23]:

$$S_n(u|s, a, g) \propto e^{U_n(u|s,a,g)} \quad \textbf{[1]}$$

We wish to capture the notion that the speaker aims to be informative about a topic of discussion, while minimizing cost. If the topic is represented by a projection $g : M \to X$ from the full space of meanings to a relevant subspace, then the speaker cares only about the listener's distribution over the subspace,

$$L_n(x|u) = \sum_{s',a'} \delta_{x=g(s',a')} L_n(s', a'|u).$$

Following the Rational Speech Act model, we formalize informativity of an utterance as the negative surprisal of the intended meaning under the listener's distribution; here the listener's distribution over the topical subspace X. Hence:

$$U_n(u|s, a, g) = \log L_n(g(s, a)|u) - C(u),$$

where $C(u)$ represents the utterance cost. Substituting into equation 1, this gives:

$$S_n(u|s, a, g) \propto \sum_{s',a'} \delta_{g(s,a)=g(s',a')} L_n(s', a'|u) \cdot e^{-C(u)} \quad \textbf{[2]}$$

In our situations, the speaker may have the goal of communicating information along the price dimension, the affect dimension, or both. This gives three possible projections r:

$$r_s(s, a) = s$$
$$r_a(s, a) = a$$
$$r_{s,a}(s, a) = s, a.$$

The speaker may also want to communicate the price either exactly or approximately (we assume that no such distinction exists for affect, since we have already binarized it). When the speaker wants to communicate the price approximately, she projects numbers to their closest round neighbors. For example, such a speaker will represent the prices 51 and 1001 as 50 and 1000, respectively. This gives two projections (exact and approximate), f, is defined as:

$$f_e(s) = s$$
$$f_a(s) = \text{Round}(s),$$

where Round(s) denotes the multiple of 10 which is closest to s. The two types of goal, f and r, can be composed to make the goal g of the speaker: $g(s, a) = r(f(s), a)$, which results in $2 \times 3 = 6$ possible goals (though note that $r_a(f_e(s), a)$ and $r_a(f_a(s), a)$ are equivalent).

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:

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

propto?

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.

$$L_n(s, a|u) \propto \sum_g P_S(s) P_A(a|s) P_G(g) S_{n-1}(u|s, a, g) \quad \textbf{[3]}$$

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 $P_G(g)$ is defined to be uniform. 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. We obtained a posterior distribution for all possible meanings s, a given an utterance u. Raw data for model predictions are here: `http://stanford.edu/~justinek/hyperbole-paper/data/model-predictions.csv`. Figure S1 shows the full posterior distributions for all utterances.

Experiment 1: Halo and hyperbole. todo: seek and destroy remaining broken urls 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 $\times$ 10 Utterances). The stimuli for Experiment 1 can be found here: `http://stanford.edu/~justinek/hyperbole-paper/materials/experiment1.html`

We normalized participants' ratings across price states for each trial to sum up to 1. There are a total of 300 average normalized ratings (3 Items $\times$ 10 Utterances $\times$ 10 Price States). The average normalized ratings across participants for each item/utterance pair is shown in Figure 5B. 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 power-law transformation on the model's distribution: We multiplied the predicted probability for each meaning by a free parameter λ and renormalized the probabilities to sum up to 1 for each utterance. Fitting λ to the behavioral data to optimize correlation, we obtained the best fit with $\lambda = 0.35$, resulting in a correlation between model predictions and participant ratings of $r = 0.973$ (see main text). All figures and analyses that we report in the main text are with this transformation. Without 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 item's 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 subject's 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 ($F(1, 28) = 18.94, p < 0.001$).

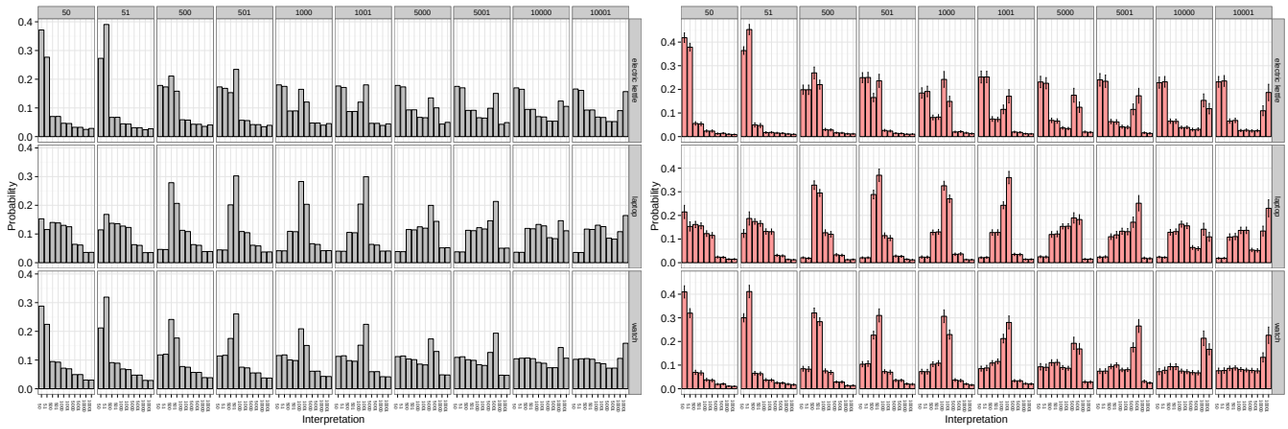

Fig. 5. (a)Full posterior meaning distribution predicted by the model for each utterance. Each column of panels is an utterance, and each row of panels is an item type. Each panel represents the interpretation distribution given an utterance for an item. (b) Full meaning distribution produced by humans 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.

Experiment 2: 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 $\times$ 60 $\{u, s\}$ pairs where $u \geq s$). The stimuli for Experiment 2 can be found here: `http://stanford.edu/~justinek/hyperbole-paper/materials/experiment2.html`; the raw data is here: `http://stanford.edu/~justinek/hyperbole-paper/data/experiment2-raw.csv`. 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 3a: Price prior. To obtain people's 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: `http://stanford.edu/~justinek/hyperbole-paper/materials/experiment3a.html`. 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: `http://stanford.edu/~justinek/hyperbole-paper/data/experiment3a-normalized.csv`

Experiment 3b: 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: `http://stanford.edu/~justinek/hyperbole-paper/materials/experiment3b.html`. 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: `http://stanford.edu/~justinek/hyperbole-paper/data/experiment3b-raw.csv`

ACKNOWLEDGMENTS. This work was supported in part by an NSF Graduate Research Fellowship to JTK and by a John S. McDonnell Foundation Scholar Award and grants from the ONR to NDG.

1. Roberts, R.M. and Kreuz, R.J., Why do people use figurative language?, (1994), Psychological Science. 5(3), pp. 159–163
2. Dews, S. and Winner, E. Obligatory processing of literal and nonliteral meanings in verbal irony, (1999), Journal of Pragmatics. 31(12), pp 1579–1599.
3. Glucksberg, S. Understanding figurative language: From metaphors to idioms, (2001), Oxford Univ. Press.
4. Gibbs, R. Figurative language, (1999), The MIT encyclopedia of the cognitive sciences, pp. 314–315.
5. Grice, H.P. Logic and conversation, (1975), pp. 41–58.
6. Clark, H.H. Using language (1996) Cambridge University Press, Vol. 4
7. Sperber, D. and Wilson, D. and Ziran, H. Relevance: Communication and cognition, (1986).
8. Wilson, D. and Sperber, D. Relevance theory, (2002), Handbook of pragmatics.
9. Sperber, D., and Wilson, D. A deflationary account of metaphors, (2008), The Cambridge handbook of metaphor and thought, pp. 84–105.
10. Wilson, D. and Carston, R. Metaphor, relevance and the emergent property' issue, (2006), Mind and Language, 21(3), pp. 404–433.
11. Sperber, D. and Wilson, D. Loose talk, (1985), In Proceedings of the Aristotelian society, pp. 153–171.
12. Frank, M.C. and Goodman, N.D., Predicting pragmatic reasoning in language games, Science, 336(6048), (2012), pp. 998
13. Goodman, N.D. and Stuhlmüller, A. Knowledge and implicature: Modeling language understanding as social cognition, Proceedings of CogSci conference, (2012)
14. 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)
15. Jäger, G. and Ebert, C., Pragmatic rationalizability, Proceedings of Sinn und Bedeutung, 13, (2009), pp. 1–15
16. McCarthy, M. and Carter, R, There's millions of them: hyperbole in everyday conversation, Journal of pragmatics, 36(2), (2004), pp. 149–184.
17. Gibbs, R.W., Irony in talk among friends, Metaphor and symbol, (2000), pp. 5–27
18. Gibbs, R.W. and O'Brien J., Psychological aspects of irony understanding, Journal of pragmatics,16(6), (1991), pp. 523–530
19. Lasersohn, P., Pragmatic halos, Language, (1991), pp. 522–551
20. Bastiaanse, H., The rationality of round interpretation, Vagueness in communication, (2011), pp. 37–50
21. Krifka, M., Approximate interpretation of number words: A case for strategic communication, Cognitive foundations of interpretation, (2007), pp. 111–126
22. Van der Henst, J. and Carles, L. and Sperber, D. Truthfulness and Relevance in Telling the Time, Mind and Language, (2002), 17(5)
23. Sutton, R.S. and Barto, A.G., Reinforcement learning: An introduction, 28, (1998)