***Science* Manuscript Template**

**General Instructions on using this template and submitting a manuscript to *Science*:** Thank you for preparing a manuscript for submission to *Science*. Using this template, or following the guidelines below, will help us in processing your paper. Our goal is to be able to identify each section of your manuscript so that we can accurately record the title, authors, abstract, etc. and also be able to enrich it by including reference links and an accurate layout.

Please use the actual template, which starts on page 2. When you are ready to submit, please delete the text on this cover page.

You can submit your paper at [www.submit2science.org](http://www.submit2science.org)

Additional instructions are available at <http://www.sciencemag.org/site/feature/contribinfo/index.xhtml>

If you are using LaTeX , please convert your paper into a Word or RTF file if possible. If this is not possible, please use our LaTeX template and upload a PDF version of your paper. Some conversion approaches are available here: <http://www.tug.org/utilities/texconv/textopc.html>

So that we can identify the parts of your paper, and even if you do not use our template, please begin each section with the specific key words listed below, some of which are followed by a colon. Please do not use paragraph breaks in the title, author list, or abstract.

**Title:**

**Authors:**

**Affiliations:**

**One Sentence Summary:**

**Abstract:**

**Main Text:**

**References and Notes** (followed by a numbered list)

**Acknowledgements:**

**Fig. #:** (Begin each figure caption with a label, “**Fig. 1**.” for example, as a new paragraph.).

**Table #:** (Begin each table caption with a label “**Table 1.”**, etc.) as a new paragraph.

**Supplementary Materials:** a list of the supplementary materials, followed by the actual text of the Supplementary Materials.

Several of these headings are optional, for example, not all papers will include a one-sentence summary, figures or tables, or supplementary material.

Please use the .docx format if possible (all version after Word 2007).

More specific formatting instructions are provided in the actual template, which follows.

Hyperbolically Speaking:

A Computational Model of Nonliteral Language Understanding

**Authors:** Justine T. Kao1\*, Jean Wu1, Leon Bergen2, Noah D. Bergen1

**Affiliations:**

1Stanford University.

2Massechusets Institute of Technology.

\*Correspondence to: justinek@stanford.edu

†Additional author notes should be indicated with symbols (for example, for current addresses).

**Abstract**: One of the most puzzling and important facts about communication is that people do not always mean what they say. In everyday situations, speakers use imprecise, exaggerated, or otherwise literally false descriptions to communicate their experiences and opinions. Here we focus on the nonliteral interpretation of number words, in particular pragmatic halo (the imprecise interpretation of round numbers) and hyperbole (interpreting exaggerated and unlikely numbers as conveying affect). We model number interpretation as rational social inference regarding the communicative goal, meaning, and affective subtext of an utterance and show that it accurately predicts humans’ pragmatic interpretation of number words. We present our model as one of the first computational approaches to quantitatively capture a range of effects in nonliteral language understanding.

**One Sentence Summary:** We present a computational model of nonliteral language understanding and show that it accurately predicts pragmatic and hyperbolic interpretation of number words.

**Main Text:**

Imagine a conversation with a friend about 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 your friend 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 hours to get a table.” You will probably interpret this to mean that the wait was shorter than a million hours, but importantly that she thinks it took much too long. One of the most puzzling and important facts about communication is that people do not always mean what they say. As a result, a crucial part of a listener’s job is to understand an utterance even when its literal meaning is extremely unlikely. The ubiquity of nonliteral language and the ease with which people are able to interpret it present a puzzle for language understanding research. Although there is a rich body of literature examining motivations for using nonliteral language and psychological effects of processing it (*1, 2, 3, 4*), there has been little work on building formal models that capture these effects quantitatively.

Many linguists and psychologists have traditionally viewed communication as an interaction between rational, cooperative agents (*5, 6*). A recent body of work formalizes these views by modeling pragmatic language understanding as probabilistic inference over recursive social models (*7, 8, 9, 10*). While these models are able to quantitatively explain a range of phenomena in human pragmatic reasoning, they are unable to handle utterances where the intended meaning directly contradicts the literal meaning, as is the case in metaphor (“Juliet is the sun”) and hyperbole (“It took a million hours to get a table”). Here we propose that nonliteral language understanding relies on considering communicative goals that are distinct from what is conveyed by the literal meaning of an utterance. 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. By modeling language understanding as social inference regarding the communicative goal, meaning, and affective subtext of an utterance, we show that our model produces nonliteral interpretations that match humans’ in the case of number words.

We focus on number words for three reasons: first, despite their flexible and nonliteral usages in everyday language, numbers have precise literal meanings that can be easily formalized, unlike more complex concepts such as “Juliet” or “the sun.” Second, number words can be systematically manipulated on a continuous scale to yield quantitative predictions. Third, there are two particular well-known phenomena regarding number interpretation: pragmatic halo and hyperbole. Pragmatic halo refers to people’s tendency to interpret simple number expressions imprecisely and complex number expressions precisely (*11*). While this effect has been formalized via game theory as a rational choice given different costs of utterances (*12, 13*), our model captures these arguments within a Bayesian framework for pragmatic inference. Hyperbole is defined as a figure of speech that uses exaggeration to create emphasis. While hyperbolic utterances are literally false, listeners can successfully infer the intended meaning and regard hyperbole as a source of humor or signal of interpersonal closeness (*1, 14, 15, 16*). Our model shows that common prior knowledge about the relevant topic plays an important role in interpreting hyperbolic statements. Furthermore, when a speaker makes a hyperbolic utterance, our model captures the intuition that the listener will infer an affective subtext beyond its literal meaning.

Here we provide the intuition for how a recursive model captures pragmatic halo. The model begins with a naïve listener who interprets a number utterance literally; e.g., interprets “30 minutes” as *30 minutes*. A speaker reasons about the literal listener and anticipates his interpretation of her utterance. Suppose the actual wait time was 32 minutes and she wants to communicate it imprecisely. By saying “30 minutes,” she successfully communicates the imprecise wait time to the literal listener without having to use a longer and more costly utterance like “32,” which makes the utterance optimal for her communicative goal. A pragmatic listener recursively reasons about the speaker and anticipates her choice of utterance given a state of the world and a possible communicative goal. The pragmatic listener hears “30 minutes” and knows that the speaker would optimally choose this utterance if she waited 32 minutes and wanted communicate imprecisely. However, if she waited 30 minutes and wanted to communicate imprecisely, the costlier utterance “32” is not an optimal utterance. As a result, a pragmatic listener will interpret “30 minutes” as *roughly 30 minutes*, and “32 minutes” as *exactly* *32 minutes*. We now provide the intuition for how a recursive model captures hyperbole. The model begins with a naïve listener who interprets a number utterance literally and has background knowledge about the affect associated with a particular wait time. The literal listener interprets “a million hours” as *a million hours*. Since the conditional probability of a speaker being upset after waiting for a million hours is very high, the literal listener understands that the speaker is very likely upset. A speaker reasons about the literal listener and anticipates his interpretation of her utterance. Suppose the actual wait time was 32 minutes and she wants to communicate she was upset. By saying “a million hours,” she successfully communicates her affect to the literal listener, which makes the utterance optimal for her communicative goal. A pragmatic listener recursively reasons about the speaker and anticipates her choice of utterance given a state of the world, her affect, and her communicative goal. The pragmatic listener hears “a million hours” and knows its literal meaning is extremely unlikely to be literally true, but that it satisfies the speaker’s communicative goal if her goal is to communicate affect. The pragmatic listener thus infers that “a million hours” means shorter than a million hours and that the speaker was upset.

Our model combines both sets of intuitions described above to pragmatically interpret number expressions (see Supplementary Materials for formal description). We test the model on the prices of three kinds of everyday items: electric kettles, watches, and laptops. We chose price because it is a common topic of conversation in which people use number expressions, and selected the three items because they are everyday products with varying price distributions. The set of possible price states for the three kinds of items is defined as S={50, 50', 500, 500', 1000, 1000', 5000, 5000', 10000, 10000'}. Each price state is either “round” (divisible by 10) or “sharp” (not divisible by 10). We assume that the set of utterances U is equivalent to the set of price states S. A speaker can say, “That electric kettle cost *u* dollars,” for *u* U, and a listener can interpret this to mean that the kettle cost *s* dollars, for *s* S. We obtained posterior meaning distributions for the ten numerical utterances using the price priors and affect priors for each of the three items (see Experiment 3a and 3b in Supplementary Materials). Figure SSS in the Appendix shows the full meaning distribution for each utterance across the three item kinds. Figure FFF summarizes this distribution into four types of interpretations: exact (e.g., “1000” interpreted as 1000), fuzzy (e.g. “1000” interpreted as 1001 or “1001” interpreted as 1000), hyperbolic (e.g. “1000” interpreted as “100”), and affective (e.g. “1000” interpreted as conveying affect about the price being too expensive). We see that round utterances such as “500” and “1000” are interpreted less exactly and more fuzzily than their sharp counterparts, which demonstrates the pragmatic halo effect. Utterances that 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). We also see an interaction between halo and hyperbole, where round utterances such as “5000” and “10000” are more likely to be interpreted hyperbolically than their sharp counterparts. Finally, utterances whose literal meanings are associated with higher affect priors (such as “10000” and “10001”) are more likely to be interpreted as conveying affect.

We conducted Experiment 1 to examine humans’ interpretation of number words using the same set of items, price states, and utterances as described earlier. Subjects read scenarios in which a buyer produces an utterance *u* about the price of an item he just bought. They then rate the likelihood that the item actually cost *s* dollars for *s* S (see Experiment 1 in Supplementary Materials). Figure FFF in the Appendix shows the full interpretation distribution for each utterance. To examine the halo effect in humans, we computed the difference between the probability of an exact interpretation and the probability of a fuzzy interpretation for each utterance. Collapsed across items and utterances, this difference is significantly smaller for round numbers than for sharp numbers (F(1, 4)=16.31, p < 0.05), which indicates that round numbers tend to be interpreted more approximately and have a weaker bias towards exact interpretation than their sharp counterparts. To examine the hyperbole effect, we show that utterances whose literal meanings have low probabilities under the item’s prior price distribution are more likely to be interpreted hyperbolically (F(1, 10) = 44.06, p < 0.0001). We compared model and human interpretation probabilities for the 300 items (I X U X S) and show that model predictions correlate significantly with human interpretations of number words (r=0.974, p<0.0001) (Figure FFF).

We conducted Experiment 2 to examine the affective subtext conveyed using hyperbolic versus literal utterances. Subjects read scenarios in which a speaker bought an item that cost *s* dollars and tells his friend that it cost *u* dollars, where *u* *s*. They then rate how likely it is that the buyer thinks the item was expensive (see Experiment 2 in materials). 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) = 9.592, p < 0.01), confirming the hypothesis that listeners infer affective subtext from hyperbolic utterances. We compared model and human interpretations of affect for each of the 45 items (I X U X S where *u* *s*). While there is a significant amount of noise in the human data (average split-half correlation is 0.833), the model significantly predicts human interpretations of the utterances’ affective subtext (r=0.772, p < 0.00001).

Each component of the model is responsible for capturing different effects that we observe in the human data. The left-most panel in Figure FFF shows the halo effect in humans, i.e. the tendency for round numbers to be interpreted less exactly. This effect is replicated by the full model, but not by a version of the model where the costs of utterances are uniform, suggesting that people take into account utterance cost to infer exact versus fuzzy interpretations. The leftmost panel in Figure FFF shows humans’ probabilities of interpreting each utterance hyperbolically. These probabilities are captured faithfully by the full model, but not by a version of the model where the price prior distribution for each item is uniform. This suggests that people take into account the prior distribution of prices when inferring hyperbolic versus literal interpretations. The leftmost panels in Figure FFF show humans' probabilities of inferring affective subtext given a price state and a literal or hyperbolic utterance. Across price states, higher price states are associated with higher degrees 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 a version of the model where the probability of affect given each price state is uniform. This suggests that the affective subtext conveyed by hyperbolic utterances is driven by prior probabilities of affect given a price state. Moreover, utterances interpreted hyperbolically convey affective subtext beyond the prior affect associated with the actual state of the world, which is consistent with our intuition about the rhetoric effect of hyperbole.

Finally, we show that reasoning about a speaker's communicative goals is crucial for capturing the nonliteral interpretation of number words demonstrated in the behavioral experiment. Figure FFF compares different model interpretations of the utterance “The electric kettle cost 1000 dollars.” A listener that does not consider alternative communicative goals interprets the utterance literally and places all interpretation mass on the price state *1000*. A listener that considers a speaker whose goal may be to communicate precisely or imprecisely places a significant amount of interpretation mass on the neighboring price state *1001*. A listener that considers a speaker whose goal may be to communicate the precise price state or her affect about the price state can now interpret the utterance hyperbolically and place mass on price states with higher prior probabilities. However, he places very little mass on the neighboring price state *1001*. Finally, a listener that considers the possibility of precise, imprecise, and affect goals closely fits humans' interpretation of the utterance.

Our model and behavioral results reveal complex patterns in the nonliteral interpretation of number words. These patterns depend on the listener's prior knowledge, consideration of communication efficiency, and reasoning about the speaker's communicative goal. Besides producing quantitative predictions for pragmatic number interpretation, we also introduced an “affect” dimension that takes into account subjective aspects of communication. When we use language, we don’t always use it just to communicate the actual state of the world, but also how we feel about it, which is an important element in nonliteral language and communication in general. Our model shows that rational recursive reasoning between speaker and listener can capture rich effects in language understanding that have been difficult to model computationally. We present an innovative formal framework for examining phenomena in nonliteral language understanding, and show evidence that this framework and its extensions will advance language research with the hopes that some day--hyperbolically speaking--Bayesian models can explain “*everything.*”

References and Notes:

1. R. M. Roberts, R. J. Kreuz, Why do people use figurative language?. *Psychological Science*. **5(3)**, 159-163 (1994).
2. S. Dews, E. Winner, Obligatory processing of literal and nonliteral meanings in verbal irony. *Journal of Pragmatics*. **31(12)**, 1579-1599 (1999).
3. S. Glucksberg, *Understanding figurative language: From metaphors to idioms*. (Oxford Univ. Press, 2001).
4. R. Gibbs, *Figurative language.*(The MIT encyclopedia of the cognitive sciences, 1999), pp. 314-315.
5. H. P. Grice, *Logic and conversation*. (1975), pp. 41-58.
6. H. H. Clark, *Using language*(Vol. 4). (Cambridge Univ. Press, 1996).
7. M. C. Frank, N. D. Goodman, Predicting pragmatic reasoning in language games. *Science*. **336(6084)**, 998-998 (2012).
8. N. D. Goodman, A. Stuhlmüller, Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, **5(1)**, 173-184 (2013).
9. L. Bergen, N. D. Goodman, R. Levy, That’s what she (could have) said: How alternative utterances affect language use. *In Proceedings of the thirty-fourth annual conference of the cognitive science society* (2012).
10. G. Jäger, C. Ebert, Pragmatic rationalizability. *In Proceedings of sinn und bedeutung*. 13, 1-15 (2009).
11. P. Lasersohn, Pragmatic halos. *Language*, 522-551 (1999).
12. H. Bastiaanse, The rationality of round interpretation. *In Vagueness in communication.* 37-50 (2011).
13. M. Krifka, Approximate interpretation of number words: A case for strategic communication. *Cognitive foundations of interpretation*. 111-126 (2007).
14. M. McCarthy, R. Carter, “There's millions of them”: hyperbole in everyday conversation. *Journal of pragmatics*. **36(2)**, 149-184 (2004).
15. R. W. Gibbs, Irony in talk among friends. *Metaphor and symbol*. **15**(1-2), 5-27 (2000).
16. R. W. Gibbs Jr, J. O’Brien, Psychological aspects of irony understanding. *Journal of pragmatics*. **16**(6), 523-530 (1991).
17. R. S. Sutton, A. G. Barto, *Reinforcement learning: an introduction*, *Vol. 1, No. 1*. (Cambridge: MIT press, 1998).

**Acknowledgments:** The acknowledgments should include a statement about where the data reported in the paper are presented, archived, or available (for example, in the Supplementary Materials or in a community archive). If in an archive, include the accession number or a placeholder for it. Please also include relevant funding information such as grant numbers and funding agencies. You can also include a statement of author contributions here or in the Supplementary Materials.



**Fig. 1**. Caption *The figure caption should begin with an overall descriptive statement of the figure followed by additional text. They should be immediately after each figure. Figure parts are indicated with capital letters (****A****). If you prefer, you can place both the actual figures and captions logically through the text near where they are cited rather than at the end of the file (but not both). If a paragraph in the main text begins with the name of a figure, write out “Figure” in full (e.g., <para>“Figure 1 shows….”)*



**Fig. 2.**



**Fig. 3.**



**Fig. 4.**

Supplementary Materials:

Materials and Methods

Figures S1-S#

External Databases S1-S#

References (*##-##*)

**Materials and Methods:**

**Model.** Here we describe our model in detail. Let *u* be the utterance a speaker utters. Let *m* be the meaning that is known to the speaker. *m* has two dimensions, one concerning the actual price state *s*, and one concerning the speaker's affect *a*. Given the set of price states S and set of affect states A, the set of possible meanings M is given by M = S X A. We denote each possible meaning as *ms,a*, where *s* S and *a*  A={0, 1} (0 means no affect and 1 means with affect).

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***s,a**where **s** 2S and **a** 2A. The goal *g***s,a**is a function *g***s,a**: M {0,1}, such that *g***s,a**(*ms,a*) = 1 if and only if *s* **s**, *a* **a**. Thus, a goal specifies a subset of price states and affects, and a meaning satisfies this goal if it belongs to this subset; such a subset will be referred to as a goal state. We assume that there are two types of price-related goal states: 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. **s**={*i*} (for some *i* S), and approximate goals are represented by subsets that consist of the price states within a distance of 1 of some state, i.e. **s**={*j | j S, |j-i|* 1}.

The prior probability of a price state *s* is taken from an empirically derived price prior PS(*s*), and the probability of an affect *a* given a price state *s* is taken from an empirically derived conditional affect prior PA(*a*|*s*) (see Experiments 3a and 3b). The probability distribution PG(|*ms,a*) over goals given that the speaker knows meaning *ms,a* is defined to be uniform over goals consistent with *ms,a*, i.e. uniform over goals *g***s,a**such that *g***s,a**(*ms,a*) = 1. This is equivalent to assuming that the speaker either wants to communicate their meaning exactly or approximately.

A literal listener *L0* provides the base case for recursive social reasoning between the speaker and listener. *L0* interprets an utterance u literally without taking into account the speaker's communicative goals:

*L0* (*ms,a* | *u*) = (1)

The speaker *Sn* 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. *Sn* chooses utterances according to a softmax decision rule that describes an approximately rational planner (*17*):

*Sn* (*u* | *g***s,a**) (2)

where the constant captures the degree of optimality of the speaker (we used =1 in the model simulations described).

The speaker wants to minimize both the cost C(*u*) of the utterance and the surprisal of her goal state. The utility function is therefore defined by:

) = (3)

which combined with equation 2 leads to:

*Sn* (*u* | *g***s,a**) ( (4)

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.

(5)

After obtaining a posterior distribution for all possible meanings *m* given an utterance *u*, we performed a Luce choice transformation on the distribution and fit to the behavioral data with in order to compare with behavioral results. Figure S1 shows the full posterior distributions for all utterances.

**Experiment 1: Halo and Hyperbole.** 120 subjects were recruited on Amazon's Mechanical Turk. Each subject 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. Bob responds by saying “It cost *u* dollars,” where *u* {50, 50 k, 500, 500 k, 1000, 1000 k, 5000, 5000 k, 10000, 10000 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. Given an utterance *u*, subjects 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 k, 500, 500 k, 1000, 1000 k, 5000, 5000 k, 10000, 10000 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. We normalized subjects' ratings across price points for each trial to sum up to 1. The average normalized ratings across subjects for each item/utterance pair is shown in Figure S2.

**Experiment 2: Affective subtext.** 160 subjects were recruited on Amazon’s Mechanical Turk. Each subject 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. Bob responds by saying “It cost *u* dollars,” where *u* U and . Subjects 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. The average ratings determine the degree of affect conveyed by an utterance given the actual price state.

**Experiment 3a: Price prior.** To obtain people’s prior knowledge of the price distributions for electric kettles, laptops, and watches, 30 subjects were recruited from Amazon's Mechanical Turk. Each subject rated the probability of an electric kettle, laptop, and watch costing *s* dollars, where *s* S. Ratings for each price state were on a continuous scale from “impossible” to “extremely likely”, represented as real values between 0 and 1. We normalized subjects' ratings across price points for each trial to sum up to 1. The average normalized ratings across subjects 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.

**Experiment 3b: Affect prior.** To obtain people’s prior knowledge of the affect likelihood given a price state, 30 subjects were recruited from Amazon’s Mechanical Turk. Each subject read 15 scenarios where someone had just bought an item that cost *s* dollars (s S). 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 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.



Fig. S1.



Fig. S2.

**Any Additional Author notes:** For example, author contributions or a list of group authors.

**Main text:**

*In general, this should include a brief (1-2 paragraph) introduction, followed by a statement of the specific scope of the study, followed by results and then interpretations. Please avoid statements of future work or claims of priority, and avoid repeating the conclusions at the end. All of the Figures and Tables should be cited in order, including those in the supporting online material (which should be cited as, for example, figure S1, and table S1). You can include page breaks if you would like to place the figures within the text.*

***Subheads****. These can be included in Research Articles or Reviews and should be brief.*

*References should be cited in parentheses with an italic number (1). Multiple reference citations are separated by commas (2, 3) or if a series, dashes (4-6). References are cited in order by where they first are called out, through the text, then the notes, captions, then through the supplementary material.*

*Equations can be included. We do not recommend using the native Word 2007, 2008, 2010 or 2011 equation editor. This can in some cases produce less reliable MathML, the online markup language we use, which may result in display errors. Instead, use the legacy equation editor in word (Insert menu; select insert object; select word equation) or use Mathtype (recommended). If you enter equations in simple LaTeX, check that they will convert accurately (Word 2007 and higher can convert simple LaTeX equations).*