# Hyperbolically speaking: a computational model of non-literal language understanding

**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 intriguing and challenging properties of language understanding is that language is not always meant to be interpreted literally. In everyday situations, people often use imprecise, exaggerated, or otherwise literally false descriptions to communicate their experiences and opinions. In this paper we focus on the non-literal interpretation of number words, in particular the effects of pragmatic halo (the imprecise interpretation of round numbers) and hyperbole (the affective subtext conveyed by exaggerated and unlikely numbers). Building on recent models of pragmatics as rational inference between speaker and listener, we model number interpretation as social inference regarding the communicative goal, meaning, and affective subtext of an utterance. Our model accurately predicts humans' pragmatic interpretation of number words, and is one of the first computational models to quantitatively capture a range of effects in non-literal language understanding.**

Pragmatics | Language understanding | Computational modeling

## Introduction

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. Now suppose your friend says: "It took 32 minutes to get a table." You are more likely to interpret the number expression to mean exactly 32, and believe that she cares to communicate the exact wait time. Suppose she says: "It took a million hours to get a table." You will probably interpret her to mean that the wait was shorter than a million hours, but importantly that she thinks it took much too long.

Given the incredible flexibility of language, a crucial part of a listener's job is to understand an utterance even when its literally meaning is extremely unlikely. Non-literal language understanding is one of the biggest challenges in language research, and it has been difficult to build formal models or design empirical studies that capture effects in non-literal language understanding quantitatively. In this paper, we present a computational model that predicts people's non-literal interpretation of number words. We build on a traditional approach in language understanding that views communication as an interaction between rational, cooperative agents [1, 2], and show that non-literal interpretation of number terms can be explained as effects of probabilistic inference over recursive social models.

Recent work has shown that modeling communication as recursive social inference is able to quantitatively explain rich phenomena in human pragmatic reasoning [3, 4, 5, 6]. However, a limitation of these models is that 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 extend the model by introducing uncertainty over the speaker's communicative goal, and propose that non-literal language understanding relies on considering the possibility of communicative goals that are distinct from what is conveyed by the literal meaning of an utterance. More concretely, a listener reasons about a speaker who optimizes informativeness of her utterances given a particular communicative goal; the speaker choses the optimal utterance assuming that the listener is reasoning in this way about the speaker; and so on. In this paper, we show how this framework of recursive social inference can be applied to capture non-literal hyperbolic interpretation of number words.

We focus on number words for three reasons: first, despite their flexible and non-literal 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*–the imprecise interpretation of round numbers, and *hyperbole*–the affective subtext conveyed by exaggerated and unlikely numbers.

Pragmatic halo describes the phenomenon in which people tend to interpret simple number expressions imprecisely and complex number expressions precisely [7]. This effect has been formalized via game theory as a rational choice given different costs of utterances [8, 9]. The model we propose captures these arguments within a Bayesian framework for pragmatic inference. Given uncertainty about whether a speaker wishes to communicate precisely or approximately in addition to differential utterance costs, we show that a rational listener will interpret costlier number words as more precise.

While hyperbolic utterances are literally false, listeners usually successfully infer the speaker's intended meaning and often regard hyperbole as a source of humor or signal of interpersonal closeness [10, 11, 12]. Previous work on human and machine identification of irony and hyperbole has focused on linguistic cues such as slow speaking rate, heavy stress, nasalization, and interjections [13, 14, 15, 16]. Here we show that common prior knowledge about the relevant topic also plays an important role in identifying and interpreting hyperbolic statements. That is, part of what makes an utterance likely to receive a hyperbolic interpretation is that both speaker and listener know that the literal meaning is extremely unlikely. Furthermore, given the possibility that the speaker wishes to convey a subjective opinion instead of simply the objective state of the world, the listener can infer an affective subtext that goes beyond the literal meaning of the utterance.

By modeling language understanding as social inference regarding the communicative goal, meaning, and affective sub-

**Reserved for Publication Footnotes**

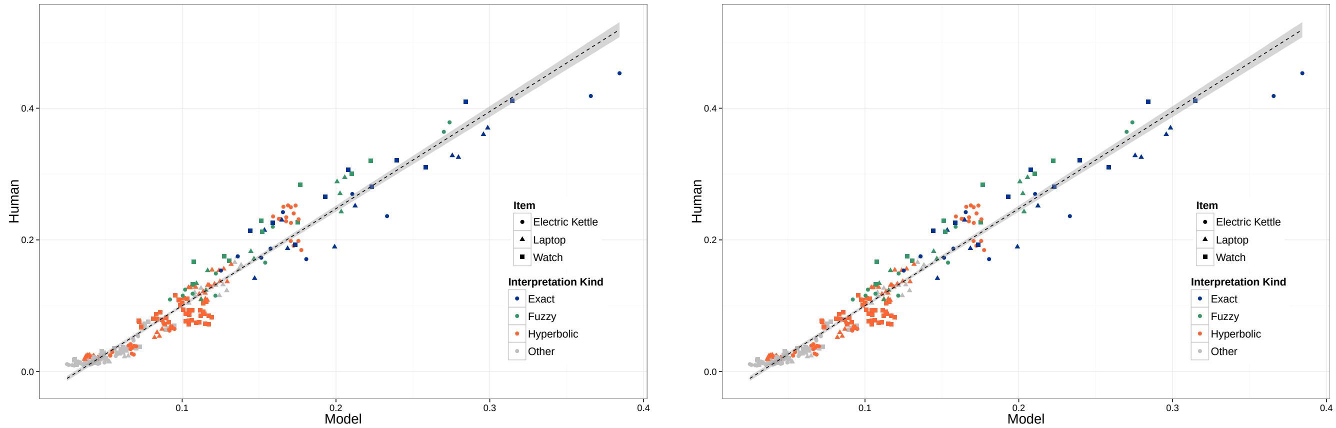


**Fig. 1.** (a) This plot summarizes pragmatic halo, hyperbole, and affective interpretations of the model. (b) The model significantly predicts human interpretations with high correlation ($r = 0.97$, $p < 0.0001$) for all 300 items.

text of an utterance, we show that our model captures the non-literal effects of halo and hyperbole.

## Results

We test our model's interpretation of number expressions regarding the prices of three kinds of everyday items: *electric kettles*, *watches*, and *laptops*. We focused on price because it is a common and natural topic of conversation in which people use number expressions. We chose the three items because they are common products whose price distributions are quite different from each other. We define the set of possible price states for the three kinds of items: $S = \{50, 50', 500, 500', 1000, 1000', 5000, 5000', 10000, 10000'\}$. Each price state is either "round" (divisable by 10) or "sharp" (which we define as $n \pm k$, $k \in \{1, 2, 3\}$, where $n$ is round). We assume that the space of possible utterances $U$ is equivalent to the set of price states $S$. A speaker can say, "That electric kettle cost $u$ dollars," for $u \in U$, and a listener can interpret this to mean that the kettle cost $s$ dollars, for $s \in S$.

**Model simulations.** Given the formal setup of our model described in the methods section, we obtained posterior meaning distributions for the ten numerical utterances using the price priors and affect priors for each of the three items. Figure 4 in the Appendix shows the full meaning distribution for each utterance across the three item kinds. We seperated interpretations into three types: *exact interpretations*, which are interpretations that are identical to the utterance (e.g. when"1000" is interpreted as meaning 1000); *fuzzy interpretations*, which are interpretations that are the sharp or round counterpart to the utterance (e.g. when "1000" is interpreted as 1001); and *hyperbolic interpretations*, which are interpretations that are much smaller than the utterance (e.g. when "1000" is interpreted as 50 or 501). The model also returns the probability of an utterance being interpreted as conveying *affect* about the price being too expensive. Figure **??** summarizes these effects. We first see a basic effect of the prior, in which utterances that are more likely given the price prior of the item are more likely to be interpreted exactly or fuzzily (e.g. "1000" is more likely to be interpreted exactly or fuzzily for laptops than for electric kettles). The model demonstrates the pragmatic halo effect, in which round utterances such as "500" and "1000" are interpreted less exactly and more fuzzily than their sharp counterparts "501" and "1001." It also demonstrates the hyperbole effect, in which 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 as $50, 51, 500$, or $501$ for electric kettles than for 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. Furthermore, utterances whose literal meanings are associated with higher affect priors (such as "10000" and "10001") are more likely to be interpreted as conveying affect.

**Behavioral experiments.** Experiment 1a examined humans' interpretation of number words. We used the same set of items, price states $S$, and utterances $U$ as described earlier. Subjects read scenarios in which a buyer produces an utterance $u$ about the price of an item he just bought. Subjects rate the likelihood of the buyer thinking that the item was expensive, as well as the likelihood that the item actually cost $s$ dollars (see Experiment 1a in materials). Figure 5 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 ($F(1, 10) = 44.06, p < 0.0001$).

Experiment 1b examined 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 $u >= s$. Subjects rate the likelihood of the buyer thinking that the item was expensive (see Experiment 1b 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.

**Comparison.** We compared model and human interpretation probabilities for the 300 items ($I \times U \times S$) and show that model predictions correlate significantly with human interpretations of number words ($r = 0.968$, $p < 0.0001$) (Figure 1). We

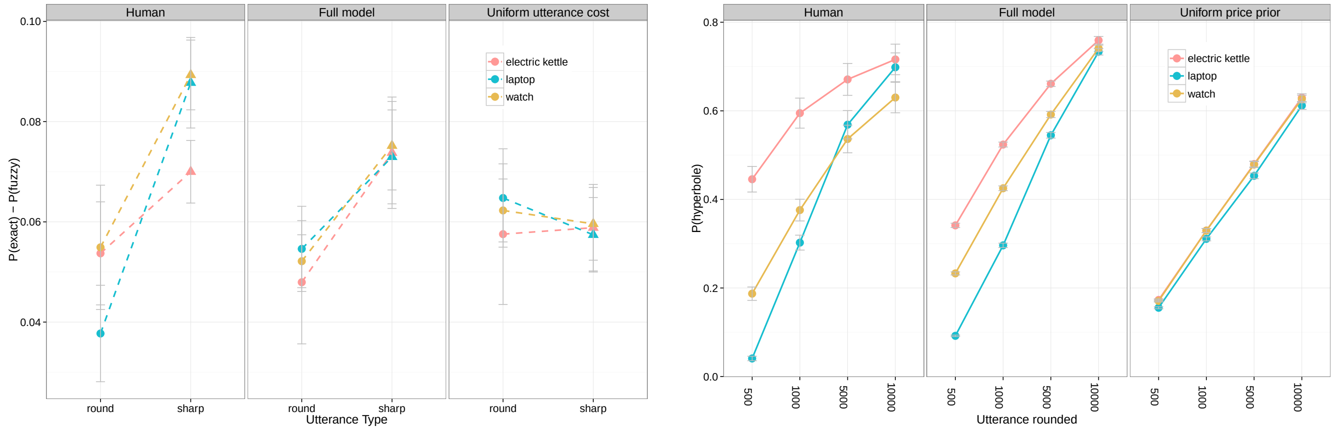


**Fig. 2.** (a) Compares pragmatic halo. (b) Compares hyperbole.

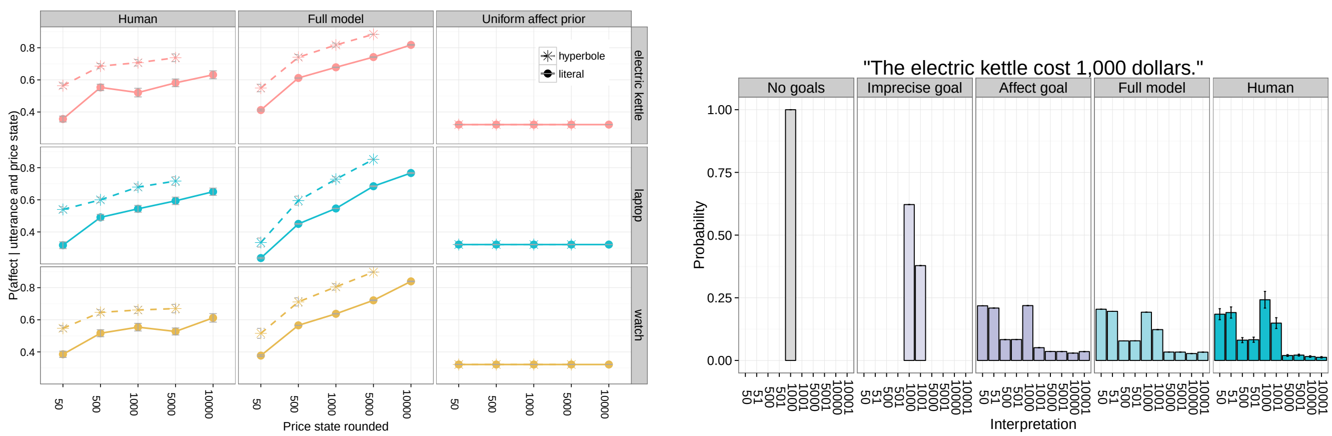


**Fig. 3.** (a) Compares affective subtext. (b) Compares communicative goals.

also show that each component of the model is responsible for capturing the effects that we observe in the human data.

The left-most panel in Figure 3 shows the halo effect in humans, namely 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. This suggests that people take into account utterance cost when inferring exact versus fuzzy interpretations. The leftmost panel in Figure 4 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 5 show humans' probabilities of inferring affective subtext given a price state and a literal or hyperbolic utterance. Within the same price state, hyperbolic utterances are interpreted as conveying more affect than literal utterances. Across price states, higher price states are associated with higher degrees of affect. 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.

Finally, we show that reasoning about a speaker's communicative goals is crucial for capturing the non-literal interpretation of number words demonstrated in the behavioral experiment. Figure 6 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.

## Discussion

Our model and behavioral results reveal complex patterns in the non-literal interpretation of number words. We show that 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 introduce an "affect" dimension that takes into account social and subjective aspects of communication. Our model shows that rational recursive reasoning between speaker and listener can capture rich effects in language understanding that have been difficult to model formally. We present an innovative formal framework for examining phenomena in non-literal language understanding, and provide evidence that this framework and its extensions will advance language research such that some day–hyperbolically speaking–Bayesian models can explain "*everything*."

## 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 $M = S \times A$. We denote each possible meaning as $m_{s,a}$, where $s \in S$ and $a \in A = \{0, 1\}$ ($0$ 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 $g_{\mathbf{s},\mathbf{a}}$ where $\mathbf{s} \in 2^S$ and $\mathbf{a} \in 2^A$. The goal $g_{\mathbf{s},\mathbf{a}}$ is a function $g_{\mathbf{s},\mathbf{a}} : M \to \{0, 1\}$, 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 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. $\mathbf{s} = \{i\}$ (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. $\mathbf{s} = \{j | j \in S, |j - i| \leq 1\}$.

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 2a and 2b). The probability distribution $P_G(\cdot|m_{s,a})$ over goals given that teh speaker knows meaning $m_{s,a}$ is defined to be uniform over goals consistent with $m_{s,a}$, i.e. uniform over goals $g_{\mathbf{s},\mathbf{a}}$ such that $g_{\mathbf{s},\mathbf{a}}(m_{s,a}) = 1$. This is equivalent to assuming that the speaker either wants to communicate their 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:

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

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]:

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

where the constant $\lambda$ captures the degree of optimality of the speaker. (We used $\lambda = 1$ 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 $U_n$ is therefore defined by:

$$U_n(u|g_{\mathbf{s},\mathbf{a}}) = \log(P(g_{\mathbf{s},\mathbf{a}}(L_n(m|u)) = 1) - C(u) \qquad [2]$$

which combined with equation 1 leads to:

$$S_n(u|m, g) \propto (P(g_{\mathbf{s},\mathbf{a}}(L_n(m|u)) = 1) \cdot e^{-c(u)})^{\lambda}. \qquad [3]$$

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.

$$L_n(m_{s,a}|u) \propto \sum_g P_S(s) P_A(a|s) P_G(g|m_{s,a}) S_{n-1}(u|g). \qquad [4]$$

After obtaining a posterior distribution for all possible meanings $m$ given an utterance $u$, we performed a Luce choice transformation on the distribution with $\lambda = 0.5$ in order to compare with behavioral results.

**Experiment 1a: 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 $u \in \{50, 50 \pm k, 500, 500 \pm k, 1000, 1000 \pm k, 5000, 5000 \pm k, 10000, 10000 \pm k\}$, where $k$ was randomly selected from the set $\{1, 2, 3\}$ for each trial. We will refer to this set of utterances as $U$. 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: $50$, $50 \pm k$, $500$, $500 \pm k$, $1000$, $1000 \pm k$, $5000$, $5000 \pm k$, $10000$, $10000 \pm k$, 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 $1$. 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 5.

**Experiment 1b: 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 $u \in U$ and $u >= s$. 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 2a: 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 $s \in S$ (see Experiment 1a). 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 2b: 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 ($s \in S$). 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.

## Appendix: Appendix title

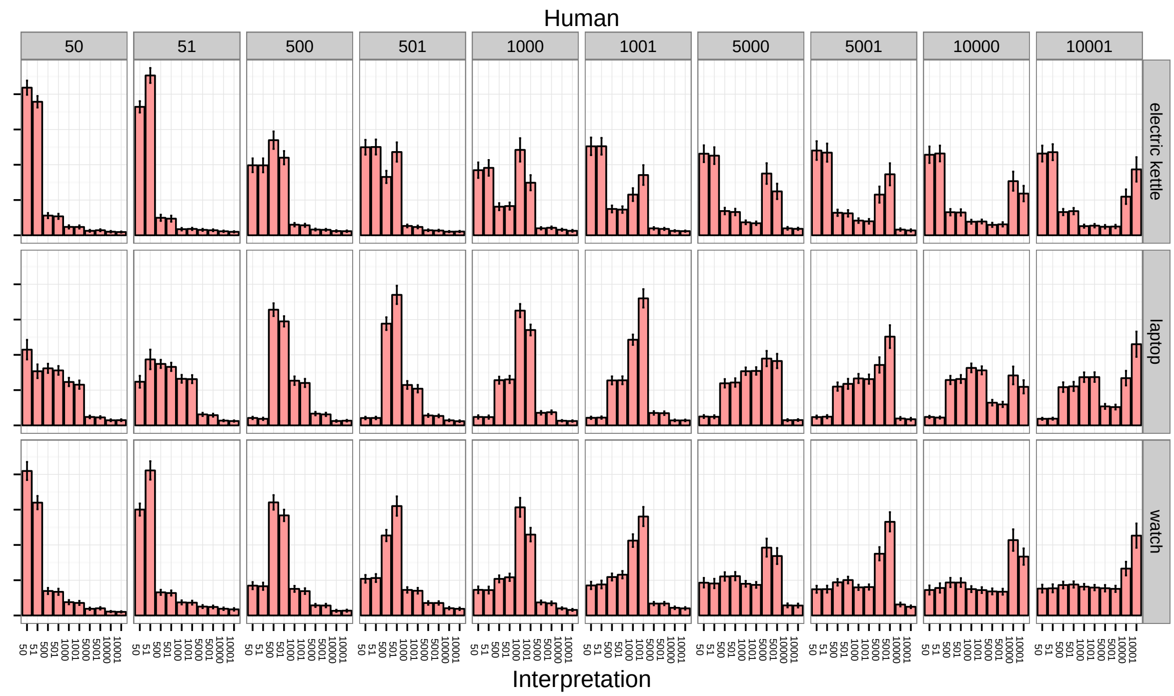


**Fig. 5.** Hello

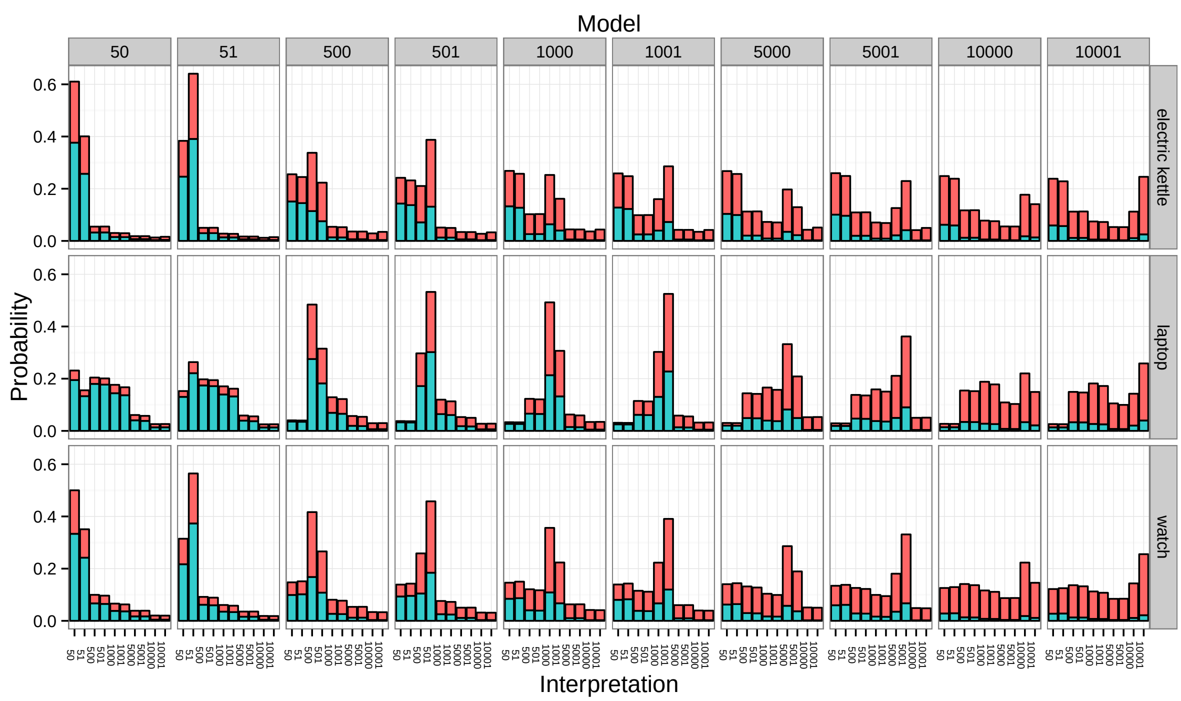


**Fig. 4.** Hello



**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)