# sorites update

## erin

### March 26, 2015

## 1 intro

We consider the following argument:

**Premise 1:** The Sears Tower is tall.

**Premise 2:** A building that is 1m shorter than a tall building is tall.

**Conclusion:** Every building in Chicago is tall.

When people hear this argument, they tend to think that premise 1 (the "concrete" premise) and premise 2 (the "inductive" premise) are both clearly true, but that the conclusion – which would naturally follow from the premises according to first order logic – is clearly false.

We explain people's reactions to these statements with a formal model of scalar adjective interpretation and show that their reactions are sensitive to the prior distribution on building heights and to the change in height $\varepsilon$ given in the inductive premise (in this case, 1m).

## 2 model

We use a model of scalar adjectives (e.g. "tall") in which a pragmatic listener must infer both the height that the speaker intended to communicate and also the threshold between "tall" and "not tall" that the speaker was using (Lassiter 2013). Because a speaker is likely to be communicating about things that are plausible in the world, but is unlikely to expend effort on uttering something uninformative, the pragmatic listener will infer that the threshold is high enough to be informative, but low enough for the communicated height to be plausible. This inference will depend on the speaker and listener's common ground on what plausible and likely heights are for a given category.

### 2.1 initial simulations of adjective model and effects of parameters

We can look at this model's inferences about the threshold $\theta$ and the degree $x$ (in the previous example, height) when the prior distribution is roughly gaussian with mean 50 and standard deviation 20. Different rationality parameters $\alpha$ and different costs $C$ for the scalar adjective utterance (e.g. "tall") result in different posterior distributions on $\theta$ and $x$.
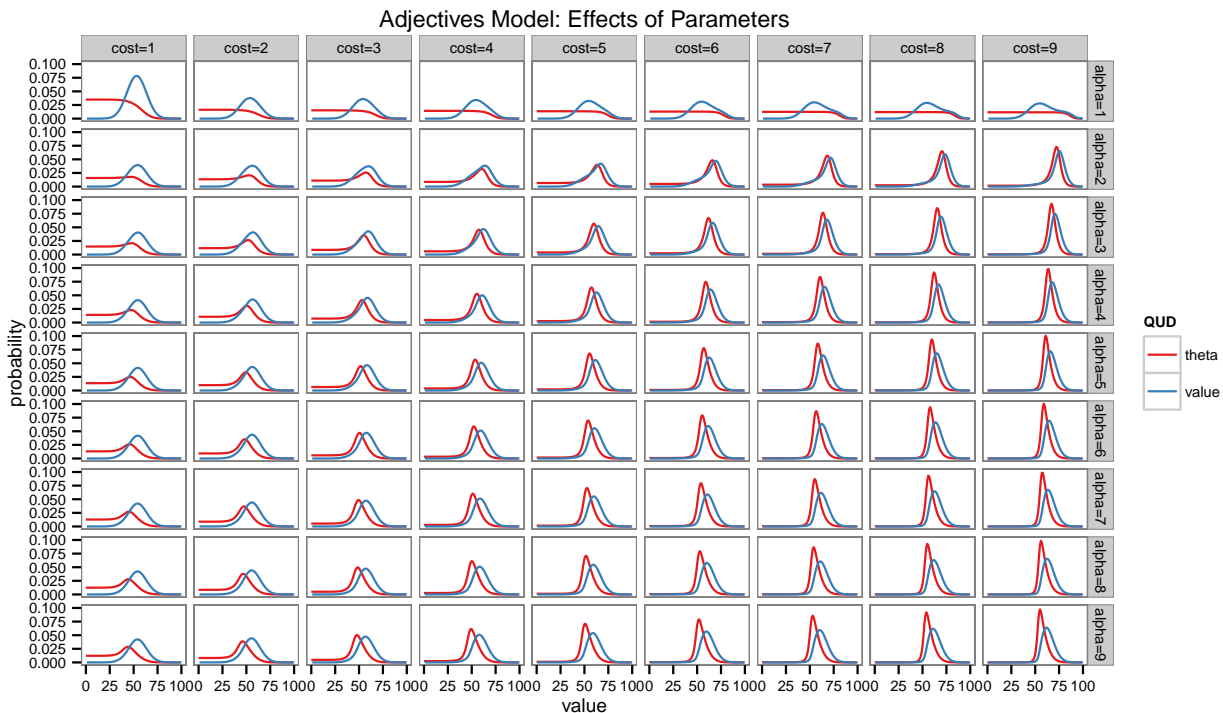
As cost increases, the model infers the threshold variable $\theta$ and the degree $x$ to be higher. As rationality parameter $\alpha$ increases, the model's posterior becomes peakier, moving probability mass away from unlikely values and towards higher probability ones.

## 2.2   applying adjective model to sorites

In order to interpret the sorites argument, our model first infers a value of the threshold variable $\theta$ and the degree $x$, conditioned on a speaker saying the scalar adjective. The model then evaluates the following expressions:

**Premise 1:** $height(\text{Sears Tower}) > \theta_{tall}$

**Premise 2:** $height(x) - 1 > \theta_{tall}$

**Conclusion:** $\forall y \in \{\text{buildings in Chicago}\}, \; height(y) > \theta_{tall}$

The expressions for premises 1 and 2 are likely (but not guaranteed) to be true. The expression for the conclusion, however, is unlikely to be true. This interpretation of the sorites argument predicts that as the amount of change in height $\varepsilon$ (in this case 1m) changes, endorsement of the inductive premise will change, since the probability of $height(x) - \varepsilon > \theta_{tall}$ becomes lower as $\varepsilon$ increases. We test this prediction, as well as sensitivity to the prior distribution.

# 3   prior elicitation(s)

In order to compare model predictions to humans, we chose to look at prices of everyday items. This is a domain where humans have relatively consistent prior knowledge and where the scale is representable as numeric quantities.

We first measured people's prior distributions over prices in a series of experiments.

## 3.1   basic setup

We gave people a number of bins, each bin representing a range of prices, and asked how likely an object would be to cost a price within each range. Within each of the three versions of this kind of prior elicitation experiment, we gave participants the same bins for each object. There were many bins, so there were multiple rows of sliders that participants had to mark.
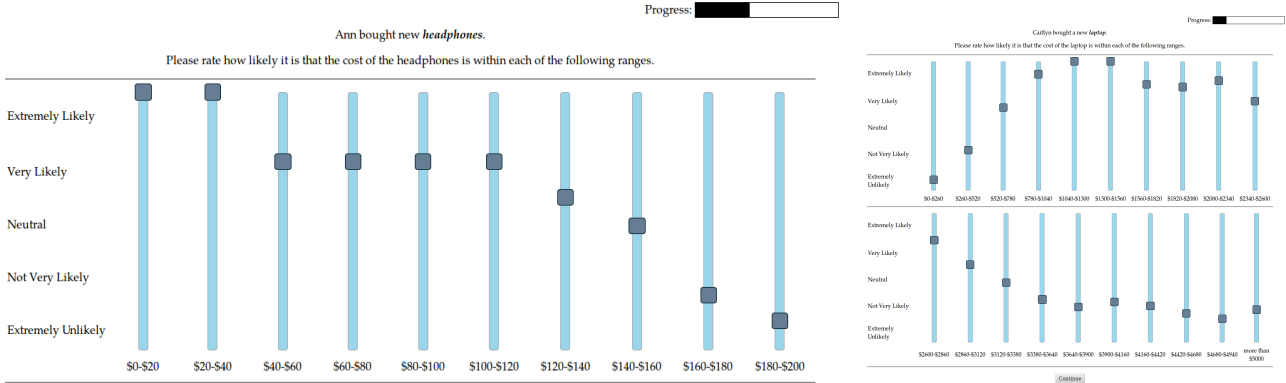


## 3.2   results

We ran three different versions of these experiments, each with different ranges and bin widths for the different items. We found similar responses across all of the experiments. Version A (10 participants, 1 outlier excluded) had 40 bins for each of the 5 objects, Version B (10 participants) had more bins (the number varied by object), and Version C (36 participants) had fewer objects in the experiment design. Because we had the best coverage of objects and ranges in Version B, and because its results are similar to those found in other versions, I use the prior elicitation data from Version B in the model simulations that follow.

Prior Elicitation Experiments
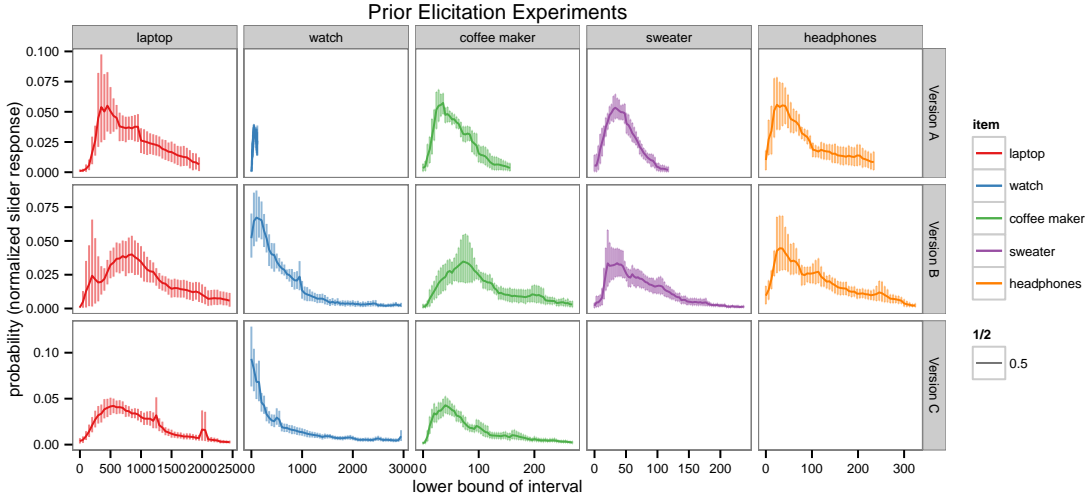
# 4 sorites experiment

We ran two different versions of the sorites experiment, with two different phrasings. In both versions, participants saw two different kinds of questions regarding 5 different categories of objects. These questions were all randomly intermixed. One of the kinds of questions represented the concrete premise and one represented the inductive premise. Participants were asked to rate each of these questions on a 9-point Likert scale from "Completely disagree" (1) to "Completely agree" (9).



**Headphones that cost $0.79 less than expensive headphones are also expensive.**

Please indicate how much you agree with the above statement.

Completely disagree   ○ ○ ○ ○ ○ ○ ○ ○ ○   Completely agree

Continue

The two versions of the experiment differed in their phrasing of the inductive premise. Version 1 (N=70) used the relative clause phrasing "An X that costs \$$\varepsilon$ less than an expensive X is also expensive". Version 2 (N=50) used the conditional statement phrasing "If an X is expensive, then an X that costs \$$\varepsilon$ less is also expensive". The results for these two versions of the sorites experiment had a correlation of 0.98.
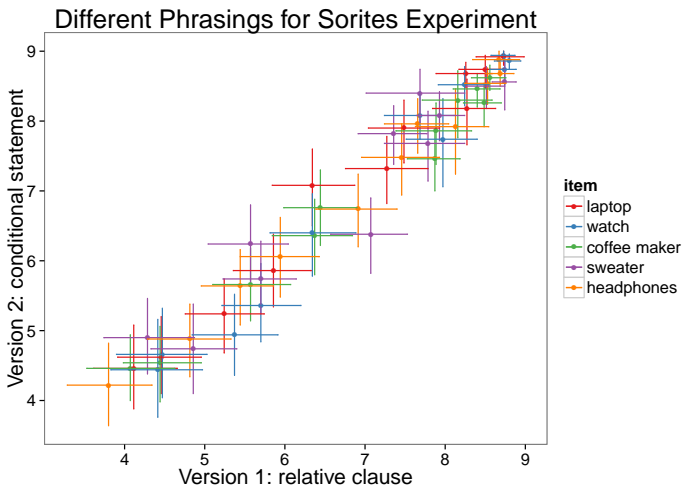


For the remainder of the analysis, I will use data aggregated from both versions of the experiment.

We varied the amount $\varepsilon$ less expensive that the item in the inductive premise was, the price $V$ given in the concrete premise, and the category of item (we used headphones, laptop, watch, coffee maker, and sweater). As predicted, participants gave graded judgements that varied by the amounts $\varepsilon$ and $V$ and by the category (different categories have different distributions over prices).

Sorites Experiment
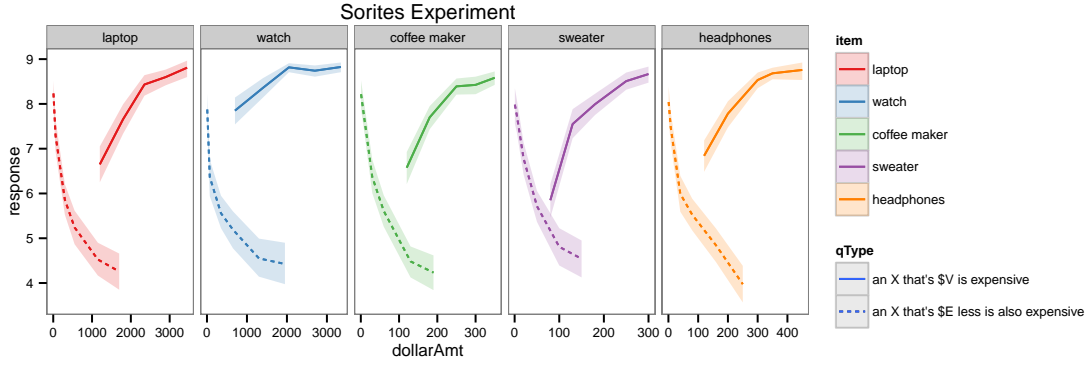
# 5 simulations

We ran the sorites model with 3 different set sof parameter values ($\alpha = C = 1$, $\alpha = C = 3$, and $\alpha = C = 5$). The simulation results were very similar, except that the range of endorsement of the concrete premise was wider for higher alpha and cost (i.e. the items we tested that had lower prices were given lower probability of being "expensive" by the model the higher alpha and cost were). I will show graphs for $\alpha = C = 1$.


Adjectives Model


Sorites Model
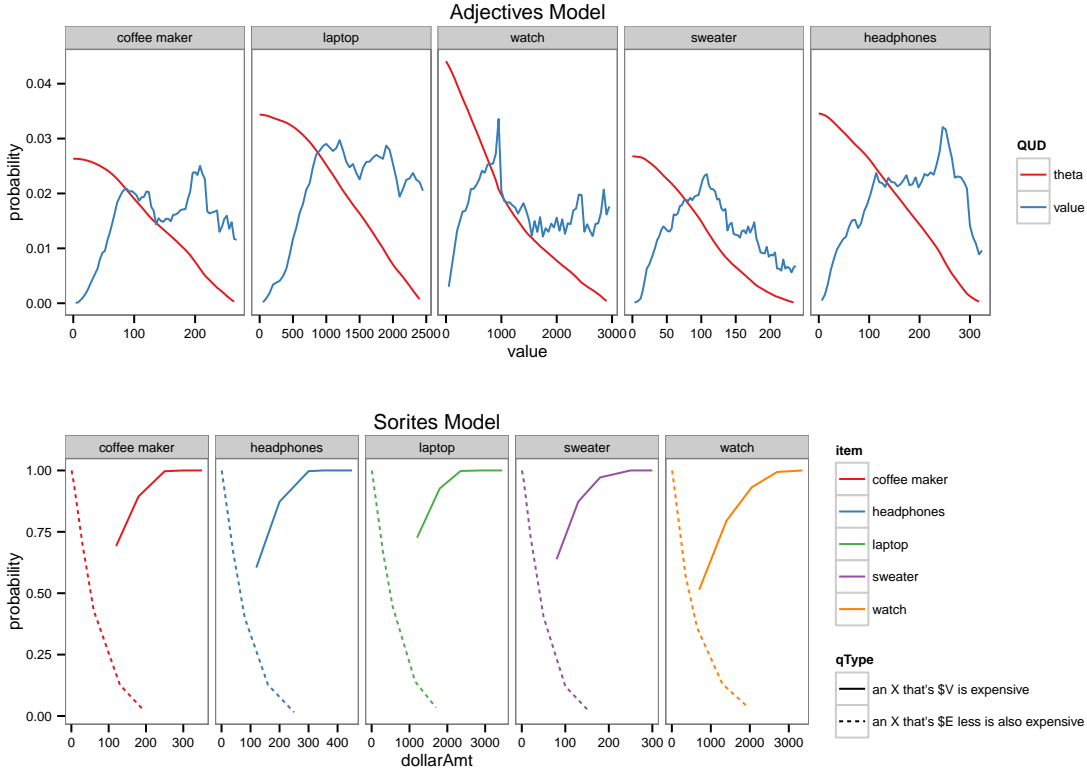
The model probabilities for the inductive premise correlated with experiment judgements on the inductive premise with Pearson's product-moment correlations of 0.96 across all parameter values. However, the model has a much wider range of probabilities than people's range of endorsements, i.e. the model goes all the way to 0 for high $\epsilon$, whereas people are still giving ratings as high as 4/9. This may be because our prior elicitation did not get as much probability mass in the tails of our distributions as people think is there. This would mean that people imagine more highly priced items than the model frequently enough that even with a high $\epsilon$, the inductive premise might still be true.
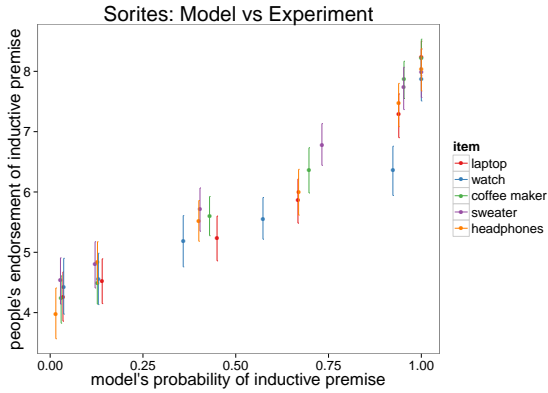
The correlation between model and human judgements on all premises aggregated together (including the concrete premises) is lower. For $\alpha = C = 1$, the correlation was 0.94; for $\alpha = C = 3$, the correlation was 0.93; and for $\alpha = C = 5$, the correlation was 0.91.
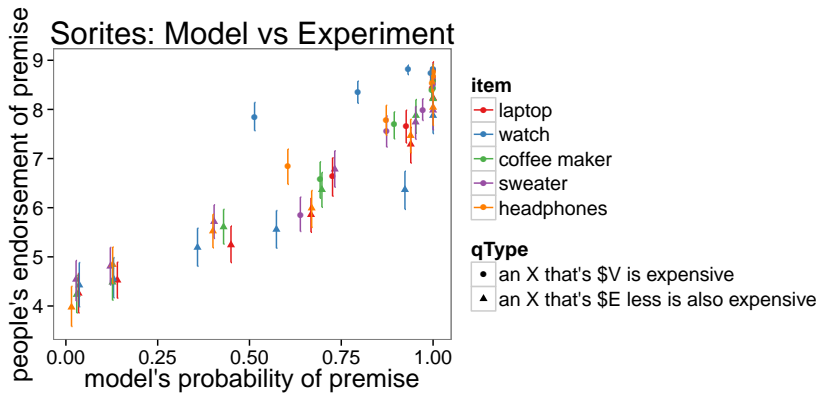


# 6   next steps

## 6.1   prior elicitation

Although its seems from the multiple versions of the prior elicitation that we have a consistent measure of people's intuitions, the N actually used for the model was only 10. We might want to increase that. The support for the model's prior also had an upper bound, which real prices do not. We might want to find a way of eliciting further into the tails of these distributions. Eliciting cummulative distribution functions (e.g. "How likely is it that a particular watch costs more than $100") might be useful here, and so would backing out the prior from multiple different measures.

## 6.2   model

**i'm missing model confidence intervals**   I intend to get bootstrapped confidence intervals on the model graphs, but it's computationally intensive, we might very well get more prior data, and I wanted to send this out, so I'm leaving out the confidence intervals for now.

**parameters**   It's also unclear to me what the best parameters are for the model.

**simpler models**   It might be nice to compare this to simpler models.