Sorites

summary by Erin Bennett

2015 June 5

Contents

Sorites

High-level summary

We ran Experiment 1 to get people's endorsements of the sorites premise with various parameters, then a prior elicitation (Experiment 2), then re-ran the sorites experiment with values that were based on cumulants of the elicited prior experiment in Experiment 3.

The data from Experiment 3 for the inductive premise seemed kind of noisy, so I simplified the language (relative clause phrasing instead of conditional phrasing) and ran Experiment 4. I think the meanings are

the same, since we have pretty high correlations between versions of the experiment with these two different phrasings.

The model fits the data from Experiment 4 very well, except for the concrete sorites premise on watches. People endorse "A watch that costs $X is expensive" more than the model predicts. People's aggregated responses also seem to have a narrower range than the model's probabilities.

Model

Adjective Model

Literal Listener's probability distribution over the values X are is prior, conditioned on the utterance being true and renormalized.

$$P_{L0}(x|u,\theta) \propto \delta_{u \text{ is true}} \cdot P(x)$$

Speaker's utility is the negative cost and the log probability of the actual state of the world under the Literal Listener's posterior. This means that the more surprised the Literal Listener would be to hear the true state of the world after already hearing the utterance, the less good the utterance would be.

$$\mathbb{U}_S(u|x,\theta) = log(P_{L0}(x|u,\theta)) - cost(u)$$

The speaker then chooses an utterance by soft-maximizing their utility funciton.

$$P_S(u|x,\theta) \propto e^{\lambda \mathbb{U}_S(u|x,\theta)}$$

The pragmatic listener infers both the threshold θ and the value x conditioning on the speaker choosing the given utterance.

$$P_{L1}(x,\theta|u) \propto P_S(u|x,\theta)P(x)P(\theta)$$

Sorites premises

Inductive premise: "An *OBJECT* that costs $E less than an expensive *OBJECT* is still expensive."

$$P(X - \varepsilon > \theta|X,\theta)$$

Where X is the inferred price of an "expensive" *OBJECT*.

Concrete premise: "An *OBJECT* that costs $X is expensive."

$$P(X > \theta|X,\theta)$$

Alternative utterances

It's not clear what the alternative utterances in the adjective model should be. I use "adjective" and "say nothing" for computational ease when fitting parameters in the simulations.

TO BE CONTINUED...

I could next run the model, with these best parameter settings, for a few different sets of alternative utterances:

- "adjective", "say nothing", "opposite of adjective" (with its own threshold, same cost)
- "adjective", "say nothing", "opposite of adjective", "not adjective" (same threshold as "adjective"), "not opposite of adjective" (same threshold as "opposite of adjective")
- "adjective", "say nothing", "intensifier adjective" (with its own threshold, double the cost)
- "adjective", "say nothing", "intensifier adjective", "opposite of adjective", "intensifier opposite of adjective"

According to our discussion in lab meeting about alternatives probably being less complex than the actual utterance, the minimal alternative set ("adjective" and "say nothing") and the alternative set containging the opposite of the adjective ("adjective", "say nothing", and "opposite of adjective") are probably the most plausible.

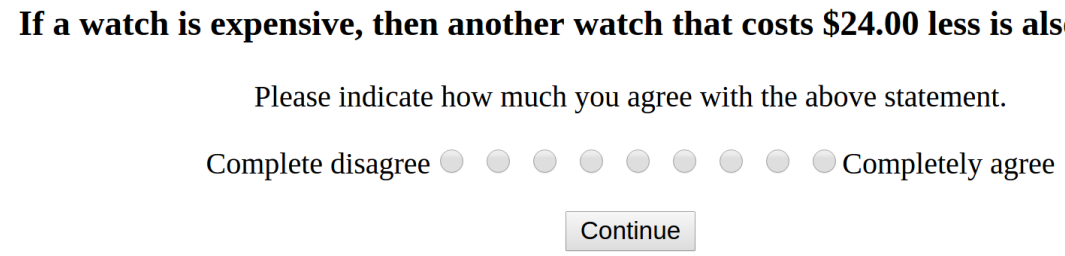
Model parameters

The free parameters in our model are the rationality parameter λ and the cost of the adjective utterance C. In our simulations, we allow both parameters to vary from 1 to 10 and fit to the combination of parameter settings that yields the smallest distance between model predictions and human responses for the sorites inductive premises.

Experiment 1: Endorsement of sorites premises

We ran two different versions of the first 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" (-4) to "Completely agree" (4).

There were two different phrasings for this experiment, which produced similar ratings from participants. (More discussion in the 2015 March 26 update.)



We varied the amount ε less expensive that the item in the inductive premise was, the price X given in the concrete premise, and the type of item. As predicted, participants gave graded judgements that varied by the amounts ε and X and by the items (different items have different distributions over prices). The exact values of the εs and Xs were chosen based on scraped prices from ebay and amazon.

More analysis of this data is detailed in `sorites_2015march26.pdf`.

Experiment 2: Priors on prices (binned histogram)

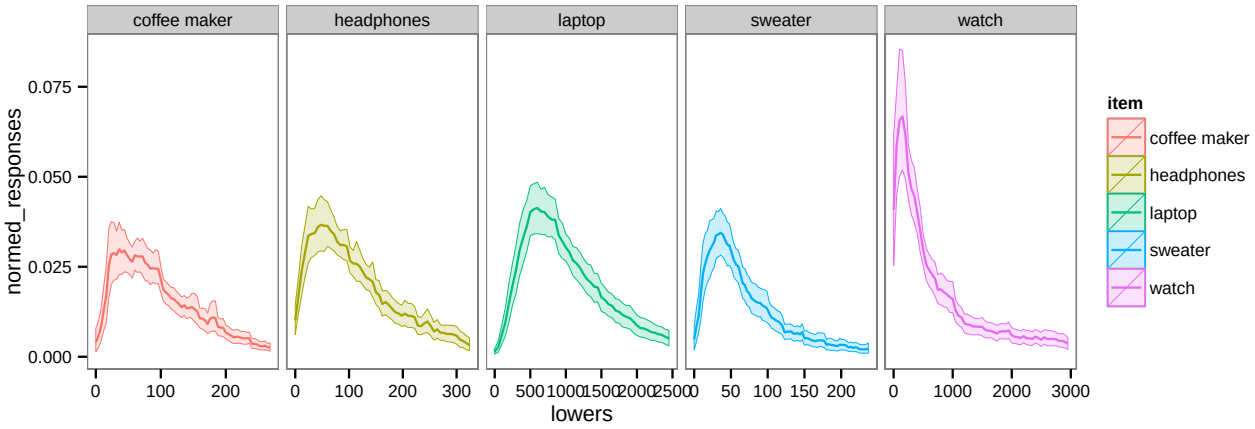
We showed 30 participants this experiment.

For each of 5 items (*coffee maker*, *laptop*, *headphones*, *watch*, and *sweater*), participants saw a page full of vertical sliders (between 50 and 80 sliders in total, depending on the item) in rows of 10 sliders each. Participants were asked to indicate, using the sliders, how likely they thought it was that the price of the

item was within various ranges. The ranges (e.g. \$0-\$50, \$50-\$100) were arranged in order from the lowest range (starting at \$0) to the highest (greater than some maximum dollar amount, which depended on the item). Participants' responses were coded from 0 ("Extremely Unlikely") to 1 ("Extremely Likely").

We chose the price ranges for each item based on pilot experiments. We wanted sufficient detail about the tails of the distributions, so we chose maximum values for each item such that the average endorsement of the highest bin was very low (in our experiment, the average endorsement for the highest bin was always less than 0.15). We also wanted sufficient granularity to address the sorites inductive premise, even for very small ε. We threfore chose the width of the bins so that, for every price X and for every ε in our original sorites premises experiment, we could confidently estimate the probability of an item ε less expensive than X. The resulting distributions are fairly smooth, allowing us to interpolate within the bins as needed. Our level of resolution also allowed us to capture detail in peaky parts of the distributions (usually the smaller ranges).

Participants' repsonses, normalized to represent probabilities, are shown in the graph below. We used these data as the prior distributions for our simulations.



Some potential problems with this experiment design

For each item, we always showed the same bins, and the same linebreaks between rows of bins. So for every participant, some pairs of adjacent bins were always farther away from each other than every other pair of adjacent bins were. This doesn't seem obvious from the responses that people gave, but it could have caused variance in responses or confusion for participants. We could have avoided this by making the sliders horizontal rather than vertical.

Experiment 3: Endorsement of wider range of sorites premises

link to Experiment 3

In our original sorites experiment, participants on average never rejected any of the premises very strongly. We wanted to see if giving participants a wider range of values for the premises would change that.

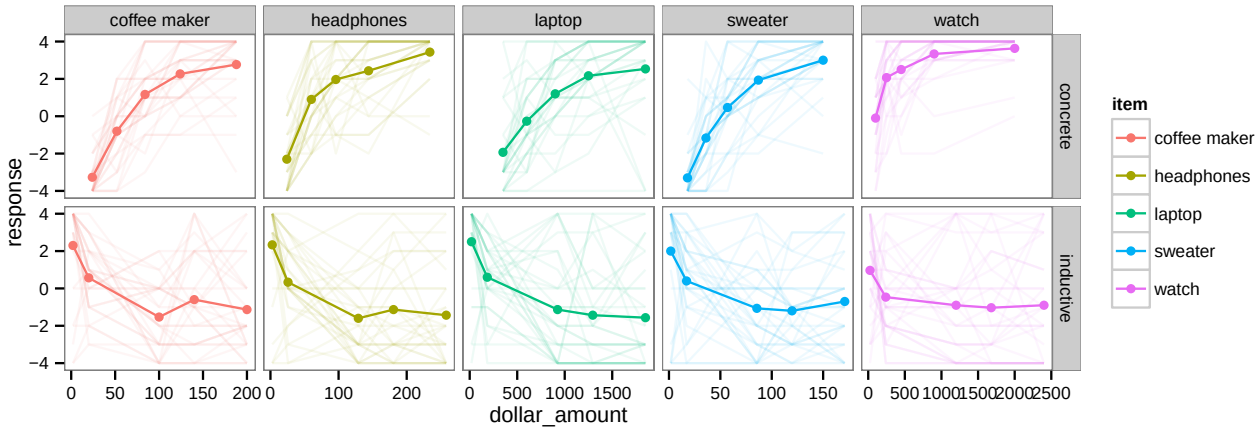
We wanted a wide range of prices for the concrete sorites premise, so that some would be obviously expensive and others would be obviously not expensive. For each item, and for each cumulative probability 0.1, 0,3, 0.5, 0.7, and 0.9, we approximated a price with that cumulative probability by chosing the highest price below that cumulative probability (e.g. since the sum of the average normed slider values for all bins up to "\$18-\$21" for *sweater* was the first such sum that was greater than or equal to 0.1, we chose the price \$18 as the smallest price for the concrete premise for sweaters). The prices were usually a little bit lower than the actual prices with those cumulative probabilities, but within a bin-width of the actual price with that cumulative probability.

We also wanted a wide ranges of prices for the inductive premise. For these, we chose proportions (0.01, 0.1, 0.5, 0.7, and 1) of the approximate 90% confidence intervals (calculated by subtracting the 0.05 comulative probability as above from the 0.95 cumulative probability).

```
##           item X0.1 X0.3 X0.5 X0.7 X0.9 E0.01  E0.1   E0.5   E0.7   E1
## 1 coffee maker   24   52   84  124  188  2.00  20.0  100.0  140.0  200
## 2   headphones   24   60   96  144  234  2.58  25.8  129.0  180.6  258
## 3       laptop  350  600  900 1250 1850 18.50 185.0  925.0 1295.0 1850
## 4      sweater   18   36   57   87  150  1.71  17.1   85.5  119.7  171
## 5        watch  100  250  450  900 2000 24.00 240.0 1200.0 1680.0 2400
```

Experiment 3 was otherwise identical to the version of Experiment 1 with conditional statements.

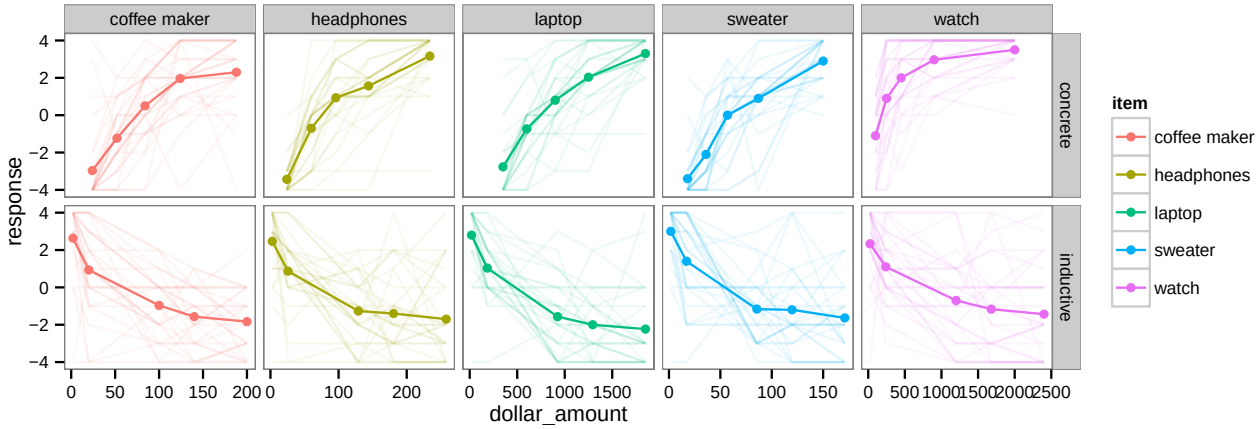
Looking at the individual responses (the lighter lines in the plot), it seems like there was more variation in participants' responses to the inductive premise questions, perhaps because participants did not understand the prompt as well.



Experiment 4: Relative clause version of Experiment 3

link to Experiment 4

On the off-chance that people were confused by the wording in Experiment 3, and that they would not be confused by the wording as much if I used relative clauses, I ran Experiment 4. Instead of using the conditional wording for the inductive premsise (e.g. "If a sweater is expensive, then another sweater that costs $171.00 less is also expensive."), I used the relative clause wording (e.g. "A sweater that costs $171.00 less than an expensive sweater is also expensive.").



The split-half correlation on the inductive premises for the relative clause version (0.9503) was higher than for the conditional version (0.8904) but not as high as the split-half correlation for the concrete premise in both experiments (0.9705 and 0.9646).

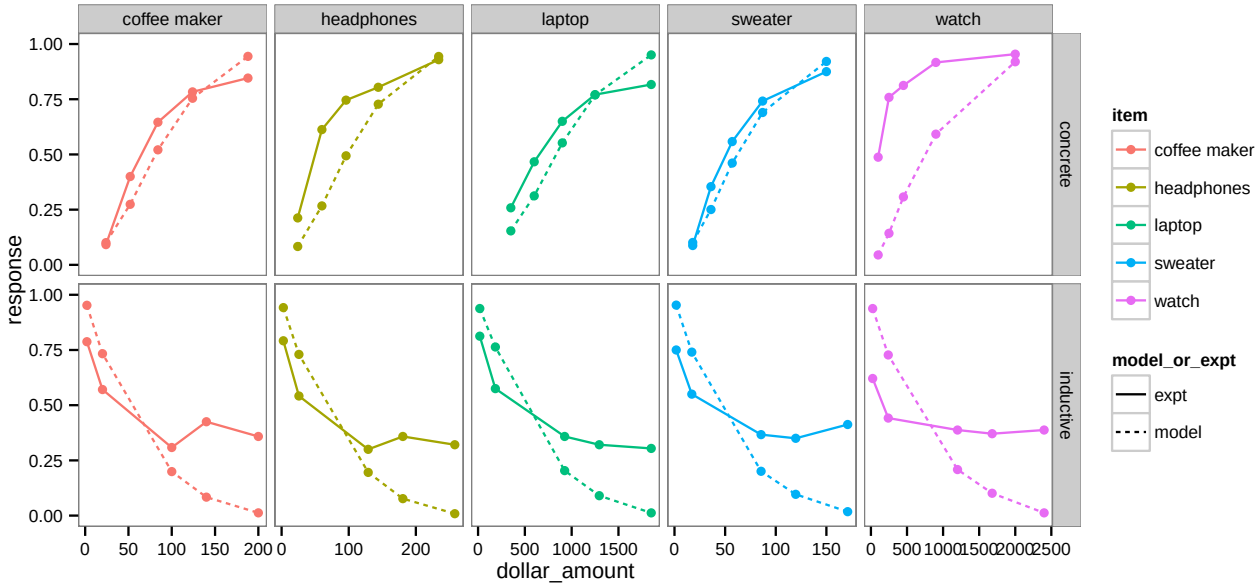
The overall correlation between responses to Experiment 3 and responses to Experiment 4 was 0.9411.

Simulations

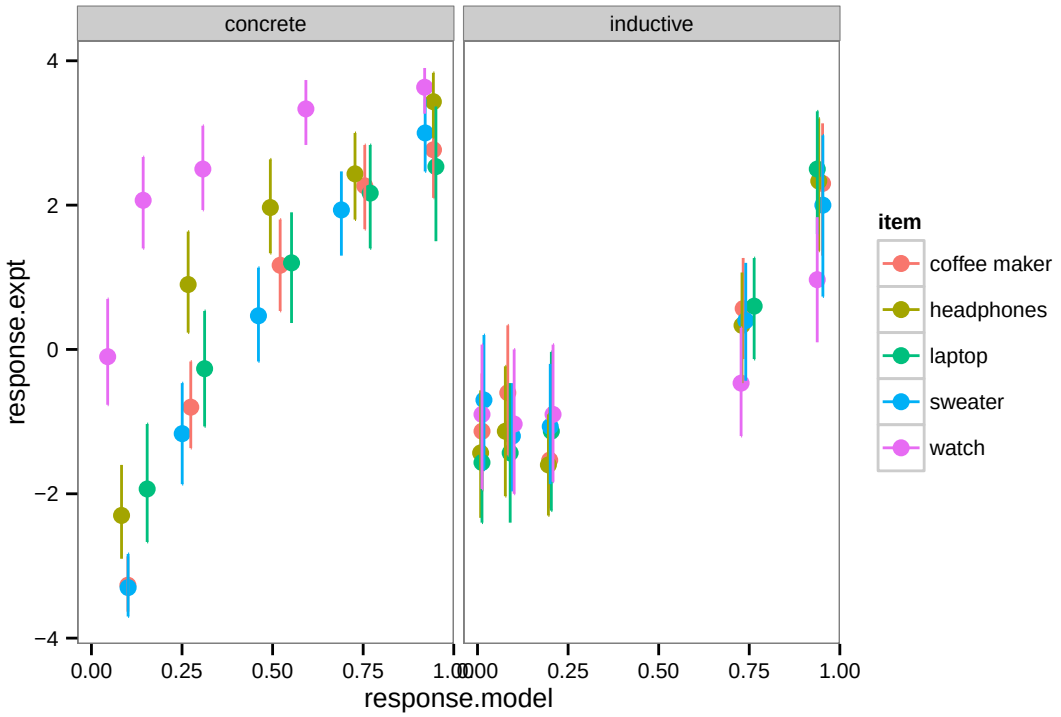
Model results fit to Experiment 3 (conditional)

For Experiment 3, where the phrasing was a conditional statement, the best-fit parameters were $C = 1$ and $\lambda = 6$ (I fit by choosing the parameters with the best correlation with the data). The lowest correlation between model and experimental data was 0.5344 and the highest correlation was 0.8008. The correlation for only the inductive premises with the parameters fit for all the experimental data was 0.9072.

The model probabilities with the highest-correlation parameters and the (rescaled to [0, 1]) responses from Experiment 3 (conditional statements) are shown below. (CONFIDENCE INTERVALS FOR MODEL NEEDED)



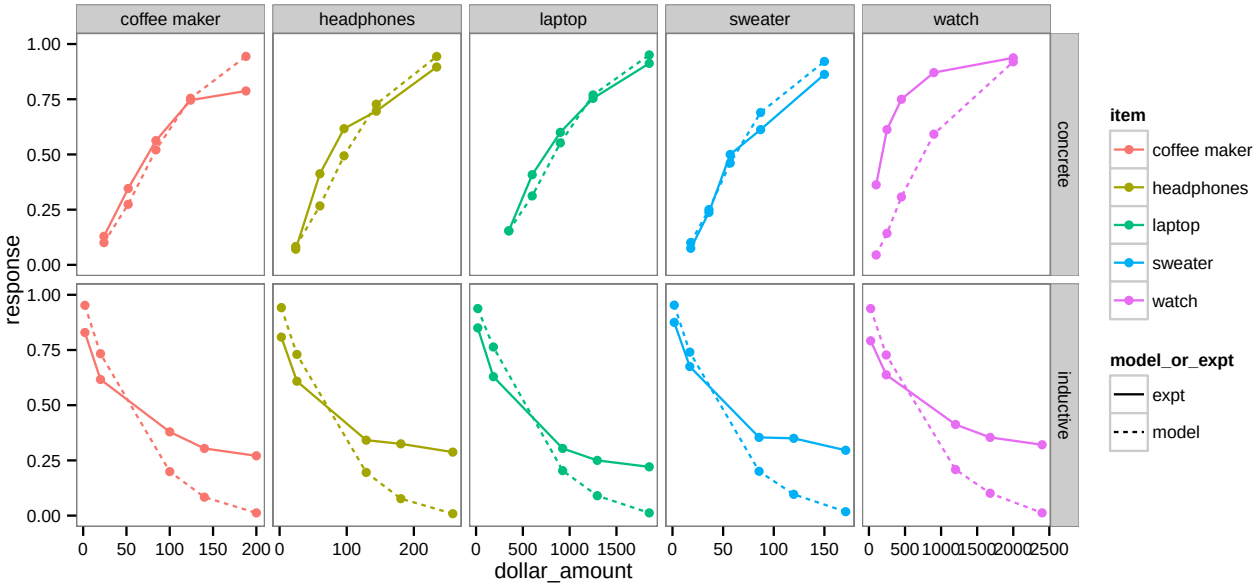
A scatterplot of the same data is shown below. (CONFIDENCE INTERVALS FOR MODEL NEEDED)



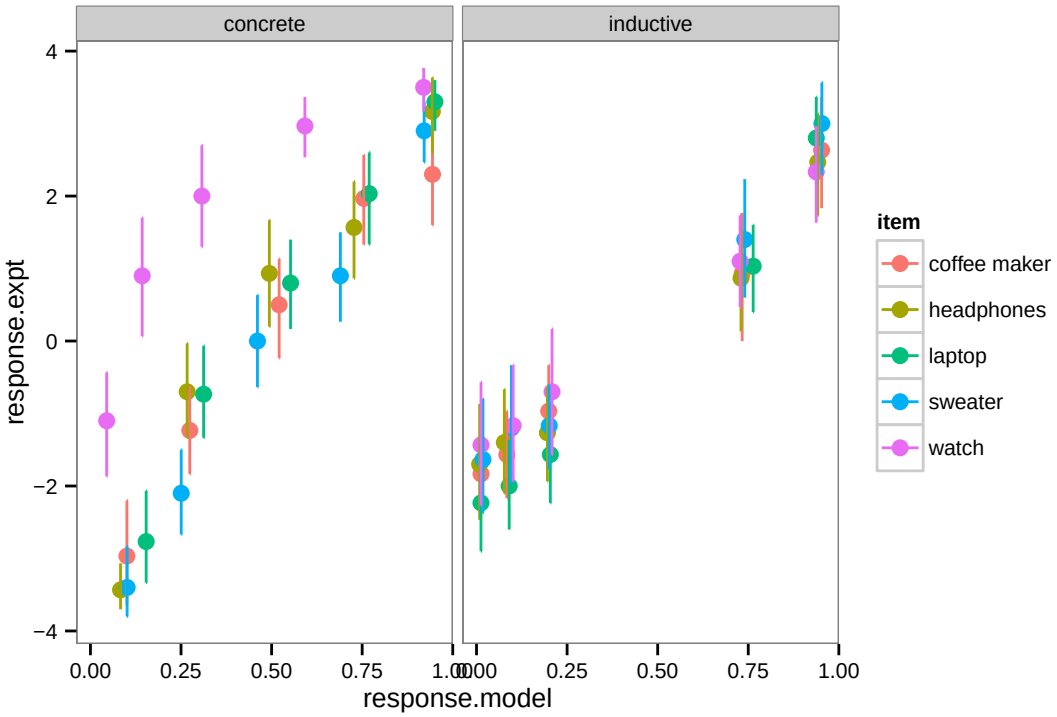
Model results fit to Experiment 4 (relative clause)

For Experiment 4, where the prhasing was a relative clause, the best-fit parameters were $C = 1$ and $\lambda = 6$ (I fit by choosing the parameters with the best correlation with the data). The lowest correlation between model and experimental data was 0.7271 and the highest correlation was 0.9062. The correlation for only the inductive premises with the parameters fit for all the experimental data was 0.9812.

The model probabilities with the highest-correlation parameters and the (rescaled to [0, 1]) responses from Experiment 4 are shown below. (CONFIDENCE INTERVALS FOR MODEL NEEDED)



A scatterplot of the same data is shown below. (CONFIDENCE INTERVALS FOR MODEL NEEDED)



Issues to resolve

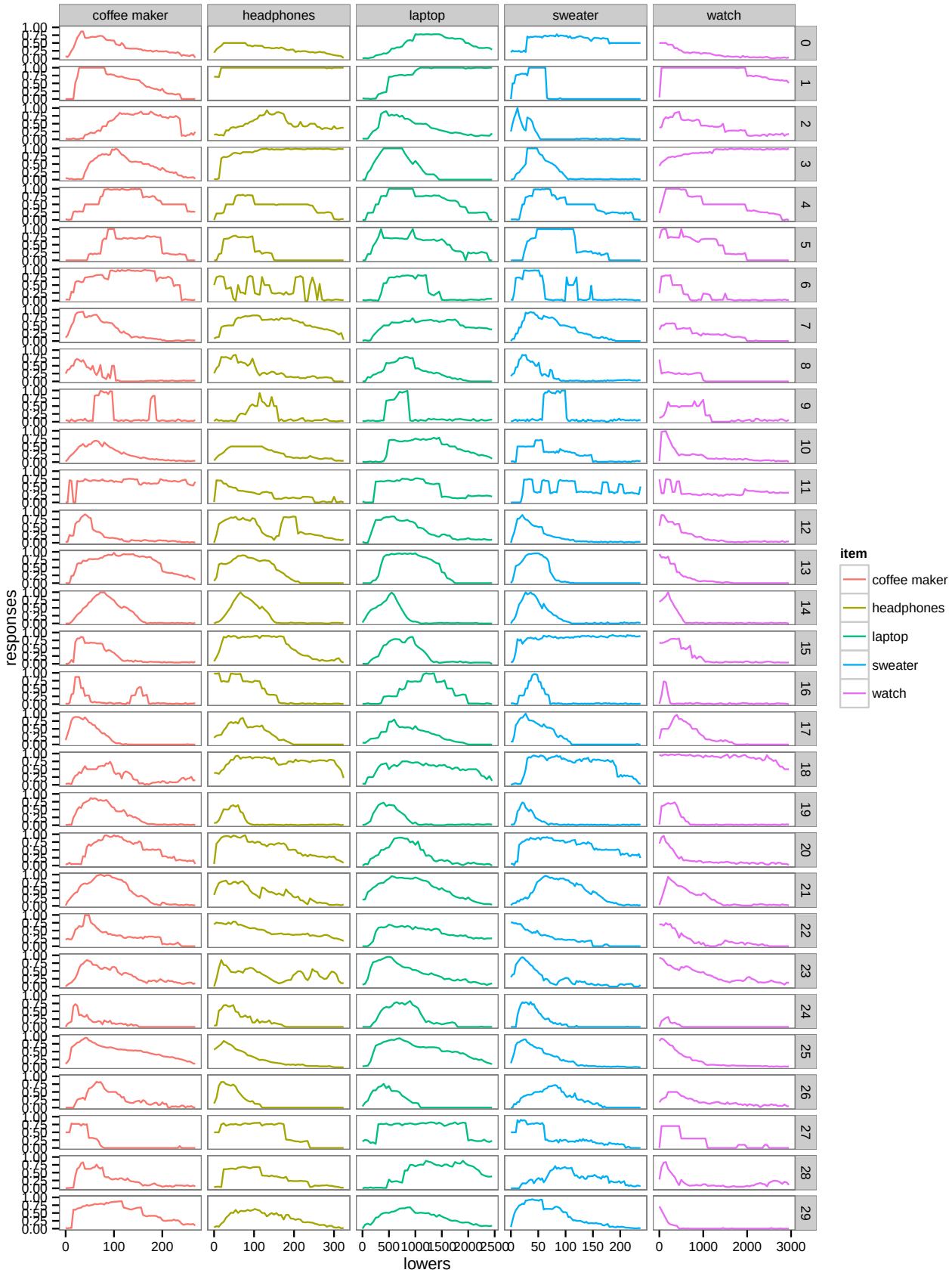
The concrete premise for watches is weird.

The model has a wider range of responses than people do.

More detail

Priors by worker

Here are our responses on the prior experiment by worker. At a glance, it looks like there are some individual differences on how people are using the scales but no one is giving answers that don't apear to be sensitive to the item and price (it looks like most poeple are not clicking randomly).



Participants 6 and 11 gave really jittery responses for some reason. Not sure why.

Prior resolution and range

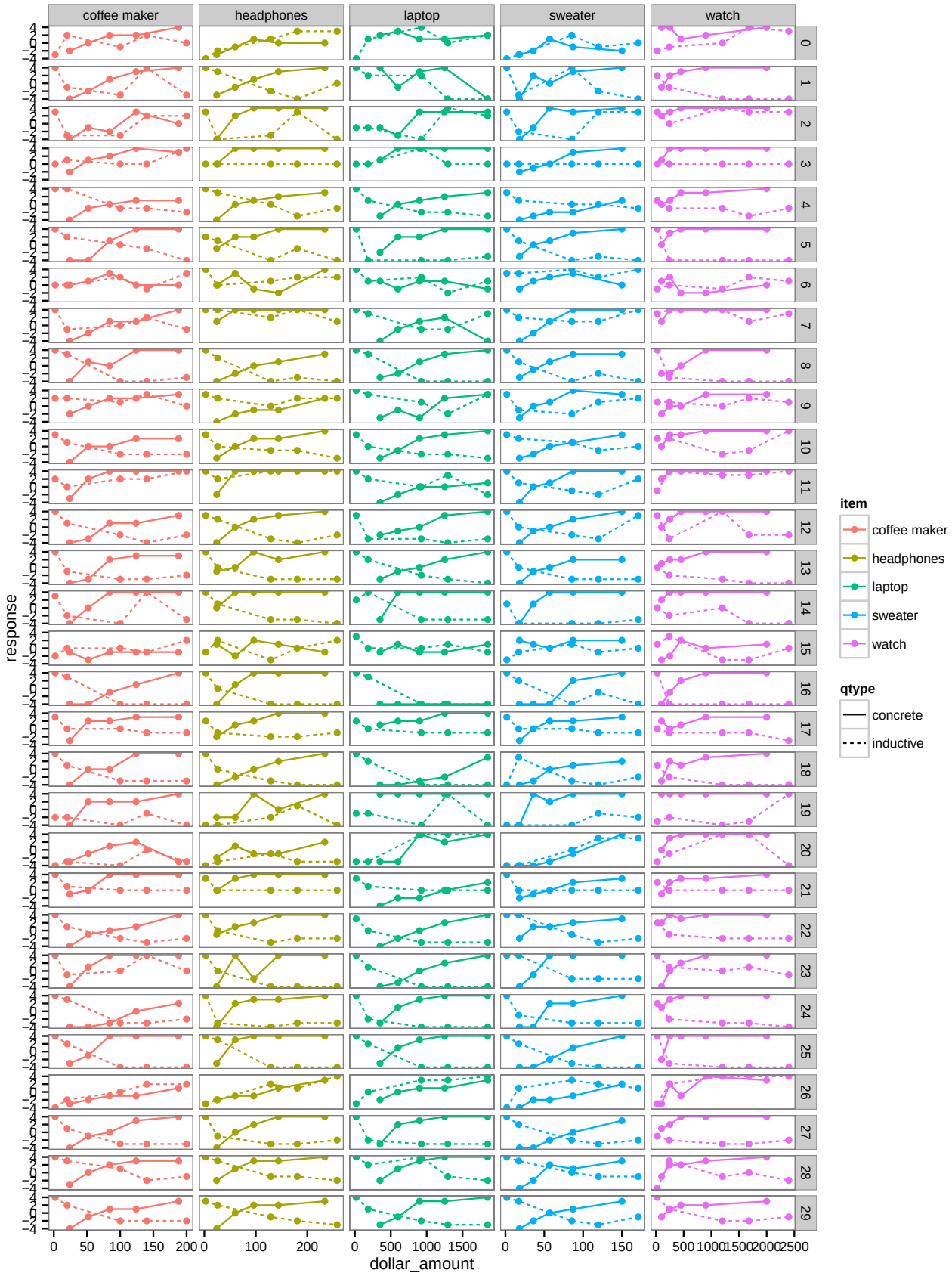
For all of the highest bins in the prior elicitation experiment, the average endorsement for that bin was less than 0.15.

```
##           item max_bin_price max_bin_response
## 1 coffee maker           268          0.07467
## 2   headphones           324          0.11267
## 3       laptop          2450          0.12400
## 4      sweater           237          0.08567
## 5        watch          2950          0.10167
```

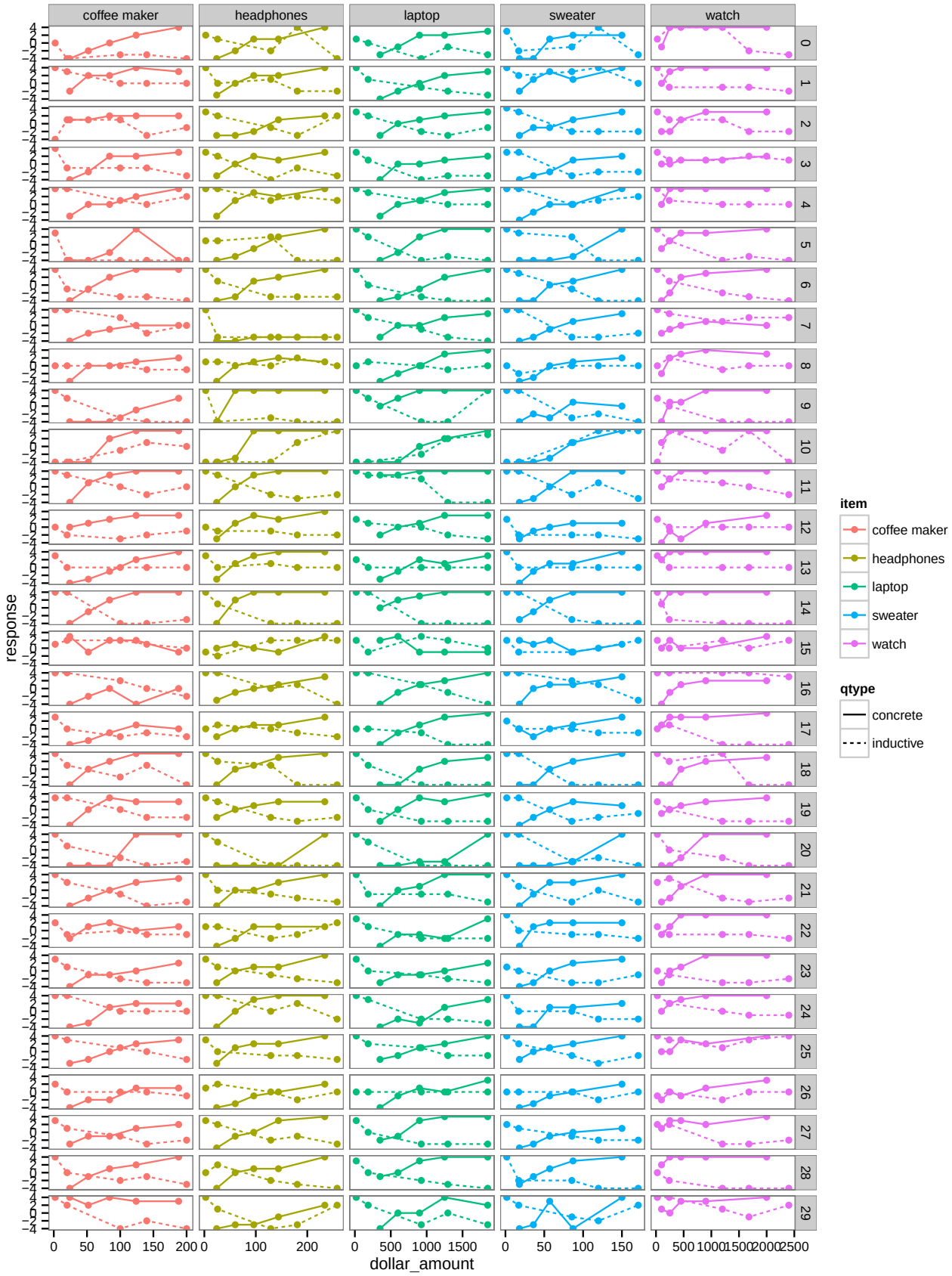
for each item, the smallest epsilons in Experiment 3 were always less than half and greater than a third of the bin width.

Sorites (Experiments 3 and 4) by worker

Here are the responses to the second sorites experiment by worker. It looks like maybe not all participants understood the task. For example, participants 0 and 26 gave the same responses for both the inductive and concrete premises.



When I used the relative clause ratings, it looks like people might be less confused.


coffee maker
headphones
laptop
sweater
watch
response
dollar_amount
item
coffee maker
headphones
laptop
sweater
watch
qtype
concrete
inductive

Model code in WebPPL

```
// webppl sorites.wppl --require-js ./simpleCSV.js

var alternatives = "adjnull";

var items = ["coffee maker", "headphones", "laptop", "sweater", "watch"];

var get_parameters = function(item) {
  var prior_csv_data = simpleCSV.readCSV("mean_priors.csv").data;
    var item_data = filter(
        function(trio) {
            return item == trio[0];
        }, prior_csv_data.slice(1, prior_csv_data.length - 1)
    );
    var values = map(function(trio) {return trio[1];}, item_data);
    var probabilities = map(function(trio) {return trio[2];}, item_data);
    return {
        "values": map( function(prob) {return global.parseFloat(prob);} , values),
        "probabilities": map( function(prob) {return global.parseFloat(prob);} , probabilities)
    };
}

var get_prior = cache(function(item) {
    var parameters = get_parameters(item);
    var values = parameters.values;
    var probabilities = parameters.probabilities;
    return function() {
        return values[discrete(probabilities)];
    }
})

var get_theta_prior = cache(function(item) {
    var values = get_parameters(item).values;
    return function() {
        return uniformDraw(values);
    };
})

var is_true = cache(function(utterance, value, theta) {
    if (utterance == "adj") {
        return value >= theta;
    } else if (utterance == "null") {
        return true;
    } else {
        console.log("err 55");
    }
})

var literal_listener = cache(function(utterance, theta, item) {
    return Enumerate(
        function() {
            var prior = get_prior(item);
            var value = prior();
```

```
            factor(is_true(utterance, value, theta) ? 0 : -Infinity)
            return value;
        }
    )
})

var utterance_prior = cache(function(cost) {
    return Enumerate(function() {
        var utterances = ["adj", "null"];
        var costs = [cost, 0];
        var probabilities = map(function(c) {return Math.exp(- c);}, costs);
        return utterances[discrete(probabilities)];
    });
})

var speaker = cache(function(value, theta, item, cost, lambda) {
    return Enumerate(
        function() {
            var utterance = sample(utterance_prior(cost));
            var literal_interpretation = literal_listener(utterance, theta, item);
            var score = literal_interpretation.score([], value);
            factor(score * lambda);
            return utterance;
        }
    )
})

var listener = function(utterance, item, cost, lambda) {
    return Enumerate(
        function() {
            var prior = get_prior(item);
            var theta_prior = get_theta_prior(item);
            var value = prior();
            var theta = theta_prior();
            var score = speaker(value, theta, item, cost, lambda).score([], utterance);
            factor(score);
            return [value, theta];
        }
    )
}

var costs = [1, 2, 3, 4, 5, 6];
var lambdas = [1, 2, 3, 4, 5, 6];

map(function(cost) {
    map(function(lambda) {
        map(function(item) {
            console.log("running " + item + ": " + alternatives + "_cost" + cost + "_lambda" + lambda);
            var listenerERP = listener("adj", item, cost, lambda);
            console.log("finished running " + item + ": " + alternatives + "_cost" + cost + "_lambda" +
            var model_output = {
                "data" : map(function(pair) {
                    var value = pair[0];
                    var theta = pair[1];
```

```
                    var score = listenerERP.score([], pair);
                    var probability = Math.exp(score);
                    return [item, cost, lambda, value, theta, score, probability];
                }, listenerERP.support())
            }
            simpleCSV.writeCSV(model_output, "model_output/" + item + "_" + alternatives + "_cost" + cos
            return 1;
        }, items);
        return 1;
    }, lambdas);
    return 1;
}, costs)
```

Varying alternative utterances

TO BE CONTINUED....

COMPARE MODEL TO EXPERIMENT 4 WITH DIFFERENT ALTERNATIVE UTTERANCES AND LAMBDA 6 AND COST 1.