Notes on sinking marbles

Judith Degen

August 5, 2014

Contents

1 Motivation

From Bart Geurts's book "Quantity Implicatures": "The cancellation argument presupposes that preferred scalar inferences will quietly withdraw whenever they happen to be implausible, and this pre- supposition is doubtful, as it is contradicted by the behaviour of bona fide scalar inferences. Let me give a couple of examples to establish this point.

(1) Some of the liberal parliamentarians voted against the bill.

In some democracies, parliamentary fractions tend to vote en bloc. In the Netherlands, for instance, it would be quite unlikely that some but not all liberal parliamentarians voted against a bill. Nevertheless, as far as I can tell, this wouldnt diminish the likelihood that (58) is interpreted as implying that the liberal fraction was divided.

(2) In order to prevent the rinderpest from spreading through the herd, some of the cows were vaccinated.

Since rinderpest is a highly contagious disease, it would be decidedly odd if only some of the cows were vaccinated, yet that is how we would understand (59).

(3) Cleo threw all her marbles in the swimming pool. Some of them sank to the bottom.

No doubt, it would be very odd if some of Cleos marbles failed to sink, yet according to my intuitions that is precisely what (60) conveys.

What these examples show is that genuine scalar inferences are not so easy to cancel, at all (cf. Geurts 1999a). From a Gricean perspective, this is to be expected: scalar implicatures arise because the speaker goes out of his way to make a statement that is weaker than what he could have said with equal effort, so it stands to reason that it should require special circumstances to suppress a scalar implicature. But in particular, lack of plausibility will generally be insufficient for doing so."

Our questions are: do Bart's intuitions pan out empirically? And if so, how can we model this effect?

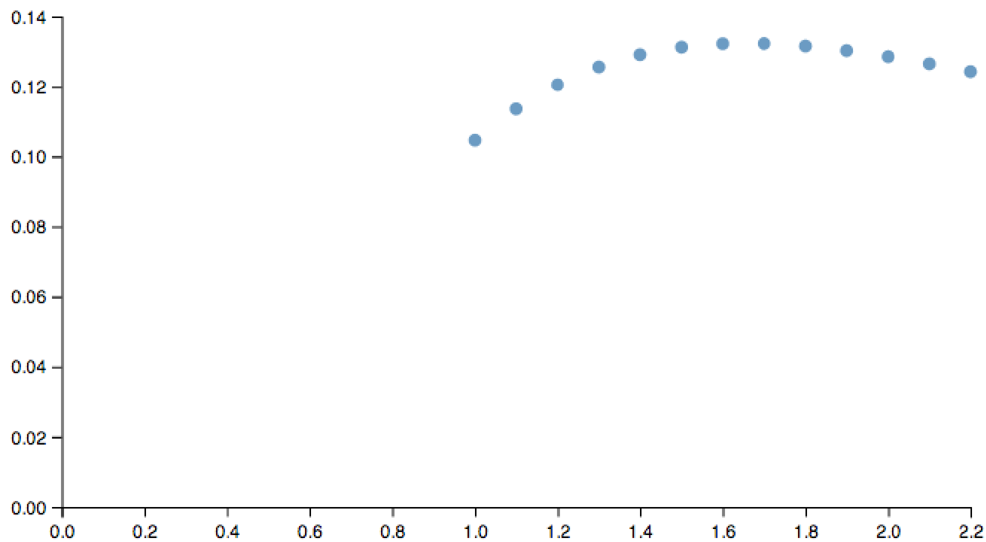

Figure 1: Model predictions for ∀ state after observing *Some of the X sank* with different prior probabilities of ∀ state (i.e. different X types) on the x-axis. Plot generated in play-space with alternative weights for *null; none; some; all* = 20; 1; 1; 1.

General modeling idea: if the prior probability of the ∀ state is high enough and there's a cheap enough alternative utterance –say, silence, or in order to refer to it: *null*– to *some* (crucially, cheaper than *some*) that signifies something like "everything is as usual", then the fact that the speaker went to the (extra) effort of saying *some* instead of *null* means things must really be different than the normal case. But if the normal case is ∀, then by saying *some* the speaker must really mean $\exists \wedge \neg\forall$. That is, there are two counter-acting forces: a) higher prior probability of ∀ state *increases* the posterior probability of ∀ state after observing *some*; b) presence of *null* alternative *decreases* posterior probability of ∀ state after observing *some* (if prior probability of ∀ is very high (how high is an empirical question)).

A preliminary version of the model is here: `http://forestdb.org/models/SFVSome.html` It's got some unnecessary bits leftover from when the model was intended to capture QUD effects. The general shape of the posterior curve for ∀ that it predicts is shown in Figure 1.

2 Experiment 1

The experiment can be found here: `http://stanford.edu/~jdegen/57_sinkingmarbles/alternatives.html`

2.1 Method

2.1.1 Participants

30 participants were recruited over Amazon's Mechanical Turk.

2.1.2 Procedure and materials

The experiment consisted of two phases: prior and posterior elicitation. The prior phase was intended to assess participants' prior beliefs about how likely different types of objects are to either *break/sink* prior to observing any additional information. Object types were marbles, feathers, balloons, books, wine glasses, iPhones. See Figure 2 for an example of a prior trial (including slider dependent measure).

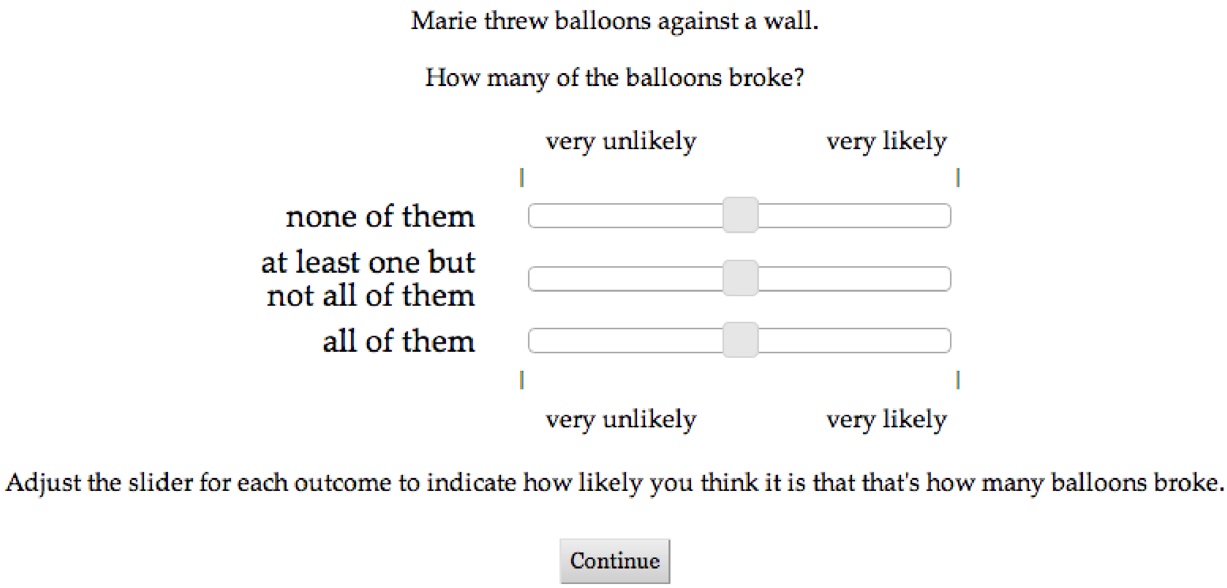

Figure 2: Example of a display in the prior elicitation phase.

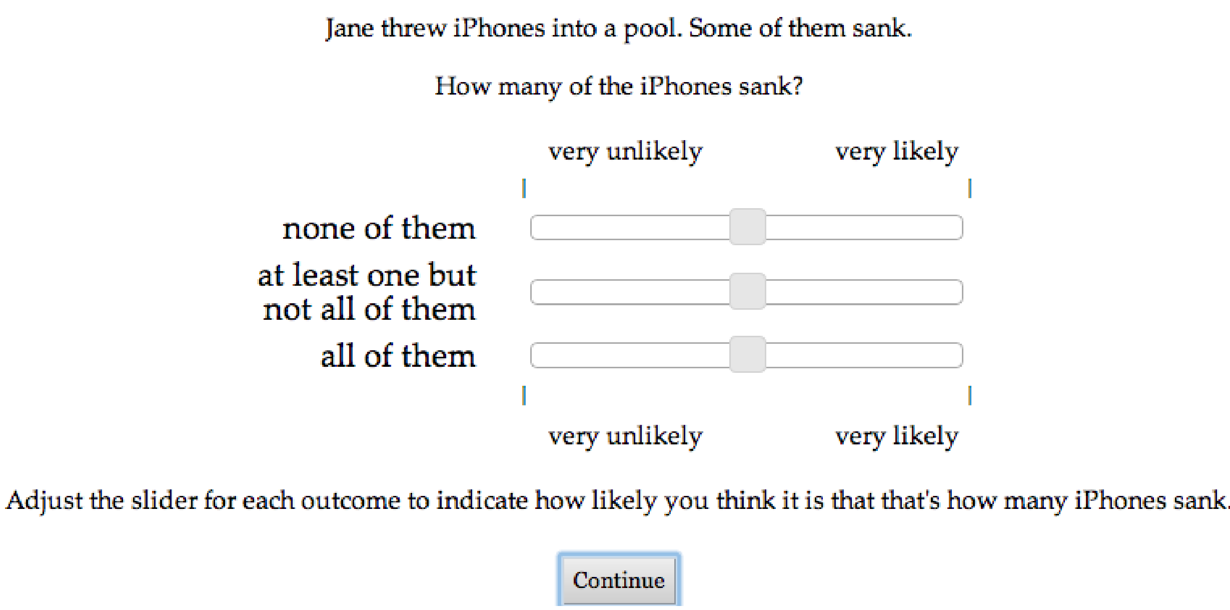

Figure 3: Example of a display in the posterior elicitation phase.

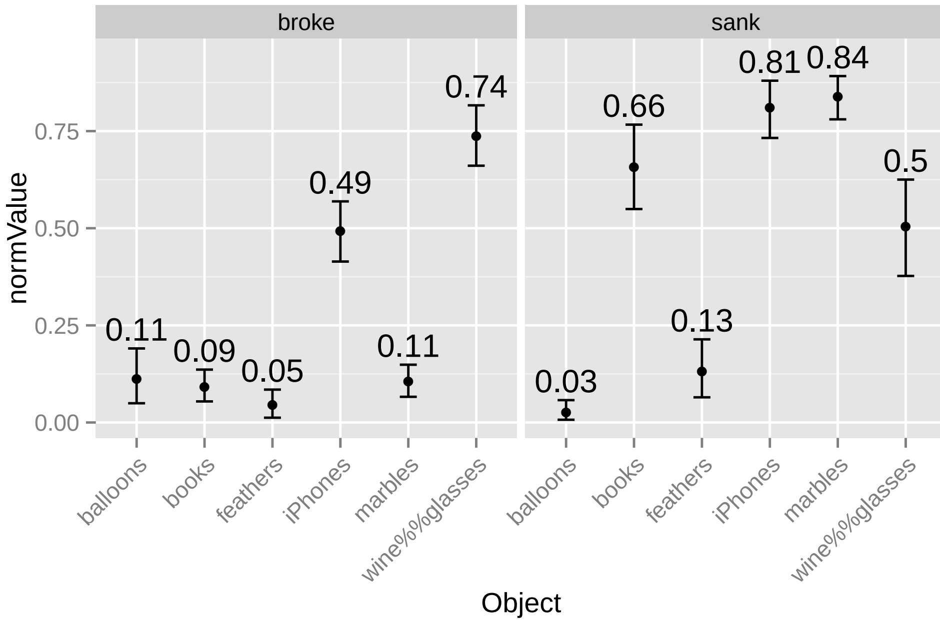

Figure 4: Mean prior probability of all objects breaking (left) or sinking (right) by object type.

In the posterior elicitation phase, additional information about how many of the objects sank/broke was given by adding an additional sentence: either *All* or *Some of the X broke/sank*. See Figure 3 for an example of a posterior trial.

Each phase of the experiment consisted of 12 trials: each of the six object types occurred with each verb (break/sink) in each phase. In the posterior phase, an object occurred once with *all* and once with *some*, but *all/some* were randomly paired with *break/sink*. That is, for each combination of verb/object we obtained 30 prior probabilities and for each verb/object/quantifier combination we obtained 15 posterior probabilities.

2.2 Results

Slider values were normalized for each trial to yield a probability distribution over states $\neg\exists$, $\exists \wedge \neg\forall$, and $\forall$ for each verb/object combination in the prior phase and each verb/object/quantifier combination in the posterior phase. Figure 4 shows the mean prior probabilities of the $\forall$ state for each object/verb combination.

The interesting question was whether we would see something like the shape in Figure 1. The empirical results are shown in Figure 5 (unaggregated) and Figure 6.

2.3 Discussion

The unaggregated data look super noisy, the aggregated ones seem to at least sort of show the right shape in the *sink* but not the *break* condition, but the error bars are huge. That it doesn't show up even a little bit in the *break* condition might be because none of the object types has a prior probability of breaking that's high enough (the highest we got was .74 for wine glasses).

Generally we have the problem so far that we're over-sampling the low- and under-sampling the high-prior-probability space. That is, we need more object types with a better spread, and preferably clustered around the higher range.

Another thing: one of the reasons the results might be so noisy is that the *all/some* information is not presented as part of a speaker's utterance, but as part of the instructions on each trial.

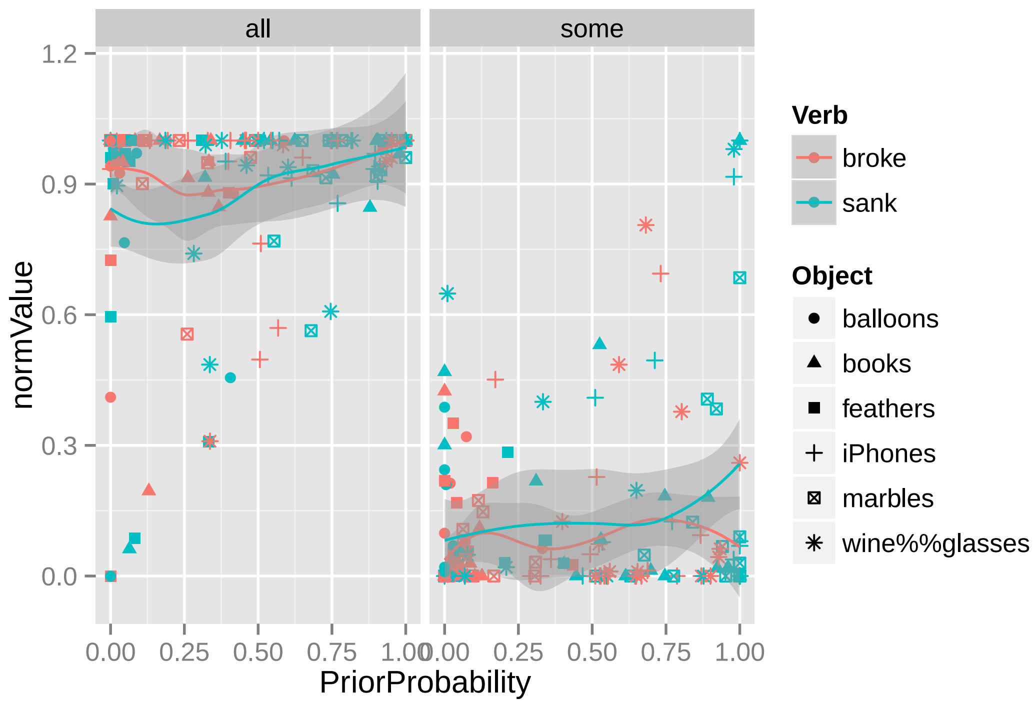

Figure 5: Posterior probability of each object breaking/sinking after being informed that *all* (left) or *some* (right) of them broke, by unaggregated (by-subject) prior probability of all of them breaking/sinking.

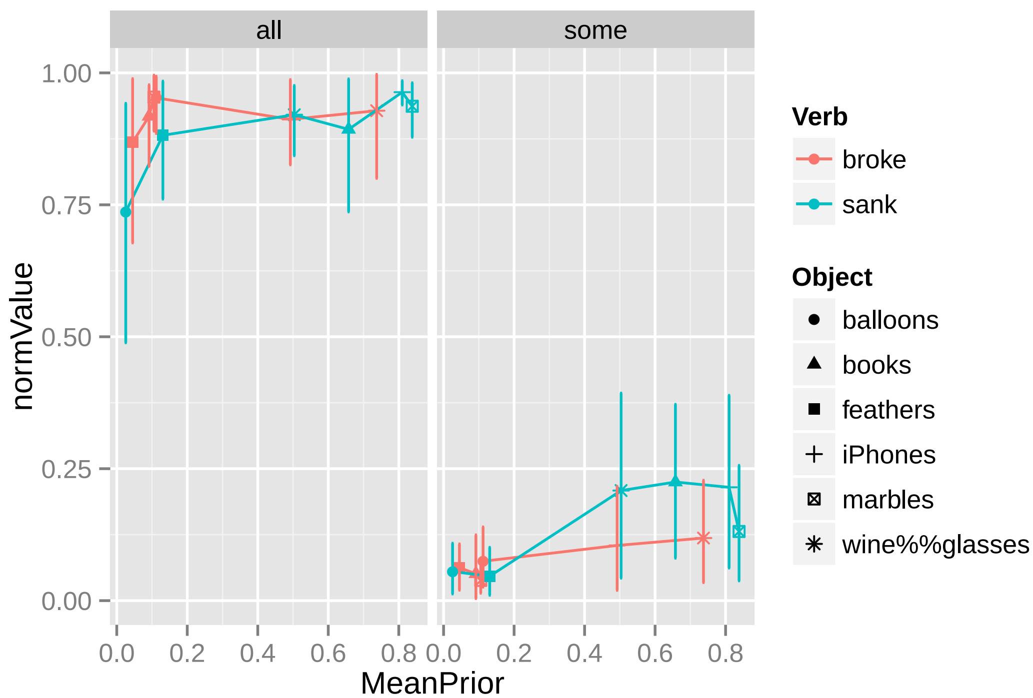

Figure 6: Posterior probability of each object breaking/sinking after being informed that *all* (left) or *some* (right) of them broke, by aggregated prior probability of all of them breaking/sinking. Error bars indicate bootstrapped 95% confidence intervals.

So...

3 To do

- Do additional piloting to get objects with a wider range of priors (and more high ones).
- Set up version of experiment where the additional information *Some/All of the X sank/broke* is provided as an utterance produced by a speaker who observes the situation.