The Effect of Container Context on the Interpretation of "full," a Gradable Adjective

Jacob J. Dalder (jjdalder@stanford.edu)
Department of Symbolic Systems
Stanford, CA 94305 USA

April Yang (april.yang@stanford.edu)
Department of Symbolic Systems
Stanford, CA 94305 USA

Abstract

This study seeks to capture differences in standards of value for three different container types (a cup, a wine glass, and beer stein) in regard to the maximum standard gradable adjective "full." Study participants are asked to participate in an experimental paradigm involving establishing their prior beliefs of fullness for the different container types and demonstrating their standard values of fullness for these container types. The experimenters' sought to capture the assumed results from such methods by adapting a Church model to predict the theta levels of standard value based on simulated prior beliefs.

Keywords: gradable adjectives, adjective model, cognitive model, full, standard value, theta

Background

After realizing we had taken the same linguistics class on semantics and pragmatics with Professor Chris Potts a year apart, we decided that we would probably be able to identify one of the topics we covered in that class for which we could build a Church model. We specifically ended up exploring the pragmatic interpretation of adjectives.

Gradable adjectives (GAs) refer to an attribute of a noun in relation to or in the context of other objects that may be ordered according to some scalar concept (McNally 2011; Lassiter & Goodman 2013). Examples of gradable adjectives include adjectives pertaining to height such as "tall" or "small" or those pertaining to weight such as "heavy" or "light." We can also breakdown gradable adjectives in terms of whether the scale associated with them is "closed" or "open." "Open" adjectives such as "tall" are unbounded—some noun can always be taller or less tall relative to another noun. Whereas closed adjectives, such as "closed" for example, are bounded at one or both ends—a closed door cannot become more closed. In conceptualizing of gradable adjectives in this way, it seems logical to understand them as properties, which implies that there must then be some degree or *standard value* of the property at which we determine that a particular adjective is applicable (McNally 2011). This idea is crucial for our research proposal.

As a result of the classification of adjectives as "open" or "closed," we need some framework to explain why "closed" adjectives have standards that are bounding endpoints in comparison to the context-dependent standards of "open" adjectives. It does not seem there is anything forcing the standards to be endpoints for "closed" adjectives. To account for this, Kennedy (2007) proposed the Interpretive Economy, which holds that language must "maximize the contribution of the conventional meanings of the elements of a sentence to the computation of its truth conditions." Therefore, for adjectives with a "closed" scale, their standards must be either maximal or minimal, and this way of determining the standard is preferable over context-dependence.

GAs may be parsed into the two categories: maximum standard absolute GAs ("straight," "flat," and "dry" are examples) or minimum standard absolute GAs ("spotted," "bumpy," and "wet" are examples). Maximum standard absolute GAs are all-or-nothing: it doesn't make sense for something to be "partially dry," it's either dry or it's not. Minimum standard absolute GAs require only some slight existence of the property in order to apply: we can use "bend" to describe an object that has some degree of bend to it. Specifically, our research project will delve into an in-depth analysis of "full," one particular maximum standard absolute GA.

Originally, our research and modeling exploration was primarily motivated by the study "Meaning and context in children's understanding of gradable adjectives," carried out by Kristen Syrett and colleagues (2009). The experiment sought to understand how the conception of "fullness" and other GAs would change in in a context of comparison, and how responses would different across child and adult study populations. One of the fundamental predictions of the Syrett study was that child participants should correctly avoid shifting their conceptions of the applicability absolute GAs even when faced with comparison.

For example, in one circumstance, the experimenter presented two cups to the subject, one three-quarters of the way full and the other half full. The subject was then instructed to "hand me the full one." Since neither cup is full, we call this an infelicitous request because it is not possible for the subject to successfully fulfill the request given the choice at hand. Children were must less likely to identify this question as infelicitous than adults were—67% of children responded in such a way as to tell the experimenter that the request was impossible, while 100% of adult subjects answered this way. Although this

significant difference in responses between children and adult subjects was noted for the adjective “full,” no other adjectives resulted in such a sharp contrast. In the discussion section of the paper, several theories are hypothesized to explain this difference. Therefore, after analyzing these study results, we decided that “full” would be a particularly interesting gradable adjective to explore further.

When someone is served a “full” cup of water, the amount of liquid rarely comes close to the volume capacity of the cup. So calling such a cup “full” is strictly speaking false, but because a nearly-full cup is accepted as “full” as long as it’s close enough to the maximal standard, McNally (2011) states that “interlocutors are willing to speak loosely.” However, at what point does speaking loosely become speaking incorrectly when we consider different containers that typically hold different types of liquid? This is the question our research seeks to answer.

Research Question and Hypotheses

How do subjects interpret “full” (an absolute, maximum standard GA) in a relative manner based on context clues from a container? The overarching research purpose is to see if there are shifting standards for the applicability of the adjective “full” depending on the context. Our means of testing this research question involves three different containers: an opaque plastic cup, a wine glass, and a beer stein. We hypothesize that our study participants will identify different thresholds for fullness for these three container types. Here, *threshold* is being used a synonym for *standard value*, as discussed in the previous section of this paper. The threshold for “full” for the beer stein will be the at more than 90% full (since beer steins are often served overflowing), the threshold for “full” for a wine glass will be at about half of the volume capacity for a wine glass, and the threshold for full for the cup will be somewhere between the other two.

Methods

Experimental Design

There are two components to our experiment. First, we use subjects to determine what the intuitive prior is for fullness of cups, wine glasses, and beer steins. Second, subjects demonstrate possible standard values for fullness of cups, wine glasses, and beer steins.

Our subjects are 20 Stanford freshmen. Subjects will be tested individually. Materials for each subject are two opaque plastic cups, two wine glasses, and two beer steins. Water is provided for cups, red wine for glasses, and beer for steins. The same containers will be used across subjects.

The experimenter will run the subject through six pouring tasks. Tasks within each set are randomized, but task set 1 is always performed first. For each task, the subject is provided with one empty container and a large, opaque dispenser full of the liquid corresponding to that container type. After each task, the materials are removed and materials for the next task are placed before the subject. The experimenter will turn their back to the subject to give task instructions. Experimenter's instructions for the tasks are as follows:

Task Set 1

a. I have a cup. Please pour the amount of water, if any, that you believe is in my cup.
b. I have a wine glass. Please pour the amount of wine, if any, that you believe is in my wine glass.
c. I have a beer stein. Please pour the amount of beer, if any, that you believe is in my beer stein.

Task Set 2

d. I have a full cup. Please pour the amount of water, if any, that you believe is in my cup.
e. I have a full wine glass. Please pour the amount of wine, if any, that you believe is in my wine glass.
f. I have a full beer stein. Please pour the amount of beer, if any, that you believe is in my beer stein.

The level of liquid poured for each task will be recorded as a percentage of the maximum liquid volume the container can hold (as measured with a measuring cup).

Modeling Approach

We are modifying a Forest DB repository model in Church for interpretation of positive-form adjectives. At a high level, this model assumes a pragmatic listener (experiment subject) interpreting the utterances of a speaker (experimenter) who speaks with a literal listener in mind. To clarify, the pragmatic listener would be willing to accept a container that is not volumetrically full as “full,” while a literal listener would only accept a container as “full” that is volumetrically full. There is a cost to the speaker for using an adjective as compared to no utterance at all. We hope the model will predict the threshold level of liquid that constitutes "full" in different container contexts.

Listeners' prior beliefs about how much liquid might be expected to be in a cup (fullness prior) are based on results of task set 1. The level of liquid subjects poured will be rounded to the nearest tens-percentile; these percentiles are used as a list from which to sample. The number of subjects who fall into each tens-percentile are used as the probabilities for sampling (normalized through the multinomial function). We will assume a probability of 0.1 (before normalization) for unused percentiles to compensate for our limited subject pool. The model representation for the prior on cup fullness might look like so:

```
(('cup) '((0 10 20 30 40 50 60 70 80 90 100 105)
          (5 .1 1 .1 .1 3 .1 2 3 6 .1 .1)))
```

The literal listener's prior threshold for fullness (theta) is assumed to be uniformly distributed over all possible fullness values (however, we later experiment with a non-uniform theta prior that is heavily skewed toward maximum liquid levels, which may more realistically reflect listener and speaker beliefs).

The literal listener assumes that the speaker's utterance is true. Any uncertainty about the threshold level for fullness is passed up to the pragmatic listener. The speaker chooses an utterance that is most informative for the literal listener,

keeping in mind the cost incurred by using an adjective. We did not perform our own cost calculations for this model; we used the calculations in the existing model. The pragmatic listener takes into account prior probabilities and what the speaker would be likely to say when interpreting the speaker's utterance.

The results of task set 2 form the basis of our data for human posteriors fullness thresholds. The human posterior can be used as a point of comparison with the model's predictions when judging the effectiveness of the chosen modeling technique.

Results and Discussion

We have not been able to run this experiment with twenty subjects. We have made assumptions about the outcomes of task set 1 (see Experimental Design) to simulate data. See the figures 1.1 through 4.2 representing our simulated data and modeling results for the different container types below.

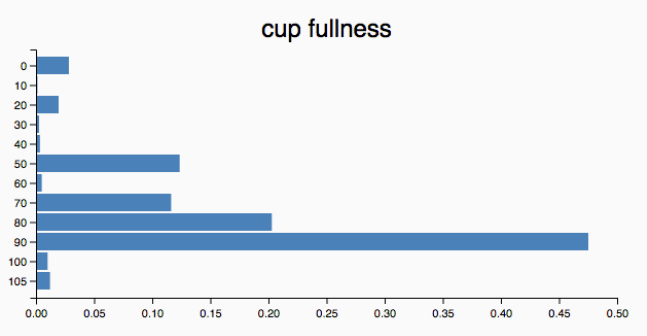

Figure 1.1

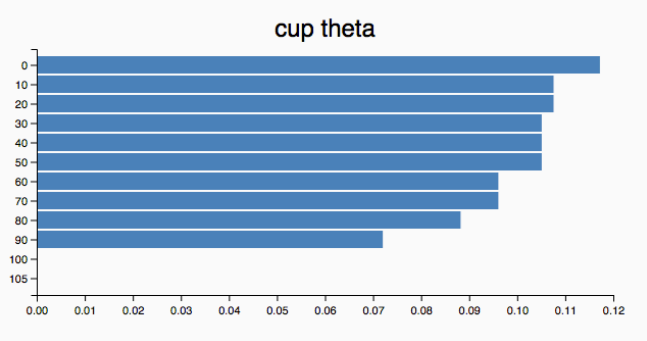

Figure 1.2

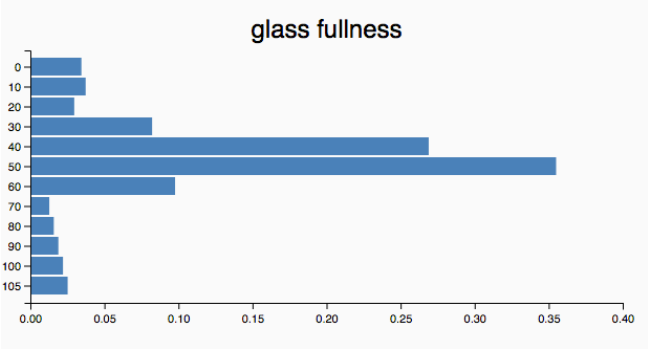

Figure 2.1

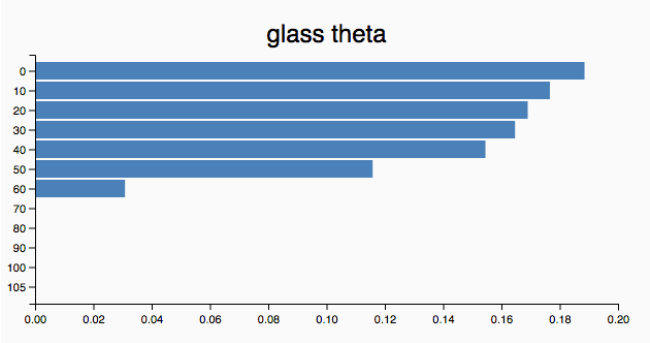

Figure 2.2

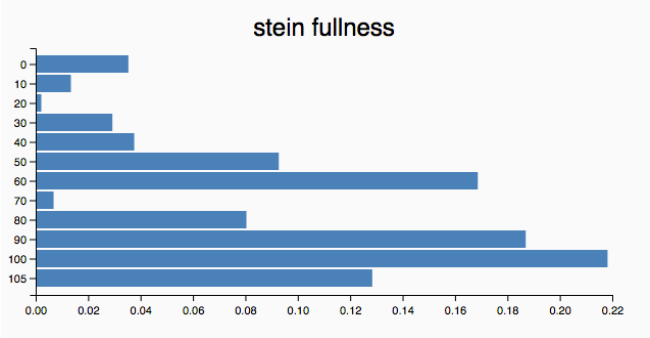

Figure 3.1

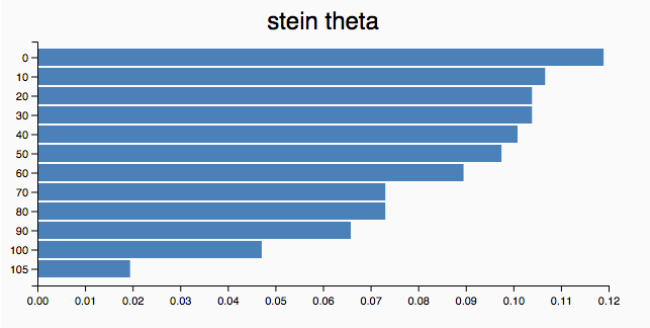

Figure 3.2

We have also experimented with using a non-uniform theta prior. Results for the beer-stein are shown in Figure 4.1 and 4.2.

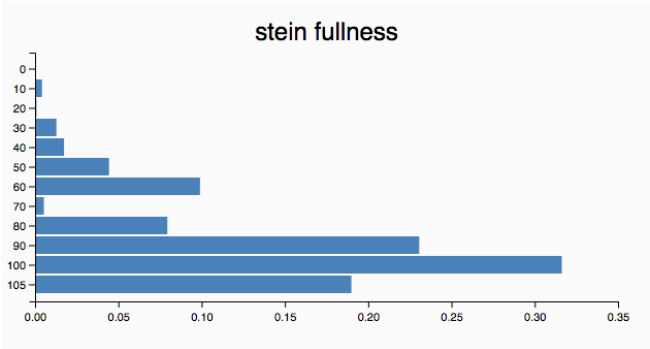

Figure 4.1

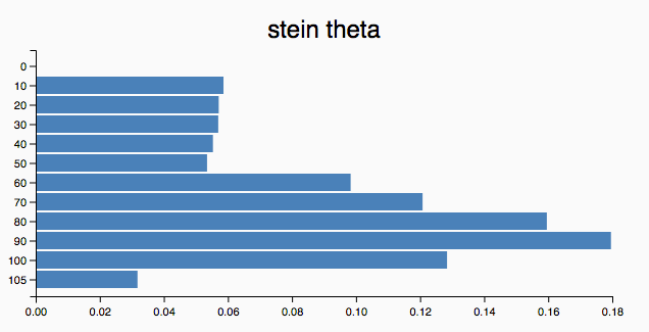

Figure 4.2

Considering the graphical representations of the results of our model above and based on the assumptions we made about the outcomes of task set 1, using a non-uniform theta prior seems to produce a result we can make better sense of. This is the modeling approach used by Lassiter, D. and Goodman, N. (2013). Therefore, averaging the stein thetas from figure 4.2, we get a theta or standard value of between 80 and 90 percent. Following the hypotheses we initially delineated for this study, this result seems to support our hypothesis that "full" level for a beer stein would be close to the volumetric capacity of the container.

One of the most significant limitations of our study is that we were not able to run participants in order to use real human priors to see how that would influence our model. The reason for this is that had problems understanding how to best adjust Forest DB repository adjectives model for use in our study such that we were forced to revise our experimental methods up to the submission of this paper. In terms of future directions, we would like to consider adapting this model to replicate Syrett et al. study involving a paradigm of comparison in which we test for shifting standard values for different container types, which was the original intention of our research project. Additionally, this body of knowledge would benefit from modeling how absolute adjectives are anchored to their endpoints, modeling alternatives to the utterance "full" excluding non-utterances, and modeling a similar study as the one we propose here but involving participants and containers from other cultures.

References

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*, 1-45.

Lassiter, D. and Goodman, N. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. *SALT*, *23*, 587-610.

Lassiter, D. and Goodman, N. Adjectival vagueness in a Bayesian model of interpretation. To Appear in *Synthese*.

Stuhlmueller, A., Goodman, N. (2013). Reasoning about reasoning by nested conditioning: Modeling theory of mind with probabilistic programs. *Cognitive Systems Research, 28*, 80-99.

Syrett, K., Kennedy, C., Lidz, J. (2009). Meaning and context in Children's understanding of gradable adjectives. *Journal of Semantics, 27*, 1-35. doi:10.1093/jos/ffp011