

Great variability exists in the cardinal values of linguistic quantifiers like *many*, *few*, *several*, etc. For example, the sentence “There are *many* students in Erin’s class” maps to a very different set of cardinal values than “Andy had *many* cups of coffee last week.” This is a problem for semanticists, who seek to formally describe meaning, because there is no clear way to resolve this context-dependent variability. Following the work of [1], we develop and experimentally validate a Bayesian model of quantifier semantics which represents these quantifiers as cumulative density thresholds along a probability distribution of expected values. Considering the previous examples, Figures 1a and 1b illustrate these expected value distributions and the cardinal threshold values for the lower-bound of *many* used in these contexts. The semantics of *many* here are defined with respect to  $\theta_{many}$ , a stable cumulative probability density shown as the area underneath these curves (Equation 1a). Our hypothesis is that this threshold remains stable across contexts, and differences between probability distributions over expected values introduce the contextual variation (Hypothesis 1). As the figures reveal, the shapes of the expected value curves shift the cardinal threshold value ( $x_{min}$ ). For example, with the threshold  $\theta_{many} = 0.4$ , then the expected values imply that “many cups of coffee” is at least 8 cups, while “many students” is at least 17 students.

In addition to testing this preliminary hypothesis, we also examine whether or not the upper- and lower- bounds of *several* ( $\theta_{several-max}$  and  $\theta_{several-min}$ ) align respectively with  $\theta_{many}$  (Hypothesis 2) and the upper-bound of *few* ( $\theta_{few}$ ) (Hypothesis 3). Experiment 1 empirically elicits values for the imputation of the expected value curve for sixteen different contexts, two of which are described above. Experiment 2 then probes participants’ truth-conditional semantics regarding the quantified utterances in these contexts.

In order to compare the viability of these hypotheses, we perform three independent analyses, all using Bayesian hierarchical models to estimate these threshold values from the experimental data. Five models were fit to test three hypotheses. One model implements the null hypothesis, which rejects the claims of all three hypotheses and fits individual thresholds for each quantifier for each context (Model A). The remaining four models accept Hypothesis 1 and implement all possible combinations of acceptance or rejection of Hypotheses 2 and 3. Models were fit using rstan and technical specifications are given on page 3.

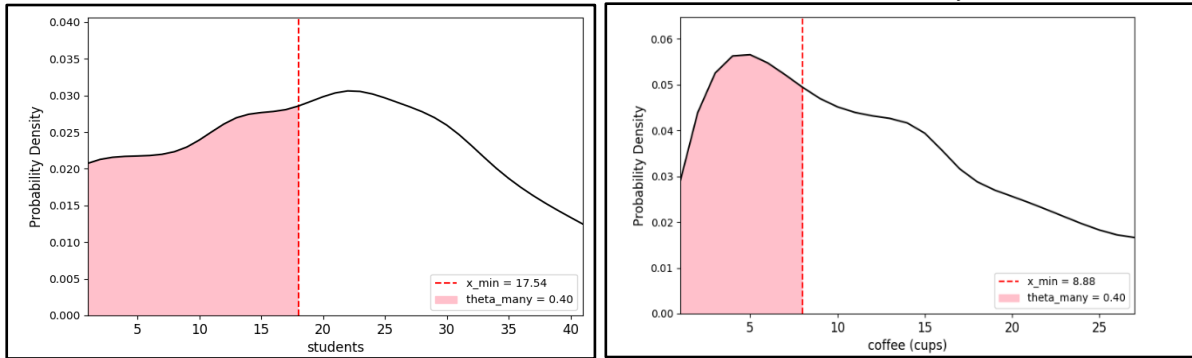
In our first analysis, we investigate what overlap exists between the context-independent thresholds implemented in Model A for a given threshold value. We find that some thresholds are more consistent than others and might represent more context-stability, namely  $\theta_{many}$  (Figure 3) and  $\theta_{several-min} \cdot \theta_{few}$  also demonstrates an overlapping interval for most contexts, although not quite as many. This provides significant evidence for the existence of a context-stable threshold for these quantifiers. However,  $\theta_{several-max}$  (Figure 4) is much less consistent than the other three. This leads us to believe that probabilistic threshold values for bounds on quantifiers might exist on a spectrum of context-dependence.

Our second analysis computed the WAIC Information Criterion [2] to perform model comparison. While Model A achieved the best performance, Model B did not perform significantly worse. This result, in combination with the findings of [1], leaves open the question of whether a context-stable threshold model is more appropriate than the null-hypothesis.

Our final analysis compares our context-stable thresholds against the combined threshold values  $\theta_{several-few}$  and  $\theta_{several-many}$ . Figure 2 shows that these thresholds are all significantly different from one another, strong evidence against Hypotheses 2 and 3.

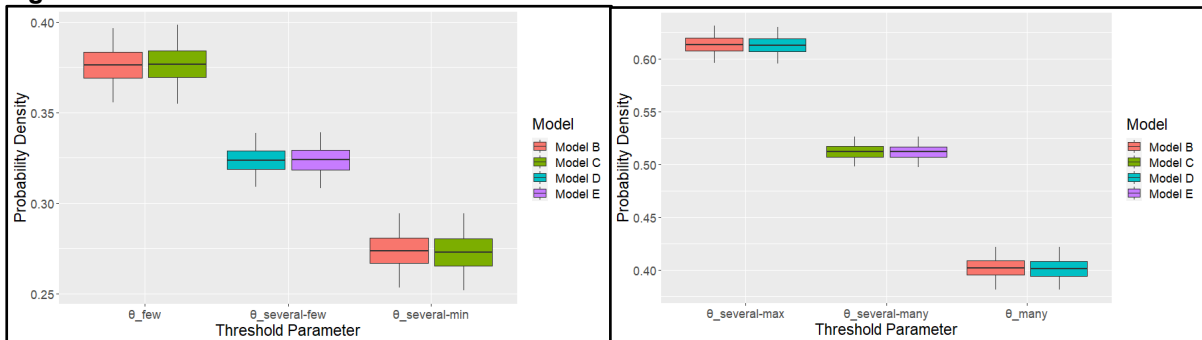
The results of our modeling and analysis suggest some degree of context-stability in cumulative density thresholds, albeit with varying degrees of stability by quantifier and by upper- versus lower-bound. We plan in future to investigate a larger, more diverse set of quantifiers to explore this spectrum in further detail. Our study does not however provide support for Hypotheses 2 and 3 as semantic phenomena. Further research will provide a pragmatic approach to the coincidence or lack thereof between thresholds of this nature.

**Figure 1 – Expected Value Distributions and Context-Stable Threshold  $\theta_{many}$**

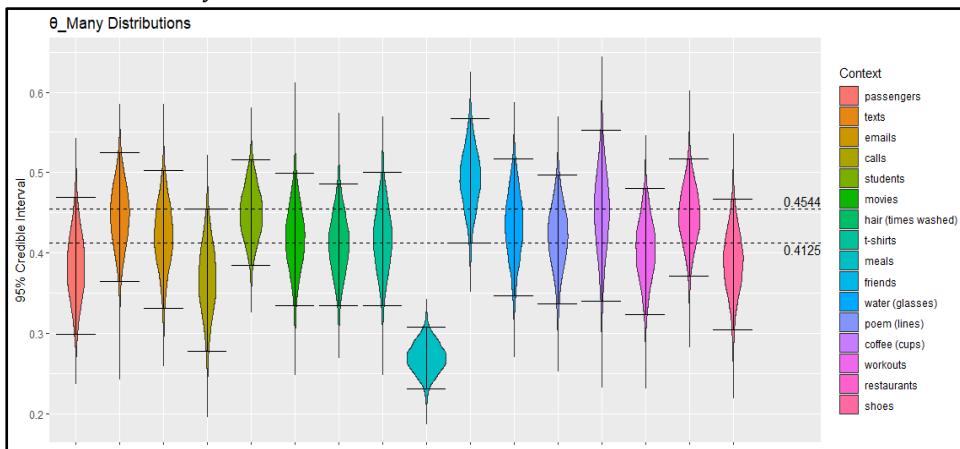


(a) “There are many students in Erin’s class.” (b) “Andy had many cups of coffee last week.”

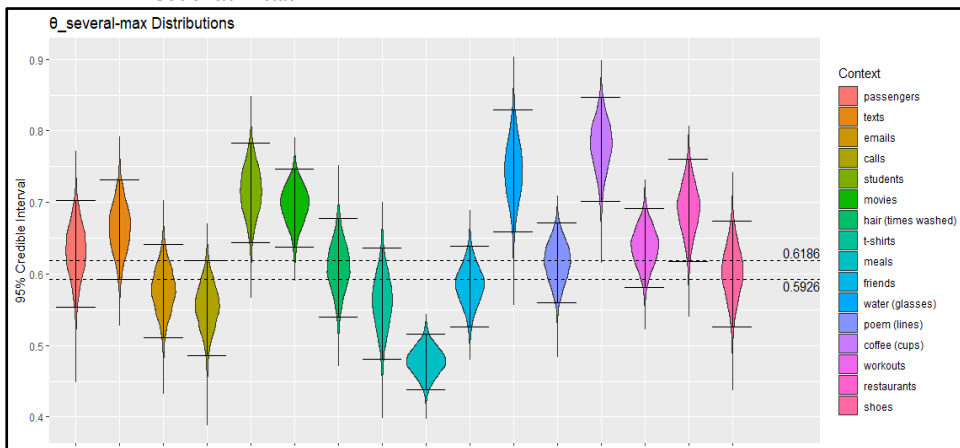
**Figure 2 – 95% Credible Intervals for Stable Thresholds in Models B-E**



**Figure 3 –  $\theta_{many}$  Distributions for Model A – ITM**



**Figure 4 –  $\theta_{several-max}$  Distributions for Model A – ITM**



**References**  
 [1]Schöller & Franke (2017)  
 [2]Gelman, Hwang & Vehtari (2013)