Bayesian Modeling of Quantifier Cardinal Reference Variability:
The Case of English Few, Several, and Many

Skyler Jove Reese - UC Davis (mwreese@ucdavis.edu), Masoud Jasbi - UC Davis (jasbi@ucdavis.edu), Emily Ida Popper Morgan - UC Davis (eimorgan@ucdavis.edu)

Great variability exists in the cardinal values of linguistic quantifiers like *many*, *few*, *several*, etc. For example, "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 [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 θ_{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 $\theta_{many} = 0.4$, "*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 θ_{many} (Hypothesis 2) and the upper-bound of *few* (θ_{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.

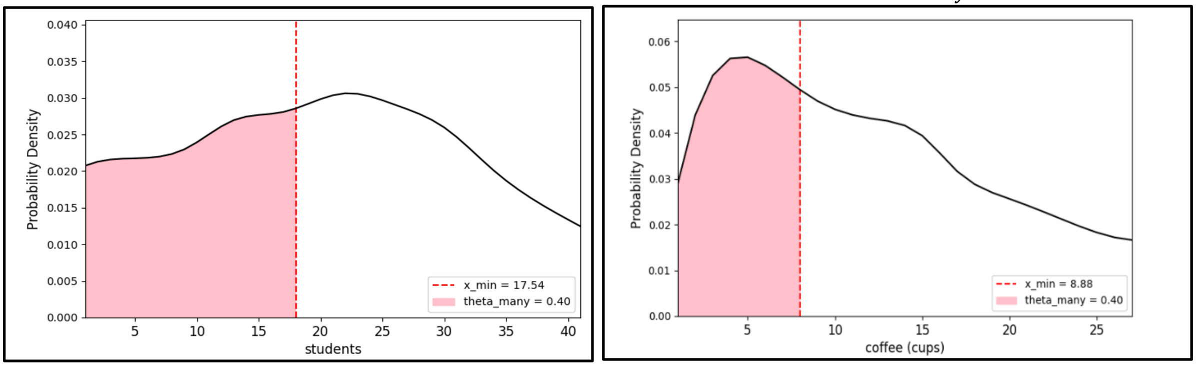
In our first analysis, we investigate what overlap exists between the context-independent thresholds implemented in Model A for a given threshold value. We find that some thresholds are more consistent than others and might represent more context-stability, namely θ_{many} (Figure 3) and $\theta_{several-min}$. θ_{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 θ_{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

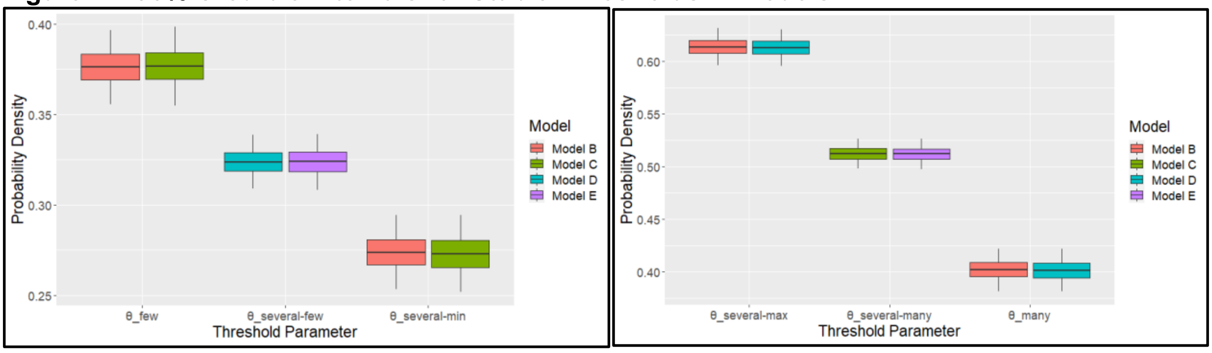


Figure 3 – θ_{many} Distributions for Model A – ITM

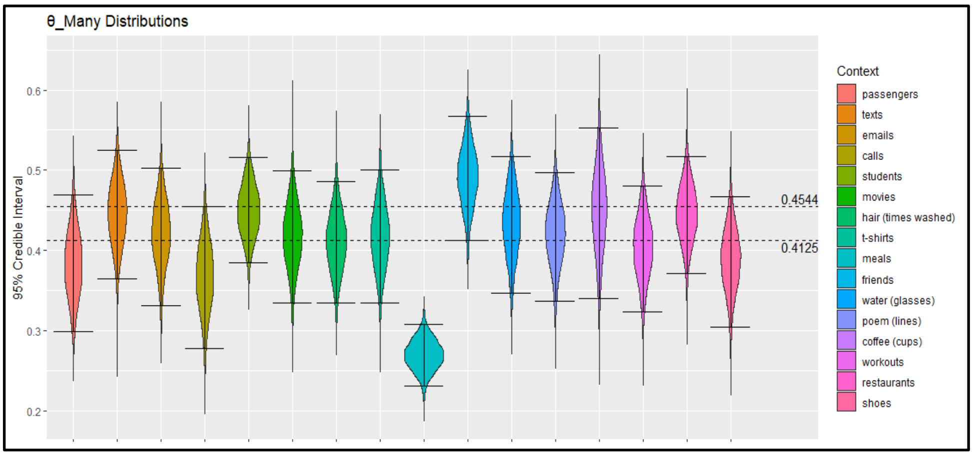
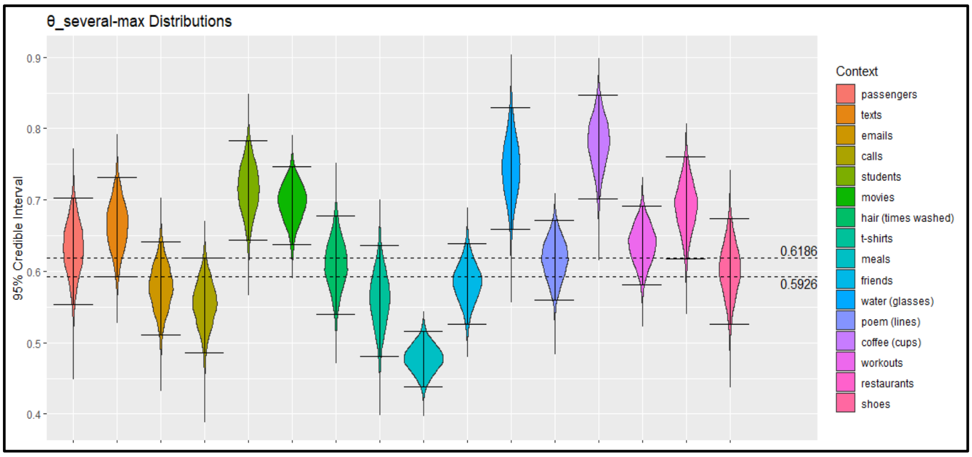


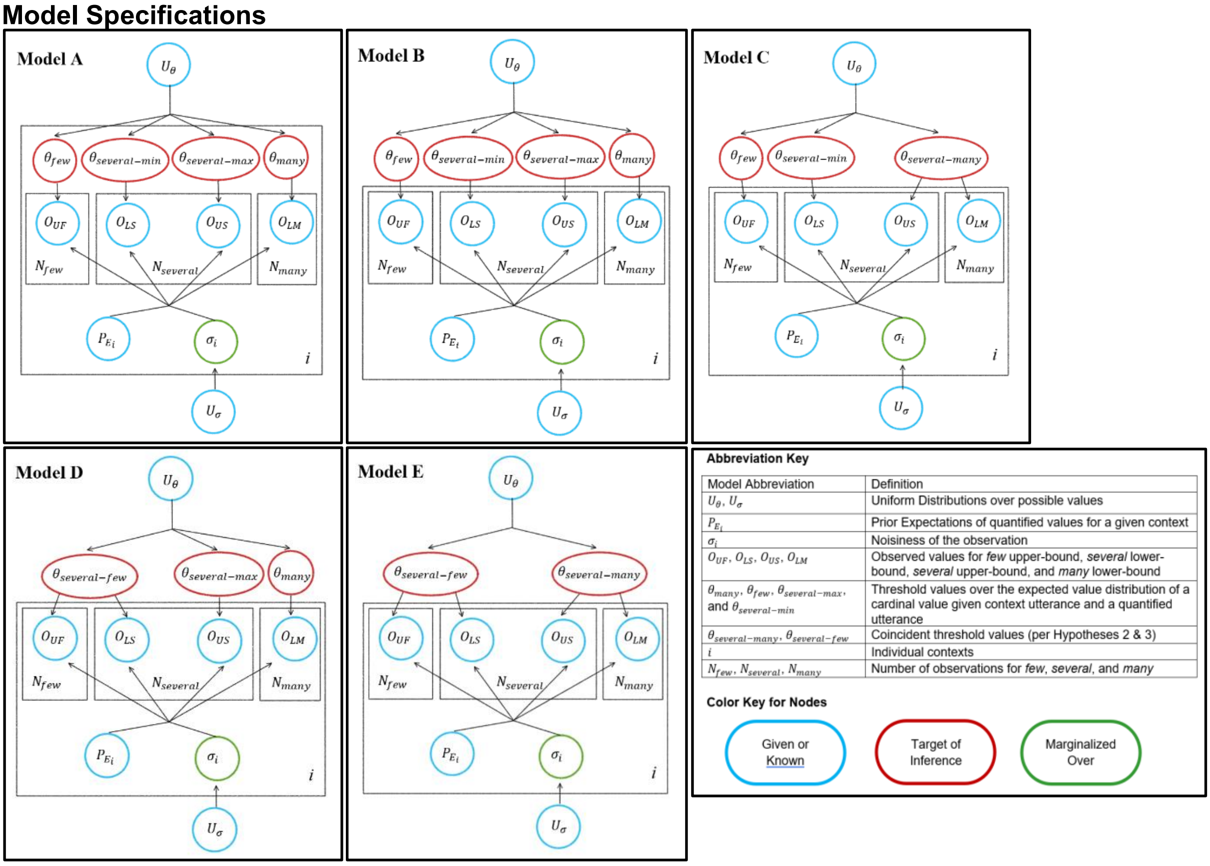
Figure 4 – $\theta_{several-max}$ Distributions for Model A – ITM



References

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

Model Specifications



Abbreviation Key

Model Abbreviation	Definition
U_θ, U_σ	Uniform Distributions over possible values
P_{E_i}	Prior Expectations of quantified values for a given context
σ_i	Noisiness of the observation
O_{UF}, O_{LS}, O_{US}, O_{LM}	Observed values for *few* upper-bound, *several* lower-bound, *several* upper-bound, and *many* lower-bound
θ_{many}, θ_{few}, $\theta_{several-max}$, and $\theta_{several-min}$	Threshold values over the expected value distribution of a cardinal value given context utterance and a quantified utterance
$\theta_{several-many}$, $\theta_{several-few}$	Coincident threshold values (per Hypotheses 2 & 3)
i	Individual contexts
N_{few}, $N_{several}$, N_{many}	Number of observations for *few*, *several*, and *many*

Bayesian Models and Hypotheses

		Hypothesis 1 Context-Stable Thresholds	Hypothesis 2 Several-Many Coincident Threshold	Hypothesis 3 Several-Few Coincident Threshold
Model A	Individualized Threshold Model (ITM)	Reject	Reject	Reject
Model B	Generalized Threshold Model (GTM)	Accept	Reject	Reject
Model C	GTM: Several-Many	Accept	Accept	Reject
Model D	GTM: Several-Few	Accept	Reject	Accept
Model E	GTM: Several-Many & Several-Few	Accept	Accept	Accept

Example Posterior Distributions for a Given Context *i* in Model A (Correlates with Stable Thresholds Implemented for Models B-E)

Equation 1

(a) $[\![\text{Many As are B}]\!] = 1 \text{ iff } |[\![A]\!] \cap [\![B]\!]| \geq O_{LM} \text{ where } \mathrm{cdf}_{P_{E_i}}(O_{LM}) = \theta_{many_i}$

(b) $[\![\text{Few As are B}]\!] = 1 \text{ iff } |[\![A]\!] \cap [\![B]\!]| \leq O_{UF} \text{ where } \mathrm{cdf}_{P_{E_i}}(O_{UF}) = \theta_{few_i}$

(c) $[\![\text{Several As are B}]\!] = 1 \text{ iff } O_{LS} \leq |[\![A]\!] \cap [\![B]\!]| \leq O_{US}$
$\text{where } \mathrm{cdf}_{P_{E_i}}(O_{US}) = \theta_{several-max_i} \text{ and } \mathrm{cdf}_{P_{E_i}}(O_{LS}) = \theta_{several-min_i}$

Equation 2

(a) $\theta_{many_i} = \mathrm{cdf}_{P_{E_i}}(O_{LM}) \xrightarrow{inverse} O_{LM} = \mathrm{cdf}^{-1}_{P_{E_i}}(\theta_{many_i})$
$\xrightarrow{add\ noise} O_{LM} \sim Normal(\mathrm{cdf}^{-1}_{P_{E_i}}(\theta_{many_i}),\ \sigma_i)$

(b) $\theta_{few_i} = \mathrm{cdf}_{P_{E_i}}(O_{UF}) \rightarrow O_{UF} \sim Normal(\mathrm{cdf}^{-1}_{P_{E_i}}(\theta_{few_i}), \sigma_i)$

(c) $\theta_{several-min_i} = \mathrm{cdf}_{P_{E_i}}(O_{LS}) \rightarrow O_{LS} \sim Normal(\mathrm{cdf}^{-1}_{P_{E_i}}(\theta_{several-min_i}), \sigma_i)$

(d) $\theta_{several-max_i} = \mathrm{cdf}_{P_{E_i}}(O_{US}) \rightarrow O_{US} \sim Normal(\mathrm{cdf}^{-1}_{P_{E_i}}(\theta_{several-max_i}), \sigma_i)$