

Supplementary Material For: “Number-Knower Levels in Young Children: Insights from a Bayesian Model”

Michael D. Lee and Barbara W. Sarnecka
Department of Cognitive Sciences
University of California, Irvine

Abstract

This supplementary note applies the approximate-meanings model to the Fast Cards data. Results are presented for the inferred knower-levels, coefficients of variation, and base-rate, and for the posterior predictive assessment of model fit.

Analysis and Results

We applied the approximate-meanings model from Figure 5 of the main article to the Fast Cards data. Our results consider the inferred knower-levels, coefficients of variation across the 56 children, the inferred base-rate and posterior predictive assessments of fit for a representative subset of the children.

Knower-Levels. The main paper reports that the heuristic classification into knower-levels gave the following distribution: 4 pre-number-knowers; 7 one-knowers; 21 two-knowers; 15 three-knowers; and 9 four-knowers/CP-knowers. (The difference between four-knowers and CP-knowers cannot be reliably detected by the Fast Cards task, as the difference depends on counting.). The posterior distributions for the latent knower-level states z_i classified the 56 children into 3 PN-knowers, 8 1-knowers, 21 2-knowers, 17 3-knowers, 7 4-knowers/CP-knowers. The kappa measure of agreement between the heuristic and model assessments is 0.72.

Coefficient of Variation. Figure 1 shows the distribution of the inferred coefficient of variation (i.e., the posterior expectation of σ) over all children. These coefficients range from 0.29 up to 13.4, with the majority being around 1. As noted in the main article, the expected range, based on previously reported estimates, is around 0.33 to 0.5. In this analysis, however, more than half of the children (32 out of 56) have estimated coefficients greater than 1.

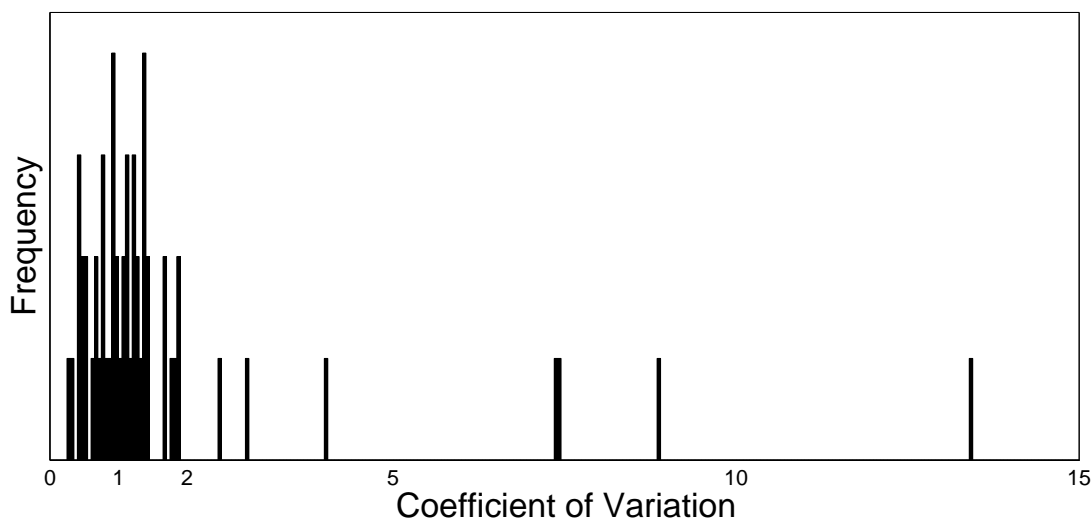


Figure 1. Distribution of the coefficient of variation across the 56 children’s Fast Cards data, as determined by the approximate-meanings model.

Base Rate. Figure 2 shows the inferred base-rate (i.e., the posterior predictive distribution of π) for the approximate-meanings model. The base-rate is expressed over the range of answers from 1 to 100, since 100 was the largest answer given by a child in this data set. Recall the base-rate for the knower-level model on the same data, as reported in Figure 9 of the main article, was highly interpretable, giving priority to single-digit answers, and prominent numbers like 10, 20 and 50. In contrast, the base-rate in Figure 2 is difficult to interpret. It appears simply to give roughly uniform, but noisy, base-rate to all possibilities.

Posterior Predictive Fits. Figure 3 shows the posterior predictive distributions and observed data for six selected children. The three children in the top row are those highlighted in Figure 10 of the main article. The three children in the bottom row are chosen to demonstrate the range of fits observed across all children. In particular, children 41 and 42 are good examples of the model describing the data well, and producing a reasonable coefficient of variation. The results for child 39, in contrast, show the difficulties the approximate-meanings model has in account for answers like 10 to questions like 2. It explains this behavior by using large coefficient of variation, but this leads to a wide range of answers being predicted for most potential questions.

As in the main article, we used Cohen’s kappa as a simple means of quantifying fit across all of the children on the Fast Cards data. For the knower-level model, kappas ranged from 0.36 to 0.75, with a mean of 0.41. For the approximate-meanings model, kappas ranged from 0.01 to 0.40 with a mean 0.22.

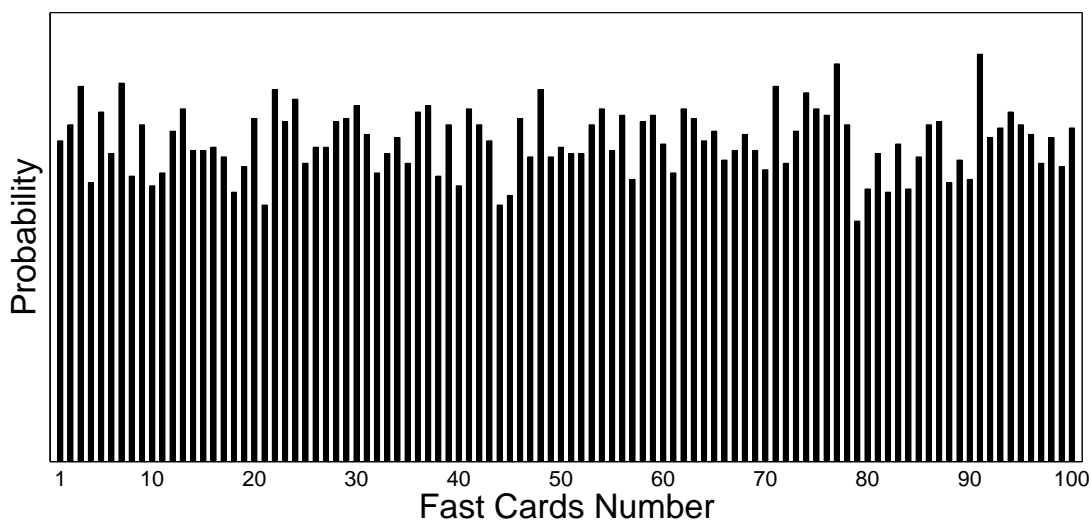


Figure 2. Inferred base-rate for the approximate-meanings model from the Fast Cards data.

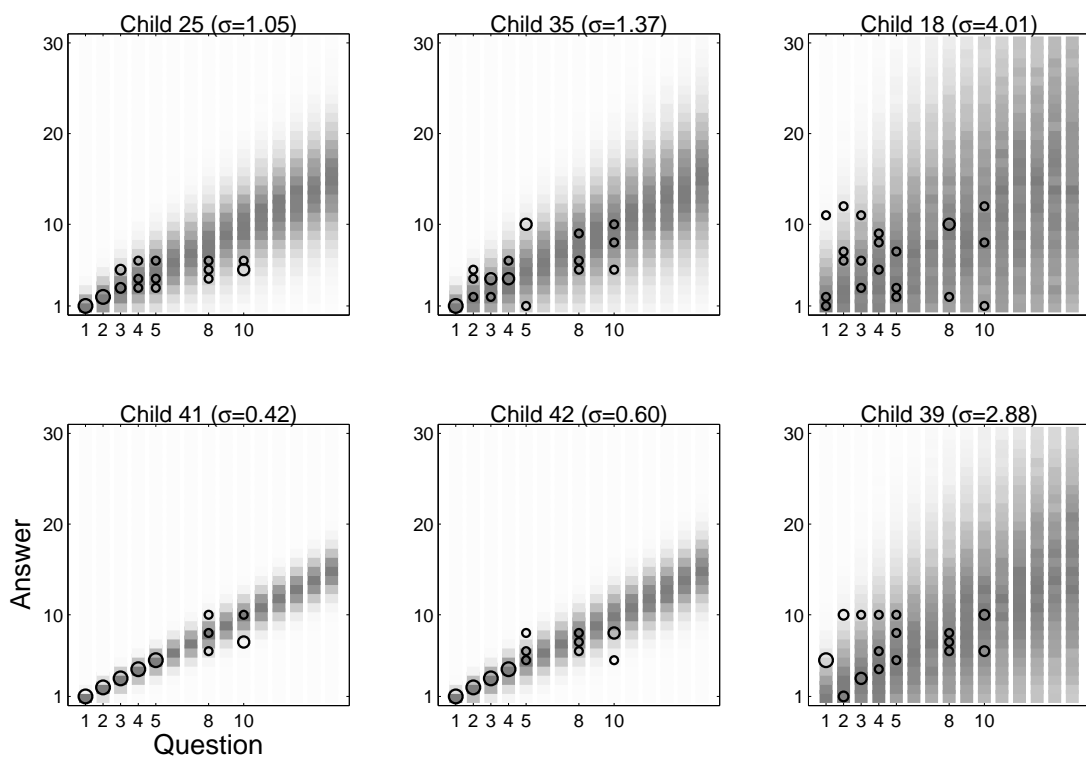


Figure 3. Posterior predictive fit to task behavior of the approximate-meanings model, to the Fast Cards data for six selected children.