Bo-NO-bouba-kiki: Picture-word mapping but not sound symbolism in a language trained bonobo

Supplementary material

Konstantina Margiotoudi   Manuel Bohn   Natalie Schwob   Jared Taglialatela
Friedemann Pulvermüller   Amanda Epping   Ken Schweller   Matthias Allritz

Data and code availability

All data and code associated with the study can be found in the following online repository: https://github.com/manuelbohn/bonoboubakiki

Power analysis

We conducted a power analysis to specify the number of test trials in the experiment. The details of this analysis can be found in the file '/analysis/power_analysis.Rmd. The results of this power analysis suggested that when assuming a true effect of 65% correct responses with 300 test trials our model would support the conclusion that performance was above chance in 88/100 cases (see Figure 1).
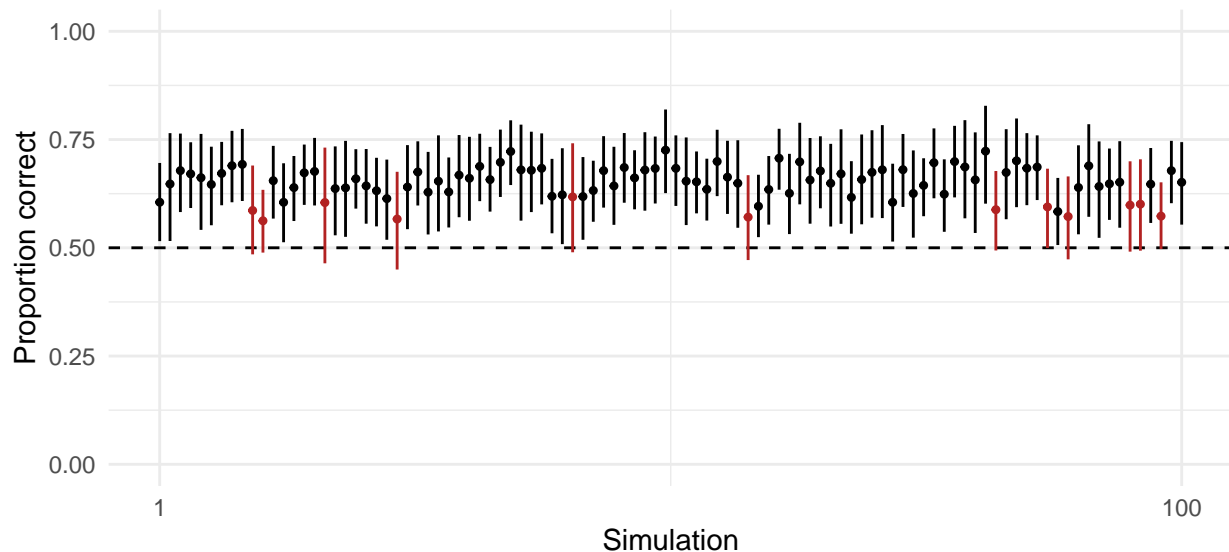

Figure 1: Power analysis based on 100 simulations with a true effect of 65% correct responses and 300 test trials. Points show the model intercept for each simulation (with with 95% CrI). Simulations in red yielded a wrong conclusion.

Table 1: Number of trials per condition.

condition	n
regular	2698
test	300

Descriptives

Number of Trials

Table 1 gives the number of trials per condition. There were 30 sessions in total, each with 10 test trials. Because of a software error, two regular trials were missing in session 1 and 2 (see main paper for detail).

Reaction times

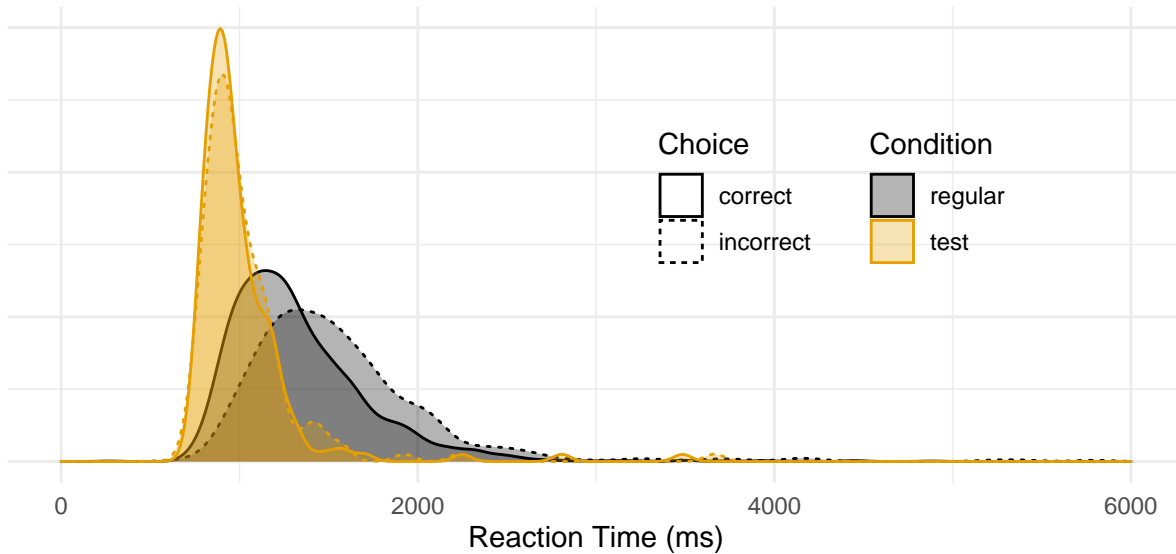Figure 2 visualizes the reaction times by correct choice and trial type.

Figure 2: Reactions times for regular and test trials split by correct and incorrect responses.

Figure 3 shows the change in reaction time over time. Reaction times are relatively stable across time and changed in similar ways in regular and test trials.

This visual impression was confirmed in a linear mixed model in which we predicted reaction time by condition and trial. The model had the following structure:

```
RT ~ condition* z_trial + (z_trial | SampleSound )
```

Because reaction times tend to be skewed and have an offset, we modeled the response as a `shifted_lognormal` distribution. Figure 4 visualizes the model estimates, showing a negative effect of condition and a small negative effect of trial but no interaction between condition and trial. Thus, reaction times were generally faster in test trials compared to regular trials, but they did not change differently over time for the two conditions.
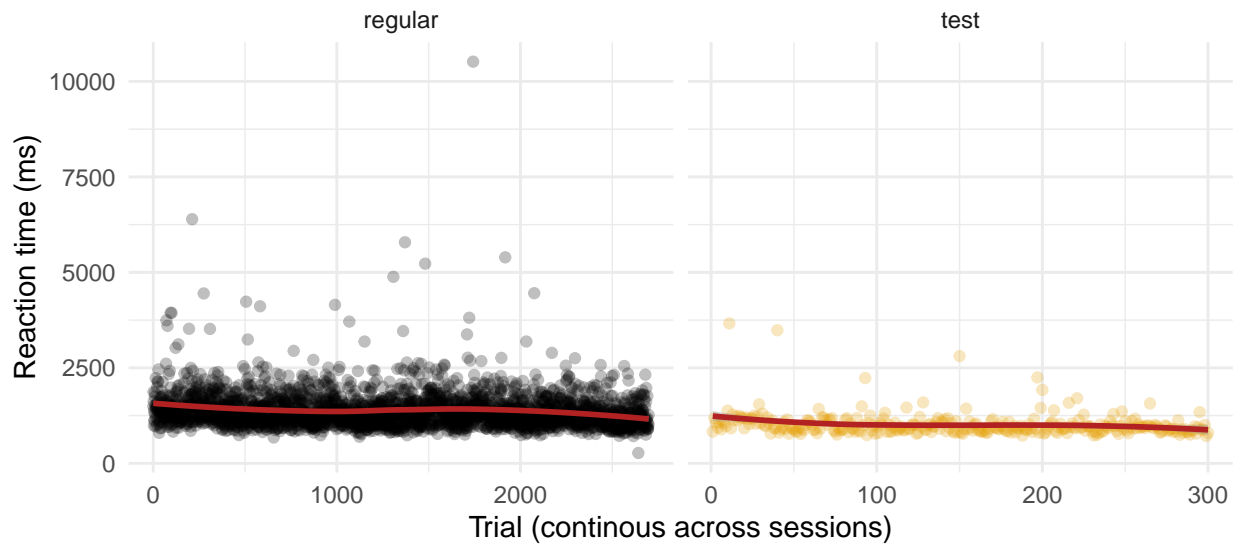
Figure 3: Reactions time across trials (continous across sessions) for regular and test trials. Light dots represent individual trials. Regression line gives a smoothed conditional mean.
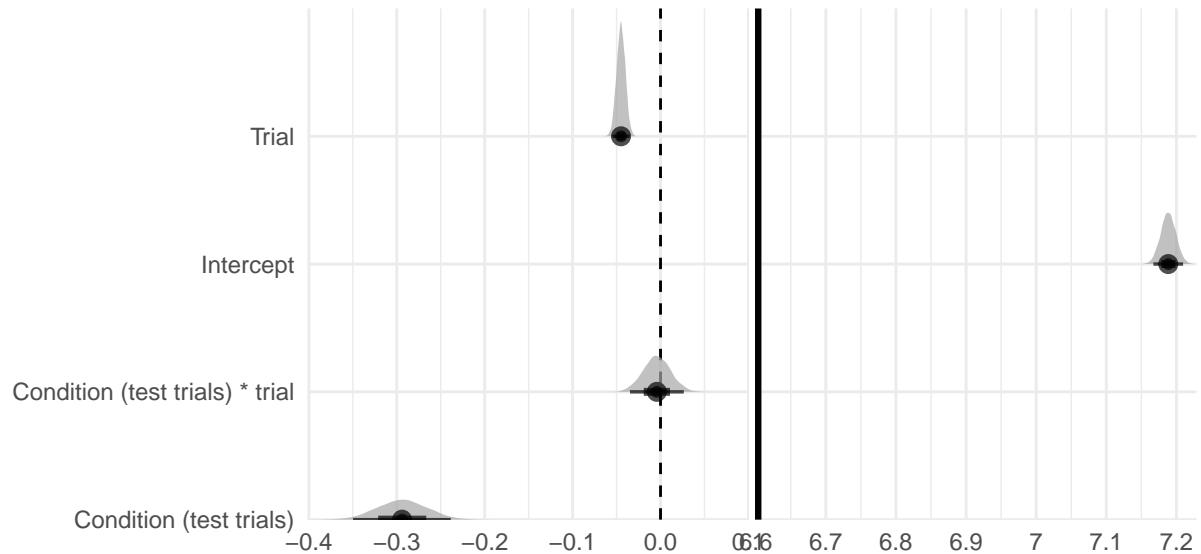

Figure 4: Estimates for condition, trial and interaction between trial and condition (with 95% CrI) in the model for reaction times. Dashed line shows chance level. Solid black line highlights a break in the x-axis

Results

Test trials

Comparison to chance

We used a Bayesian generalized linear mixed model with a logit link to analyze performance. Our model hat the following structure:

```
correct ~ 1 + (z_trial | SampleSound )
```

In this a model, the intercept models the average rate of correct responses. A value of 0 in link space corresponds to a proportion of correct choices of 0.5. Thus, we inferred that if the 95% credible interval (CrI) around the posterior estimate for the intercept did not include 0, performance was reliably above chance. Figure 5 visualizes the posterior distribution based on our model. It shows that the 95%CrI overlaps with 0. We therefore concluded that Kanzi's performance in test trials was not reliably above chance.
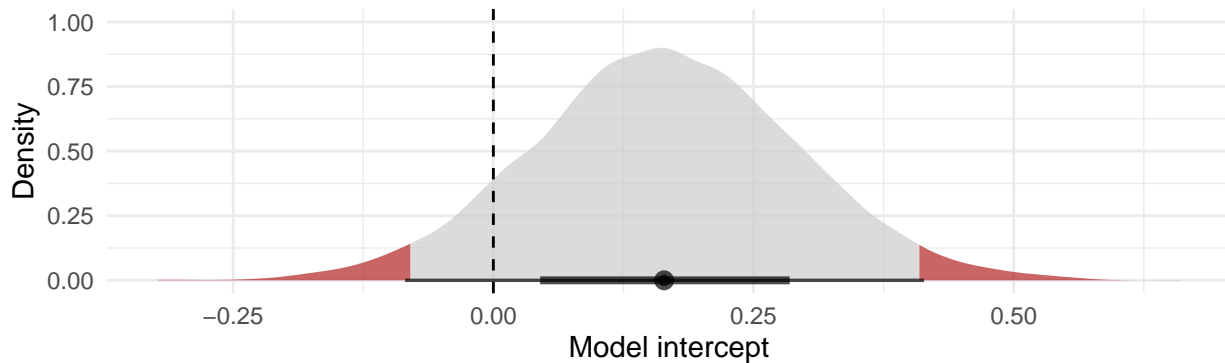

Figure 5: Estimate for intercept term (with 95% CrI) in the model for test trials. Dashed line shows chance level (in link space).

Prior sensitivity analysis The model reported above used default priors as implemented in the `brms` package. Here we explore the effect of different prior distributions for the intercept estimate on the results. The default prior in `brms` is `student_t(3, 0, 2.5)`. We compare this to six other priors. First, three priors that are also centered at `0` (i.e. assume no effect) and then three priors that assume a very strong effect (centered at `1` which corresponds to 73% correct responses). For every effect type we compare a weakly informative prior: `normal(0/1, 10)`, an informative prior: `normal(0/1, 5)`, and a highly informative prior: `normal(0/1, 1)`. Figure 6 visualizes the estimates given the different prior distributions and shows that the prior settings had little to no influence on the posterior estimate in the model.

Subset analysis In McCormick et al. (2015), some pseudo words were rated as more sound symbolic than others. These words were: `keke,kiki,pepe,pipi,lulu,lolo,mumu,momo,nono,nunu`. Given that these words are perceived as more sound symbolic in humans, Kanzi's performance might have been higher in trials with just these words. We therefore re-ran the model on a subset of the test trials that included only these words. This reduces the number of trials to 150. Figure 7 shows that the average rate of correct responses was the same when estimated based on trials with only highly sound symbolic pseudo words compared to all trials.

Shape preference

For humans, it has been repeatedly found that sound symbolism effects are stronger for round shapes compared to edgy shapes. We therefore ran a model estimating the rate of correct responses on test trials with round compared to edgy shapes. This model had the following structure:
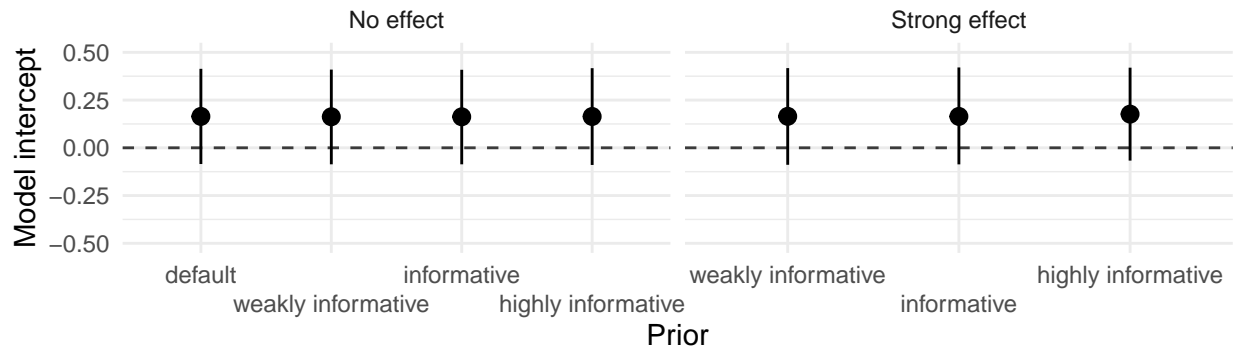
Figure 6: Estimate for intercept term (with 95% CrI) in the model for test trials with different prior distributions. Dashed line shows chance level (in link space). *No effect* refers to prior distributions centered at zero, *strong effect* refers to prior distributions centered at one.
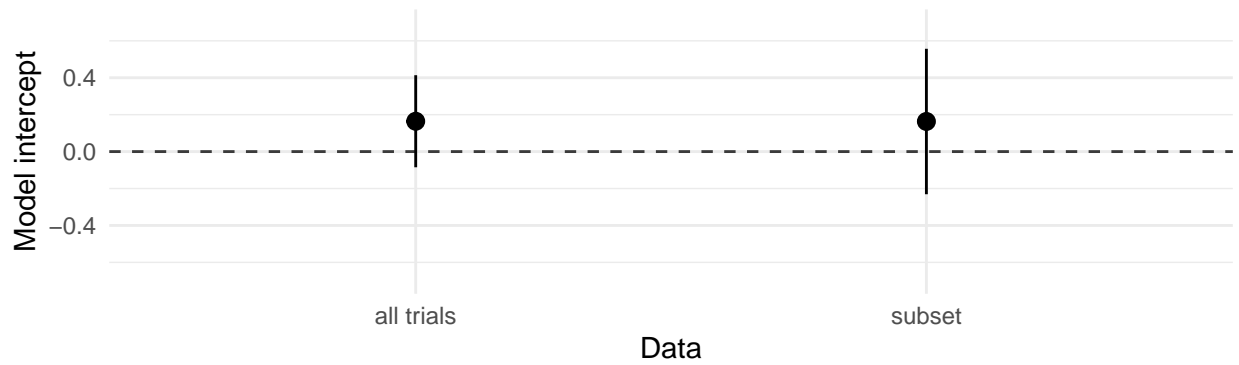

Figure 7: Estimate for intercept term (with 95% CrI) in the model for test trials based on all trials (left) and a subset of trials with words that had high sound symbolic ratings in McCormick et al. (2015). Dashed line shows chance level (in link space).

```
correct ~ test_shape + (z_trial | SampleSound )
```

Figure 8 shows that while performance was slightly higher with round compared to edgy shapes, it was nevertheless not reliably different from a level expected by chance.
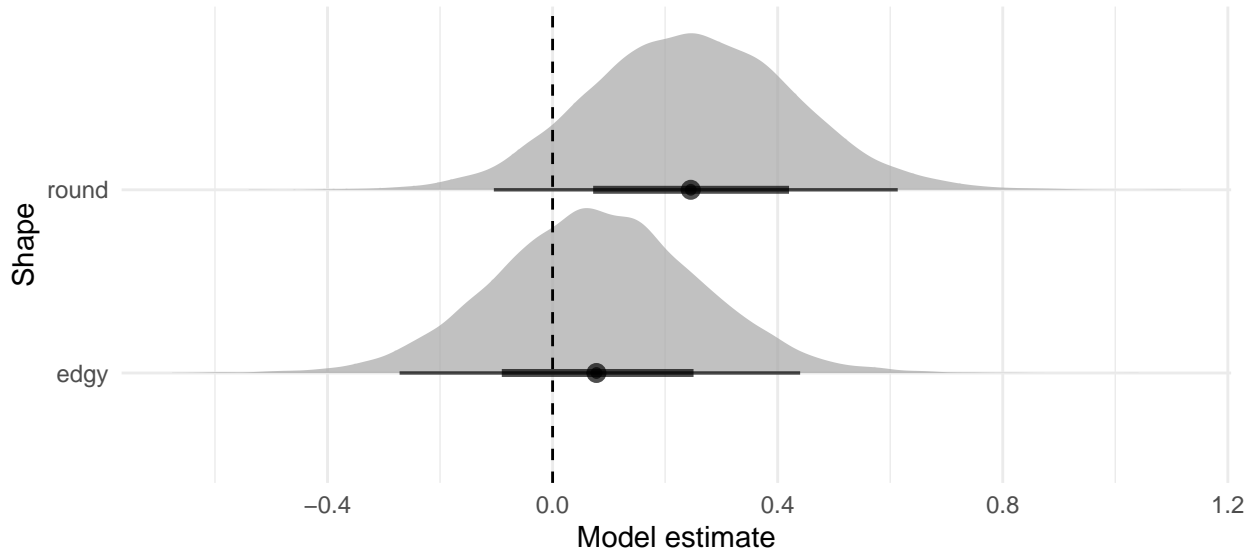

Figure 8: Posterior estimates (with 95% CrI) for test trials with round or edgy shapes. Dashed line shows chance level (in link space).

Regular trials

Comparison to chance

Test trials were embedded in regular trials in which Kanzi had to match a word to a picture (cropped and converted to gray scale). We analysed these regular trials in the same way as the test trials using the following model:

```
correct ~ 1 + (z_trial | SampleSound )
```

Figure 9 shows that the posterior estimate for the intercept was clearly positive and the 95% CrI did not overlap with 0. Thus, Kanzi was clearly able to matcht the sounds to the correct pictures.
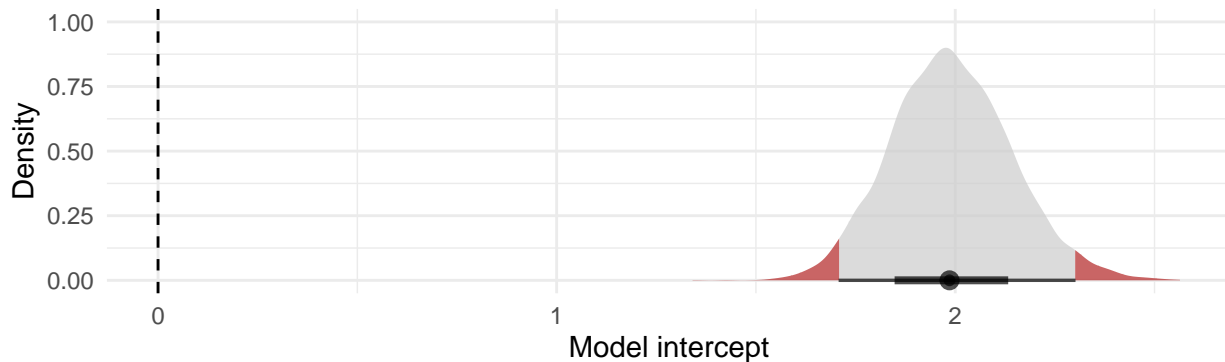

Figure 9: Estimate for intercept term (with 95% CrI) in the model for regular trials. Dashed line shows chance level (in link space).

Results by word

Figure 10 shows the proportion of correct responses for each word used in regular and test trials. Each regular word appeared between 26 and 28 times and each sound symbolic test word appeared 15 times.

Visualizations for manuscript

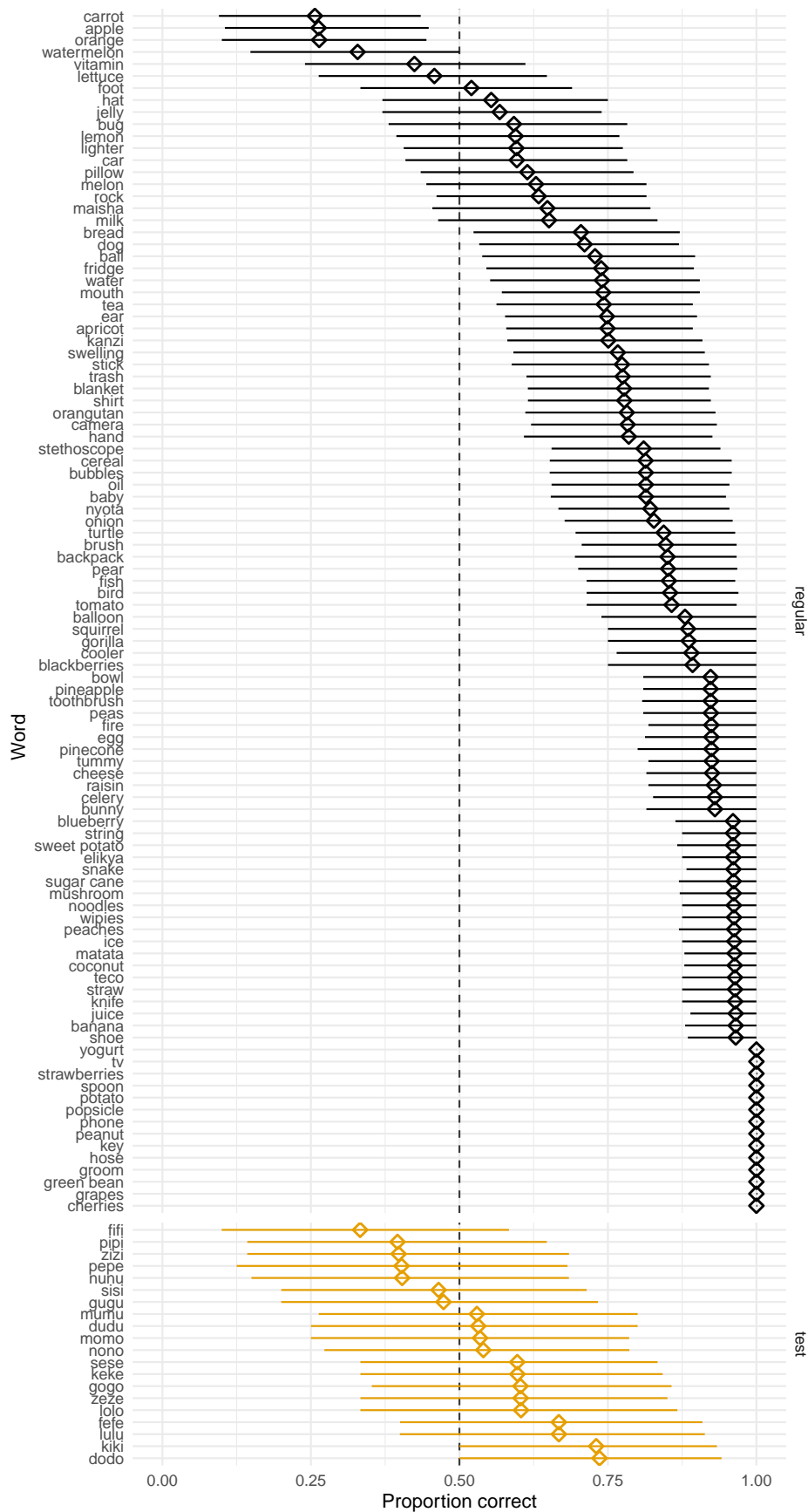Figure 11 reproduces Figure 2 from the main manuscript.

Figure 10: Proportion of correct responses (with 95% confidence interval) for each word for regular and test trials.
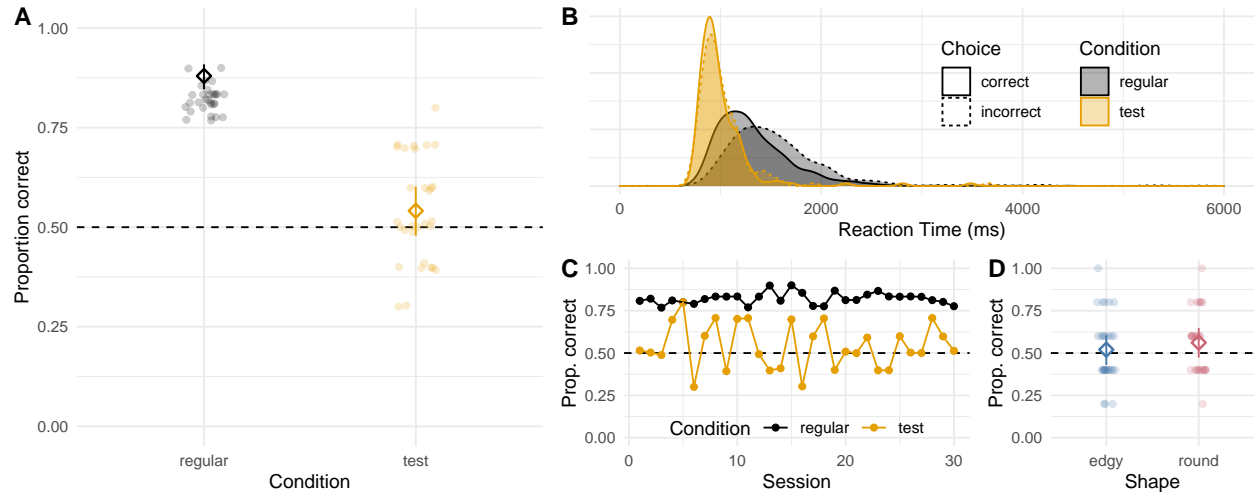
Figure 11: (A) Proportion of correct choices in regular and test trials. (B) Distribution of reaction times in regular and test trials for correct and incorrect choices. (C) Proportion of correct choices in each session for regular and test trials. (D) Proportion of correct choices in test trials with either edgy or round target shapes. In A and D: Diamonds and error bars represent the mean and 95% CrI based on the models' posterior distribution. Transparent dots show horizontally jittered session means of the data.