1   **Supplementary Materials**

2   **Analysis of Observed Variables**

3   In addition to our evidence accumulation modelling analysis, we first analysed accuracy and

4   reaction time (RT). Analyses was conducted using a Bayesian estimation approach to

5   multilevel modelling (McElreath, 2020). Specifically, we used the Bayesian modelling

6   package "brms" to run multilevel models in the R (Bürkner, 2017).

7   Models were built incrementally towards the most complex model. This meant that all fixed

8   and varying effects that the design would permit were included in the full model (Barr et al.,

9   2013). Model 0 for both accuracy and reaction time were intercept only models so that we

10   could compare all subsequent models that included effects of interest to a model without any

11   predictors. The most complex model included an effect of training type (naming, tying, both,

12   untrained). Shifted lognormal model was used to fit reaction time, while accuracy data was

13   fit with a Bernoulli model. Priors were set using a weakly informative approach (see Table 1

14   for all priors used for modelling) (Gelman, 2006). The formula for the full model (Model 1)

15   used to fit reaction time data is specified below:

$$rt \sim 1 + trainingtype + (1 + trainingtype|pid), ndt \sim (1|pid)$$

16   Note: RT = reaction time (ms); training type = training condition (untrained vs. naming

17   vs. tying vs. both); pid = subject/participant identifier; ndt = non-decision time.

18   The priors we used followed a weakly informative approach and are reported in

19   Supplementary Table 1.

20   There was some evidence for relatively small effects of training type on accuracy (see

21   Supplementary Figure 1A). The posterior distribution for the effect of each training type

22   overlapped with zero, but did have values which fell mainly below or above zero. This

Table 1

Priors used for models.

Variable	Prior	Class	dpar
Accuracy	normal (0,1)	Intercept	
	normal (0,.5)	sd	ndt
	normal (0,.5)	b	
	Lkj(2)	cor	
Reaction time	normal (6.68,0.5)	Intercept	
	normal (5.99,0.5)	Intercept	ndt
	normal (0,.5)	b	
	normal (0,.5)	sd	
	normal (0,.5)	sd	ndt
	normal (0,.5)	sigma	
	Lkj(2)	cor	

Note. dpar = distributional parameter; sd = standard deviation; b = fixed effect; cor = correlation; ndt = nondecision time.

23 suggests that relative to the untrained condition accuracy was lower in both the tying and

24 naming conditions, but slighly higher in the combined condition.

25 There also appeared to be an effect of training type on RT (see Supplementary Figure 1B).

26 Inspection of the posterior distribution for the fixed effects of training type revealed values

27 which primarily fell below zero. This suggests that relative to knots that received no

28 training, participants responded faster to knots that received naming, tying and both naming
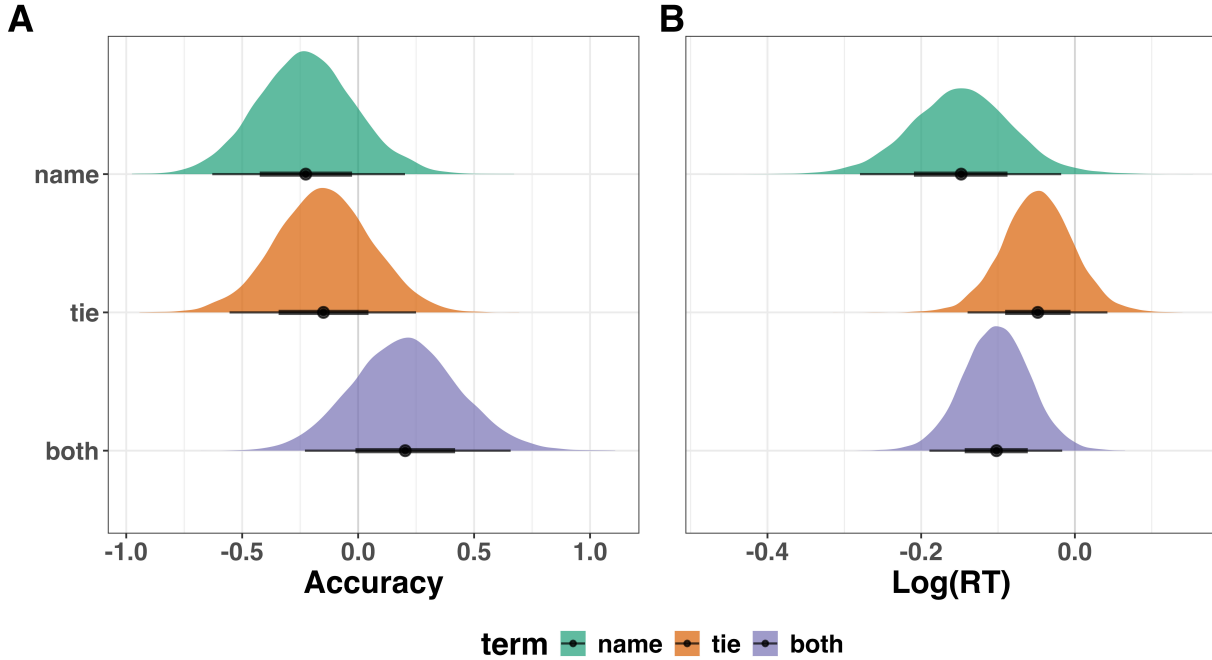
29 and tying training.

Fixed effects of condition for accuracy and RT

Figure 1. Fixed Effects for the Most Complex Model for (A) Accuracy and (B) Reaction Time. Points represent the median value of the posterior distribution for that estimate, the thicker line represents 66th percentile, while the thinner line is the 95th percentile of the distribution. The untrained condition was the reference group.

LBA Analysis

Sampling. For each model we used three times as many chains were used as model parameters. Sampling was carried out in two steps. First, sampling was carried out separately for individual participants in order to get reasonable start points for hierarchical sampling. The results of this step were then used as starting points for sampling the full hierarchical sample. During the initial burn-in-period there was a probability of .05 that a crossover step was replaced with a migration step. After burn in only crossover steps were used and sampling continued until the proportional scale reduction factor ($\hat{R}$) was less than 1.1 for all parameters, and also the multivariate version was less than 1.1 (Brooks & Gelman,

₃₉ 1998). Hierarchical estimation assumed independent normal population distributions for each

₄₀ model parameter. Population-mean start points were calculated from the mean of the

₄₁ individual-subject posterior medians and population standard deviation from their standard

₄₂ deviations, with each chain getting a slightly different random perturbation of these values.

₄₃ Hierarchical sampling used probability .05 migration steps at both levels of the hierarchy

₄₄ during burn in and only crossover steps thereafter with thinning set at 10 (i.e., only every

₄₅ 10th sample was kept), with sampling continuing until $\hat{R}$ for all parameters at all levels, and

₄₆ the multivariate $\hat{R}$ values, were all less than 1.1. The final set of chains were also inspected

₄₇ visually to confirm convergence.

₄₈ **Priors.**   Priors were chosen to have little influence on estimation. Priors were normal

₄₉ distributions that were truncated below zero for B, A and sv parameters, and truncated at

₅₀ 0.1s for the t0 parameter (assuming that responses made in less than 0.1s are implausible).

₅₁ The t0 parameter was truncated above by 1s. There were no other truncations, so the v

₅₂ prior was unbounded. The prior mean for B was 1 and for A 0.5. The v parameter for the

₅₃ true accumulator was given a prior mean of 1, while the mismatching accumulator was given

₅₄ a prior mean of 0. The sv parameter for the matching accumulator had a prior mean of 0.5.

₅₅ The t0 parameter had a prior mean of 0.3s. All priors had a standard deviation of 2. Mean

₅₆ parameters of population distributions were assumed to have priors of the same form as for

₅₇ individual estimation, and the standard deviations of hyper parameters were assumed to

₅₈ have exponential distributions with a scale parameter of one. Plots of prior and posterior

₅₉ distributions revealed strong updating (i.e., posteriors dominated priors), making it clear

₆₀ that the prior assumptions had little influence on posterior estimates (see Supplementary

₆₁ Figure 2 for an example of how the priors were updated for hyperparameters).
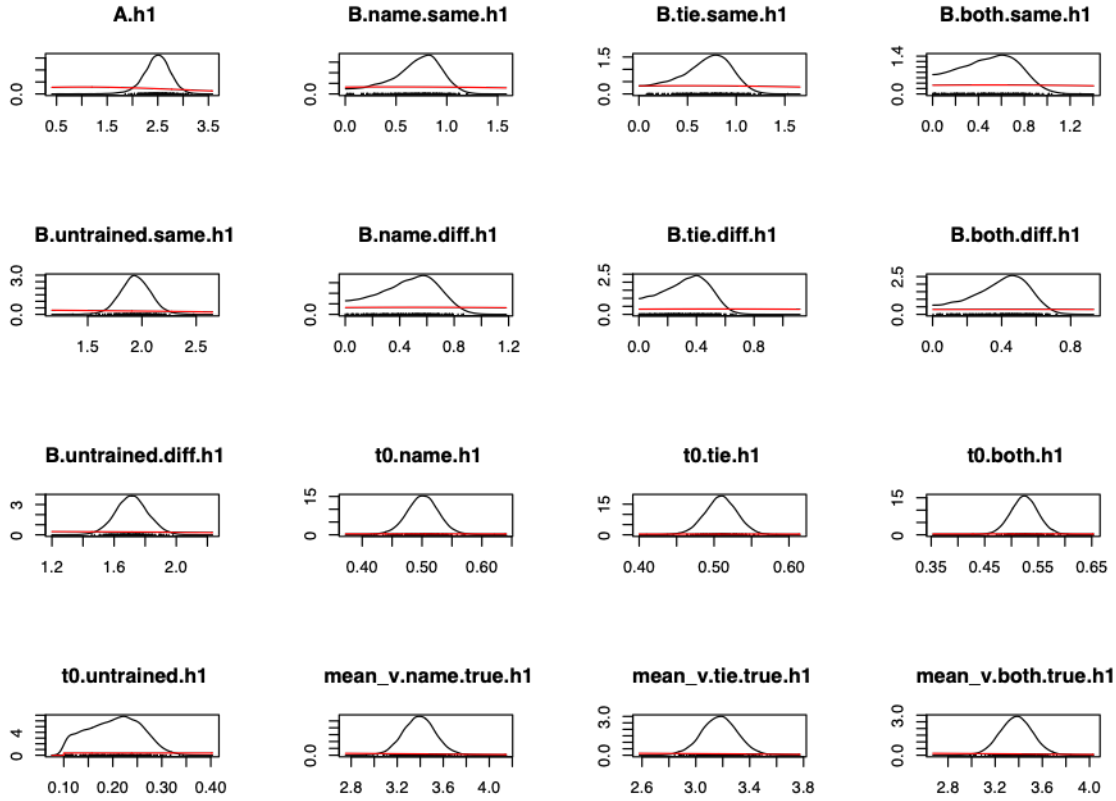
Figure 2. Prior and Posterior Graphs. Red lines represent priors, while black lines represent posteriors. Only a selection of all the parameters are shown here.

Model fit.   Supplementary Figures 3 display the fits of the LBA model to the data in terms of defective cumulative distribution functions (lines) and 10th, 30th, 50th, 70th and 90th percentiles (points from left to right) averaged over participants. Thick black lines and open points correspond to the data and the think grey lines solid black points to the model prediction averaged over posterior samples. The grey points correspond to percentile predictions for 100 randomly selected sets of posterior parameter samples, so their spread gives an idea of the uncertainty in the model's predictions. As can be seen from the figure the average fit of the selected LBA model was reasonable.
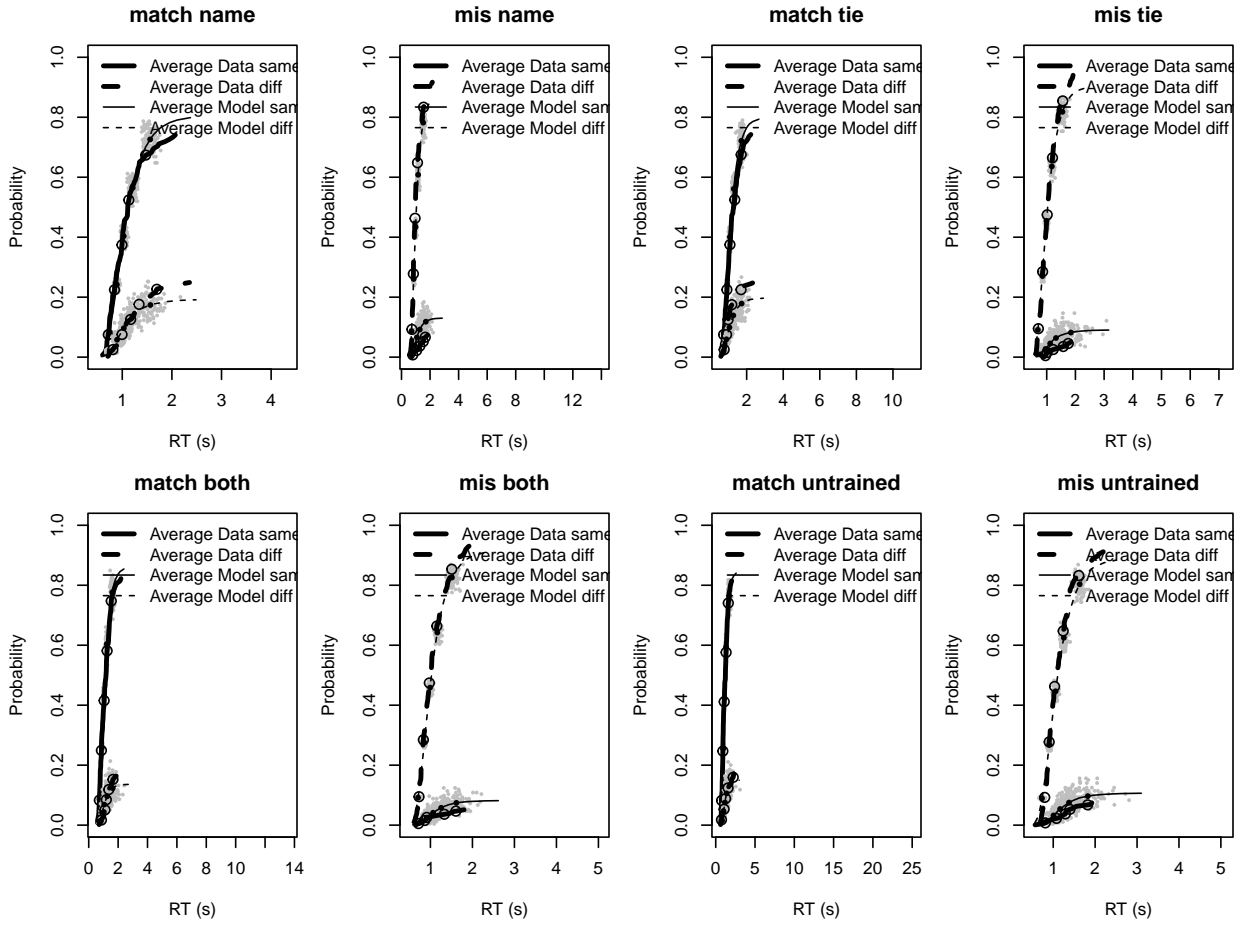
Figure 3. Cumulative distribution functions for data (thick lines) and fits (line grey lines) of the LBA model. Each panel contains results for both same and different responses at each level of stimulus (match and mismatch) and training type (name, tie, both, untrained). Symbols mark the 10th, 30th, 50th, 70th and 90th percentile (solid for average fits, open for data). Grey points are 500 percentile estimates from fits for random draws from posterior samples; the grey line and black solid points are the average of these 500 fits.