Fitting some data

In two experiments Spanton and Berry (2021) manipulate encoding variability and and memory strength at two levels (high, low) in a 2 × 2 factorial design, within-subjects, between-blocks.

Strength was manipulated by using a one-back digit judgement task concurrent to study (absent: high strength; present: low strength). Encoding variability was manipulated by selecting words for study that maximize the variance in four characteristics: word frequency, concreteness, age of acquisition and word length. In experiment 1, variance was assumed to be Gaussian. In Experiment 2, this was relaxed.

Model fitting

We fitted all individual data sets (all participants, all conditions) in maximum likelihood estimation for Gaussian-UVSDT (unequal-variance), Gaussian-EVSDT (equal-variance) and Gumbel-EVSDT (equal-variance, more precisely: equal-scale) to individual response frequencies. We fitted the models 20 times to each individual data set, optimizing using nlminb. Response frequencies of 0 for a given confidence category were replaced by 1/6 for model fitting for all models.

We implemented the small-extremes version of the Gumbel distribution using the ordinal package (Christensen, 2019), with "max = FALSE" and the normal distribution using base-R. For all models, we fixed $\mu_{new} = 0$ and $\sigma_{new} = 1$ ($\beta_{new} = 1$ for the Gumbel model) for all models, we allowed μ_{old} and the five decision criteria to vary (precisely, we estimated the lowest criterion freely, and estimated the differences to the next highest criterion, with lower bounds of 0). The models differed in the estimation of the variance/scale of the old-items distribution (Gaussian-EVSDT: $\sigma_{old} = \sigma_{new}$; Gaussian-UVSDT: σ_{old} estimated freely; Gumbel-EVSDT: $\beta_{old} = \beta_{new}$).

Results

We first examined the stability of the model fits. Figure 1A shows the stability of all models in all conditions and experiments as the difference between the two best fits of the model. With the correction for empty cells (e.g., response frequencies of 0 for some confidence ratings), the models are generally stable.

Figure 1B shows the model fit results as ΔAIC of individual AIC and the best fit (within experiment and condition, across models and participants), averaged across individuals. The data of the majority of participants across all conditions is best described by the Gumbel model, with the Gumbel model also resulting in the best generalizability on average.

As Figure 1C shows, the Gumbel model performs more strongly than the EVSDT model for many individuals. In comparison to the UVSDT model, the Gumbel model wins on AIC at relatively equal fit (the difference between them is approximately equal to the penalty of the additional parameter in the unequal variance model). This means the Gumbel model does not fit the data better than the UVSDT model, rather that for many data sets, it fits the data equally well but does so at lower cost.

Examining model wins more closely

While the Gumbel model performs well in a majority of cases, the Gaussian models win in the comparison by AIC for a number of individual data sets across all conditions and experiments. We examined the conditions of those fits more closely. Figure 2 and 3 show comparison of fits, and estimates of σ_{old} and μ_{old} for subgroups by winning model for all conditions.

Figure 4 shows the residuals of model predictions to observed response proportions, averaged across data sets where EVSDT, UVSDT or Gumbel won respectively.

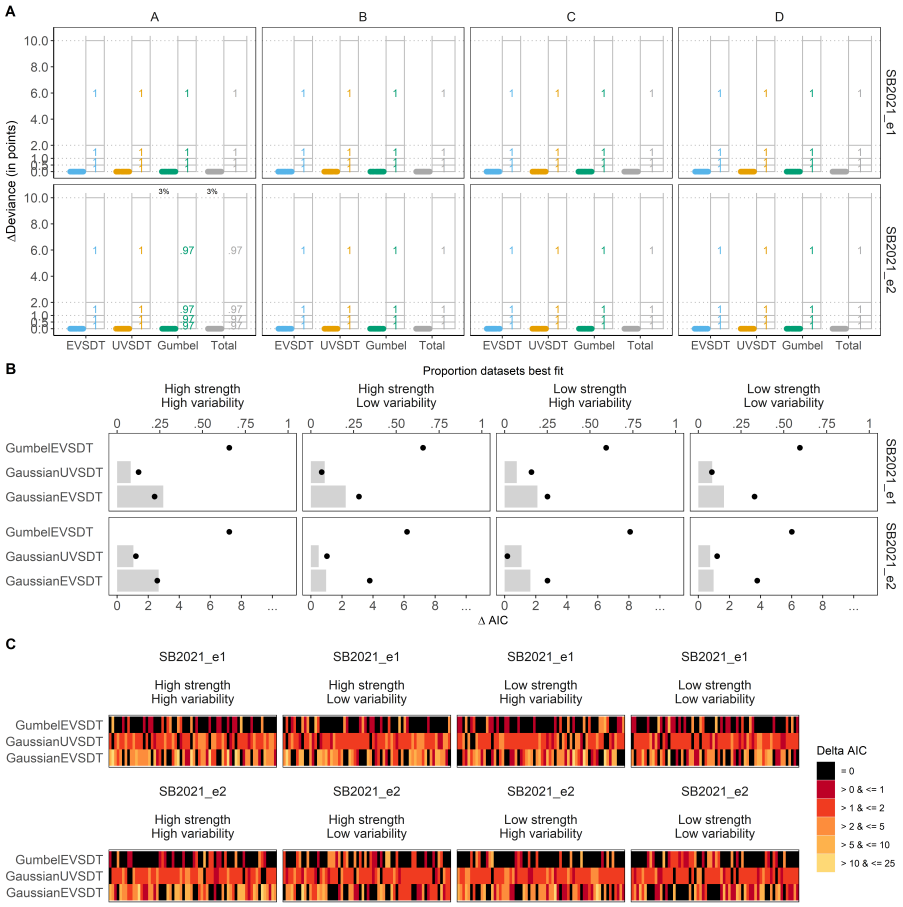

Figure 1

Model comparison of EVSDT, UVSDT and Gumbel. ***A****. Stability of model fits as difference between best and second-best fit of each model to each data set.* ***B.*** *Model comparison by average* ΔAIC *(bars) and proportion of individual data sets best fit (points),* ***C.*** *Individual* ΔAIC *for all conditions and experiments*

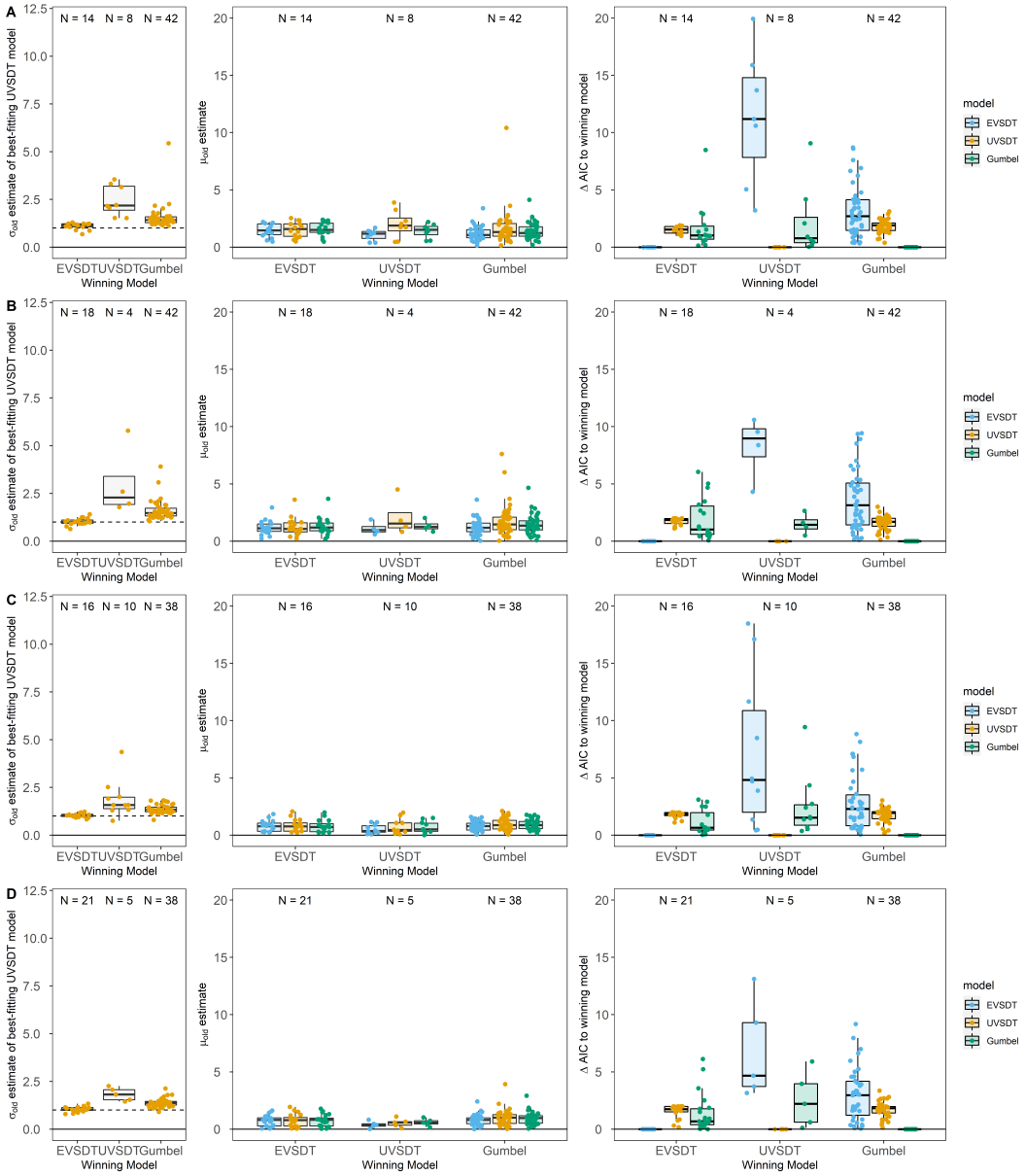

Figure 2

Fits split by winning model in Exp 1. ***A****. Estimate of* σ_{old} *in best-fitting UVSDT model.* ***B.*** *Estimate of* μ_{old} *in all models (in Gumbel:* $-\mu_{old}$*)* ***C.*** *Model comparison by* ΔAIC *(*$\Delta AIC > 20$ *omitted).*

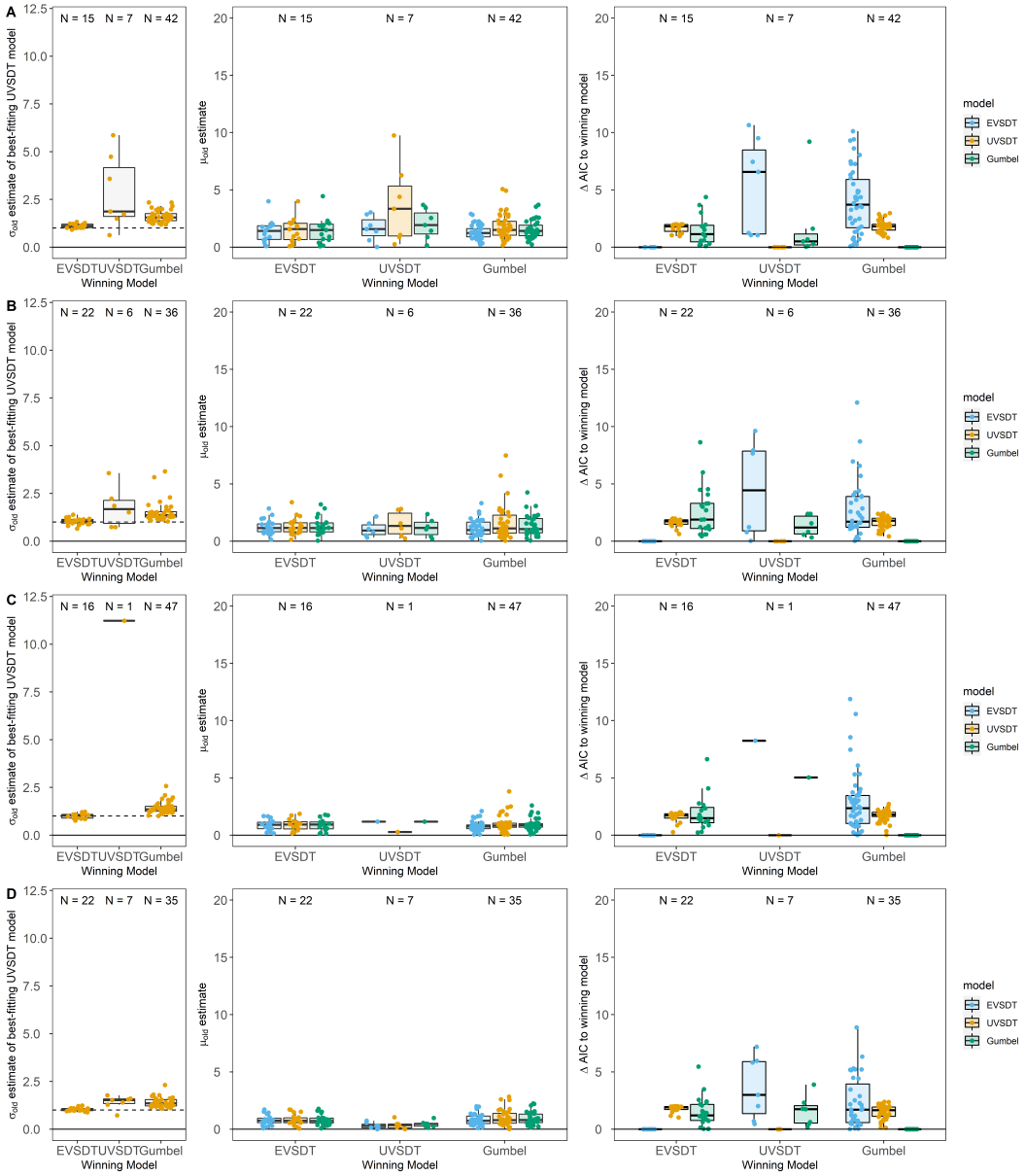

Figure 3

Fits split by winning model in Exp 2. ***A****. Estimate of* σ_{old} *in best-fitting UVSDT model.* ***B.*** *Estimate of* μ_{old} *in all models (in Gumbel:* $-\mu_{old}$*)* ***C.*** *Model comparison by* ΔAIC *(*$\Delta AIC > 20$ *omitted).*

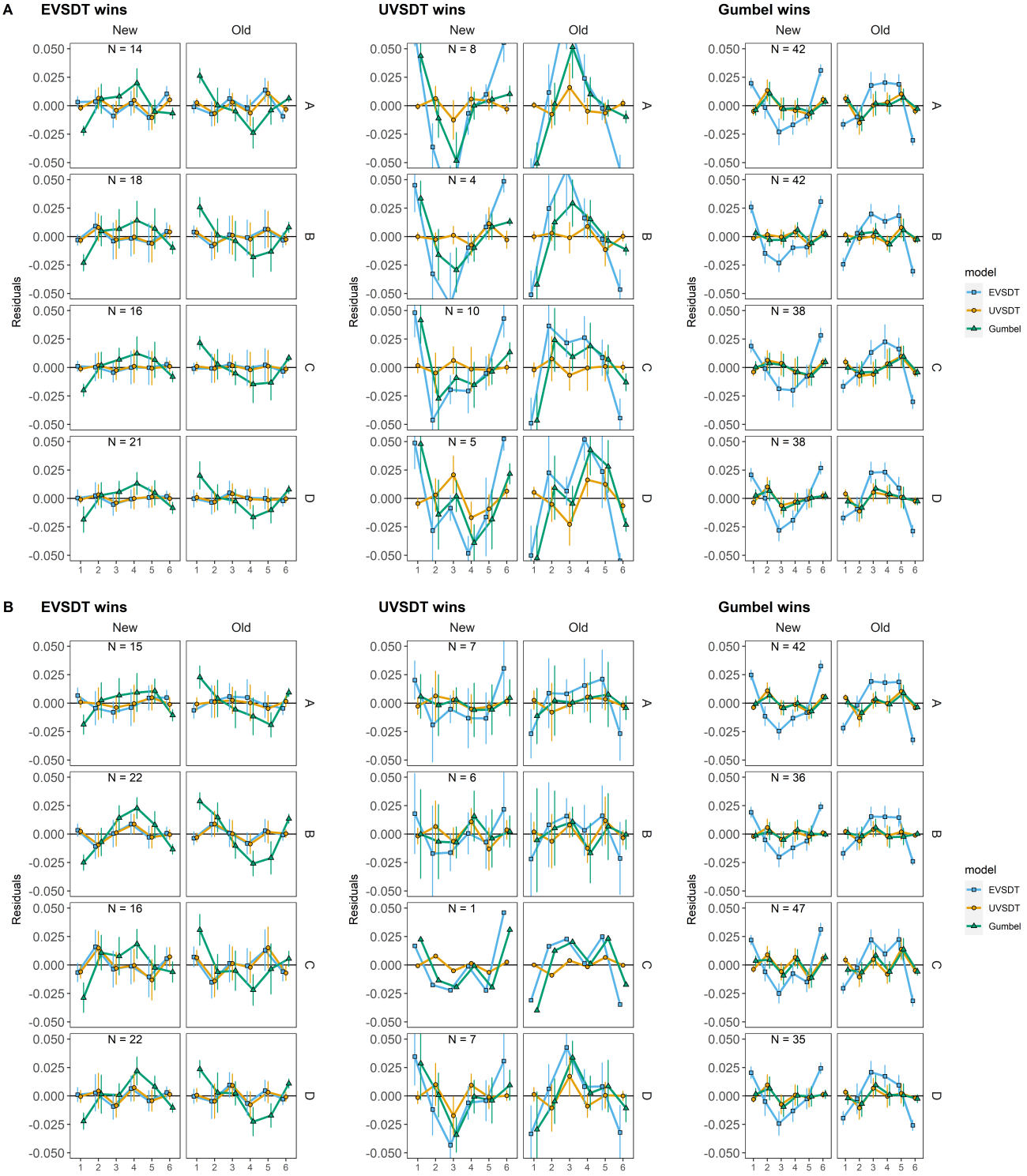

Figure 4

Residuals of response proportions averaged across conditions in Experiment 1 (A) and Experiment 2 (B), split by winning model. Error bars indicate 95% confidence interval.