Cognitive Science Special Issue: 2013 Rumelhart Prize Honoring Linda B. Smith

Highlighting in Early Childhood: Learning Biases through Attentional Shifting: *Supplementary Materials*

Joseph M. Burling
University of California, Los Angeles

Hanako Yoshida
University of Houston

Contents

Corresponding authors: jmburling@ucla.edu, yoshida@uh.edu

Experiment 1 Supplement

The following is a set of supplementary materials specific to Experiment 1 from the main text. See the accompanying text in the methods section from the main article that references each Table and Figure, as well as the Results section from the main article which describes results taken from the analysis and presented here in the supplement. Cue-referent assignments are displayed in Table S4, task instructions in Table S5, conditional results in Table S6, average training length in Table S7, model fitted correlation/standard deviation matrix in Table S8 and all image sets shown in Figure S4.

Stimuli & Materials

This section contains a set of additional tables and figures that show the full extent of the stimuli used in the image-based task, as outlined in Experiment 1. All stimuli were designed in Adobe Flash CS5 Software. Instructions used for the highlighting task in Experiment 1 are also displayed, which were administered verbally.

Symbol	Group 1	Group 2	Group 3
Cue: A	cup	apple	chair
Cue: B	glasses	spoon	shoe
Cue: C	strawberry	hat	cake
Referent: X	duck	elephant	cat
Referent: Y	cow	monkey	dog

Table S4
A complete list of illustrations used for the image-based task design. Each child was randomly assigned one of three groups, while adults were given all three groups in a random order.

Phase	Image Version Instruction
Familiarization	"First let's try dragging the triangles down to the same box. After you put them in there, press the *red* button at the top. Make sure you put them *inside* the box."
Training	"Now it's time to learn where some other pictures go. Drag the top pictures down to *one* of the boxes. Press the red button when you are done."
Testing	"Now you will see some new pictures and also some old ones. Try your best to put them into the right box. You can only choose *one* box to put them in."

Table S5
List of instructions used for the image-based version of the highlighting task.

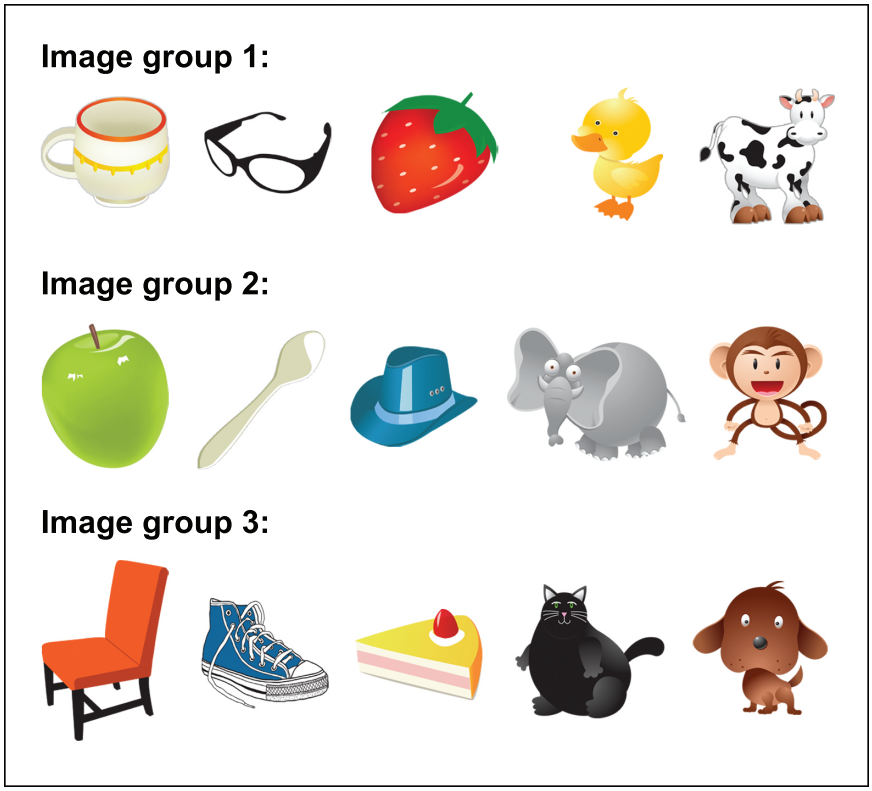

Figure S4. Full sample of illustrated images used in the image-based version of the highlighting task. Refer to Table S4 for the corresponding cue symbols.

Analysis

The fully Bayesian hierarchical logistic regression model was fitted using Stan version 2.5 (Stan Development Team, 2014b) from within the R programming language (R Core Team, 2015; Stan Development Team, 2014a). Twelve chains were run in parallel. Each chain allowed for a warm-up period of 10,000 samples in order to tune and adapt the Hamiltonian sampling algorithm. Each chain collected 210,000 samples, and were thinned every 100 samples to reduce autocorrelation. All chains had a high effective sample size and $\hat{R} < 1.002$. Even though convergence criteria were met for all chains, the two chains with the lowest average effective sample size were automatically dropped and the rest of the samples from each chain were merged. This ultimately led to a total of 20,000 posterior samples used for the analysis (2,000 from 10 chains).

Model formulation. The outcome variable y_i indicates the probability of success and is an estimate of the proportion of correct trials answered out of the total number of trials administered. This proportion can be estimated given indicators for the type of test cue and any interactions with the test cue, such as how the set of stimuli and the person's age can interact with the probability of success for a particular type of testing item. The coefficients estimated for each testing cue are allowed to vary by subject. Fixed effects such as the main effect of stimulus set and the main effect of age are also estimated but do not vary by subject.

Let j be the index for a specific person out of 33 participants, k be the indicator of a

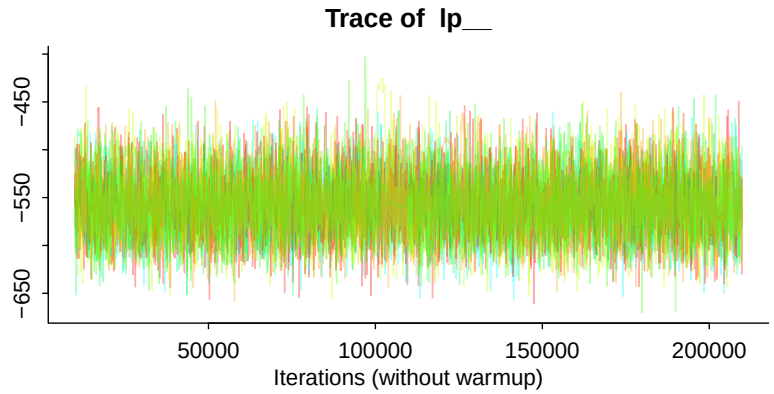

Figure S5. Traceplot of the samples over time for each chain. The figure displays the degree of mixing for the log probability of the model in Experiment 1.

particular testing item out of the five possible test items (the test cue $A.B$ is used as the baseline and folded into the intercept), and l be the indicator for one of the stimulus sets out of two possible sets of images (stimulus set 2 is used as the baseline and also folded into the intercept).

The regression equation for the probability of success is as follows for $i = 1, \ldots, n$, where n is the total number of observations in the data:

$$\begin{aligned} &\Pr(y_i = 1) = \text{logit}^{-1}(\hat{y}_i) \\ &\hat{y} = \alpha_{0,j} + \alpha_{k,j} * cue_k + \beta_l * stim_{l,j} + \beta_3 * age_j \end{aligned} \tag{1}$$

The function logit^{-1} is the inverse-logit/sigmoid function which constrains values in log-odds scale to be in the range of 0 to 1. The varying intercept $\alpha_{0,j}$ is defined by the following equation, which includes any interactions with the intercept:

$$\alpha_{0,j} = \gamma_{0,0} + \gamma_{0,l} * stim_{l,j} + \gamma_{0,3} * age_j \tag{2}$$

The additive effect for one of the other testing cues k is defined by the following (which equals zero if the effect of interest is only for cue $A.B$ and stimulus set 2, the intercept):

$$\alpha_{k,j} = \gamma_{k,0} + \gamma_{k,l} * stim_{l,j} + \gamma_{k,3} * age_j \tag{3}$$

The equation presented above may be reformulated using matrix notation for the parameters and data, which directly corresponds to the model code listed below and better illustrates the full extent of the model and estimated parameters. Let $\boldsymbol{Y}$ stand for the vector of successes out of $\boldsymbol{N}$, the vector of the number of trials for each combination of stimulus set, testing cue type, and participant. The notation $\hat{\boldsymbol{Y}}$ will denote the predicted values in the unconstrained log-odds scale. Thus, the likelihood for the model is ...

$$\boldsymbol{Y} \sim \text{Binomial}(\text{logit}^{-1}(\hat{\boldsymbol{Y}}), \boldsymbol{N})$$

where ...

$$\hat{\boldsymbol{Y}} = \boldsymbol{XB} + \boldsymbol{ZA}$$

The data matrix $\boldsymbol{X}$ is a 198×4 design/indicator matrix for the fixed effects which do not vary by subject. Let i be the index for one of the 198 rows of data (6 cue frequencies for each of the 33 participants). The stimulus set indicator variables made up of ones and zeros and are $CW.DK$ for the cow/duck set, and $CT.DG$ for the cat/dog set.

$$\boldsymbol{X} = \begin{bmatrix} 1 & stim_1^{CW.DK} & stim_1^{CT.DG} & age_1 \\ 1 & stim_2^{CW.DK} & stim_2^{CT.DG} & age_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & stim_i^{CW.DK} & stim_i^{CT.DG} & age_i \end{bmatrix}$$

and the 4×1 vector of regression coefficients corresponding to these predictors is ...

$$\boldsymbol{B} = \begin{bmatrix} \beta_{Int} \\ \beta_{CW.DK} \\ \beta_{CT.DG} \\ \beta_{age} \end{bmatrix}$$

The indicator matrix for the varying subject level coefficients is denoted as $\boldsymbol{Z}$, a 198×198 or $i \times (j * 6)$ matrix of ones and zeros (1 intercept indicator and 5 testing cue indicators for each subject), and the 1×198 vector $\boldsymbol{A}$ corresponds to the testing cue parameters that vary by subject accounting for the repeated measures for each type of cue.

$$\boldsymbol{Z}_{i,j,k} = \begin{bmatrix} Int_{1,1} & cue_{1,1,1} & \dots & cue_{1,1,k} & \dots & Int_{1,j} & cue_{1,j,1} & \dots & cue_{1,j,k} \\ Int_{2,1} & cue_{2,1,1} & \dots & cue_{2,1,k} & \dots & Int_{2,j} & cue_{2,j,1} & \dots & cue_{2,j,k} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ Int_{i,1} & cue_{i,1,1} & \dots & cue_{i,1,k} & \dots & Int_{i,j} & cue_{i,j,1} & \dots & cue_{i,j,k} \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} \alpha_1^{Int} = \gamma_{0,0} + \gamma_{0,1} * stim^{CW.DK} + \gamma_{0,2} * stim^{CT.DG} + \gamma_{0,3} * age_1 \\ \alpha_{1,1}^{Cue} = \gamma_{1,0} + \gamma_{1,1} * stim^{CW.DK} + \gamma_{1,2} * stim^{CT.DG} + \gamma_{1,3} * age_1 \\ \vdots \\ \alpha_{1,k}^{Cue} = \gamma_{k,0} + \gamma_{k,1} * stim^{CW.DK} + \gamma_{k,2} * stim^{CT.DG} + \gamma_{k,3} * age_1 \\ \vdots \\ \alpha_j^{Int} = \gamma_{0,0} + \gamma_{0,1} * stim^{CW.DK} + \gamma_{0,2} * stim^{CT.DG} + \gamma_{0,3} * age_j \\ \alpha_{j,1}^{Cue} = \gamma_{1,0} + \gamma_{1,1} * stim^{CW.DK} + \gamma_{1,2} * stim^{CT.DG} + \gamma_{1,3} * age_j \\ \vdots \\ \alpha_{j,k}^{Cue} = \gamma_{k,0} + \gamma_{k,1} * stim^{CW.DK} + \gamma_{k,2} * stim^{CT.DG} + \gamma_{k,3} * age_j \end{bmatrix}$$

The vector of varying subject-level coefficients $\boldsymbol{A}$ was constructed by converting the 6×33 matrix form of $\boldsymbol{A}$ into a single column vector. Where $\boldsymbol{A}_{k,j}$ is determined by the matrix of interaction parameters $\boldsymbol{G}_{k,l}$, which consists of regression coefficients that influence each cue type. $\boldsymbol{G}$ is multiplied by the group level matrix $\boldsymbol{U}_{l,j}$, which denotes the indicators for stimulus type and values for age for each subject.

$$\boldsymbol{A} = \boldsymbol{G}\boldsymbol{U}$$

Where the subject-level intercepts and slopes that interact with each type of test cue are contained in the 6×4 matrix $\boldsymbol{G}$ ...

$$\boldsymbol{G} = \begin{bmatrix} \gamma_{0,0} & \gamma_{0,1} & \gamma_{0,2} & \gamma_{0,3} \\ \gamma_{1,0} & \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{k,0} & \gamma_{k,1} & \gamma_{k,2} & \gamma_{k,3} \end{bmatrix}$$

... and multiplied by the 4×33 subject-level design matrix $\boldsymbol{U}$.

$$\boldsymbol{U} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ stim_1^{CW.DK} & stim_2^{CW.DK} & \dots & stim_j^{CW.DK} \\ stim_1^{CT.DG} & stim_2^{CT.DG} & \dots & stim_j^{CT.DG} \\ age_1 & age_2 & \dots & age_j \end{bmatrix}$$

Model priors. Normal distribution priors were used for all regression coefficients with a mean of zero. All standard deviations were obtained by assuming a weakly informative folded-t prior with $\nu = 5$, $\mu = 0$, and $\sigma = 3$, which are based on values in the unconstrained log-odds scale. A non-informative prior was used to model the participant-level correlation matrix, with the single parameter set to 1 and is uniform across all possible positive definite correlation matrices. The priors for all estimated parameters start with modeling the standard deviations for the regression coefficients. Each standard deviation is distributed as a folded-t distribution with the β (4) coefficients having standard deviations σ, the α (6) coefficients having standard deviations τ, and the γ (24) coefficients having standard deviations ζ.

$$\sigma, \tau, \beta \sim \text{Folded-}t(\nu = 5, \mu = 0, \sigma = 3)$$

The regression coefficients can then be estimated using the standard deviations given the following:

$$\begin{aligned} \beta &\sim \mathcal{N}(0, \sigma) \\ \alpha &\sim \mathcal{N}(0, \tau) \\ \gamma &\sim \mathcal{N}(0, \zeta) \end{aligned}$$

The 6×6 correlation matrix which models the dependencies between types of testing cues due to the same person generating responses from each cue was modeled using a uniform prior over all possible positive definite correlation matrices.

$$\boldsymbol{\rho} \sim \text{LKJcorr}(\nu = 1)$$

Along with the standard deviation priors τ, the covariance matrix used to model the varying subject-level coefficients can be computed given the following:

$$\boldsymbol{\Sigma} = \text{Diag}(\tau)\boldsymbol{\rho}\text{Diag}(\tau)$$

And thus the varying coefficients for each subject can be modeled under the assumption of a multivariate normal distribution with mean vector $\boldsymbol{GU}_j$ for subject j (corresponding to each of the α's for subject j) and covariance matrix $\boldsymbol{\Sigma}$.

$$\boldsymbol{A}_j \sim \mathcal{N}(\boldsymbol{GU}_j, \boldsymbol{\Sigma})$$

Condition	n	Mean Age	A	B	C	$A.B$	$A.C$	$B.C$
$B > A$	11	57.8	0.46	0.76	0.70	0.76	0.65	0.46
$A > B$	14	54.5	0.88	0.46	0.80	0.74	0.68	0.62
$A = B$	9	54.6	0.80	0.76	0.81	0.83	0.75	0.67

Table S6

Posterior means for each test item separated by the three conditions used in Figure 1c from Experiment 1 in the main text.

Phase	Avg. n $A.B \rightarrow X$	Avg. n $A.C \rightarrow Y$
Early	5.15	0
Mixed	3.18	1.06
Late	3.64	10.91

Table S7

Average number of training trials completed across all children for each phase and for each outcome from Experiment 1.

	Int	A.C	B	C	A	B.C
Int	**1.21**	-	-	-	-	-
A.C	-0.05	**1.11**	-	-	-	-
B	-0.04	-0.06	**1.49**	-	-	-
C	0.03	0.07	-0.30	**2.04**	-	-
A	-0.17	0.12	-0.47	0.28	**2.12**	-
B.C	0.04	0.44	-0.36	0.48	0.26	**1.65**

Table S8

Correlation matrix for the coefficients that vary by subject. The standard deviations for each parameter are shown in bold on the diagonal (logit scale).

Experiment 2 Supplement

The following contains additional materials from the main text regarding Experiment 2, which was conducted with adults. See main text for further discussion of methods and analysis. The text-based stimuli layout is shown in Figure S6, list of words used for the text task in Table S9, instructions used for the text task in Table S10, text-based results in Table S11, and image-based results from in Table S12.

Stimuli & Materials

This section displays additional tables and figures regarding the names of stimuli used for the text-based version of the highlighting task, the layout of the experimental screen, and the instructions that were shown to the adult participants for the text-based task.

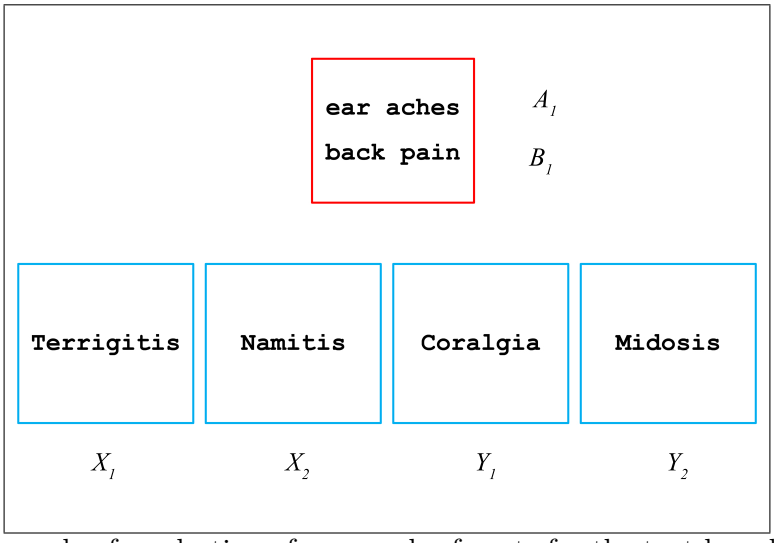

Figure S6. Example of a selection of cues and referents for the text-based version of the task. The symbol for each type of cue is marked with a letter next to the text (not shown in the actual task). Paired cues are symptoms (red box) while referents are novel diseases (blue boxes). Only adults were administered the text-based task.

List of symptoms	List of diseases
earaches, yellow eyes, rash, dizziness, sore muscles, nausea, hair loss, coughing, fever	Burlosis, Namitis, Terrigitis, Coralgia, Gouphosis, Midosis, Althrax

Table S9
A complete list of words used in the text-based task design. Symptoms and diseases were assigned at random during task onset.

Phase	Text Version Instruction
Training	"In this experiment you will see some common symptoms on the top of the computer screen and fictional diseases on the bottom of the screen. Your job is to learn which symptoms indicate which disease. When the symptoms are presented, you can make a guess by touching one of the diseases. You can press any of the diseases."
Testing	"Now you will diagnose diseases based on previous symptoms, some combinations may be new. You will choose the appropriate disease based on the given symptom/s. Please make an informed choice. You will touch the disease on the screen to make your choice."

Table S10
List of instructions used for the text-based version of the highlighting task.

Analysis

The model implementation in terms of program used, type of model fitted, the number of samples collected, and convergence criteria were the same for Experiment 2 as described in Experiment 1. The major differences are the structure of the hierarchical regression equation and priors used for the adult results. Since adults performed both the image-based and text-based version of the highlighting task, each participant has repeated measurements for each combination of task type and cue type. We analyzed the accuracy of cues that were common to both types of highlighting tasks for comparison purposes, with task type (text vs. image), cue type ($A.B, A.C, \ldots, B.C$), and the their interactions set as random factors that are allowed to vary by subject.

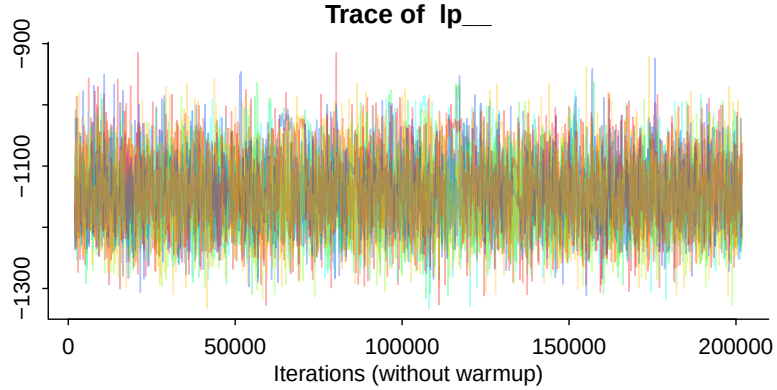

Figure S7. Traceplot of the samples over time for each chain. The figure displays the degree of mixing for the log probability of the model in Experiment 2.

Model formulation. As in the analysis for Experiment 1, let $\boldsymbol{Y}$ be a vector of the number of successes out of $\boldsymbol{N}$ trials. The probability of success is a binomial distribution with the likelihood expressed as the following:

$$\boldsymbol{Y} \sim \text{Binomial}(\text{logit}^{-1}(\hat{\boldsymbol{Y}}), \boldsymbol{N})$$

Where the predicted values in the logit scale, $\hat{\boldsymbol{Y}}$, are estimated by the equation:

$$\hat{\boldsymbol{Y}}_i = \boldsymbol{X}_i \boldsymbol{A}_j$$

And where $\boldsymbol{X}$ is the $n \times k$ design matrix of indicators based on the main and interaction effects, $task + cue + task * cue$, and $\boldsymbol{A}$ is the $k \times j$ matrix of coefficients which vary by person, with $n = 396,\ k = 12,\ j = 33$.

$$\boldsymbol{A}' = \begin{bmatrix} \alpha_{1,1} & \alpha_{2,1} & \dots & \alpha_{k,1} \\ \alpha_{1,2} & \alpha_{2,2} & \dots & \alpha_{k,2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1,j} & \alpha_{2,j} & \dots & \alpha_{k,j} \end{bmatrix}$$

Model priors. The subject-level regression coefficients from $\boldsymbol{A}$ are determined from the following set of k hyper-priors:

$$\begin{aligned} \beta &\sim \mathcal{N}(\mu, \sigma) \\ \sigma &\sim \text{Half-cauchy}(0, 1) \\ \mu &\sim \text{Uniform}(-\infty, \infty) + \mathcal{N}(0, 1) * \sigma \end{aligned}$$

And the subject-level coefficients from $\boldsymbol{A}$ are distributed as multivariate-normal with a mean vector of β hyper-parameters and covariance matrix $\boldsymbol{\Sigma}$.

$$\boldsymbol{A}_j \sim \mathcal{N}(\beta_{1,\dots,k}, \boldsymbol{\Sigma})$$

As in Experiment 1, the covariance matrix can be constructed given the standard deviations for each β and the correlation matrix $\boldsymbol{\rho}$.

$$\begin{aligned} \boldsymbol{\rho} &\sim \text{LKJcorr}(\nu = 2) \\ \boldsymbol{\Sigma} &= \text{Diag}(\sigma)\boldsymbol{\rho}\text{Diag}(\sigma) \end{aligned}$$

Test item	X	X_o	Y	Y_o
AB	0.94	0.01	0.03	0.02
AC	0.07	0.00	0.91	0.01
B	0.92	0.03	0.02	0.02
C	0.06	0.00	0.93	0.01
A	0.61	0.08	0.21	0.09
BC	0.35	0.02	0.64	0.00
AB_o	0.17	0.69	0.10	0.04
AC_o	0.09	0.05	0.11	0.76
BB_o	0.33	0.65	0.00	0.02
CC_o	0.03	0.03	0.56	0.38
AA_o	0.35	0.29	0.17	0.18
BC_o	0.29	0.03	0.02	0.66
ABC	0.37	0.03	0.58	0.02
AB_oC	0.05	0.21	0.72	0.02
ABC_o	0.46	0.02	0.04	0.48
AB_oC_o	0.26	0.07	0.60	0.07

Table S11

Adult accuracy given the text-based version of the highlighting task in Experiment 2. Proportions are displayed for each possible outcome given different cue combinations. The subscript o denotes that the cue belongs to the other *set than what is being tested.*

Test item	Accuracy $p(correct \mid cue)$
AB	0.99
AC	0.96
B	0.97
C	0.99
A	0.64
BC	0.68

Table S12

Adult accuracy for the image-based version of the highlighting task. Scores are in proportion correct (based on expected response) given each type of test cue administered.

Model code

Experiment 1 STAN code

The script used for Experiment 1 in Stan formatted code:

```
data {
int<lower=1>                nObs;      // n observations (198)
int<lower=1>                nSubs;     // n participants (33)
int<lower=1>                nAlpha;    // n cue type coef. (6)
int<lower=1>                nBeta;     // n unmodeled coef. (4)
int<lower=1>                nGamma;    // n interaction coef. (4)
matrix[nObs,nBeta]          X;         // obs lvl design matrix
matrix[nGamma,nSubs]        U;         // group lvl design matrix
matrix[nObs,nSubs*nAlpha]   Z;         // random design matrix
int                         Yi[nObs]; // vector of correct freq.
int                         Yn[nObs]; // vector of n trials
}
transformed data {
// none
}
parameters {
vector<lower=0>[nBeta]        sigma_fixed; // priors for unmodeled std. dev.
vector<lower=0>[nAlpha]       tau_cue;     // priors for cue type std. dev.
vector<lower=0>[nGamma]       zeta_intxn;  // priors for interaction std. dev.
vector[nBeta]                 Beta;        // priors for fixed coef.
matrix[nAlpha,nGamma]         Gamma;       // priors for grp lvl intxn.
matrix[nAlpha,nSubs]          Alpha;       // priors for varying grp lvl coef.
cholesky_factor_corr[nAlpha] Rho_chol;    // prior for cue type corr. matrix
}
transformed parameters {
cholesky_factor_cov[nAlpha] Sigma_chol; // cholesky factored covariance matrix
corr_matrix[nAlpha]         Rho;        // correlation matrix

// transform cholesky corr. matrix to cholesky cov. matrix
Rho <- tcrossprod(Rho_chol);
Sigma_chol <- cholesky_decompose(
      diag_matrix(tau_cue) * Rho * diag_matrix(tau_cue)
    );
}
model {
vector[nObs]          y_hat;        // predicted values on logit scale
matrix[nAlpha,nSubs] alpha_hat;    // temp. matrix

// folded-t priors with df=5, mean=0, sd=3
sigma_fixed ~ student_t(5.0, 0.0, 3.0);
zeta_intxn ~ student_t(5.0, 0.0, 3.0);
tau_cue ~ student_t(5.0, 0.0, 3.0);

// normal priors for regression coefficients, mean=0
Beta ~ normal(0, sigma_fixed);
for (k in 1:nAlpha) {
    for (l in 1:nGamma) {
        Gamma[k,l] ~ normal(0, zeta_intxn[l]);
    }
}

// uniform prior across pos. def. corr. matrices
Rho_chol ~ lkj_corr_cholesky(1.0);

// restructure: alpha matrix, varying subj. coef.
alpha_hat <- Gamma * U;

// multivariate normal prior for subject level coefficients
```

```
for (j in 1:nSubs)
    col(Alpha, j) ~ multi_normal_cholesky(col(alpha_hat,j), Sigma_chol);

// compute predicted values, logit scale
y_hat <- X * Beta + Z * to_vector(Alpha);

// binomial likelihood
for (i in 1:nObs)
    Yi[i] ~ binomial_logit(Yn[i], y_hat[i]);
}
generated quantities {
vector[nAlpha]  Alpha_mu; // means for cue type coef.

for (k in 1:nAlpha)
    Alpha_mu[k] <- mean(Alpha[k]);
}
```

Experiment 2 STAN code

The script used for Experiment 2 in Stan formatted code:

```
data {
int<lower=1>         nObs;      // n observations (396)
int<lower=1>         nSubs;     // n participants (33)
int<lower=1>         nBeta;     // n coef. (12)
matrix[nObs,nBeta]   X;         // design matrix
int                  Yi[nObs];  // vector of sucesses
int                  Yn[nObs];  // vector of n trials
int                  sid[nObs]; // subject index
}
parameters {
vector<lower=0>[nBeta]      beta_sigma;      // hyper-prior for std. dev. of beta
vector[nBeta]               beta_mu_raw;     // hyper-prior for unscaled beta means
real                        beta_adj;        // hyper-prior for beta adjust, uniform
vector[nBeta]               beta_sub[nSubs]; // prior for sub. lvl beta coef.
cholesky_factor_corr[nBeta] rho_chol;        // prior for beta coef. correlations
}
transformed parameters {
cholesky_factor_cov[nBeta] sigma_chol;  // cholesky factored covariance matrix
corr_matrix[nBeta]         rho;         // regular correlation matrix
vector[nBeta]              beta_mu;     // hyper-prior for means of beta coef.

// to avoid stuck chain segments
beta_mu <- beta_adj + beta_mu_raw .* beta_sigma;

// transform cholesky corr. matrix to cholesky cov. matrix
rho <- tcrossprod(rho_chol);
sigma_chol <- cholesky_decompose(
  diag_matrix(beta_sigma) * rho * diag_matrix(beta_sigma));
}
model {
beta_mu_raw ~ normal(0, 1);        // standard normal prior
beta_sigma ~ cauchy(0, 1);         // half-cauchy prior
rho_chol ~ lkj_corr_cholesky(2);   // diag. prior corr. matrix

// multivariate normal prior for subject level coefficients
for (j in 1:nSubs)
    beta_sub[j] ~ multi_normal_cholesky(beta_mu, sigma_chol);

// binomial likelihood
for (i in 1:nObs)
    Yi[i] ~ binomial_logit(Yn[i], X[i] * beta_sub[sid[i]]);
}
```

```
generated quantities {
vector[nObs]  log_lik;

// log-likelihood for each observation for WAIC
for (i in 1:nObs) {
    log_lik[i] <- binomial_logit_log(
        Yi[i], Yn[i], X[i] * beta_sub[sid[i]]
    );
}
}
```

References

R Core Team. (2015). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Stan Development Team. (2014a). Rstan: The R interface to Stan, version 2.5.0.

Stan Development Team. (2014b). Stan: A C++ library for probability and sampling, version 2.5.0.