Supplemental Information

Supplemental Note 1 – Bayesian data analysis

We constructed Bayesian hierarchical generalized linear models (GLMs) for both experiments and for both DVs per experiment. For log-reaction time (log-RT) and log-localization error, we used a Gaussian family and identity link function, whereas for accuracy, we used a Bernoulli family and logit link. All experimental factors of interest included the full random effects structure (both intercepts and slopes) for subjects as well as for items. For Experiment 1, this yields the following GLM equation:

$$DV \sim 1 + Congruency + (1 + Congruency \mid Subject) + (1 + Congruency \mid Item)$$

For Experiment 2, this yields the following:

$$DV \sim 1 + Congruency * Probe + (1 + Congruency * Probe \mid Subject) + (1 + Congruency * Probe \mid Item)$$

Models were constructed using Bambi, which was also responsible for setting appropriate weakly informative priors. Posterior probability distributions were obtained using Markov Chain Monte Carlo (MCMC) based on the No U-Turn Sampler (NUTS), as implemented in PyMC3. Four chains were sampled for each model, with 3,000 samples per chain, after a 6,000-sample tuning period, using a target acceptance ratio of 90%. Starting values were determined using Automatic Differentiation Variational Inference (ADVI; option init = 'advi+adapt_diag'), run for 35,000 time steps or until plateau. We checked chain convergence through visual inspection, as well as through the Gelman-Rubin statistic $\hat{r}$. Posterior results are primarily summarized using 94% Highest Density Intervals (HDI_{94}), based on the combined samples of all chains. Probability of parameters lying above/below a critical value is summarized by the proportion of combined samples above/below that value.

Supplemental Note 2 – Congruency costs might not always depend on spatial attention

In Experiment 2, if the congruency cost in the Probe-Key condition is due to an incongruent item 'grabbing attention' (either through foveation or covertly) more strongly than a congruent item, then we would expect a congruency *benefit* in the Probe-Other condition. Also in that condition, a by definition irrelevant incongruent item (present only in the incongruent condition) would grab attention, thereby impairing performance on discriminating the (always congruent) Probe-Other target item in the scene. As described in Results, our data do not provide evidence for such an effect.

As a further test of the attentional locus of this effect, we might ask whether stimulus items that yield a strong Probe-Key congruency effect, also yield a strong Probe-Other congruency effect. Such a correlation might exist even in the absence of a Probe-Other congruency effect in the average. However, also here we find evidence to the contrary,

both for 2AFC accuracy ($r_{(60)} = -.094$, $p = .235$, $CI_{95} = [-0.34, 0.16]$, $BF_{10} = 0.31$) and for reaction times ($r_{(60)} = .13$, $p = .154$, $CI_{95} = [-0.12, 0.37]$, $BF_{10} = 0.083$).

Taken together, the absence of a Congruency effect in Probe-Other, as well as the absence of a relationship between Probe-Key and Probe-Other effects, raise the interesting possibility that identified congruency costs in the exemplar identification task are not entirely mediated by attentional factors. However, given the large body of literature interpreting congruency costs as attentional in other tasks (see Introduction), we do not wish to make strong conclusions here; particularly since we did not record eye movements.

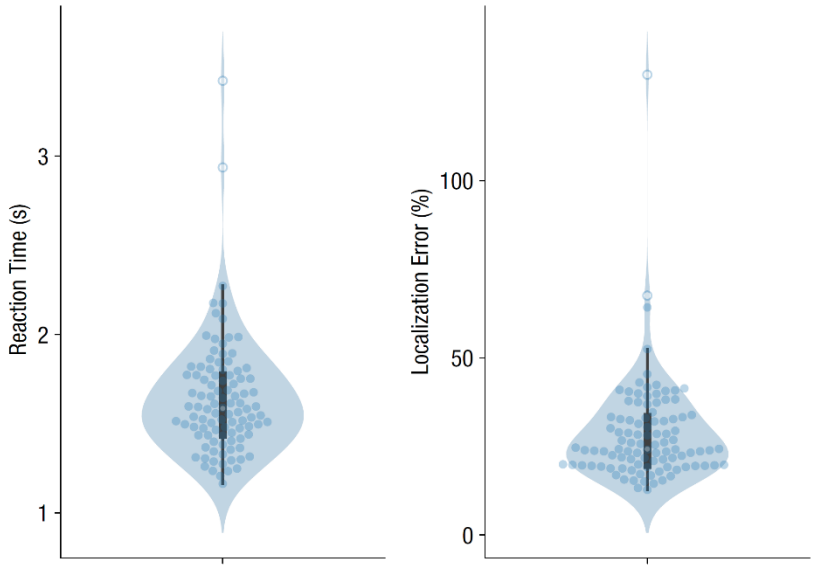

Figure S1. Overall reaction time and localization error distributions for Experiment 1. Dots are participants; hollow circles are outliers (removed from all analyses).

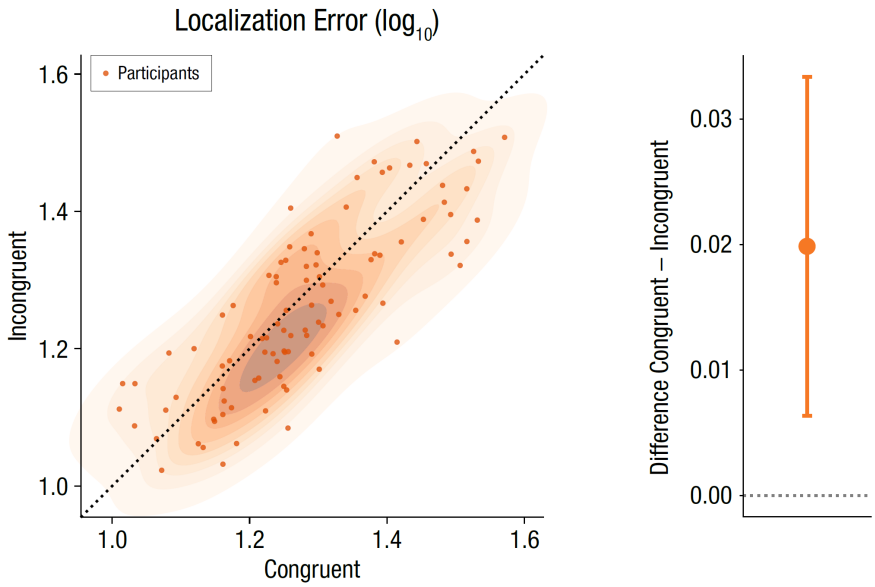

Figure S2. Localization errors across all participants in Experiment 1, for Congruent and Incongruent trials (analogous to Figure 1c). Full distribution in scatterplot, mean ± 95% confidence interval of difference scores on the right.

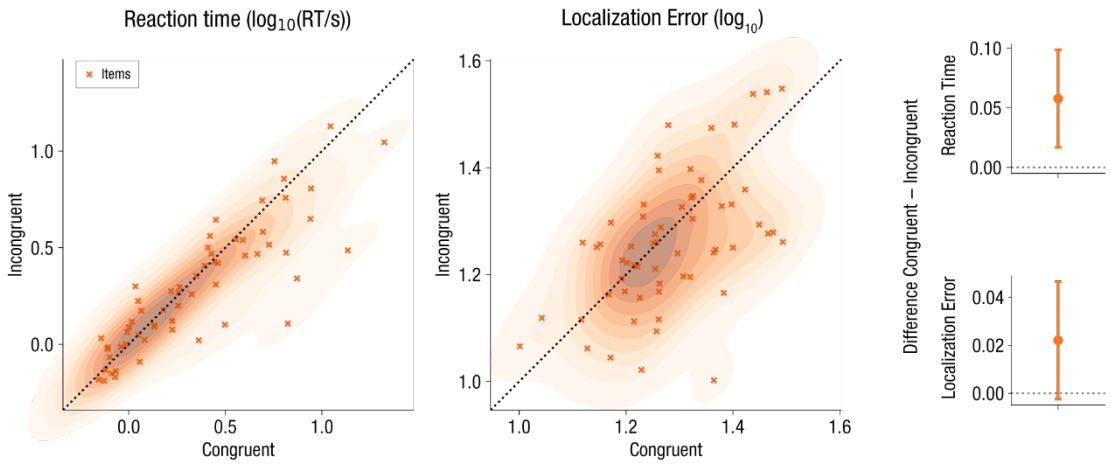

Figure S3. Reaction times and localization errors across experimental items (averaged over participants) for Experiment 1. Crosses are individual items. Full distributions in scatterplot, mean ± 95% confidence interval of difference scores on the right.

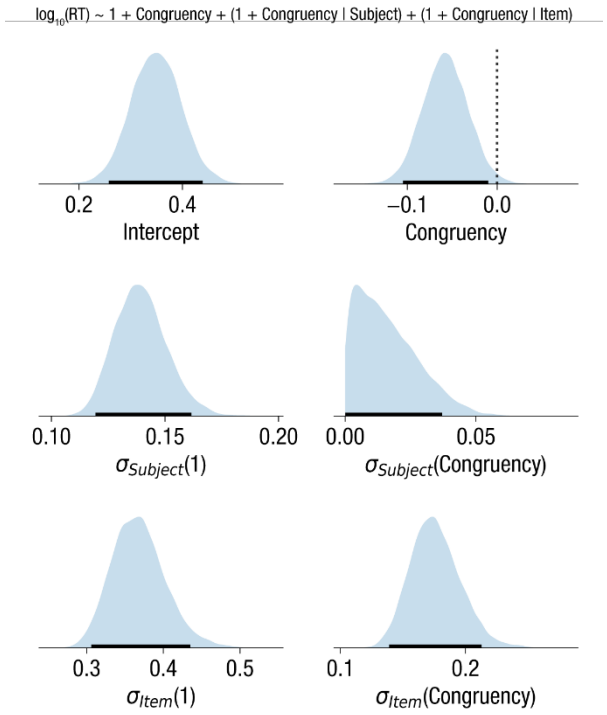

Figure S4. Posterior distribution after MCMC sampling of Bayesian regression model for reaction time in Experiment 1. Top row shows parameter estimates for fixed effects coefficients, middle row corresponds to subject-level random effects (i.e. standard deviation of effect over subjects), bottom row corresponds to item-level random effects. Intercepts and slopes for individual items/subjects are not shown (as these are very many), only the parameters related to their spread are included. Shading reflects a kernel density estimate of the marginal full posterior for one parameter; black horizontal bars indicate 94% HDI; dashed vertical lines, if present, correspond to a reference value (i.e., 0 for a fixed effect).

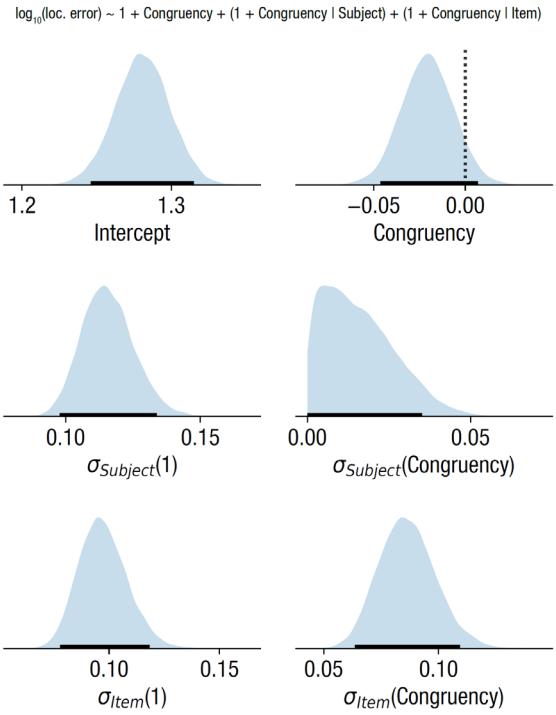

Figure S5. Posterior distribution after MCMC sampling of Bayesian regression model for localization error in Experiment 1. All panels and conventions as in Figure S4.

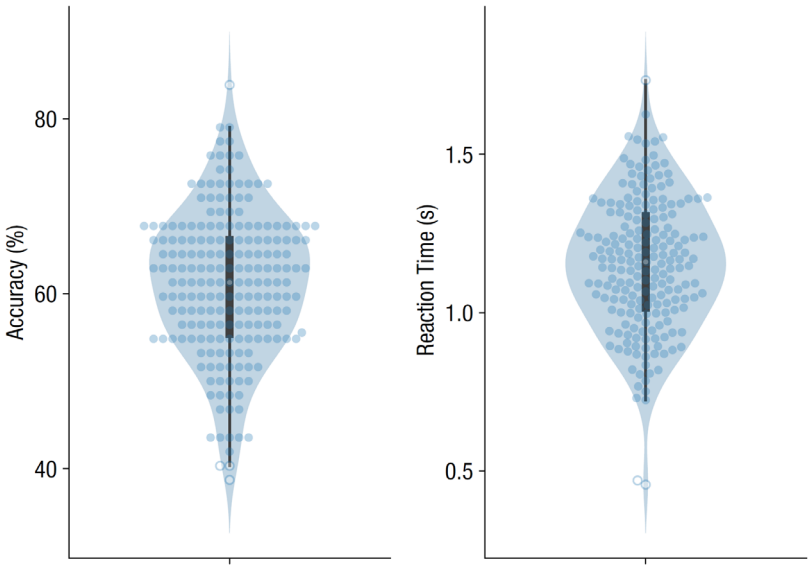

Figure S6. Overall accuracy and reaction time distributions for Experiment 2. Dots are participants; hollow circles are outliers (removed from all analyses).

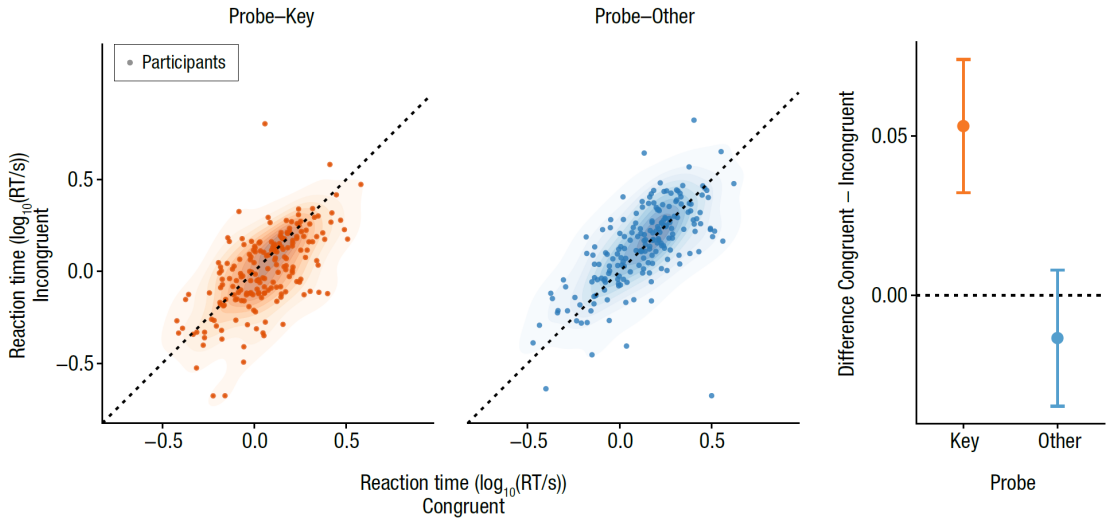

Figure S7. 2AFC reaction times in Congruent and Incongruent trials, separately for Probe-Key (orange, left) and Probe-Other (blue, right). Dots are individual participants. Right panel shows mean ± 95% confidence interval of difference scores for both Probe conditions. (Presentation analogous to Figure 3.)

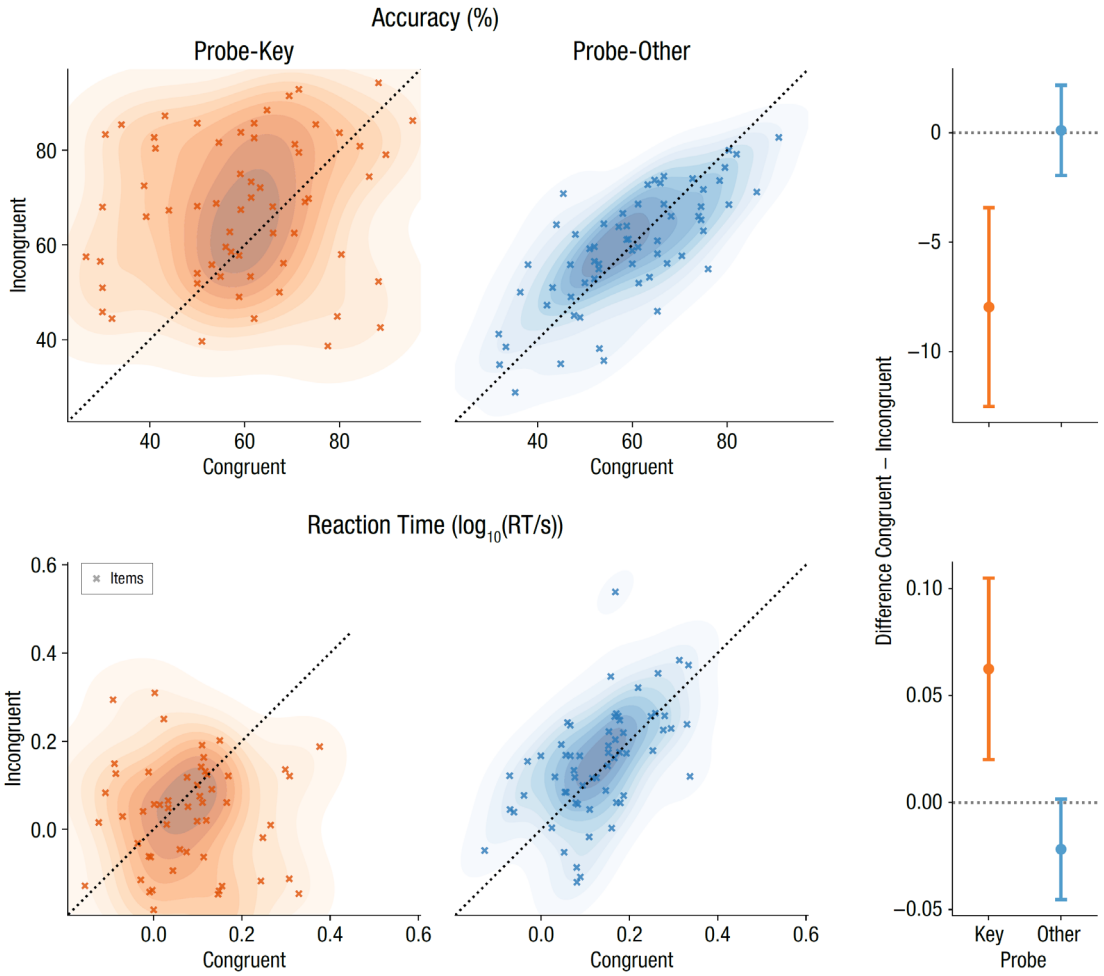

Figure S8. 2AFC accuracy (top) and reaction times (bottom) for Experiment 2, across experimental items (averaged over participants), for the Incongruent/Congruent and Probe-Key/Probe-Other conditions. Crosses are individual items. Full distributions in scatterplots, mean ± 95% confidence interval of difference scores in right panels.

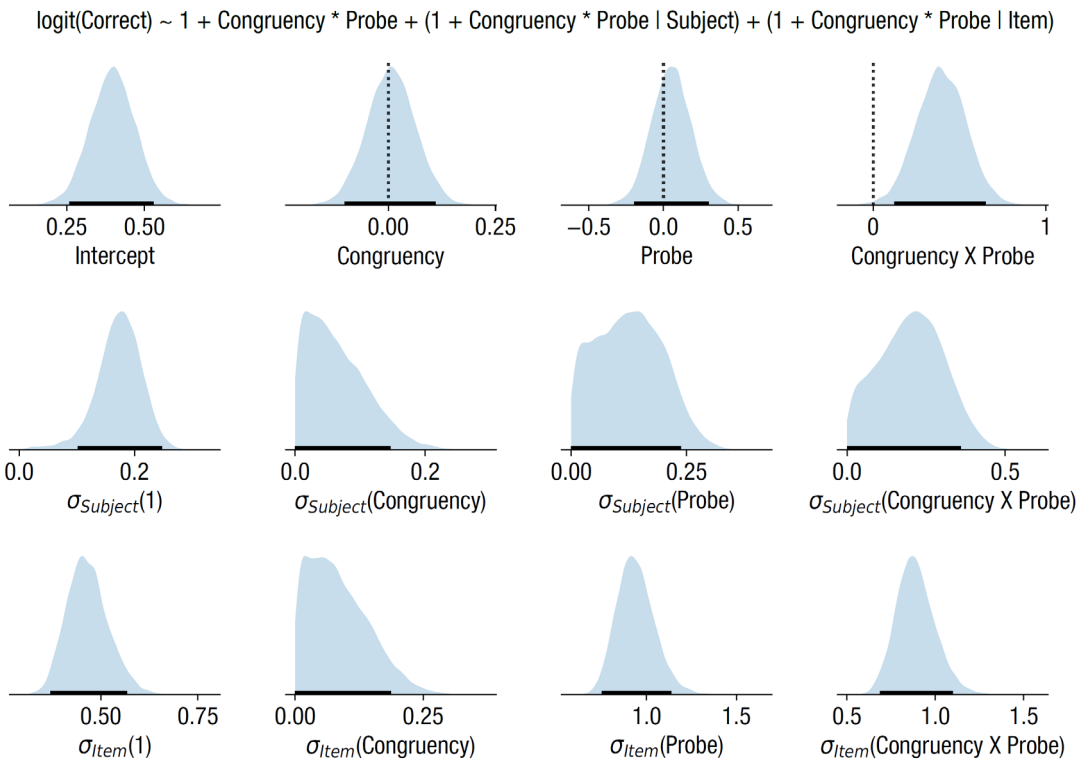

Figure S9. Posterior distribution after MCMC sampling of Bayesian logistic regression model for accuracy in Experiment 2. All panels and conventions as in Figure S4.

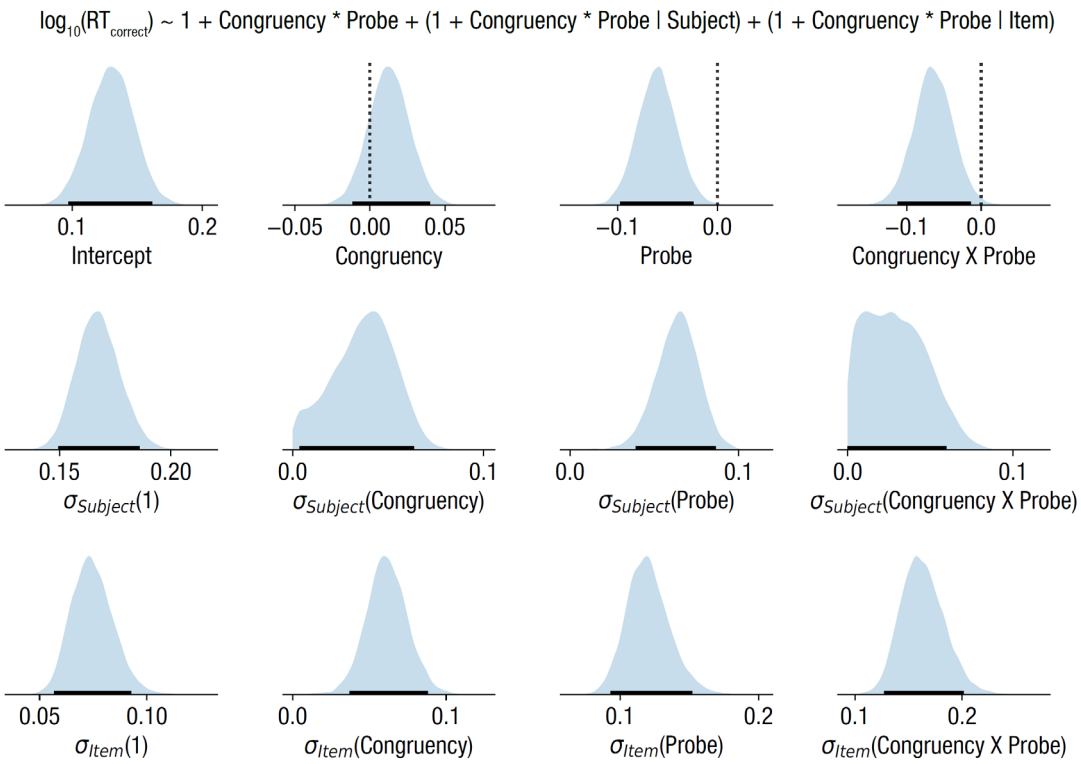

Figure S10. Posterior distribution after MCMC sampling of Bayesian regression model for reaction times in Experiment 2. All panels and conventions as in Figure S4.

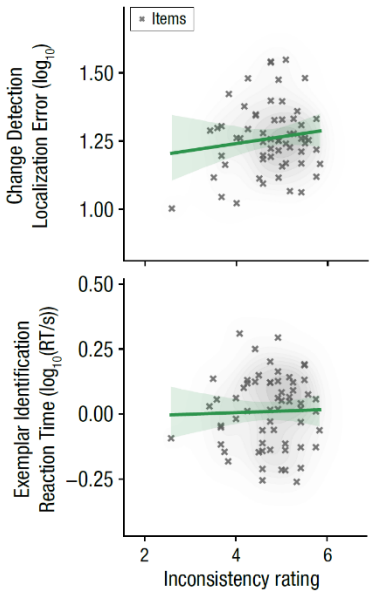

Figure S11. Correlations of secondary dependent variables with subjective inconsistency ratings across all incongruent items. Lines indicate best-fitting regression, shading indicates 95% confidence interval of regression line.

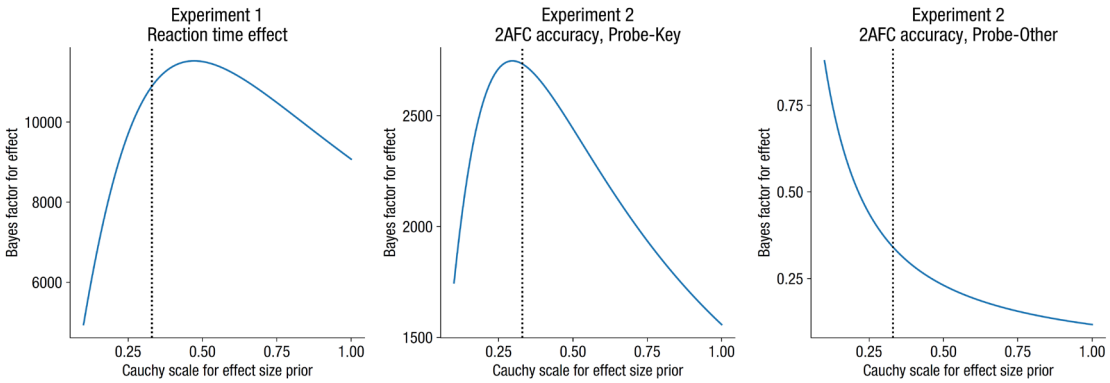

Figure S12. Robustness analysis for *t*-statistic-based Bayes factors (BF_{10}). BF_{10} is shown as a function of the scale parameter *r* used for the Cauchy prior over effect sizes. Different panels correspond to the three primary paired hypotheses tested in the manuscript (see panel headings). The dashed vertical line corresponds to the scale parameter $r = 0.33$ used for all reported analyses.

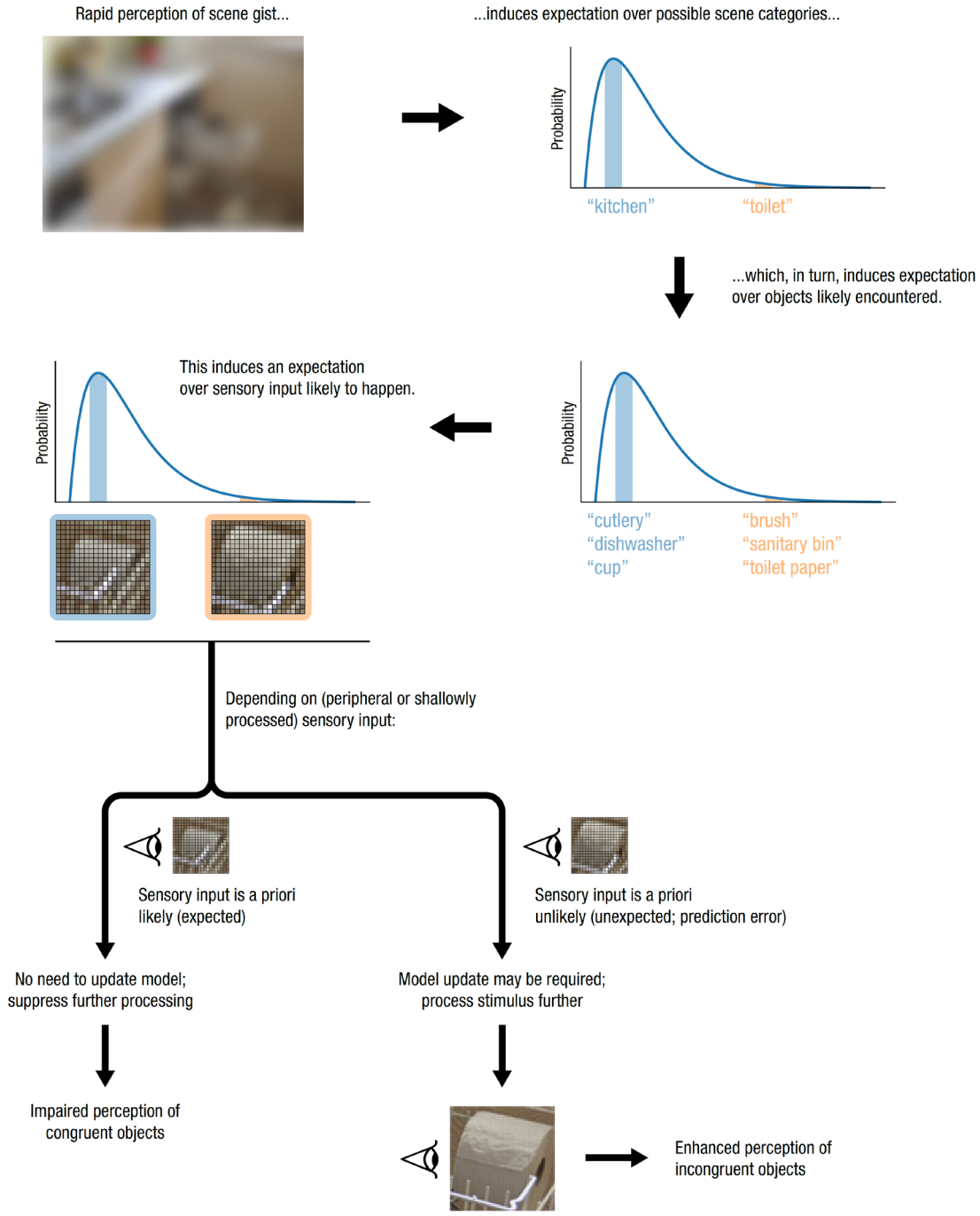

Figure S13. Outline of how semantic scene context may influence the perception of congruent/incongruent objects within that scene, through the induction of hierarchical prior expectations over sensory input that should be associated with scenes of a particular category.