**Supplementary Information**

****

**Figure S1**: Reaction times and confidence overlap. **Upper row**. Histogram and density functions of reaction times (left) and confidence ratings (right) following correct (in green) and incorrect type 1 responses (in red), for a representative participant in the auditory (A), tactile (T), and visual (V) conditions. **Lower row.** Correlation between the reaction times and confidence overlaps following correct and incorrect type 1 responses across participants. The metacognitive efficiency (meta-d’/d’) for each modality is mentioned on the top-right corner of each plot. Note that the correlation strength varies accordingly to metacognitive efficiency. The auditory condition is represented in red, the tactile condition in green, and the visual condition in blue.

*Individual model fits*



**Figure S3:** Individual model fits for the unimodal visual condition. Each plot shows the response rates for a participant in experiment 2. Two kinds of trials were presented: congruent (represented with black lines) and incongruent (represented with grey lines). After a median split of each participant’s confidence ratings, four possible responses resulted from the combination of the congruency judgment (“congruent”/”incongruent”) and the confidence judgment (“high confidence”/”low confidence”, represented here as “sure”/”unsure”). The red lines represent the response rates predicted by the unimodal model. These model predictions are presented here for illustration purposes only. They are the result of searching the parameter space (σ, c) for the values that best predicted the 8 response rates. While they suggest that the unimodal model we used was plausible, they were not subjected to any model comparison.



**Figure S4:** Individual model fits for the unimodal auditory condition. See legend in figure S2 for details.



**Figure S5:** Observed and predicted response rates for each participant, for the three best models (namely, the Integrative Average model, the Comparative Minimum model and the Single-modality Maximum metacognition model). Each model predicted response rates for 8 categories that result from combining the three factors: stimulus congruency (congruent/incongruent), response (correct/incorrect) and binned confidence (high/low). High and low confidence responses were defined as above or below each participant’s median confidence, across all conditions.



**Figure S6:** **Left panel**: confidence model predictions for all models tested. Ellipses represent the partially overlapping bivariate internal signal distributions for each of the stimuli combinations, represented at a fixed density contour. The models allowed the auditory and visual modalities to have different σ parameters, leading to ellipses instead of circles. Within each schematic of model predictions, the top right quadrant corresponds to congruent stimuli, where the stimuli of both pairs were stronger on the right side. The colors represent the predicted relative confidence for every combination of internal signal strength for each stimulus pair (XA, XV). Note that in all models except the two single-modality ones, confidence increases with increasing evidence in all four quadrants, because the congruency judgment is based on the conjunction of stimuli, and cannot be performed on basis of one stimulus only (see Aitchison et al., 2015 for similar work) **Right panel**: BIC values for the model fits in the audiovisual condition.