Supplementary Information for

Automatic multisensory integration follows subjective confidence rather than objective performance

Yi Gao[1]*, Kai Xue[1], Brian Odegaard[2], Dobromir Rahnev[1]

[1] School of Psychology, Georgia Institute of Technology, Atlanta, GA, 30332, USA

[2] Department of Psychology, University of Florida, Gainesville, FL, 32611, USA

Corresponding author: Yi Gao (yi.gao0525@outlook.com)

Supplementary Results

We employed a linear mixed-effects model to assess multisensory effects across different participant groups. The analysis revealed no statistically significant interaction between the group and condition (F(2,192) = 0.60, p = 0.55, partial η^2 = .0062). The main effect of the group was significant (F(2,192) = 8.41,p = 0.0003, partial η^2 = .08), indicating significant differences in multisensory effect among the three groups. However, there was no statistically significant main effect of condition (F(1,192) = 0.69, p = 0.408, partial η^2 = .004). These results suggest that while the groups differ in their multisensory effect, these differences are consistent across both high and low-energy conditions. Post-hoc comparisons indicate that there are significant differences in multisensory effect between Group 1 and Group 2 (F(1,192) = 8.92, p = 0.0032, partial η^2 = .04) and between Group 2 and Group 3 (F(1,192)=16.23, p = 0.08, partial η^2 = .08), but not between Group 1 and Group 3 (F(1,192) = 0.36, p = 0.5495, partial η^2 = .002). Given that the group differences were consistent across conditions, we opted to combine the groups in the main analysis.

Supplementary Figures

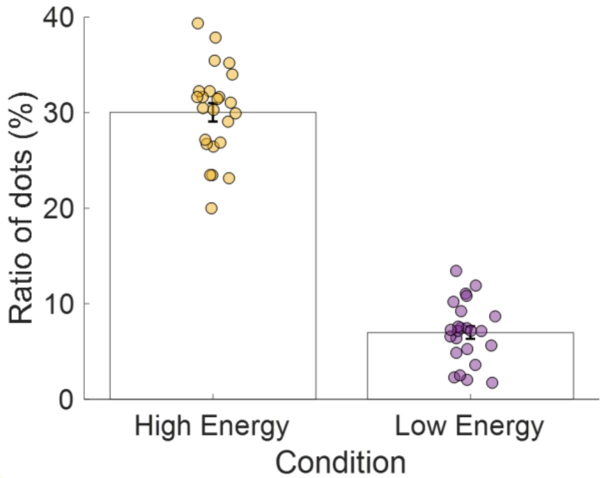

Figure S1. Distributions of percentages of dots moving in the non-dominant direction for high- vs. low-energy conditions across participants in Group 1 (N=24). In the high-energy condition, the mean percentage of dots moving in the non-dominant direction was 30.04% (SD = 4.75), while in the low-energy condition, the mean percentage was 6.98% (SD = 3.18). For Group 2 (N = 25), the percentage of dots moving in the non-dominant direction was fixed at 45% for high-energy stimuli and 0% for low-energy stimuli. For Group 3 (N = 50), the percentage of dots in the non-dominant direction was slightly decreased to 31% for high-energy stimuli and again set to 0% for low-energy stimuli. Colored symbols indicate individual data. Error bars represent the standard error of the mean (SEM).

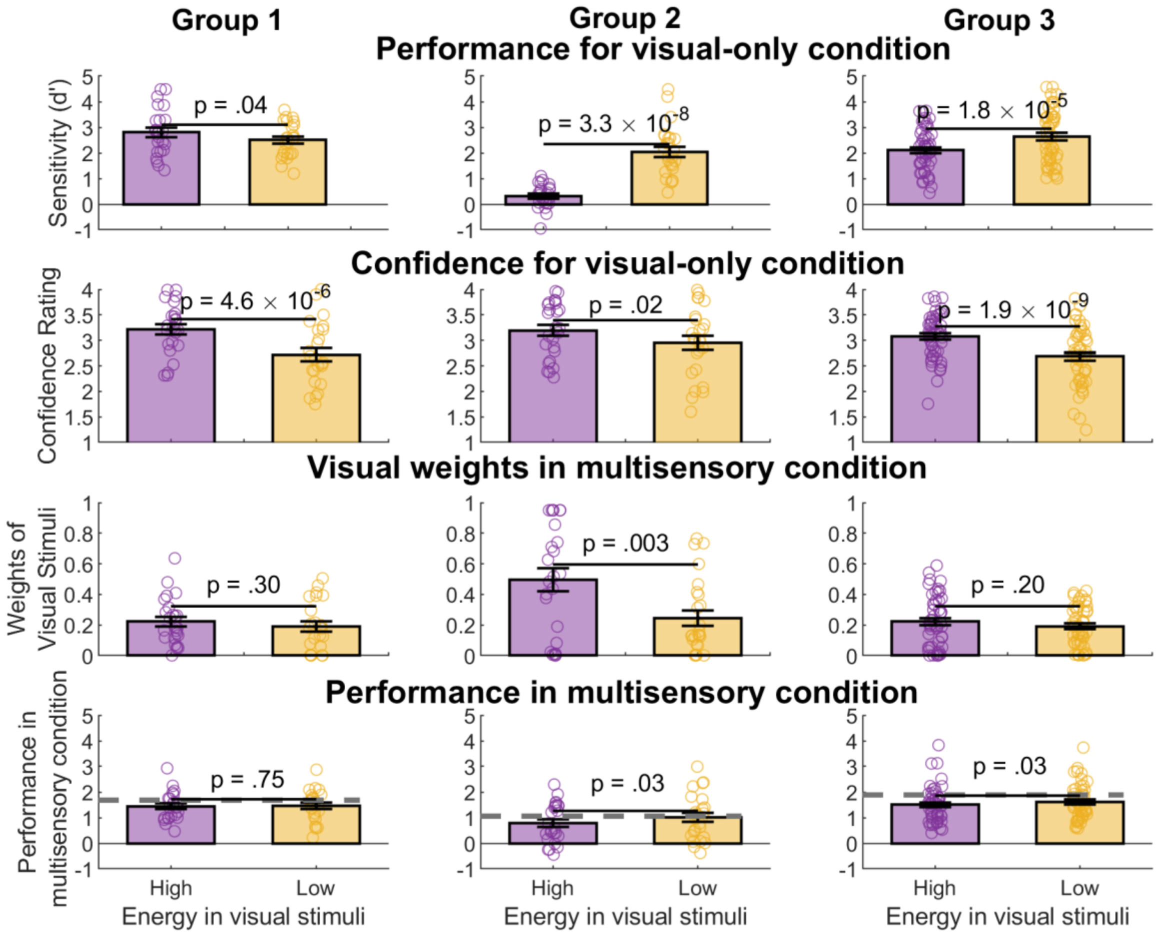

Figure S2. Results for the three participant groups. In the first group of participants (left column, N = 24), we attempted to match the performance for high- vs. low-energy stimuli by running a 3-down-1-up staircases on the proportion of dots moving in the non-dominant direction separately for the high- vs. low-energy stimuli. The staircase worked reasonably well, but performance was nonetheless slightly but significantly higher for the high-energy visual stimuli ($t(23) = 2.18$, $p = .04$, 95% CI [.01, .56], Cohen's $d = .44$; top row). Nevertheless, confidence was much higher for the high-energy visual stimuli ($t(23) = 5.95$, $p = 4.6*10^{-6}$, 95% CI [.32, 67], Cohen's $d = 1.21$; second row), similar to what we observed in the pooled data in the main paper. The directions of the visual weights and performance in multisensory condition are the same as in the pooled data but not significant in this group alone. Specifically, the visual stimuli were weighted numerically higher for the high- compared to the low-energy visual stimuli, but this effect was not statistically significant ($t(23) = 1.05$, $p = .30$, 95% CI [-.03, .10], Cohen's $d = .22$; third row). Similarly, the performance in the multisensory condition was lower for the high- compared to the low-energy stimuli, but this effect was also not statistically significant ($t(23) = .32$, $p = .75$, 95% CI [-.19, .14], Cohen's $d = .07$; last row). In the second group of participants (middle column, N = 25), we fixed the percent of dots moving in the non-dominant direction to 45% for the high-energy stimuli and 0% for the low-energy stimuli. All results in this group are equivalent to the results in the main text where we pooled data from all three groups. High-energy visual stimuli led to lower performance ($t(24) = 7.98$, $p = 3.3*10^{-8}$, 95% CI [1.29, 2.18], Cohen's $d = 1.60$) but higher confidence ($t(24) = 2.61$, $p = .02$,95% CI [.05, 43], Cohen's $d = .52$). The visual stimuli were weighted more heavily in multisensory judgments for the high- compared to the low-energy visual stimuli ($t(24) = 3.32$, $p = .003$, 95% CI [.09, 41], Cohen's $d = .66$). The performance in the multisensory condition was lower for the high- compared to the low-energy stimuli ($t(24) = 2.25$, $p = .03$, 95% CI [.02, 42], Cohen's $d = .45$). In the third group of participants (right column, N = 50), we fixed the percent of dots moving in the non-dominant direction to 31% for the high-energy stimuli and 0% for the low-energy stimuli. Results for this group are equivalent to the results in the main text where we pooled the data from all three groups, except that the difference in visual weights used in the multisensory condition for high- vs. low-energy stimuli did not reach significance. High-energy visual stimuli led to lower performance ($t(49) = 4.76$, $p = 1.8*10^{-5}$, 95% CI [.31, 77], Cohen's $d = .67$) but higher confidence ($t(49) = 7.35$, $p = 1.9*10^{-9}$, 95% CI [.28, 49], Cohen's $d = 1.04$). The visual stimuli were weighted numerically higher for the high- compared to the low-energy visual stimuli, but this effect was not statistically significant ($t(49) = 1.31$, $p = .20$, 95% CI [-.02, 08], Cohen's $d = .19$). The performance in the multisensory condition was statistically significantly lower for the high- compared to the low-energy stimuli ($t(49) = 2.21$, $p = .03$, 95% CI [.01, .21], Cohen's $d = .31$). Shaded symbols indicate individual data. Error bars indicate SEM, dashed line indicates average performance in the auditory-only condition.

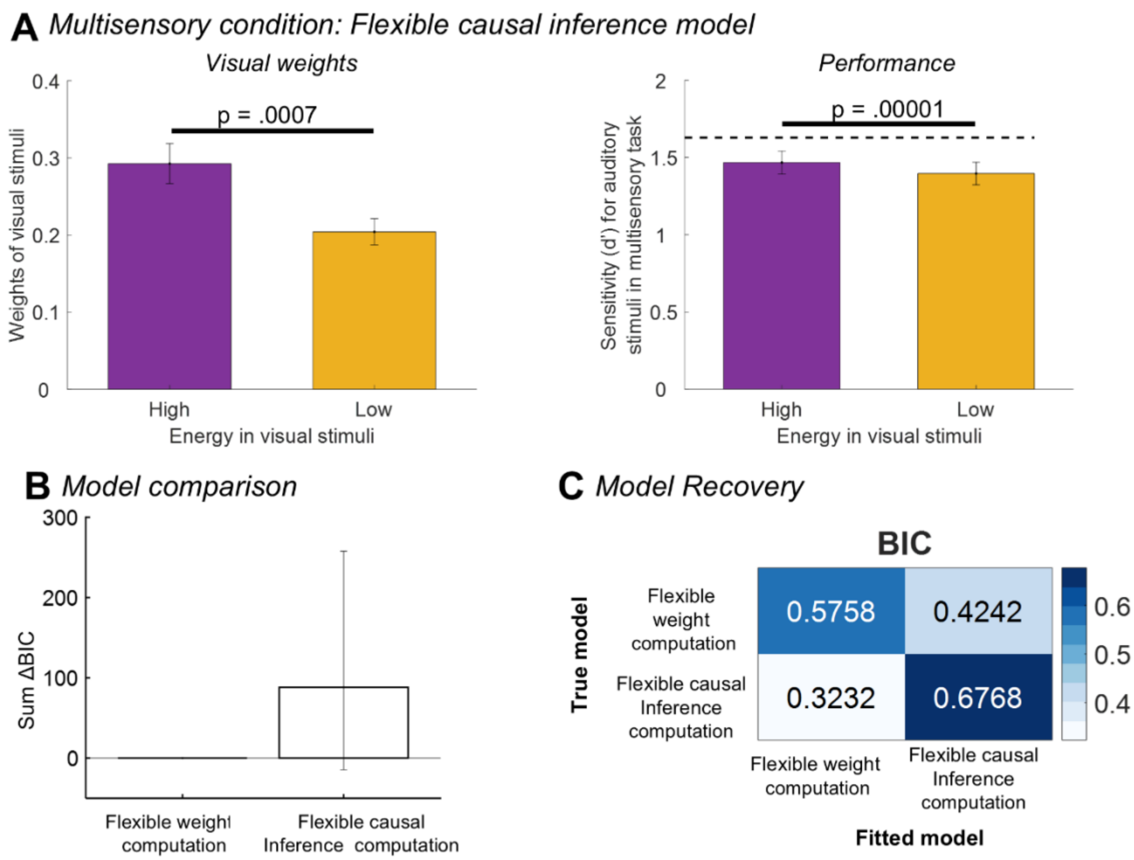

Figure S3. Results for the Flexible causal inference computation. The Flexible causal inference computation postulates that multisensory judgments are made by considering the absolute difference between visual and auditory evidence. If this difference exceeds a certain threshold T (a free parameter representing the point at which the cues are considered too dissimilar), the model relies exclusively on the auditory signal. However, when the absolute difference is less than or equal to T, the model combines both visual and auditory signals using a reliability-weighted approach. Specifically, the weight assigned to the visual cue is calculated as $\frac{d'_{visual}}{d'_{visual}+d'_{auditory}}$, where d′ values represent the sensitivity of each sensory modality. (A) The Flexible causal inference computation reproduced the higher estimated weight for the high- compared to the low-energy stimuli ($t(98) = 3.51$, $p = .0007$, 95% CI [-0.03, -.01], Cohen's d = .35, $BF_{10} = 31.26$). However, in contrast to the data, the model produced larger multisensory d' for the high- compared to the low-energy visual stimuli ($t(98) = 4.61$, $p = .00001$, 95% CI [-0.03, -.01], Cohen's d = .46, $BF_{10} = 1339.6$). (B) Comparison of model performance between the Flexible weight computation and the Flexible causal inference computation based on BIC values. The two computations provided relatively similar model fits. The Flexible weight computation won by a total of 88.2 points, which is an average of .89 points per participant). Error bars indicate 95% confidence intervals (CIs) generated using bootstrapping. (C) Model recovery analysis comparing the Flexible weight and Flexible causal inference computations. The model recovery analysis showed that the two models are not easily distinguishable from each other, with an average model recovery rate of 62.63%. Note that because the Flexible weight and Flexible causal inference computations have the same number of free parameters (one for each), using AIC instead of BIC in panels B and C produces the same results.

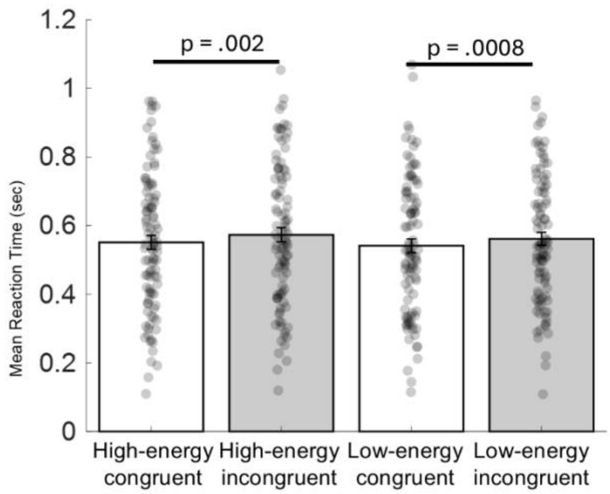

Figure S4. Reaction times for congruent vs. incongruent trials. The reaction times were shorter when visual motion direction and auditory motion direction were congruent compared to incongruent. This effect was significant for both high- and low-energy conditions (High-energy: t(98) = 3.26, p = . 002, 95% CI [.009, .036], Cohen's d = .33; Low-energy: t(98) = 3.46, p = .0008, 95% CI [.009, .032], Cohen's d = .35, N = 99). Error bars indicate SEM.

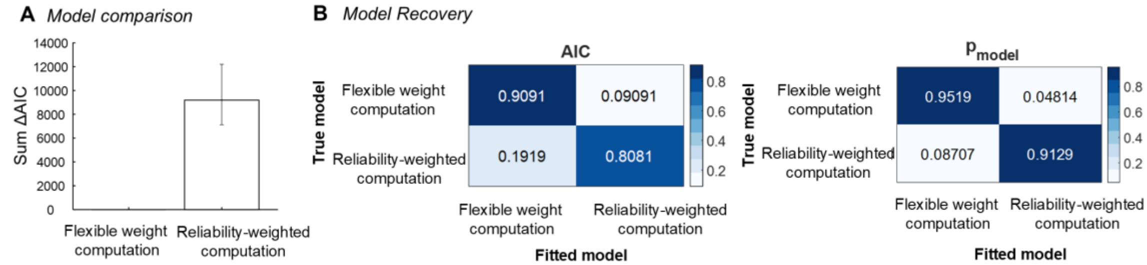

Figure S5. Model comparison and model recovery based on AIC between the Flexible weight computation and the Reliability-weighted computation. (A) Comparison of model performance between the Flexible weight computation and the Reliability-weighted computation based on AIC values. The Flexible weight computation outperformed the Reliability-weighted computation by a total of 9188 AIC points. Error bars indicate 95% confidence intervals (CIs) generated using bootstrapping. (B) Model recovery analysis comparing the Flexible weight and Reliability-weighted computations. The model recovery analysis shows a slight bias towards recovering the Flexible weight computation more frequently. Consequently, to avoid this bias, we used BIC in the main text (BIC showed a slight bias towards the Reliability-weighted computation).