CoT Locality: Add/Lesion Effects (Layers 25-27)

Baseline Accuracy: 65.0%
Δ = Intervention Accuracy − Baseline Accuracy
*p-values: McNemar's Test | * p<0.05, ** p<0.01, *** p<0.001 | Bold = Significant*