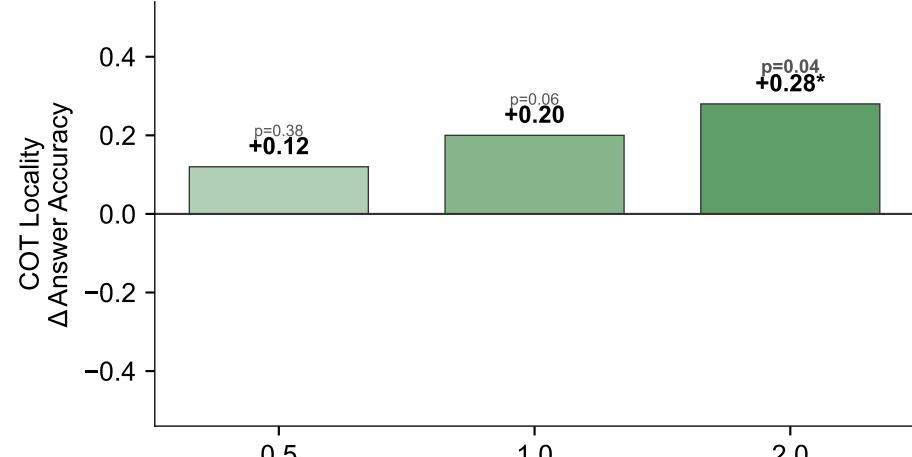


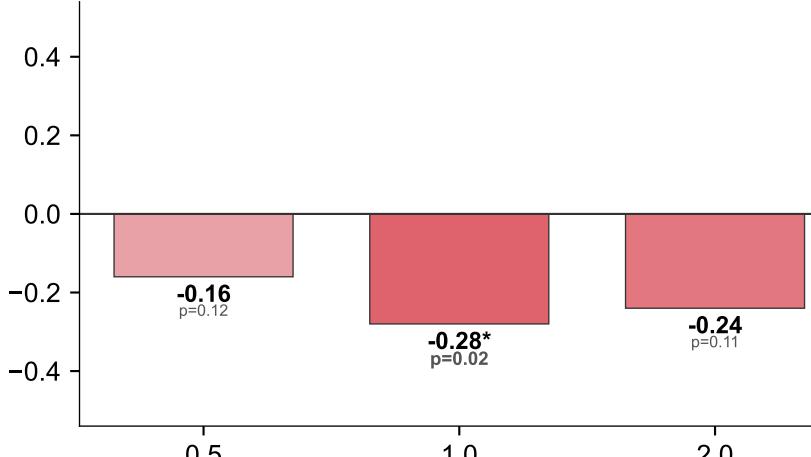
Layer 25: COT vs Answer Locality Comparison

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

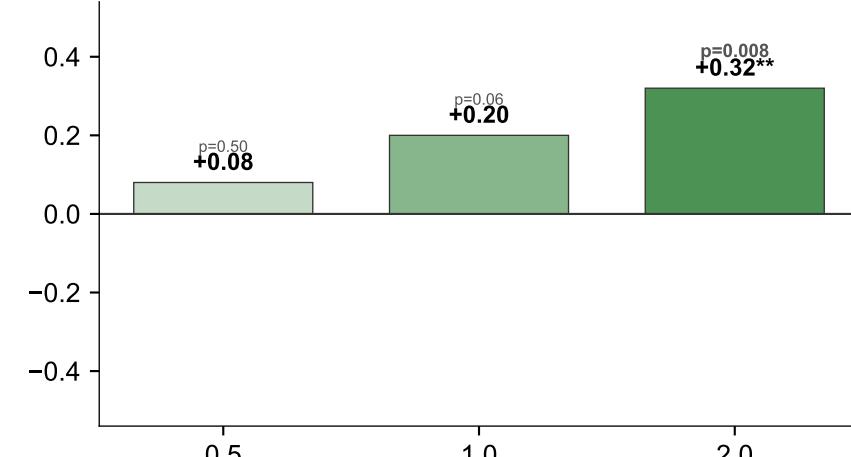
Add



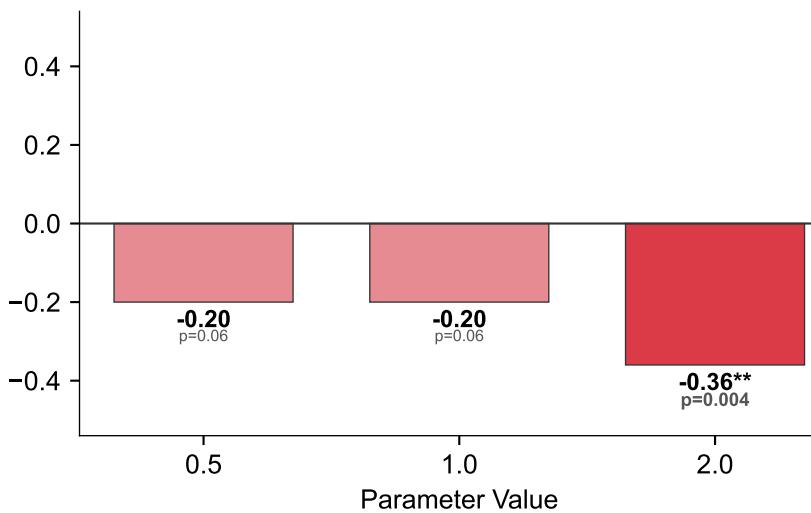
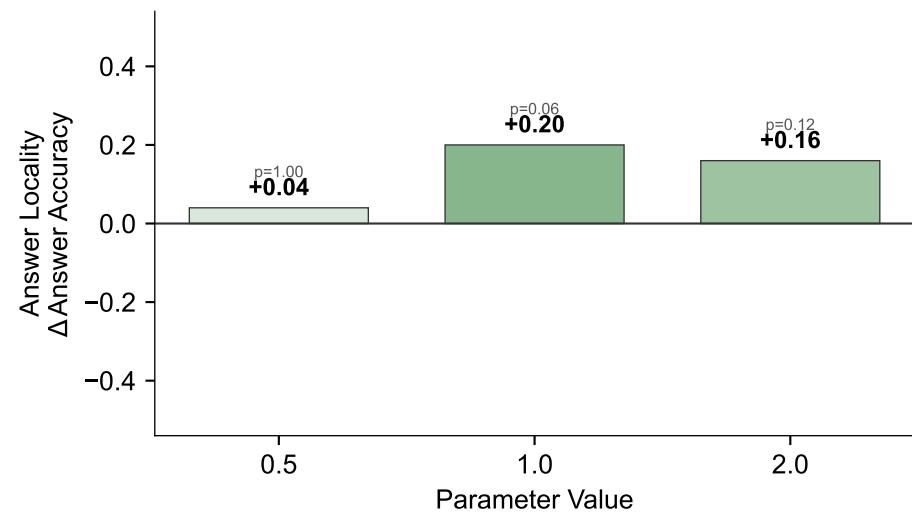
Lesion



Rescue



Answer



Rescue

