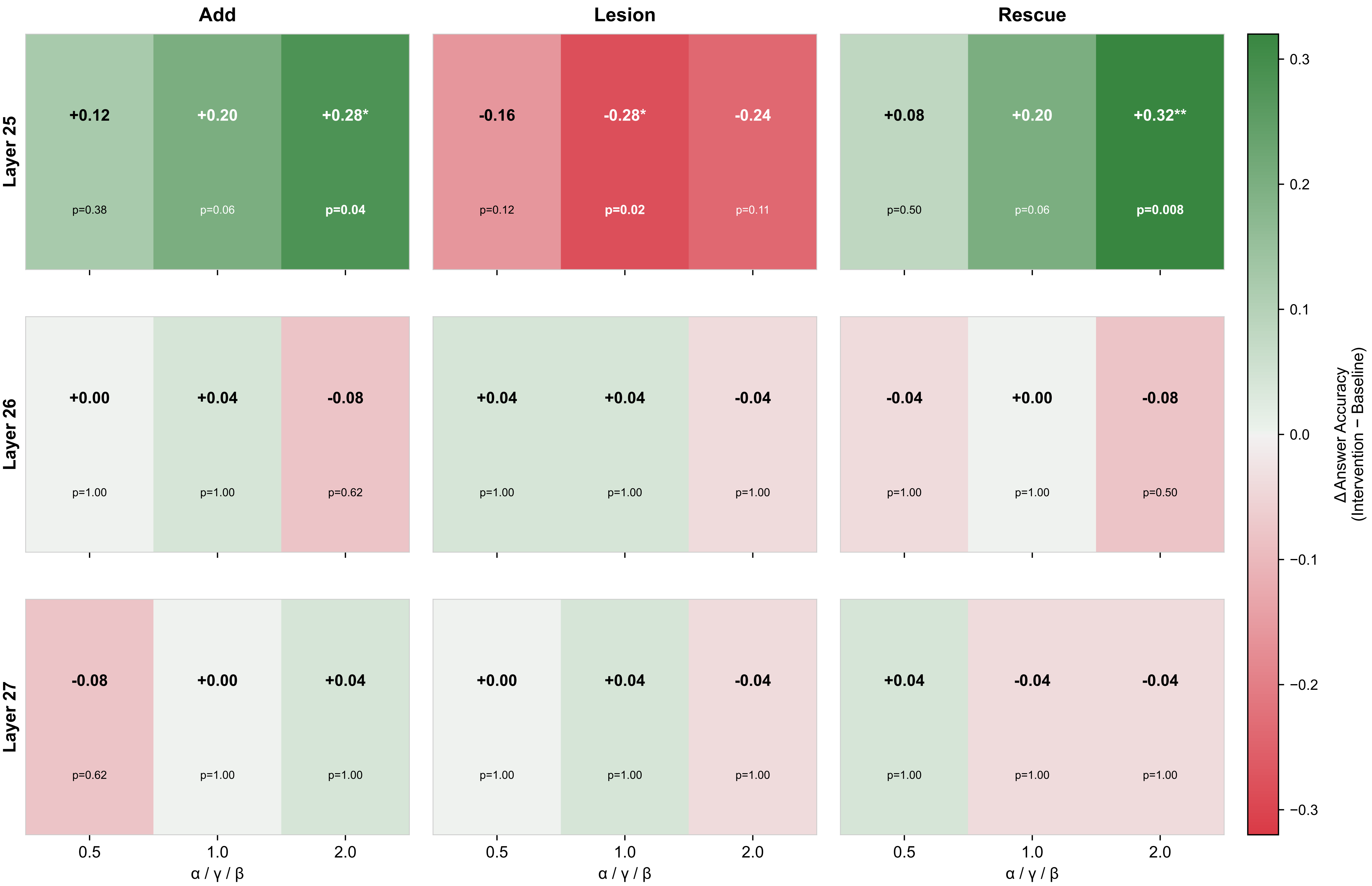
CoT Locality: Intervention Effects by Mode and Layer

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
Layer 25
p=0.38
p=0.06
+0.28*
p=0.04
p=0.12
-0.28*
p=0.02
p=0.11
p=0.50
p=0.06
+0.32**
p=0.008
Layer 26
p=1.00
p=1.00
p=0.62
p=1.00
p=1.00
p=1.00
p=1.00
p=1.00
p=0.50
Layer 27
p=0.62
p=1.00
p=1.00
p=1.00
p=1.00
p=1.00
p=1.00
p=1.00
p=1.00
α / γ / β
α / γ / β
α / γ / β
Δ Answer Accuracy
(Intervention − Baseline)