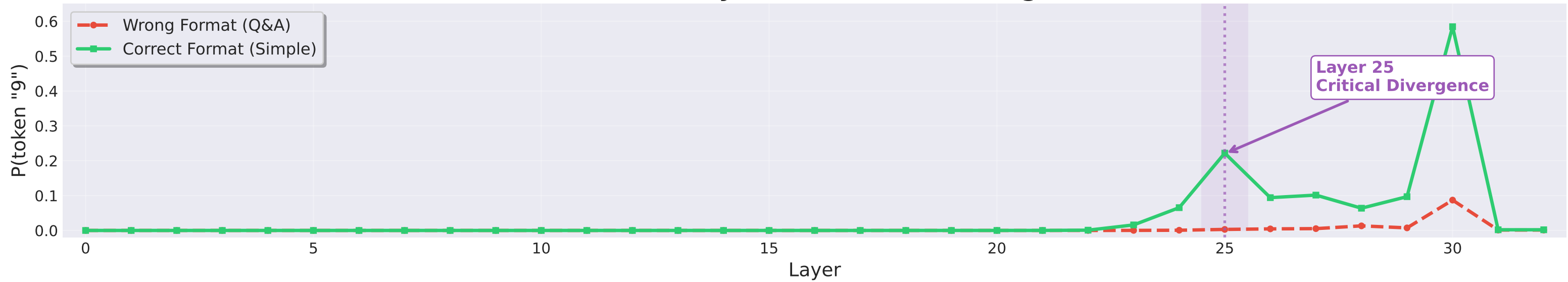


9.11 vs 9.8 Bug: Complete Token Evolution Analysis

Token "9" Probability Evolution: The Divergence Point



Wrong Format (Q&A → 9.11)

Label	Value
L0	greg 0.025
L5	strup 0.026
L10	tual 0.022
L15	PCP 0.014
L20	Both 0.089
L25	Both 0.365
L28	neithe 0.286
L30	They 0.127
L32	0.464

Correct Format (Answer → 9.8)

Label	Value
L0	0.02
L5	0.02
L10	0.18
L15	0.01
L20	0.02
L25	0.22
L28	0.32
L30	0.58
L32	0.87

Probability Difference Heatmap (Correct - Wrong Format)

