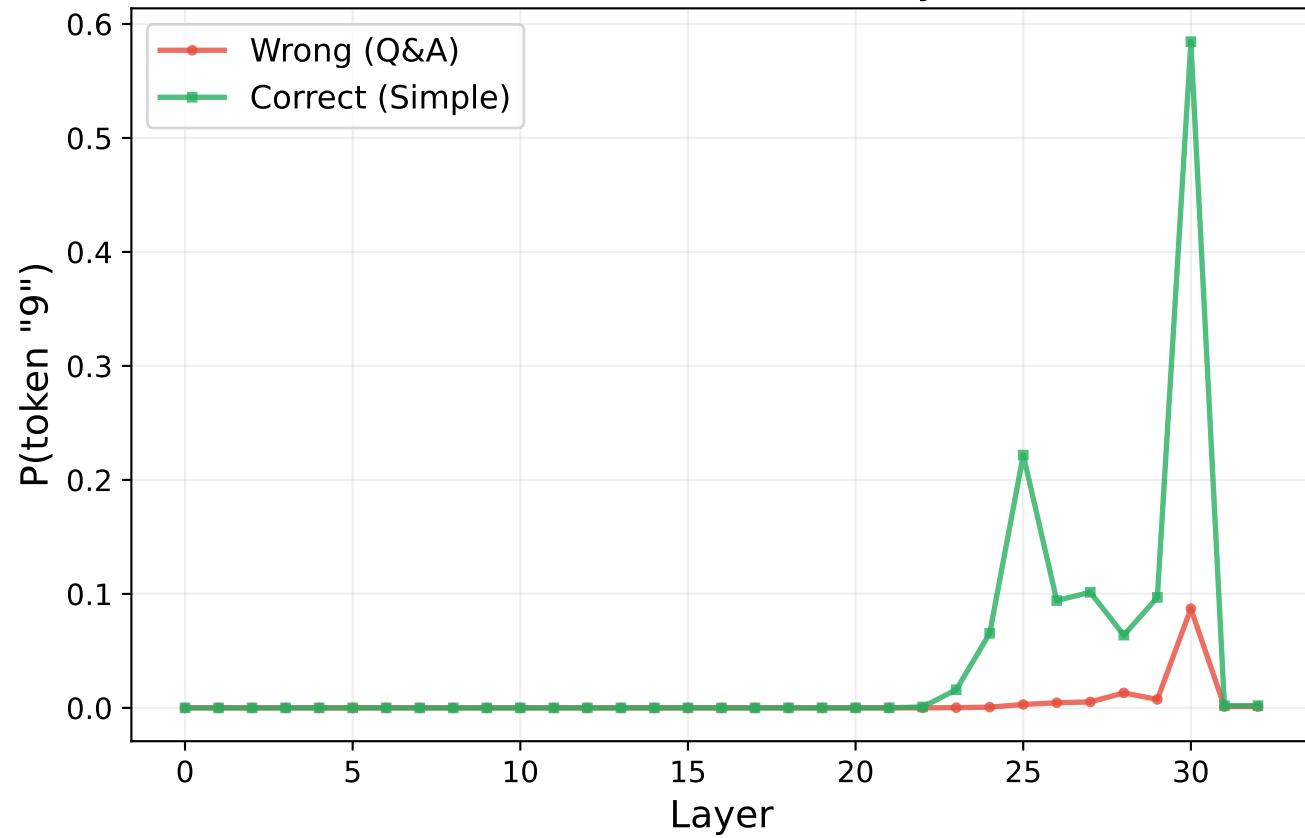
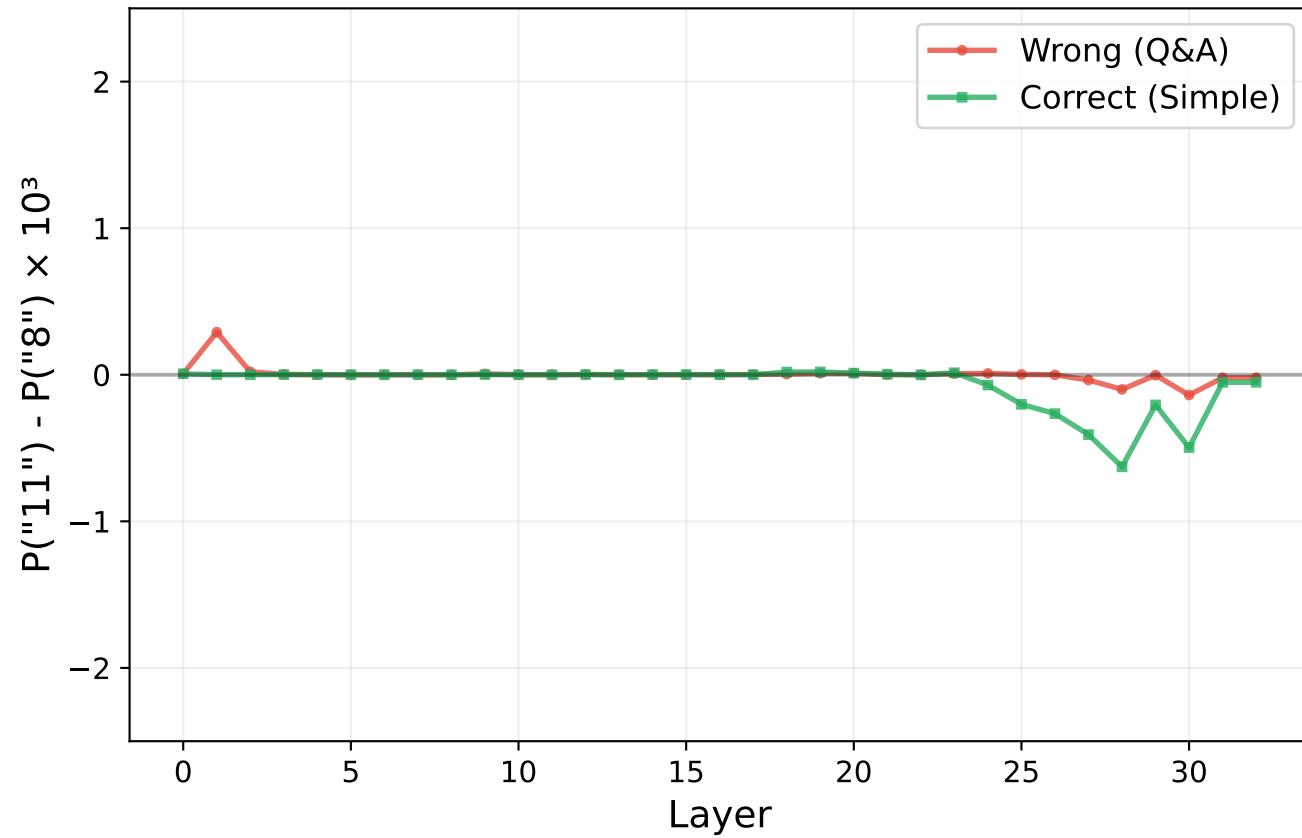


Logit Lens Analysis: Format-Dependent Processing

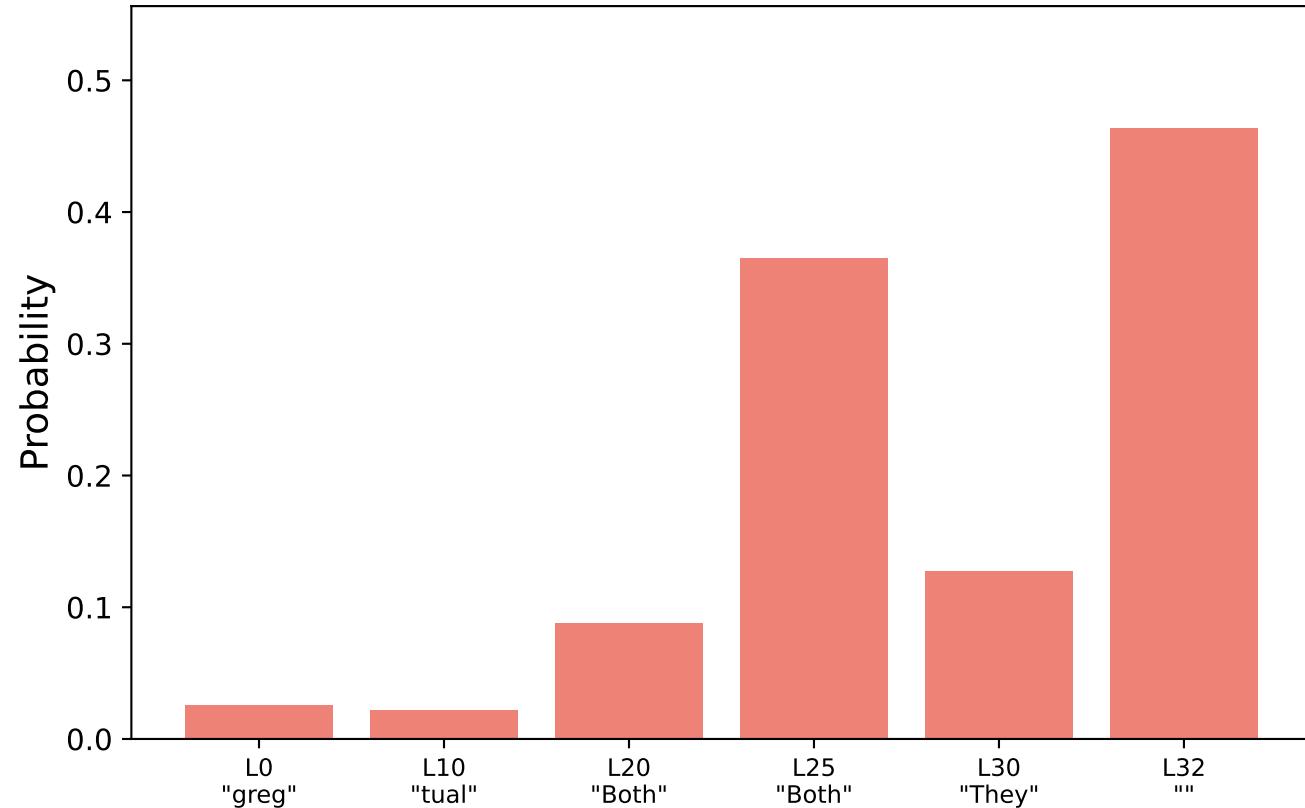
Token "9" Probability



Token Preference



Wrong Format - Top Tokens



Correct Format - Top Tokens

