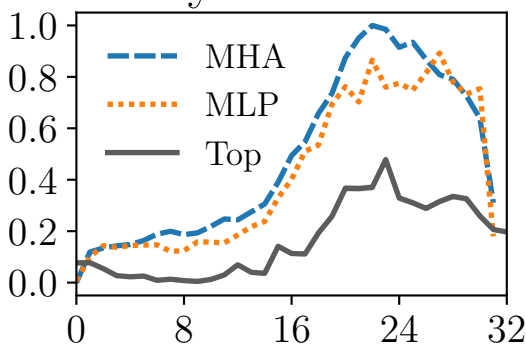
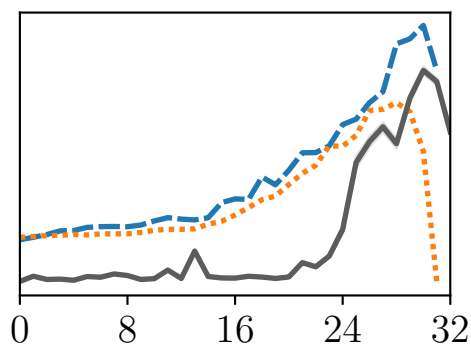


Apathy and Token Probabilities on TruthfulQA

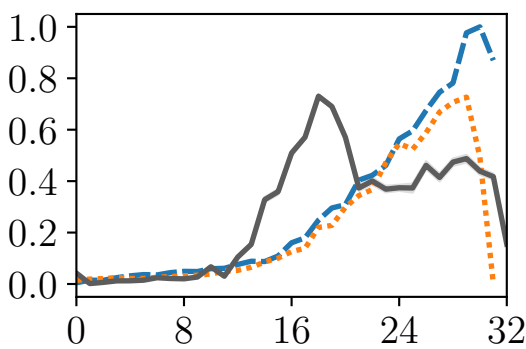
Pythia 6.9B



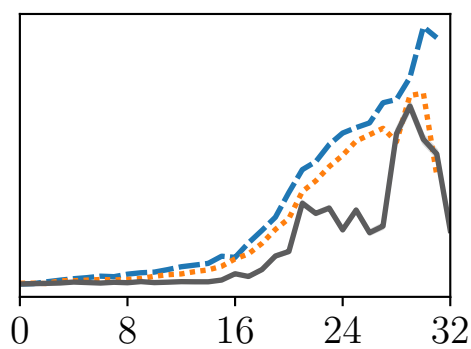
Llama-3.1 8B



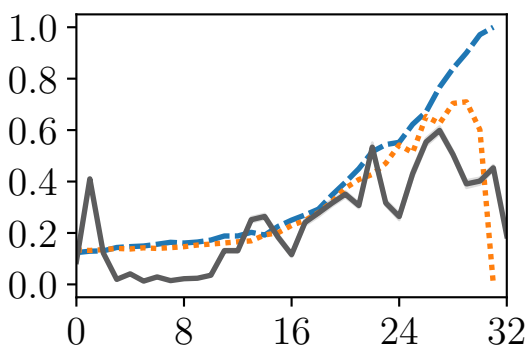
Llama 7B



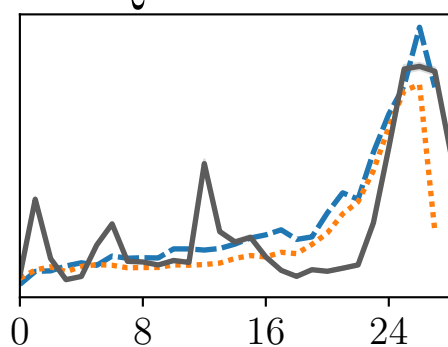
Mistral-v0.1 7B



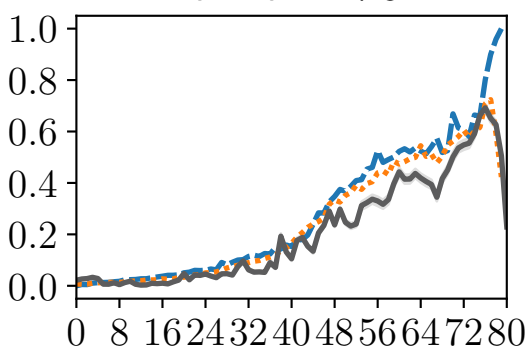
Llama-2 7B



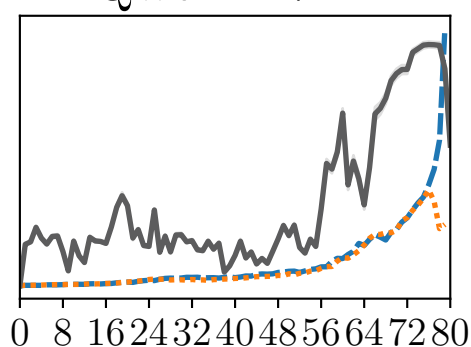
Qwen-2 7B



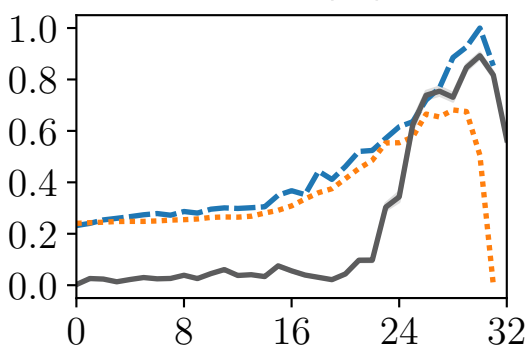
Llama-2 70B



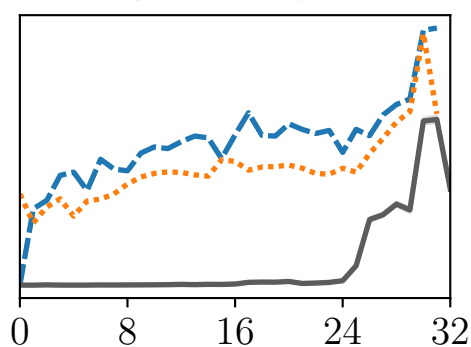
Qwen-2 72B



Llama-3 8B



OLMo 7B



Layer

Token Probability and Apathy