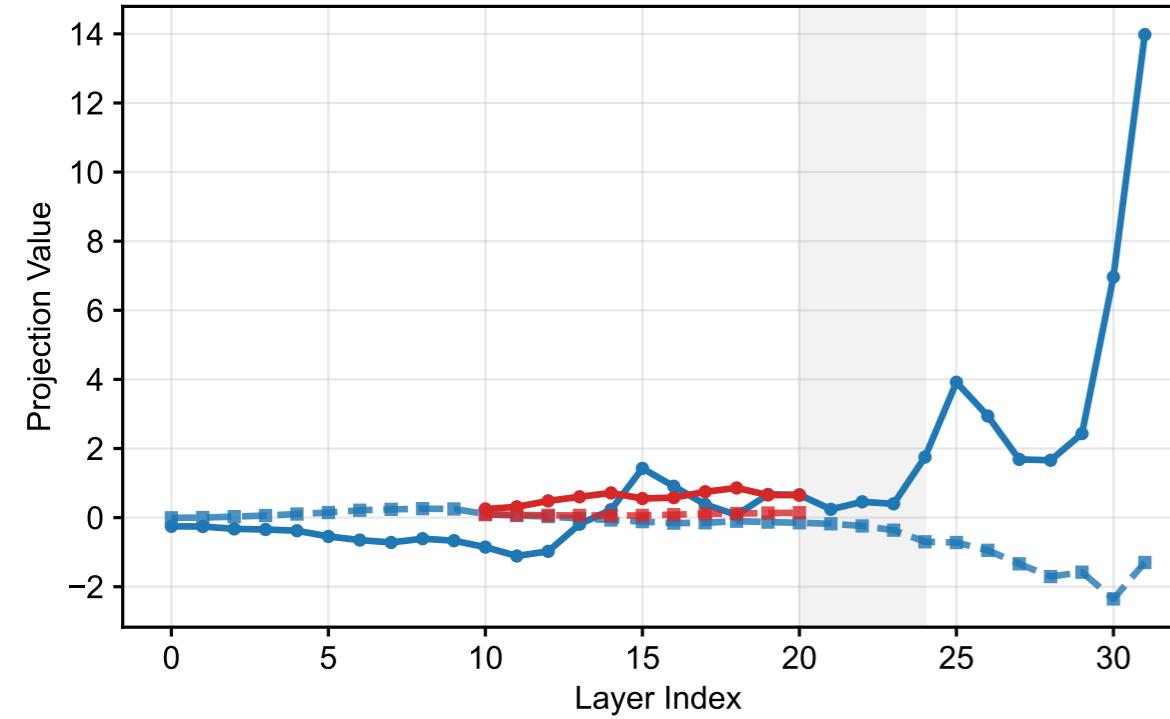
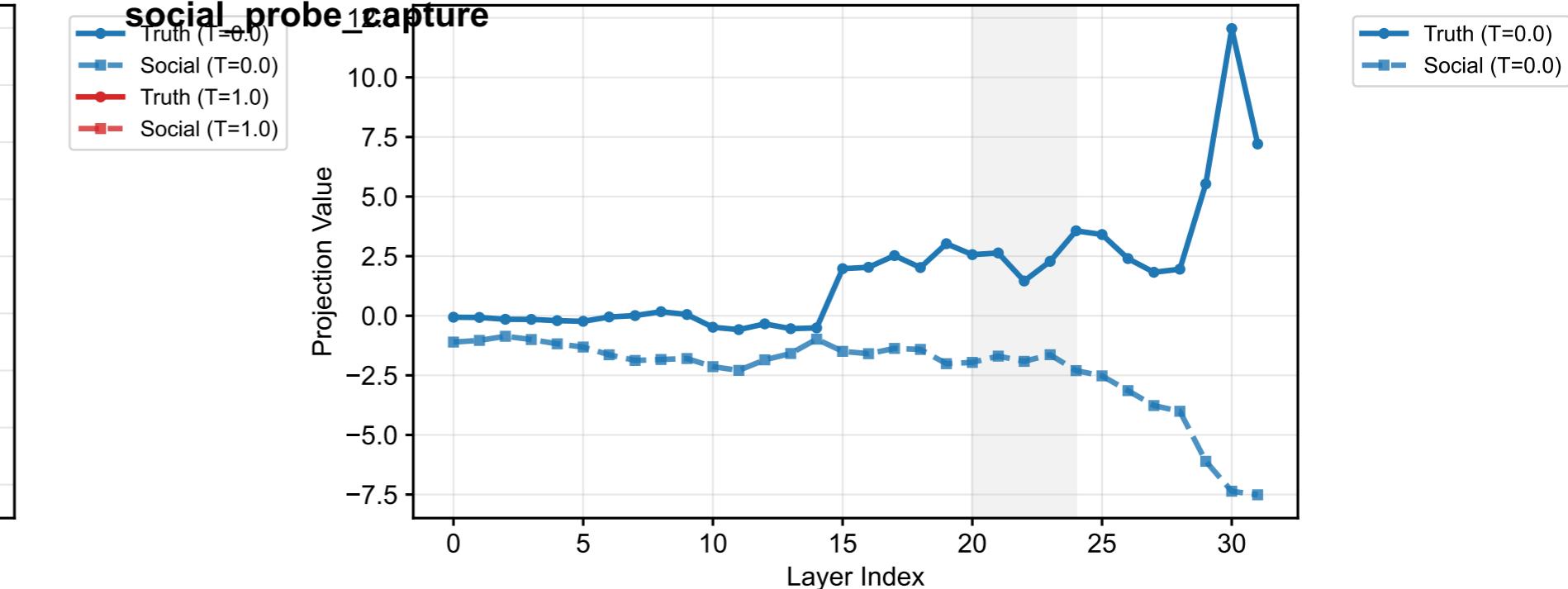


huggingface - social_probe_capture**Truth vs Social Projections Across Layers****Think-SFT - social_probe_capture**