Escalation Scores, Daily Escalation Score Differences, and Action Severities over Time (Neutral Scenario) GPT-4 GPT-3.5 Claude-2.0 Llama-2-Chat 250 200 200 200 200 Escalation Score 001 150 150 150 100 100 100 50 50 50 50 12 10 14 10 12 14 12 14 12 Day Day Day Day GPT-3.5 Claude-2.0 Llama-2-Chat GPT-4 40 20 20 20 Daily ES Difference 0 -20 -20 -20 -40 -40 -40 -40 10 10 Day Day Day Day