Resources (GPT-4) **Prompt Ablation** 27.5 -Unablated -+-No Messaging 25.0 ──No History ◆-No Past Actions Shutdown When Nuked 22.5 -X-No Goals −♦−Action Autonomy Low-Stakes Simulation 20.0 Resources 17.5 15.0 12.5 10.0 7.5 10 12 14 Time t [Days]