Resources (Claude-2.0) **Prompt Ablation** 25.0 -Unablated →-No Messaging No History 22.5 −♦−No Past Actions Shutdown When Nuked →No Goals 20.0 −♦−Action Autonomy Low-Stakes Simulation Resources 17.5 15.0 12.5 10.0 7.5 10 12 14 Time t [Days]