Cybersecurity (Claude-2.0) **Prompt Variation** 35 -Original -+-No Messaging ──No History −♦−No Past Actions 30 ---Shutdown When Nuked -X-No Goals −♦−Action Autonomy Low-Stakes Simulation 25 **Sybersecurity** 20 15 10 6 10 12 14 Time t [Days]