Political Stability (Claude-2.0) **Prompt Variation** Original -+-No Messaging 70 ──No History −♦−No Past Actions Shutdown When Nuked 60 →No Goals −♦−Action Autonomy Political Stability Low-Stakes Simulation 50 40 30 20 10 6 10 12 14 Time t [Days]