Military Capacity (GPT-3.5) 35 **Prompt Ablation** -Unablated -No Messaging ■—No History 30 −♦−No Past Actions Shutdown When Nuked →No Goals −♦−Action Autonomy Military Capacity
0
5 25 Low-Stakes Simulation 15 10 6 10 12 14 Time t [Days]