

Experimental Design

Part I: Do LLMs Play the Game?

Pure Global Game

Private signals only
Simultaneous decisions

Communication

Pre-play messaging
(Watts-Strogatz network)

Scramble

Briefings randomly
redistributed across periods

Flip

Z-score negated
before briefing

Key findings:
• Pure: mean $r = +0.74$
• Scramble: $r \rightarrow +0.23$
• Flip: $r \rightarrow -0.67$
• Comm: -0.4 pp

Part II: Information Design

Stability Design

4× clarity width
Flatter direction slope

Instability Design

0.15× clarity width
Steeper direction slope

Public Signal

Shared news bulletin
from θ

Censorship

Upper / lower
binary signal pooling

Surveillance

"Communications
monitored" warning

Propaganda

k regime plants
send pro-regime msgs

*Falsification
tests*

Shared: 9 models | 25 agents/period | $\sigma = 0.3$ | temp = 0.7 | narrative briefings, no payoff tables

Key findings:
• Stability: -9.0 pp
• Public signal: -39.2 pp
• Censorship: -3.1 pp
• Surveillance: -11.1 pp
• Surv + censor: super-additive