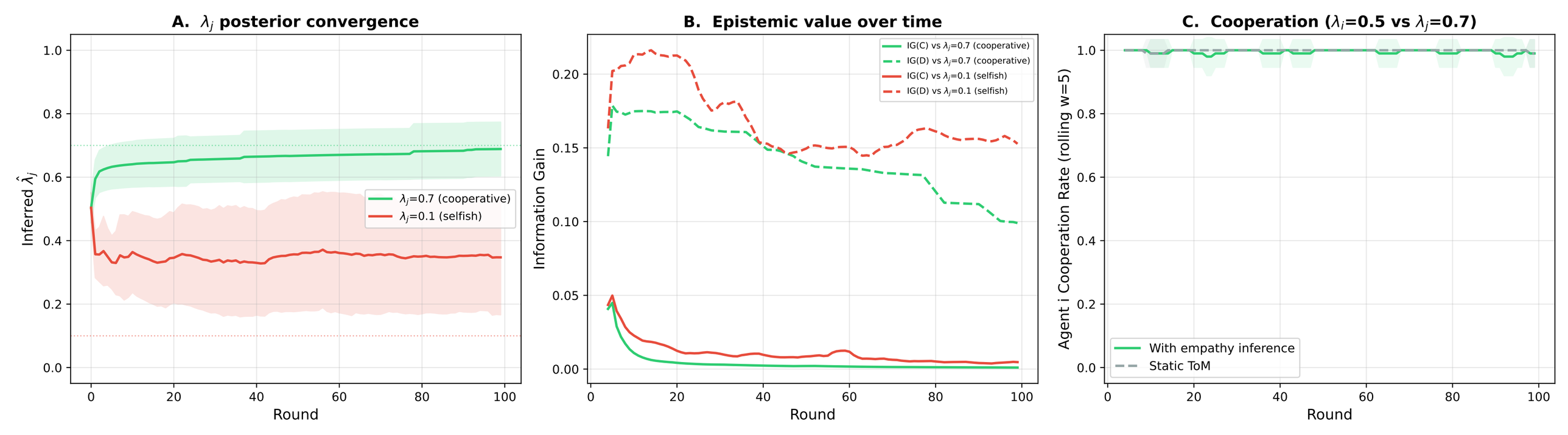
A. λ_j posterior convergence
Inferred $\hat{\lambda}_j$
Round
λ_j=0.7 (cooperative)
λ_j=0.1 (selfish)
B. Epistemic value over time
Information Gain
Round
IG(C) vs λ_j=0.7 (cooperative)
IG(D) vs λ_j=0.7 (cooperative)
IG(C) vs λ_j=0.1 (selfish)
IG(D) vs λ_j=0.1 (selfish)
C. Cooperation (λ_i=0.5 vs λ_j=0.7)
Agent i Cooperation Rate (rolling w=5)
Round
With empathy inference
Static ToM