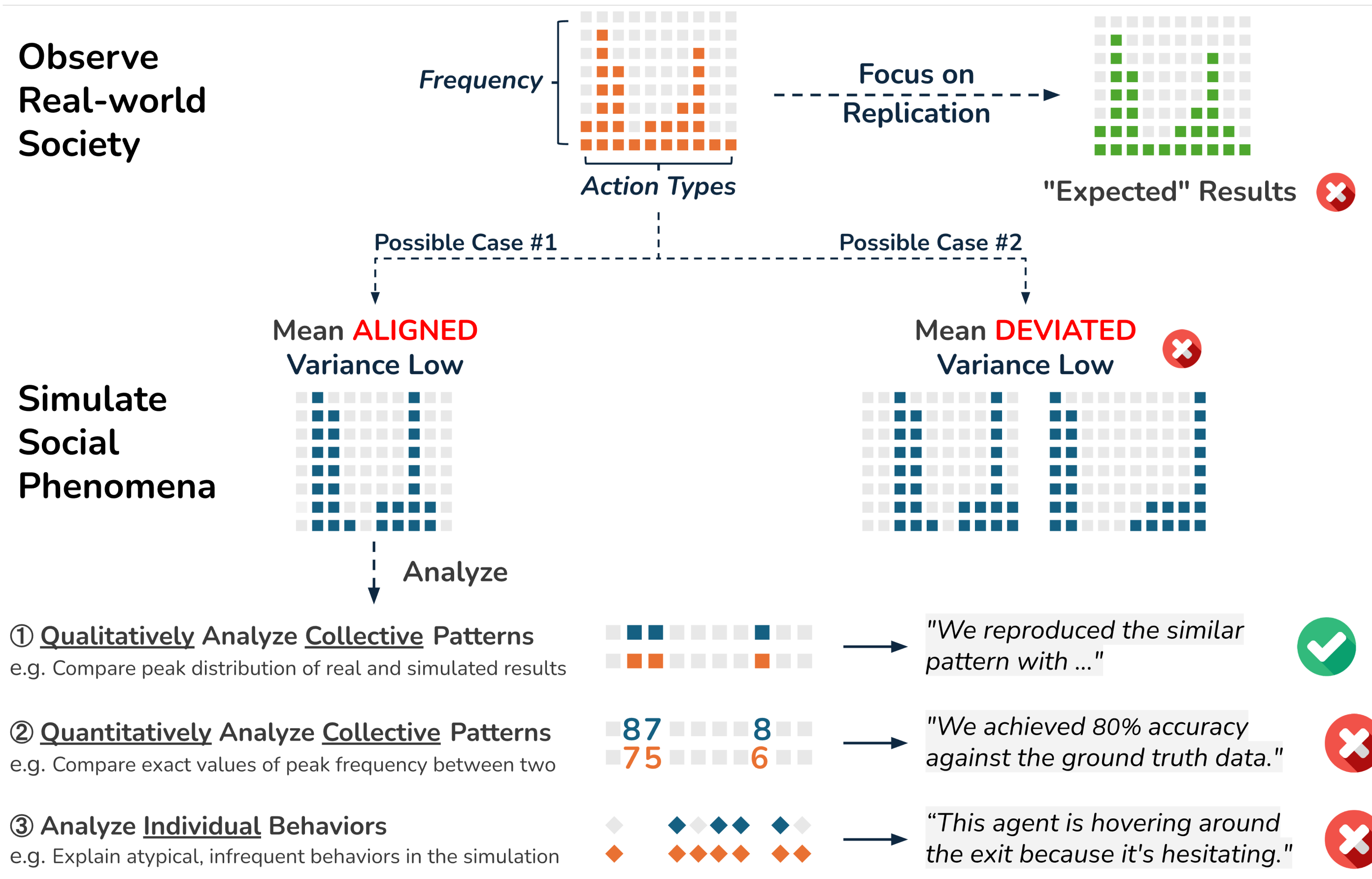


What we CAN and CANNOT simulate?



Motivation

Why Define Boundaries?

Issues in current practice

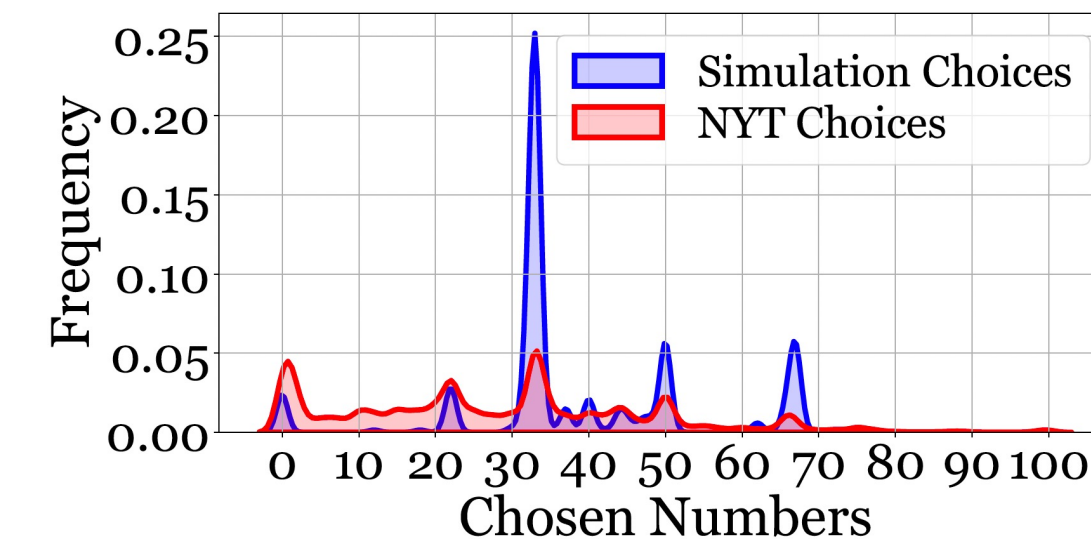
- ❌ Over-emphasis on perfect **REPLICATION** of reality
- ❌ Overstated **PREDICTIVE** capabilities
- ❌ Neglecting the importance of **HETEROGENEITY**

Research Questions

1. What can LLM agents really simulate for social science?
2. Where should we draw the boundary?

"Average Persona" & Heterogeneity

Keynesian Beauty Contest (Guessing 2/3 of the average from 0 to 100)



✅ Low **VARIANCE** (frequency differs)

❌ **MEAN** aligned (Reproduced peak values)

Aim at **collective behavior** and **qualitative patterns**, as these may be well-reflected despite *low individual variance*.

Three Key Boundaries

Boundary	Key Question	Takeaway
Alignment & Heterogeneity	"Average Persona" problem: LLMs tend toward mainstream behavior	Collective alignment does not guarantee individual fidelity.
Temporal Consistency	Can agents maintain role consistency in long-term simulations?	Apparent emergence may stem from inconsistency, not interaction.
Robustness	Are findings stable across different prompt settings?	A valid pattern must survive prompt and context perturbations.