Managing the Machine:

Does Organization Theory Matter When Organizing AI Agents?

Prasanna (Sonny) Tambe

Wharton School, University of Pennsylvania

with Yuan Gao (Boston University), Sharique Hasan (Duke), David Hsu (Penn), and Dokyun Lee (Boston University)

Columbia MAD

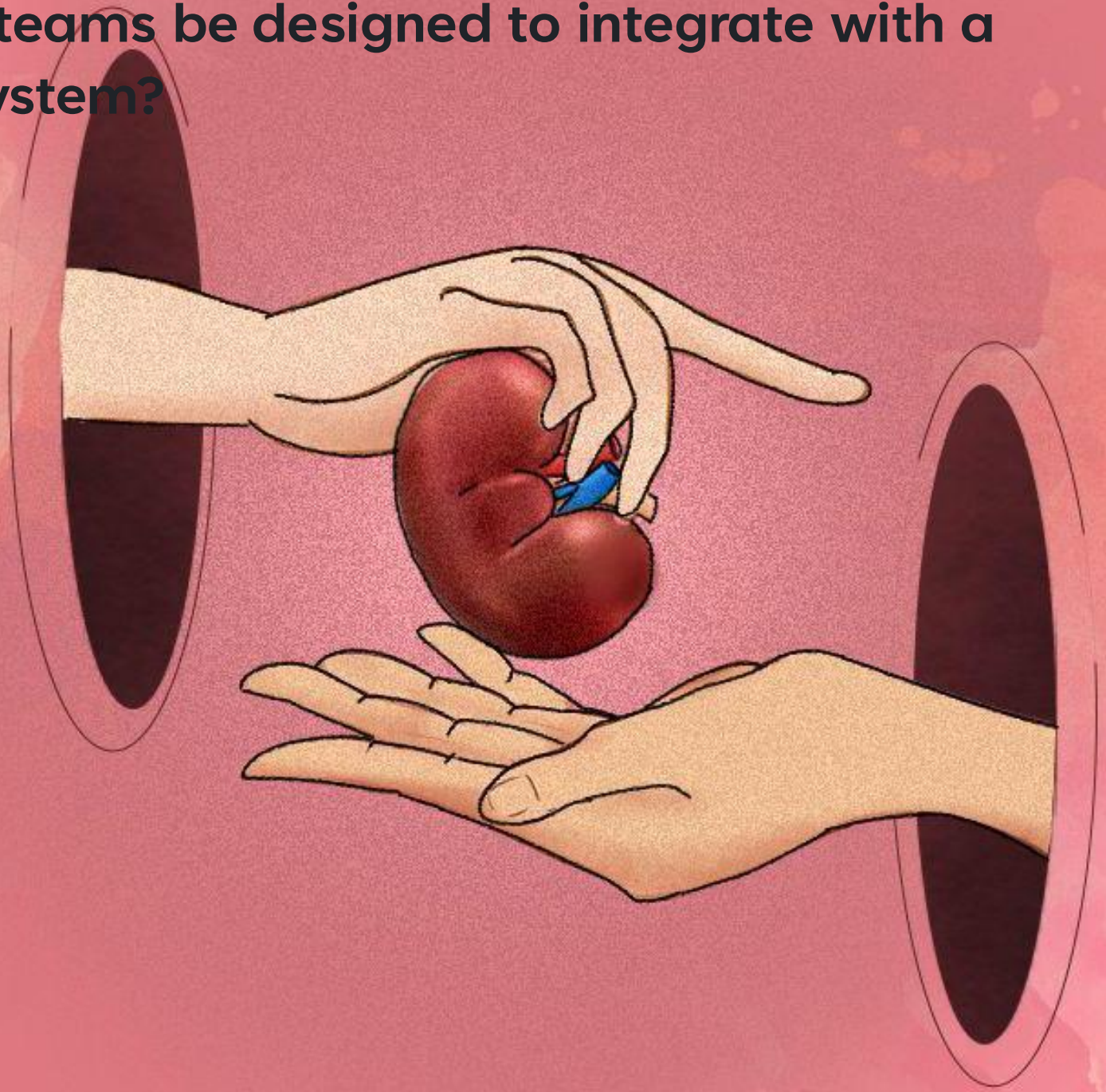
OpenAI Imagines Our AI Future

Stages of Artificial Intelligence

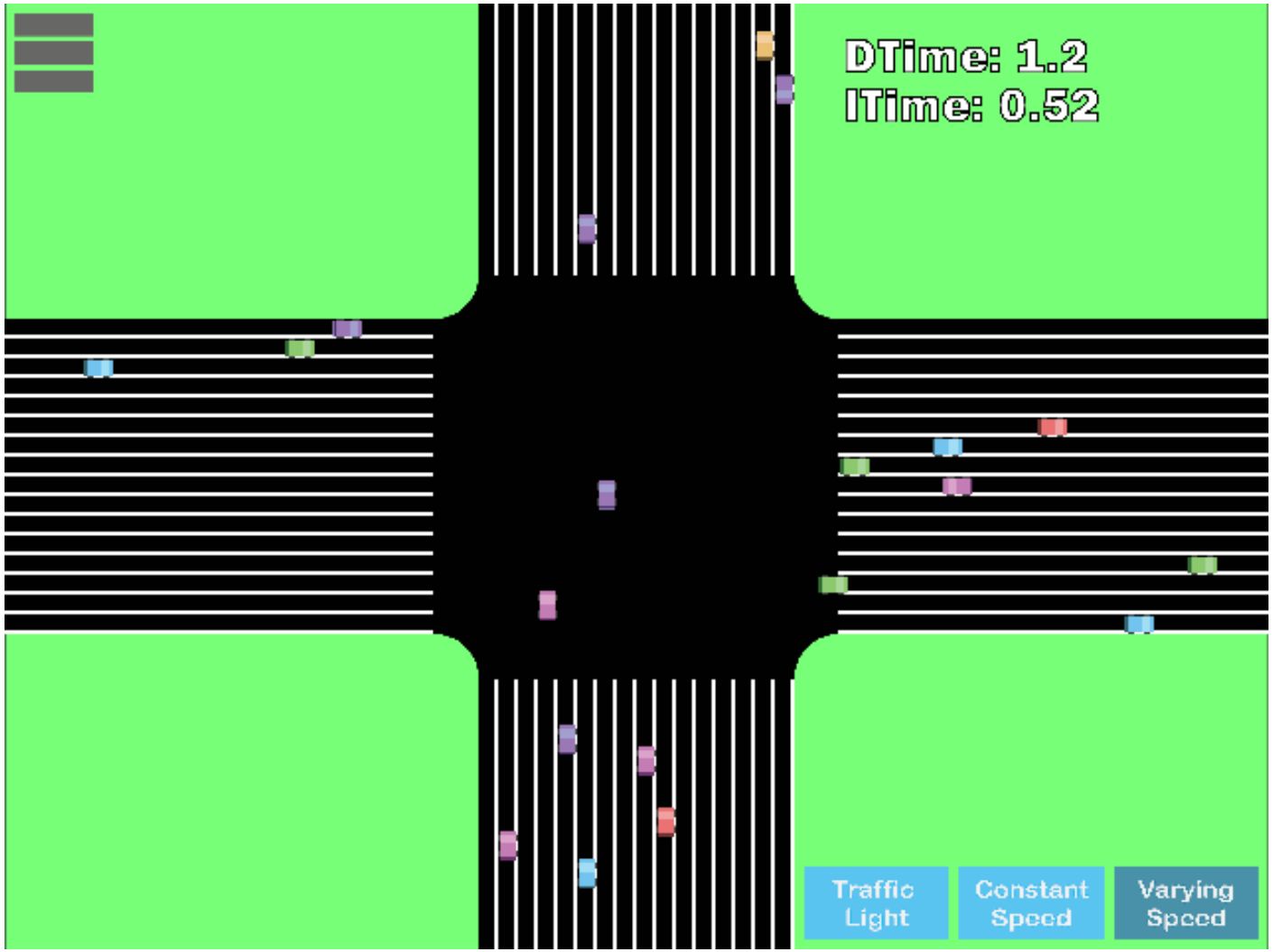
Level 1	Chatbots, AI with conversational language
Level 2	Reasoners, human-level problem solving
Level 3	Agents, systems that can take actions
Level 4	Innovators, AI that can aid in invention
Level 5	Organizations, AI that can do the work of an organization

Source: Bloomberg reporting

How should agent teams be designed to integrate with a complex, human system?

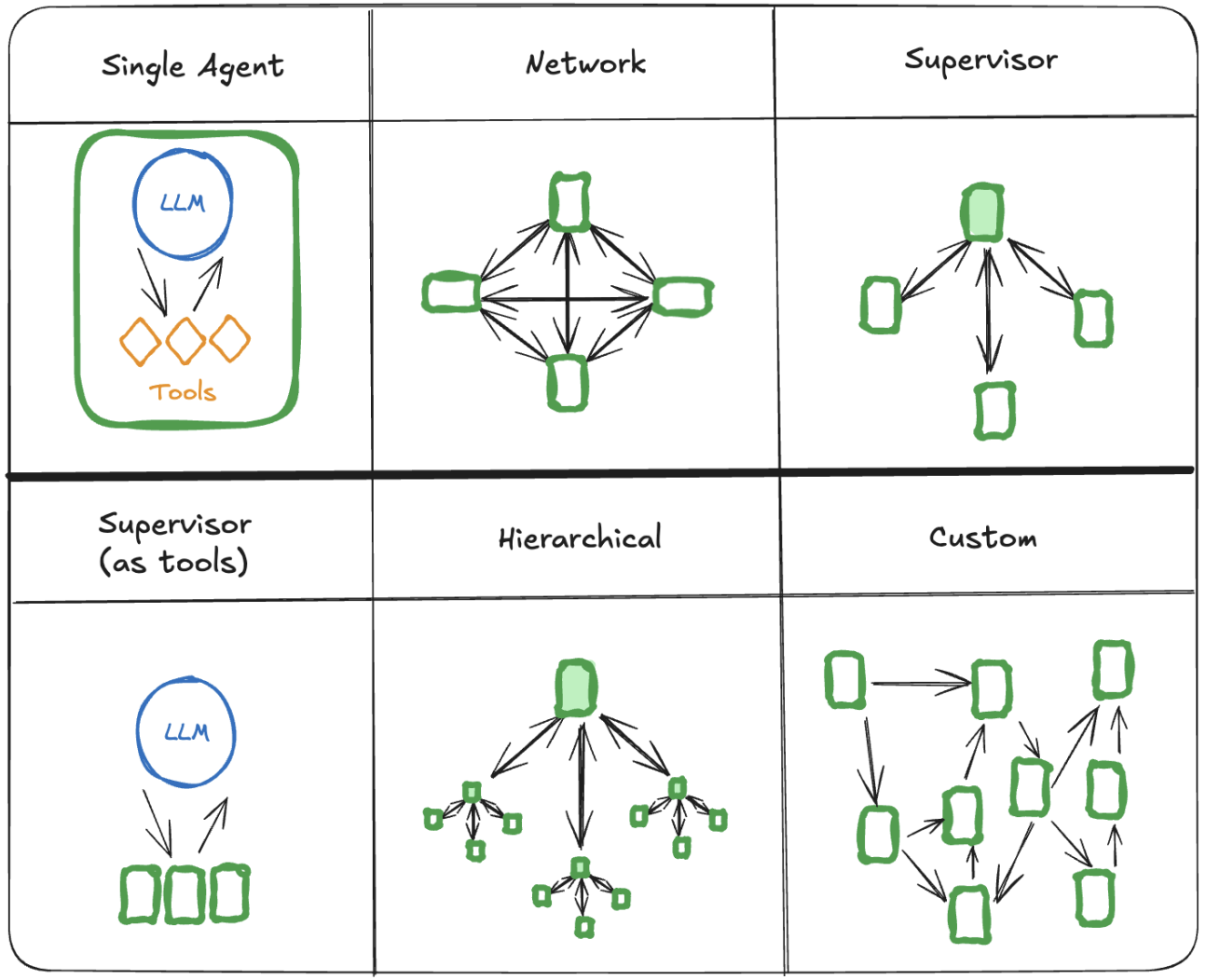


And so that human teams can interact with them?



Agentic platforms understand this challenge

from LangGraph



As you expand your system with more agents, the complexity can overwhelm a single supervisor, leading to suboptimal decisions and difficulty in managing context.

To mitigate this, consider a hierarchical design. Form specialized teams of agents, each managed by an individual supervisor, with a top-level supervisor overseeing the entire operation.

-- LangGraph

Research question: How can organizational theory inform the design of LLM teams?

This is an alignment question.

It is the "inverse" of the classic IT-org problem.

Note: Our unit of observation is configurations of LLM teams, not human-AI collaboration.

Why this is different for LLMs

🧠**Organizational theory** is meant to deal with human limitations; e.g. bounded rationality, moral hazard, etc.

🤖**LLMs EXHIBIT SIMILAR TRAITS**; e.g. not fully rational (goal stability, inconsistent reasoning), they can be deceptive, they exhibit biases, they “forget”, etc

Contribution ...

1. **HUMAN-AI INTEGRATION**
    - Organizational structures prevent fragmented/contradictory insights
    - Framing AI systems in organizational terms enhances transparency and intuition for humans, building trust.
2. **THEORETICAL FOUNDATIONS**
    - Theoretically-founded approaches enable adaptive, flexible structures
    - Better than fixed optimizations for single scenarios
    - More durable findings than combinatorial approaches

Background:

A great deal of thought has gone into the effective design of organizations.

MTM / Tambe

The Principles of Scientific Management

BY

FREDERICK WINSLOW TAYLOR, M.E., Sc.D.

PAST PRESIDENT OF THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS

MTM / Tambe

MTM / Tambe

Key insight: Effective organization → competitive advantage

Leadership & Organizational Behavior
(LEAD)

This course focuses on how managers become effective leaders by addressing the human side of enterprise.

The first modules examine teams, individuals, and networks in the context of:

- The determinants of group culture.
- Managing the performance of individual subordinates.
- Establishing productive relationships with peers and seniors over whom the manager has no formal authority.

The intermediate modules look at successful leaders in action to see how they:

- Develop a vision of the future.
- Align the organization behind that vision.
- Motivate people to achieve the vision.
- Design effective organizations and change them to achieve superior performance.

Leadership & Organizational Behavior
(LEAD)

This course focuses on how managers become effective leaders by addressing the human side of enterprise.

The first modules examine teams, individuals, and networks in the context of:

- The determinants of group culture.
- Managing the performance of individual subordinates.
- Establishing productive relationships with peers and seniors over whom the manager has no formal authority.

The intermediate modules look at successful leaders in action to see how they:

- Develop a vision of the future.
- Align the organization behind that vision.
- Motivate people to achieve the vision.
- Design effective organizations and change them to achieve superior performance.

How does this need to change when the team is comprised of LLMs?

We focus on TWO organizational choices:

1. **HIERARCHY** (managers vs flat structure)
2. **TEAM DIVERSITY** (heterogeneity in knowledge among agents)

Typical organizations

	Low Diversity	High Diversity
High Hierarchy	• Military units • Traditional manufacturing	• Large consulting firms • Healthcare systems
Low Hierarchy	• Worker cooperatives • Unions	• Open source projects • Creative agencies

Which problems are best suited to which configurations?

	Low Diversity	High Diversity
High Hierarchy	• SPEED + SCALE	• COMPLEXITY
Low Hierarchy	• TRUST	• CREATIVITY

empirical predictions

benchmark task: Hierarchy is bad for creativity
benchmark task: Hierarchy is good for complexity
real task: Hierarchy is bad for evaluating innovation outcomes

Impact of diversity on team performance

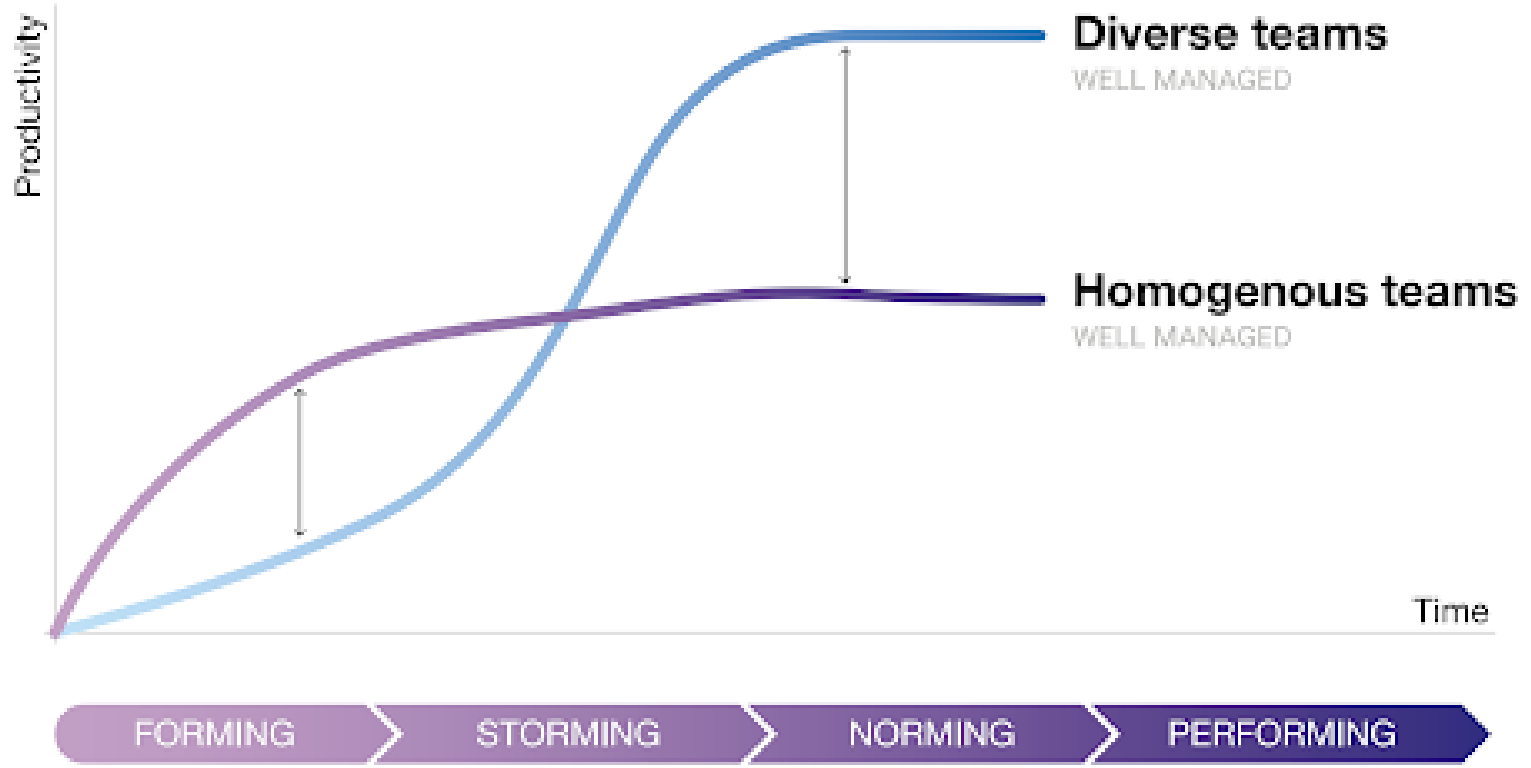


Figure 1: Impact of diversity on team performance
Source: Korn Ferry Institute, 2019

Setting 1 We can use LLM Benchmarks (e.g. LiveBench)

Evaluating LLM performance across **18 tasks** in **6 categories** of real-world challenges

"LiveBench.ai is a benchmark for LLMs designed with test set contamination and objective evaluation in mind."

Key LLM Challenges:

- Ambiguity in task definitions
- Varying degrees of task complexity
- Need for multi-step reasoning

Key Properties of LiveBench & BigBench questions for our use case


Grounded in real-world scenarios and practical applications


Support multiple solution strategies and approaches

✓ Feature clear, objective evaluation criteria

Implementing hierarchy and diversity:

Team Diversity

No diversity / Model Diversity (GPT 3.5, 4o, LLaMa) / Thinking styles (CoT, Reflection)

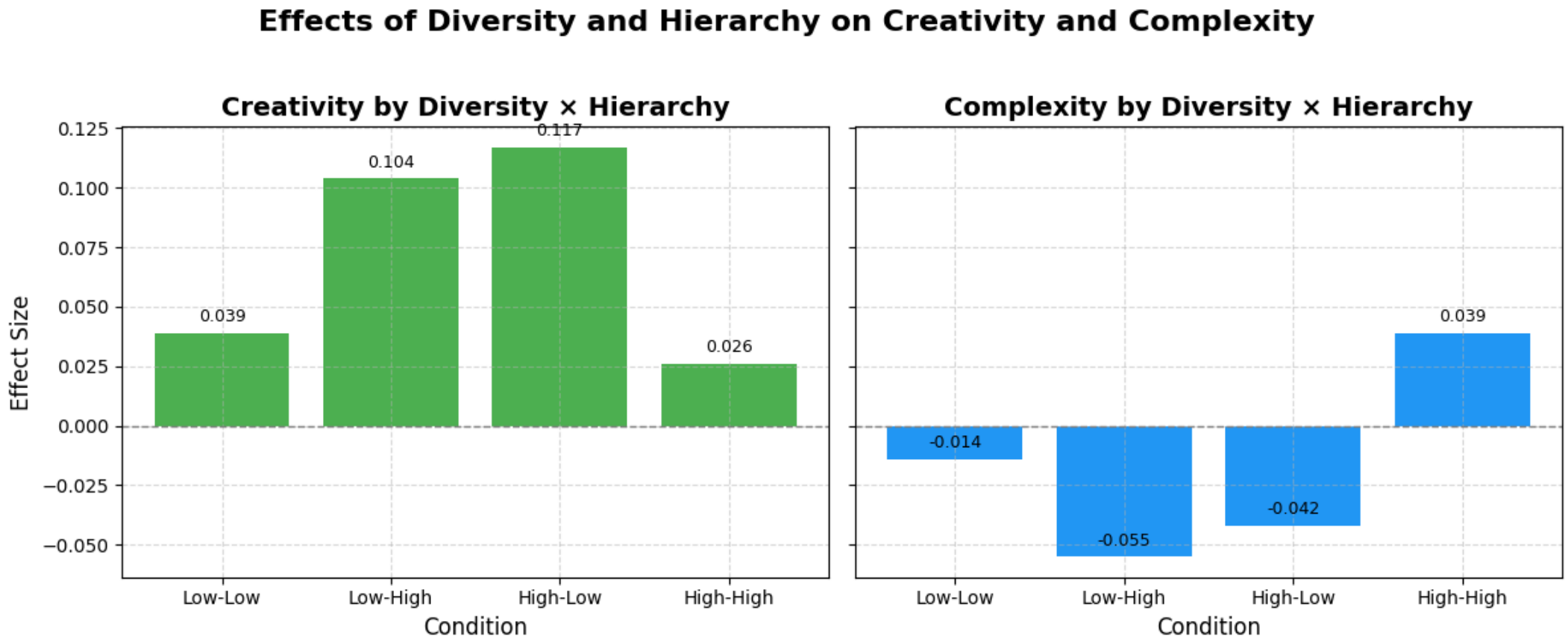
Team Hierarchy & Roles

Structure: Flat (peer-to-peer) / Hierarchy (leader-member)

Member: "Collaborate and share ideas." **Leader:** "Facilitate discussion."

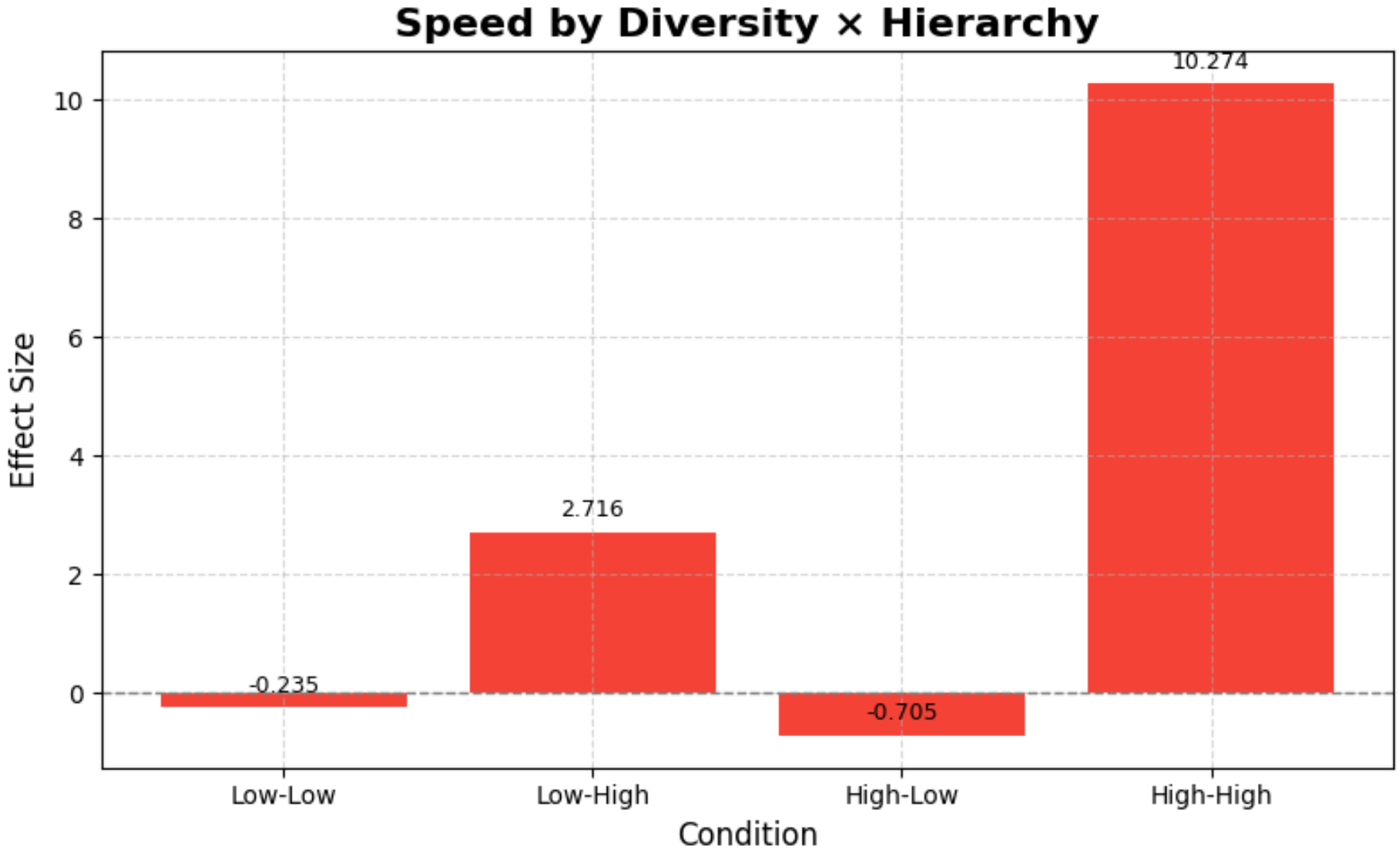
Consensus Mechanism

Flat: Any member can determine consensus **Hierarchy:** Only leader decides



Creative tasks favor low hierarchy, high diversity

Complex tasks favor high hierarchy, high diversity



But diversity comes at a cost

We can examine the token trail to investigate why errors occur in different configurations (ongoing)

Leader:

- **Analyzed** the vertices A, C, G, E and guided the team to see they form a **square**.
- **Prompted** the team to confirm the conclusion.

Member 2:

- **Directly answered** that the shape is a **square**, confirming the analysis.

Member 3:

- **Agreed** with the answer and **explained why** it is a square, adding supporting details.

SETTING 2 Commercializing Scientific Innovations

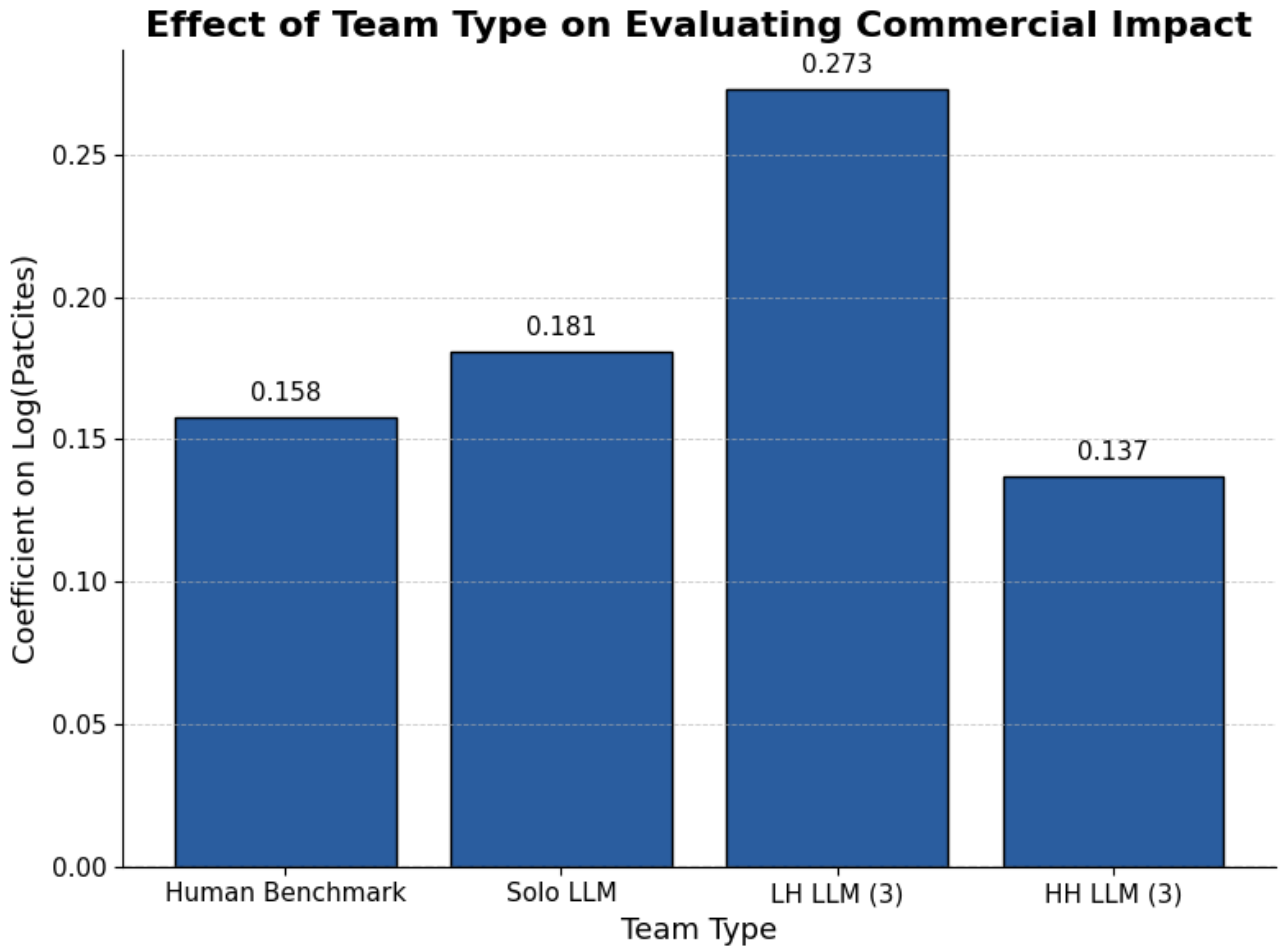
- We ask human and LLM teams to assess the potential of scientific innovations
- This task is critical for directing R&D investments and resource allocation efficiently
- Helps bridge the gap between scientific discovery and market impact
- Enables early identification of high-potential innovations that warrant commercialization support

MTM / Tambe

A known hard problem: mRNA vaccine

Key advantages for our setting

1. This is an open-ended task with no CORRECT answer
2. It involves team-based decision making. Often have to rally experts from different domains. TTO is a team-based decision process.
3. We can compare predictions with existing human benchmarks of commercial impact (Reliance on Science patenting database (Marx))



Evaluating Scientific Innovations: Flat > Hierarchy

Thank you. We have several “design” choices to make and your comments and suggestions are greatly appreciated!

Prasanna (Sonny) Tambe.

tambe@wharton.upenn.edu.