*Lecture 1.*

# A Tour of Simulations: Past, Present, and Future

## CS 222: AI Agents and Simulations
## Stanford University

## Joon Sung Park

# Welcome to CS 222!

# Nice to meet everyone!

My name is Joon — I am your instructor for this course!

Currently a 5th-year Ph.D. candidate in the computer science department, researching AI agents and simulations.

CS 222 is a brand-new course in the core AI lineup, and we are excited to offer it this quarter!

# Class Logistics

# Course information

**Course website:**

https://joonspk-research.github.io/cs222-fall24/index.html

**Course structure:**

This is a 200-level CS course that satisfies the "Learning and Modeling" breadth requirement for CS Ph.D. students.

It is designed to be a mix of seminal classes, some assignments, and a final project.

This course does require coding abilities (in Python).

# Assignments and Grading

## Reading commentaries *(30%)*

- There are two required readings per lecture.
- Please write one commentary that cover both readings (3 ~ 4 paragraphs).
- These are due at **10:00 PM the day before the lecture** on Canvas.
- The commentaries are graded on a {0, 1, 2} scale.
- We will drop two-lowest grade from your reading commentaries.

# Assignments and Grading

## <u>Two simulation assignments</u> *(30% each)*

- There are two simulation assignments this quarter.
- They will account for 15% of your grade each.

## <u>One final group project</u> *(30%)*

- There is one final project for this quarter (in groups).
- The proposal will account for 5% of your grade, and the final submission will account for 25% of your grade.

# Assignments and Grading

## Class participation *(10%)*

- Please note that attendance is mandatory!
- I strongly encourage you to participate!

# In case useful...

**My office hour:**

Office Hours: Friday 1:00-2:00 pm;
Location: Gates 3B Common Area

**Commentary guidelines**

https://joonspk-research.github.io/cs222-fall24/commentaries.html

- The commentaries are graded on a {0, 1, 2} scale.
- 1 == A.
- More info on other assignments will be provided later!

# Important: Class interest form

*This class has received significantly more registrations than we had planned for. We are currently discussing how best to accommodate the interest while ensuring a good course experience.*

But important for right now: please respond to this interest form by **10 pm on Tuesday, September 24, 2024.**

What are simulations?

Why should you care?

Why now?

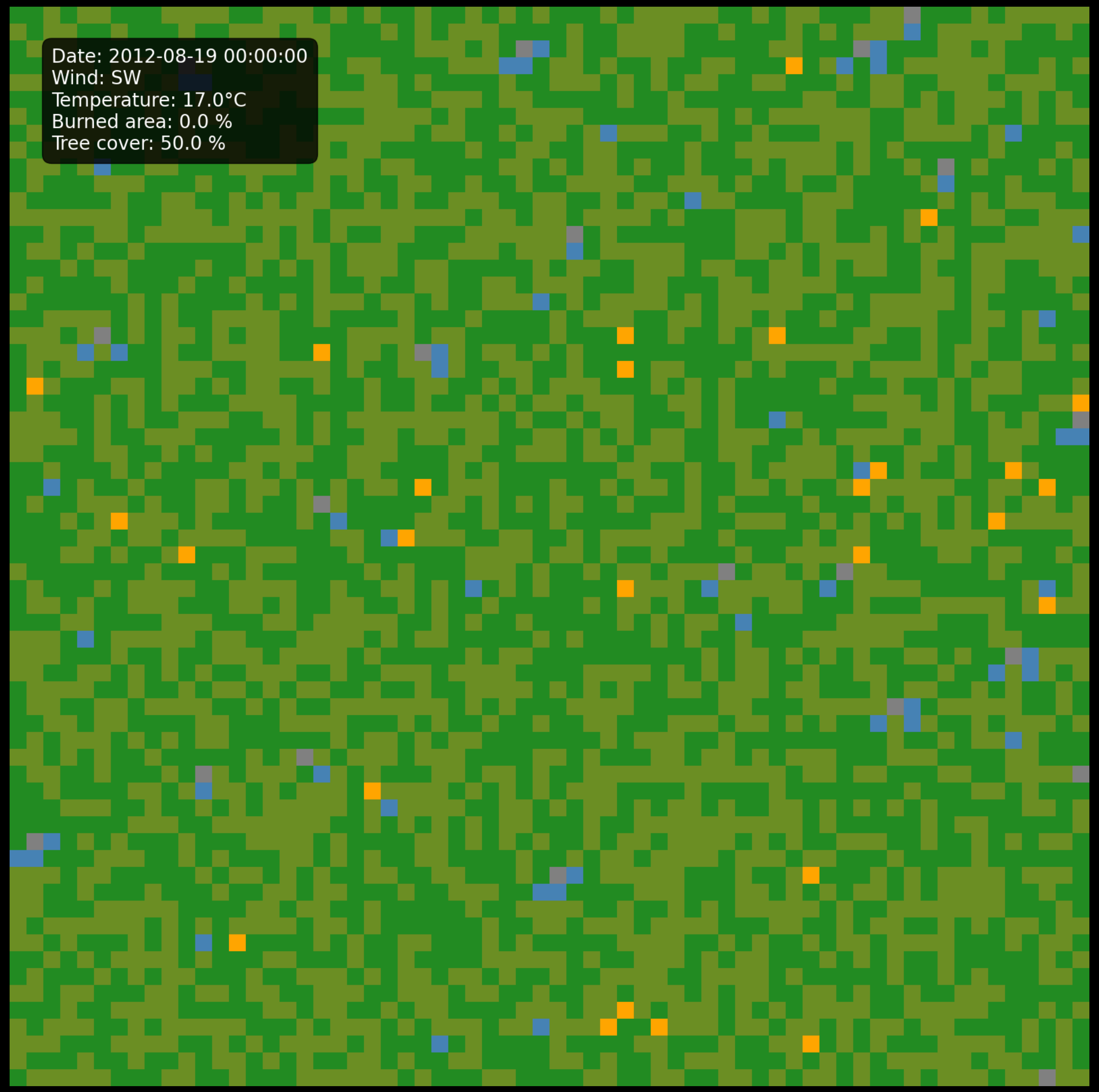
# Q: What are simulations?

## Where have you seen them before?

# Examples of simulations you may have seen before.

In games (e.g, The Sims)

In movies (e.g, The Matrix)



```
Date: 2012-08-19 00:00:00
Wind: SW
Temperature: 17.0°C
Burned area: 0.0 %
Tree cover: 50.0 %
```

In your courses (e.g, forest fire)

# In short, simulations are...

A program that defines an *environment* and the behaviors of *individuals*, then outputs the resulting world.

# More formally...

$W(t)$: The world's state over time $t$.

$E$: The environment, defined by a set of state $S_E$ and rules $R_E$ that govern the environment's dynamics.

$A_i$: Individual agent $i$ in the environment, where $i = 1{,}2{,}...{,}N$ for $N$ agents.

Simulations are a recursive function:

$$W(t) = \left(S_E(t), S_{A1}(t), S_{A2}(t), \ldots, S_{AN}(t)\right)$$

where $W(t+1)$ is recursively defined by the interactions of the environment and agents according to the rules $R_E$ and behaviors $B(A_i)$.

# User-facing features of simulations

We can run them multiple times from the same initial state. (Do you think they are deterministic?)

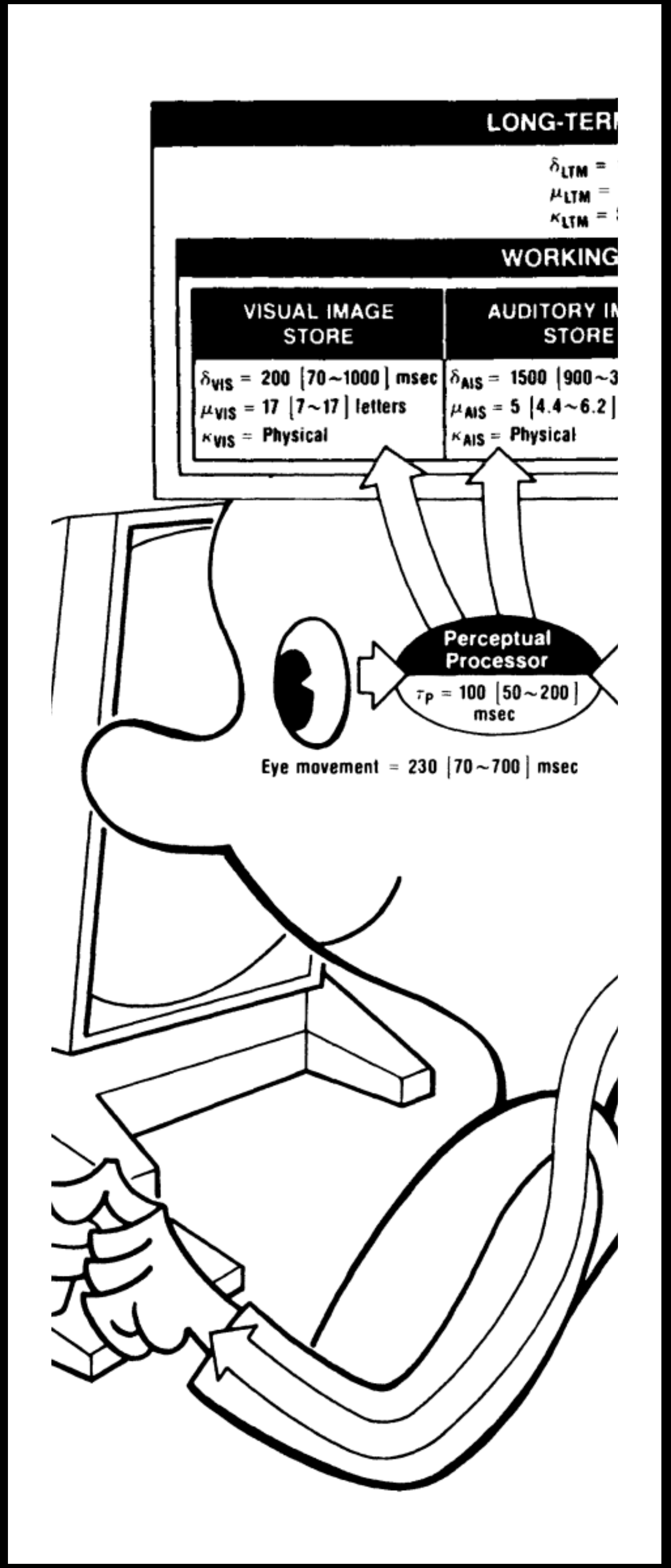
We can influence the state of the simulations.

And, in return, simulations surprise us. (The fact that they surprise us should be surprising, given that we know the rules!)

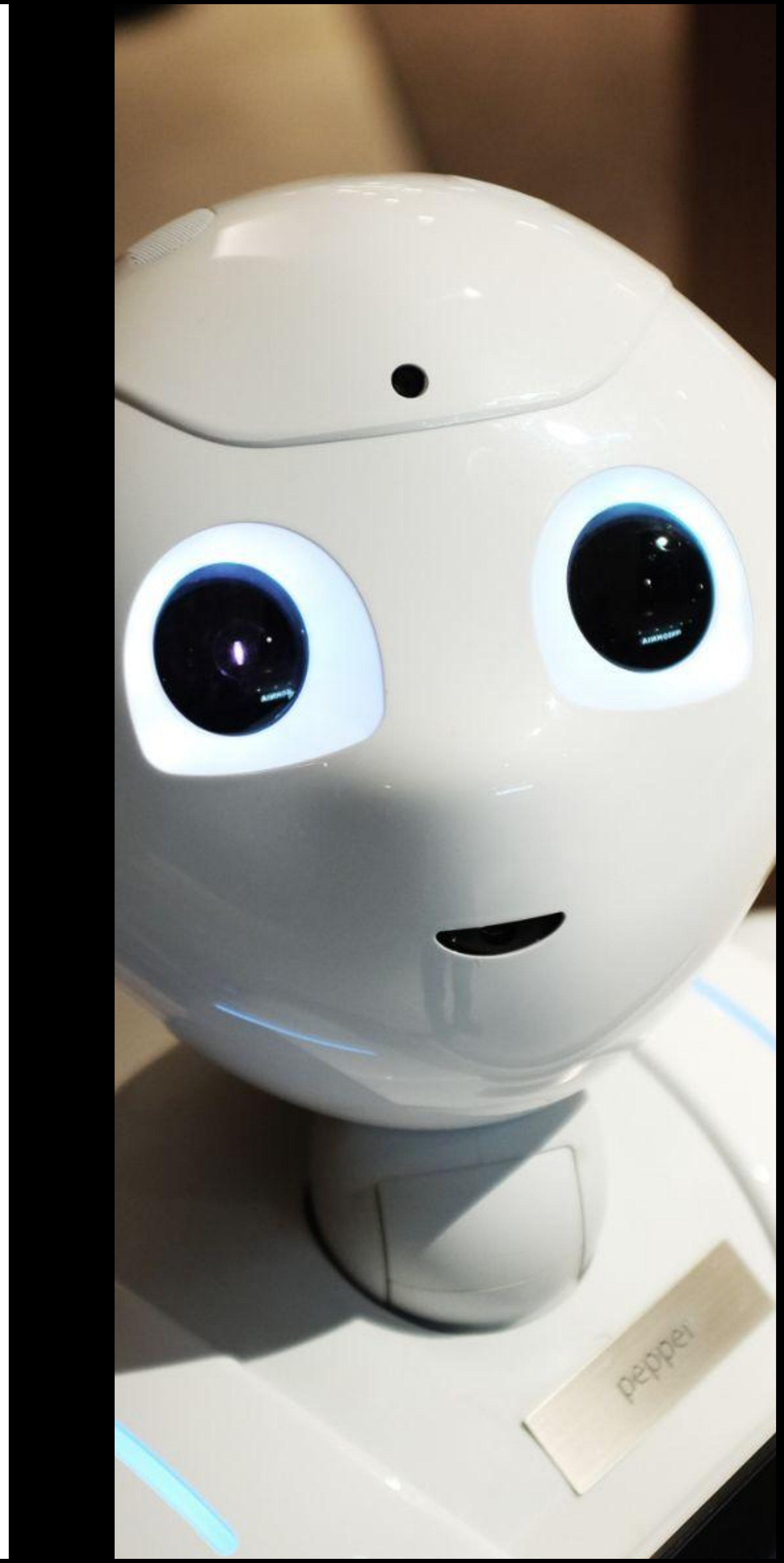
# Q: Why should you care?

## What can you do or answer with simulations?

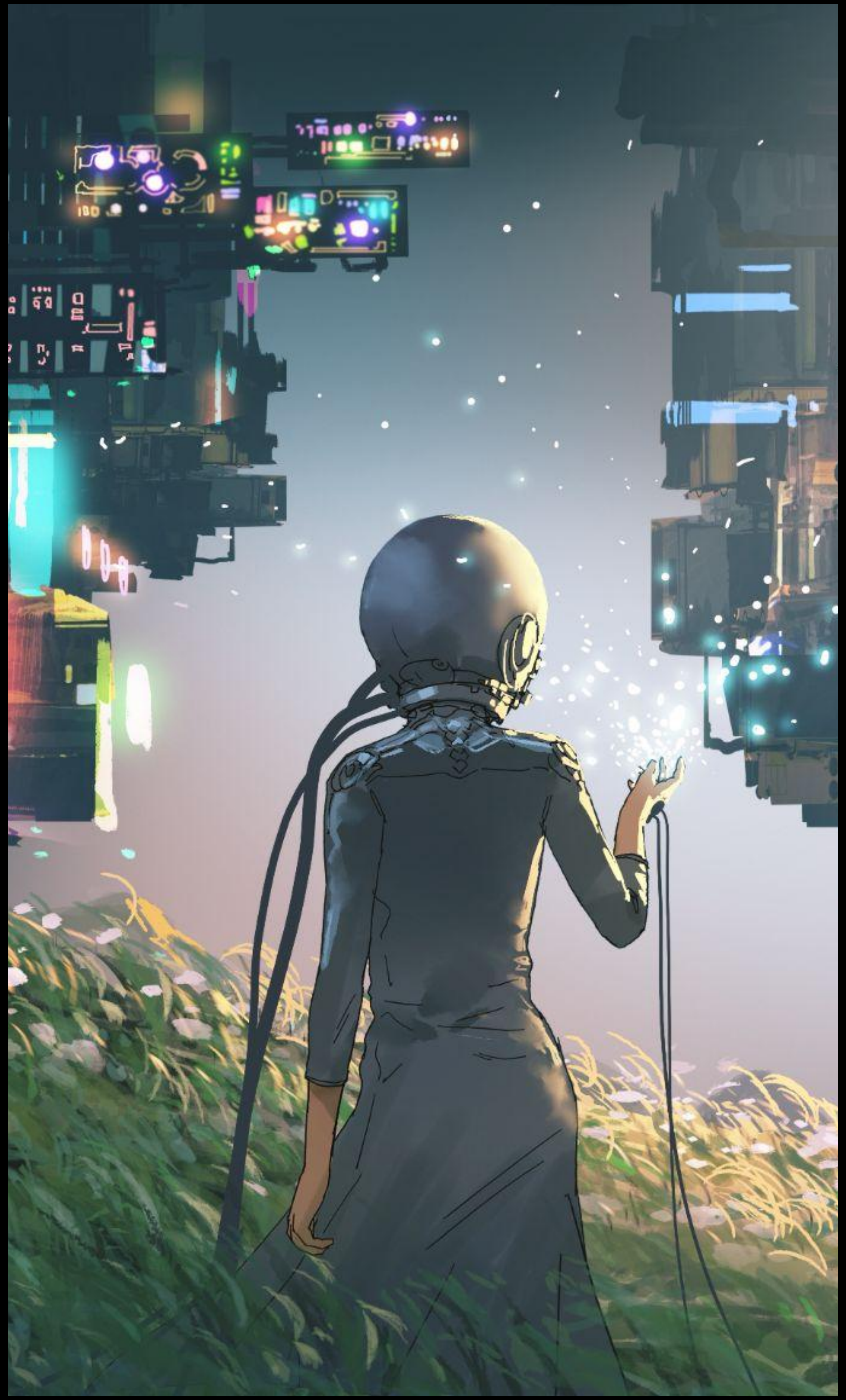
Simulations allow us to ask 'what-if' counterfactual questions by creating a multiverse of possibilities.


LONG-TERI
$\delta_{LTM}$ =
$\mu_{LTM}$ =
$\kappa_{LTM}$ =
WORKING
VISUAL IMAGE STORE
$\delta_{VIS}$ = 200 [70~1000] msec
$\mu_{VIS}$ = 17 [7~17] letters
$\kappa_{VIS}$ = Physical
AUDITORY IM STORE
$\delta_{AIS}$ = 1500 [900~3
$\mu_{AIS}$ = 5 [4.4~6.2]
$\kappa_{AIS}$ = Physical
Perceptual Processor
$\tau_P$ = 100 [50~200] msec
Eye movement = 230 [70~700] msec

**Behavioral Models**



**Social Robots**

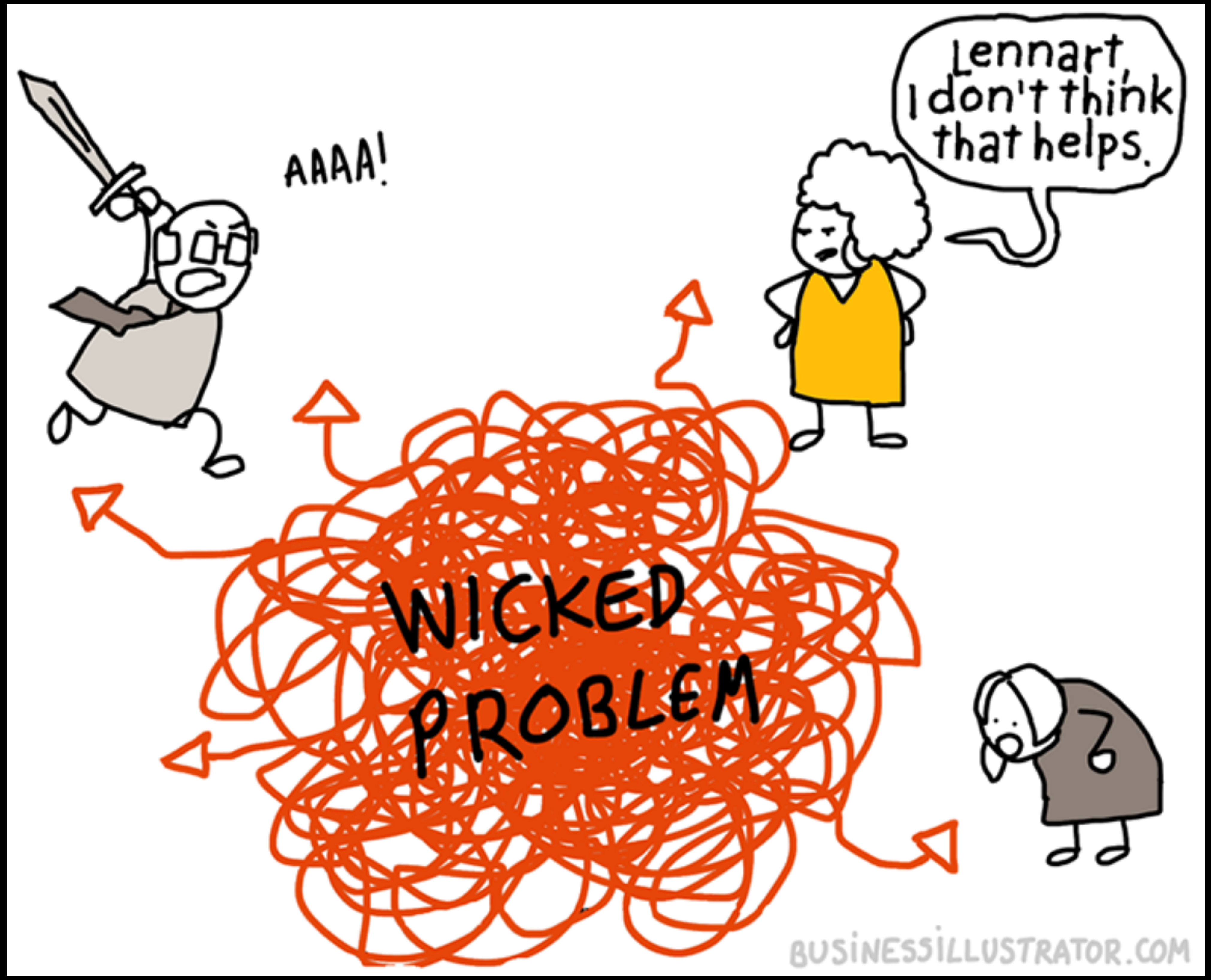


**Non-Playable Characters**

**Agent-Based Models**

SK Card, TP Moran, and A Newell. 1983. The psychology of human-computer interaction. (1983).
Mark Weiser. 1999. The Computer for the 21st Century. SIGMOBILE Mob. Comput. Commun. Rev. 3, 3 (jul 1999).
Allen Newell. 1990. Unified Theories of Cognition. Harvard University Press, Cambridge, Massachusetts.

Many problems in the world are wicked, characterized by complex equilibria and real-world constraints.



H. W. J. Rittel, M. M. Webber, Dilemmas in a general theory of planning. Policy Sciences 4, 155-169 (1973).

# As an individual...

What class do I want to take?

What major should I pursue?

# As a group...

How do I rehearse for a difficult conversation?

How do I coordinate when there are conflicting values or goals between people?

# As a society...

How do we organize collective action for sustainability?

How do we mitigate the spread of misinformation?

Many challenges we face require us to explore complex counterfactuals that cannot be tested in the real world.

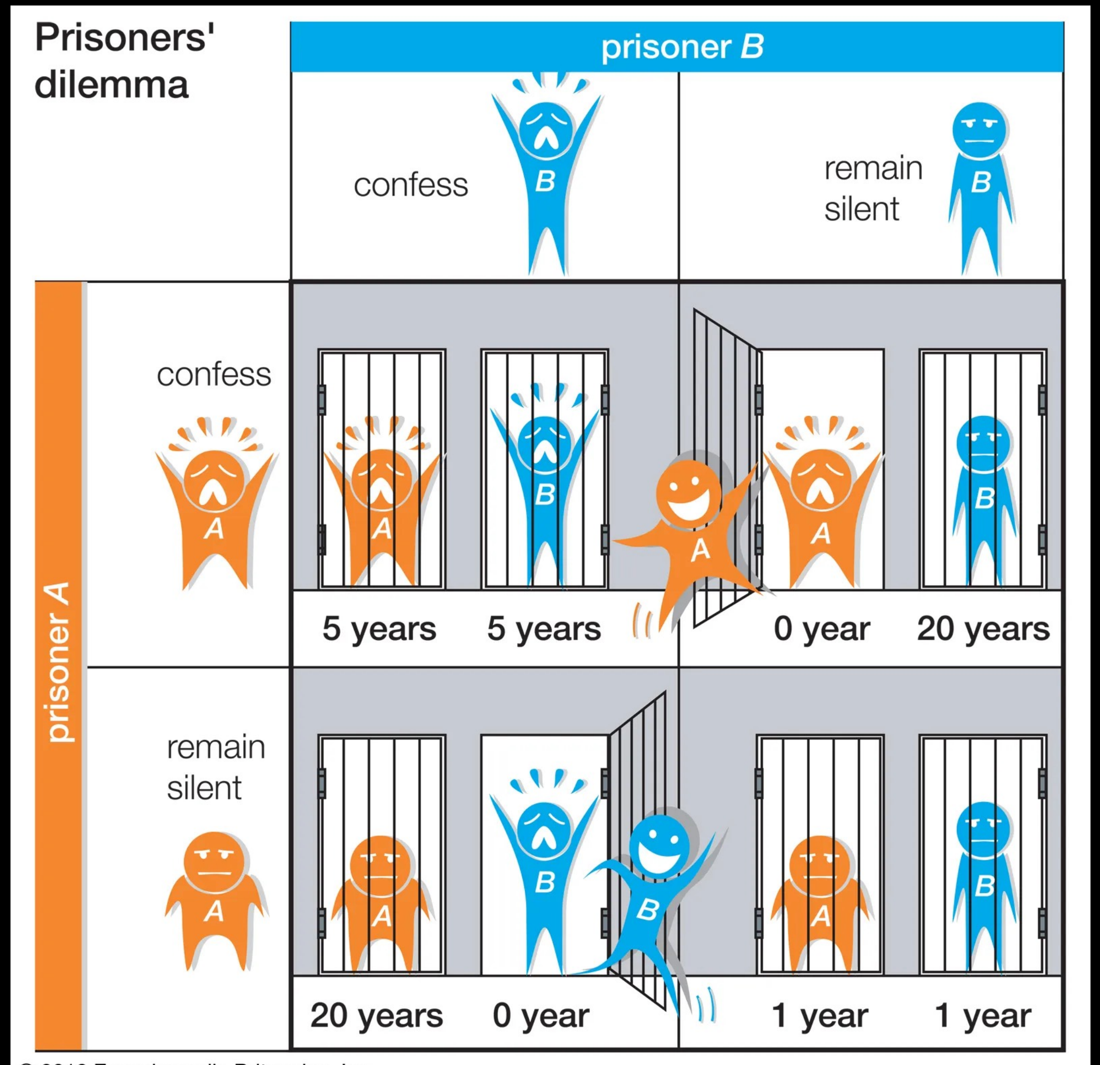
Simulations offer the potential to answer questions we previously had no way of answering.

# Q: Why now?

## Is it a particularly exciting time for simulations?

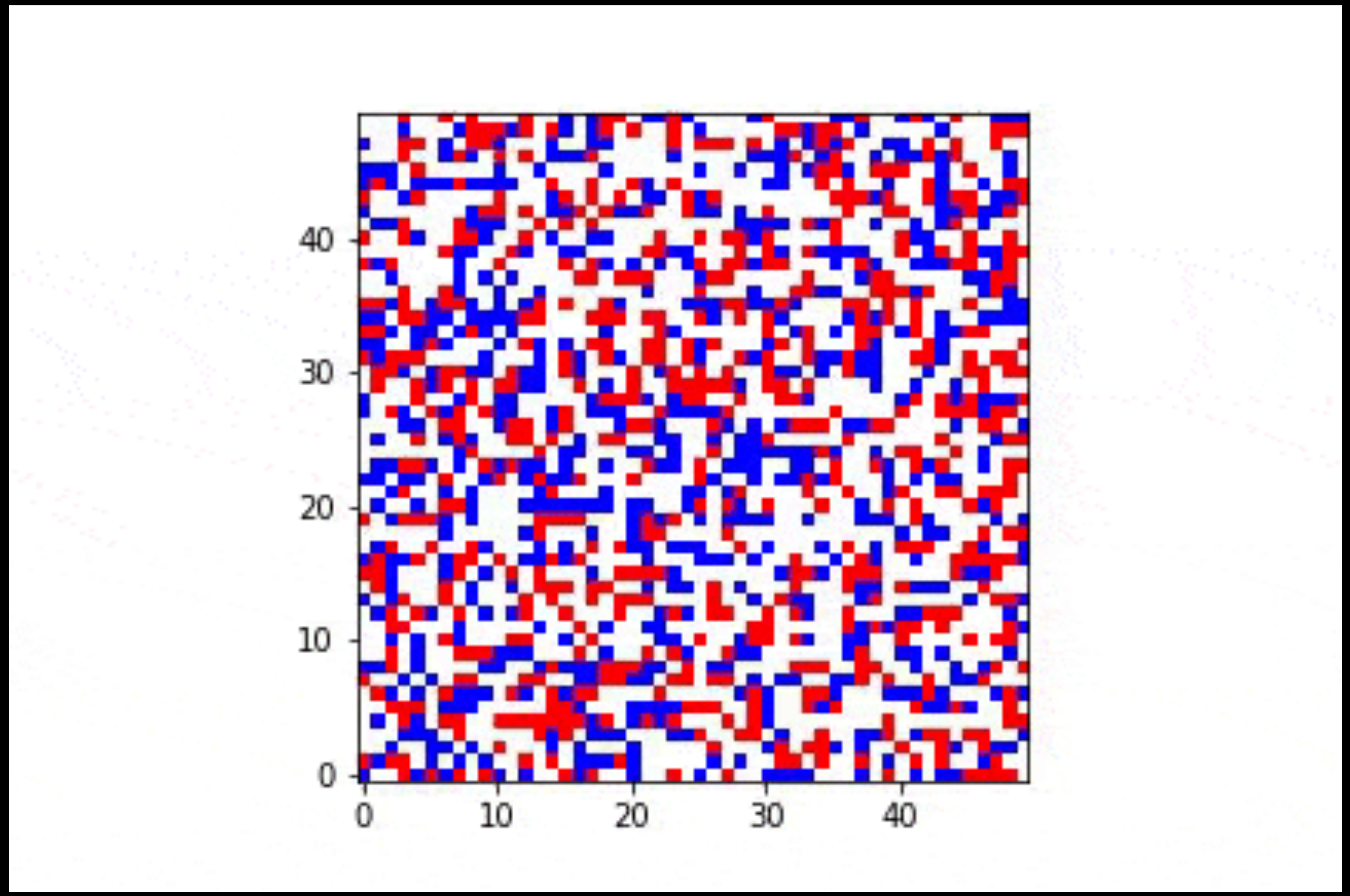
# The idea of simulation is not new.

Cellular automata



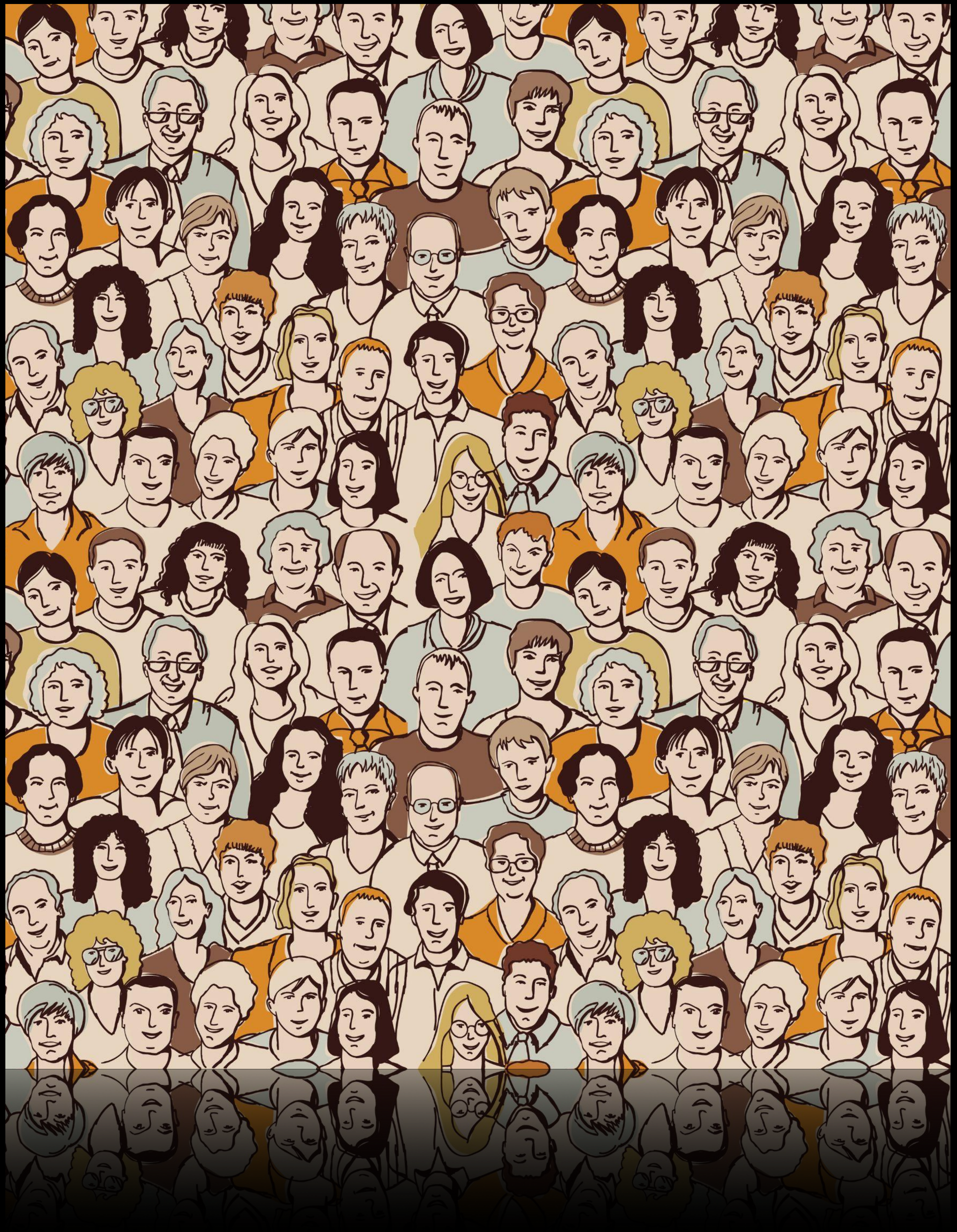
© 2010 Encyclopædia Britannica, Inc.

Game theory



Agent-based models

A new paradigm shift allows us to revisit old problems with fresh insights.



## Large language models can be prompted to generate human behavior conditioned on a variety of experiences.

GPT

`“[name] is a [description]”`

**Social Simulacra (UIST ’22)**

Joon Sung Park, Lindsay Popowski, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Social Simulacra: Creating Populated Prototypes for Social Computing Systems. UIST 2022.

LW:
RP: 💤
AC: 💤
FL: 💤
HJ: 💤
AB:
IR:
RP: 💤
GR: 💤
CG: 💤
WS:
SM:
KM: 💤
JL:
AK:
ML: 💤
AS:
YY: 💤
JM:
TT:
JM:
CO: 💤
TM: 💤
ML: 💤
EL: 💤

# What I want you to get out of this class

# Simulation is a new emerging field

Learn about it!

I am excited, and I think you should be too.

# Seminal (70%) + practice (30%) course

What are the key challenges the field of simulations is grappling with? For example, how do we build and evaluate these simulations?

What's the history of simulations, and where are we headed?

And try your hand at building agents and simulations.

# This is a new course!

We will learn together as we develop this course throughout the quarter.

Participate! This is meant to be a discussion-heavy class!

By the end of the quarter, I hope you see simulations as an exciting emerging direction and envision all the cool things you might be able to do with them.

# A tour of simulations

(a quick teaser for the quarter)

# Act 1: Past

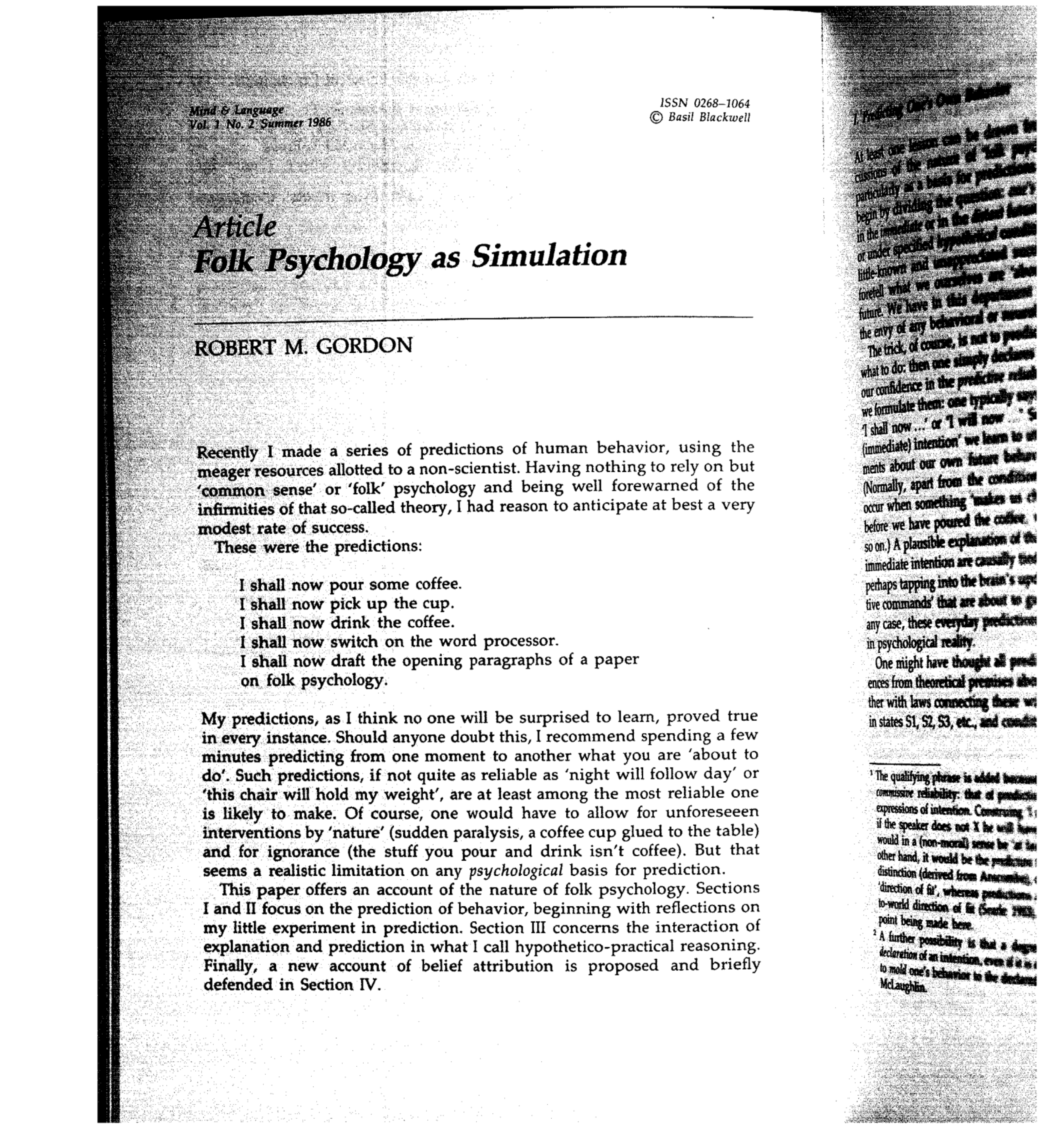
## Examples of simulations pre-generative AI

For each method, discuss:

1) How did we define “agents”?

2) How did we define “environment”?

# Theory of Mind (ToM)



*Mind & Language*
*Vol. 1 No. 2 Summer 1986*

*ISSN 0268–1064*
*© Basil Blackwell*

*Article*

## *Folk Psychology as Simulation*

ROBERT M. GORDON

Recently I made a series of predictions of human behavior, using the meager resources allotted to a non-scientist. Having nothing to rely on but 'common sense' or 'folk' psychology and being well forewarned of the infirmities of that so-called theory, I had reason to anticipate at best a very modest rate of success.

These were the predictions:

I shall now pour some coffee.
I shall now pick up the cup.
I shall now drink the coffee.
I shall now switch on the word processor.
I shall now draft the opening paragraphs of a paper
on folk psychology.

My predictions, as I think no one will be surprised to learn, proved true in every instance. Should anyone doubt this, I recommend spending a few minutes predicting from one moment to another what you are 'about to do'. Such predictions, if not quite as reliable as 'night will follow day' or 'this chair will hold my weight', are at least among the most reliable one is likely to make. Of course, one would have to allow for unforeseeen interventions by 'nature' (sudden paralysis, a coffee cup glued to the table) and for ignorance (the stuff you pour and drink isn't coffee). But that seems a realistic limitation on any *psychological* basis for prediction.

This paper offers an account of the nature of folk psychology. Sections I and II focus on the prediction of behavior, beginning with reflections on my little experiment in prediction. Section III concerns the interaction of explanation and prediction in what I call hypothetico-practical reasoning. Finally, a new account of belief attribution is proposed and briefly defended in Section IV.

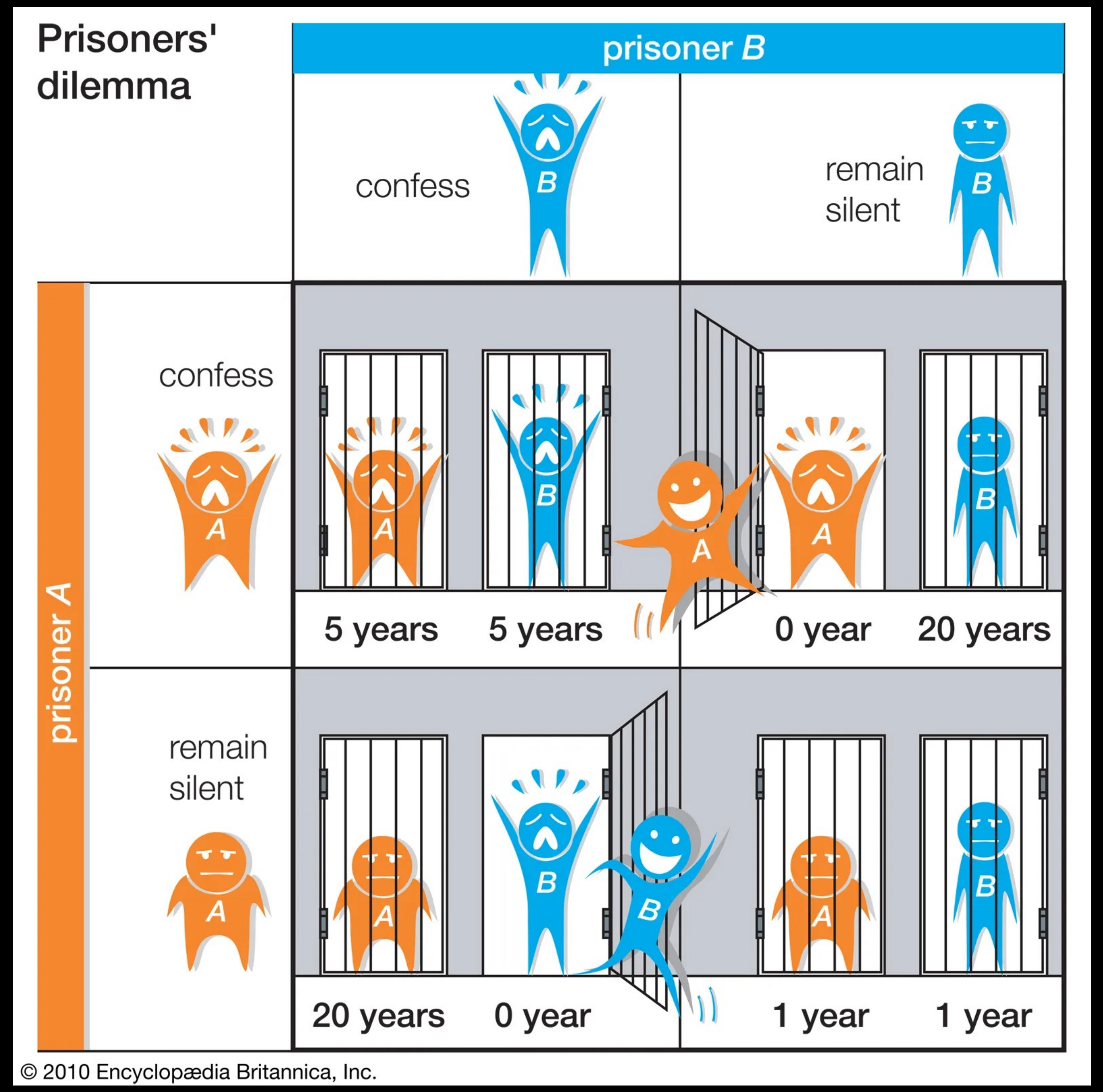
Gordon, R. M. (1986). Folk psychology as simulation. Mind & Language, 1(2), 158-171.

# Cellular automata

J. von Neumann, Theory of Self-Reproducing Automata, A. W. Burks, Ed. (University of Illinois Press, 1966).

S. Wolfram, A New Kind of Science (Wolfram Media, 2002).

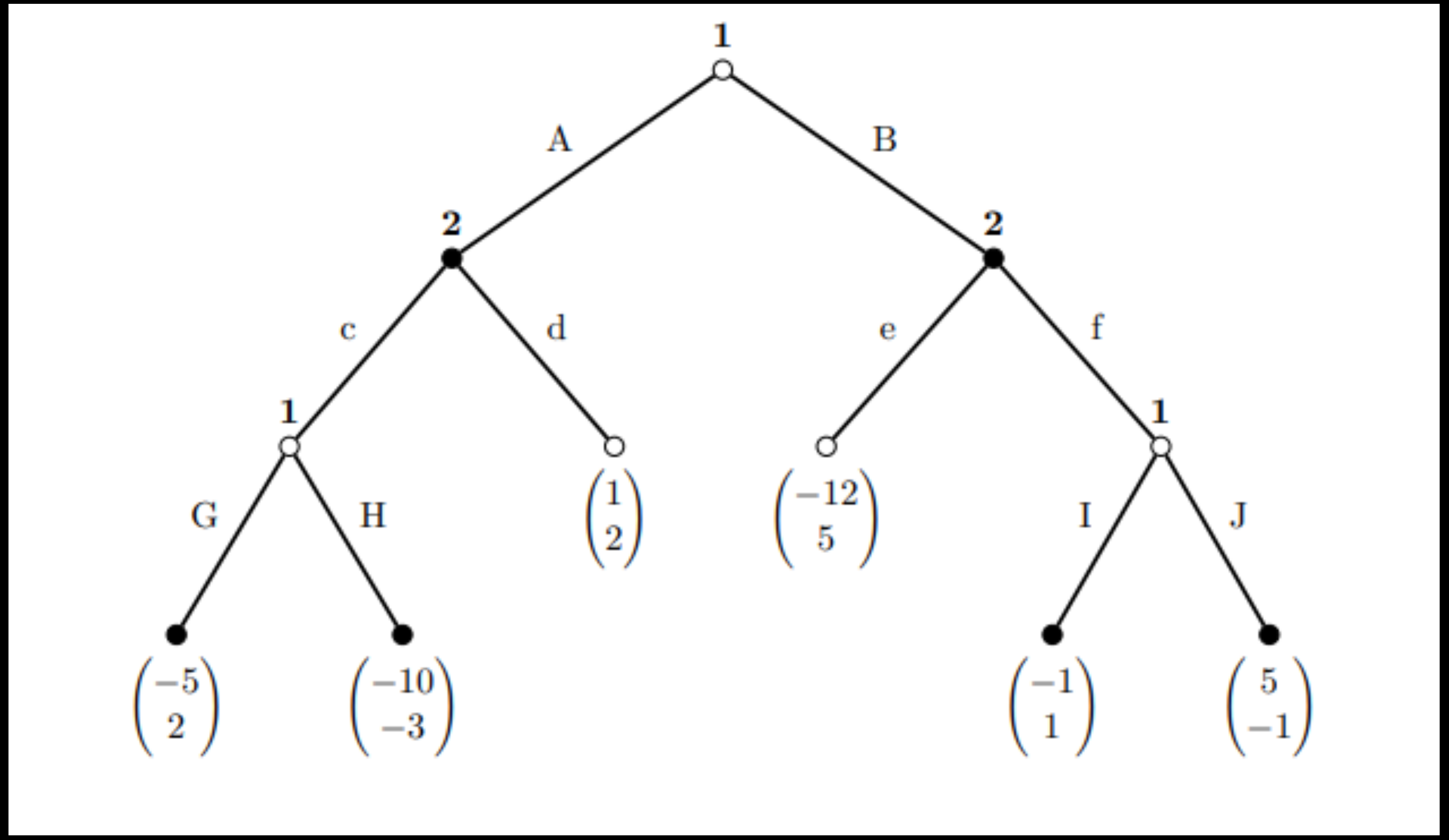
# Game theory



Prisoners' dilemma

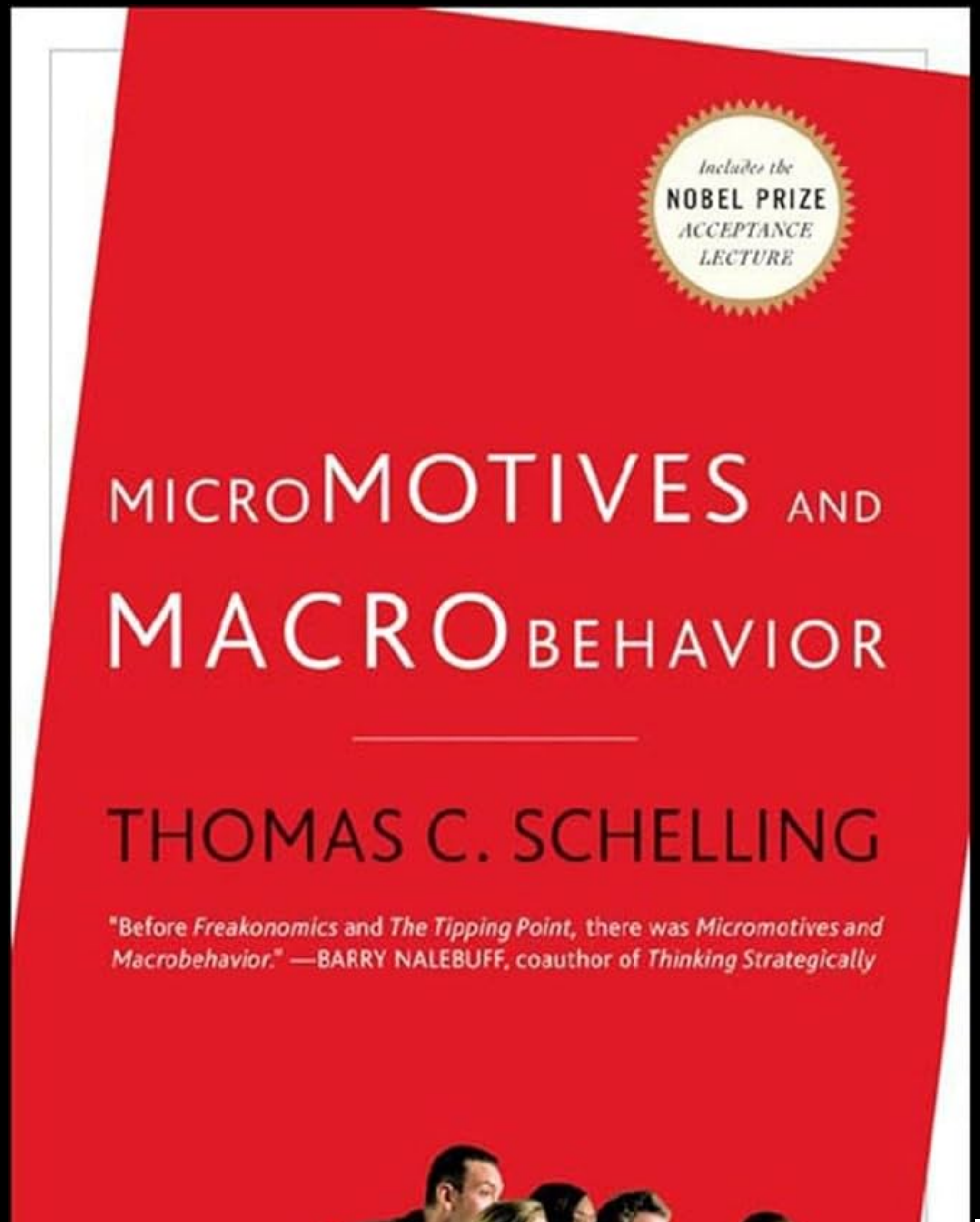
| prisoner A | prisoner B: confess | prisoner B: remain silent |
|---|---|---|
| confess | 5 years, 5 years | 0 year, 20 years |
| remain silent | 20 years, 0 year | 1 year, 1 year |

© 2010 Encyclopædia Britannica, Inc.


1
A
B
2
2
c
d
e
f
1
1
G
H
I
J
$\begin{pmatrix}1\\2\end{pmatrix}$
$\begin{pmatrix}-12\\5\end{pmatrix}$
$\begin{pmatrix}-5\\2\end{pmatrix}$
$\begin{pmatrix}-10\\-3\end{pmatrix}$
$\begin{pmatrix}-1\\1\end{pmatrix}$
$\begin{pmatrix}5\\-1\end{pmatrix}$

J. von Neumann, O. Morgenstern, Theory of Games and Economic Behavior (Princeton University Press, 1944).

# Agent-based models

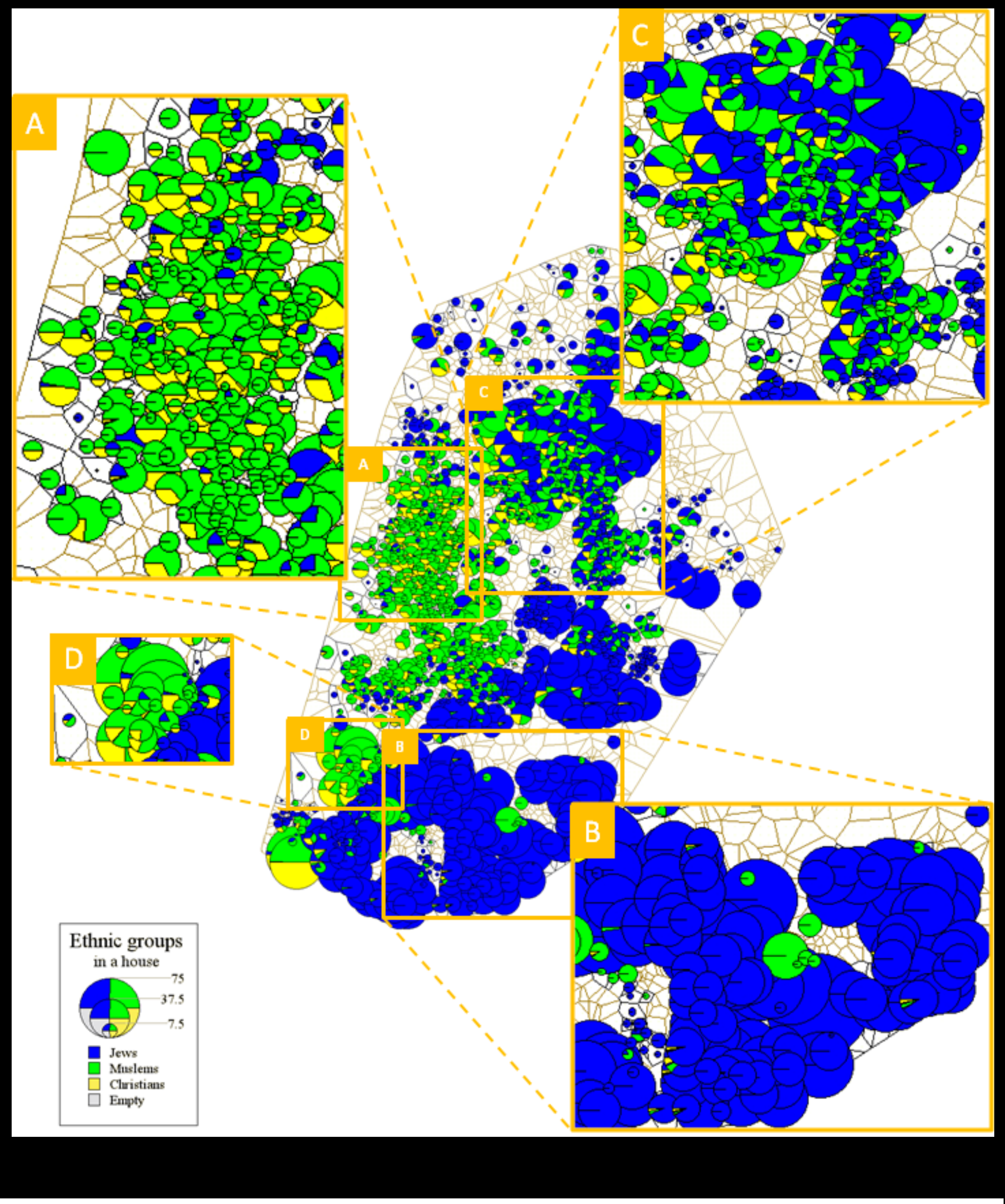


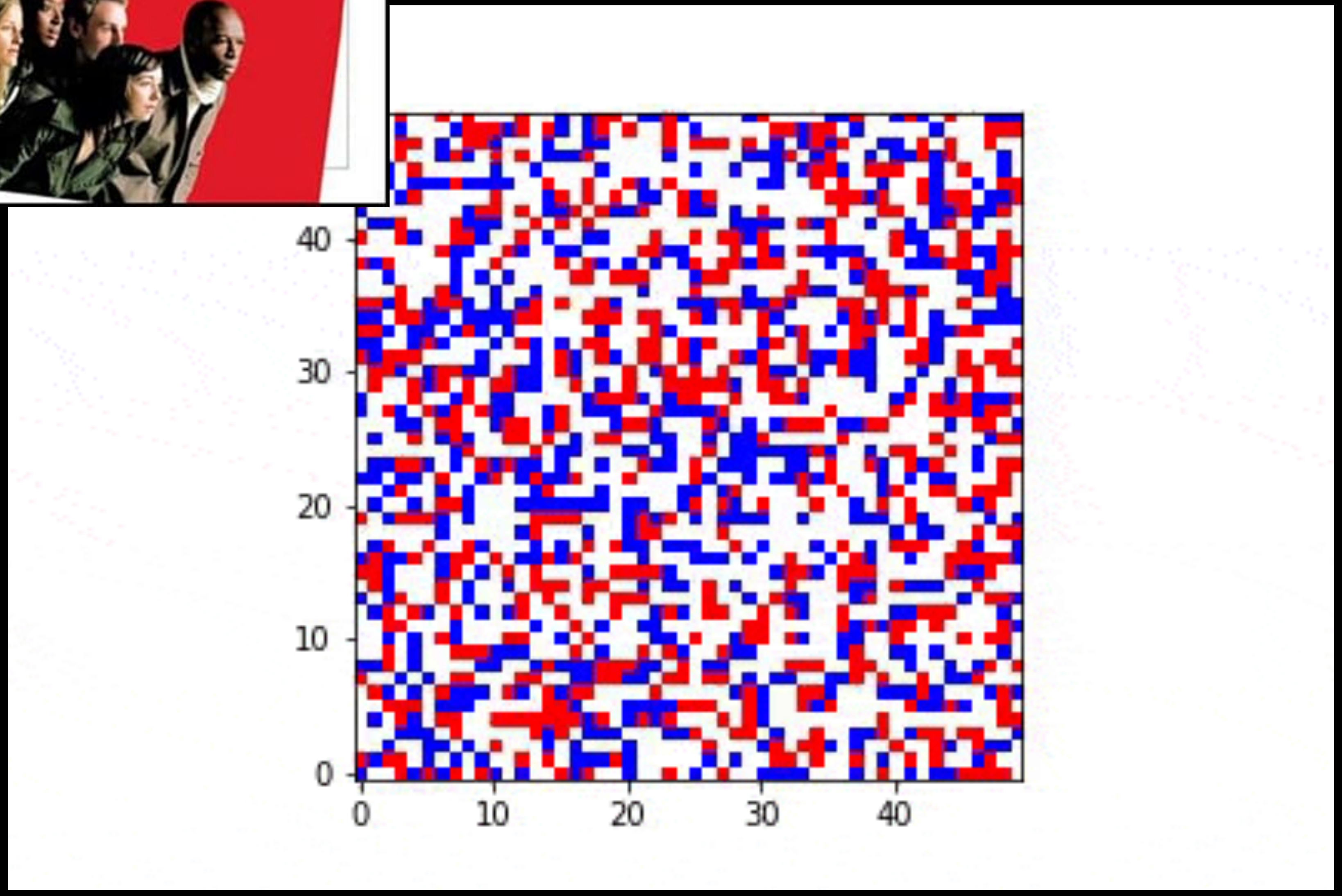
Includes the NOBEL PRIZE ACCEPTANCE LECTURE

MICROMOTIVES AND MACROBEHAVIOR

THOMAS C. SCHELLING

"Before *Freakonomics* and *The Tipping Point*, there was *Micromotives and Macrobehavior*." —BARRY NALEBUFF, coauthor of *Thinking Strategically*





T. C. Schelling, Dynamic models of segregation. Journal of Mathematical Sociology 1, 143-186 (1971).

# How might we characterize traditional simulations?

*Strength:*

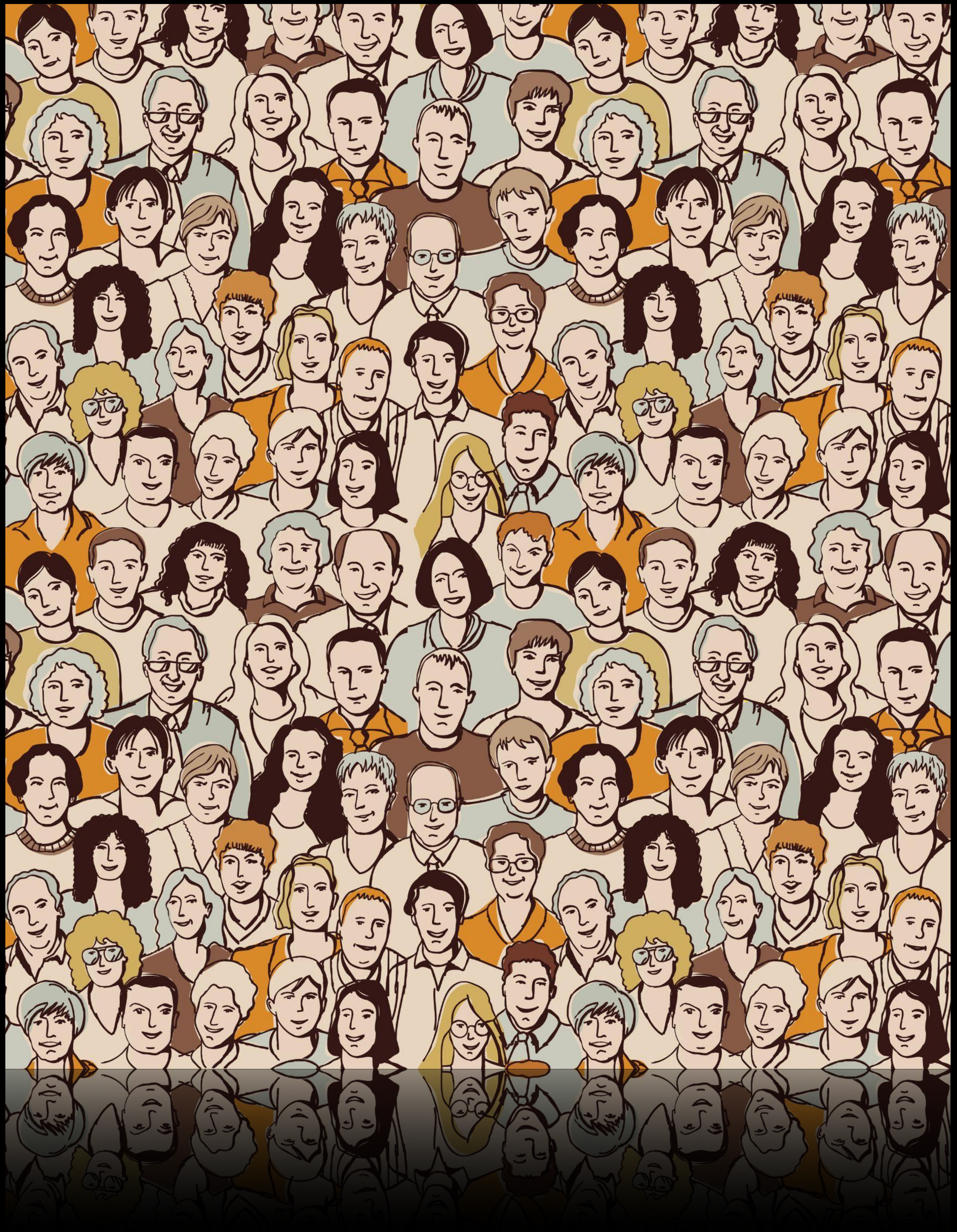
Simple and interpretatble

*Weakness:*

Oversimplifies human contingencies

# Act 2: Present

## Simulations with generative agents



Large language models can be **prompted** to generate human behavior conditioned on a variety of experiences.

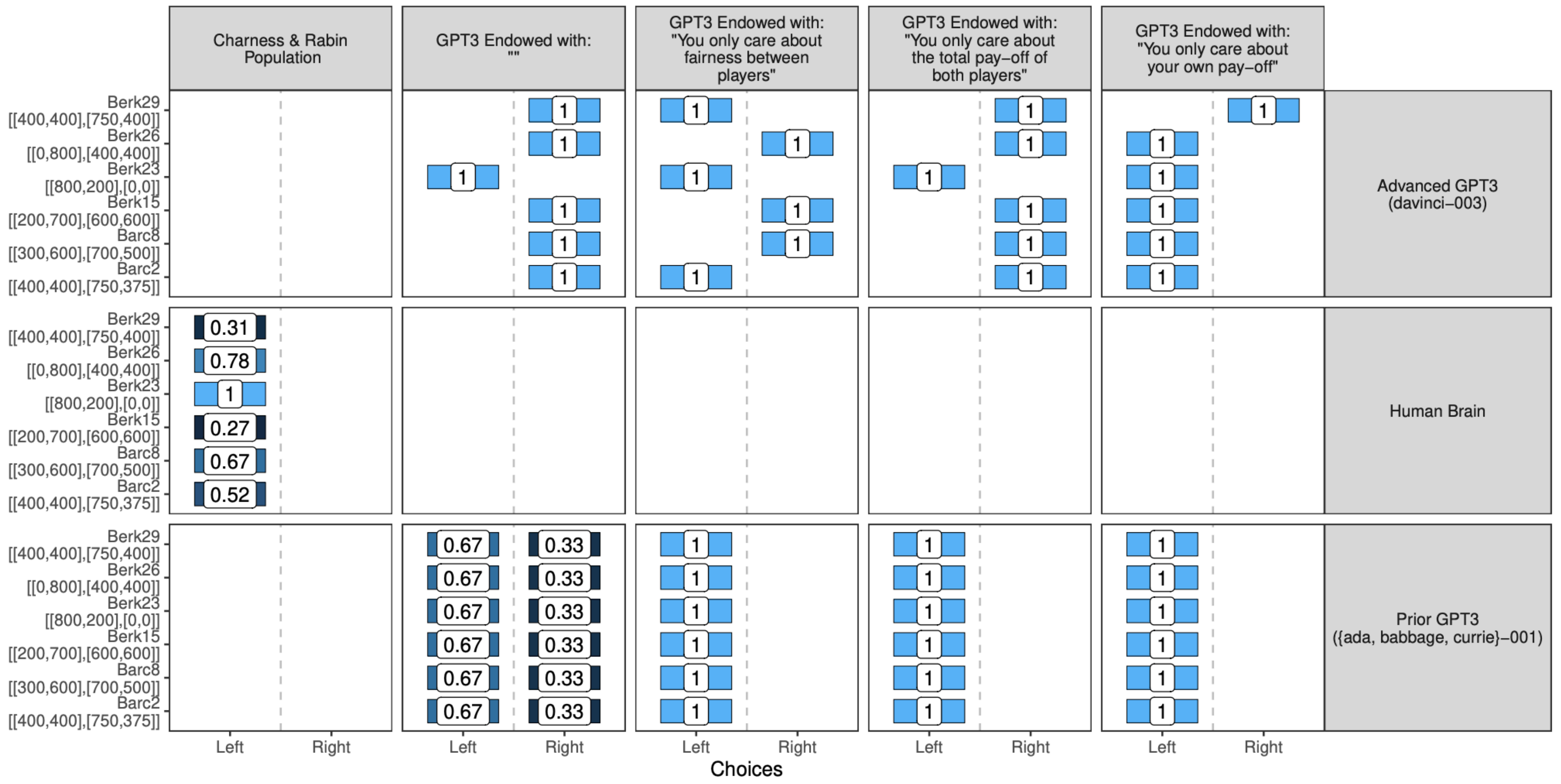
`“[name] is a [description]”`

**Social Simulacra (UIST '22)**

Joon Sung Park, Lindsay Popowski, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Social Simulacra: Creating Populated Prototypes for Social Computing Systems. UIST 2022.

# Replicating surveys and experiments

Figure 1: Charness and Rabin (2002) Simple Tests choices by model type and endowed "personality"



*Notes: This shows the faction of AI subjects choosing each option, by framing.*

J. J. Horton, "Large language models as simulated economic agents: What can we learn from homo silicus?" (2023).

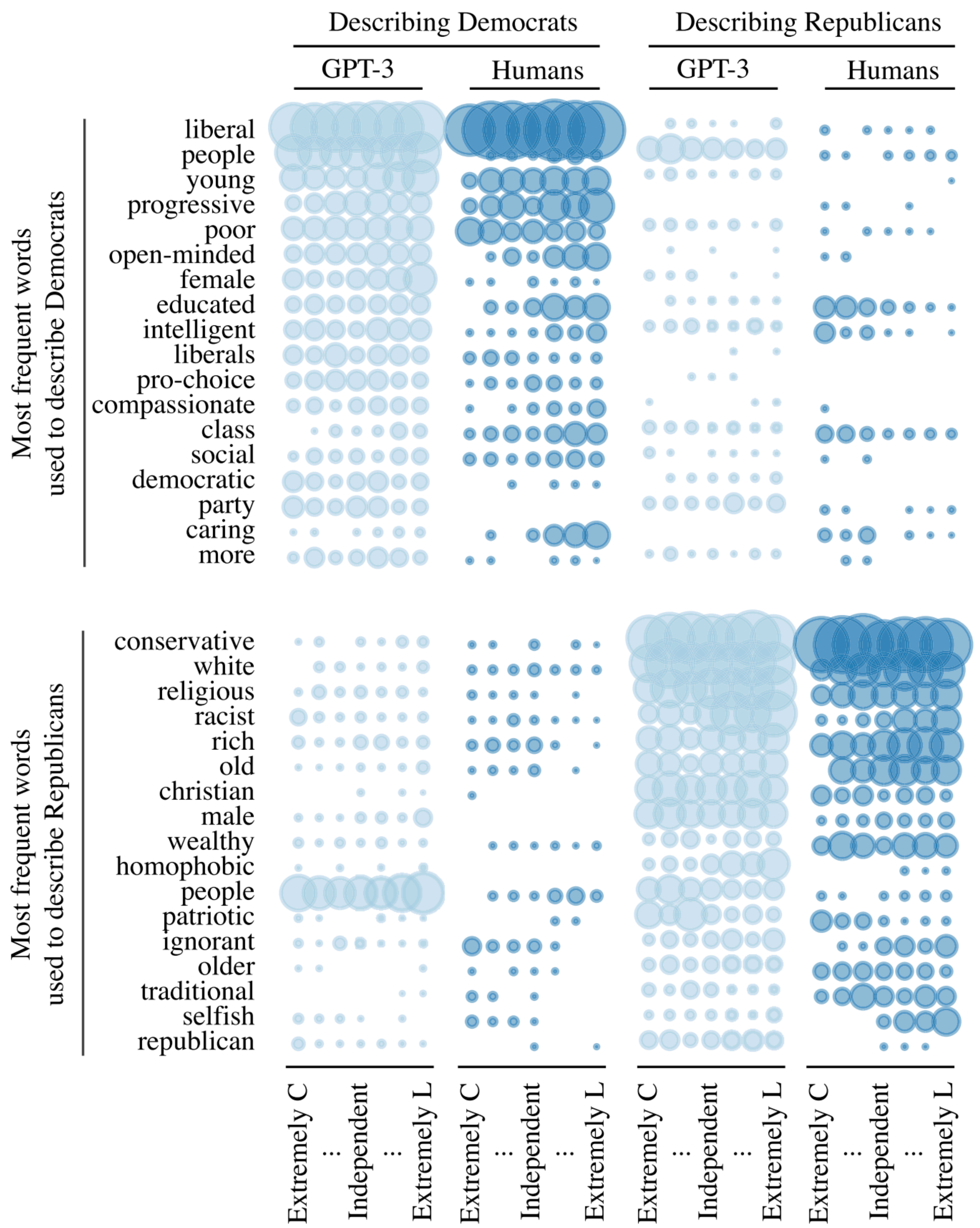
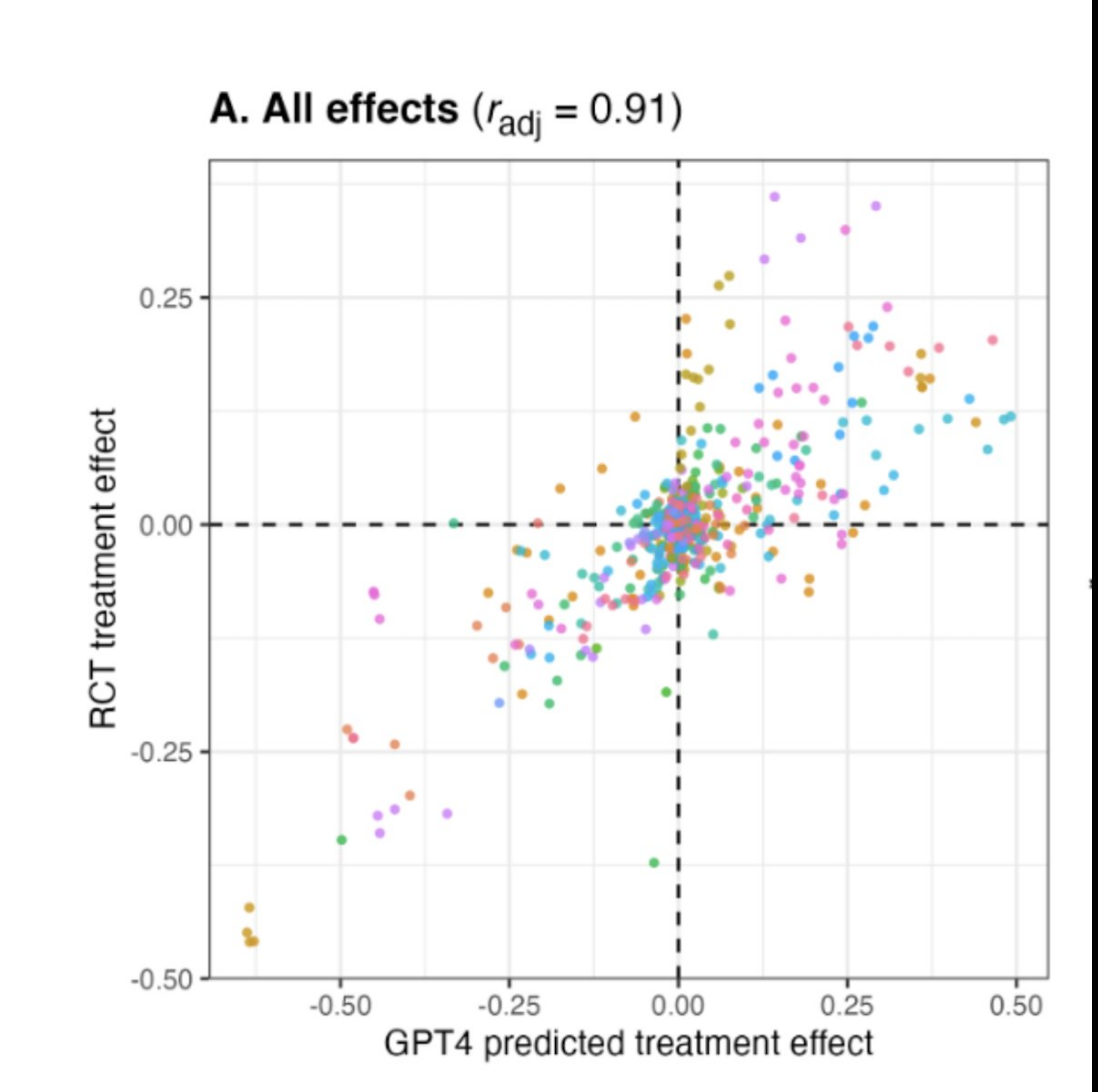


**Figure 2.** The original Pigeonholing Partisans dataset and the corresponding GPT-3-generated words. Bubble size represents relative frequency of word occurrence; columns represent the ideology of list writers. GPT-3 uses a similar set of words to humans.

L. P. Argyle et al., Out of one, many: Using language models to simulate human samples. Political Analysis 31, 337-355 (2023).

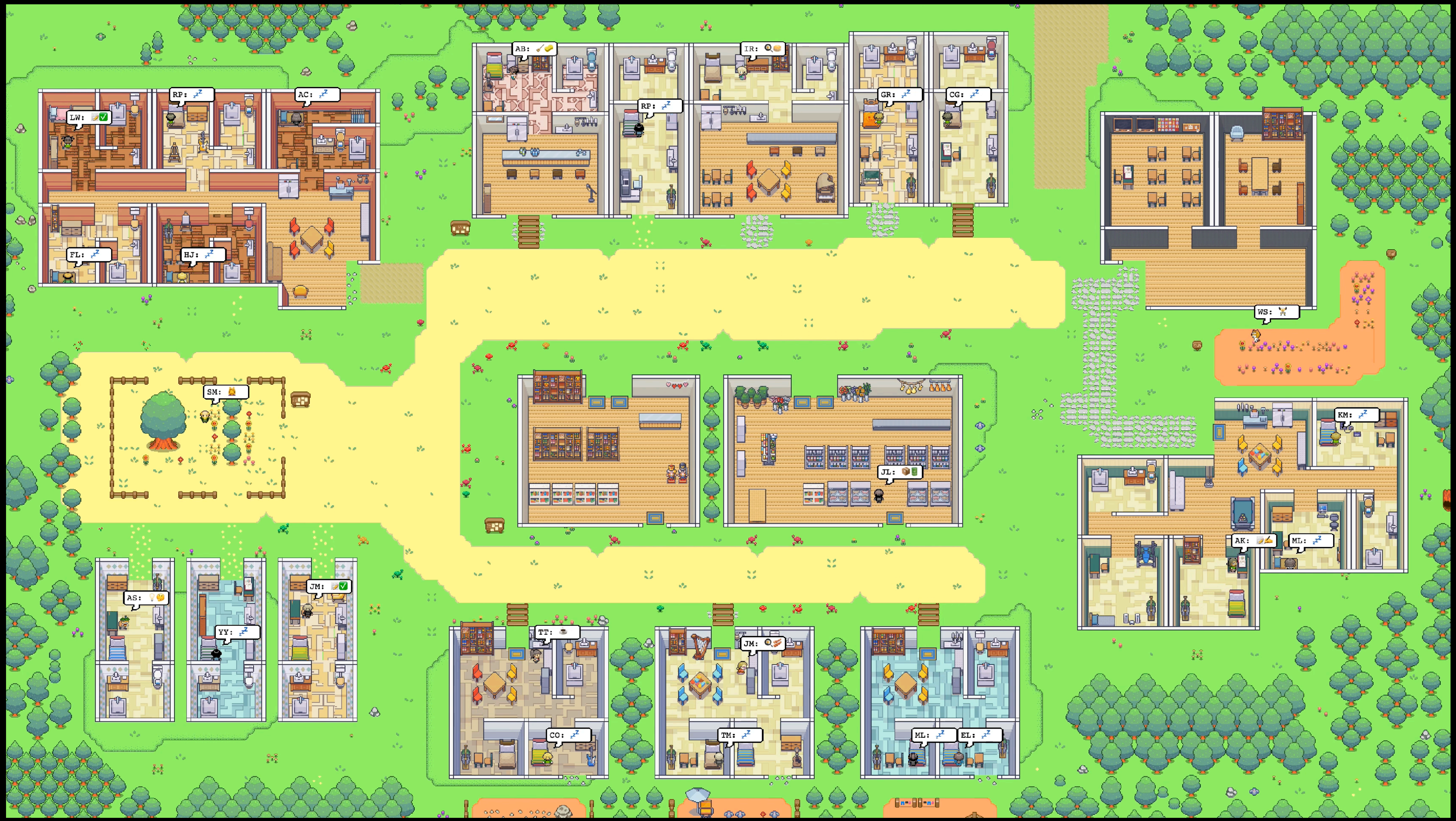
# Replicating treatment effects



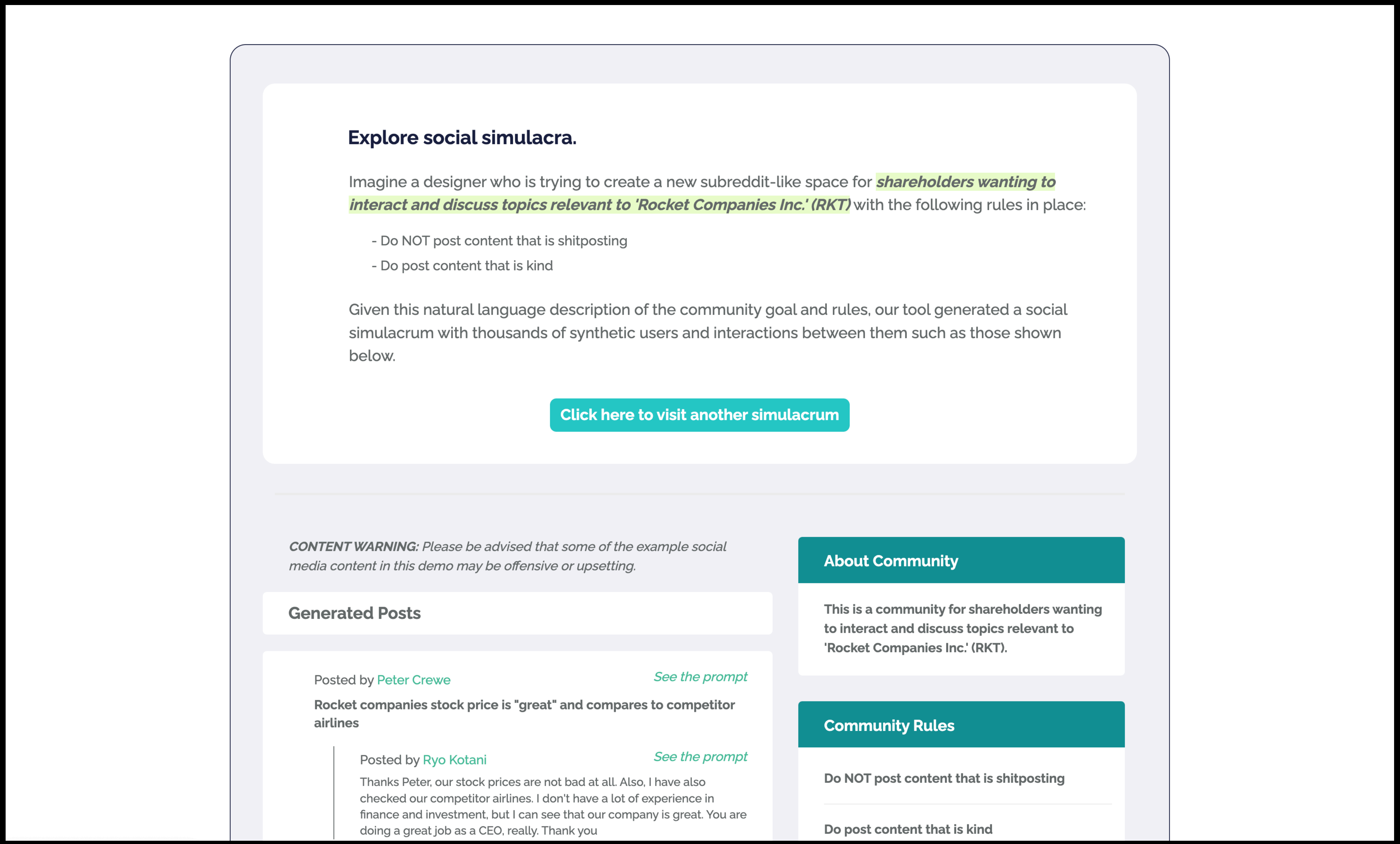
A. Ashokkumar, L. Hewitt, I. Ghezae, R. Willer, "Predicting Results of Social Science Experiments Using Large Language Models" (2024).

# Generative agents and social simulacra

https://social-simulacra.herokuapp.com/
https://reverie.herokuapp.com/arXiv_Demo/



J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, Generative agents: Interactive simulacra of human behavior, in Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (ACM, 2023).



**Explore social simulacra.**

Imagine a designer who is trying to create a new subreddit-like space for ***shareholders wanting to interact and discuss topics relevant to 'Rocket Companies Inc.' (RKT)*** with the following rules in place:

- Do NOT post content that is shitposting
- Do post content that is kind

Given this natural language description of the community goal and rules, our tool generated a social simulacrum with thousands of synthetic users and interactions between them such as those shown below.

***CONTENT WARNING:*** *Please be advised that some of the example social media content in this demo may be offensive or upsetting.*

**Generated Posts**

Posted by Peter Crewe *See the prompt*

**Rocket companies stock price is "great" and compares to competitor airlines**

Posted by Ryo Kotani *See the prompt*

Thanks Peter, our stock prices are not bad at all. Also, I have also checked our competitor airlines. I don't have a lot of experience in finance and investment, but I can see that our company is great. You are doing a great job as a CEO, really. Thank you

**About Community**

This is a community for shareholders wanting to interact and discuss topics relevant to 'Rocket Companies Inc.' (RKT).

**Community Rules**

Do NOT post content that is shitposting

Do post content that is kind

J. S. Park, L. Popowski, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, Social simulacra: Creating Populated Prototypes for Social Computing Systems, in Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (ACM, 2022).

# How might we characterize generative agent-based models?

*Strength:*

Open-ended and captures the nuance

*Weakness:*

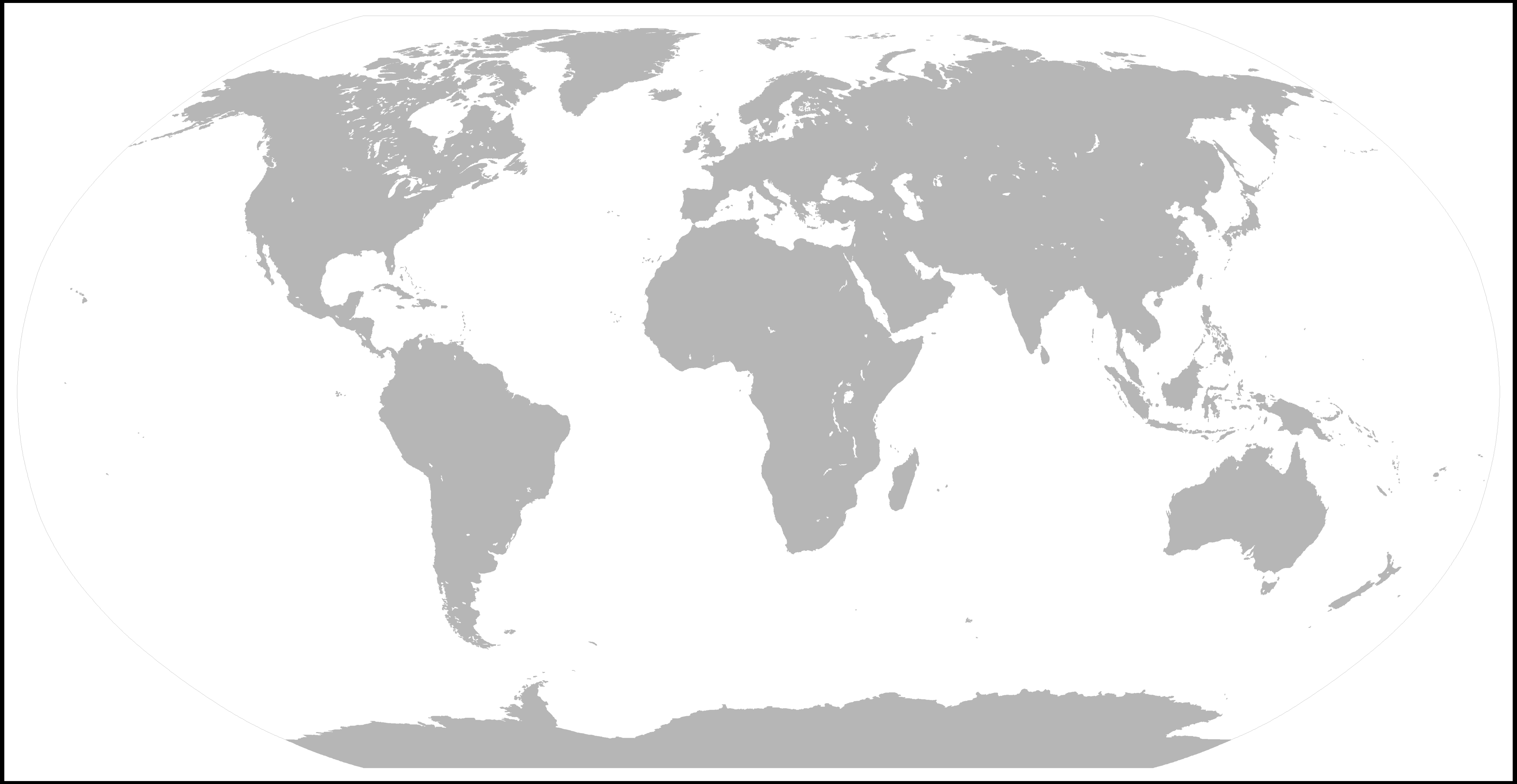
Complex

# Act 3: Future

## Towards an accurate simulation of our world

What do *you* think they ought be?

# One vision: a world simulator of 8 billions.



# References

- SK Card, TP Moran, and A Newell. 1983. The psychology of human-computer interaction. (1983).
- Mark Weiser. 1999. The Computer for the 21st Century. SIGMOBILE Mob. Comput. Commun. Rev. 3, 3 (jul 1999).
- Allen Newell. 1990. Unified Theories of Cognition. Harvard University Press, Cambridge, Massachusetts.
- H. W. J. Rittel, M. M. Webber, Dilemmas in a general theory of planning. Policy Sciences 4, 155–169 (1973).
- Social Simulacra: Creating Populated Prototypes for Social Computing Systems. UIST 2022.
- Gordon, R. M. (1986). Folk psychology as simulation. Mind & Language, 1(2), 158-171.
- J. von Neumann, Theory of Self-Reproducing Automata, A. W. Burks, Ed. (University of Illinois Press, 1966).
- S. Wolfram, A New Kind of Science (Wolfram Media, 2002).

# References

- J. von Neumann, O. Morgenstern, Theory of Games and Economic Behavior (Princeton University Press, 1944).
- T. C. Schelling, Dynamic models of segregation. Journal of Mathematical Sociology 1, 143-186 (1971).
- J. J. Horton, "Large language models as simulated economic agents: What can we learn from homo silicus?" (2023).
- L. P. Argyle et al., Out of one, many: Using language models to simulate human samples. Political Analysis 31, 337-355 (2023).
- A. Ashokkumar, L. Hewitt, I. Ghezae, R. Willer, "Predicting Results of Social Science Experiments Using Large Language Models" (2024).
- J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, Generative agents: Interactive simulacra of human behavior, in Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (ACM, 2023).

# **CS 222:** AI Agents and Simulations

**Stanford University**

Joon Sung Park