Against theory-motivated experimentation

Marina Dubova

About me

BS in **Psychology**
Saint Petersburg University

PhD student in **Cognitive Science**
Indiana University

Visiting PhD student in **Philosophy & Formal Epistemology**
Carnegie Mellon University

Arseny Moskvichev
UCI & Santa Fe Institute

Kevin Zollman
CMU

Outline

1. Motivation: how to choose the best next experiment?
2. Multi-agent model of learning through collecting and explaining the data
3. Evaluation of experimentation strategies:
   a. Across contexts
   b. Across time
   c. With prior knowledge
4. Brief attempt to convince you that the results make sense
5. Discussion time!

Motivation

How to design the best next experiment?

How to design the best next experiment?

1. Falsification

How to design the best next experiment?

1. Falsification
2. Theoretical disagreement (conducting crucial experiments)

How to design the best next experiment?

1. Falsification
2. Theoretical disagreement (conducting crucial experiments)
3. Verification

How to design the best next experiment?

1. Falsification
2. Theoretical disagreement (conducting crucial experiments)
3. Verification
4. __insert your favourite heuristic__

How to design the best next experiment?

1. Falsification
2. Theoretical disagreement (conducting crucial experiments)
3. Verification
4. __insert your favourite heuristic__

theory-motivated

No model to study this question

No model to study this question

The current models:

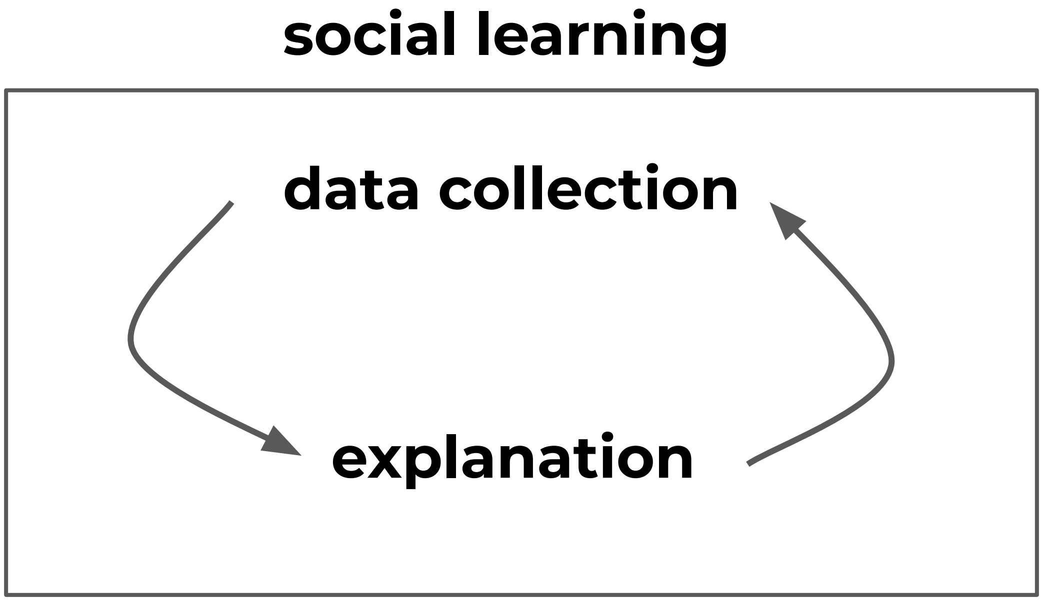
1. Are designed to test the efficiency of a particular scientific practice (e.g. replication studies: Smaldino & McElreath, 2016)

No model to study this question

The current models:

1. Are designed to test the efficiency of a particular scientific practice (e.g. replication studies: Smaldino & McElreath, 2016)
2. Do not formalize active data collection, explanation, and social learning processes at the same time (e.g. the world provides all observations or the theory building is minimized, as in Zollman, 2007; Smaldino & McElreath, 2016)

We designed a new model



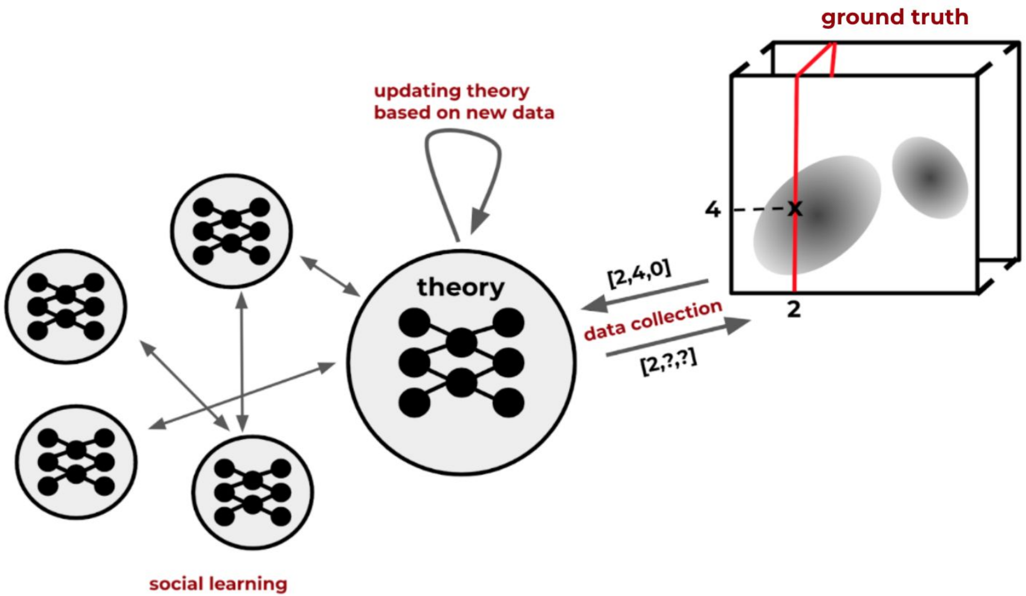
The model

Main principles behind the model

Minimal formalization of the essential scientific processes (data collection, explanation, social learning)

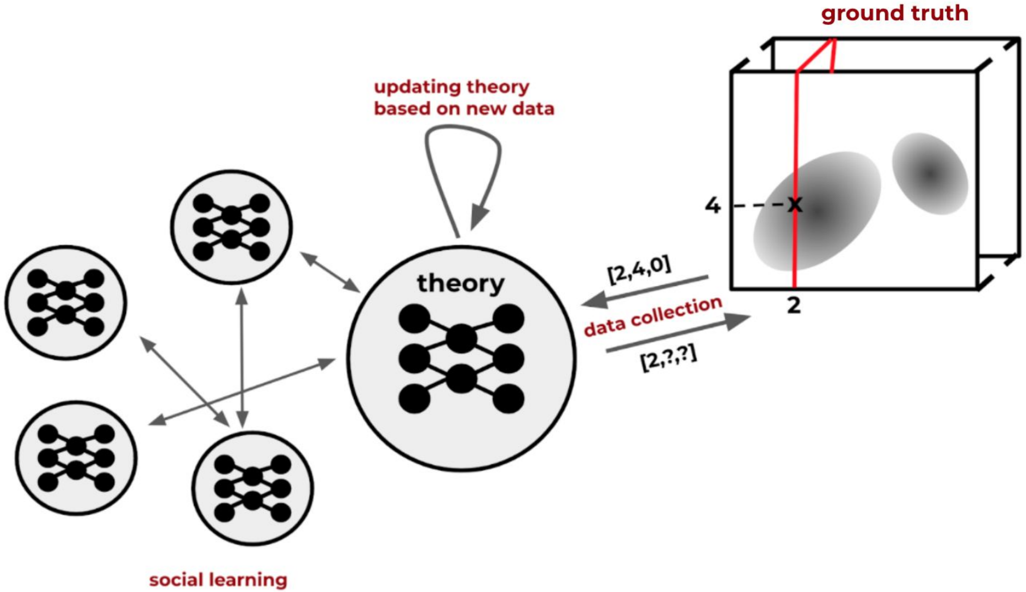
Main principles behind the model

Minimal formalization of the essential scientific processes (data collection, explanation, social learning)



Main principles behind the model

Minimal formalization of the essential scientific processes (data collection, explanation, social learning)



Main principles behind the model

Minimal formalization of the essential scientific processes (data collection, explanation, social learning)

Possibility to widely vary the learning context

Main principles behind the model

Minimal formalization of the essential scientific processes (data collection, explanation, social learning)

Possibility to widely vary the learning context

“Ideal world”

The model

The model

- Agent-based: agents (“scientists”) use particular experimentation strategies to study the “ground truth”

The model

- Agent-based: agents (“scientists”) use particular experimentation strategies to study the “ground truth”
- Agents acquire evidence from the “ground truth”

The model

- Agent-based: agents ("scientists") use particular experimentation strategies to study the "ground truth"
- Agents acquire evidence from the "ground truth"
- Agents build explanations for their evidence

The model

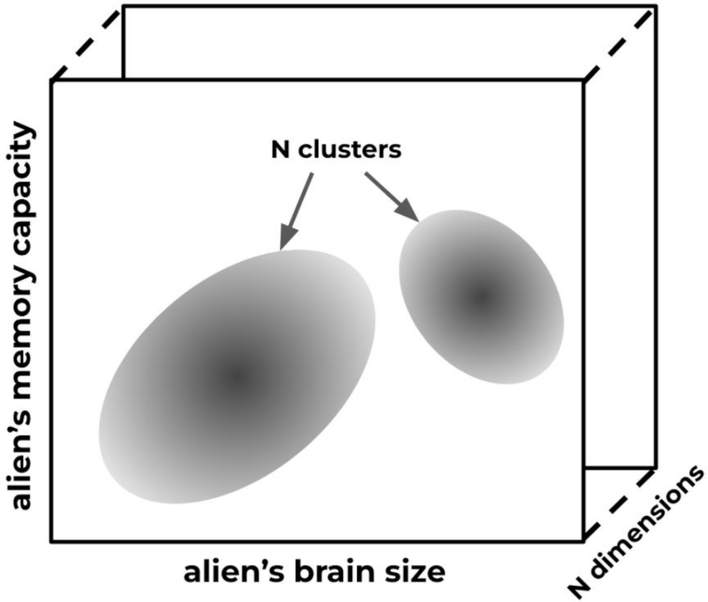
- Agent-based: agents ("scientists") use particular experimentation strategies to study the "ground truth"
- Agents acquire evidence from the "ground truth"
- Agents build explanations for their evidence

Collecting and explaining the data

- The “ground truth” is a high-dimensional space of potential observations distributed in a particular way

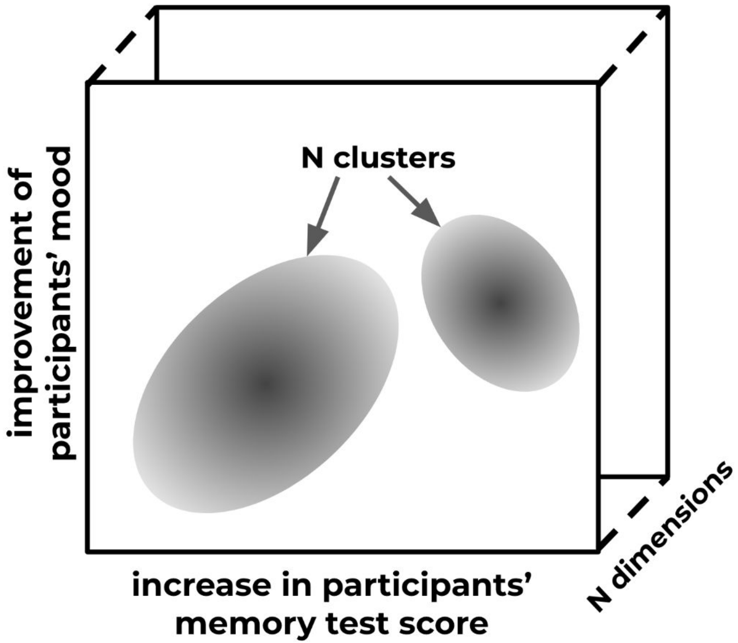
Collecting and explaining the data

- The “ground truth” is a high-dimensional space of potential observations distributed in a particular way



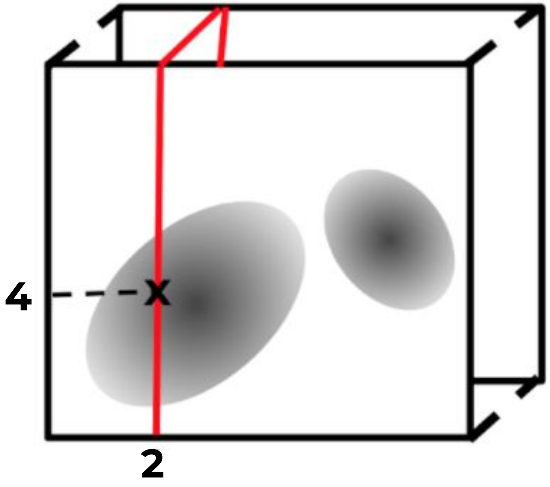
Collecting and explaining the data

- The “ground truth” is a high-dimensional space of potential observations distributed in a particular way



Collecting and explaining the data

- The “ground truth” is a high-dimensional space of potential observations distributed in a particular way
- At each timestep an agent from the group samples a point from the ground truth space and records what it sees

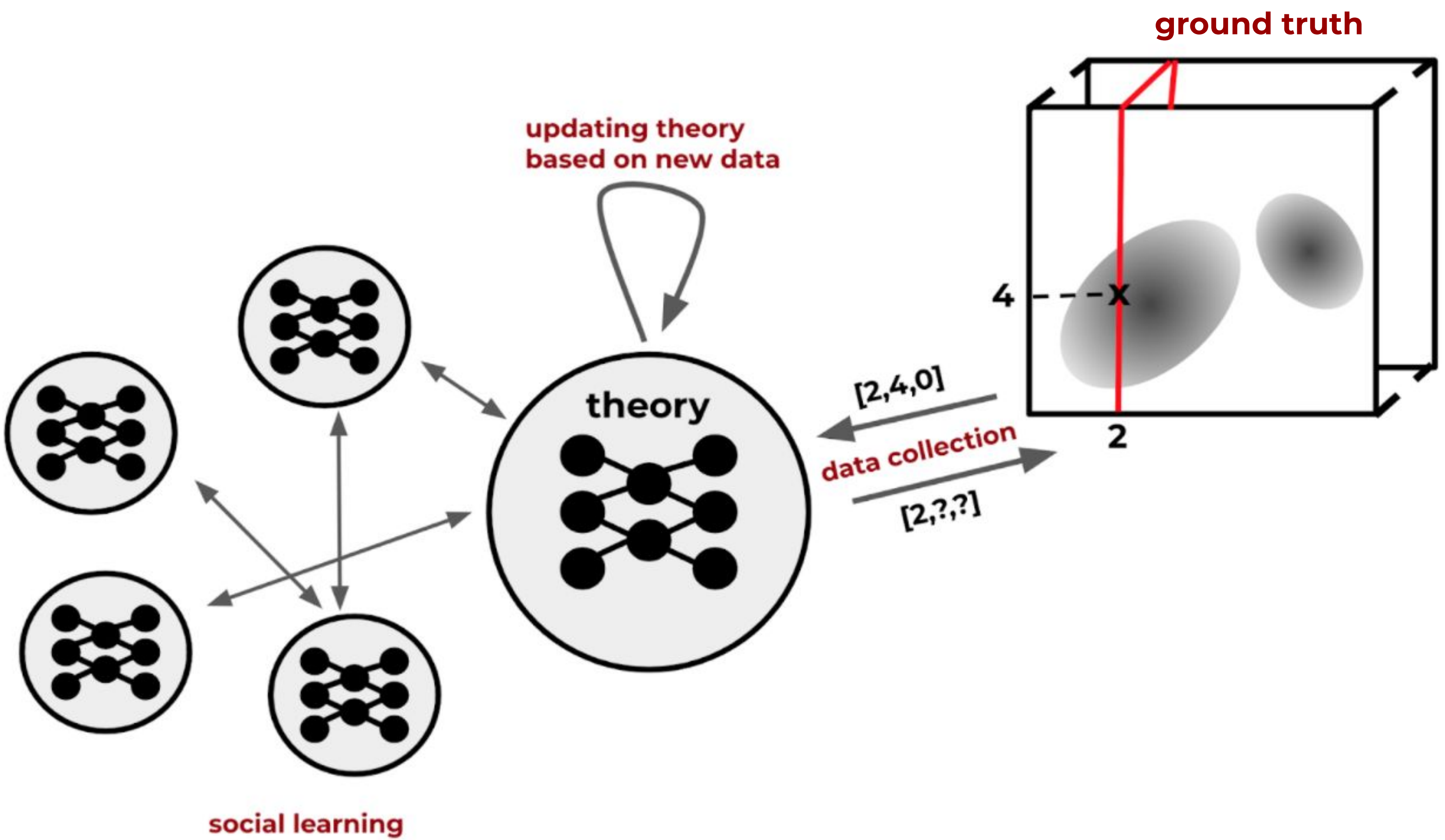


Collecting and explaining the data

- The "ground truth" is a high-dimensional space of potential observations distributed in a particular way
- At each timestep an agent from the group samples a point from the ground truth space and records what it sees
- The agents construct lower-dimensional representations (theories) to account for the evidence they collected

Collecting and explaining the data

- The “ground truth” is a high-dimensional space of potential observations distributed in a particular way
- At each timestep an agent from the group samples a point from the ground truth space and records what it sees
- The agents construct lower-dimensional representations (theories) to account for the evidence they collected
- The goal of each agent is to find a lower-dimensional representation (theory) that captures as much information about the higher-dimensional “ground truth” as possible


ground truth
updating theory
based on new data
theory
4
2
[2,4,0]
data collection
[2,?,?]
social learning

The study

Main questions

Which experimentation strategies work better (in general/in particular contexts)?

How much should a scientist’s theoretical framework influence their experimentation?

SPOILERS

Which experimentation strategies work better (in general/in particular contexts)? **Random data collection works best in all contexts we tested:)**

How much should a scientist's theoretical framework influence their experimentation? **As minimally as possible:)**

SPOILERS

Which experimentation strategies work better (in general/in particular contexts)? **Random data collection works best in all contexts we tested:)**

How much should a scientist's theoretical framework influence their experimentation? **As minimally as possible:)**

Agents’ experimentation strategies

Confirmation

Agents’ experimentation strategies

Confirmation

Falsification

Agents’ experimentation strategies

Confirmation

Falsification

Disagreement (conducting crucial experiments)

Agents’ experimentation strategies

Confirmation

Falsification

Disagreement (conducting crucial experiments)

Novelty (space-filling)

Agents' experimentation strategies

Confirmation

Falsification

Disagreement (conducting crucial experiments)

Novelty (space-filling)

Random

Agents’ experimentation strategies

Confirmation

Falsification

Disagreement (conducting crucial experiments)

Novelty (space-filling)

Random

+ hybrid strategies

Social learning strategies

1. Data sharing
2. Feature sharing
3. Explanation sharing
4. Aligned explanation sharing
5. Skeptical aligned explanation sharing
6. Teaching and learning

\+ their combinations

Group size = [5, 10]

Ground truth

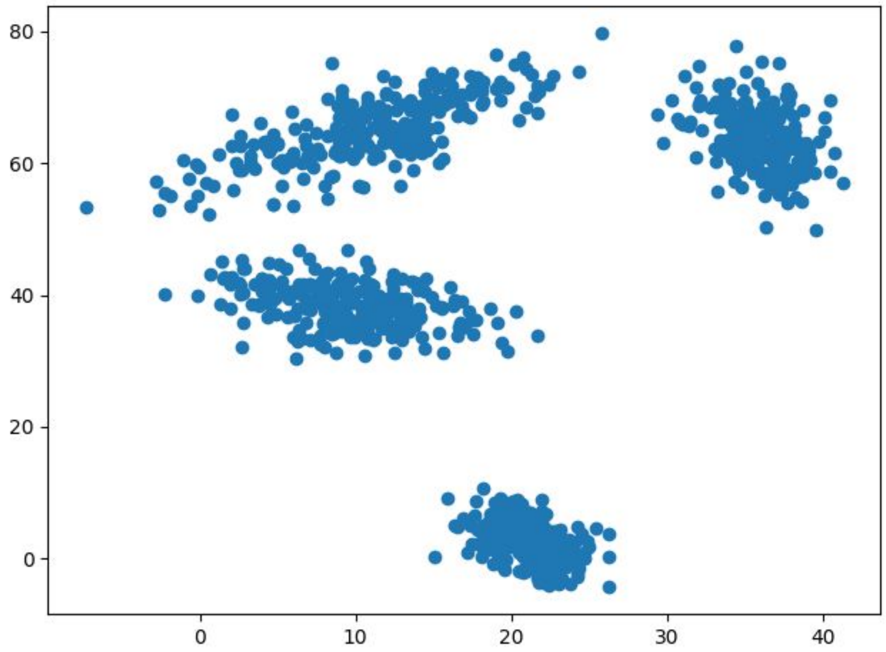
Mixture multivariate gaussian distribution

N of dimensions: [20, 100]

N of clusters: [2, 10, 30]

Agents' measurement capability:

[all dimensions, half of the dimensions]

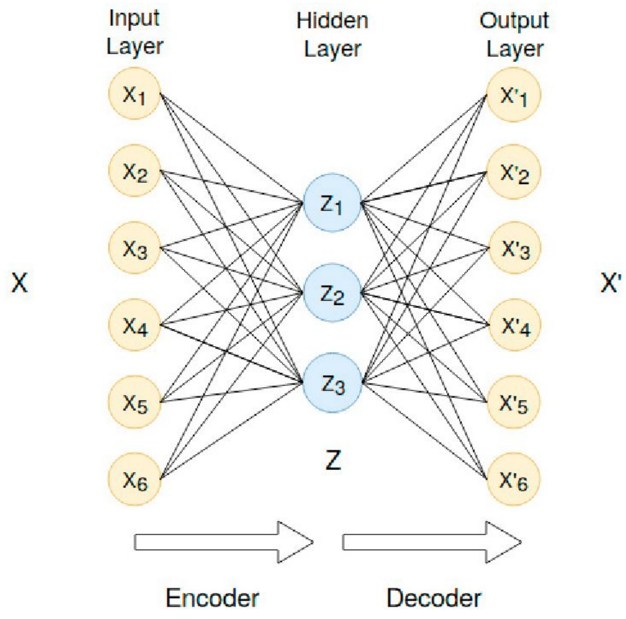


Agents' explanation strategy: NN embedding

Each agent develops a lower-dimensional representation of the ground truth

Agents' theories are simple NN autoencoders with one hidden layer

N of internal neurons = [3,6]



Evaluating the agents: reconstruction error

Evaluating the agents: reconstruction error

“Subjective” performance

How well the agents’ explanations fit the observations they have collected

Evaluating the agents: reconstruction error

"Subjective" performance

How well the agents' explanations fit the observations they have collected

"Objective" performance

How well the agents' explanations fit the observations sampled from the full ground truth distribution

Evaluating the agents: reconstruction error

“Subjective” performance

How well the agents’ explanations fit the observations they have collected

“in-sample”

“Objective” performance

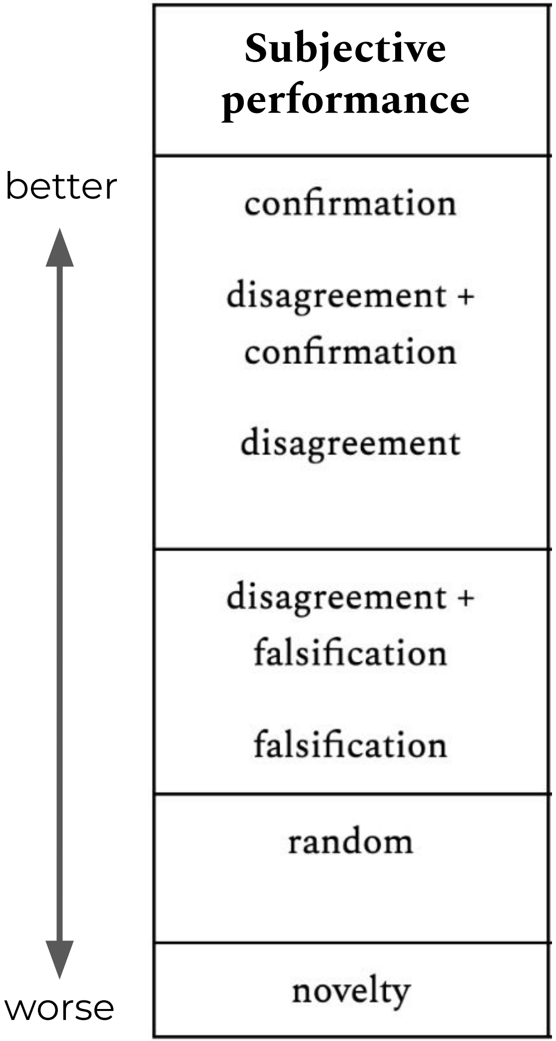
How well the agents’ explanations fit the observations sampled from the full ground truth distribution

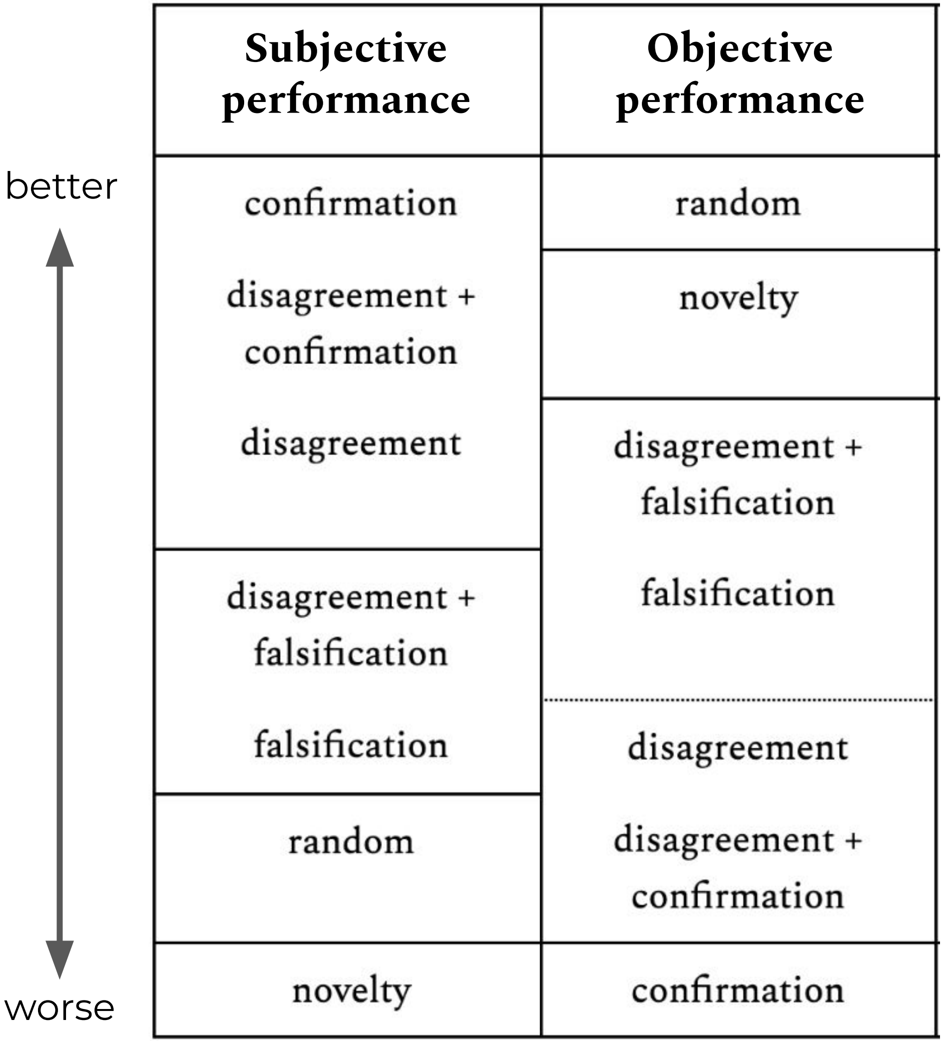
“out-of-sample”

Results

Experiment 1

1. Varying all context conditions (~4 samples per condition; 11372 simulations in total)
2. Looking at the learning results of the agents following different experimentation strategies at the end of the simulation (= after the group collects 300 observations)


Subjective performance
better
confirmation
disagreement + confirmation
disagreement
disagreement + falsification
falsification
random
novelty
worse



better
worse
Subjective performance
confirmation
disagreement + confirmation
disagreement
disagreement + falsification
falsification
random
novelty
Objective performance
random
novelty
disagreement + falsification
falsification
disagreement
disagreement + confirmation
confirmation

Experiment 1: summary of the results

1. Random experimentation helps agents to develop the best theories about the ground truth

Experiment 1: summary of the results

1. Random experimentation helps agents to develop the best theories about the ground truth
2. The agents following confirmation-, falsification-, disagreement-, novelty-based experimentation strategies develop inferior theories

Experiment 1: summary of the results

1. Random experimentation helps agents to develop the best theories about the ground truth
2. The agents following confirmation-, falsification-, disagreement-, novelty-based experimentation strategies develop inferior theories
3. If evaluated against the data available to scientists, the theory-motivated experimentation appears to produce more successful theories

Experiment 1: summary of the results

1. Random experimentation helps agents to develop the best theories about the ground truth
2. The agents following confirmation-, falsification-, disagreement-, novelty-based experimentation strategies develop inferior theories
3. If evaluated against the data available to scientists, the theory-motivated experimentation appears to produce more successful theories

an artifact of the simulation settings?

Experiment 2: what if the agents are very limited in a number of experiments they can conduct?

Experiment 2: what if the agents are very limited in a number of experiments they can conduct?

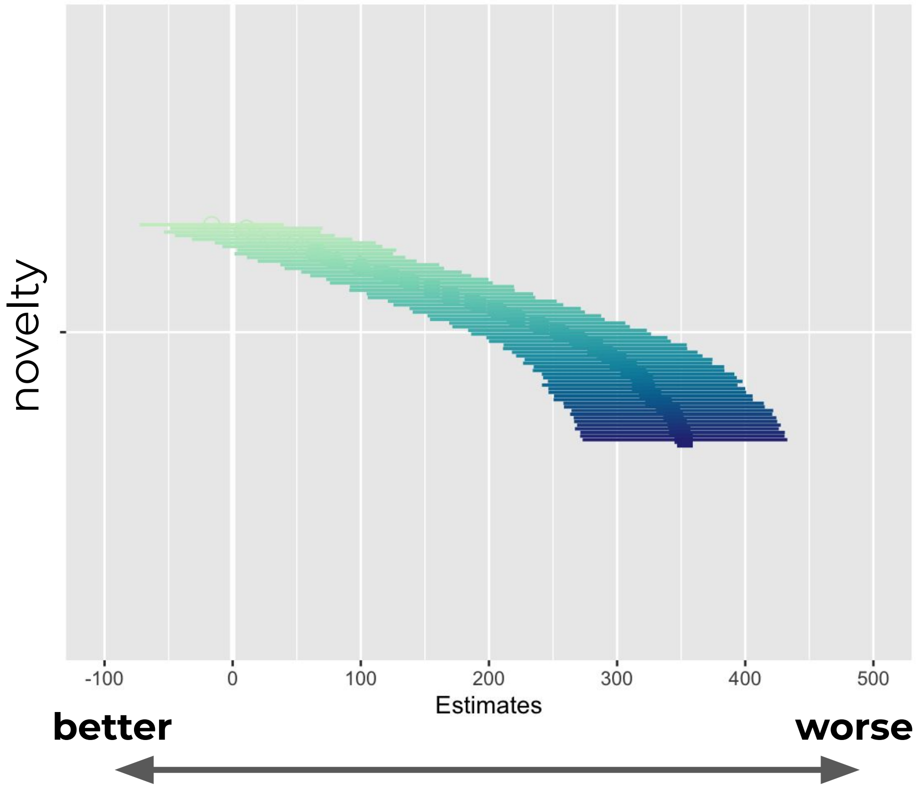
Look at the performance of the novelty-motivated & random agents at higher temporal resolution

Experiment 2: what if the agents are very limited in a number of experiments they can conduct?

Look at the performance of the novelty-motivated & random agents at higher temporal resolution

Varying all conditions as in experiment 1 (4320 simulations in total)

objective average reconstruction error



baseline: random strategy

Experiment 3: what if the agents start with better theories?

Experiment 3: what if the agents start with better theories?

Let agents have some accurate prior knowledge about the ground truth

Experiment 3: what if the agents start with better theories?

Let agents have some accurate prior knowledge about the ground truth

1. Each agent is pretrained on 10/50/100 randomly sampled observations

Experiment 3: what if the agents start with better theories?

Let agents have some accurate prior knowledge about the ground truth

1. Each agent is pretrained on 10/50/100 randomly sampled observations
2. Then, agents start learning with their target experimentation strategy (the setting & all conditions are the same as experiment 1: total of 9072 simulations)

	no pre-training (naive theories)	pre-training (informed theories)
better	random	random
	novelty	novelty
	disagreement + falsification	disagreement + falsification
	falsification	disagreement
	disagreement	disagreement + confirmation
	disagreement + confirmation	falsification
worse	confirmation	confirmation

Why don’t other strategies work?

Why don’t other strategies work?

We analyzed:

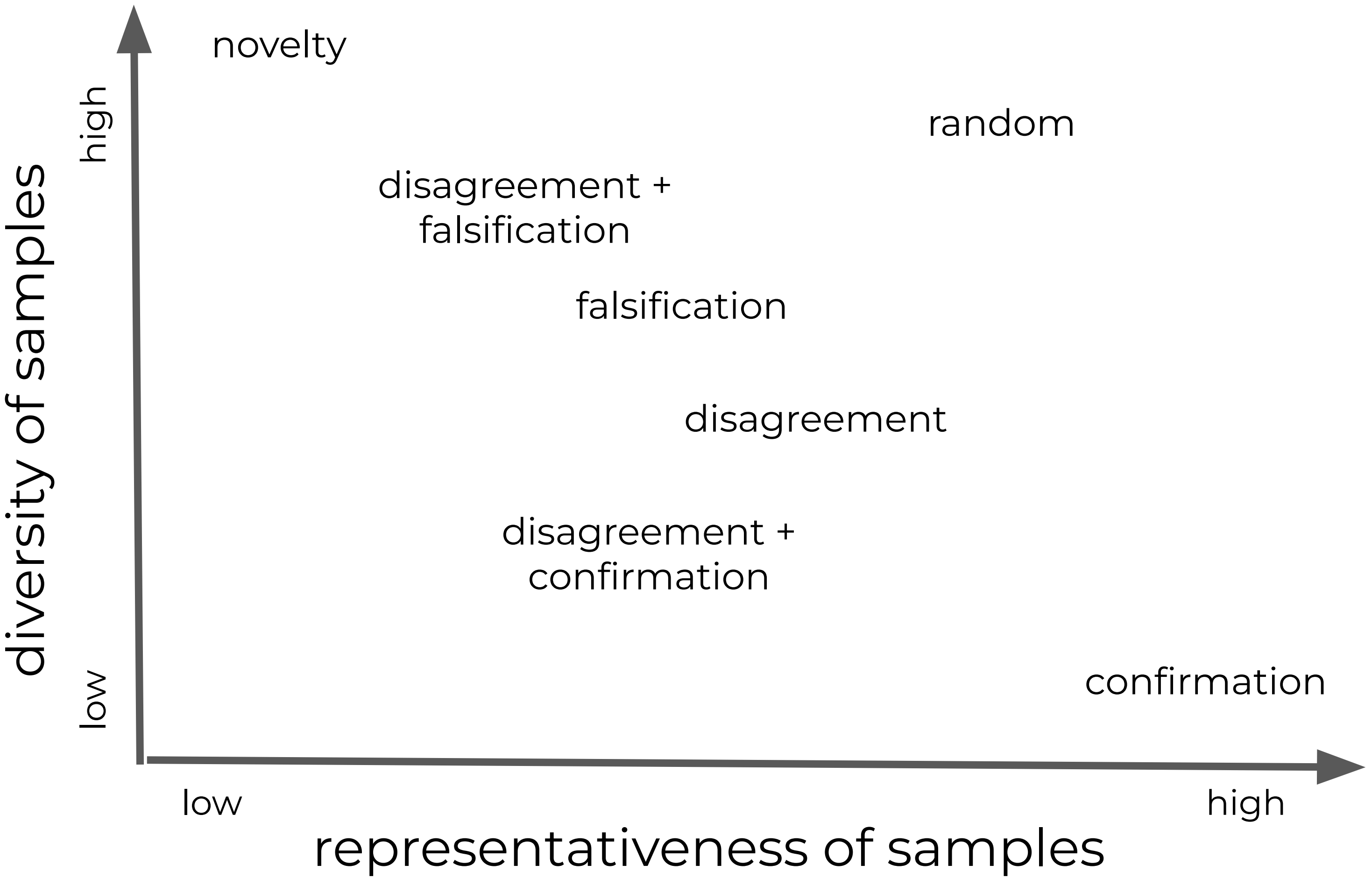
1. Heterogeneity of agents’ theories
2. Within- and between-agent diversity of samples
3. Representativeness of samples

Why don’t other strategies work?

We analyzed:

1. Heterogeneity of agents’ theories
2. Within- and between-agent diversity of samples
3. Representativeness of samples

\+ how these change over time


novelty
random
disagreement +
falsification
falsification
disagreement
disagreement +
confirmation
confirmation
diversity of samples
high
low
low
high
representativeness of samples

Results

Results

Theory-motivated experimentation creates an illusion of epistemic success by introducing a bias that prevents agents from learning about the target space of phenomena

Results

Theory-motivated experimentation creates an illusion of epistemic success by introducing a bias that prevents agents from learning about the target space of phenomena

Random experimentation enables the agents to develop the most representative accounts for the ground truth across all the studied contexts

Results

Theory-motivated experimentation creates an illusion of epistemic success by introducing a bias that prevents agents from learning about the target space of phenomena

Random experimentation enables the agents to develop the most representative accounts for the ground truth across all the studied contexts

What seems to be a good approach is not always actually a good approach

Results

Theory-motivated experimentation creates an illusion of epistemic success by introducing a bias that prevents agents from learning about the target space of phenomena

Experimentation that is uninformed by theory or previous observations (random) supports construction of the most representative accounts for the ground truth across all the studied contexts

What seems to be a good approach is not always actually a good approach

Generating more data that is well explained ≠ learning about ground truth

Results: limitations

Results are agnostic with respect to other functions that theories play in science

Results: limitations

Results are agnostic with respect to other functions that theories play in science

There are reasons to prefer theory-motivated experimentation that we did not explore

Results: limitations

Results are agnostic with respect to other functions that theories play in science

There are reasons to prefer theory-motivated experimentation that we did not explore

Perfectly random data collection is impossible

Results: limitations

Results are agnostic with respect to other functions that theories play in science

There are reasons to prefer theory-motivated experimentation that we did not explore

Perfectly random data collection is impossible

The model does not capture every kind of science

Why the results make sense

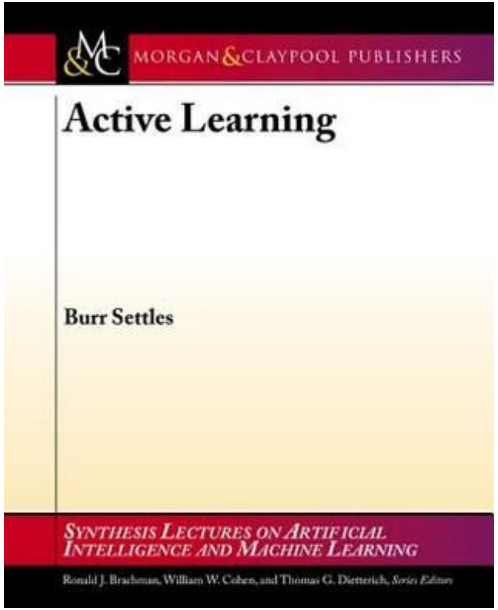
Why the results make sense

To learn something, one must have representative samples

Why the results make sense

To learn something, one must have representative samples

But....active learning!



Why the results make sense

To learn something, one must have representative samples

But....active learning!

1. The “active learning” strategies that scientists try to follow do not correspond to successful active learning strategies

Why the results make sense

To learn something, one must have representative samples

But....active learning!

1. The “active learning” strategies that scientists try to follow do not correspond to successful active learning strategies
2. The successful active learning strategies are fragile: they work only in very specific contexts (e.g. when a learner has accurate prior knowledge about the problem, etc). Otherwise, they are misleading.

A theory is not a good reason to bias experimentation!

Discussion!

1. Criticisms/concerns?
2. Scenarios where the results won't hold?
3. Strategies that might work better than random?