PARIS LODRON UNIVERSITY SALZBURG

From toy models to tactics: What user simulation is good for

Christine Bauer

Univ.-Prof. Mag. DI Dr.
Christine Bauer
Professor of
Interactive Intelligent Systems

EXDIGIT
Department of Artificial Intelligence and Human Interfaces
Faculty of Digital and Analytical Sciences

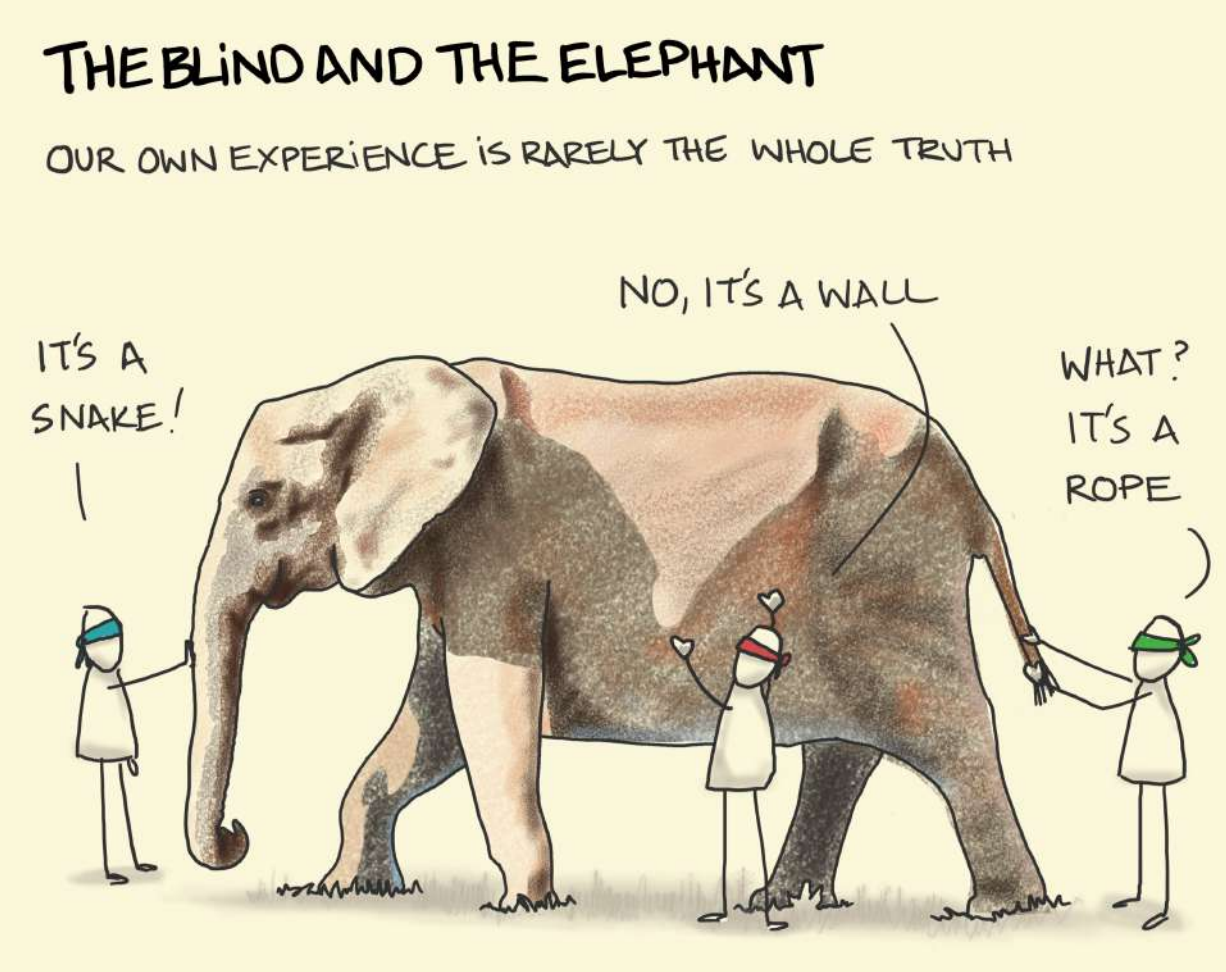
christine.bauer@plus.ac.at
https://christinebauer.eu
@christinebauer.bsky.social

17 July 2025

PARIS LODRON UNIVERSITY SALZBURG

Context

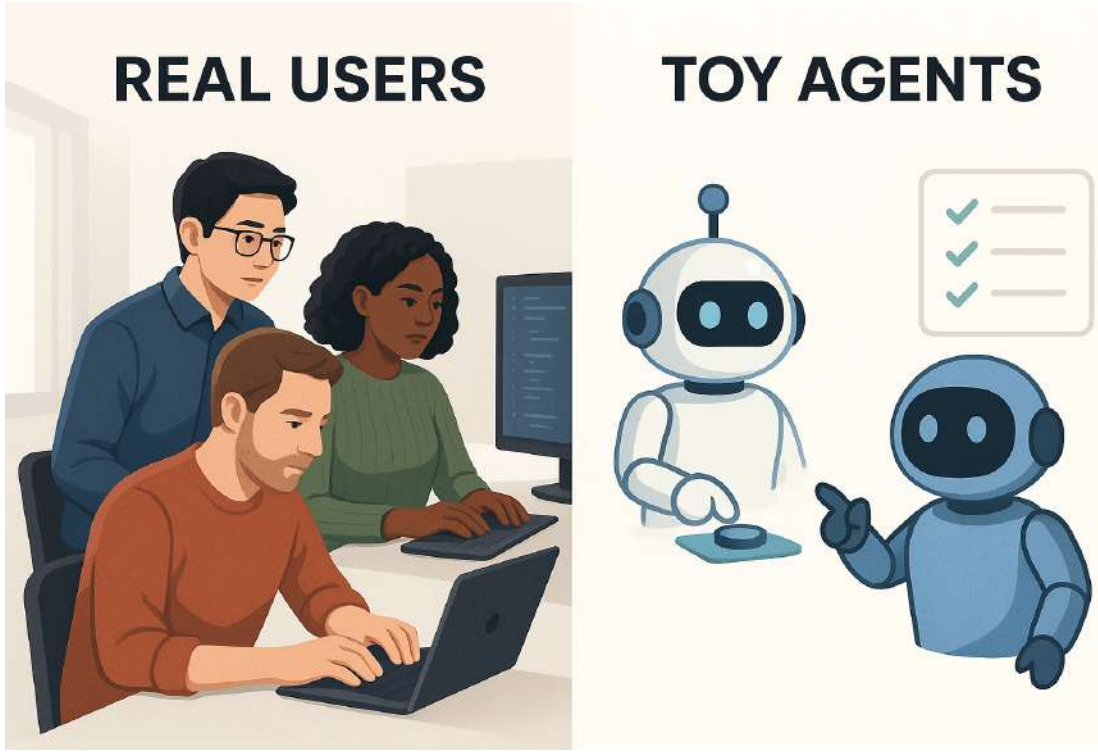
- Recommender systems
  - → I use recommender's terminology
- Interdisciplinary
  - → drawing from various disciplines
  - → also mixing terminology → I'm not perfect…
- Human-centered computing
  - → I care about humans: again and again—always!



PARIS LODRON UNIVERSITY SALZBURG

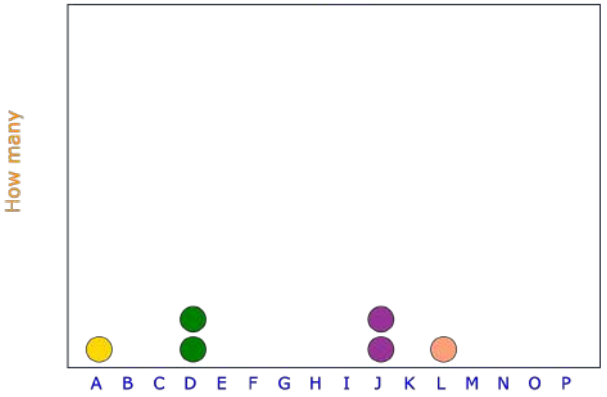
What can we learn from imagined users that real users cannot teach us?

What can we learn from real users that imagined users cannot teach us?



Motivation: Why user simulation?

~~...because we can...?!~~



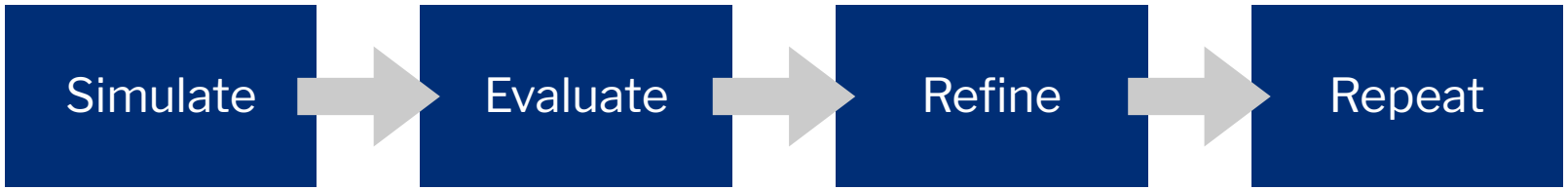
data sparsity

ethics

cost of user studies

PARIS LODRON UNIVERSITY SALZBURG

Simulation as design tool



Simulations help us design with foresight, not just afterthought.

PARIS LODRON UNIVERSITY SALZBURG

User simulation in information access

PARIS LODRON UNIVERSITY SALZBURG

Mimicking users in information retrieval research

- following *Cranfield paradigm* → mimic potential requests of users:
  - "removing actual users from the experiment but including a static user component: the ground truth"
  - simplifying assumption that a single set of judgments for a topic is representative of the user population
- mimicking user requests → essentially, this is user simulation

Cyril W. Cleverdon (1991). The significance of the Cranfield tests on index languages. Proceedings of the 14th Annual international ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '91). ACM, New York, NY, USA, 3–12. DOI: 10.1145/122860.122861

Ellen M. Voorhees (2002). The Philosophy of Information Retrieval Evaluation. In: Peters, C., Braschler, M., Gonzalo, J., Kluck, M. (eds) Evaluation of Cross-Language Information Retrieval Systems. CLEF 2001. Lecture Notes in Computer Science, vol 2406. Springer, Berlin, Heidelberg. DOI: 10.1007/3-540-45691-0_34

PARIS LODRON UNIVERSITY SALZBURG

Mimicking users in recommender systems research

- assumptions:
  - historic behavior is representative for future behavior
  - taking the average of all users
- mimicking user behavior in “offline evaluation” → essentially, this is user simulation

PARIS LODRON UNIVERSITY SALZBURG

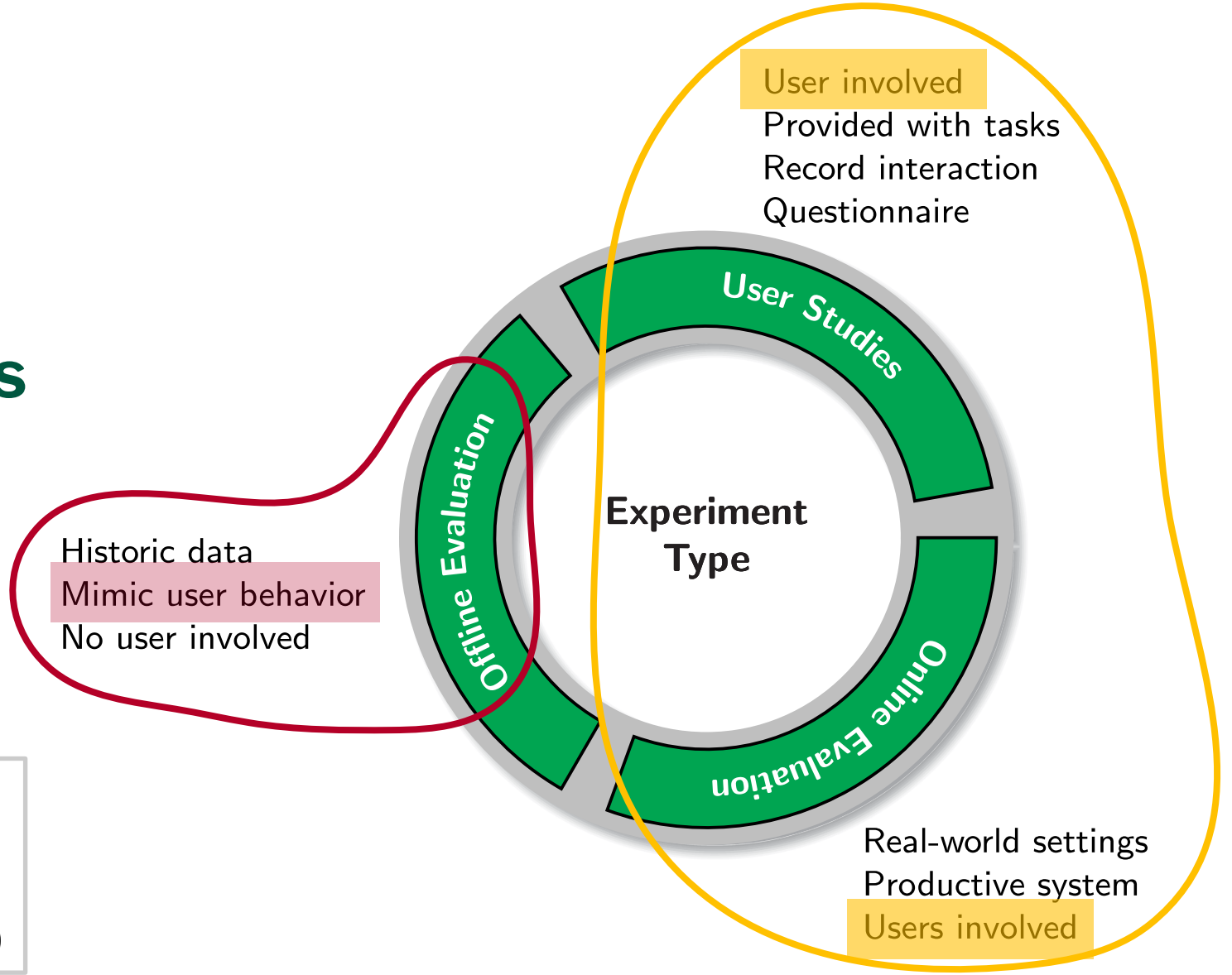
Is all what we do *user simulation* in the end?

—

No!

PARIS LODRON UNIVERSITY SALZBURG

Experiment types in recommender systems research



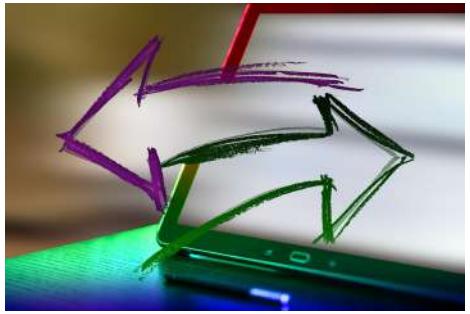
Different (sub-)communities
→different terminology
- Computational or algorithmic approaches
- User studies (in the lab or online)
- Field study/experiment (using real-world system)

Eva Zangerle & **Christine Bauer** (2022). Evaluating recommender systems: survey and framework. ACM Computing Surveys, 55(8), Art no. 170, pp 1-38. DOI: 10.1145/3556536

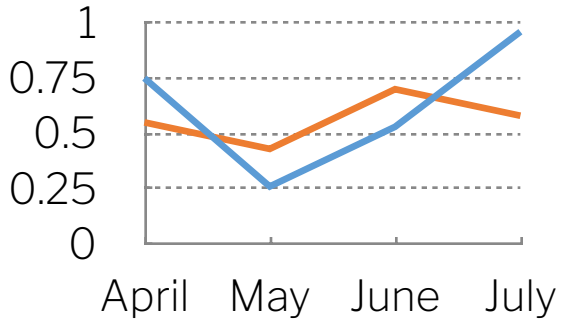
Basically:
Addressing more than mere one-shot mimicking.

e.g.,

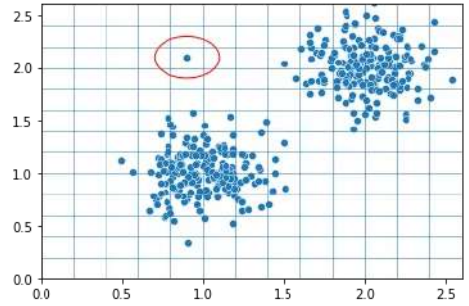
Dynamics



Longitudinal effects



Special/rare cases or extremes

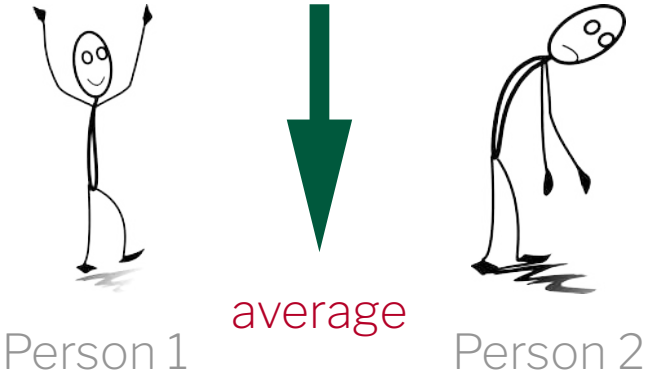


PARIS LODRON UNIVERSITY SALZBURG

We consider the “typical”, “average”, … user.

But who’s that?

The “average” might *not* even exist.



PARIS LODRON UNIVERSITY SALZBURG

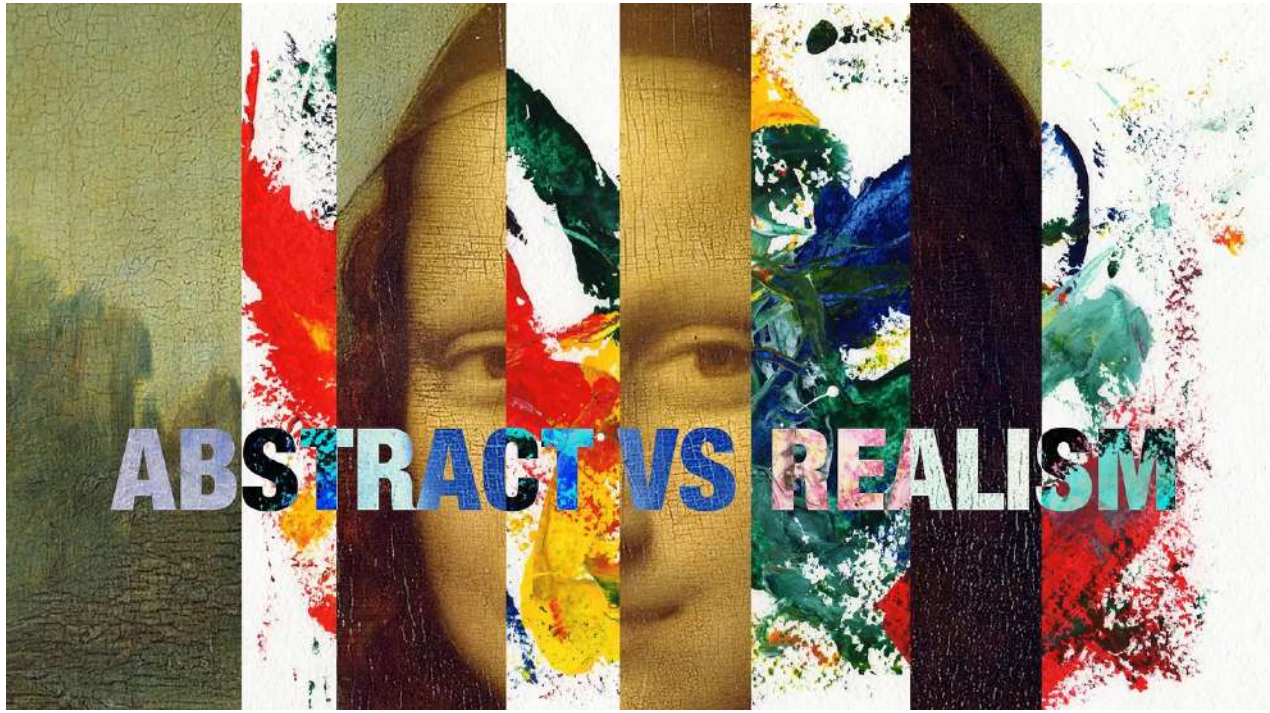
Is it all wrong?

—

Well, it depends!

PARIS LODRON UNIVERSITY SALZBURG

Accuracy/ realism vs. abstraction/ simplification



https://miro.medium.com/v2/resize:fit:1400/format:webp/1*ZCPI3n2DG5ksaWAAQ7X_gA.png

PARIS LODRON UNIVERSITY SALZBURG

A model is per definition a simplification.

Creating realistic user simulation models:

- is complex
- requires in-depth study to develop realistic assumptions
- may not always be necessary

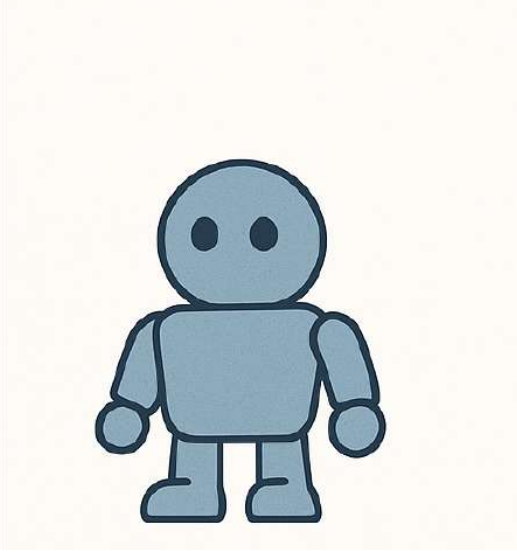
PARIS LODRON UNIVERSITY SALZBURG

Simulation ≠ faithful replication

We do not simulate
to replicate reality.

We simulate
to explore possibilities.

Realistic user model—
detail, complexity, accuracy



Toy agents—
abstraction, simplicity

PARIS LODRON UNIVERSITY SALZBURG

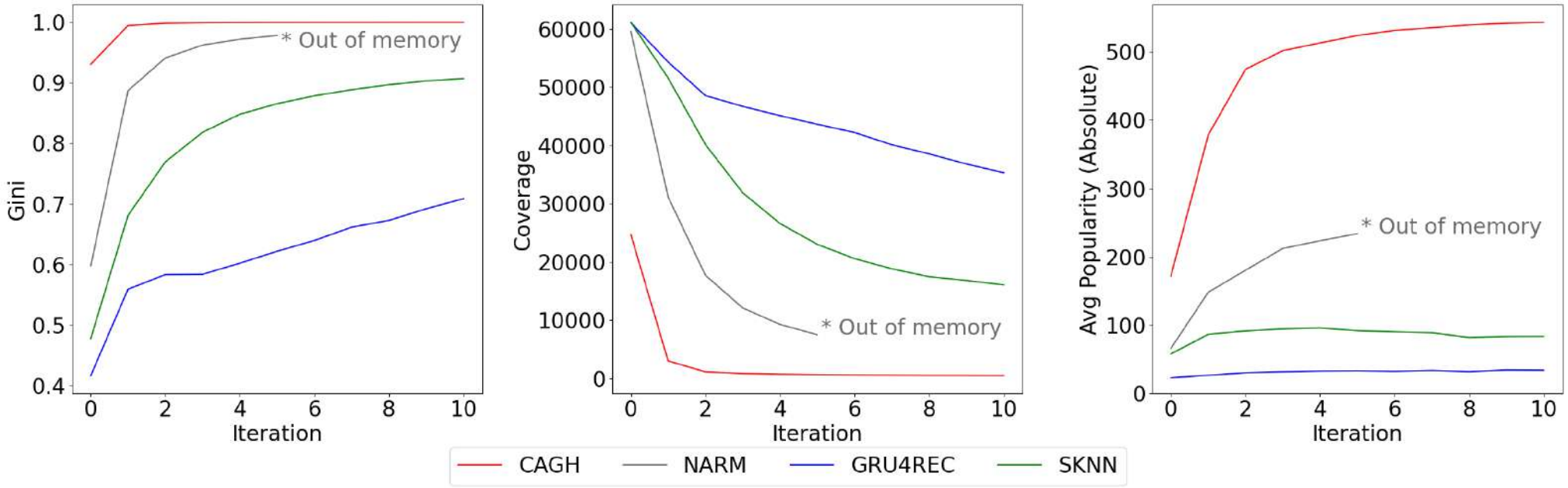
Longitudinal effects

- Simulating sessions, habits, or drift over time

- Generative science approach:
  → modeling at the micro-level, insights on the macro-level
  → simulate “what-if” scenarios (e.g., how changing one behavior rule affects the whole system)
  → Discover non-obvious dynamics (e.g., tipping points, paradoxes)

PARIS LODRON UNIVERSITY SALZBURG

Studying concentration, coverage, and popularity effects in session-based recommendation



Andres Ferraro, Dietmar Jannach, and Xavier Serra (2020). Exploring Longitudinal Effects of Session-based Recommendations. Proceedings of the 14th ACM Conference on Recommender Systems (RecSys '20). ACM, New York, NY, USA, 474–479. DOI: 10.1145/3383313.3412213

PARIS LODRON UNIVERSITY SALZBURG

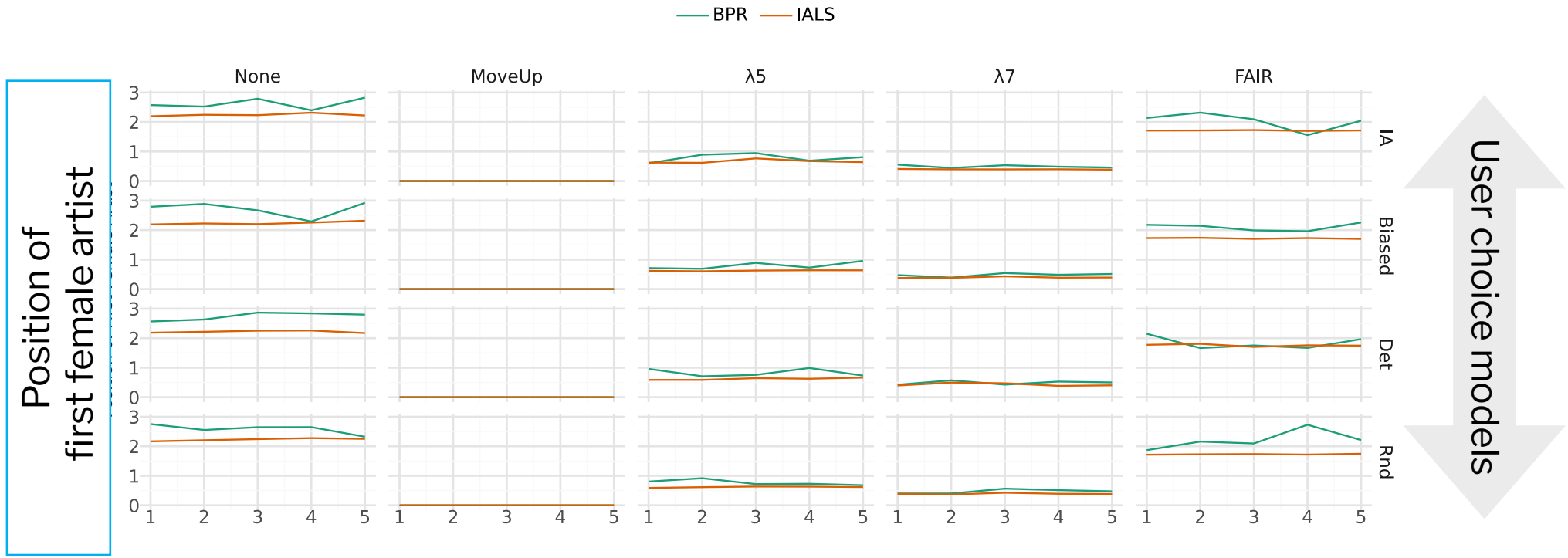
Comparing impact of (strongly) diverging behaviors

4 simple—possibly unrealistic—user choice models

Deterministic	Random	Inspection-Abandon	Biased
User listens to all top N recommendations	User listens to each item with probability 50%	User listens to each item with probability 50%, after item: stop listening entirely with probability 30%	Like Inspection-Abandon, but users are 10% more likely to listen to male artists

Andrés Ferraro, Michael D. Ekstrand, & **Christine Bauer** (2024). It's not you, it's me: the impact of choice models and ranking strategies on gender imbalance in music recommendation. Proceedings of the 18th ACM Conference on Recommender Systems (RecSys 2024). Bari, Italy, 14-18 October, pp 884-889. DOI: 10.1145/3640457.3688163

PARIS LODRON UNIVERSITY SALZBURG



More variation between columns than rows

↓

Algorithms have more impact than user choice models

Andrés Ferraro, Michael D. Ekstrand, & **Christine Bauer** (2024). It's not you, it's me: the impact of choice models and ranking strategies on gender imbalance in music recommendation. Proceedings of the 18th ACM Conference on Recommender Systems (RecSys 2024). Bari, Italy, 14-18 October, pp 884-889. DOI: 10.1145/3640457.3688163

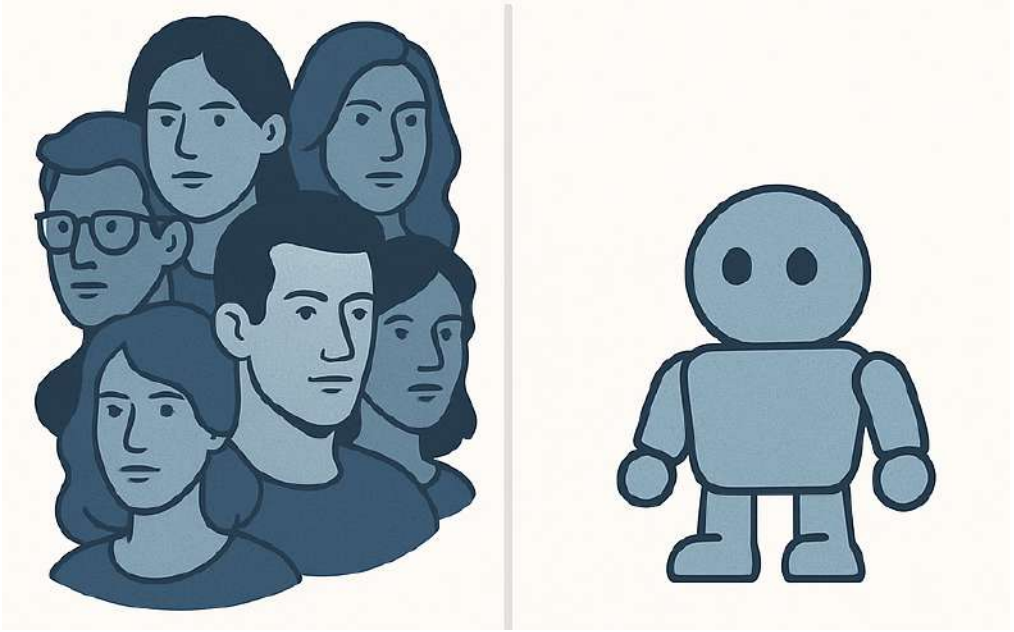
PARIS LODRON UNIVERSITY SALZBURG

From toy models to tactical insights

Toy models

→ not the whole picture

→ no subtle details



Tactical insights on what the user simulation was specifically designed for.

PARIS LODRON UNIVERSITY SALZBURG

Simulating extremes (e.g., to investigate implications)

Extremes can mean:

- Rare user behavior (e.g., a new user binge-clicking unrelated items)
- Unusual preference shifts (e.g., a user drastically changing interests overnight)
- Outlier content dynamics (e.g., one item going unexpectedly viral)
- Adversarial user behaviors (e.g., flooding clicks to boost an item)
- ...

PARIS LODRON UNIVERSITY SALZBURG

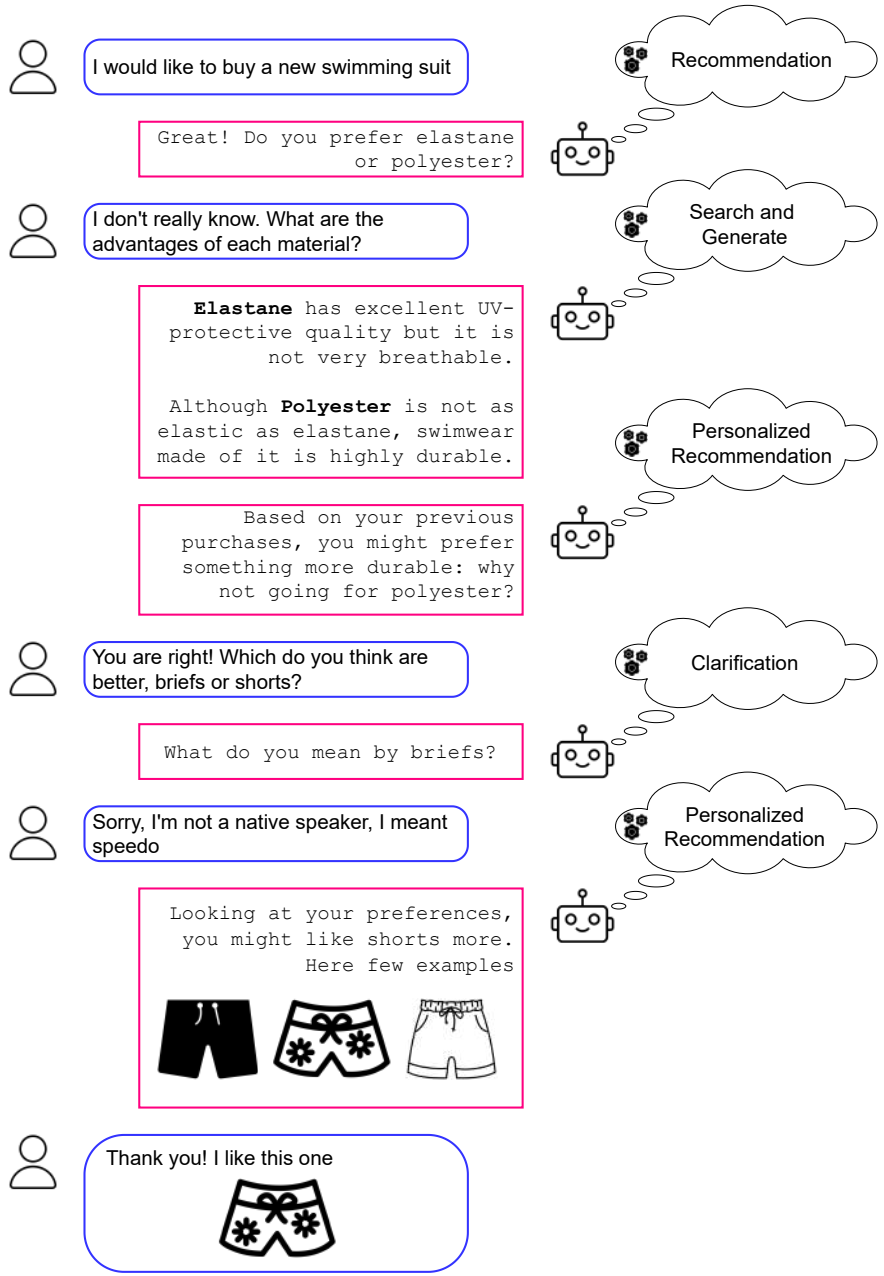
Dealing with complexity

e.g., CONversational Information ACcess agents (CONIAC)

→ Multi-turn interactions

→ Combination of retrieval, recommendation, clarification, explanations,...

→ Exploring “what ifs”



Christine Bauer, Li Chen, Nicola Ferro, Norbert Fuhr, et al. (2025). Conversational Agents: A Framework for Evaluation (CAFE) (Manifesto from Dagstuhl Perspectives Workshop 24352). DOI: 10.48550/arXiv.2506.11112

PARIS LODRON UNIVERSITY SALZBURG

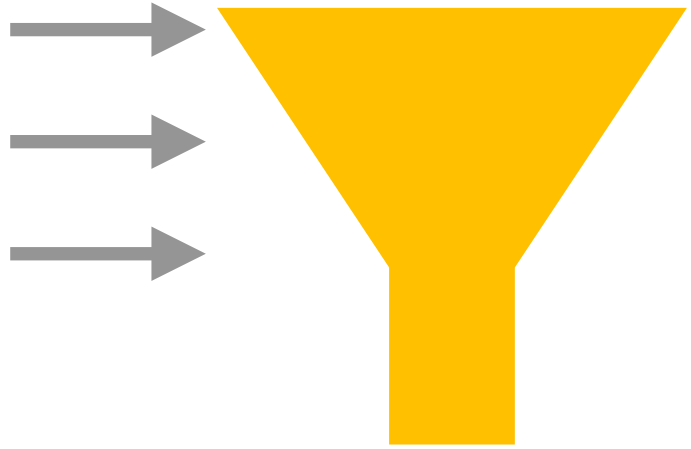
Abstraction: What we lose and gain

Aspect	Abstracted model	Realistic model
Speed	✅	❌
Fidelity	❌	✅
Interpretability	✅	❌
Generalization	✅	❌

PARIS LODRON UNIVERSITY SALZBURG

3 questions to ask

- What question are you trying to answer?
- What level of realism is necessary?
- How will you validate your insights?



design choice for user simulation

PARIS LODRON UNIVERSITY SALZBURG

Simulation with LLMs: the good, the bad, and the ugly

LLM-based simulations can be useful → growing number of research papers!

Yet, critical challenges require attention:

- Neither representative of the general population nor based on a well-defined, controlled user model
- How biased are they? → We do not know!

Speed	?
Fidelity	✗
Interpretability	✗
Generalization	?

PARIS LODRON UNIVERSITY SALZBURG

Takeaways

- Simulation is about reasoning, not realism
- Simple agents can yield deep insights
- Thoughtful abstraction is a strength, not a flaw
- Don’t you dare using “unthoughtful abstraction”!

PARIS LODRON UNIVERSITY SALZBURG

From toy models to tactics: What user simulation is good for

Christine Bauer

MuRS 2025 @RecSys 2025

Submission deadline: Aug 7th, 2025
https://sites.google.com/view/murs-2025

BEYOND 2025 @RecSys 2025

Submission deadline: Aug 7th, 2025
https://beyondrecsys.github.io

Univ.-Prof. Mag. DI Dr.
Christine Bauer
Professor of
Interactive Intelligent Systems

EXDIGIT
Department of Artificial Intelligence and Human Interfaces
Faculty of Digital and Analytical Sciences

17 July 2025

christine.bauer@plus.ac.at
https://christinebauer.eu
@christinebauer.bsky.social