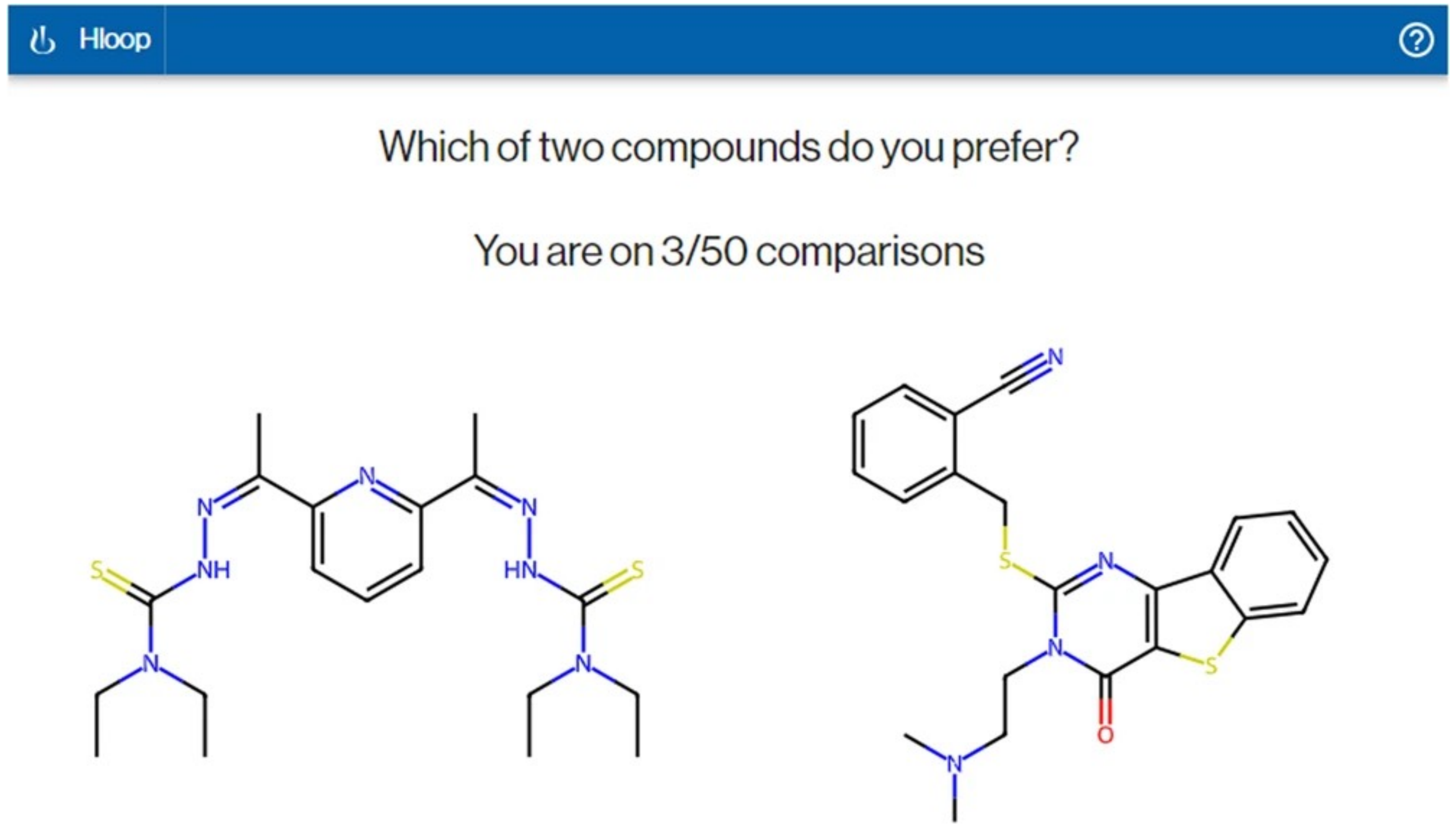
Preference Learning of Latent Decision Utilities with a Human-like Model of Preferential Choice

Sebastiaan De Peuter Shibei Zhu Yujia Guo
Andrew Howes Samuel Kaski

Aalto University School of Science
The University of Manchester
FCAI Finnish Center for Artificial Intelligence
University of Exeter

Learning from Preferences

Infer a latent utility function f_w over options x_i from human stated preferences.



Reproduced from: Choung, Oh-Hyeon, et al. "Extracting medicinal chemistry intuition via preference machine learning." *Nature Communications 14.1* (2023): 6651.

- (Graphics) Material Design [Brochu et al., NeurIPS 2007]
- Reinforcement Learning [Christiano et al., NeurIPS 2017]
- Small Drug Molecules [Choung et al., Nature Communications 2023]
- LLM Question answering [Ouyang et al., NeurIPS 2022]
- Text Summarisation [Stiennon et al., NeurIPS 2020]
- Image Generation [Xu et al., NeurIPS 2024]

Choice modelling

Prior work

	Model
Binary Choice Model	$p(y = x_1 \mid x_1, x_2) = \sigma\left(f_w(x_1) - f_w(x_2)\right)$
Bradley-Terry	$p(y = x_i \mid x_1, \ldots, x_n) \propto \exp\left(\beta f_w(x_i)\right)$

Challenge: human biases incl. context effects

Choice modelling

Context effects

Contextual preference reversal: a change in preference between two options due to a change in further options.

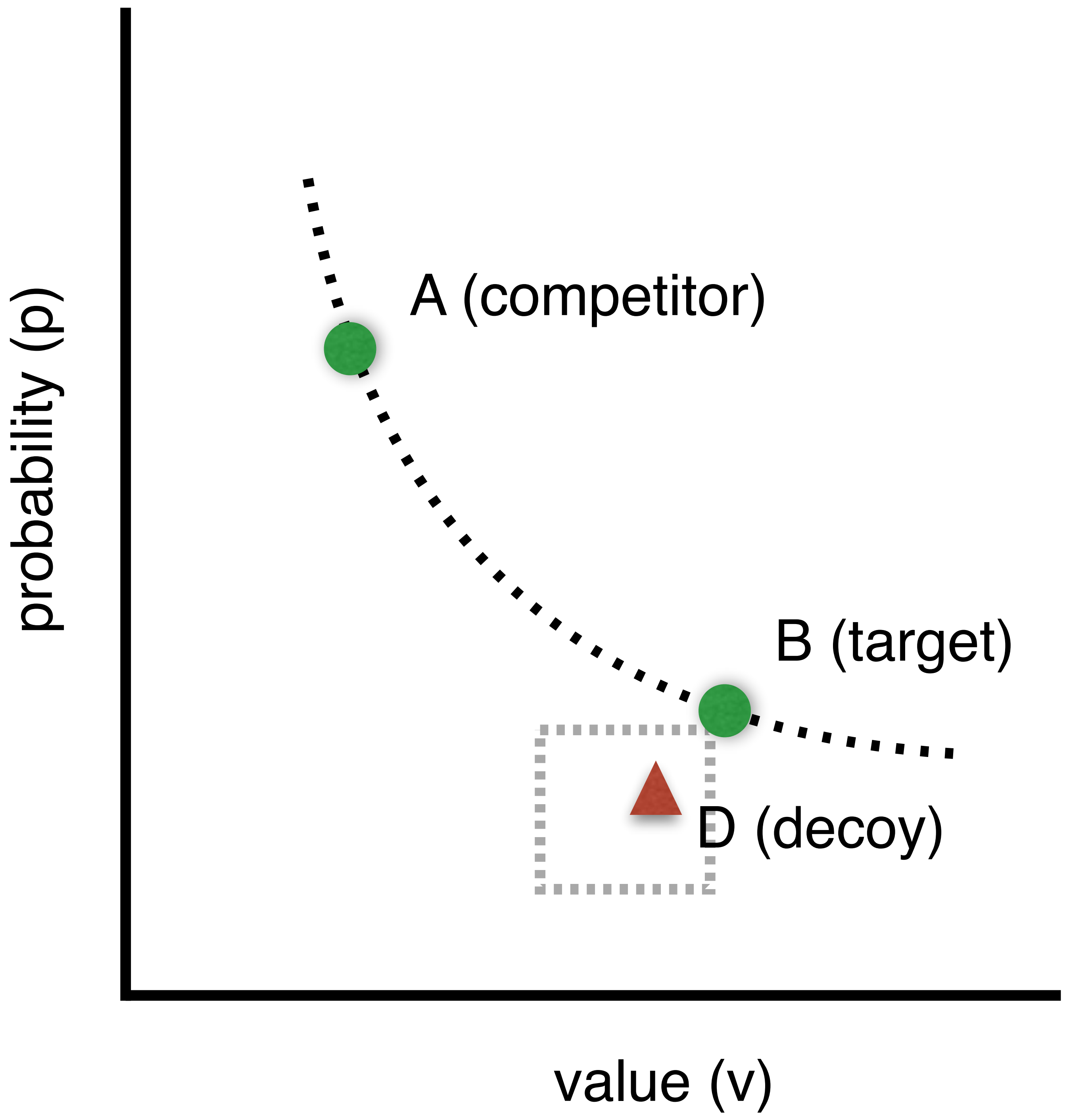


Figure: Attraction effect

Howes, Andrew, et al. "Why contextual preference reversals maximize expected value." *Psychological review* 123.4 (2016): 368.
Wollschlaeger, Lena M., and Adele Diederich. "Similarity, attraction, and compromise effects: Original findings, recent empirical observations, and computational cognitive process models." *The American Journal of Psychology* 133.1 (2020): 1-30.

Choice modelling

Prior work

	Model
Binary Choice Model	$p(y = x_1 \mid x_1, x_2) = \sigma\left(f_w(x_1) - f_w(x_2)\right)$
Bradley-Terry	$p(y = x_i \mid x_1, \ldots, x_n) \propto \exp\left(\beta f_w(x_i)\right)$
Challenge: human biases incl. context effects	
Bower & Balzano, ICML 2020	$p(y = x_i \mid x_1, \ldots, x_n) \propto \exp\left(w^T x_i^{\tau(C)}\right)$
Tomlinson & Benson, SIGKDD 2021	$p(y = x_i \mid x_1, \ldots, x_n) \propto \exp\left((w + A x_C)^T x_i\right)$
CRCS (ours)	Computationally rational

Contributions

1. Show that **computational rationality** theory can **improve inference from and prediction** of human behavior, specifically for learning from preferences.

2. Introduce **CRCS**, a surrogate of an existing cognitive choice model that allows for **practical inference**

3. **Extend** CRCS into **LC-CRCS**, which can learn **cross-feature effects** between options.

Computationally
Rational
Choice
Surrogate

A computationally rational choice model

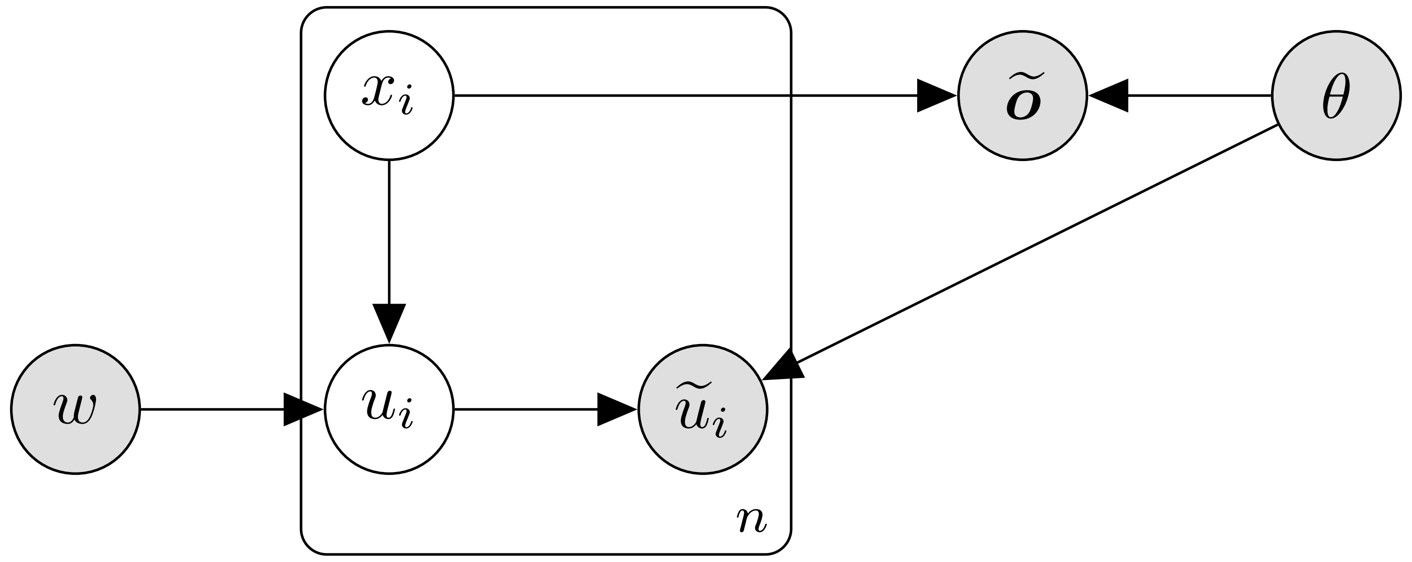
The original cognitive model

Choices are **utility-maximizing** under **observational bounds.** [Howes et al. 2016]

Options not directly observable. Instead, noisy observations:

$\widetilde{\boldsymbol{u}}$: noisy utility value

$\widetilde{\boldsymbol{o}}$: noisy pairwise attribute value comparisons



$$y = \mathrm{argmax}_{x_i \in \{x_1, \ldots, x_n\}} \mathbb{E}[u_i \,|\, \widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{o}}, w, \theta]$$

Howes, Andrew, et al. "Why contextual preference reversals maximize expected value." *Psychological review* 123.4 (2016): 368.

A computationally rational choice model

Making it tractable

$$y = \text{argmax}_{x_i \in \{x_1,\ldots,x_n\}} \mathbb{E}[u_i \,|\, \widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{o}}, w, \theta]$$

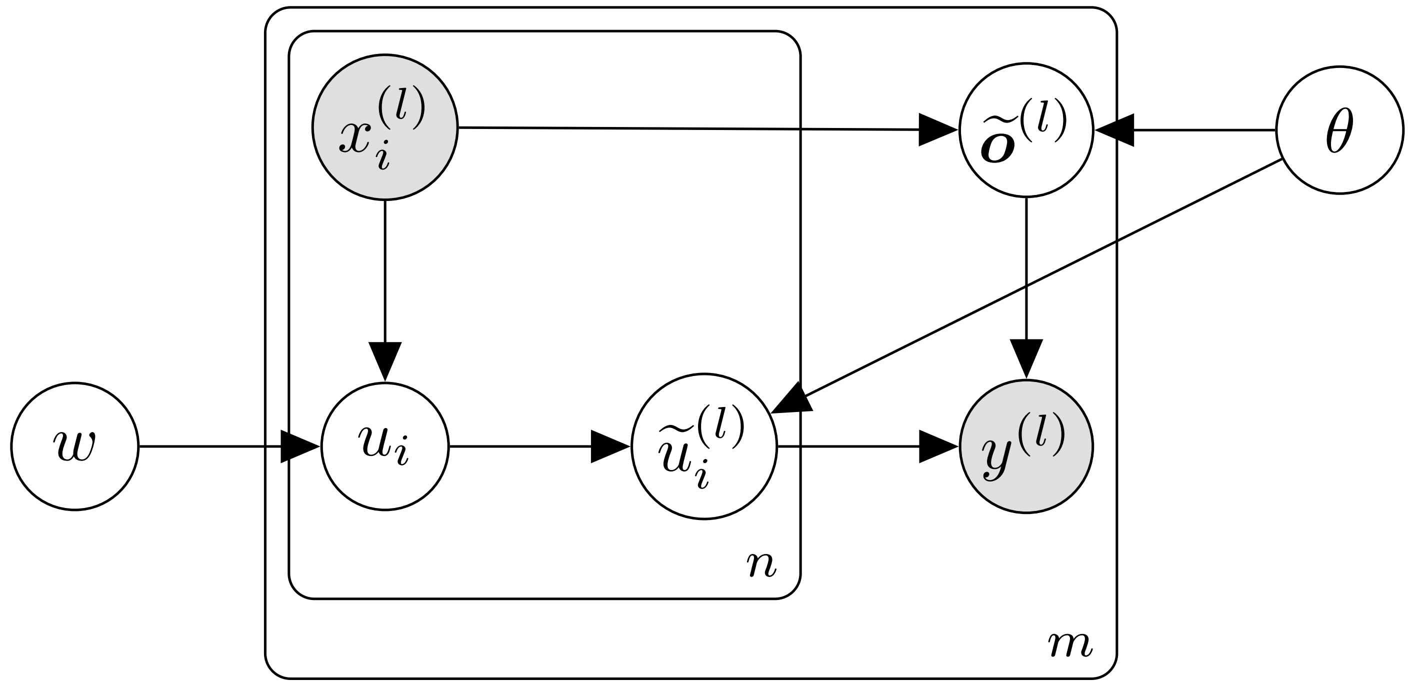
Surrogate trained to predict expected utility of options:

$$\mathscr{L}_{\text{util}}(\hat{u}) = \mathbb{E}_{p(w,\theta,\boldsymbol{u},\widetilde{\boldsymbol{u}},\widetilde{\boldsymbol{o}})} \left[\| \hat{u}(\widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{o}}, w, \theta) - \boldsymbol{u} \|_2 \right]$$

Tractable Preference learning with CRCS

Outside observer infers parameters from observed choices.

But, associated noisy observations $\widetilde{\boldsymbol{o}}, \widetilde{\boldsymbol{u}}$ are latent.



$$\mathscr{L}_{\text{pol}}(\widehat{q}) = \mathbb{E}_{p(w,\theta,\boldsymbol{x},\widetilde{\boldsymbol{u}},\widetilde{\boldsymbol{o}})}\left[-\ln \widehat{q}\left(\operatorname{argmax}_{\{x_1,\ldots,x_n\}} \widehat{u}(\widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{o}}, w, \theta) \middle| \boldsymbol{x}, w, \theta\right)\right]$$

LC-CRCS

LCL can learn cross-feature effects through Ax_C

$$\exp\left((w + Ax_C)^T x_i\right)$$

LC-CRCS introduces the same mechanism in CRCS

$$\hat{q}\left(y \middle| \boldsymbol{x}, w, \theta\right) \longrightarrow \hat{q}\left(y \middle| \boldsymbol{x}, w{+}Ax_C, \theta\right)$$

CRCS LC-CRCS

Experiments

Inference and Choice Prediction

LC-CRCS is a **better predictor** of real human choices

Table 1: Choice model NLLs on human choice data sets. Bolded digits indicate a significant ($p < 0.01$) improvement over baselines (BT, BB, LCL).

Dataset	Bradley-Terry	Bower & Balzano	LCL	CRCS (ours)	LC-CRCS (ours)
Hotels	573	573	553	**536**	**536**
District-Smart	3432	3432	3305	3371	**3276**
Car-Alt	7414	7416	7290	7322	7345
Dumbalska	103669	103711	100683	100450	**99147**

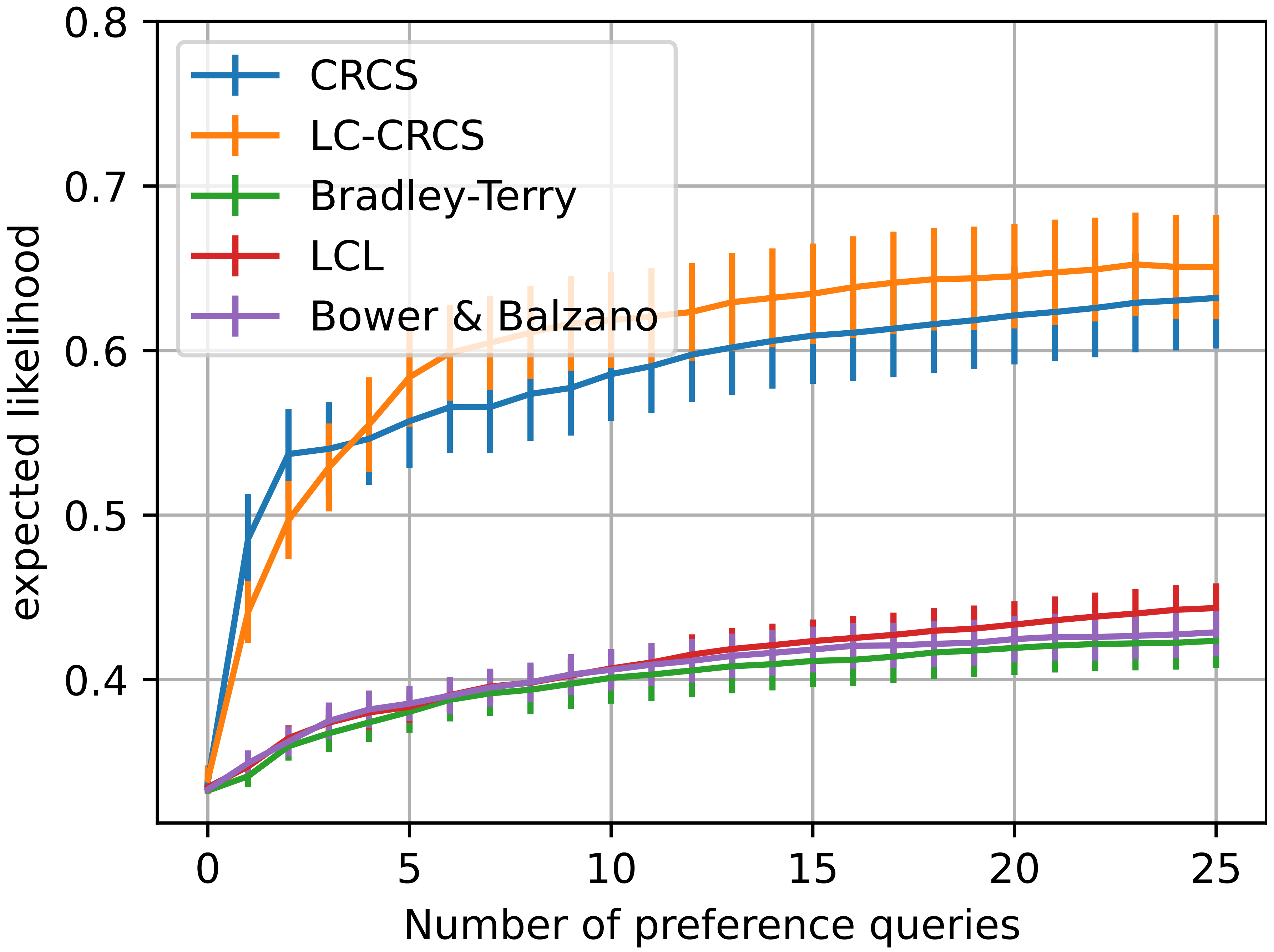
LC-CRCS and CRCS infer utility that **better aligns** with human preferences

Table 2: Consistency of inferred utility function with separately collected rankings on District-Smart. Bolded digits indicate a significant improvement over baselines (BT, BB, LCL).

Dataset	Bradley-Terry	Bower & Balzano	LCL	CRCS (ours)	LC-CRCS (ours)
District-Smart	0.162	0.217	0.286	**0.622**	**0.525**

Active Inference and Assistance

CRCS and LC-CRCS are **more data-efficient** in an active learning setting.



In-silico experiments show practicality of CRCS for real design problems:

- Structural design
- Drainage network design
- Retrosynthesis planning

Conclusion

- Show that **computational rationality** can improve learning from preferences through strong **inductive biases**.
- Introduce **CRCS**, a computationally rational surrogate for human choice making which enables **practical inference**.
- Extend CRCS to **LC-CRCS**, which can learn additional **cross-feature effects**.
- Show experimentally that CRCS and LC-CRCS have significantly **better utility inference and choice prediction** compared to baselines.

Aalto University School of Science
MANCHESTER 1824 The University of Manchester
FCAI Finnish Center for Artificial Intelligence
UNIVERSITY OF EXETER