LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Media Informatics Group

The Impact of Expertise In the Loop for Exploring Machine Rationality

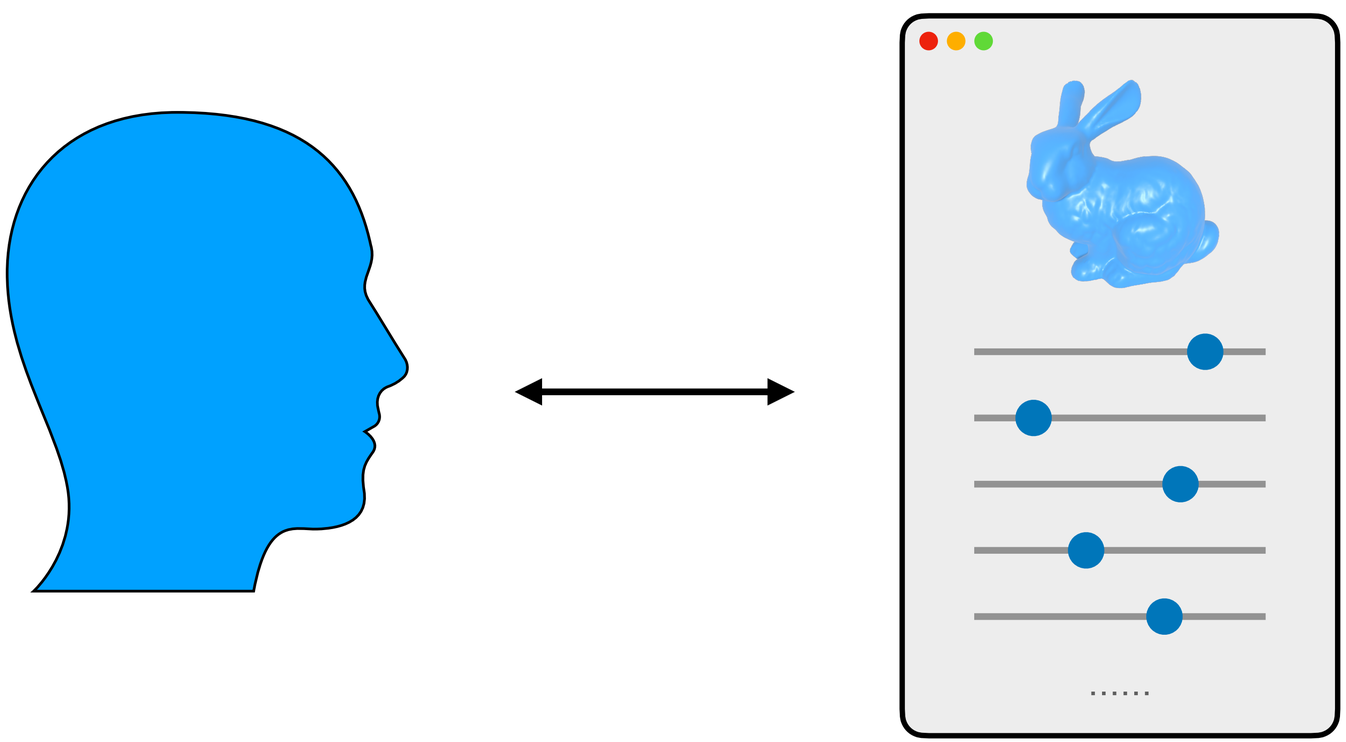
Changkun Ou, Sven Mayer, Andreas Martin Butz
Media Informatics Group
LMU Munich

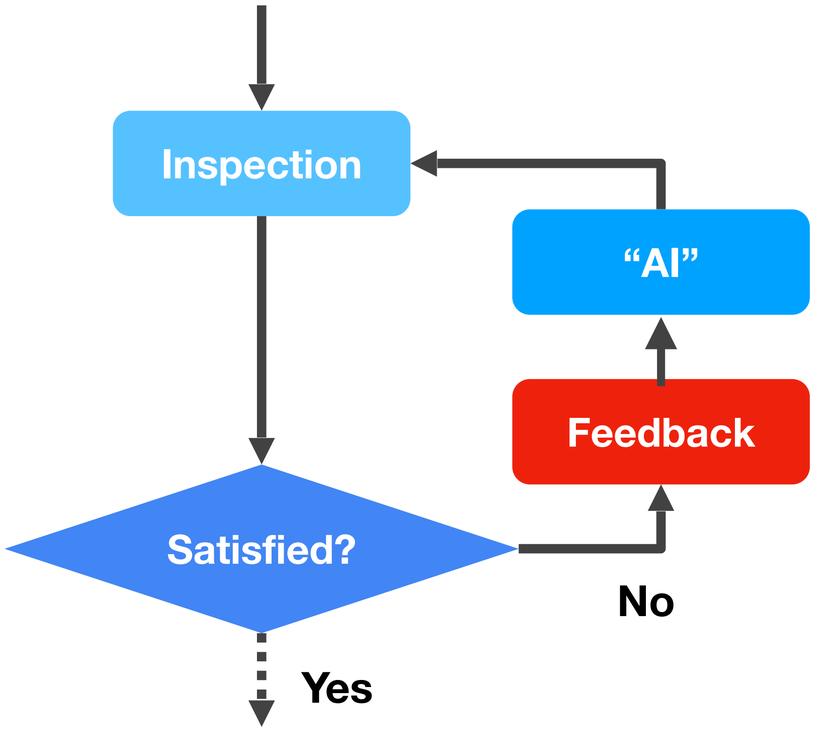
ACM Intelligence User Interfaces 2023
Session 4
March 29, Sydney, Australia

Motivation

Human-in-the-Loop (HITL) Optimization Systems

A human-in-the-loop optimization system refers systems or processes that involves an **underlying optimization** process **towards** user **expectation** or preference.





Human-in-the-Loop (HITL) Optimization Systems

A human-in-the-loop optimization system refers systems or processes that involves an **underlying optimization** process **towards** user **expectation** or preference.

Examples

Text Summarization [Simpson et al. 2020]

Photo Color Enhancement [Koyama et al. 2016]

Melody composition [Zhou et al. 2021]

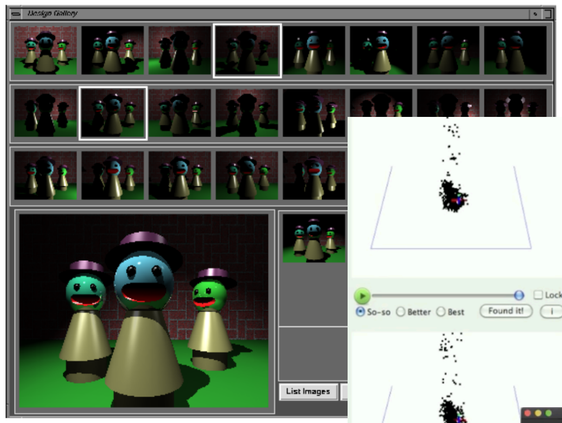
Meshing [Ou et al. 2022]

Material design [Brochu et al. 2007]

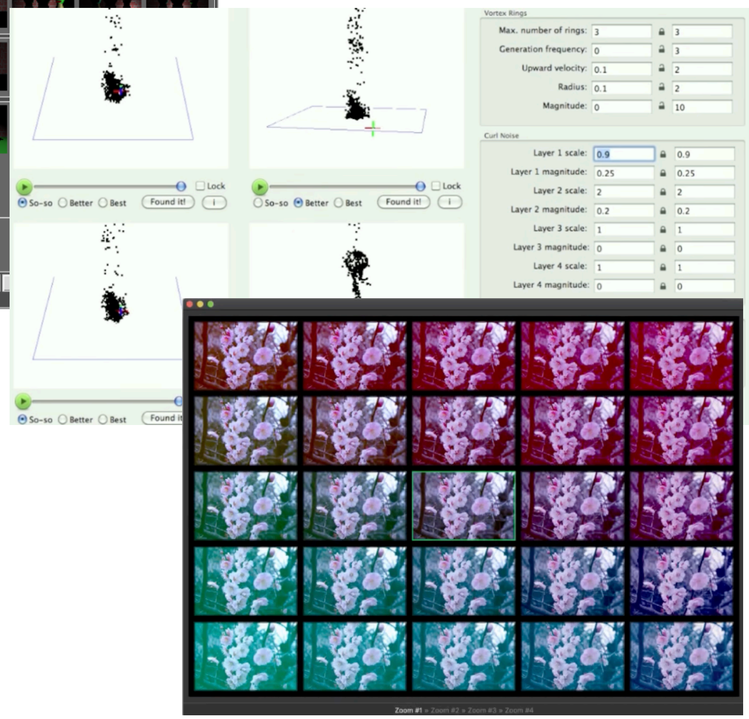
Animation [Brochu et al. 2010]

Illumination [Marks et al. 1997]

…



[Marks et al. 1997]



[Brochu et al. 2010]

[Koyama et al. 2020]

Research Questions and Hypothesis

Observations:

1. Problem context, unstable preference objectives can lead to non-satisfactory results
2. Prior HITL optimization systems are mostly evaluated on novice users and rarely report on the effect of user expertise
3. Prior HITL optimization systems are evaluated based on subjective responses

Research Questions and Hypothesis

Observations:

1. Problem context, unstable preference objectives can lead to non-satisfactory results
2. Prior HITL optimization systems are mostly evaluated on novice users and rarely report on the effect of user expertise
3. Prior HITL optimization systems are evaluated based on subjective responses

RQs:

What is the objective impact of involved expertise on the system outcomes?

What is the subjective impact of involved expertise on user satisfaction?

Research Questions and Hypothesis

Observations:

1. Problem context, unstable preference objectives can lead to non-satisfactory results
2. Prior HITL optimization systems are mostly evaluated on novice users and rarely report on the effect of user expertise
3. Prior HITL optimization systems are evaluated based on subjective responses

RQs:

What is the objective impact of involved expertise on the system outcomes?

What is the subjective impact of involved expertise on user satisfaction?

Hypothesis:

Using higher expertise leads to better results in HITL optimization

User Study

User Study and Workflow

Task:

Fill beginning survey, providing feedback to the AI results; inspect improved results;

loop until satisfaction, and fill ending survey

User Study and Workflow

Task:

Fill beginning survey, providing feedback to the AI results; inspect improved results;
loop until satisfaction, and fill ending survey

Selection Criteria:

- A task should partially involve rational, objective judgment, and subjective components.
- Each domain requires different levels of human expertise

User Study and Workflow

Task:

Fill beginning survey, providing feedback to the AI results; inspect improved results;

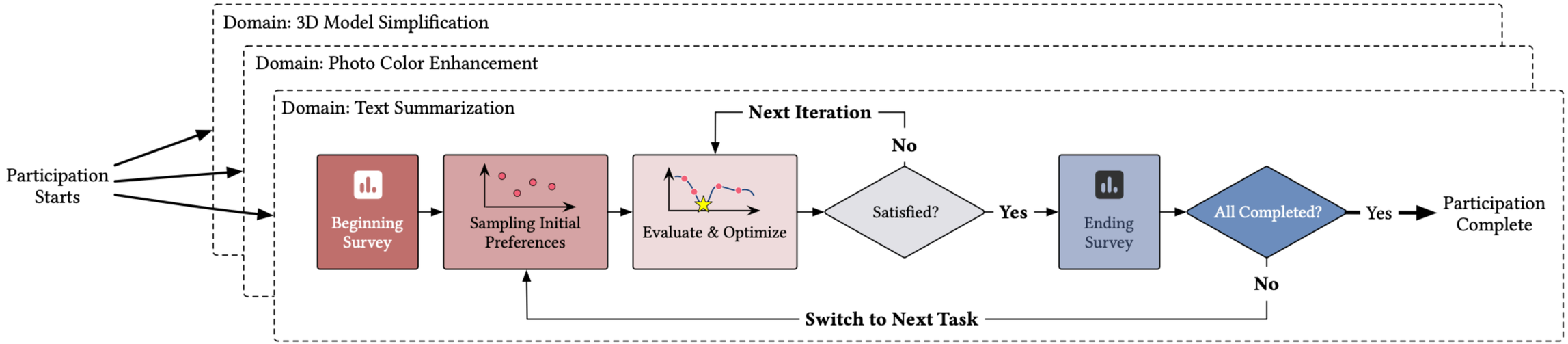
loop until satisfaction, and fill ending survey

Selection Criteria:

- A task should partially involve rational, objective judgment, and subjective components.
- Each domain requires different levels of human expertise

Participants (N=60)

3 selected domains. 20 for each domain, 31 female, 29 male, no diverse; age M=26.92, range 19-52



Apparatus: Text Summarization

Pre-trained BART model, fine-tuned for CNN, nucleus sampling

4 adjustable hyper parameters

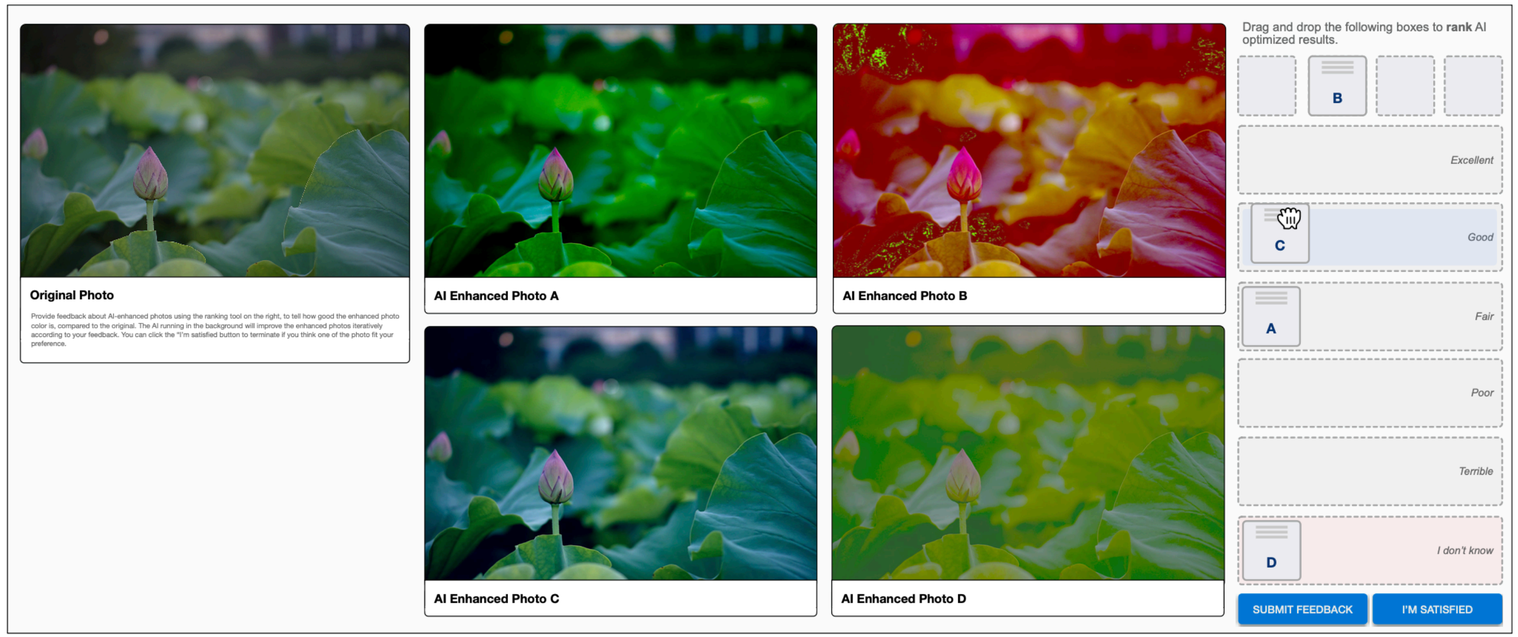
- Summarization ratio
- Length penalty
- top-p
- temperature

Apparatus: Photo Color Enhancement

A parameterized photo color enhancer [Koyama et al. 2016, 2017, 2020]

4 adjustable hyper parameters

- Brightness
- Contrast
- Saturation
- Temperature

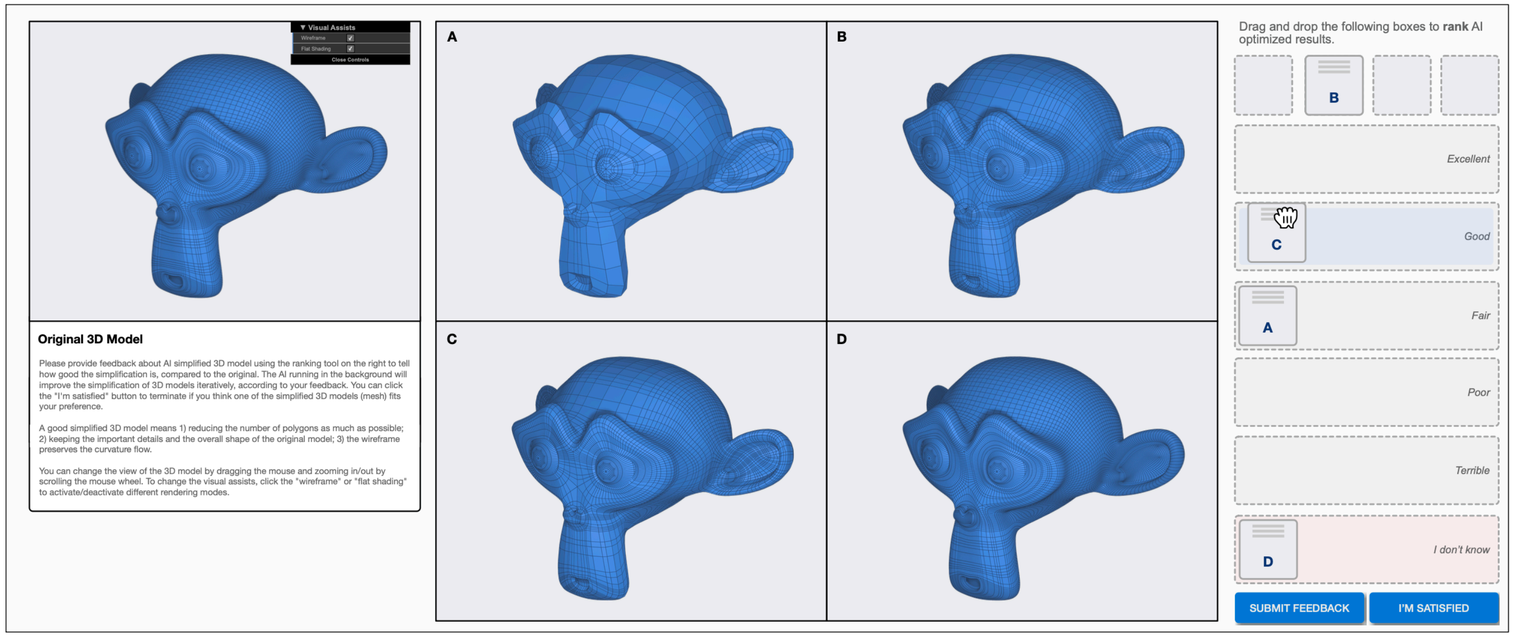


Apparatus: 3D Model Simplification

A parameterized 3D mesh simplifier [Ou et al. 2022]

5 adjustable hyper parameters

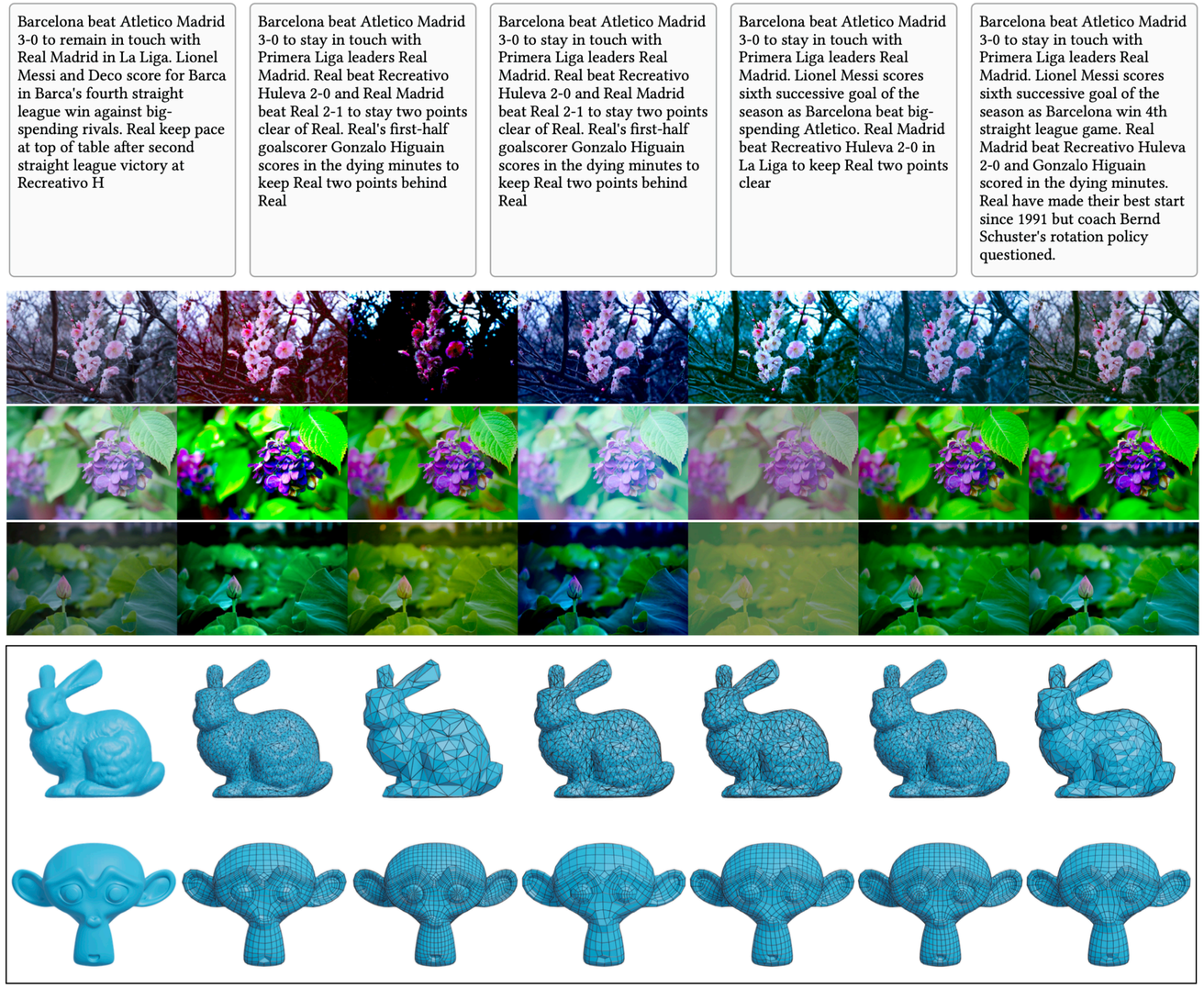
- Simplification ratio
- Border preservation
- Hard edge preservation
- Sharpness preservation
- Quadrilateral preservation



Apparatus: Bayesian Optimizer

Based on Expected Utility Bayesian Optimization [Lin et al. 2022]

Modified to fit objects ranking optimization



Example outcomes

Evaluation & Results

Evaluation Methodology

Machine performance

Objective outcome quality measures

BLEU, ROUGE; HSV, YUV; SSIM/PSNR, Jacobian Cell, Chamfer Distance

Optimizer measures

Posterior mean of the estimated ranking utility

User performance

Interaction behavior measures

decision time, iterations, incomplete/indifference preference, ranking interactions

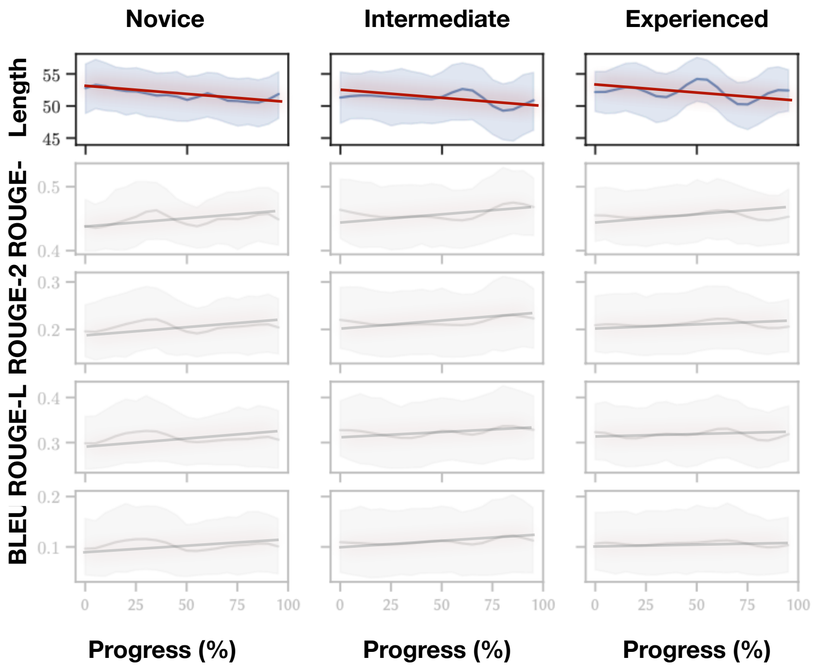
Expertise measures, hindsight questionnaire measures

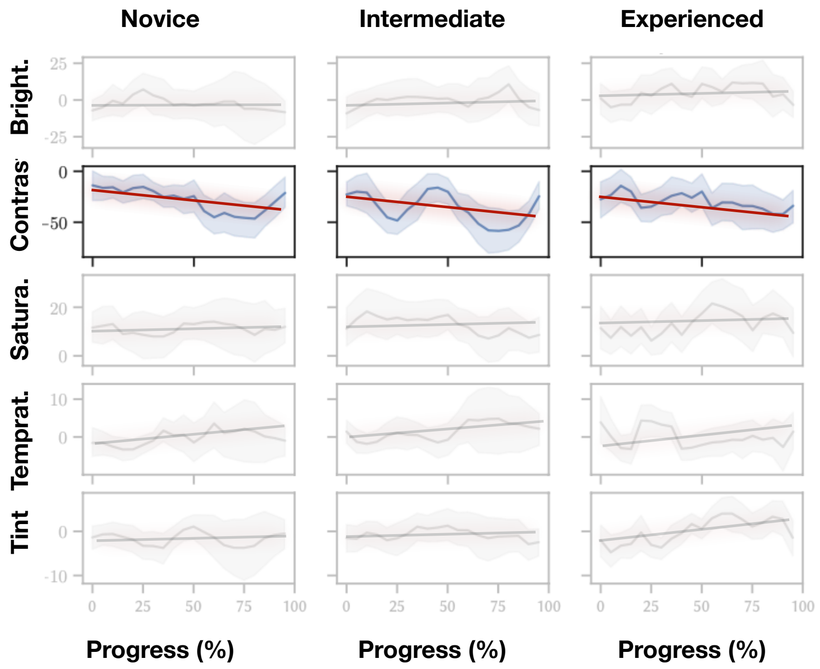
years of expertise, subjective satisfaction

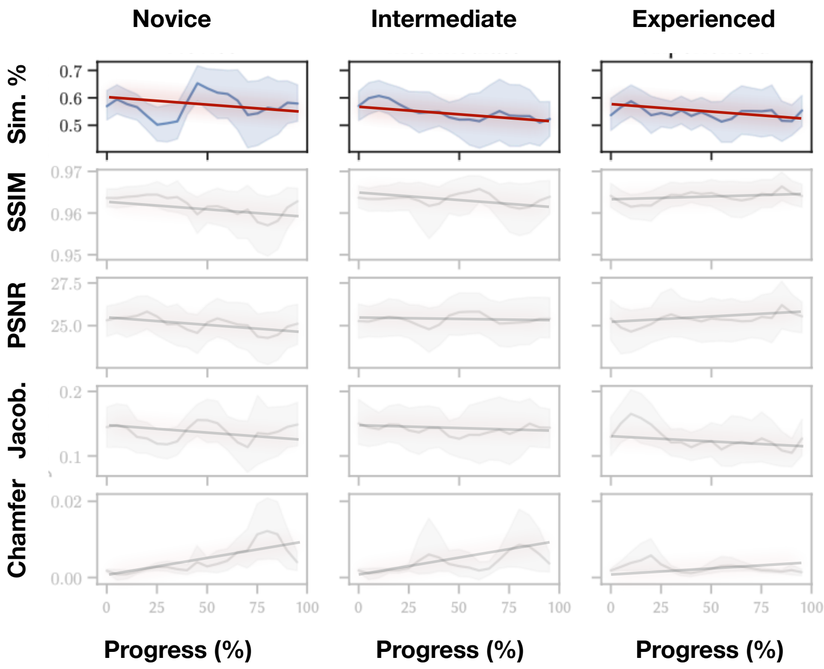
Quantile-based discretization* (relative expertise)

Machine Performance: Objective Outcome Quality

Novices and intermediates can reach expert level performance



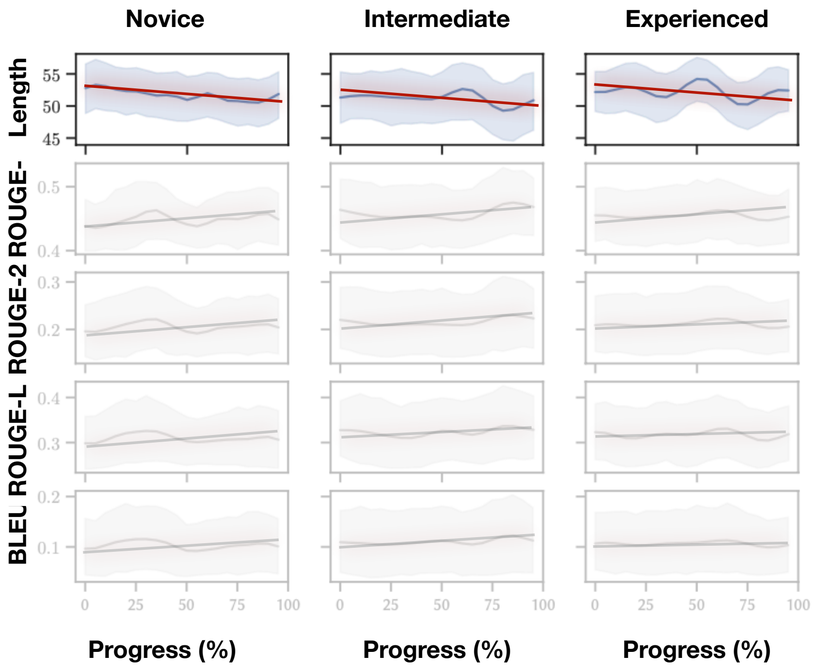




Machine Performance: Objective Outcome Quality

Novices and intermediates can reach expert level performance

The effect of iteration was statistically significant for at least one objective



Text Summarization

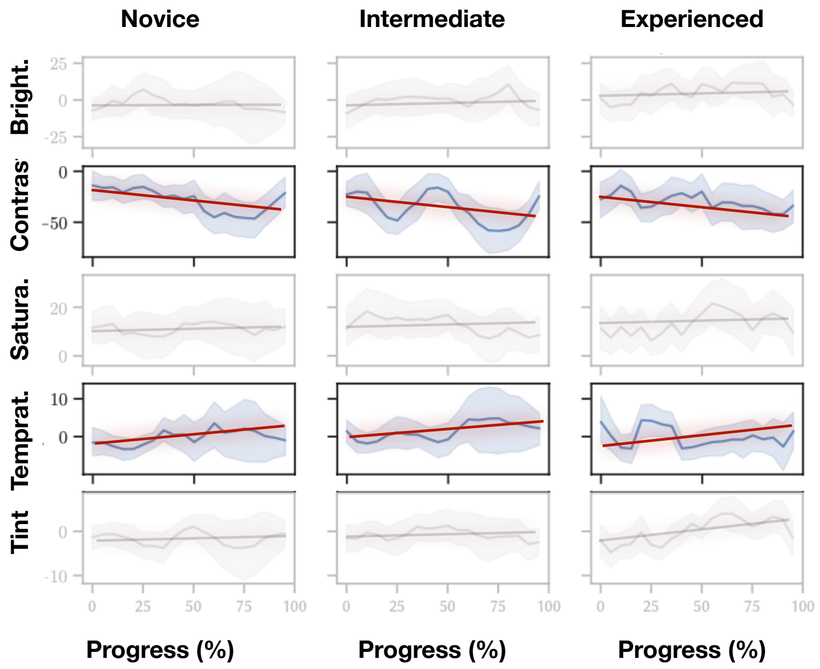
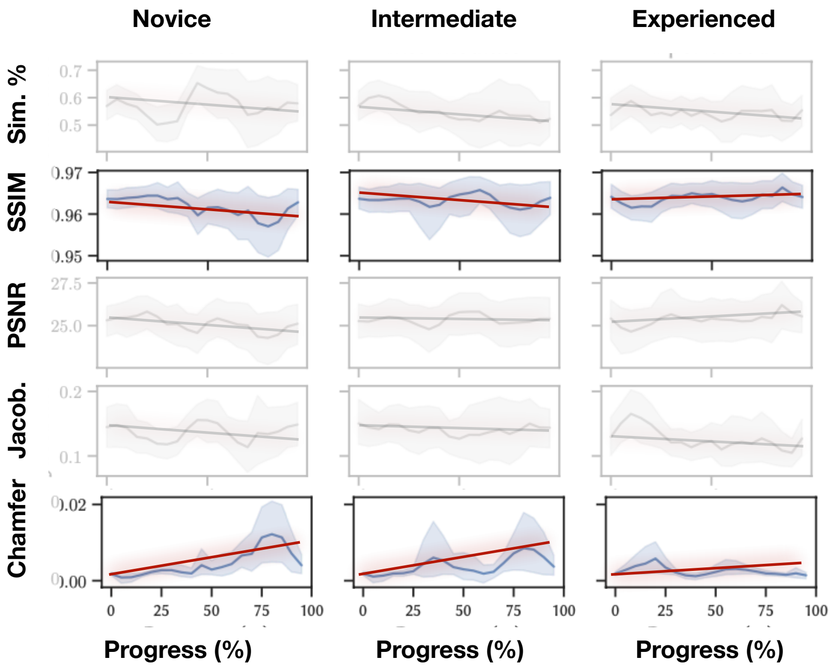


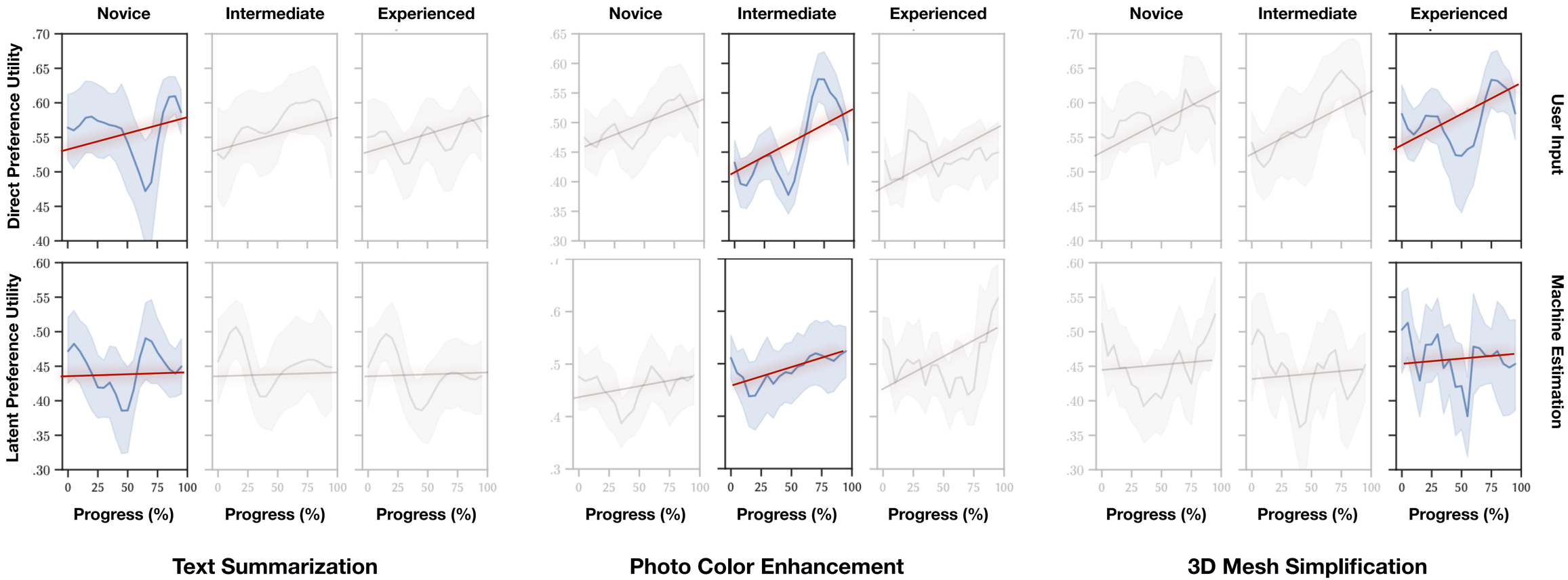
Photo Color Enhancement



3D Mesh Simplification

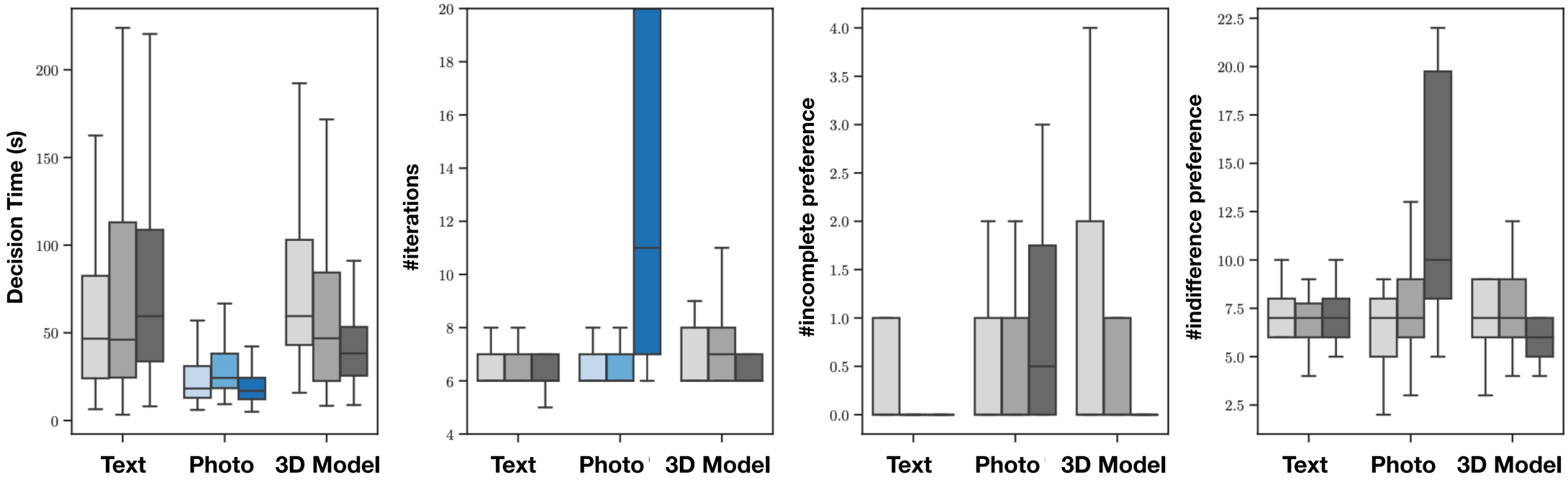
Machine Performance: Direct/Latent Utility

Direct preference utility matches latent preference utility



User Performance: Interaction Behavior

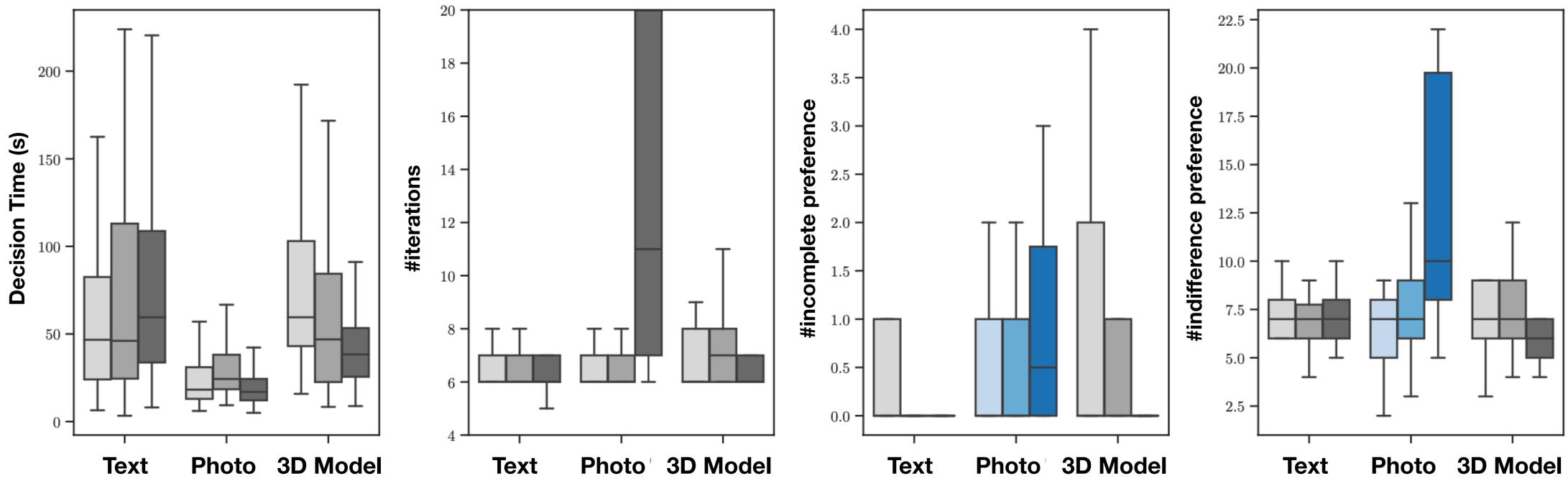
Experienced participants explore solutions more when feedback loop is more efficient



User Performance: Interaction Behavior

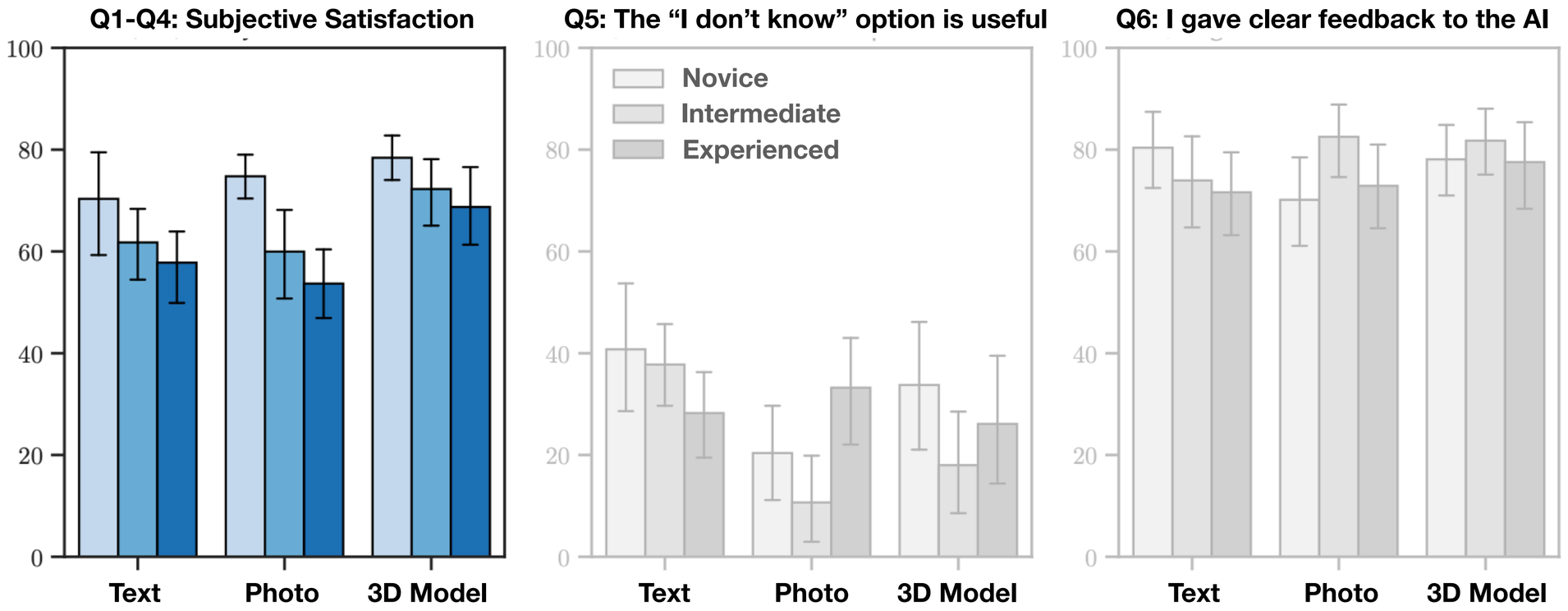
Experienced participants explore solutions more when feedback loop is more efficient

Experienced participants indicate clearer preference by showing more incomplete/indifference preference



User Performance: Subjective Satisfaction

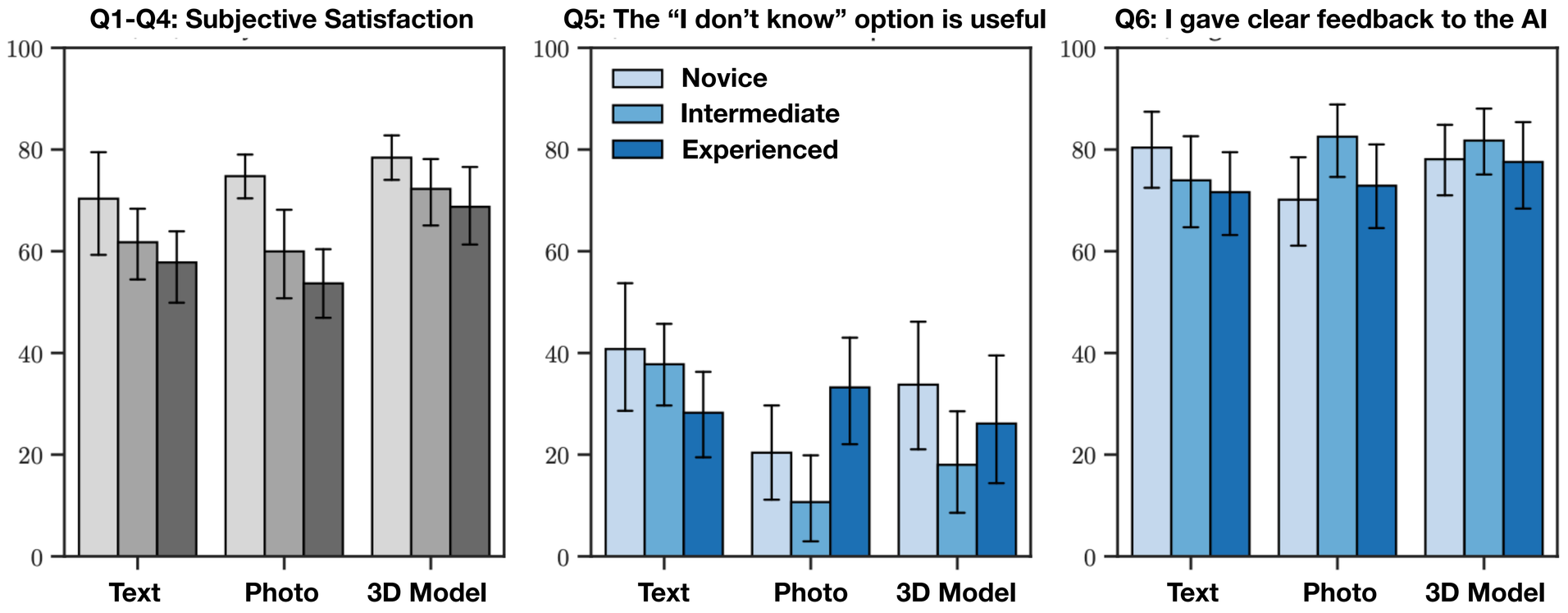
Instead, novices are significantly **more satisfied** than experienced ones



User Performance: Subjective Satisfaction

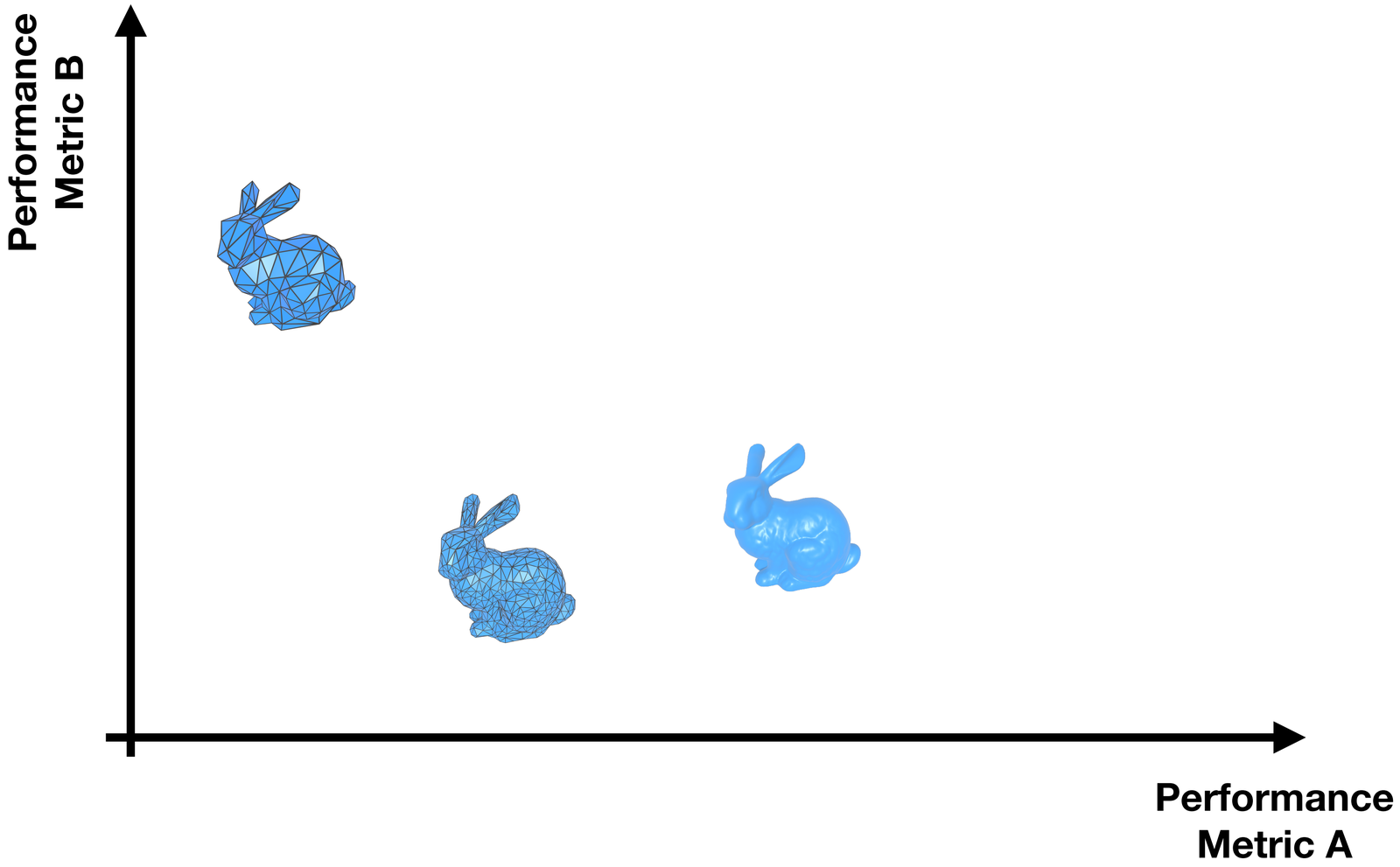
Instead, novices are significantly more satisfied than experienced ones

All participants consider they gave **clear feedback**

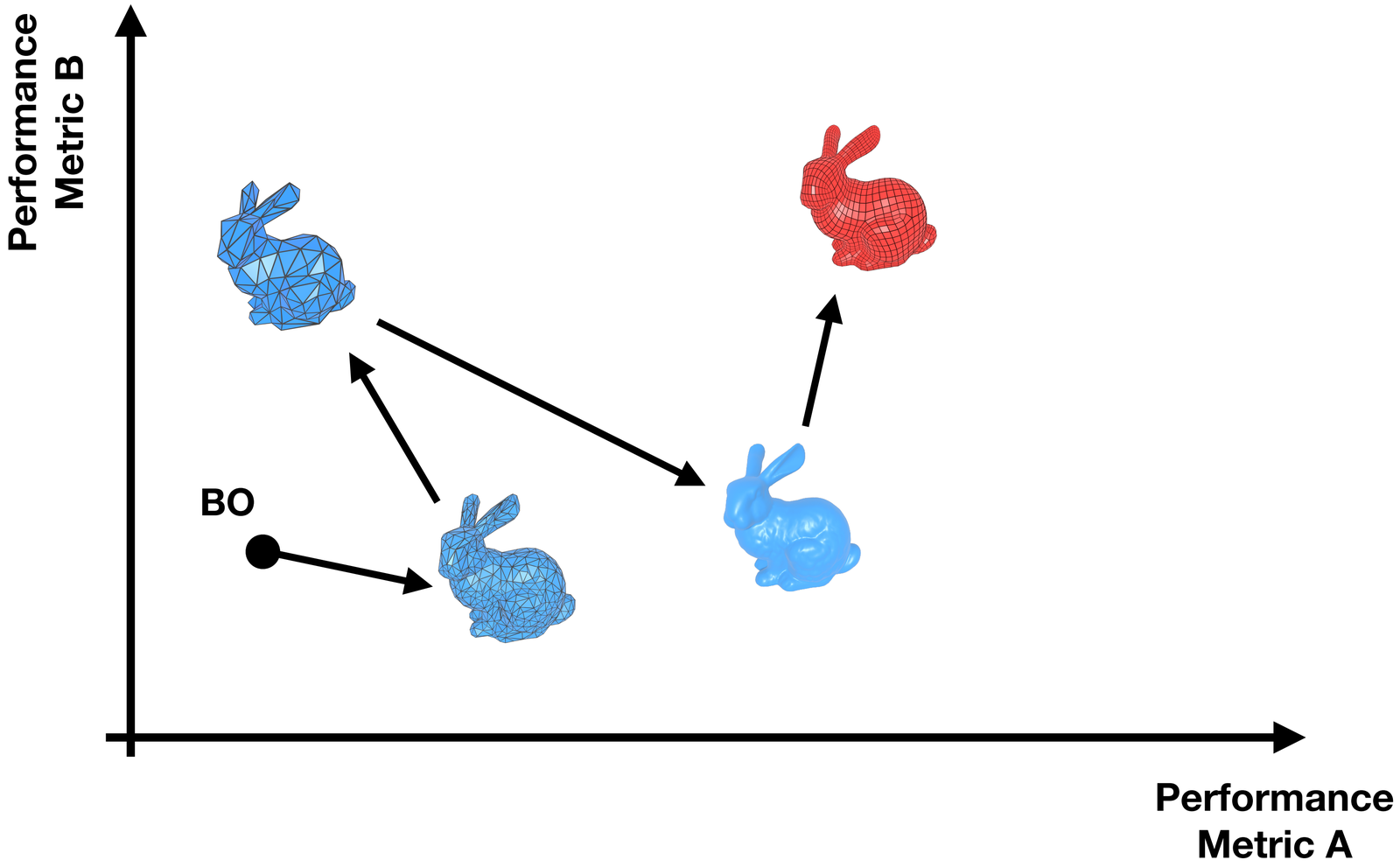


Implications

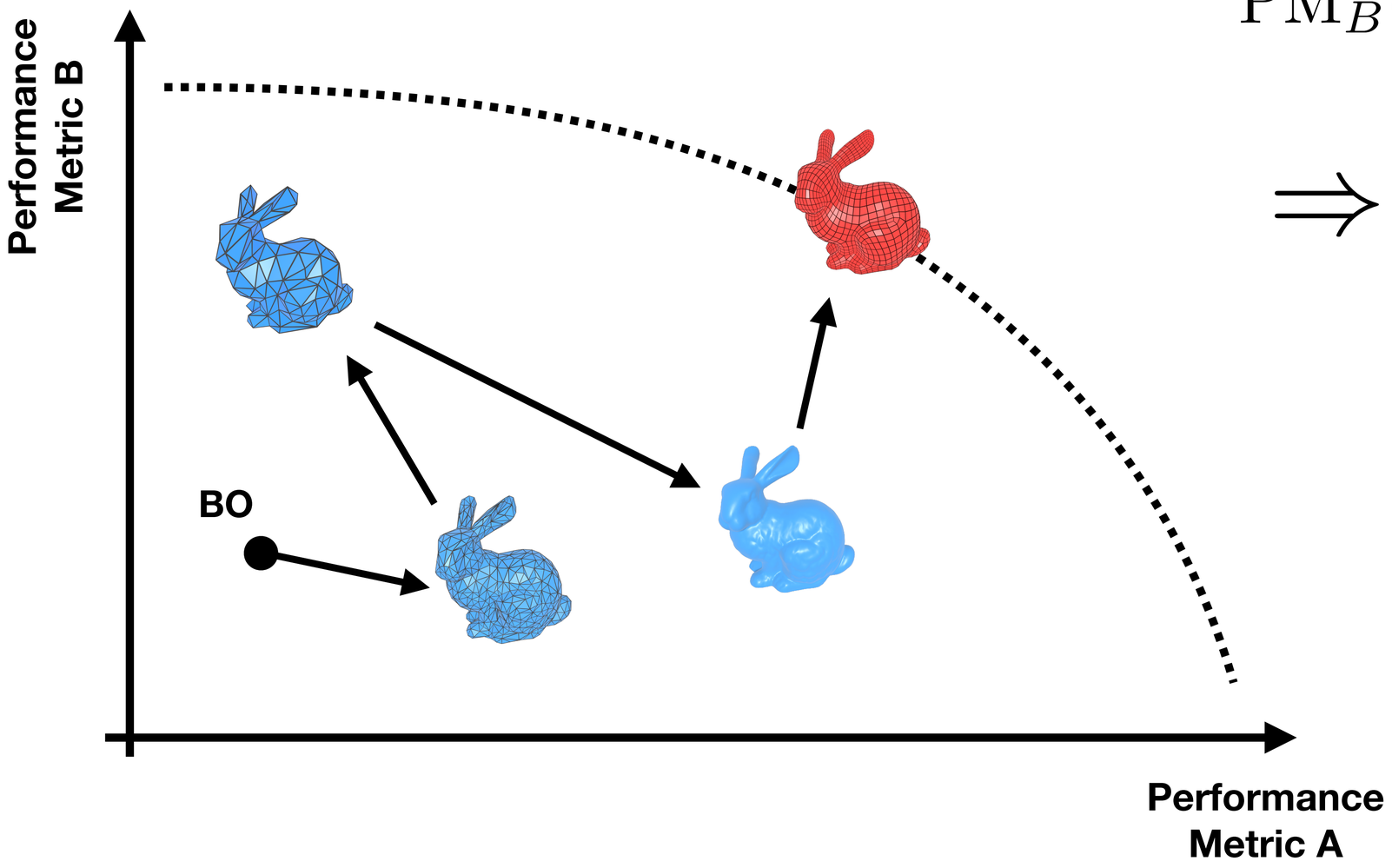
Pareto Optimality [Pareto, 1912]



Pareto Optimality [Pareto, 1912]



Pareto Optimality [Pareto, 1912]

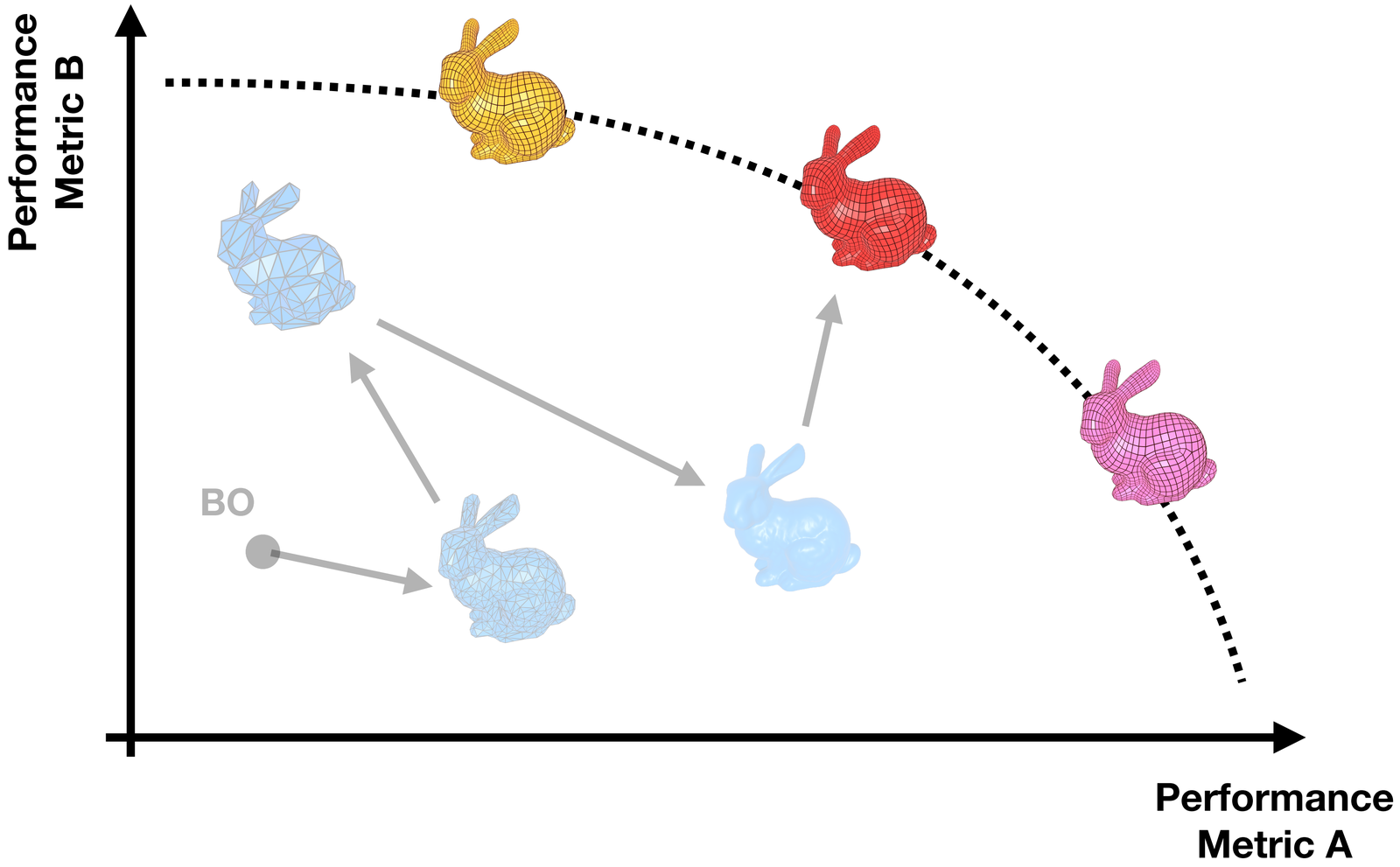


$$\mathrm{PM}_A(\ ) \geq \mathrm{PM}_A(\ )$$

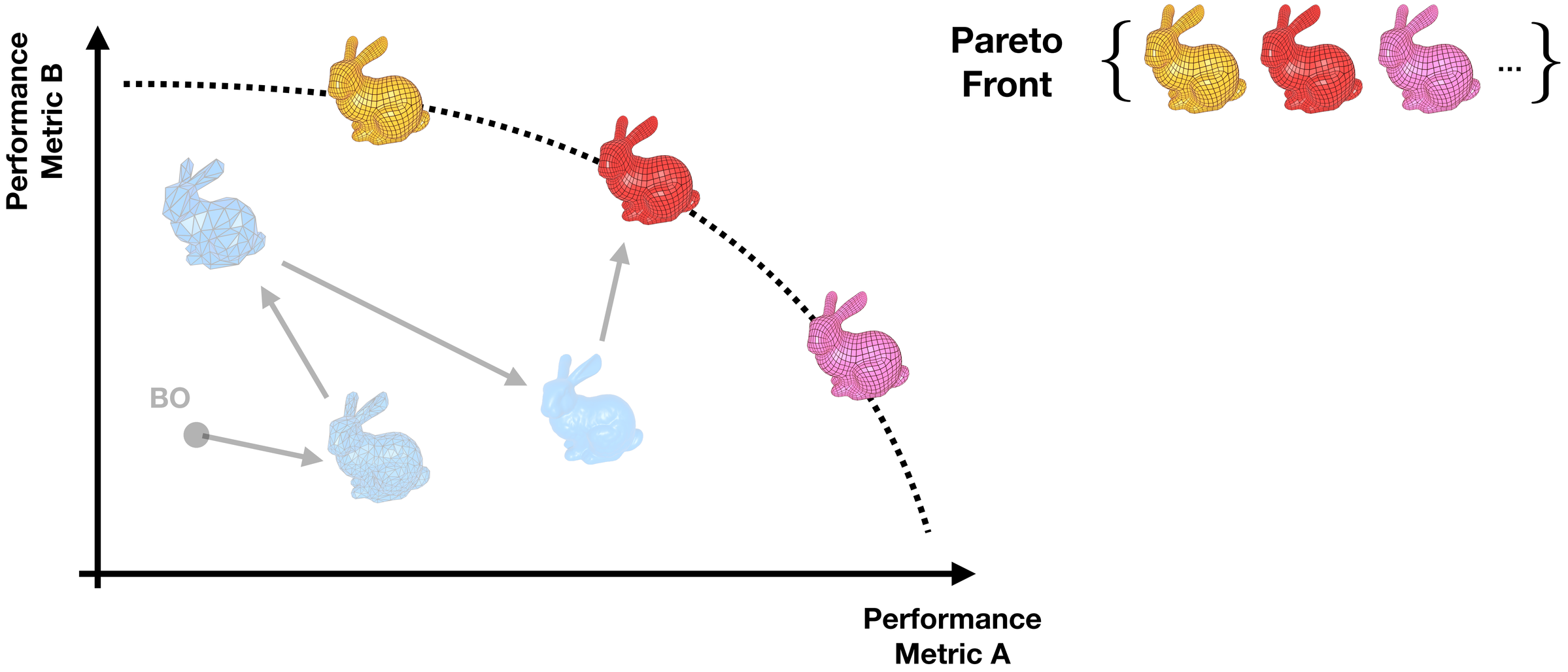
$$\mathrm{PM}_B(\ ) \geq \mathrm{PM}_B(\ )$$

$\Rightarrow$ dominants

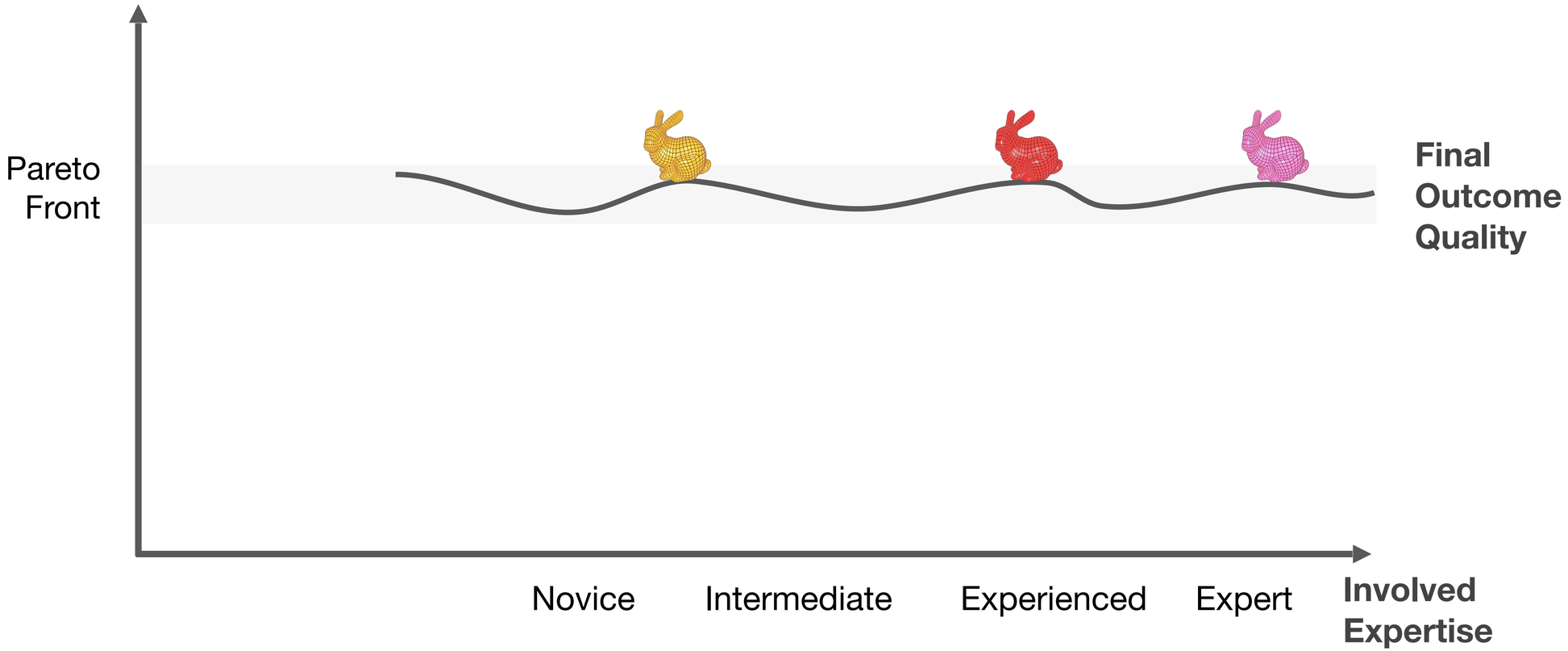
Pareto Optimality [Pareto, 1912]



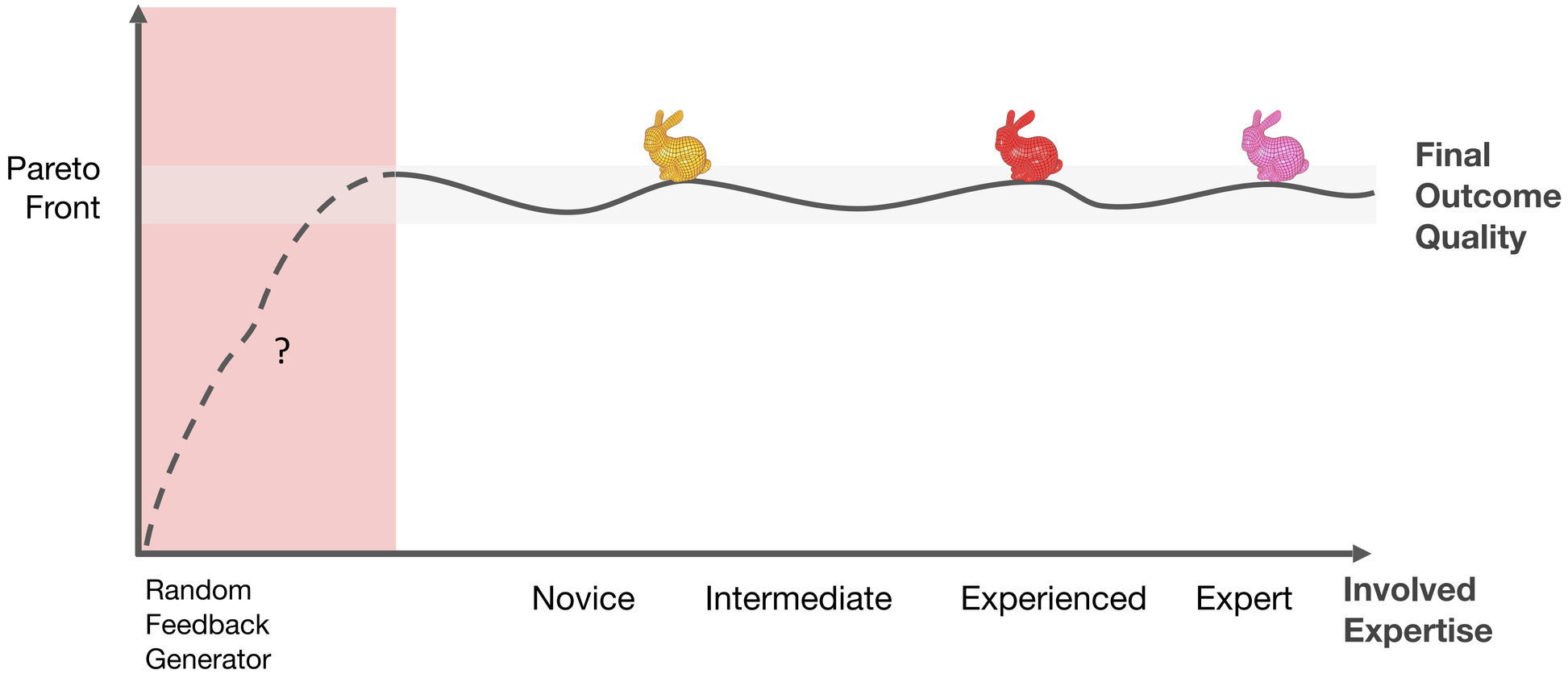
Pareto Optimality [Pareto, 1912]



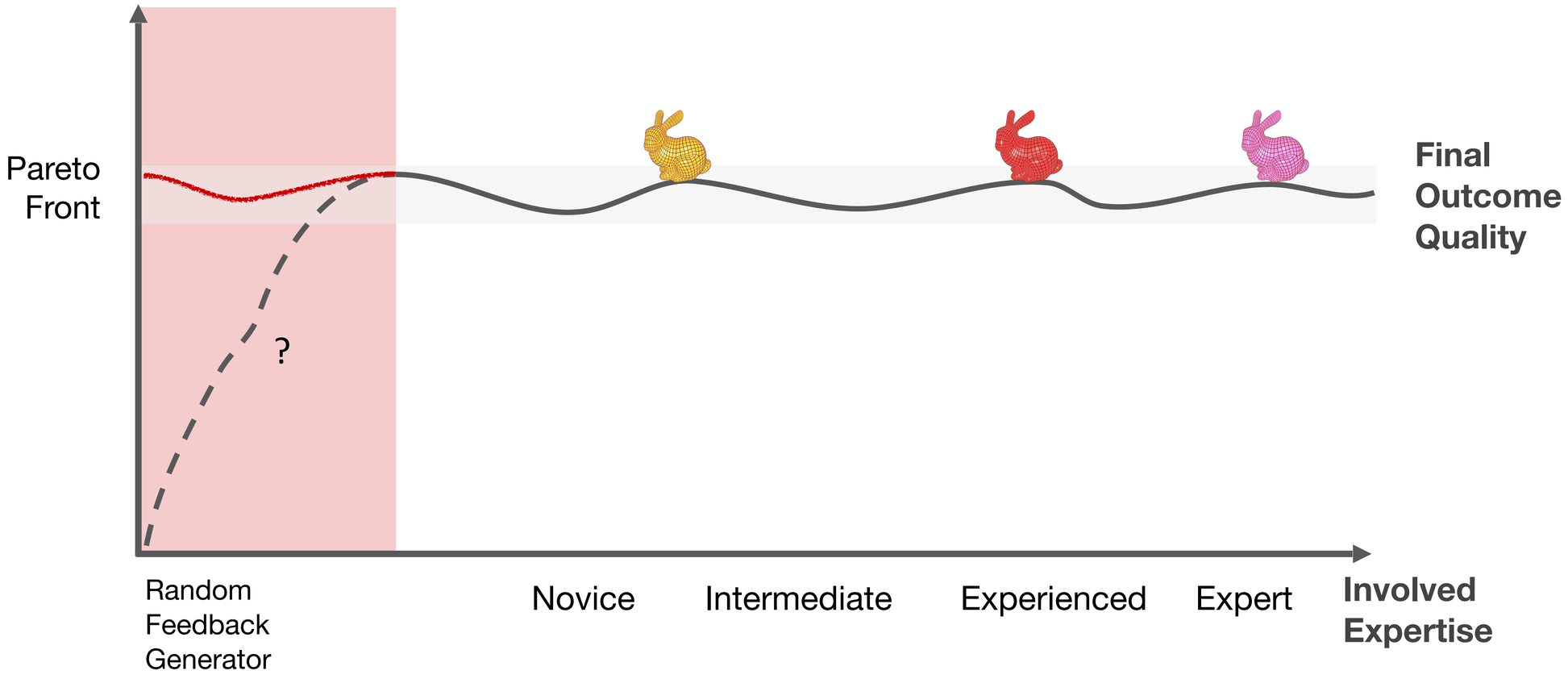
Implications on the Impact of Expertise in HITL



Implications on the Impact of Expertise in HITL

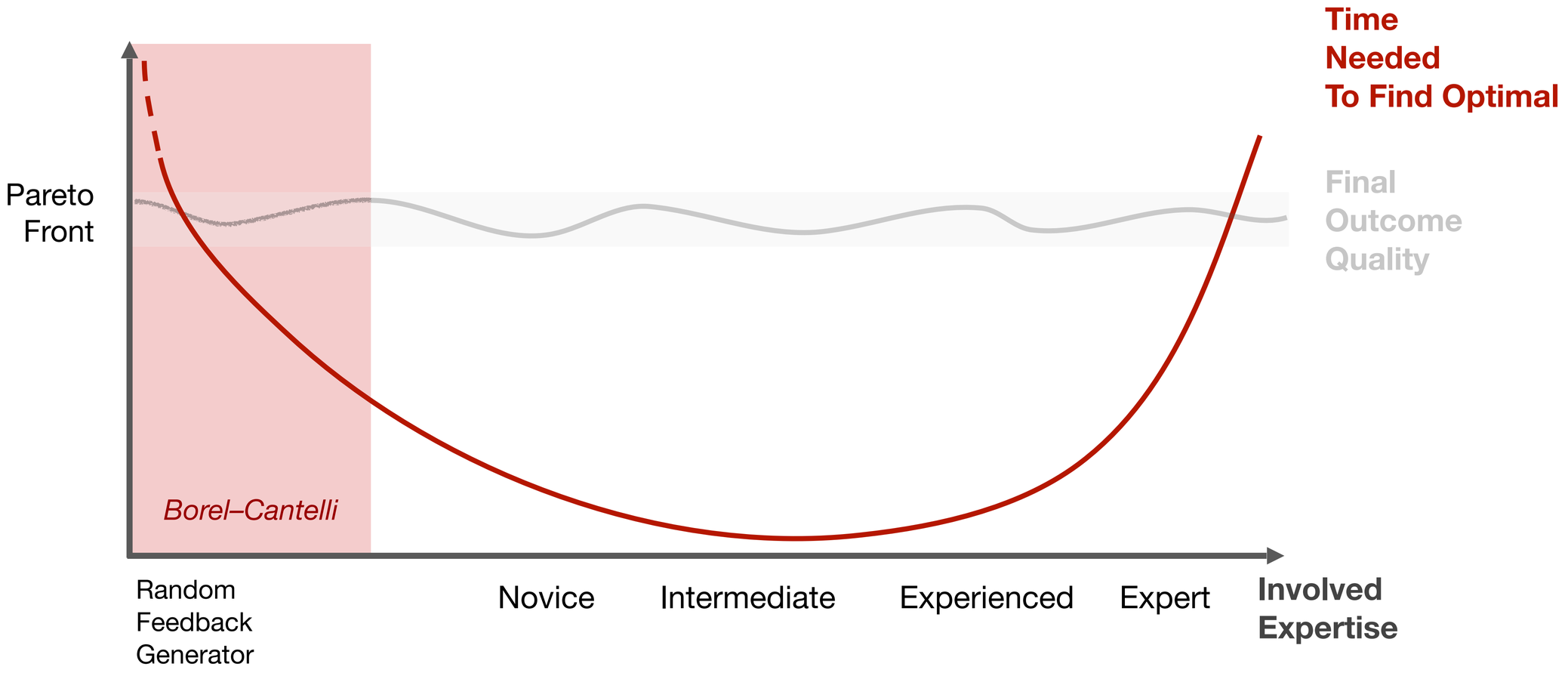


Implications on the Impact of Expertise in HITL

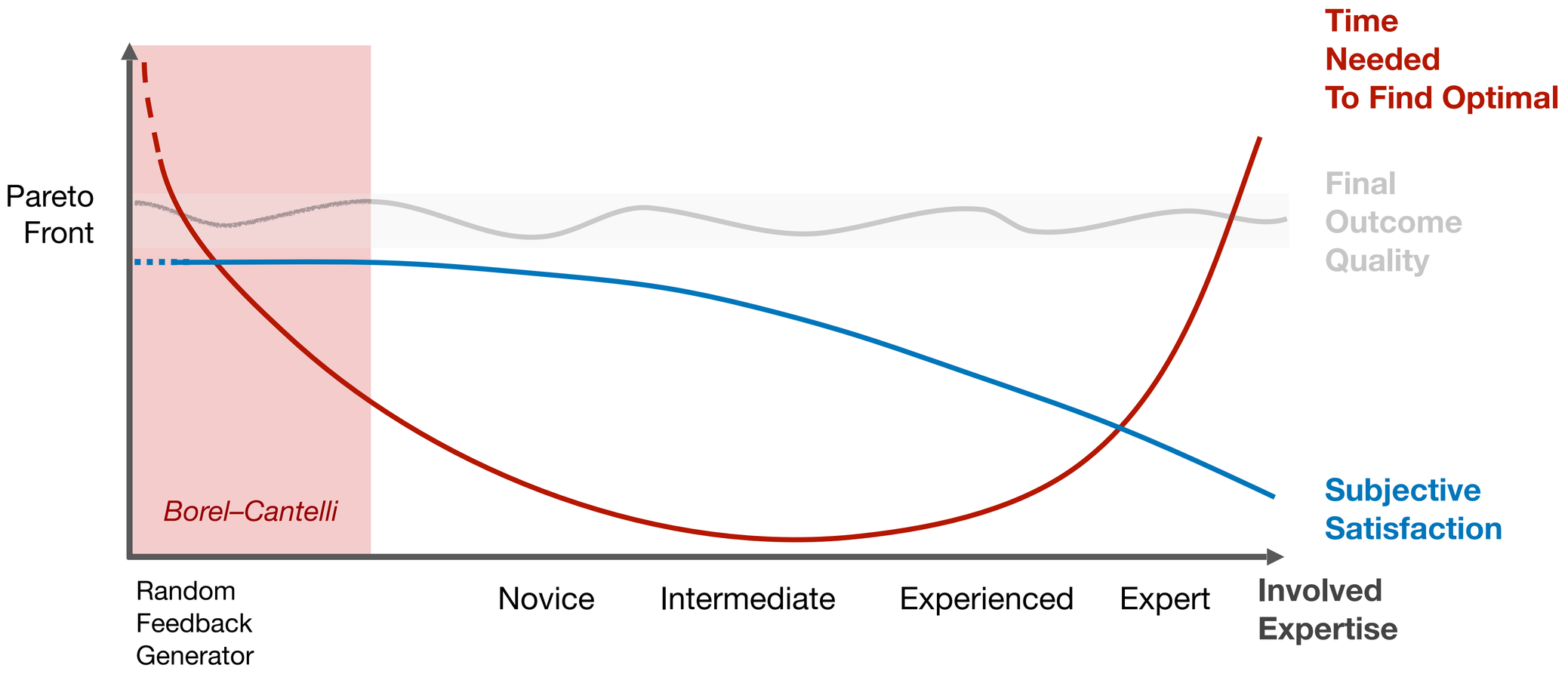


The Borel–Cantelli lemma [Borel 1909] [Cantelli 1917]
With infinite amount of events, the probability of observing any meaningful result is 1.0

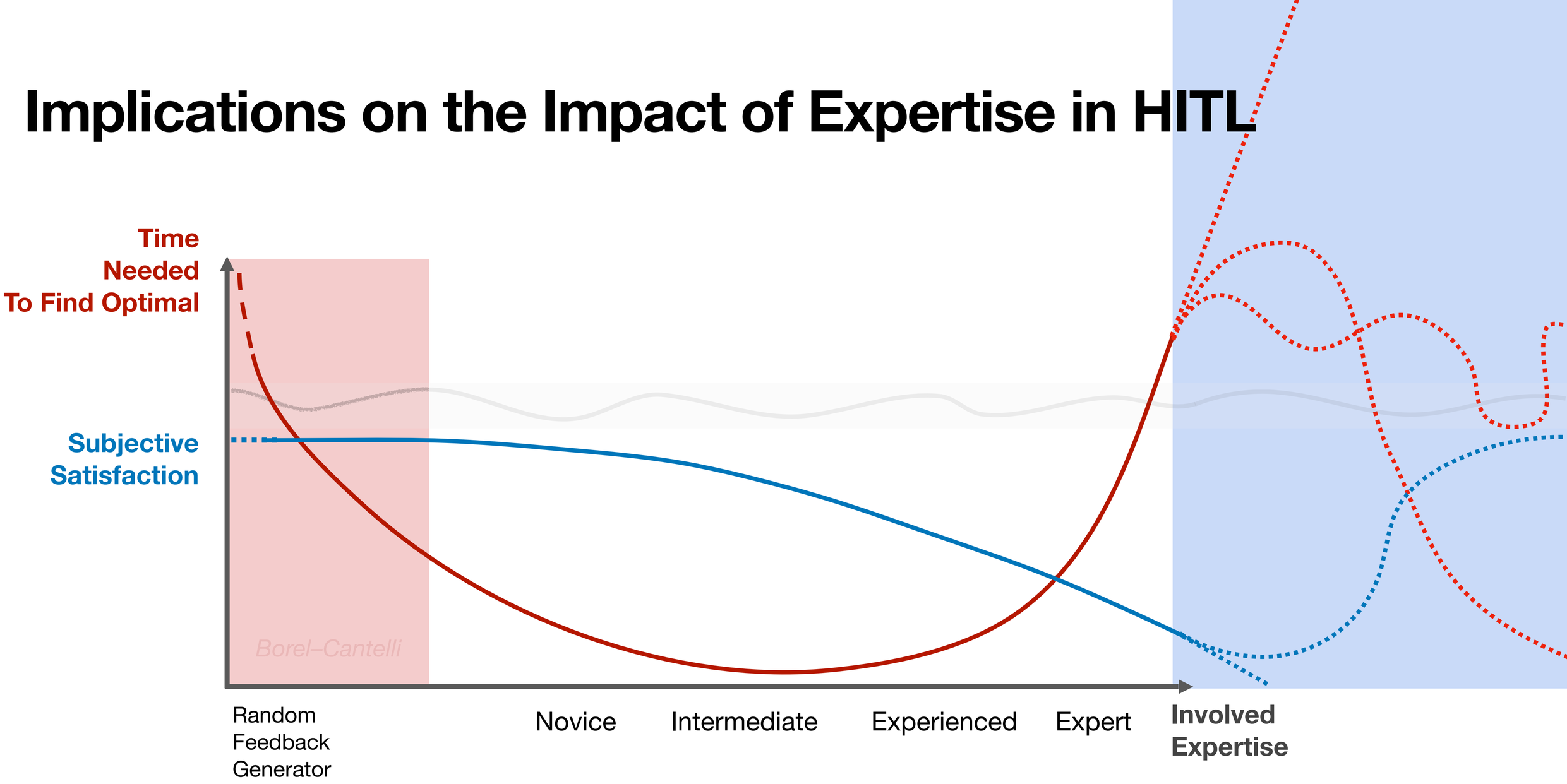
Implications on the Impact of Expertise in HITL



Implications on the Impact of Expertise in HITL



Implications on the Impact of Expertise in HITL



LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Media Informatics Group

The Impact of Expertise In the Loop for Exploring Machine Rationality

Changkun Ou, Sven Mayer, Andreas Martin Butz
Media Informatics Group
LMU Munich

ACM Intelligence User Interfaces 2023
Session 4
March 29, Sydney, Australia