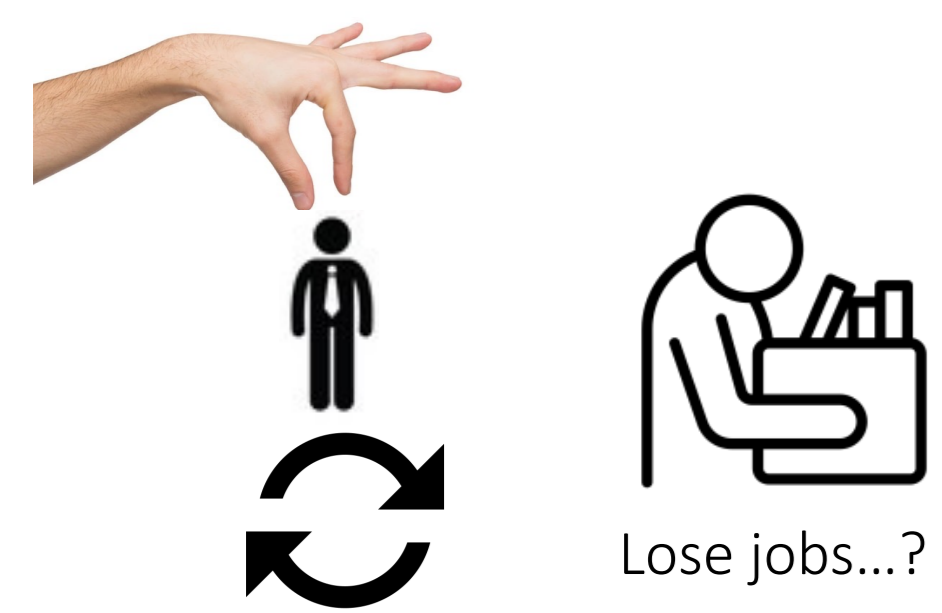


## Background: Human-AI collaboration

### What is the role of human experts in the era of AI?

If AI is a better designer of a product, will experts lose jobs?



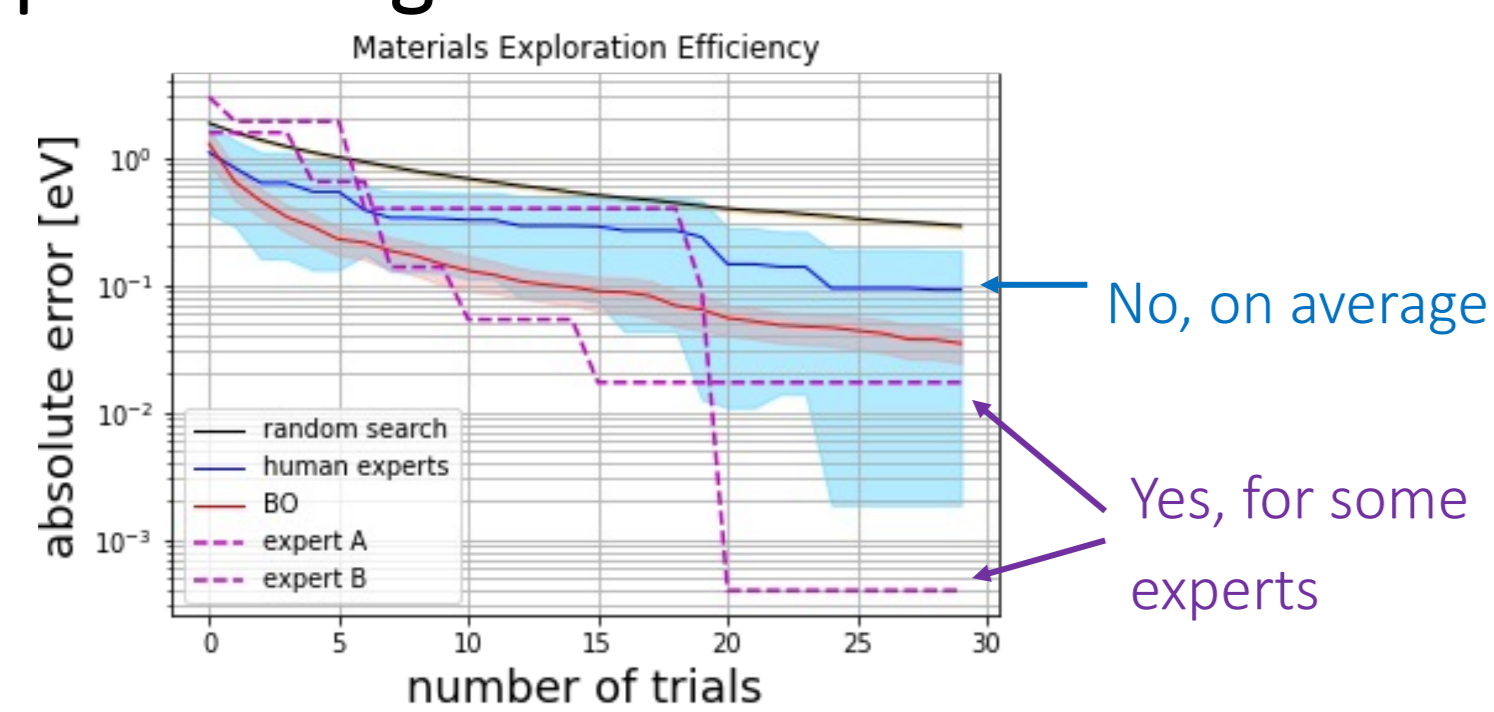
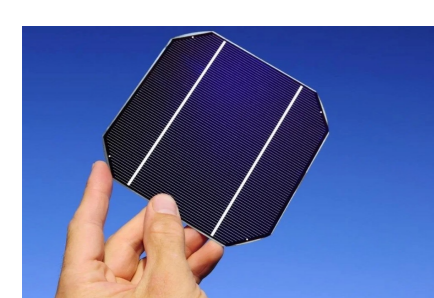
But you would think...

- You are better than baby AI.
- You can help AI.
- At least you wish to intervene in the decision-making process!

### Do human experts design better than AI?

Yes and No.

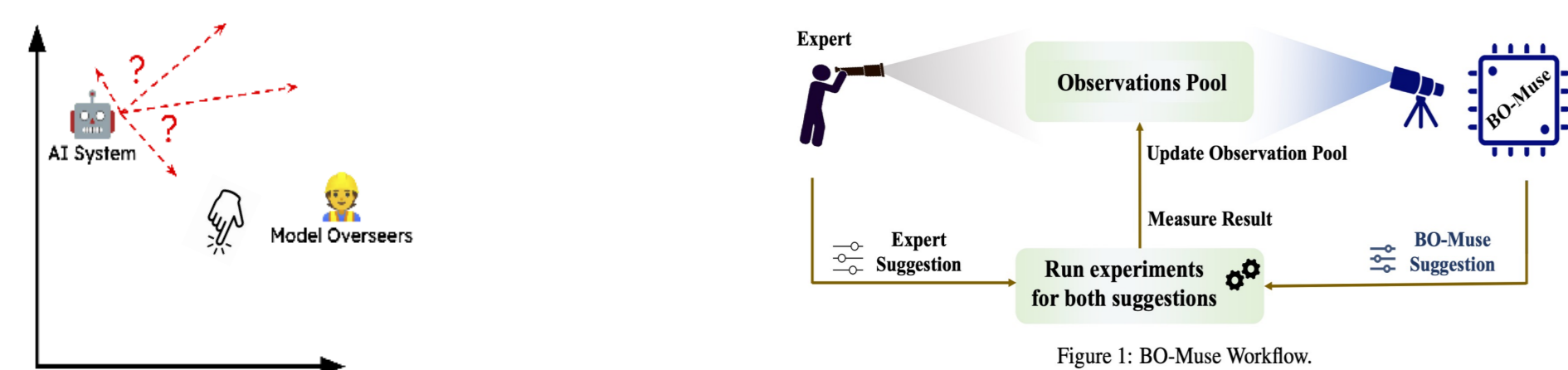
Finding the best solar cell materials<sup>[1]</sup>.



The quality of experts' advice varies.

### How can human experts help AI?

1. Human supervises BO<sup>[2]</sup>
2. Teaming up Human-BO<sup>[3,4]</sup>



Human rectifies the wrong BO suggestions. (But if human is such strong, why do we need BO?)

Both human and BO suggest the next points. (Human is just acting as a weaker BO, no communication.)

How can human experts collaborate with AI in a more balanced, theoretically grounded way??

## Paper, Code, and Reference

paper



code



### Reference

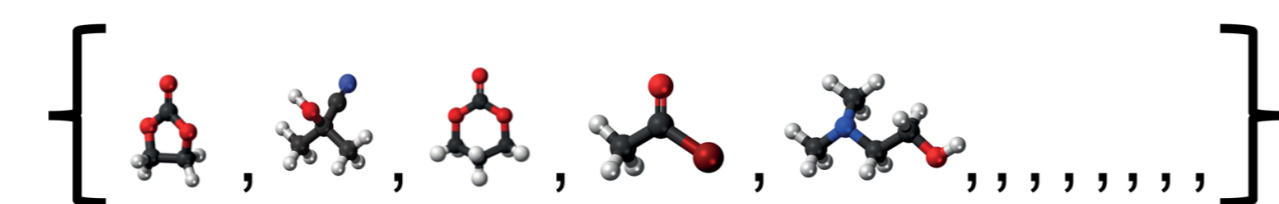
- [1] M. Adachi, NeurIPS ML4Physics (2021)
- [2] A. K. AV, et al., NeurIPS (2022)
- [3] S. Gupta, et al., arXiv:2303.01684 (2023).
- [4] A. Khoshvishkaie, et al., KDD (2023)

## Collaborative and Explainable BO (CoExBO)

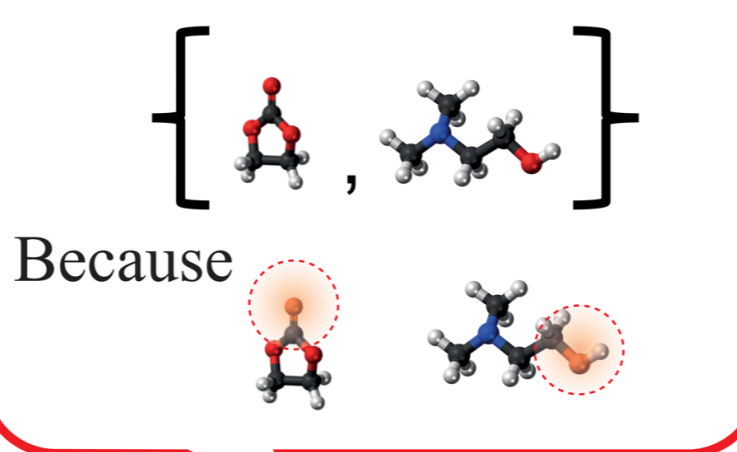
### Collaborative and Explainable BO (CoExBO)

1. BO combines experimental results and expert preferences.
2. BO generates pairwise candidates along with explanations.
3. Human interprets the acquisitions and picks their preferred one.
4. Human conducts experiments and repeat step 1.

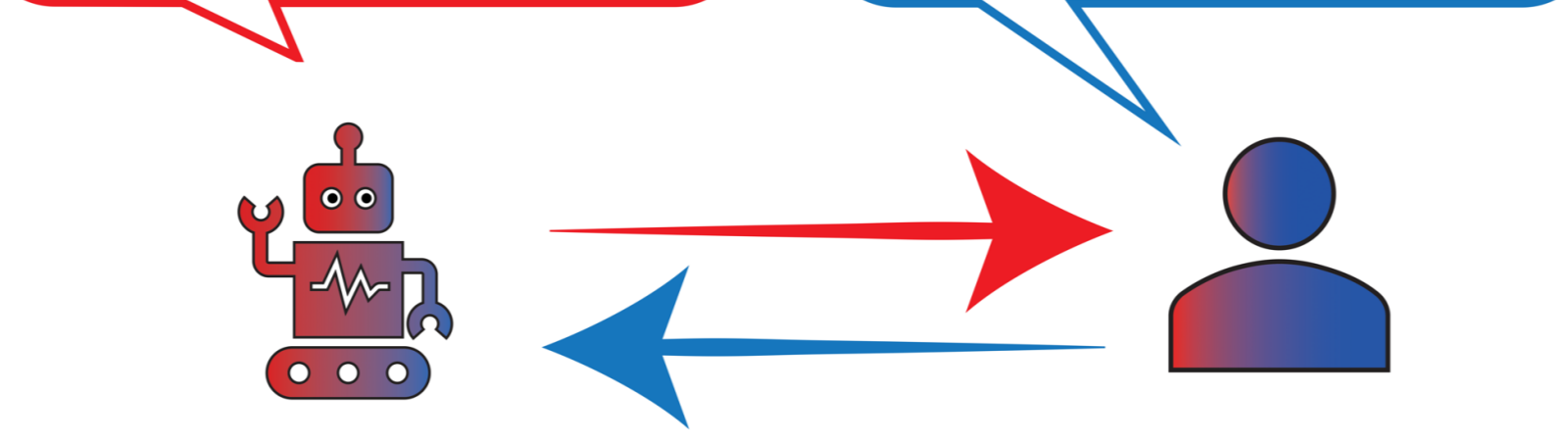
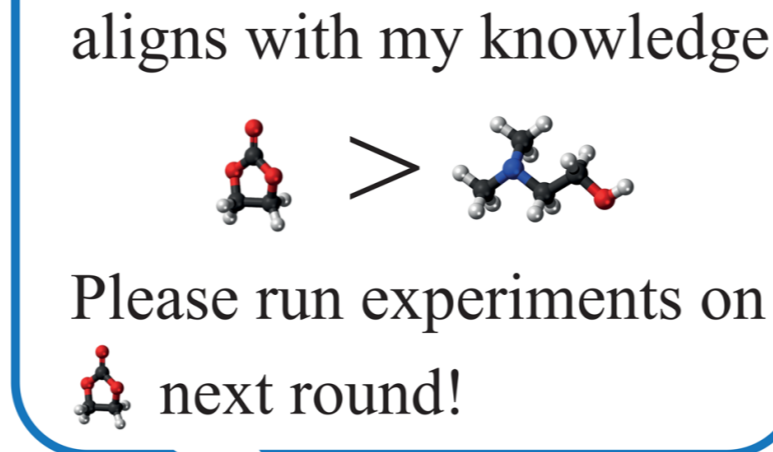
Find the best electrolyte material from the below:



I generated the candidates

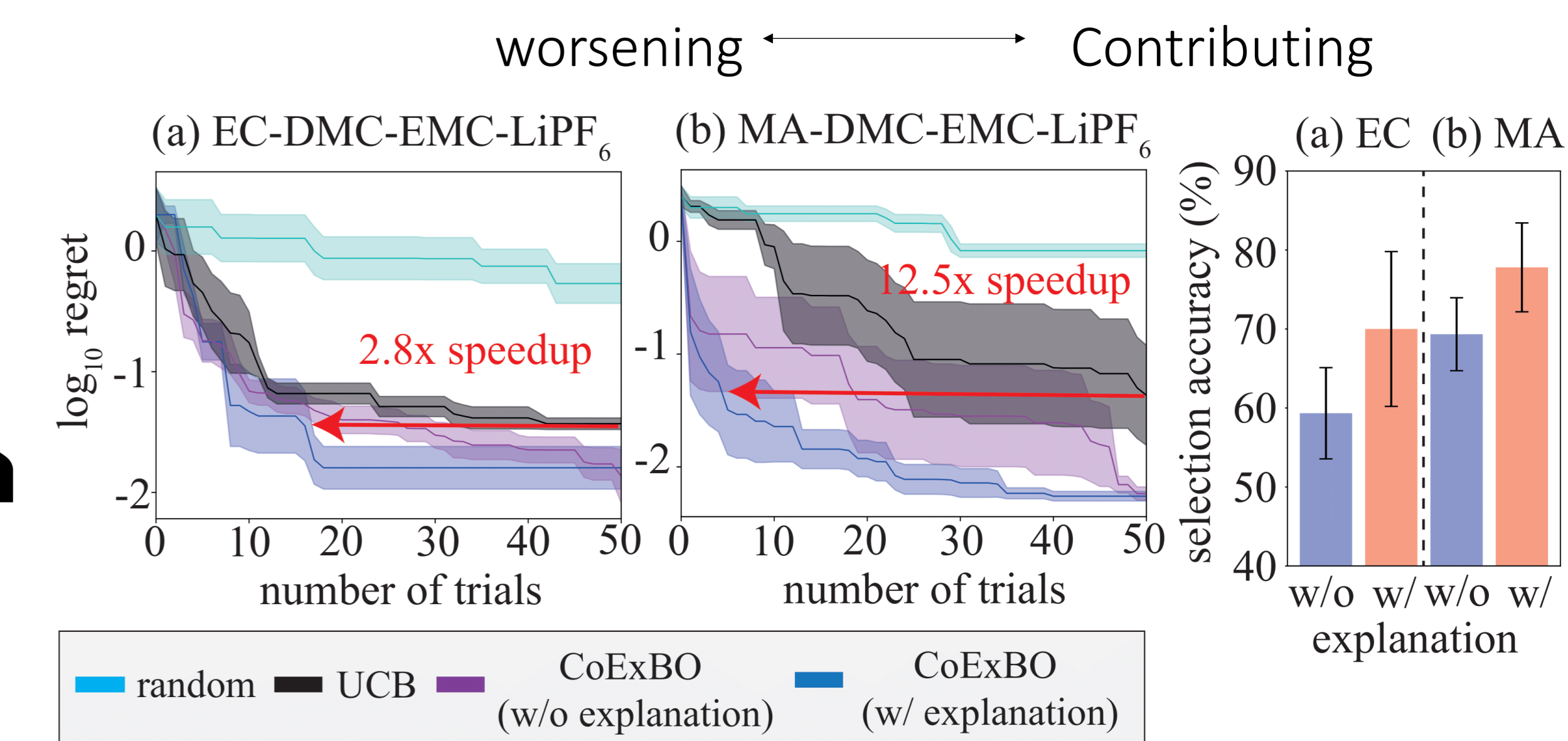
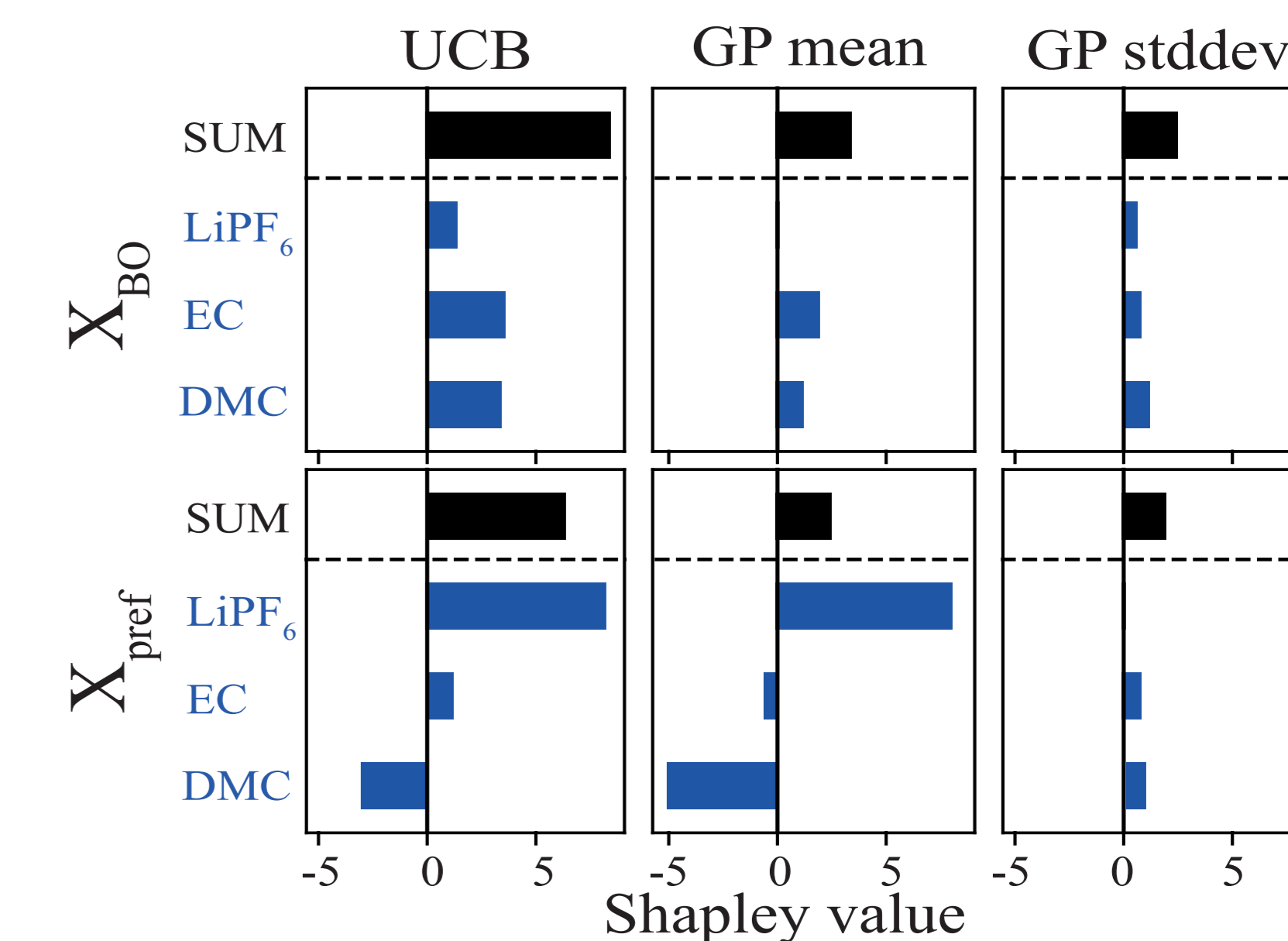
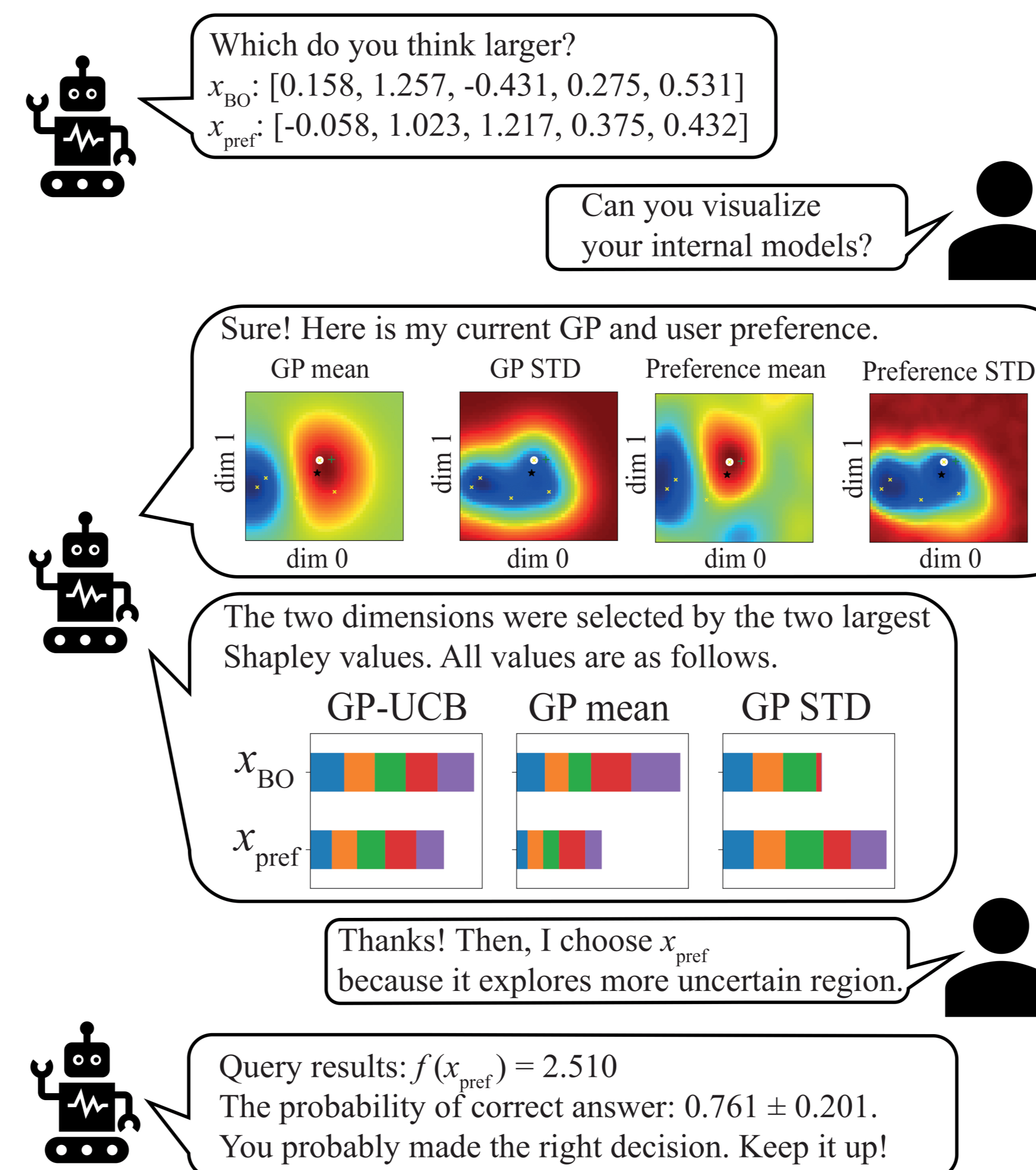


Ah carbonyl group, that aligns with my knowledge



Facilitate mutual understanding

### Explainability

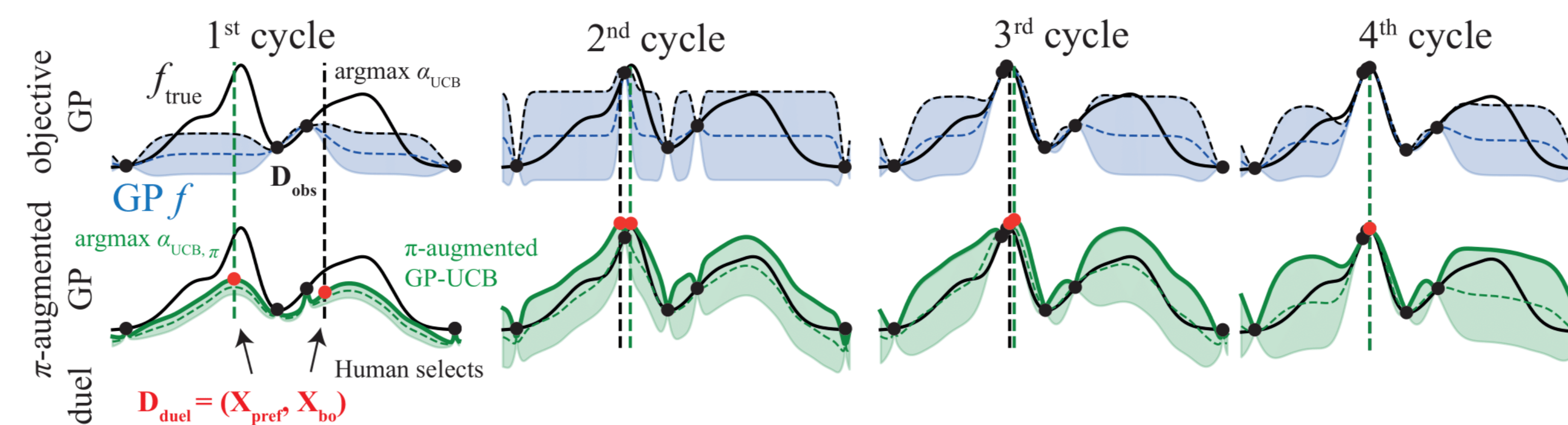


## Algorithms and Guarantee

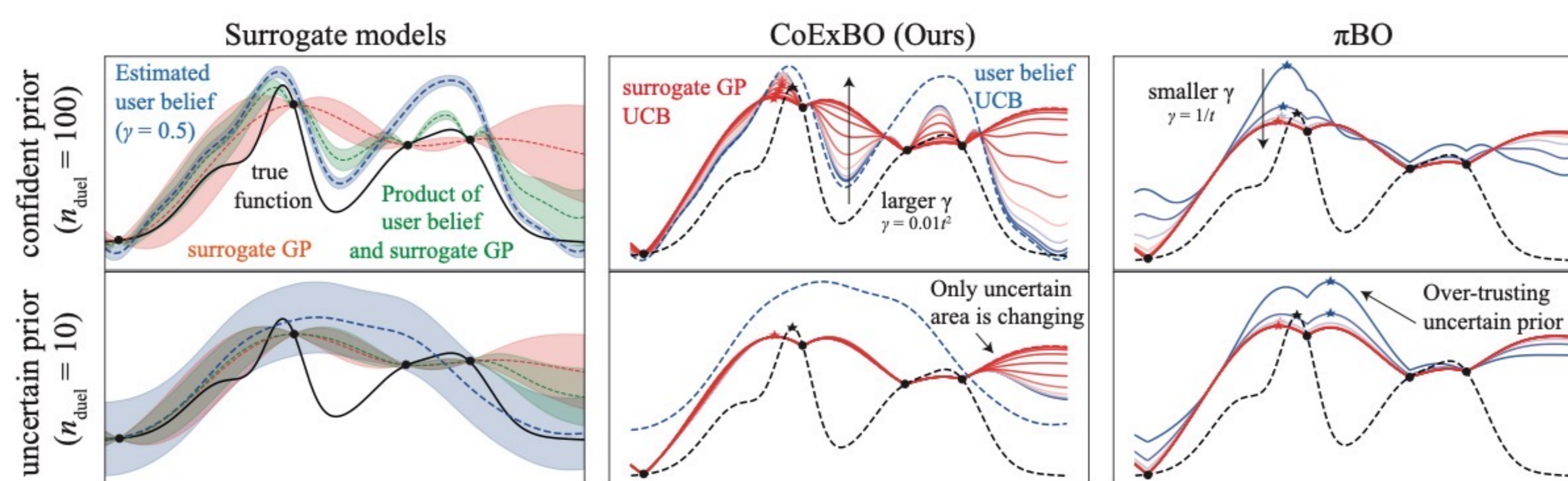
### Pairwise candidate generation

$$x_1 = \arg \max_{x \in \mathcal{X}} \alpha_{f_t}(x) \quad (\text{standard UCB})$$

$$x_2 = \arg \max_{x \in \mathcal{X}} \alpha_{f_t, \hat{\pi}_t}(x) \quad (\hat{\pi} \text{ incorporated UCB})$$



### Augment Gaussian process with experts knowledge



### Acceleration Guarantee

If human suggestions are mostly correct, it will accelerate convergence.

**Theorem (Acceleration guarantee)**  
 If augmented GP satisfies  $|f(x) - \mu_{f_t, \pi_t}(x)| \leq \beta_t^{1/2} \sigma_{f_t, \pi_t}(x)$ , i.e.  

$$\frac{r_{\text{pref}}}{r_{\text{BO}}} < 1$$
 where  $r$  is the regret.

### No-Harm Guarantee

Even if human suggestions are completely wrong, it won't be worse than vanilla BO.

**Proposition (No harm guarantee)**  
 Regardless of whether users provide informative or non-informative preferences, CoExBO asymptotically reaches the standard UCB sub-linear convergence rate, i.e.  

$$\lim_{t \rightarrow \infty} \frac{r_{\text{pref}}}{r_{\text{BO}}} = 1$$

