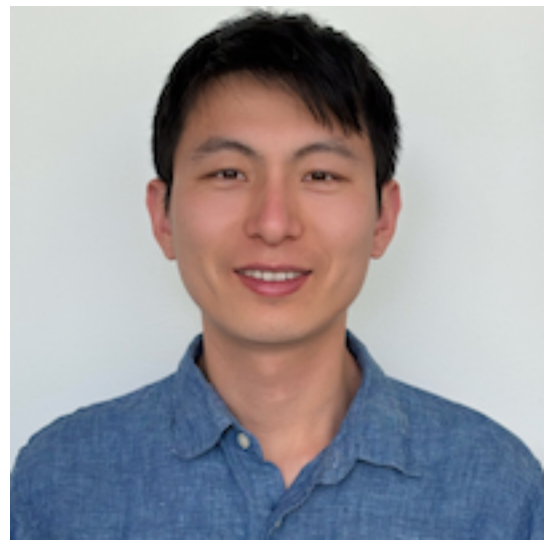
Enhancing Preference-based Linear Bandits via Human Response Time

Shen Li*[1], Yuyang Zhang*[2], Zhaolin Ren[2], Claire Liang[1], Na Li[2], Julie A. Shah[1]






[1] MIT   [2] Harvard

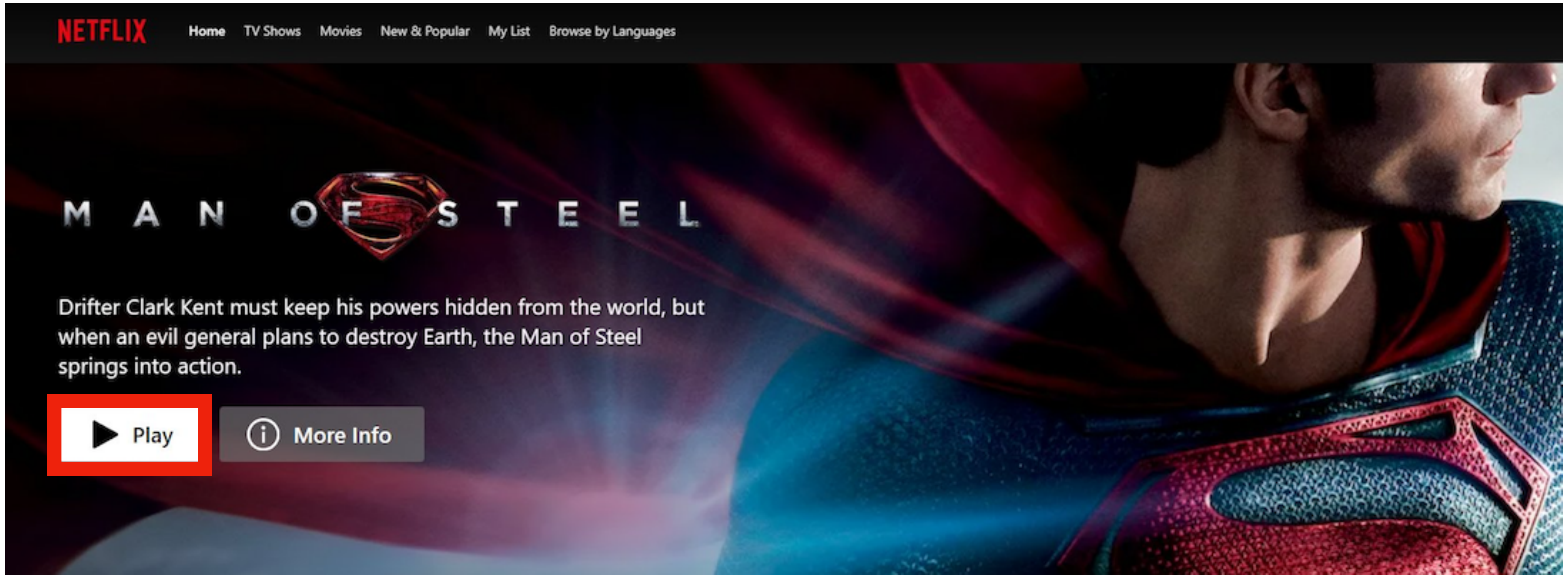
* First two authors have equal contribution

MIT CSAIL

Harvard John A. Paulson School of Engineering and Applied Sciences

NEURAL INFORMATION PROCESSING SYSTEMS

Binary Choices Are Widely Used to Align AI Systems With Human Preferences

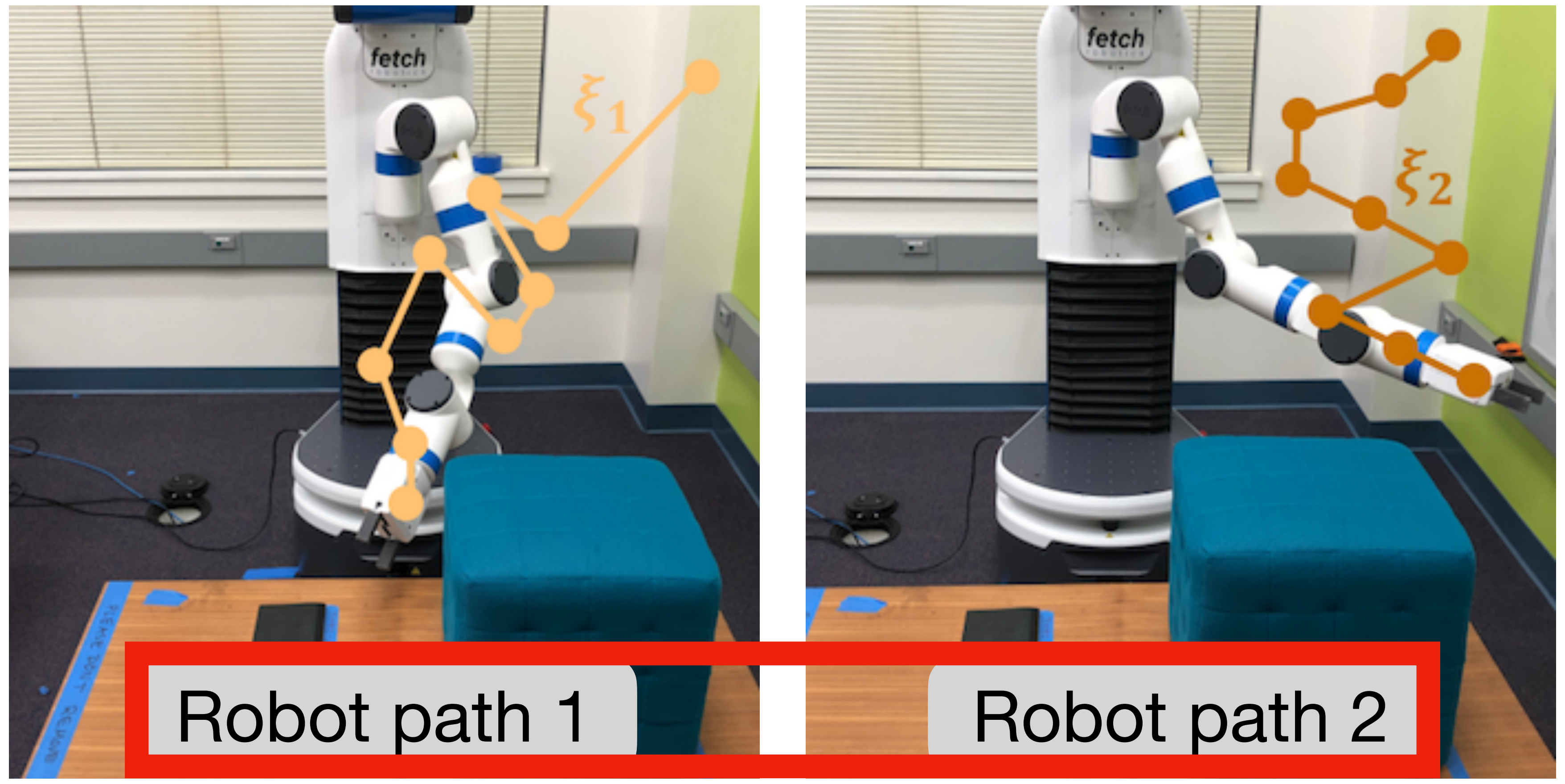


Q: why do our eyebrows and eyelashes stop growing after a certain length?

A: Hair has different cycles[1]. It grows for a certain amount of time before falling out and starting fresh. For eyelashes, the cycle is much shorter (a few weeks) whereas eyebrows grow more slowly and have a cycle closer to two months. The hair on your head has a cycle of several years, allowing it to grow much longer. Some people have longer arm or leg hair because they have a faster speed of growth or a longer cycle. Look it up, it's really interesting

[1] https://www.philipkingsley.co.uk › hair-science › hair-g.

Plausible answer? ○ Yes ○ No ○ Not sure



(Bıyık et al., 2022)

Choices Provide Limited Information About Preference Strengths. Our Work Resolves This by Incorporating Response Times.

Which one would you like during the poster session?

Long response time

Weak preference

Short response time

Strong preference

(Alós-Ferrer, et al., 2021)

Research Questions and Key Contributions

Q1. ***How*** to combine response times with choices in preference learning?

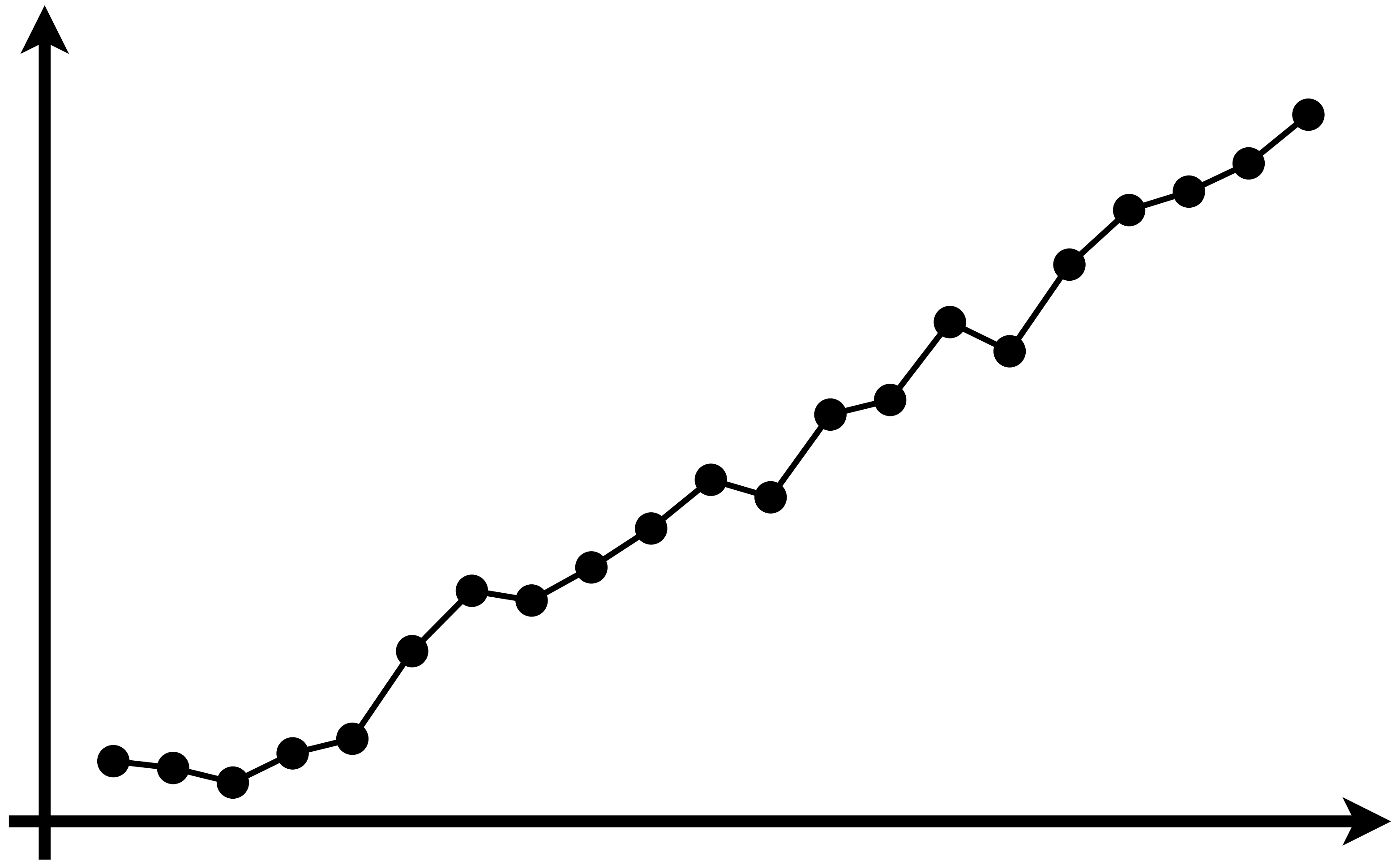
Response times

Choices

→ Preferences

Q2. ***When*** do response times improve preference learning?

Improvement due to using response times

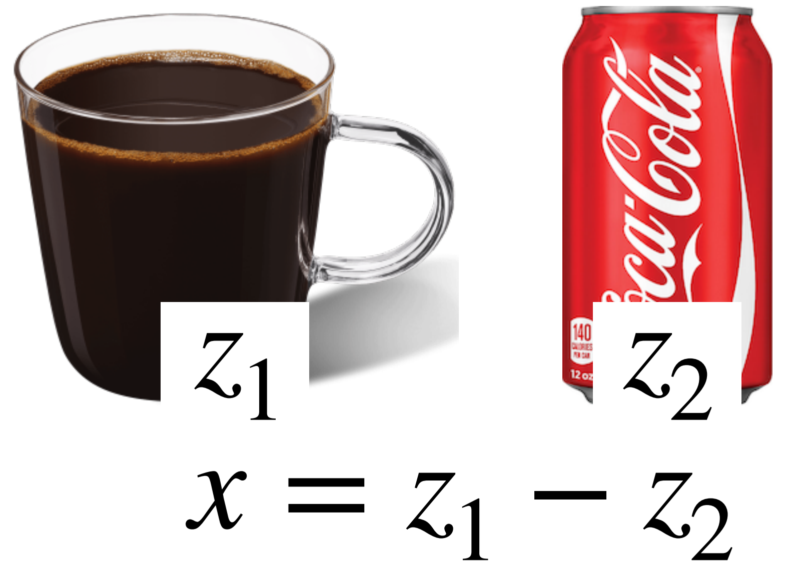


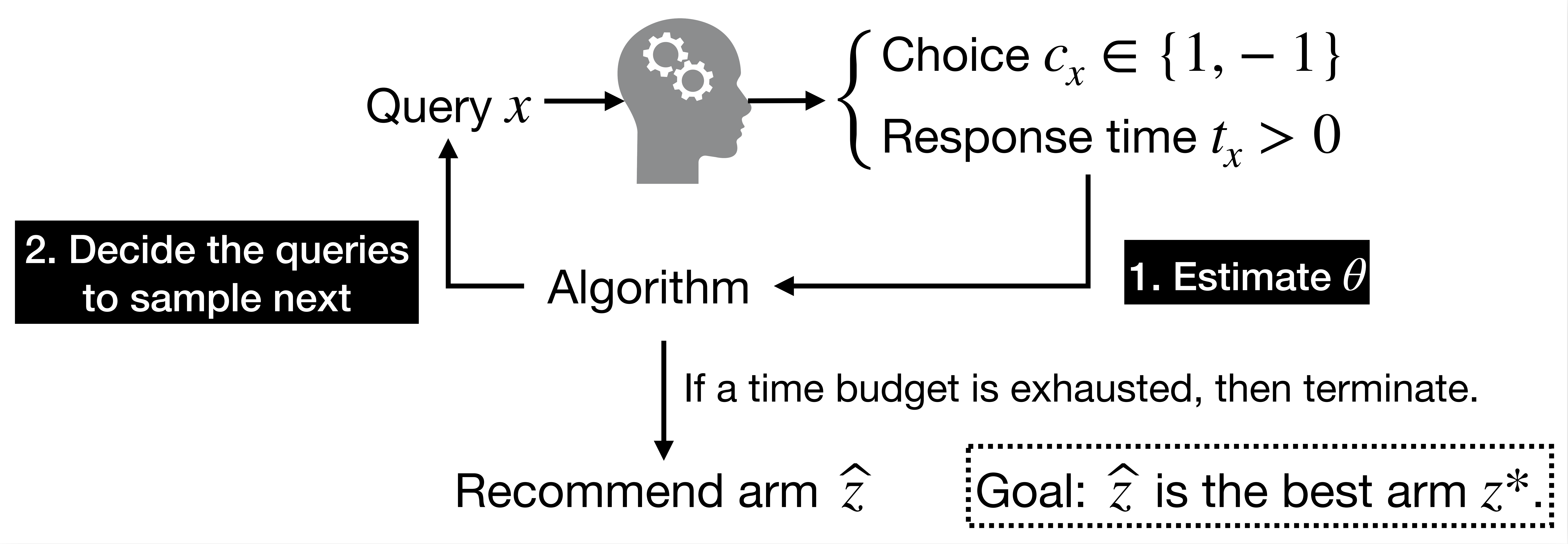
Human pref are weak

Human pref are strong

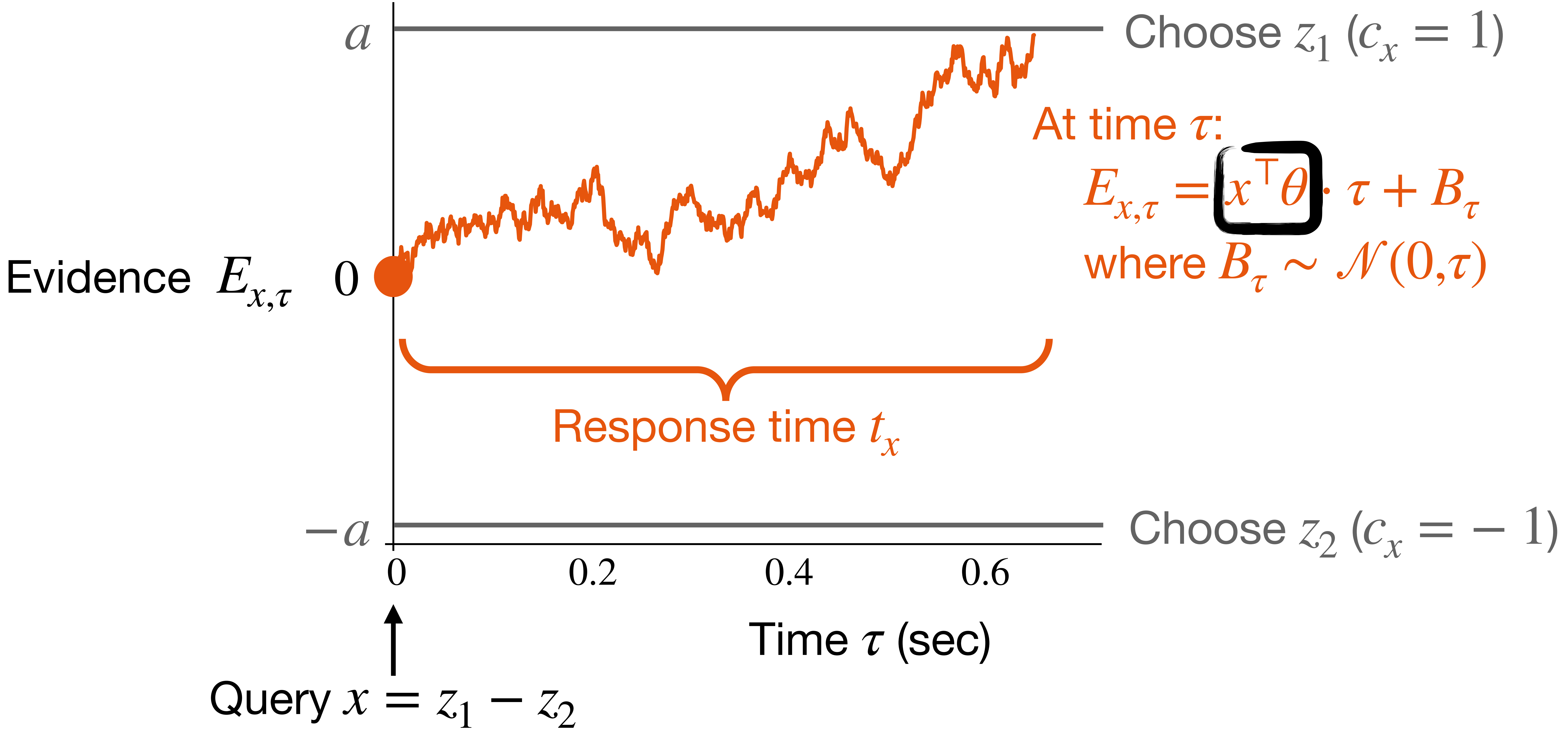
Problem Formulation: Linear Bandit With Binary Choice and Response Time Feedback

- Human preference: $\theta \in \mathbb{R}^d$
- Each arm: z with a utility $z^\top \theta$
- Each query: $x := z_1 - z_2$ with a utility difference $x^\top \theta$



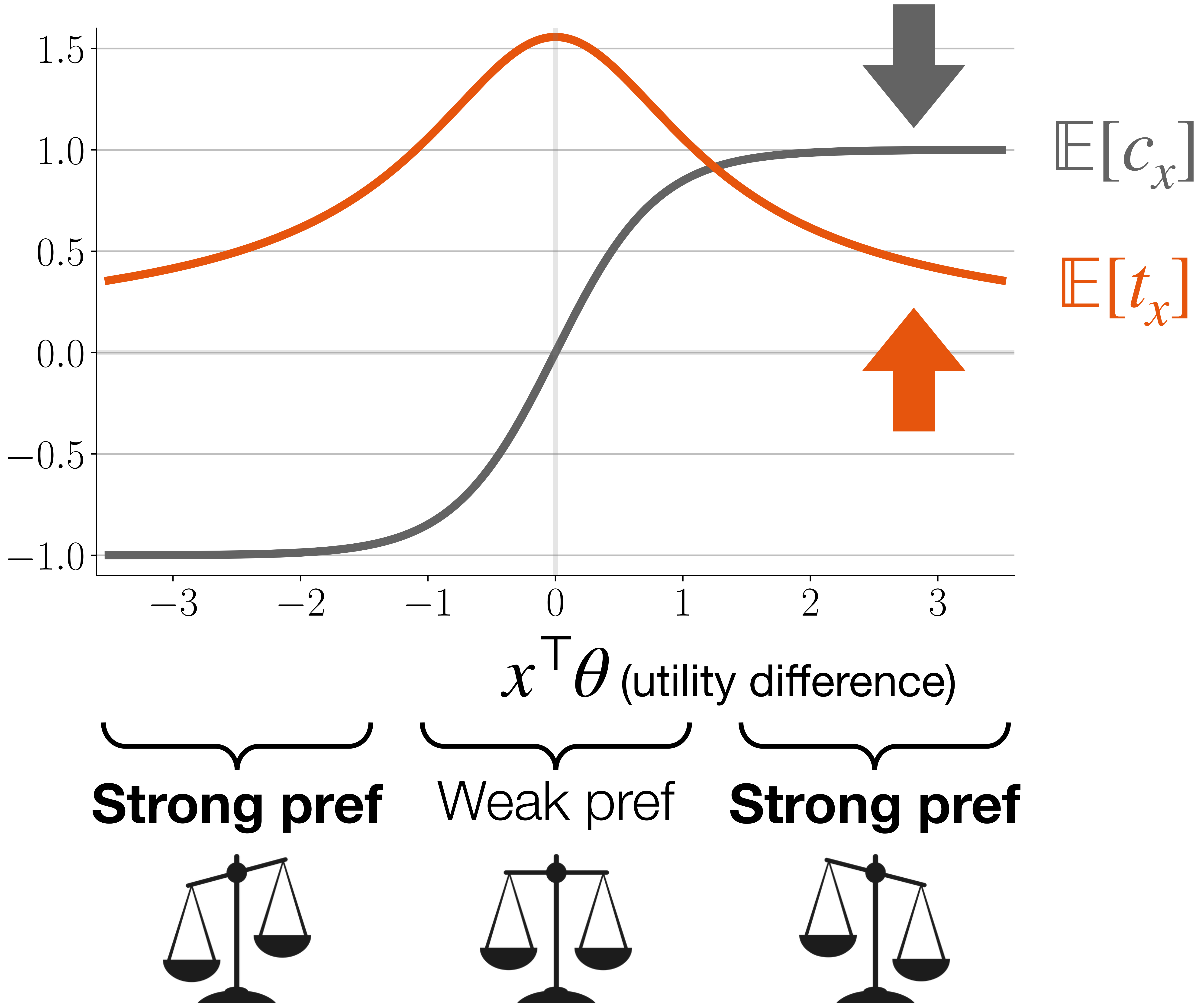


EZ-Diffusion Model Links Utility Differences, Choices, and Response Times



Wagenmakers et al., 2007.

The Magnitude of Utility Difference Is Proportional to Expected Choice and Inversely Proportional to Expected Response Time



Our Novel θ Estimator Uses Choices and Response Times, While Prior Methods Only Use Choices

Given a dataset with i.i.d. choices and response times from various queries:

Our estimator uses both choices and response times:

Input *Coefficient* *Target (estimated from data)*

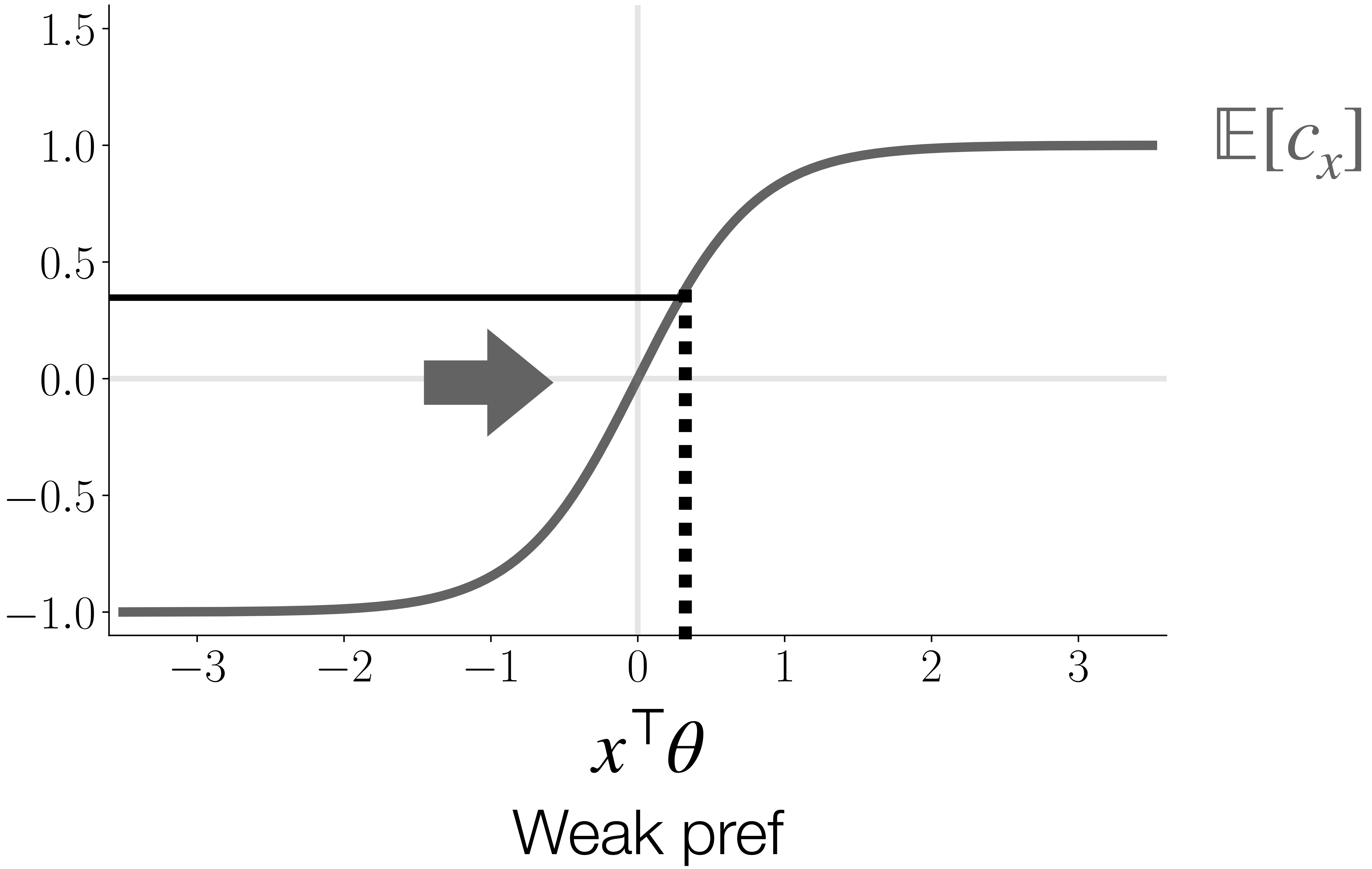
$$x^\top \frac{\theta}{a} = \frac{\mathbb{E}[c_x]}{\mathbb{E}[t_x]} \Rightarrow \hat{\theta}_{\textit{choices, times}} \leftarrow \text{linear regression with squared loss}$$

Prior methods only use choices:

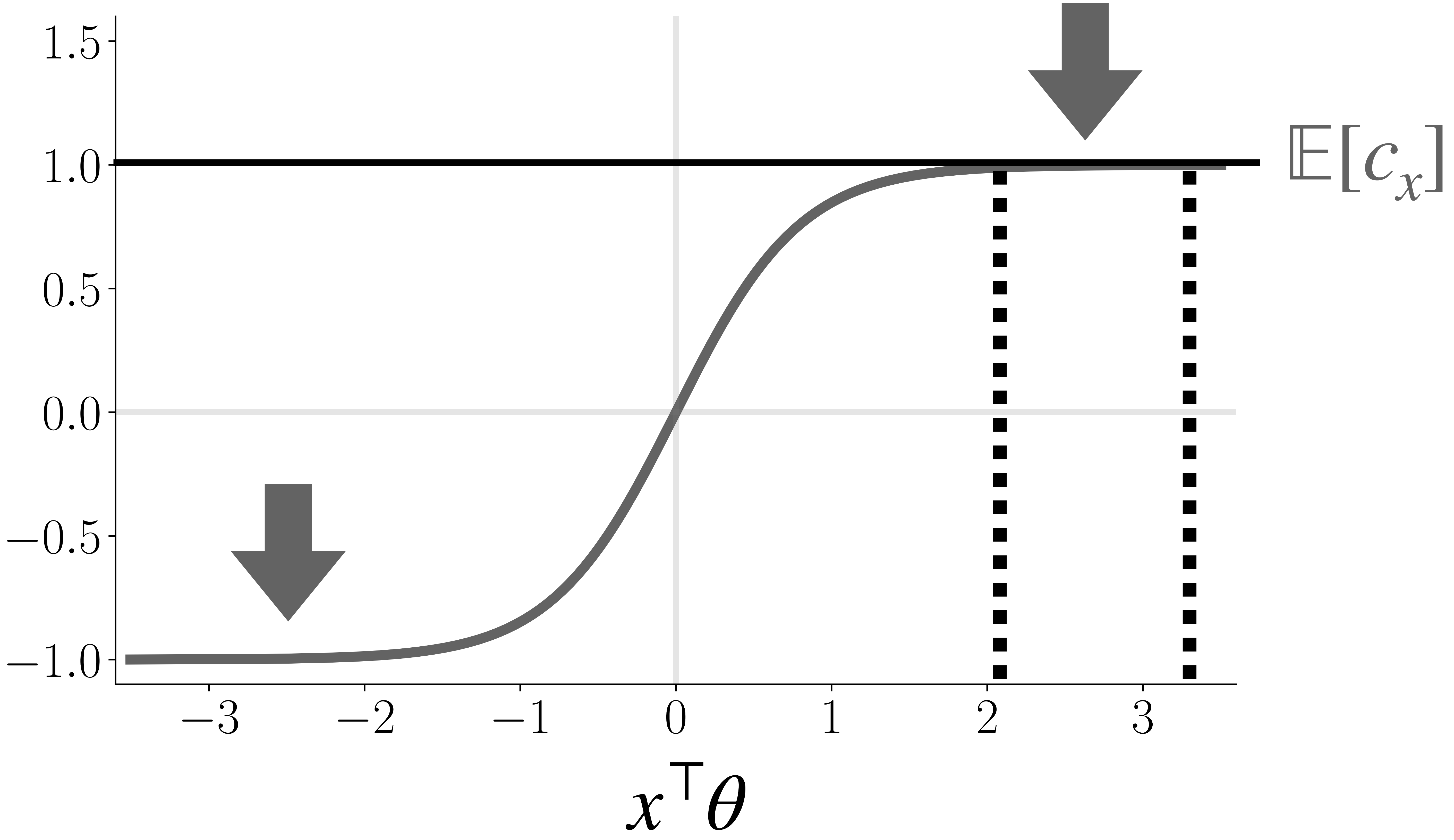
$$\mathbb{P}\left[c_x = 1\right] = \frac{1}{1 + \exp\left(-c_x \cdot x^\top \cdot 2a\theta\right)} \Rightarrow \hat{\theta}_{\textit{choices}} \leftarrow \text{logistic regression}$$

(Same as the Bradley Terry model)

Intuitively, for Queries With *Strong* Preferences, Response Times Provide Information That Complements Choices

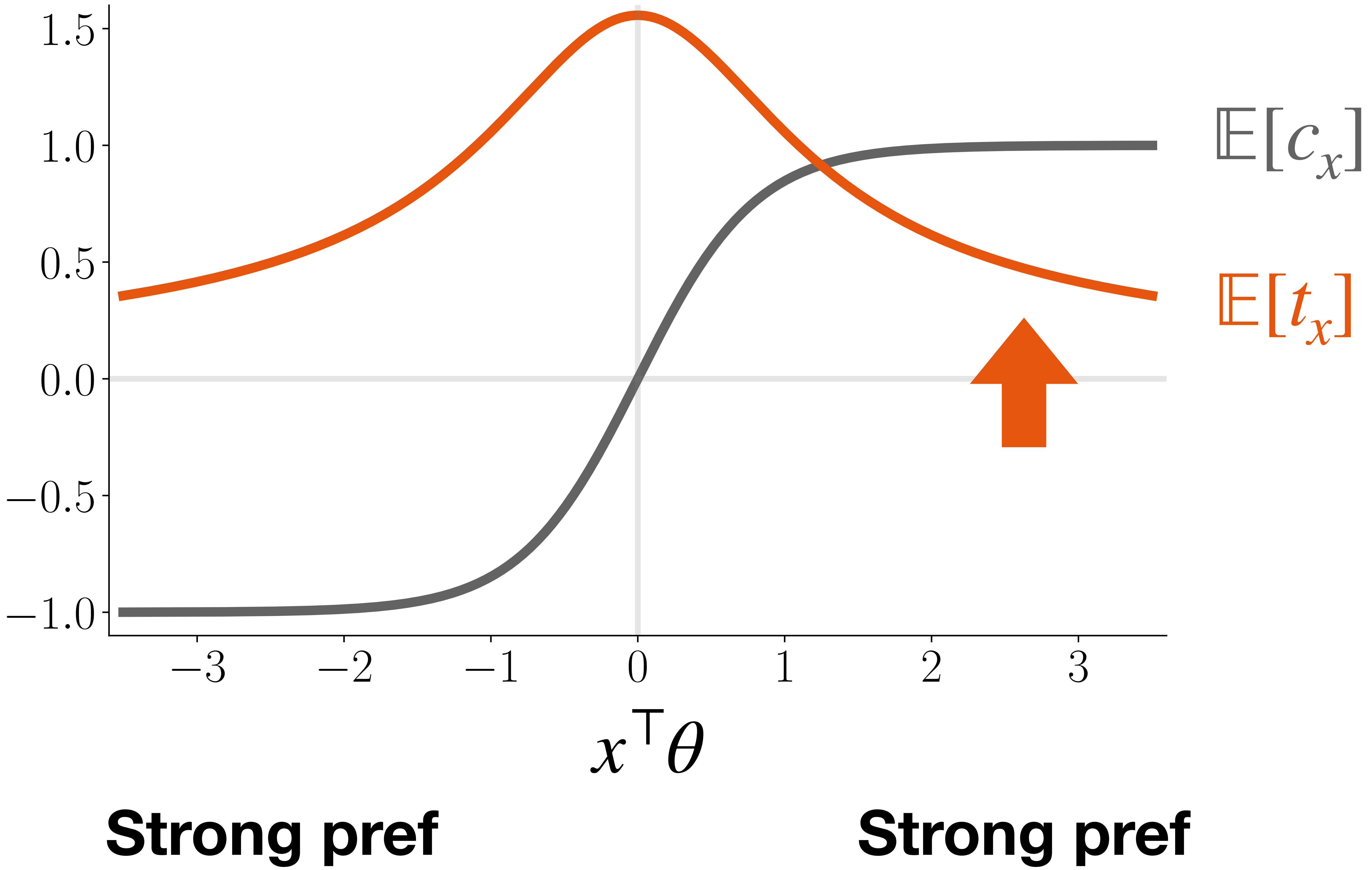


Intuitively, for Queries With *Strong* Preferences, Response Times Provide Information That Complements Choices



Strong pref **Strong pref**

Intuitively, for Queries With *Strong* Preferences, Response Times Provide Information That Complements Choices

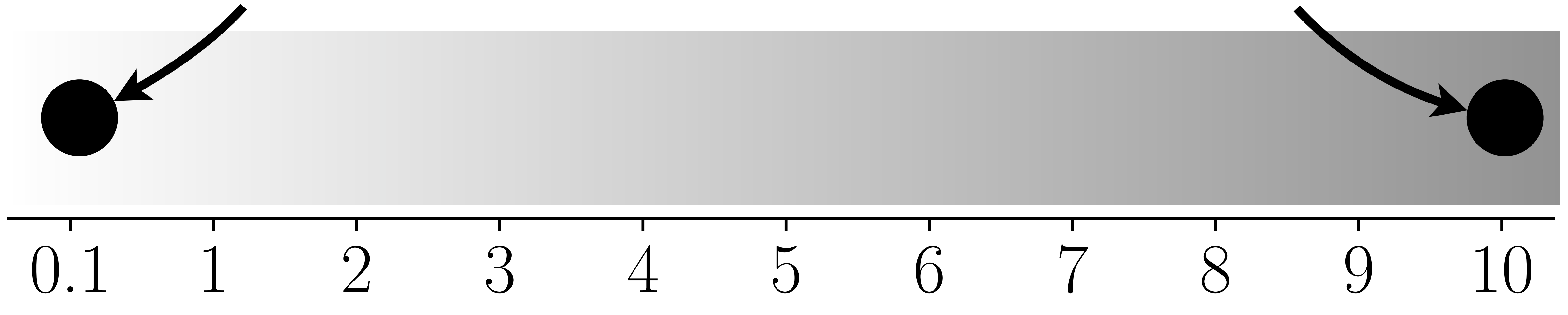


To Verify This Insight, We Use a Synthetic Problem to Compare Estimator Performance

A bandit problem with weak pref

A bandit problem with strong pref

Scaling factor c that scales each arm z to $c \cdot z$



1. Computes an experimental design:

....

2. Sample 50 queries and gather feedback

....

0.5 sec

3. Estimate $\widehat{\theta}_{choices,\ times}$ or $\widehat{\theta}_{choices}$ and recommend $\widehat{z}$

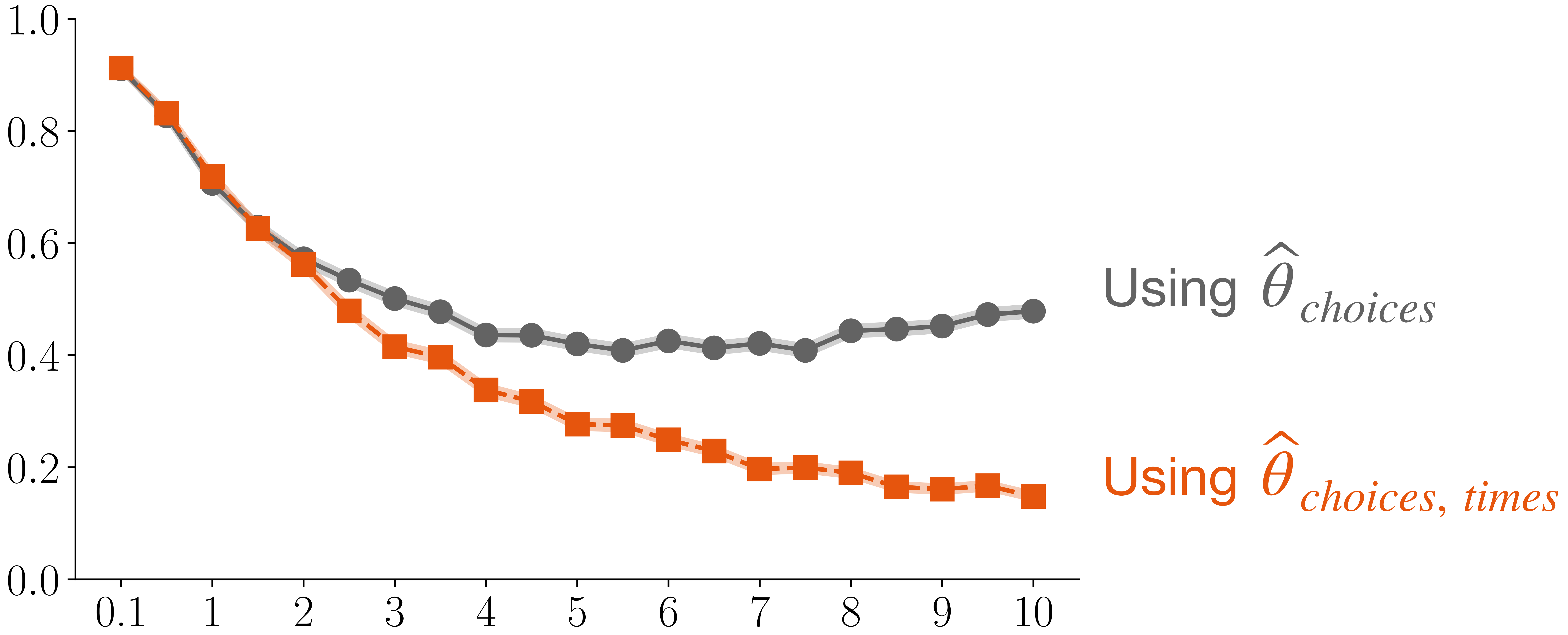
Performance measure: $\mathbb{P}\left[\widehat{z} \neq z^*\right]$

(The synthetic problem is from Tao et al., 2018)

Empirical Result Confirms That for Queries With *Strong* Preferences, Response Times Provide Information That Complements Choices

$\mathbb{P}\left[\hat{z} \neq z^*\right]$

The lower, the better



Scaling factor c that scales each arm z to $c \cdot z$

Problems with weak pref

Problems with strong pref

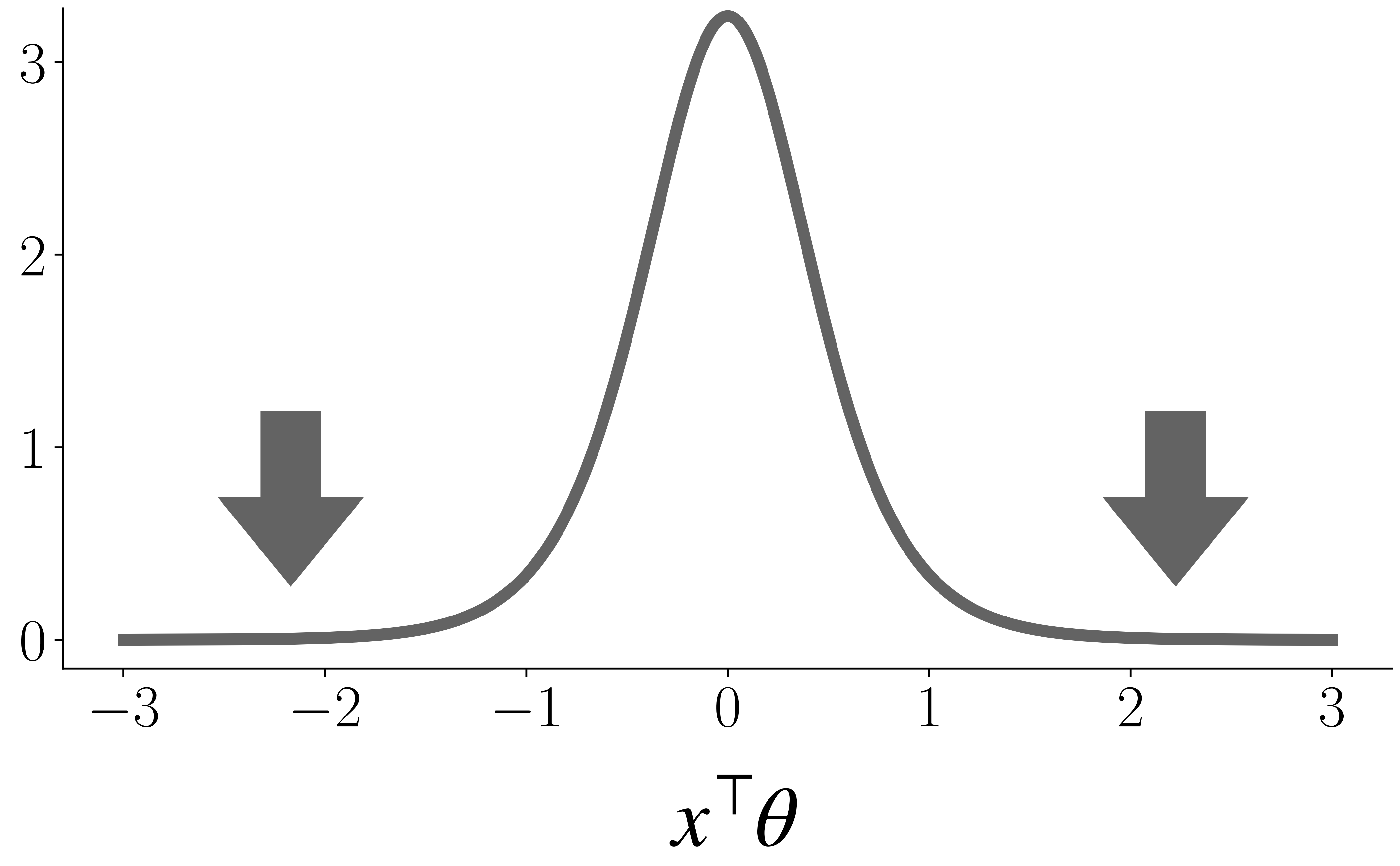
Asymptotic Variances Shows That for Queries With *Weak* Preferences, Choices Provide a Limited Amount of Information

Given a fixed dataset that contains n choices and response times for each query in $\mathcal{X}$,
then, for each arm z, the utility estimation error satisfies $\sqrt{n}\ \left(z^\top \widehat{\theta} - z^\top \theta\right) \xrightarrow{D} \mathcal{N}\left(0,\ \mathrm{AVar}_z\right)$.

If using $\widehat{\theta}_{choices}$, then

$$\mathrm{AVar}_z = z^\top \left(\sum_{x \in \mathcal{X}} \boxed{a^2\ \mathrm{Var}[c_x]}\ x x^\top \right)^{-1} z$$

The highlighted term (*The higher, the better*)



Strong pref Weak pref **Strong pref**

Asymptotic Variances Confirms That for Queries With ***Strong*** Preferences, Response Times Provide Information That Complements Choices

Given a fixed dataset that contains n choices and response times for each query in $\mathcal{X}$, then, for each arm z, the utility estimation error satisfies $\sqrt{n}\left(z^\top\widehat{\theta} - z^\top\theta\right) \xrightarrow{D} \mathcal{N}\left(0, \text{AVar}_z\right)$.

If using $\widehat{\theta}_{choices}$, then

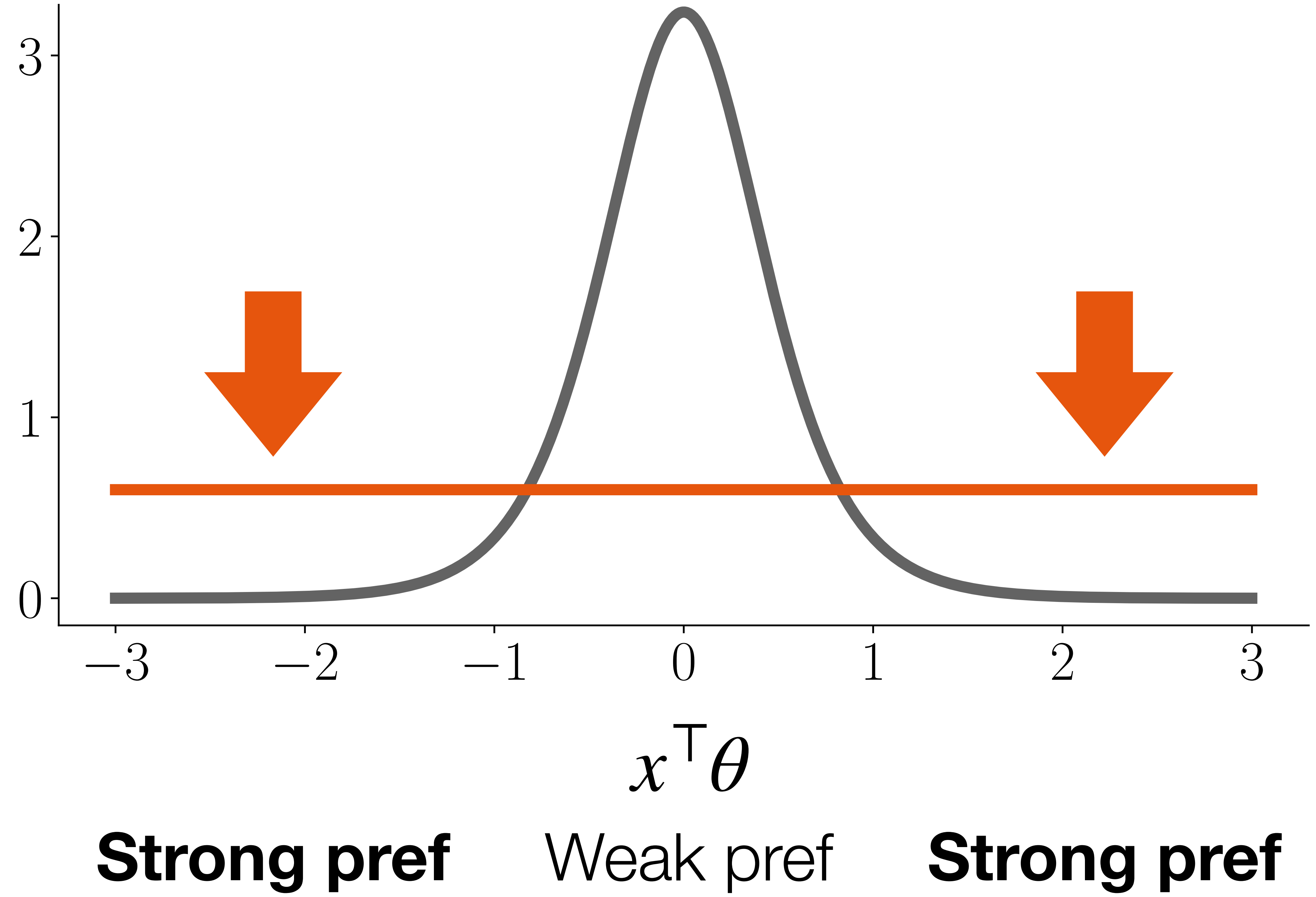
$$\text{AVar}_z = z^\top\left(\sum_{x\in\mathcal{X}} \boxed{a^2\ \text{Var}[c_x]}\, xx^\top\right)^{-1} z$$

If using $\widehat{\theta}_{choices,\ times}$, then

$$\text{AVar}_z \leq z^\top\left(\sum_{x\in\mathcal{X}} \boxed{\min_{\tilde{x}\in\mathcal{X}} \mathbb{E}[t_{\tilde{x}}]}\, xx^\top\right)^{-1} z$$

Example: assume each $x^\top\theta \in [-3,3]$:

The highlighted term (*The higher, the better*)



This Plot Does Not Provide Definitive Conclusions for Comparing Response Times and Choices for Queries With *Weak* Preferences

Given a fixed dataset that contains n choices and response times for each query in $\mathcal{X}$, then, for each arm z, the utility estimation error satisfies $\sqrt{n}\left(z^\top \widehat{\theta} - z^\top \theta\right) \xrightarrow{D} \mathcal{N}\left(0, \text{AVar}_z\right)$.

If using $\widehat{\theta}_{choices}$, then

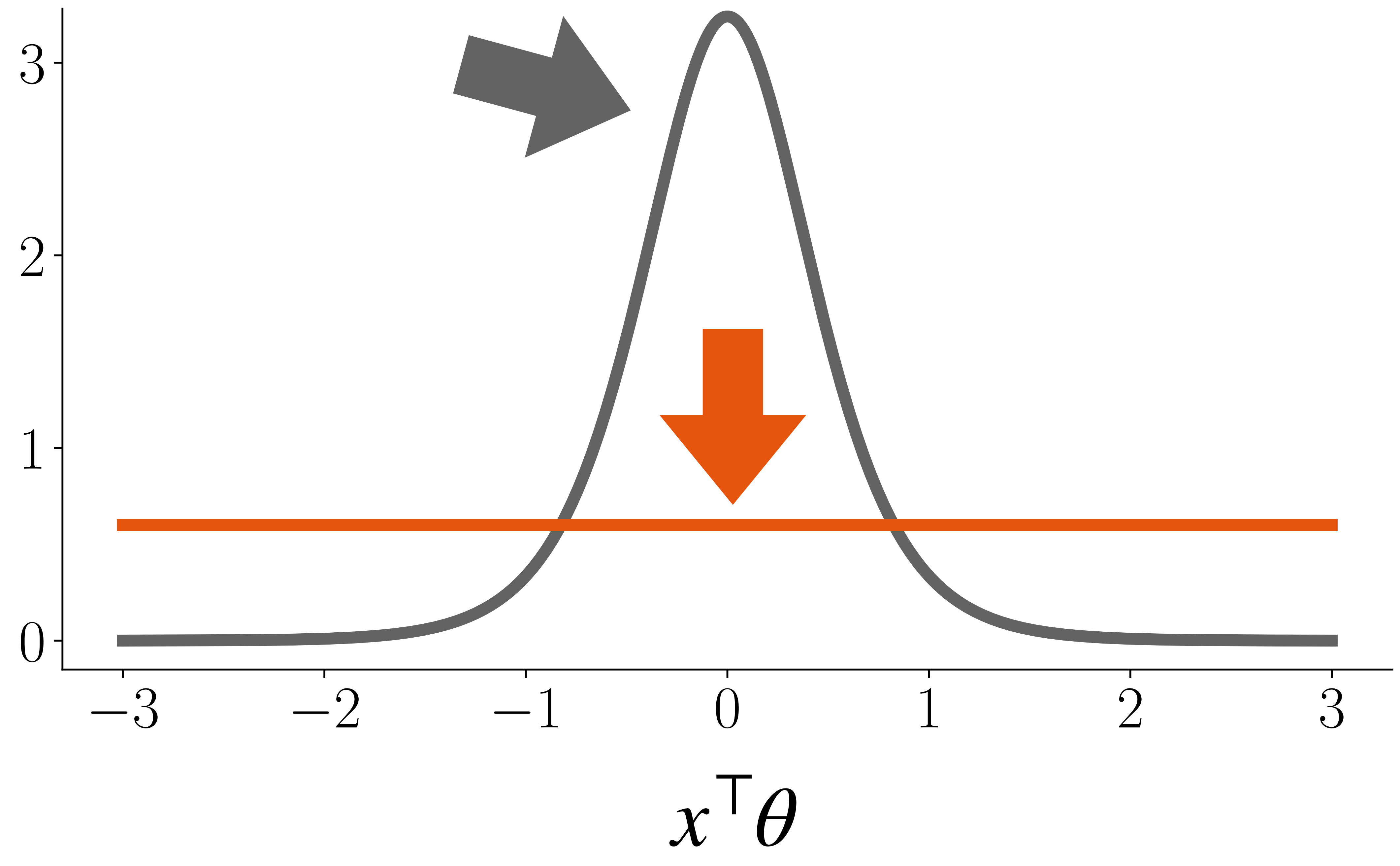
$$\text{AVar}_z = z^\top \left(\sum_{x \in \mathcal{X}} \boxed{a^2 \, \text{Var}[c_x]} \, x x^\top \right)^{-1} z$$

If using $\widehat{\theta}_{choices,\, times}$, then

$$\text{AVar}_z \leq z^\top \left(\sum_{x \in \mathcal{X}} \boxed{\min_{\tilde{x} \in \mathcal{X}} \mathbb{E}[t_{\tilde{x}}]} \, x x^\top \right)^{-1} z$$

Example: assume each $x^\top \theta \in [-3,3]$:

The highlighted term (*The higher, the better*)



Integrating Both Estimators Into the Generalized Successive Elimination Algorithm to Identify the Best Arm Within a Fixed Time Budget

- Split the total budget evenly into multiple phases.

- For each phase:

Compute an experimental design

Sample queries till the budget is exhausted

Estimate θ and eliminate arms

- Recommended the remaining arm $\widehat{z}$.

Performance measure: $\mathbb{P}\left[\widehat{z} \neq z^*\right]$

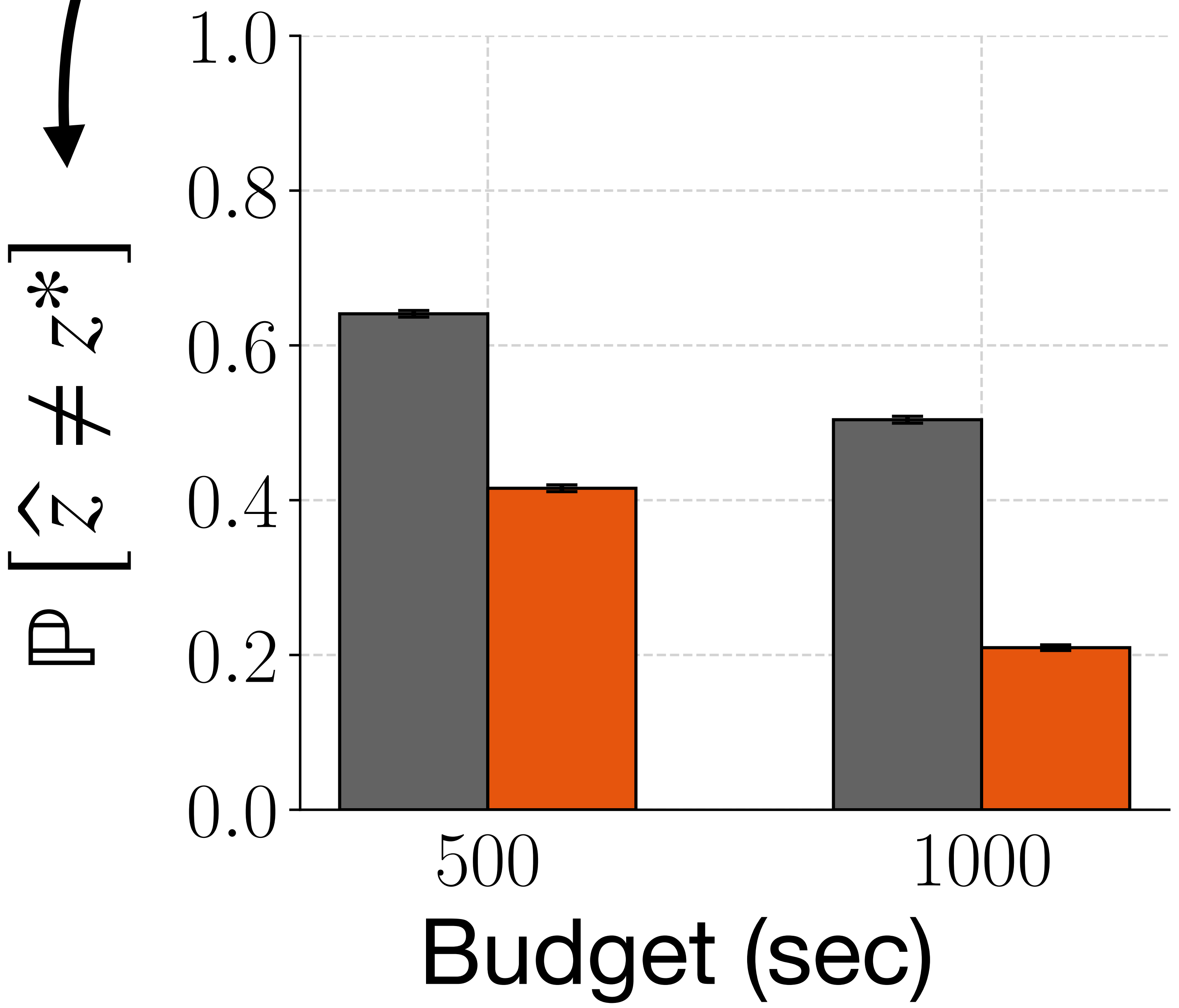
(Azizi et al., 2022)

Empirical Result of Bandit Learning Shows That Incorporating Response Times Reduces Learning Errors

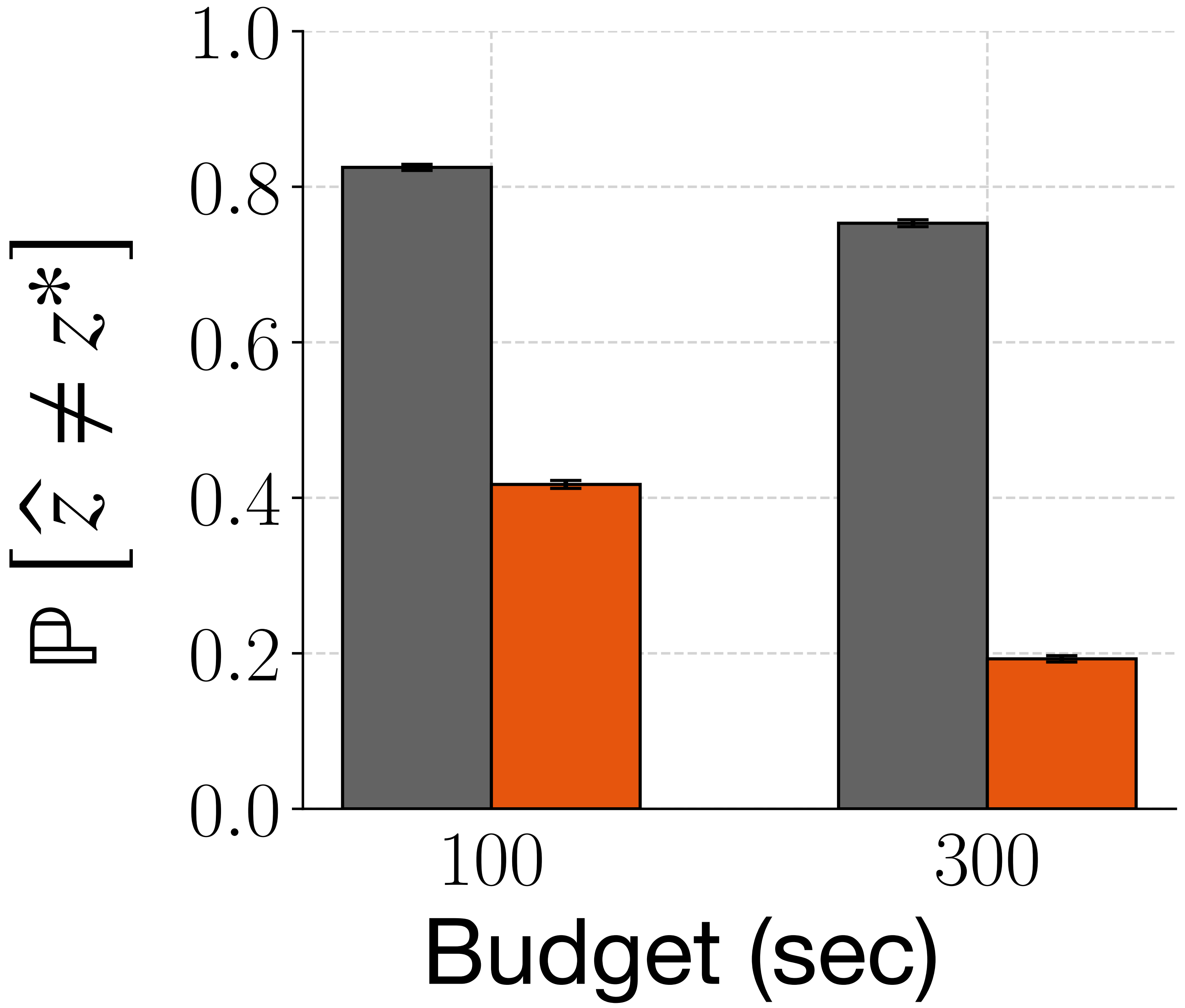
The lower, the better.

Generalized Successive Elimination with $\widehat{\theta}_{choices}$

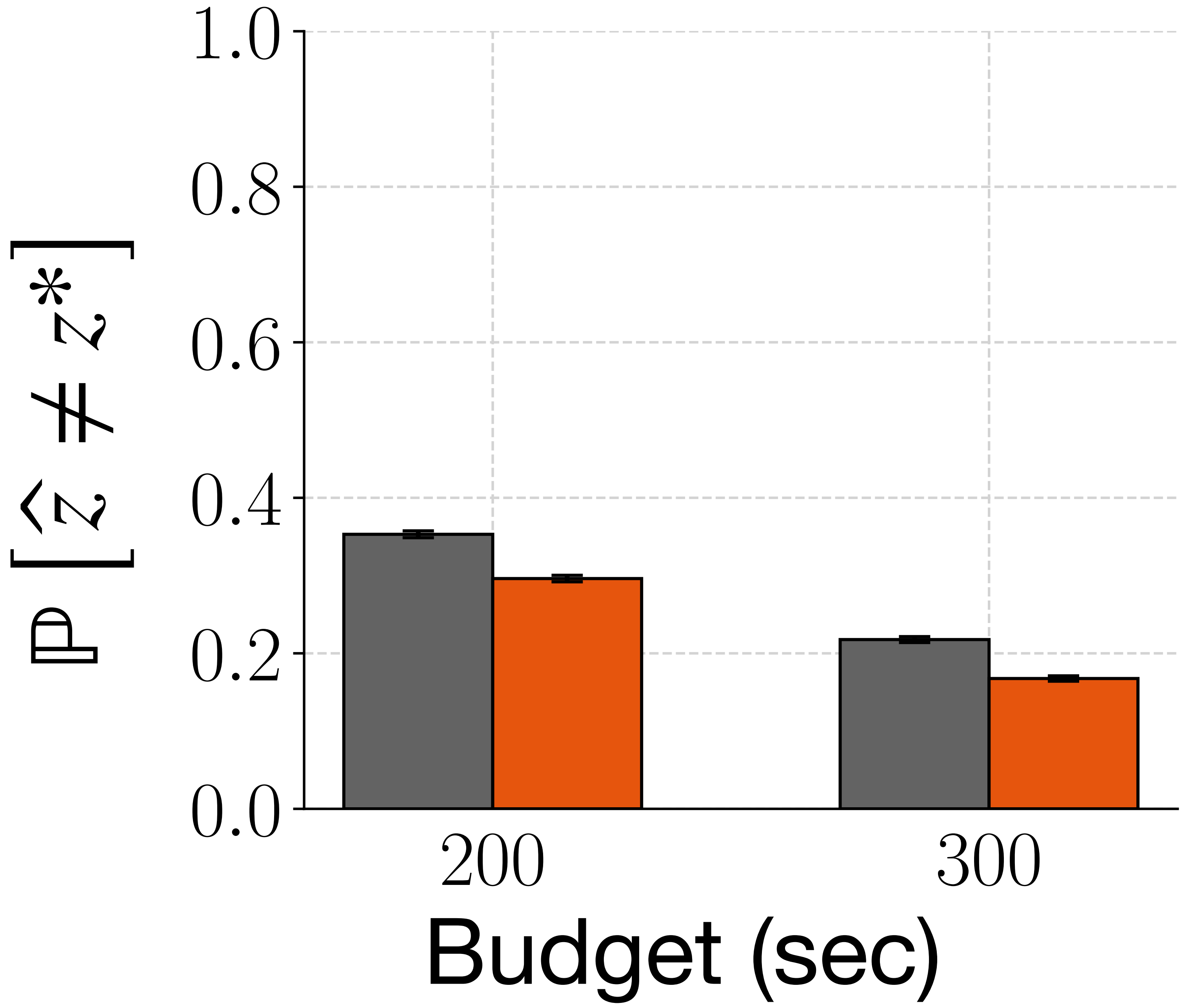
Generalized Successive Elimination with $\widehat{\theta}_{choices,\ times}$



(Smith and Krajbich, 2018)



(Clithero, 2018)



(Krajbich, et al., 2010)

Key Contributions: The *First* to Use Response Times for Preference Learning

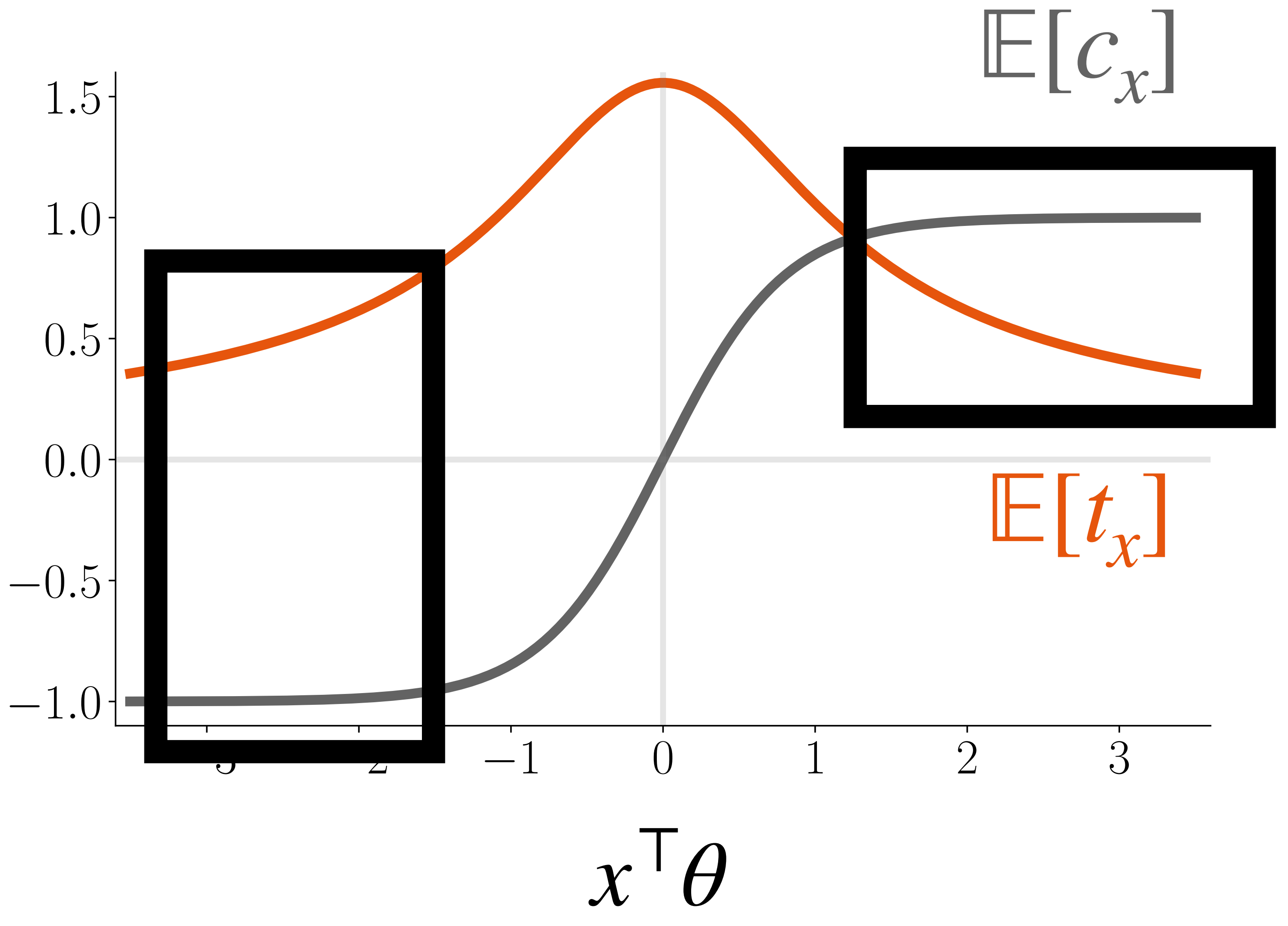
Full paper:

Poster: 4:30-7:30
@ East Exhibit Hall A-C #4901

(*Shen* is on faculty job market)

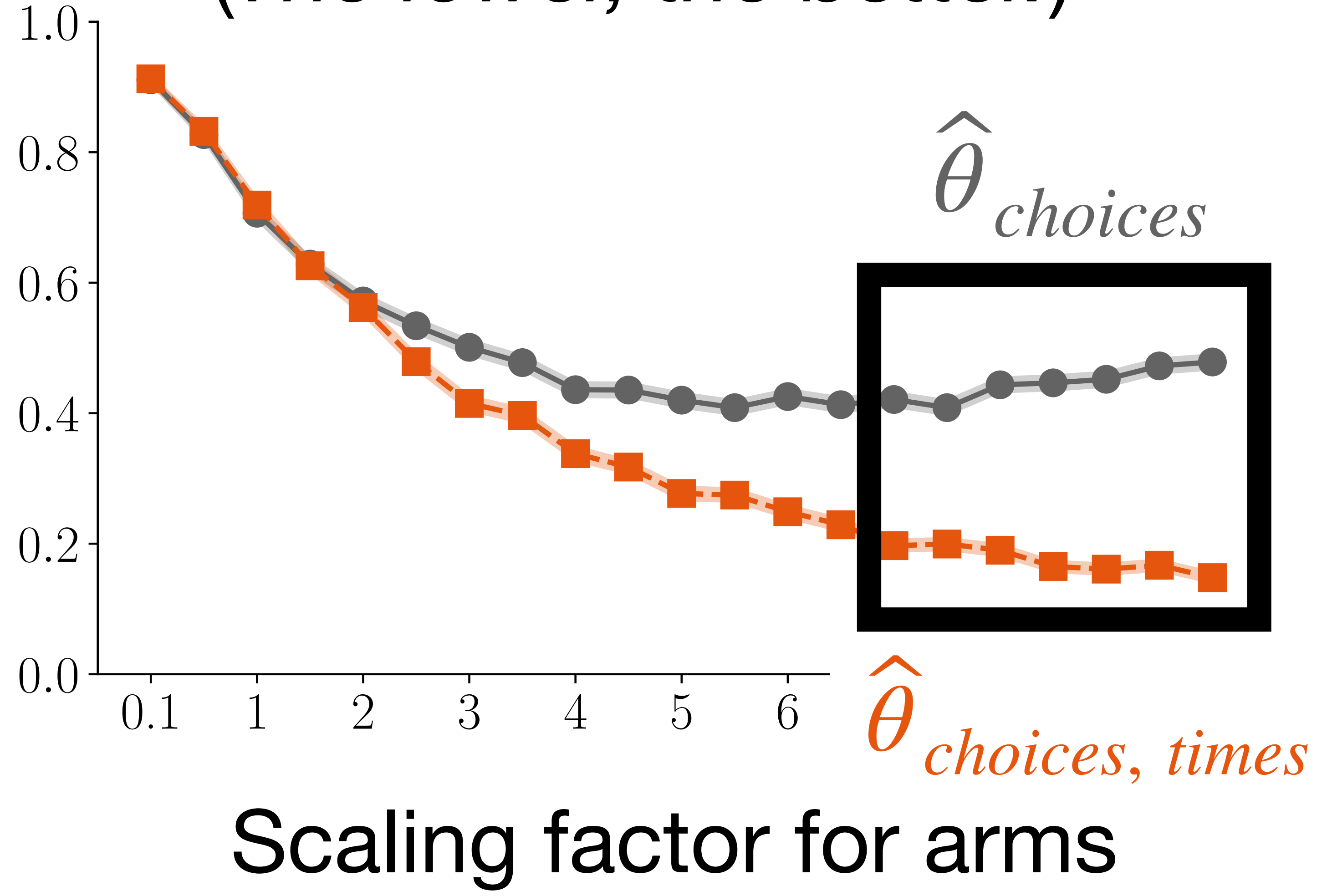
- A utility estimator using both choices and response times.
- An insight: response times from **queries with strong preferences** provide extra information that complements choices.

EZ-diffusion model



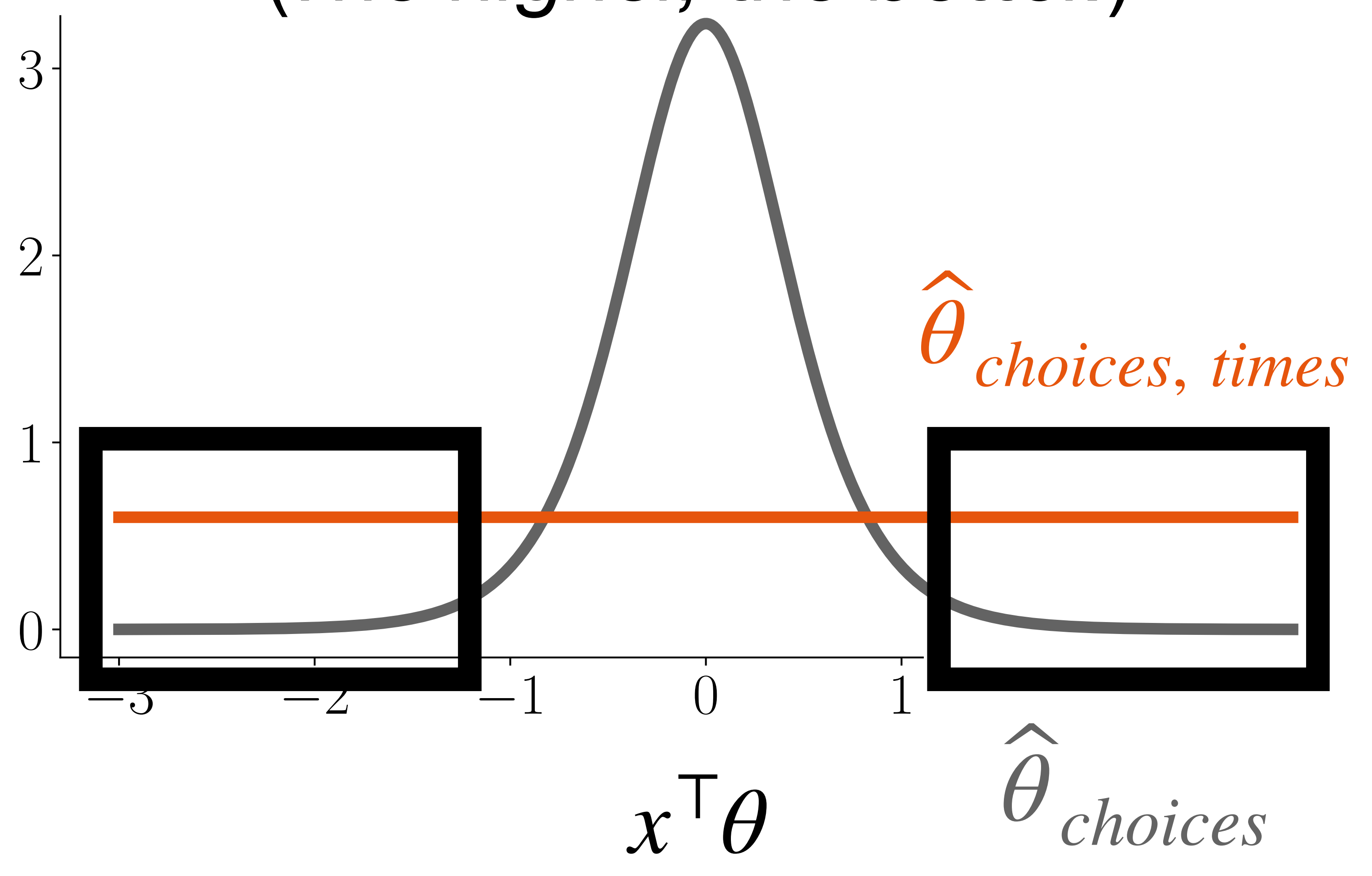
Strong pref Weak **Strong pref**

Empirical estimation error
(*The lower, the better.*)



Weak pref **Strong pref**

Asymptotic variance's key terms
(*The higher, the better.*)



Strong pref Weak **Strong pref**