Integrating Attention and Response Time Data into Cognitive Psychology Models to Understand Discrete Choices

Prateek Bansal
Presidential (Young) Assistant Professor
National University of Singapore

17th workshop on Discrete Choice Models, EPFL

Collaborators

My Research Group

- Xinwei Li (PhD Student)
- Jiaxuan Ding (PhD Student)
- Vladimir Maksimenko (Research Fellow)
- MSc Students

External Collaborators

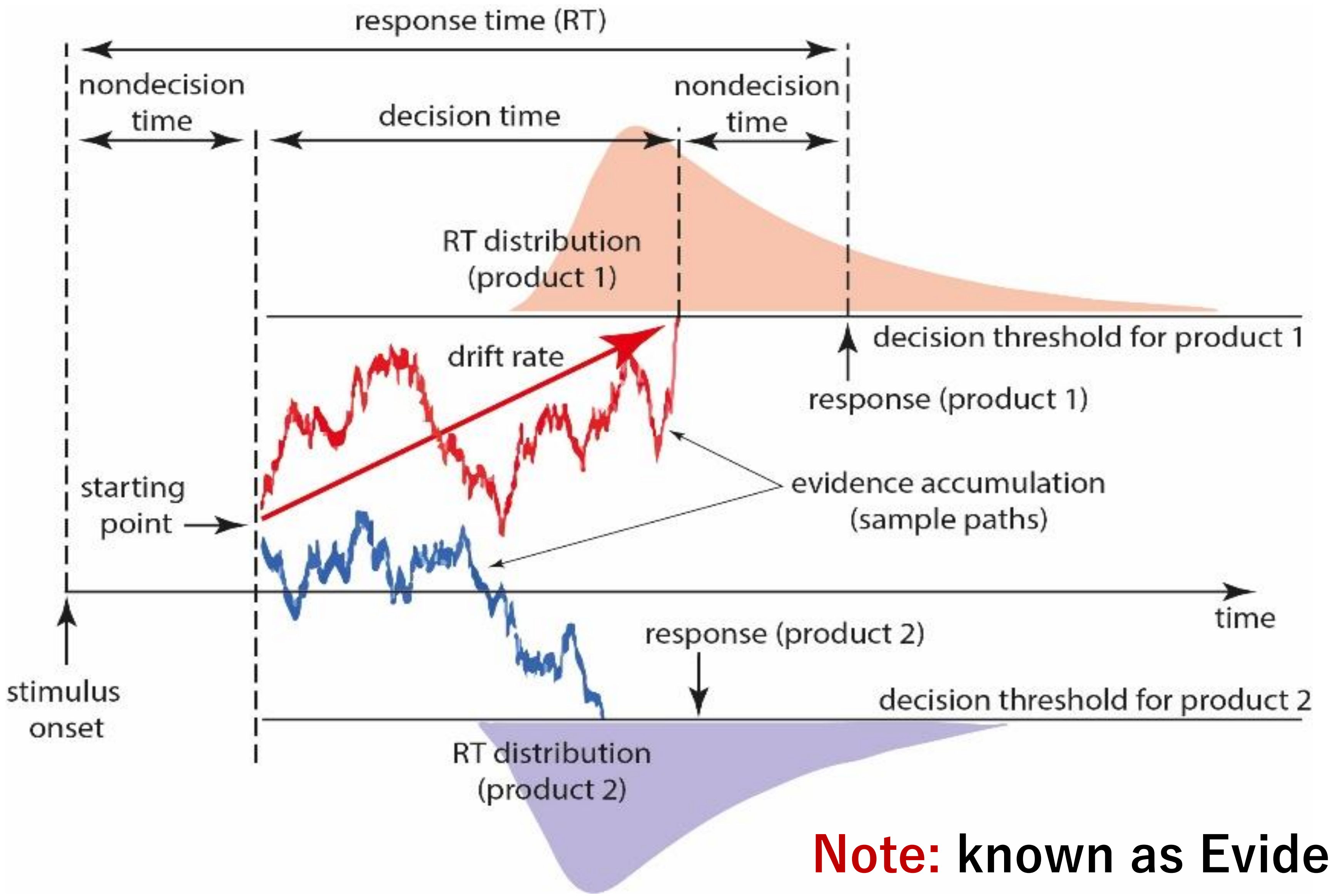
- Leonard Lee (NUS Marketing)
- David Nott (NUS Statistics)
- DongWon Oh (NUS Psychology)
- Ian Krajbich (UCLA Psychology)
- Bastian Henriquez-Jara (Universidad de Chile)
- Omar David Perez (Universidad de Chile)

Plan for the Talk

- What is the underlying formulation of models from Cognitive Psychology and what are their Advantages?
- What is the value of Response Time (RT) and how to integrate it into choice models?
- What is the value of Eye-tracking Data and how to integrate it into Choice Models?
- What are the Open Questions and pathways?

Formulation of models from Cognitive Psychology

Formulation of Sequential Sampling Models (SSMs)



General components:

- Starting point
- Drift rate
- Decision threshold
- Non-decision time

Note: known as Evidence Accumulation Models

Multi-attribute ballistic accumulator (MLBA)

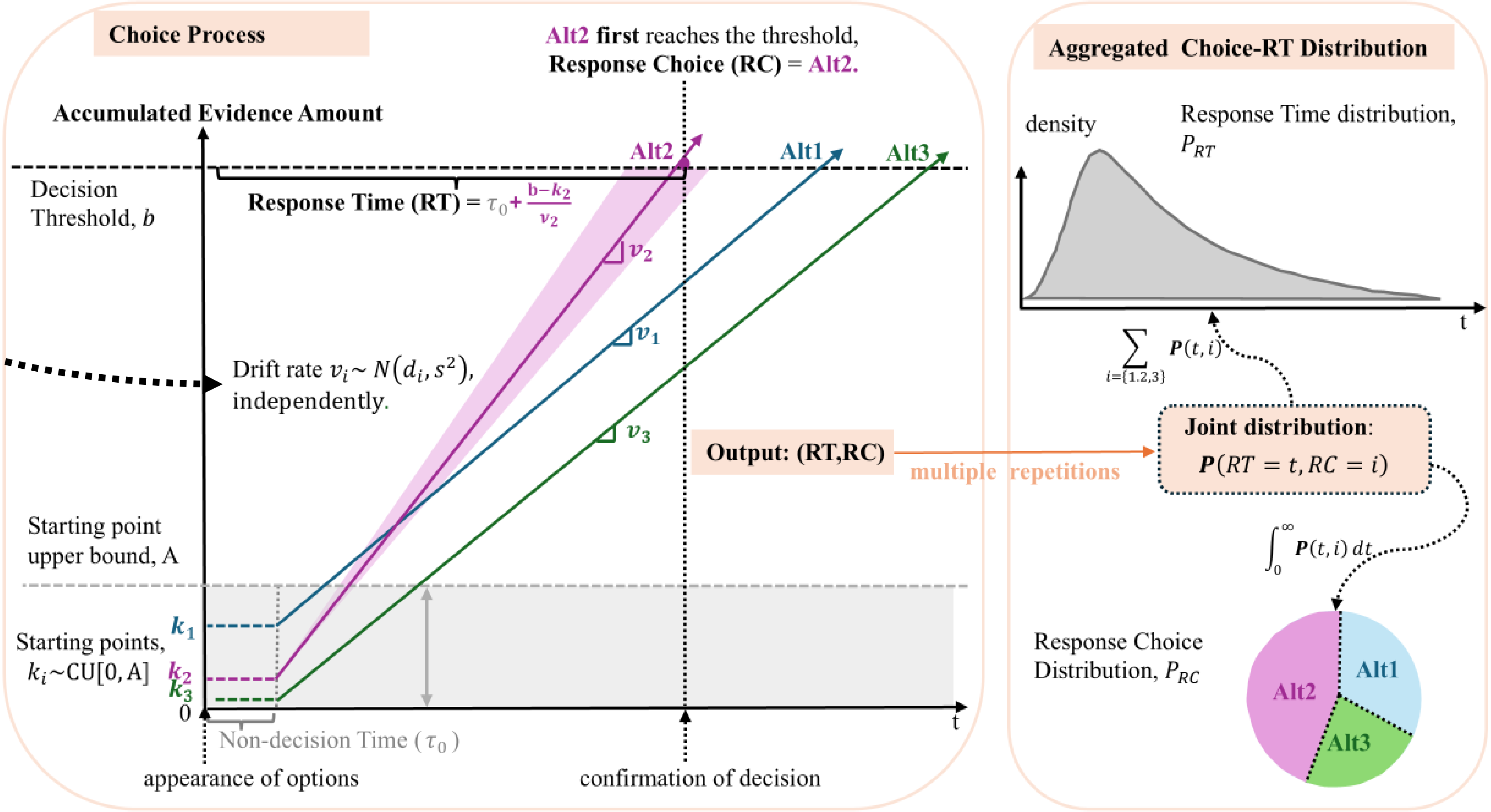
formulating drift rate mean d_i.

$$d_i = \zeta_i + \sum_{j \in \mathcal{C}, j \neq i} \sum_{k=1}^{K} \omega_{ijk} \beta_k (X_{ik} - X_{jk}),$$

where

$\omega_{ijk} = exp(-\lambda|\beta_k(X_{ik} - X_{jk})|)$ and

$$\lambda = \begin{cases} \lambda_1, \beta_k(X_{ik} - X_{jk}) > 0 \\ \lambda_2, \beta_k(X_{ik} - X_{jk}) \leqq 0 \end{cases}.$$



Notions:

X_{ik}: the value of attribute k for alt i.

----------front-end parameters------

ζ_i: alternative-specific constant.

β_k: attribute-specific coefficient.

λ_1 (λ_2): parameter for gain (loss).

-----------back-end parameters------

A: starting point upper bound;

b: decision threshold;

s: across-trial noise in drift rate;

τ_0: non-decision time.

$$MLBA\_CRT(RC = i, RT = t) = f_i(t)\Pi_{j \neq i}(1 - F_j(t))$$

$f_i(t)$is the probability density function of the time t taken for the accumulator i to reach the threshold and $F_i(t)$ is the cumulative density function.

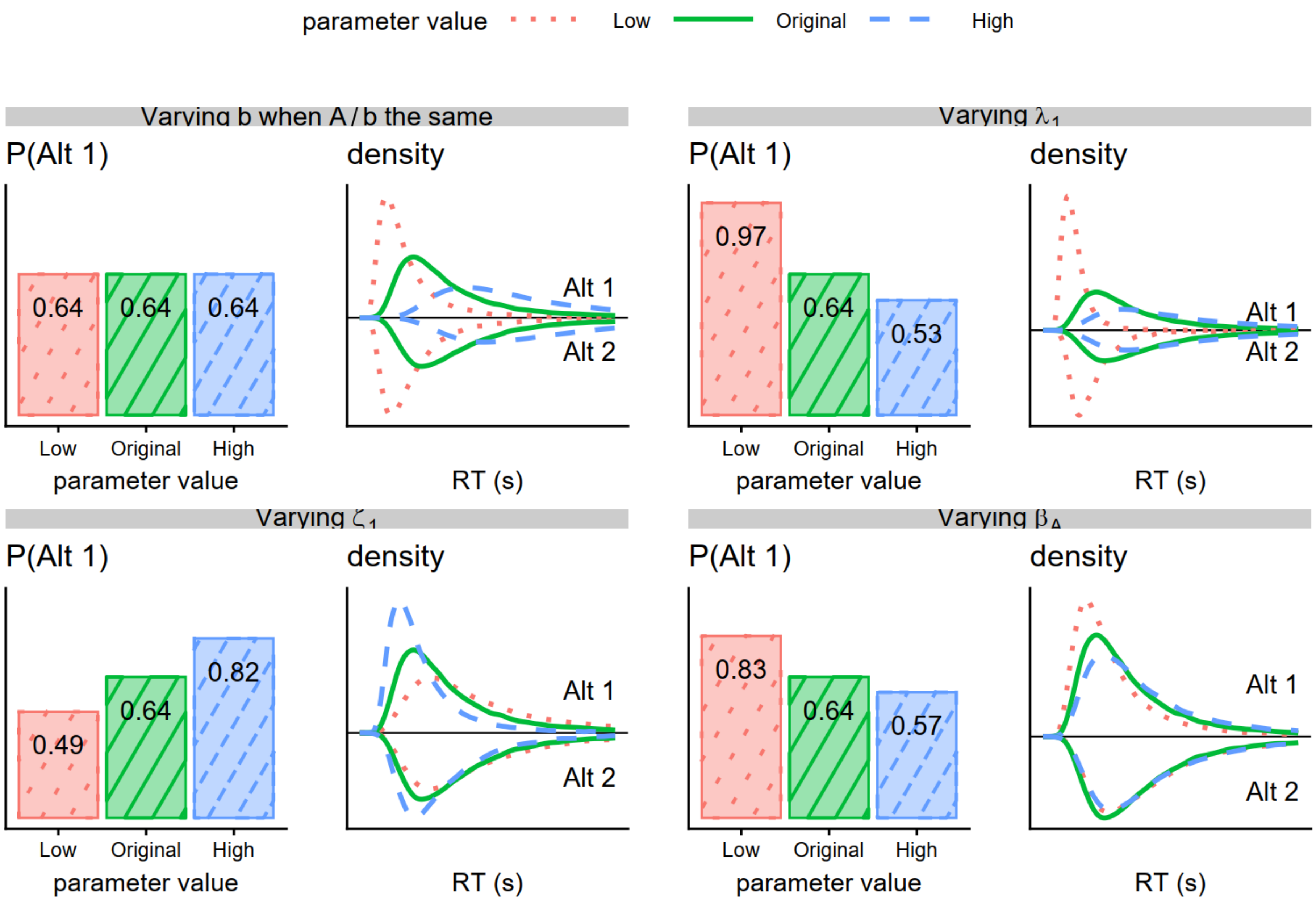
Reference: Trueblood et al. (2014) and Li and Bansal (2024).

MLBA Parameters' Impact on Choice & RT Distributio

Example:
A binary choice situation with three attributes.

	Attr A	Attr B	Attr C
Alt 1	1 unit	2 units	3 units
Alt 2	3 units	1 unit	2 units

- **Front-end** parameters consist of the *drift rate mean* affect both **choice proportion** and **RT distribution.**
- **Back-end** parameters <u>except</u> *drift rate mean* **only affect RT distribution** while keeping choice proportion the same.



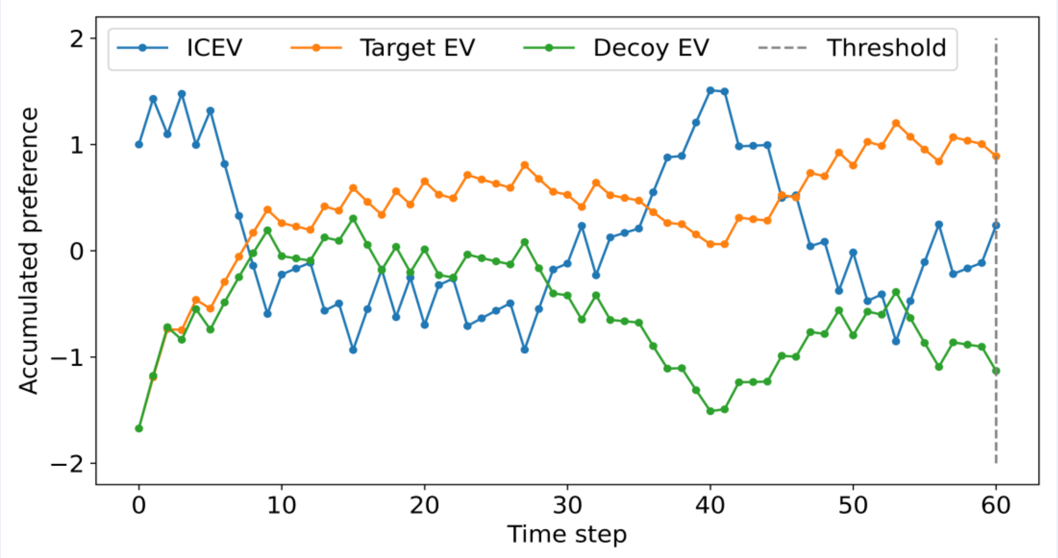
Other SSMs: MDFT and MDBS

Multi-alternative Decision-field Theory (MDFT)

Attention Shifts: Attribute-wise

Preference accumulation

- All alternatives accumulate evidence at time step t.
- **Evidence Direction & magnitude:** comparison results

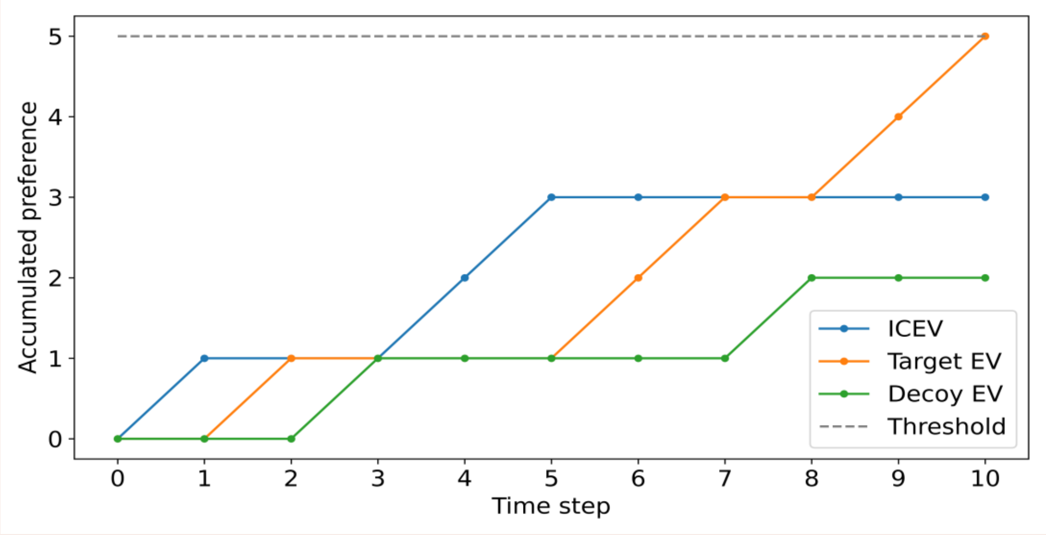


Multi-alternative Decision by Sampling (MDbS)

Attention Shifts: Pair-wise

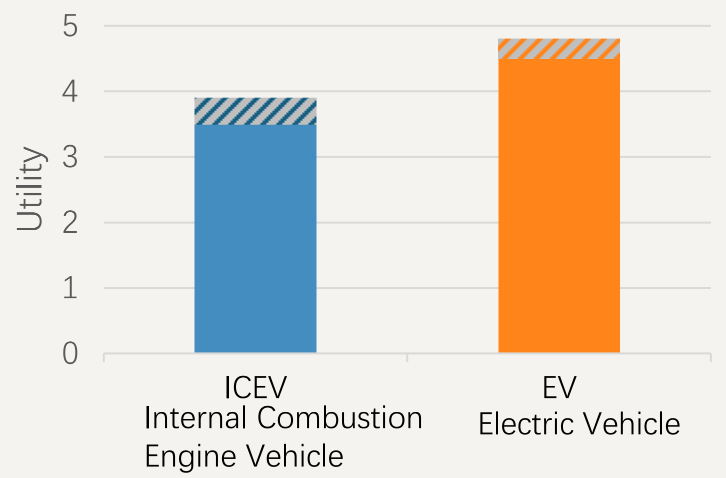
Preference accumulation

- One alt. accumulates one unit of evidence at step t.
- **Direction and magnitude:** fixed



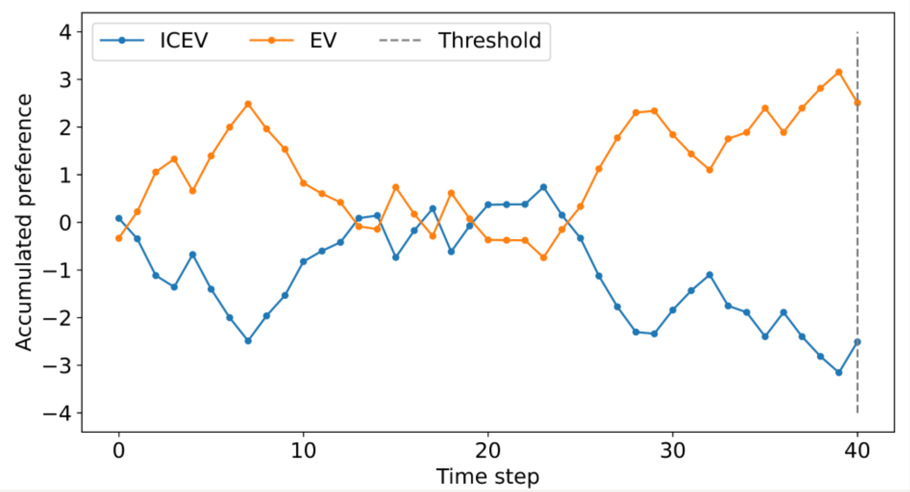
Potential Advantages of SSM over RUM?

RUM (Random utility maximization)



Explain **only WHAT** people choose

SSM (Sequential sampling model)



Explain **WHAT** people choose **+ HOW** they arrive at the choice

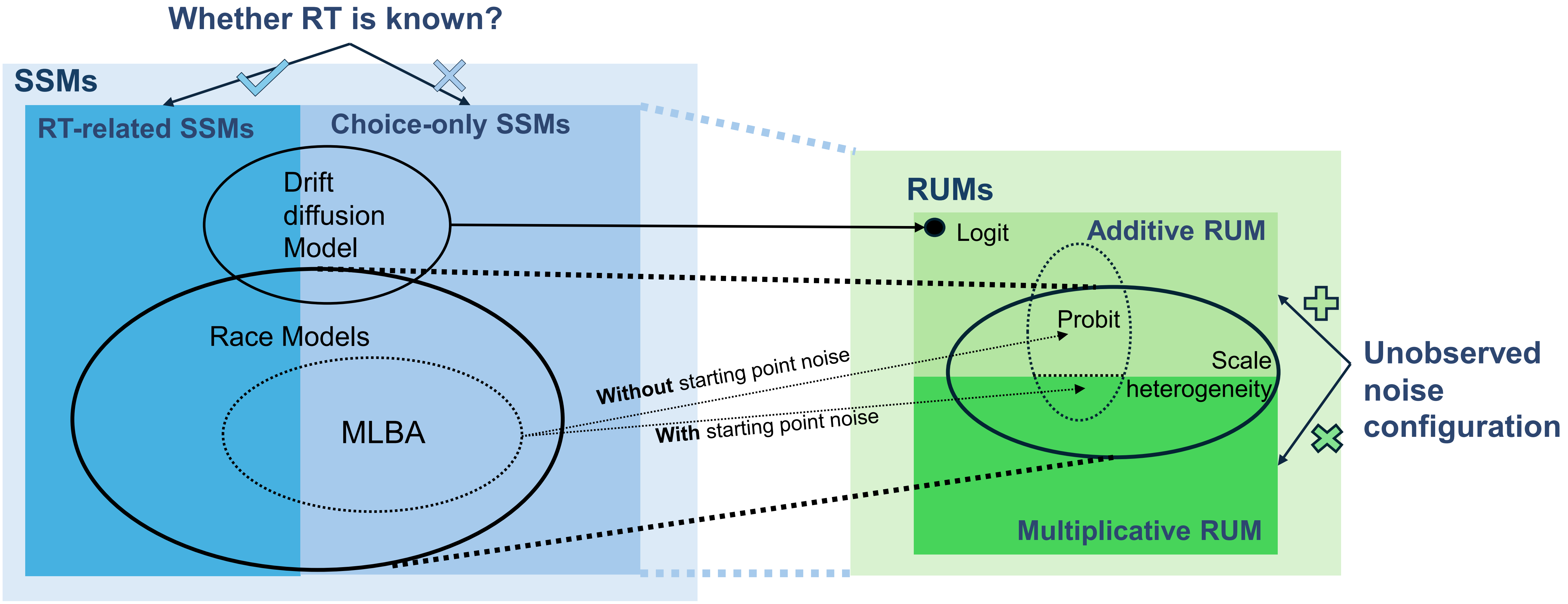
	RUM	SSM
Chosen	Highest utility	Evidence reaches a predefined threshold
Decision-making process	❌	✅
Cognitive bias	❌	✅

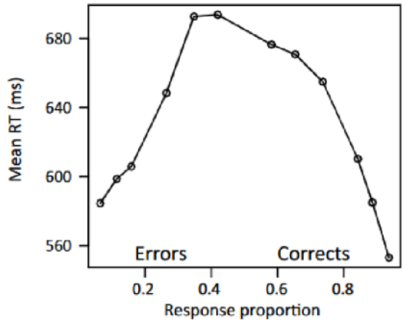
Advantages

- Can incorporate process data (e.g., eye-tracking & RT)
- Cognitive underpinning
- Can explain decoy effect better

Response Time

RUMs vs SSMs





Why do we think that RT might have some information?

- Speed-accuracy trade-off
- RT reduces standard error: empirical evidence

Reference: Webb (2019) and Li and Bansal (2024).

Models with and Without RT

(i) selecting alternative i from the choice set $\mathscr{C}$ at response time t:

$$P\_CRT_{\theta}(RC = i, RT = t), i \in \mathscr{C}, t \geq 0$$

(ii) selecting alternative i from the choice set $\mathscr{C}$ conditional on the given $RT = t$:

$$P\_RTG_{\theta}(RC = i | RT = t) = \frac{P\_CRT_{\theta}(RC = i, RT = t)}{\sum_{i \in \mathscr{C}} P\_CRT_{\theta}(RC = i, RT = t)}, \quad i \in \mathscr{C}$$

(iii) selecting alternative i from the choice set $\mathscr{C}$ after marginalizing over RT:

$$P\_CO_{\theta}(RC = i) = \int_{0}^{\infty} P\_CRT_{\theta}(RC = i, RT = t) dt, \quad i \in \mathscr{C}$$

Reference: Li and Bansal (2024).

Asymptotic Results

Result 1:
P_{CRT} provides better estimate in terms of efficiency.

- Estimates of parameters based on P_{CRT} have the least **asymptotic** *MSE*, *posterior variance*, *standard errors, CI length*, and etc., comparing to those of P_{CO} and P_{RTG}.
- For a given inference tolerance, P_{CRT} requires the least *effective sample size* compared to that of P_{CO} and P_{RTG}.

Result 2:
P_{RTG} provides more robust estimate in terms of the predicted choice accuracy.

P_{RTG} ensures a **no worse** asymptotic predicted accuracy than that of P_{CO} and P_{CRT} no matter model and parameter space are *well- or mis-specified*.

RT-related distributions contribute to both parameter inference and choice prediction.

Sketch Proof: Result 1

The intuition is that the chain rule of the Fisher Information Matrix for two jointly distributed random variables X and Y implies that:

$$I_{XY}(\theta) = I_{X|Y}(\theta) + I_X(\theta)$$

Hence, the Fisher Information matrix of three types of distribution follows:

$$I_{CRT}(\theta_0) = I_{CO}(\theta_0) + I_{RT|RC}(\theta_0)$$
$$I_{CRT}(\theta_0) = I_{RTG}(\theta_0) + I_{RT}(\theta_0)$$

Given all Fisher Information Matrices above are non-negative definite,

$$I_{CRT}(\theta_0) \geq I_{CO}(\theta_0)$$
$$I_{CRT}(\theta_0) \geq I_{RTG}(\theta_0)$$

Validation on Result 1

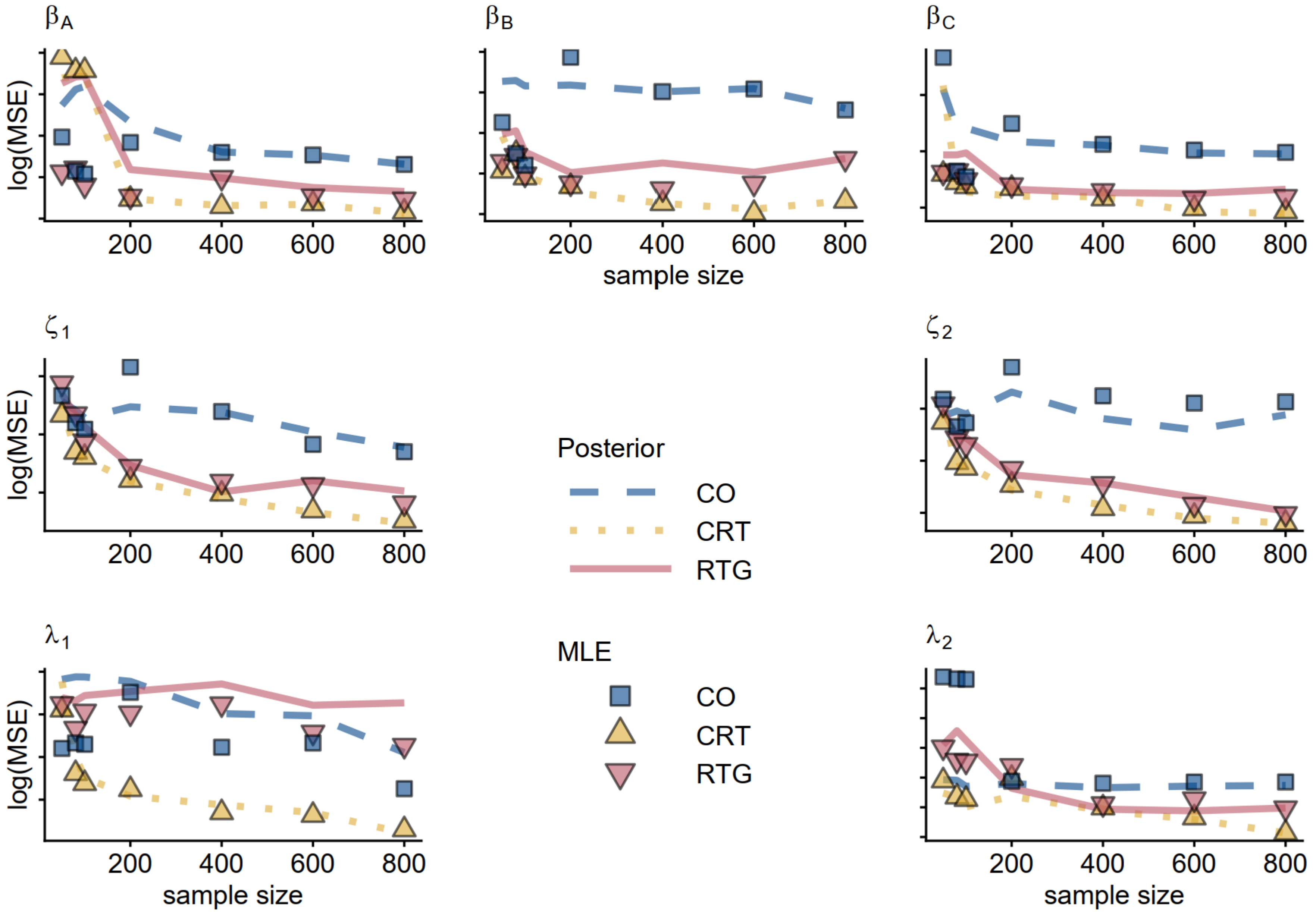
Result 1:
P_{CRT} provides better estimate in terms of <u>efficiency</u>.

$\mathrm{MSE_{CRT}} \leq \mathrm{MSE_{RTG}}$ and $\mathrm{MSE_{CRT}} \leq \mathrm{MSE_{CO}}$,

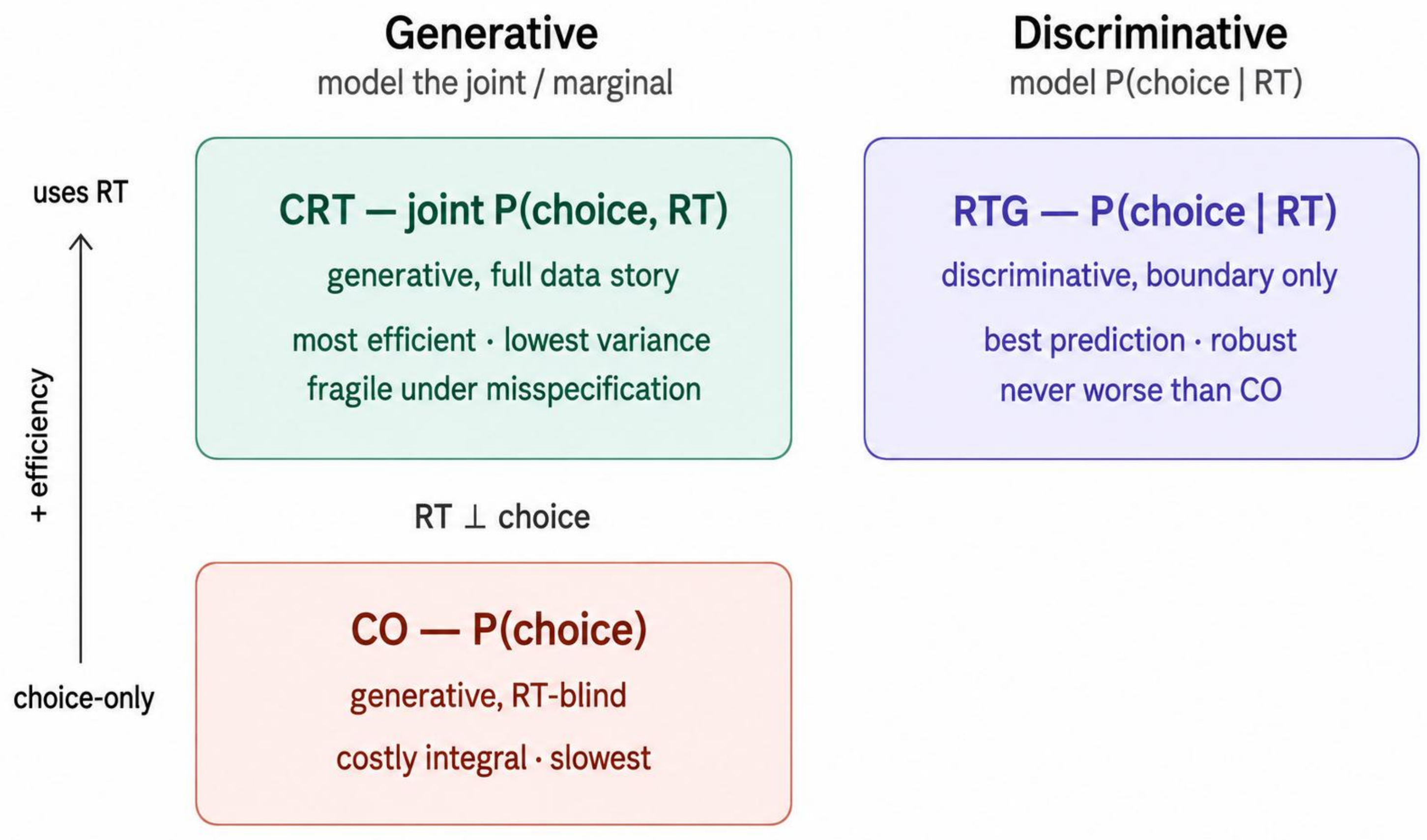
as *sample size* goes large enough, regardless estimation methods:

- Bayesian (posterior);
- frequentist's (MLE).

MLBA as an example.



Intuition of Result 2

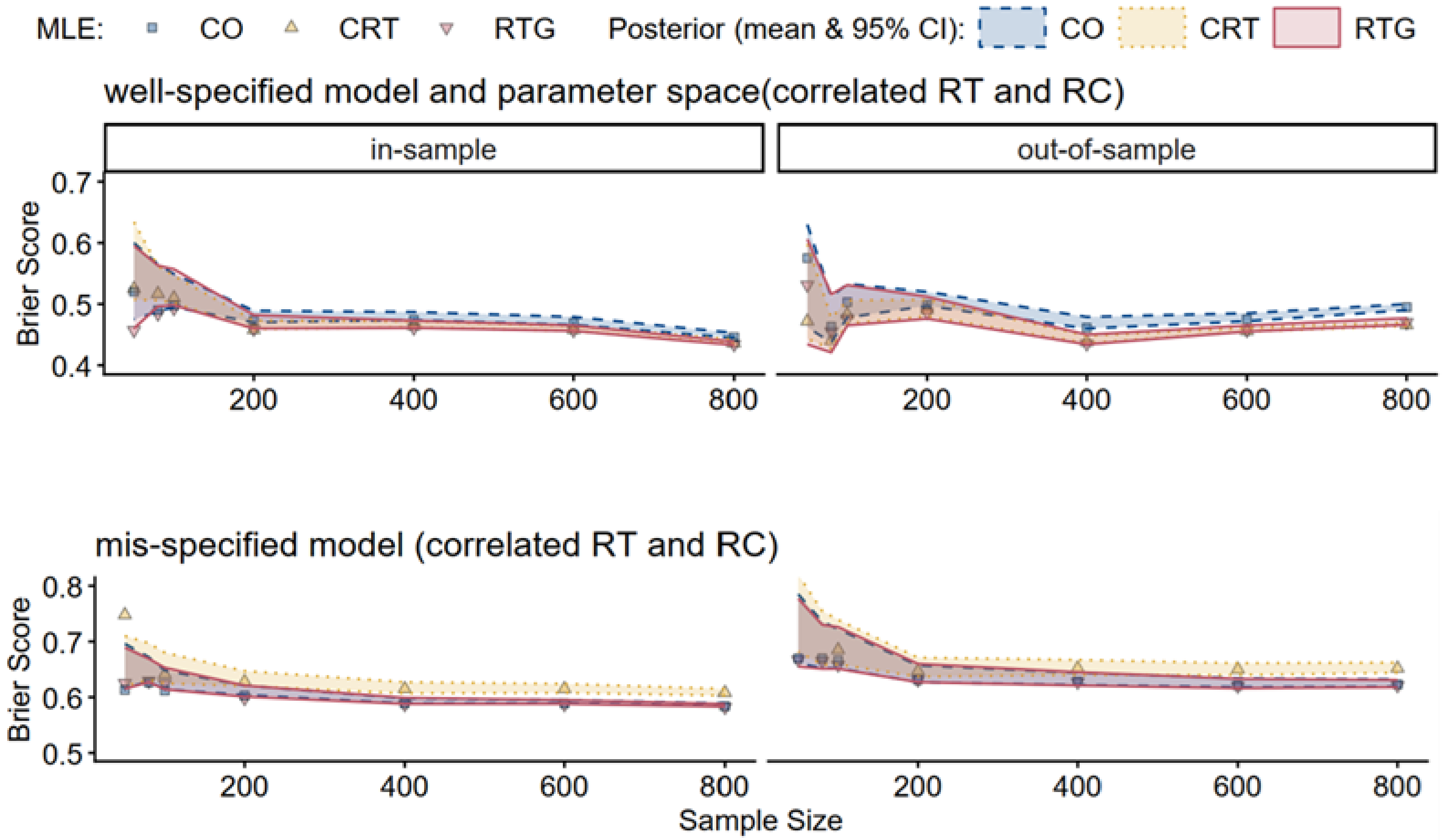


- Efficiency climbs upward (use RT); robustness grows rightward (stop modelling P(RT)).
- RTG's likelihood is $P(RC \mid RT)$ only: the RT marginal $P(RT)$ never enters.
- CRT fits the joint distribution; so a wrong RT distribution drags its estimate toward matching that bad margin at the expense of choice fit. RTG has no such term to poison.

Validation on Result 2

Result 2:
P_{RTG} provides more robust estimate in terms of the <u>predicted choice accuracy</u>.

- On *well-specified* model and parameter space:
  $$PE_{\mathrm{RTG}} = PE_{\mathrm{CRT}} \leq PE_{CO}.$$
- On *mis-specified* model:
  $$PE_{\mathrm{RTG}} \leq PE_{CO} \leq PE_{\mathrm{CRT}}.$$



Eye-tracking Data: Case Study 1

Eye-tracking for Intra-individual Preference Heterogeneity

Multiple trials

A choice situation
(eg.travel mode choice)

Multiple individuals

1st trial: travel time matters the most;
2nd trial: travel cost matters the most;
…

intra-individual heterogenuity

Not fully true in behavioral findings!

1st subject: travel time matters most;
2nd subject: travel cost matters most;
…

inter-individual heterogenuity

unobserved random noise on **constant** *personal tastes*.

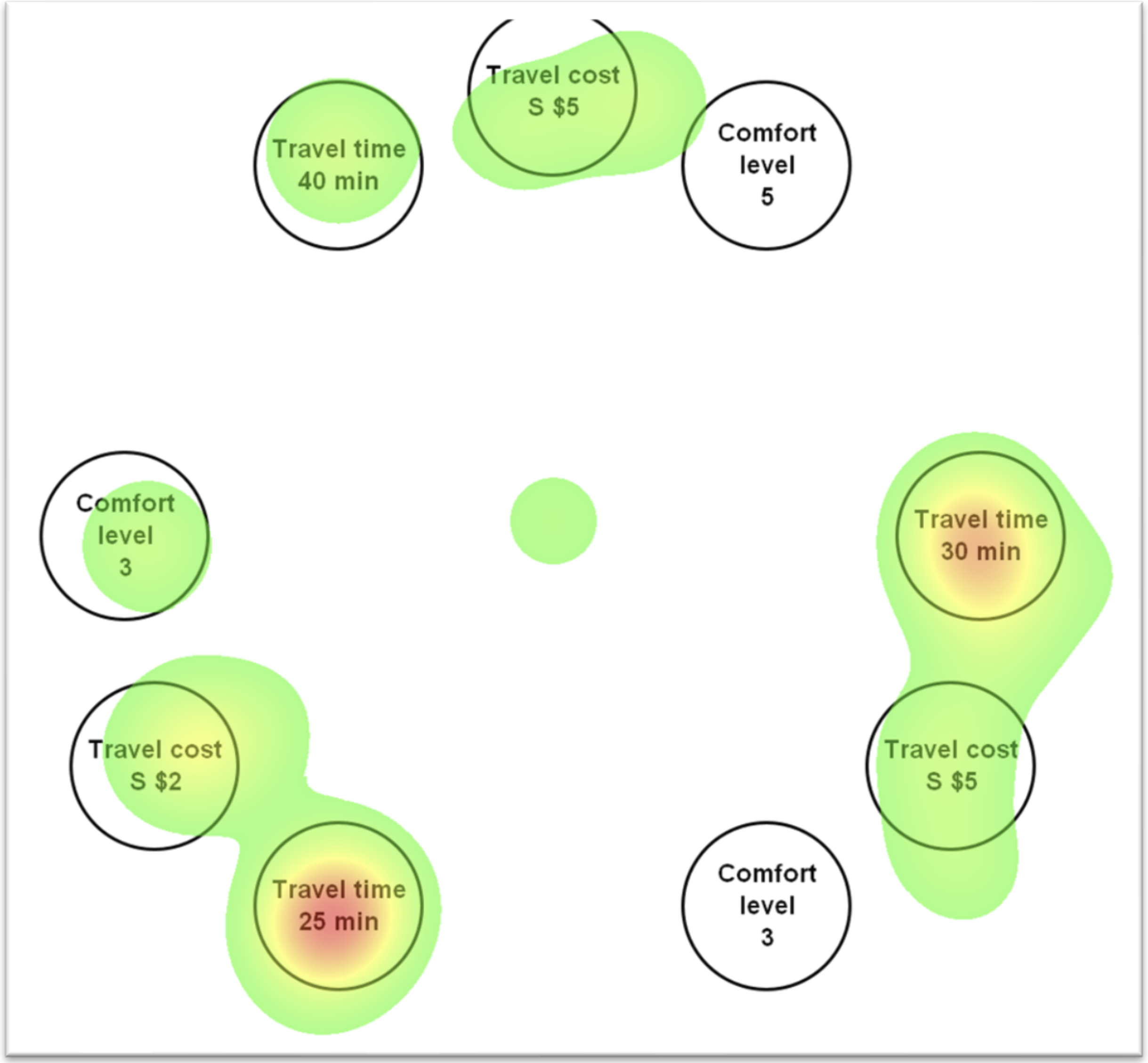
Superior in-sample fitness but not for out-of-sample prediction.

Sources of the preference perturbation based on mixed logit models & latent class latent variable models

various *personal tastes* divergent from population mean.

Reference: Hess and Rose, (2009) , Krueger et al. (2021), Kim, S. H. (2023)

Empirical Study: Experiment Design and Data collection



fixation location
Reset (2s)

Ternary choice task
(free time)

150 choice situations in total (each individual)

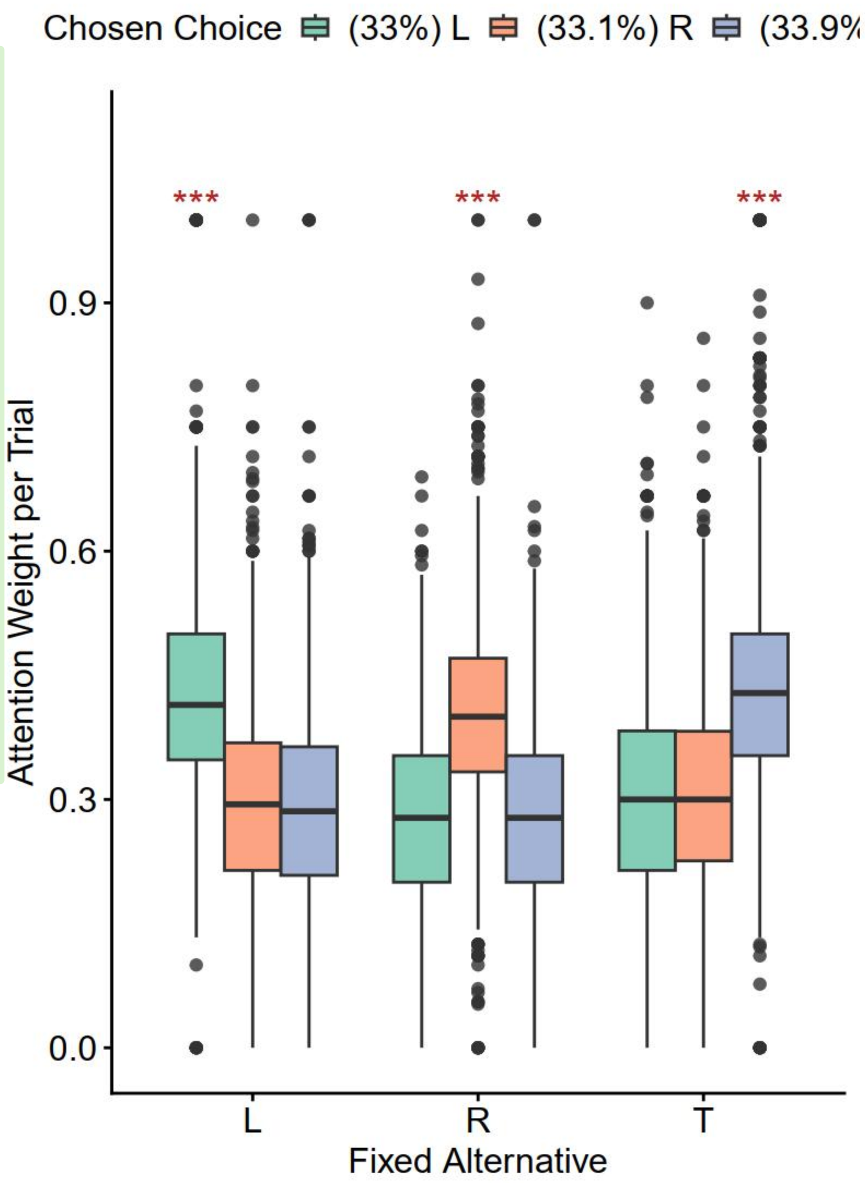
5 blocks with 4 breaks.

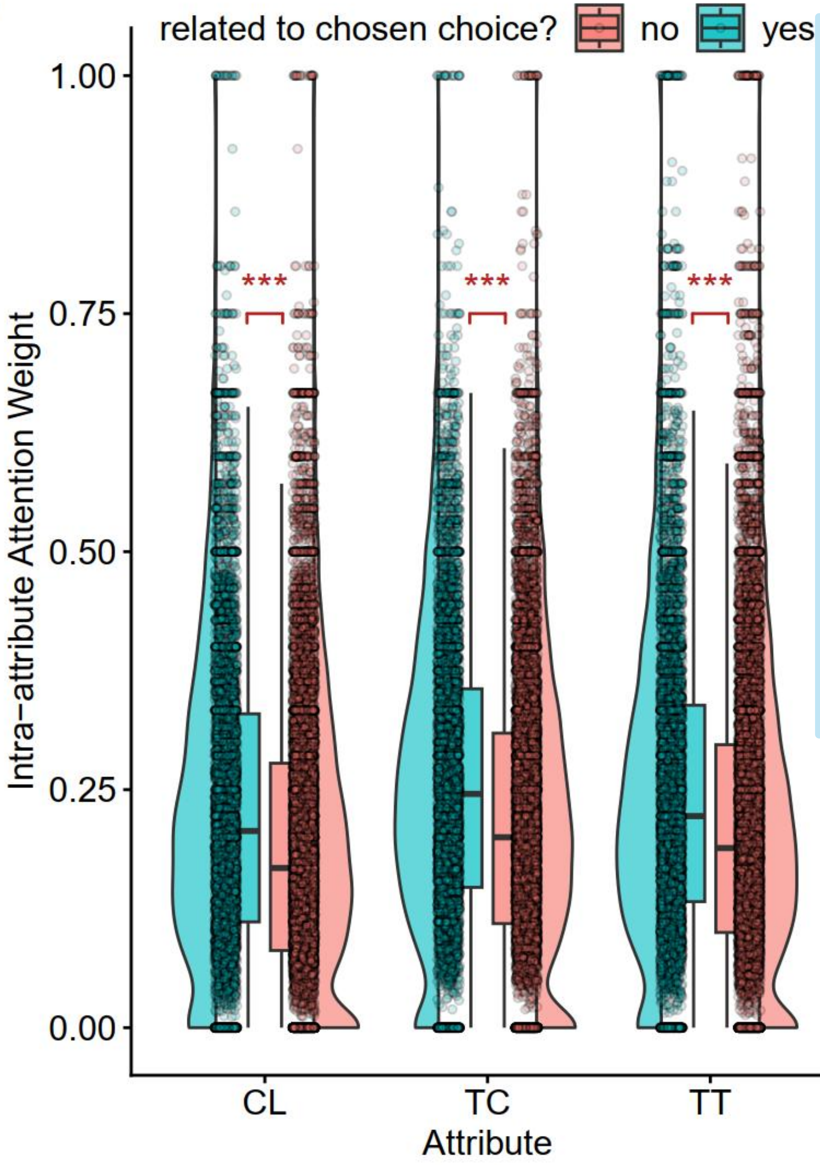
Chosen choice, RT and ET are collected.

Behavioral Study Findings (within Subject)

ET-choice link at <u>alt level</u>:

The **visual attention weight** on an option <u>positively correlates</u> to its chosen probability.





ET-choice link at <u>attribute level</u>:

The **pairwise comparison** within an attribute related to an option <u>positively correlates</u> its chosen probability.

Attentional MLBA with Trial-level ET

Canonical MLBA model

Drift rate mean for alt i of indiv n for situation s:

$$d_{ins} = \zeta_{in} + \sum_{j\in\mathcal{C},j\neq i}\sum_{k=1}^{K}\omega_{ijkns}\beta_{kn}(X_{iks}-X_{jks}),$$

where

$$\omega_{ijkns} = exp(-\lambda_{ns}|\beta_{kn}(X_{iks}-X_{jks})|)$$

and

$$\lambda_{ns} = \begin{cases}\lambda_{1n}, \beta_{kn}(X_{iks}-X_{jks}) > 0\\ \lambda_{2n}, \beta_{kn}(X_{iks}-X_{jks}) \leqq 0\end{cases}.$$

When $\lambda_{1n} = \lambda_{2n} = 0$,

Choice-only canonical MLBA

=

Mixed Probit with inter-individual heterogenuity

Notions:

X_{ikns}: the value of attribute k for alt i of s choice situation.

Personal taste parameters of individual n *(stable across trials):*

ζ_{in}: alternative-specific constant.

β_{kn}: attribute-specific coefficient

λ_{1n} (λ_{2n}): sensitivity parameter for gain (loss).

-----------*only in A-MLBA*-----------

α_n: decision strategy weight on context effect.

Trial-level ET data

(noisy but informative and observed)

w_{ikns}: visual attention weight on attribute k of option i on trial s by indiv n.

Trial-level RT data

Attentional–MLBA model

Drift rate mean for alt i of indiv n for situation s:

$$d_{ins} = \zeta_{in} + \sum_{k=1}^{K}\beta_{kn}X_{ikns}$$

isolate goal-driven evaluation

$$+\alpha_n\sum_{j\in\mathcal{C},j\neq i}\sum_{k=1}^{K}\omega_{ijkns}\lambda_n\beta_{kn}(X_{ikns}-X_{jkns}),$$

stimuli-driven context effect

Where

$$\omega_{ijkns} = normalized\{w_{ikns}w_{jkns}\}.$$

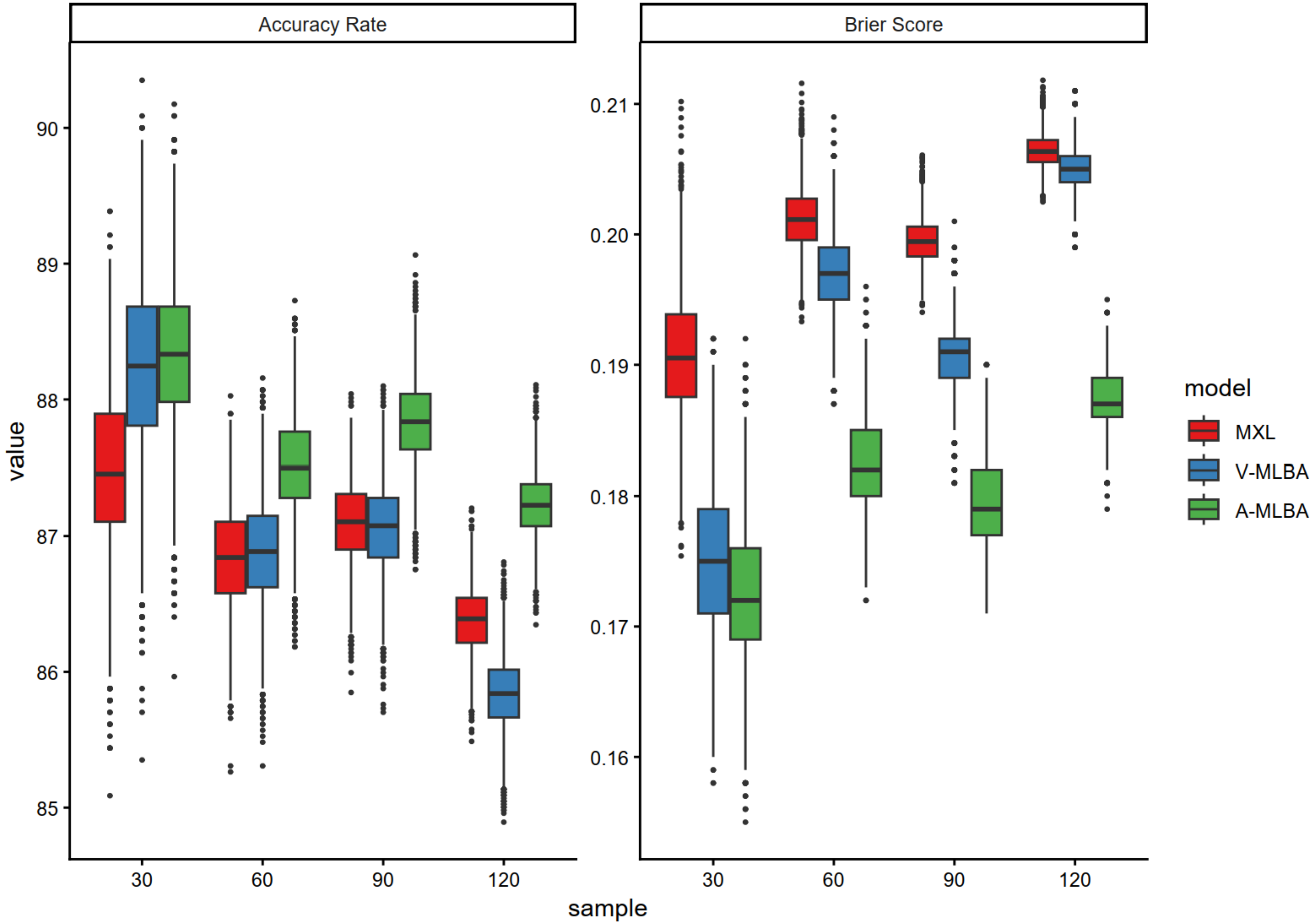
When $\omega_{ijkns} \backsim N(0,1)$ and the same over all i and j, $\lambda_{1n} = \lambda_{2n} = 1$, and $\alpha_n = 1/J$,

Choice-only A-MLBA

=

Mixed Probit with inter- and intra-individual heterogenuity.

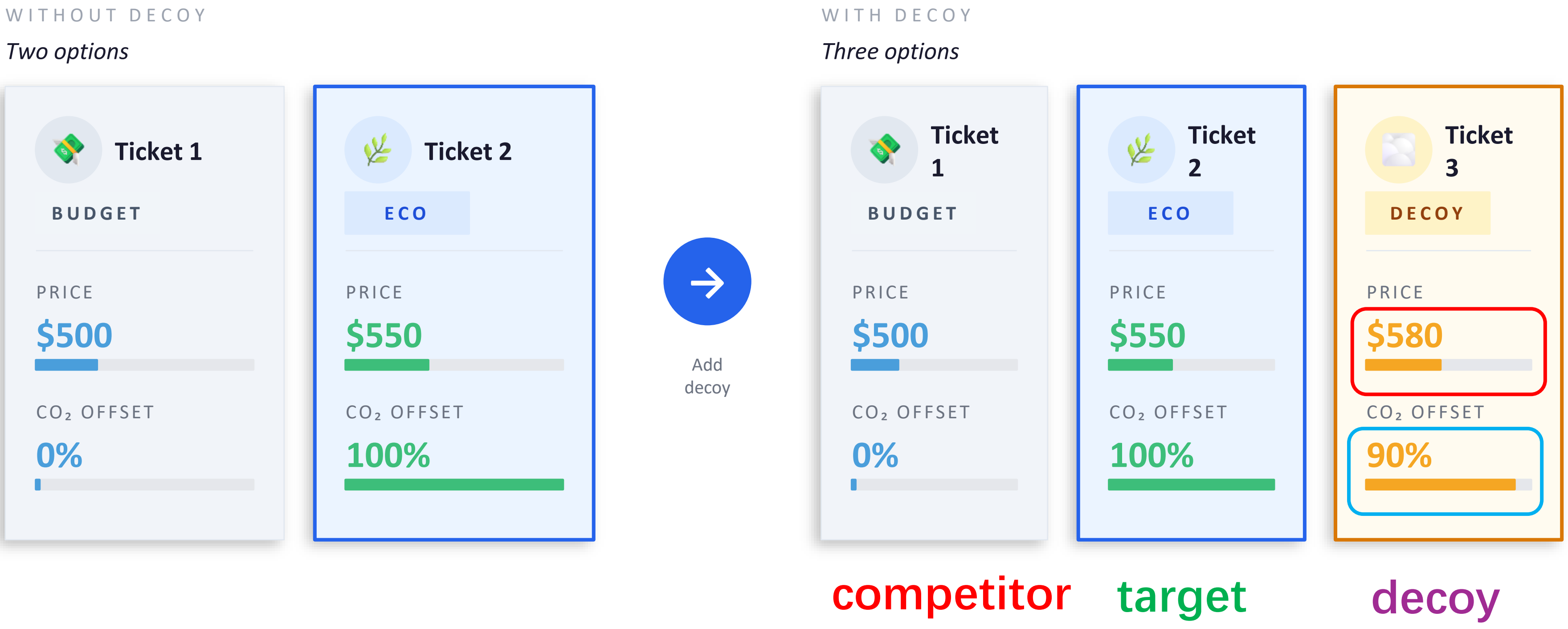
Empirical Study: Out-of-sample Choice Prediction



Eye-tracking Data: Case Study 2

Leveraging ET to Inform Optimal Decoy Configuration

Decoy is the nudge context-based strategy - where adding a third option to the binary choice set may **increase the probability of choosing** one of initial options.



- The strength of decoy effect depends on decoy parameters (here, price and offset)
- Some parameter configurations may work well while others may not work

Design of the ET Decoy Experiment in the Lab

A. Pairwise Choice Scenario

Suppose that you are planning **a 9 hour flight**. This flight produces **720.0 kg of CO_2 emissions** which is equivalent to producing **7600 1.5L plastic water bottles**.
Please choose preferred option by pressing the button below the table.

Option 1	Option 2
Amount to be paid S$ 844	Amount to be paid S$ 853
Compensated emissions 0	Compensated emissions 100%

competitor target

Option 1 Option 2

B. Choice Scenario with Decoy

Suppose that you are planning **a 9 hour flight**. This flight produces **720.0 kg of CO_2 emissions** which is equivalent to producing **7600 1.5L plastic water bottles**.
Please choose preferred option by pressing the button below the table.

Option 1	Option 2	Option 3
Amount to be paid S$ 855	Amount to be paid S$ 844	Amount to be paid S$ 853
Compensated emissions 90%	Compensated emissions 0	Compensated emissions 100%

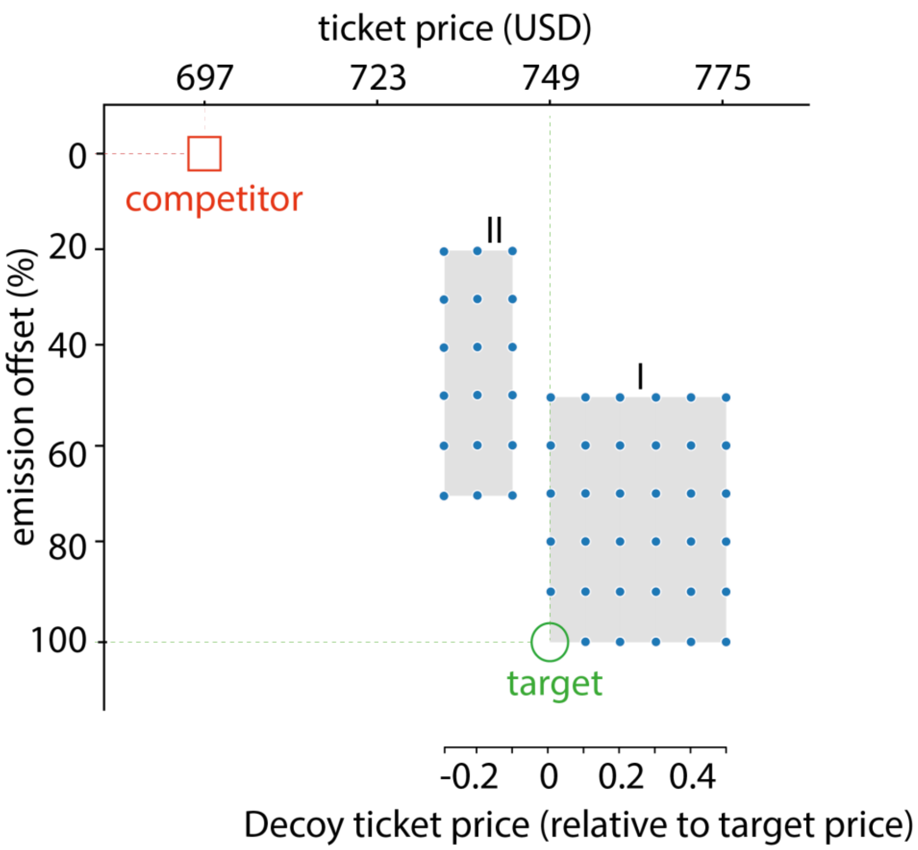
decoy

Option 1 Option 2 Option 3

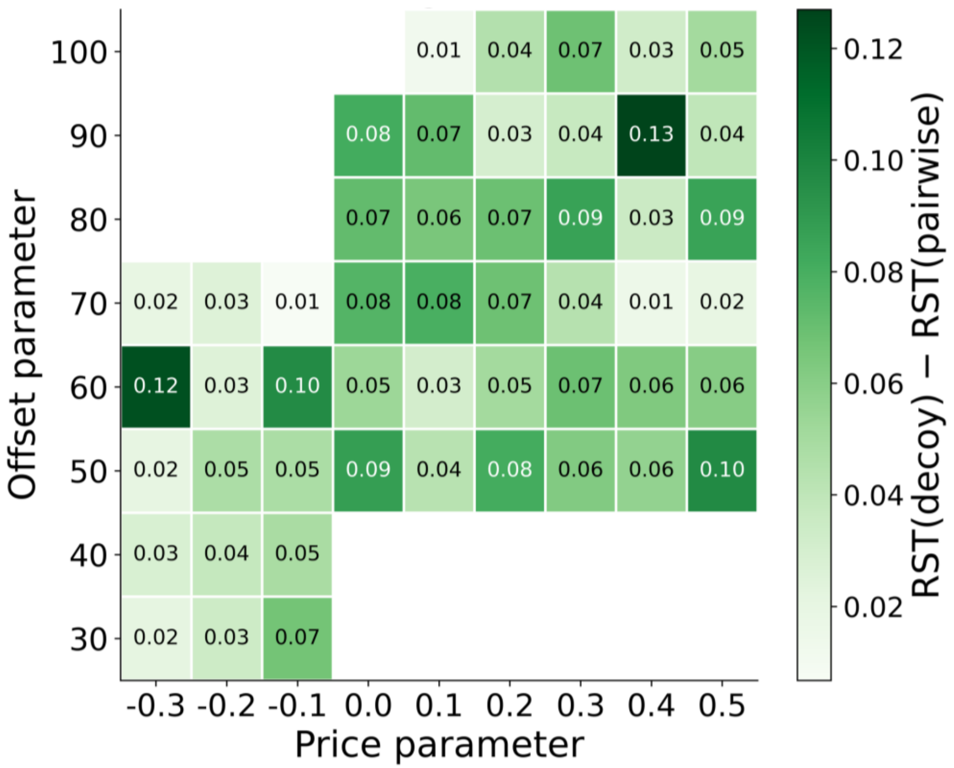
- Each participant completed **100 trials**: 50 pairwise and 50 decoy trials.
- Collected data of 100 participants

Results from the Lab-based ET experiment

Design

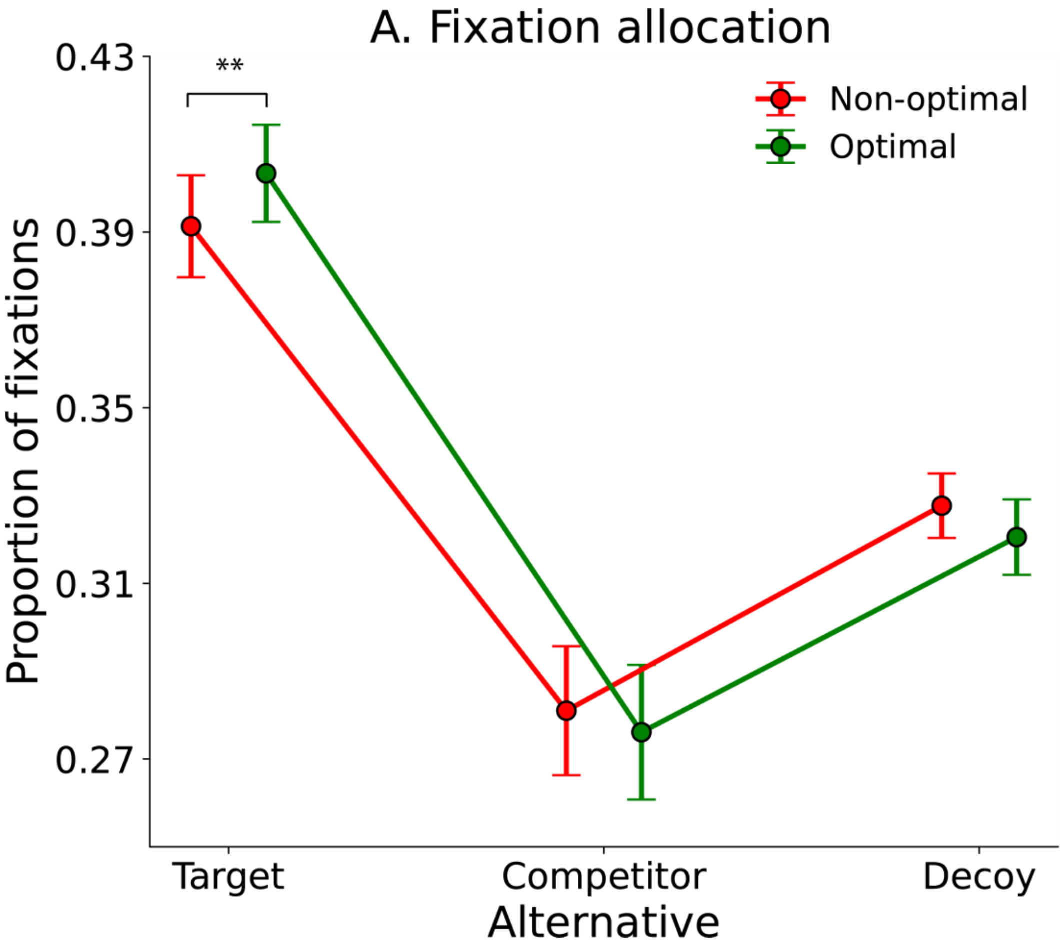


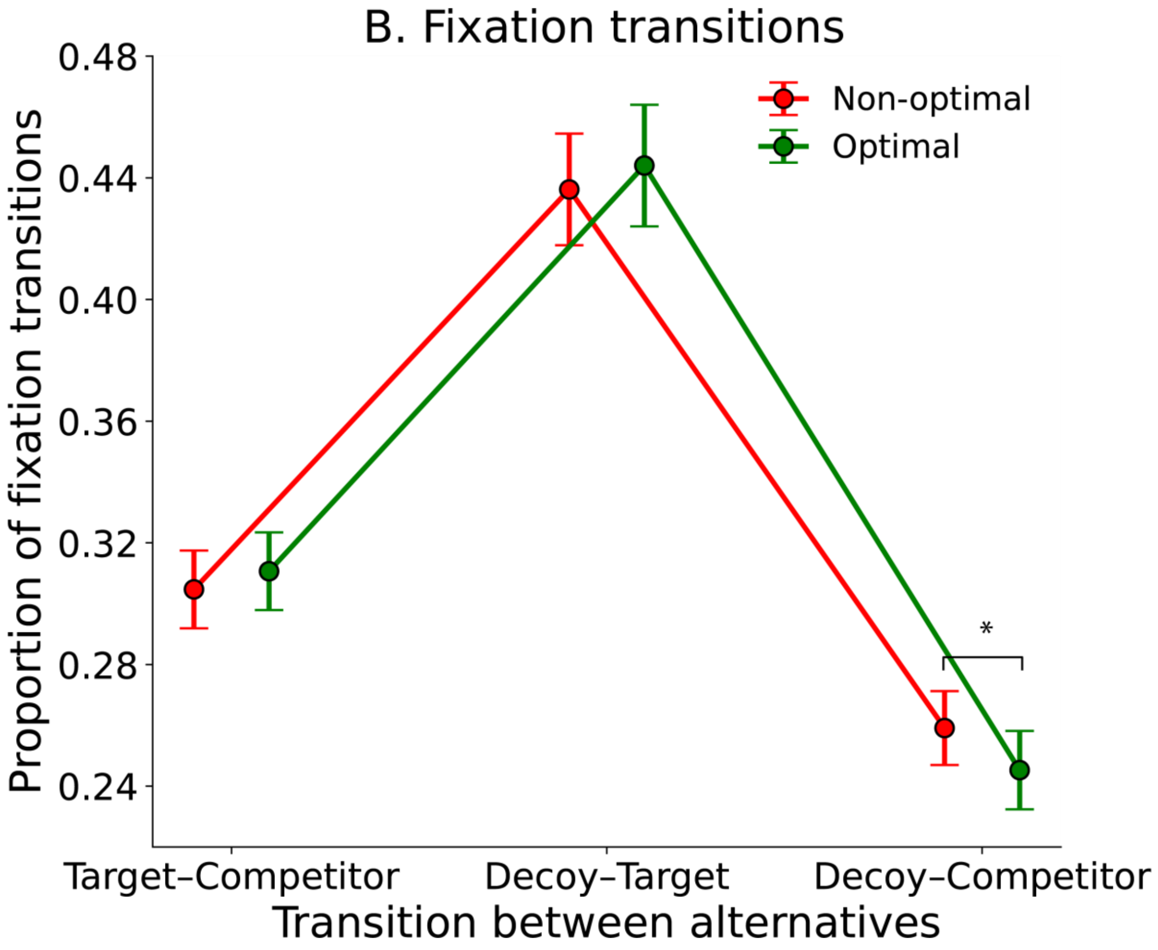
Results



$$RST = \frac{P(Target)}{P(Target) + P(Competitor)}$$

Results from the Lab-based ET experiment





Why Can MDbS Explain Decoy Effects Better?

Probability of an alternative accumulating evidence depends on its attribute **similarity** to the other alternatives in the choice set.

$$p_{n,i} = \sum_{k=1}^{Q_n} p(\text{evaluate alternative } i \text{ on attribute k})\, p(\text{ alternative } i \text{ wins a comparision on attribute } k)$$

$$p(\text{evaluate alternative } i \text{ on attribute } k) = \frac{\sum_{j=1}^{J_n} RS_{n,i,j,k}}{\sum_{i=1}^{J_n}\sum_{j=1}^{J_n}\sum_{k=1}^{Q_n} RS_{n,i,j,k}}$$

$$RS_{n,i,j,k} = e^{-\alpha D_{n,i,j,k}}$$

$$D_{n,i,j,k} = \frac{|x_{n,i,k} - x_{n,j,k}|}{|x_{n,j,k}|}$$

$$p(\text{alternative } i \text{ win a comparision on attribute } k) = \sum_{j=1}^{J_n} w_{n,i,j,k}\, p(\text{alternative } i \text{ is favored over alternative j})$$

$$p(\text{alternative } i \text{ is favored over alternative j}) = \begin{cases} \dfrac{1}{1+\exp[-\beta_1(D_{n,i,j,k} - \beta_0)]} & if\ x_{n,i,k} > x_{n,j,k} \\ 0 & \text{otherwise} \end{cases}$$

Open Questions

Key Takeaways

- ❑ There is a value of using **response time** in choice models, which is easy to collect – (i) Efficient estimation; (ii) Faster to estimate than choice-only models.

- ❑ Eye-tracking can be used for **personalized recommendations** and can also help designing **optimal nudging** strategies.

- ❑ Flexible data generating process of **SSMs** have a potential in case of predicting choices better under **contextual decisions** where cognitive biases may emerge.

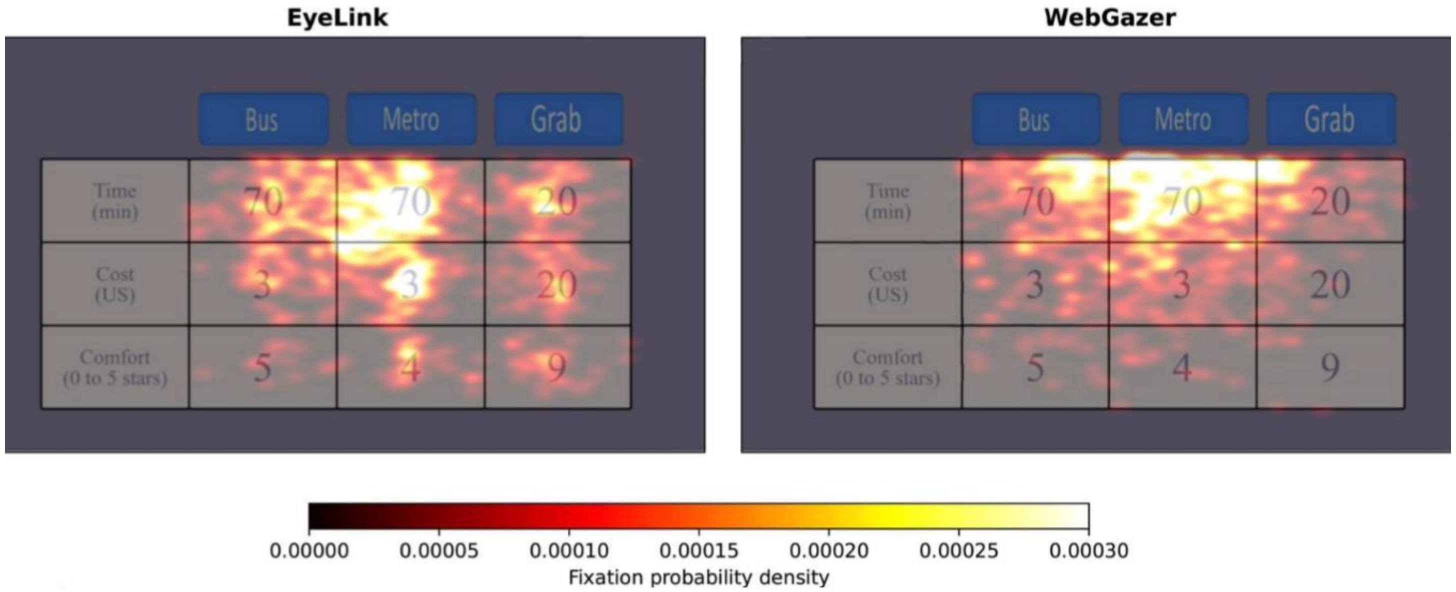
Future Research Questions

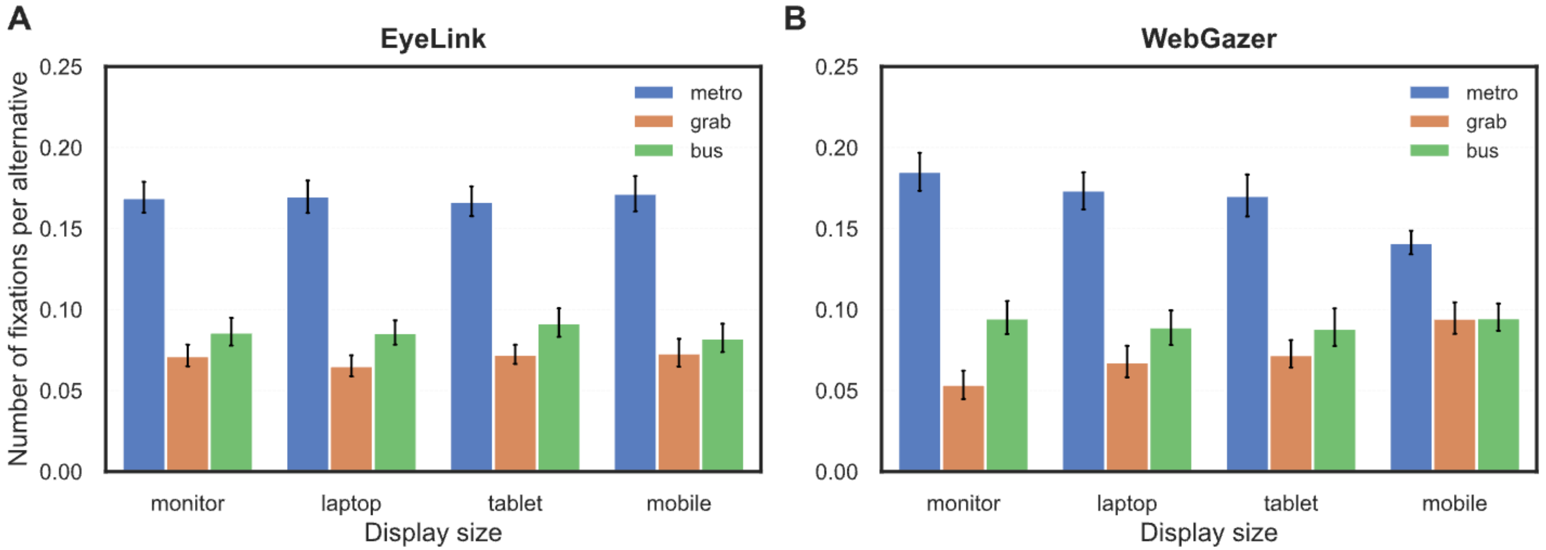
- ❓ How do we **scale** collection of eye-tracking data at scale?

- ❓ How do we compute **welfare and willingness to pay estimates** after incorporating process datasets?

Scaling Eye-tracking Through Web-Cam

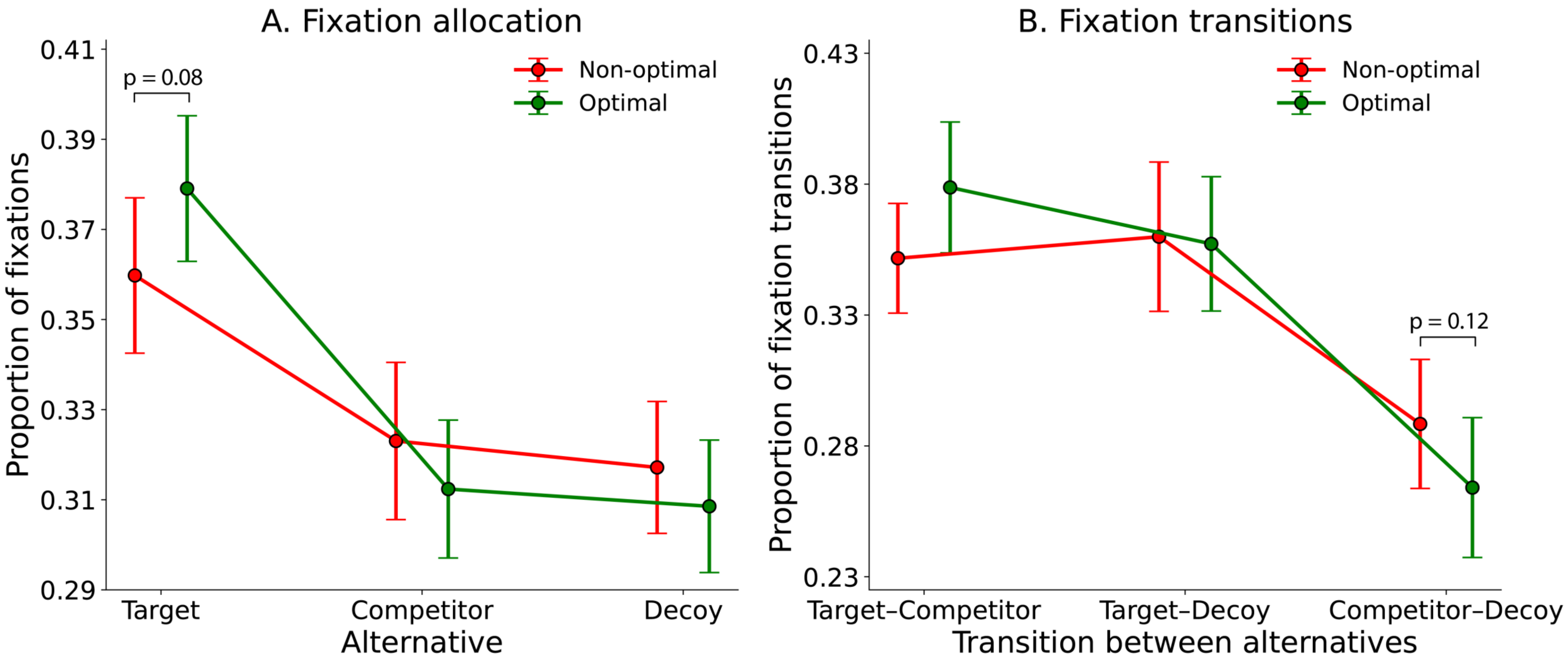
Study 1: Comparing WebGazer with Eye-link (1000 hz)





Muñoz, S., Maksimenko, V., Henriquez-Jara, B., Bansal, P., & Perez, O. D. (2025). In-lab versus web-based eye-tracking in decision-making: A systematic comparison on multiple display-size conditions mimicking common electronic devices. *Behavior Research Methods*, *57*(12), 339.

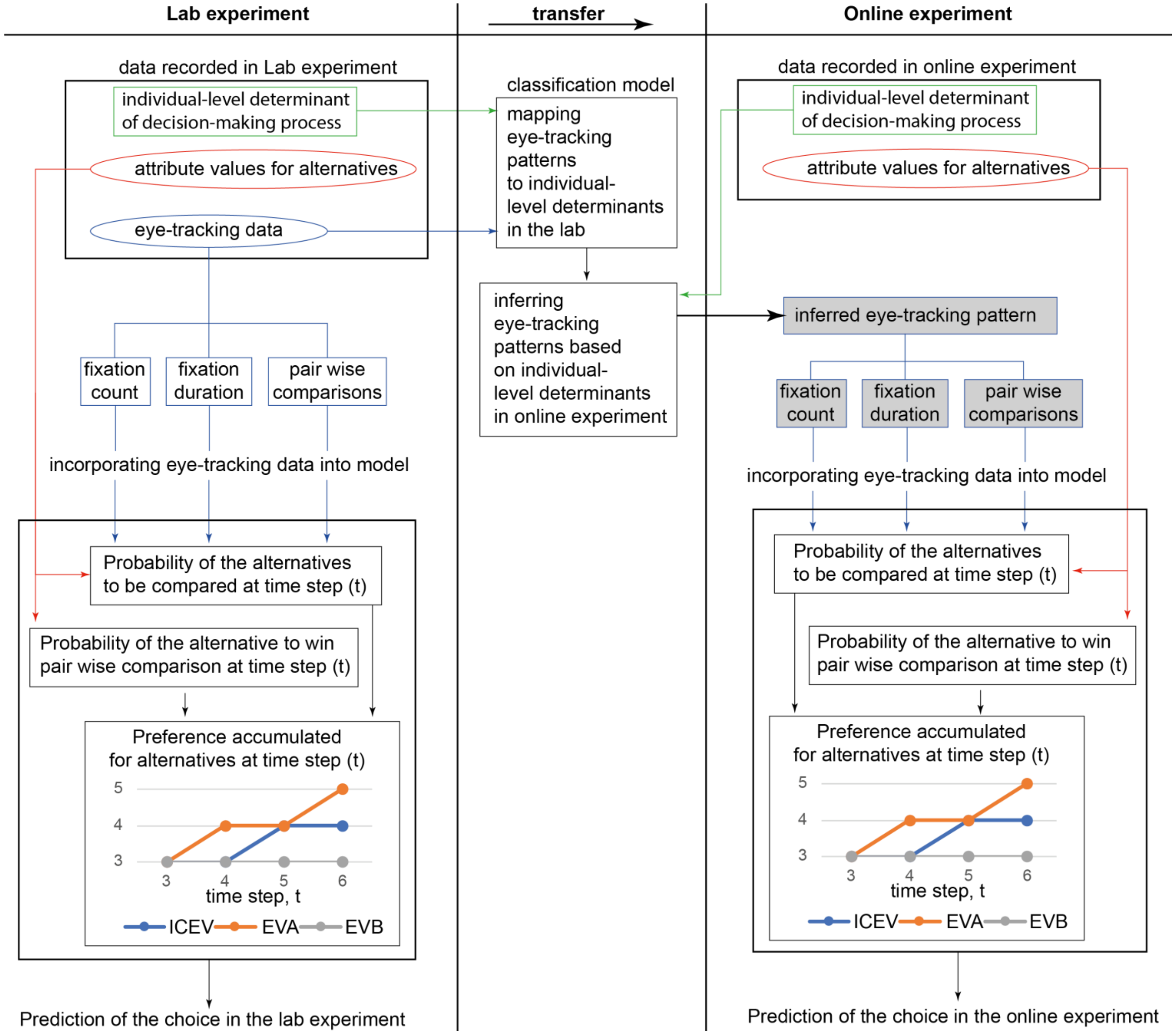
Study 2: Leveraging web-cam to find differences in gaze pattern between optimal & non-optimal decoys



How to **correct** Webcam-based eye-tracking data and find some generalizable correction factors?

Scaling Eye-tracking through Transfer from Lab to Web Surveys

Transferring Gaze Data from the Lab to Online Surveys: Framework

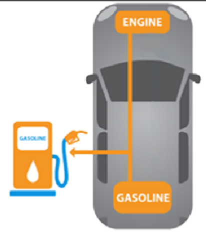
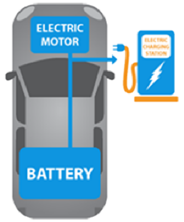
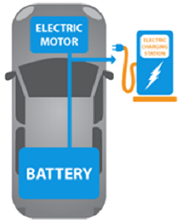


- Step 1: Conducting the same experiment in both lab and online settings
- Step 2: Segmenting respondents into different cluster using Gaussian Mixture Models (GMM) based on individual-level determinates
- Step 3: Calculating cluster-specific visual attention metrics based on lab experiment
- Step 4: Calculating cluster membership for online experiment
- Step 5: Mapping cluster-specific eye-movement patterns to corresponding respondent cluster in the online experiment

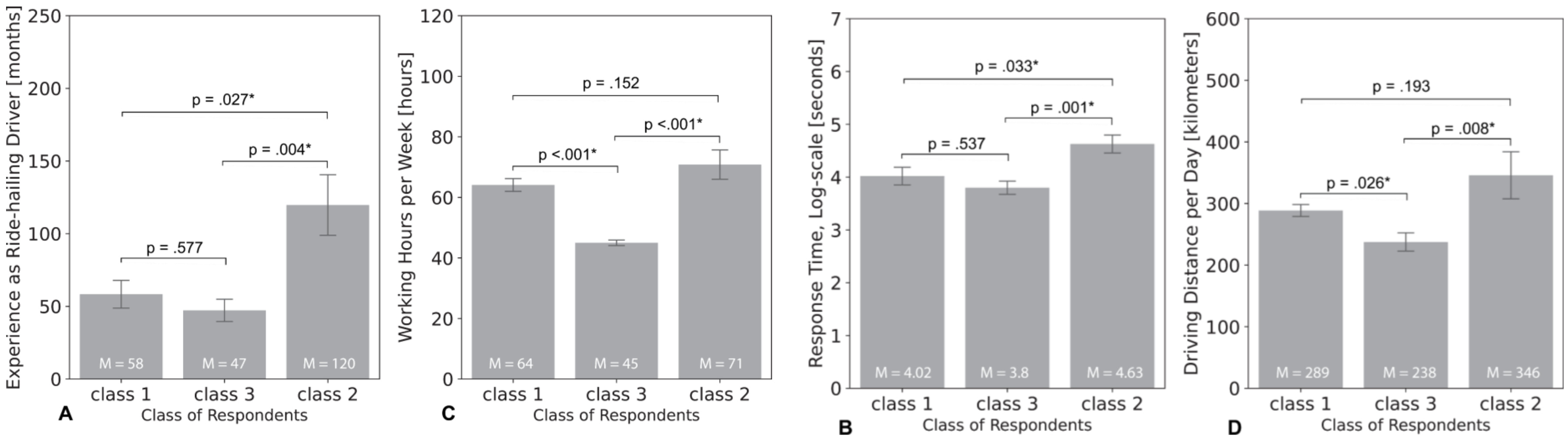
Marginal Comparability in Demographics

- Two studies exhibit marginal comparability demographic profiles

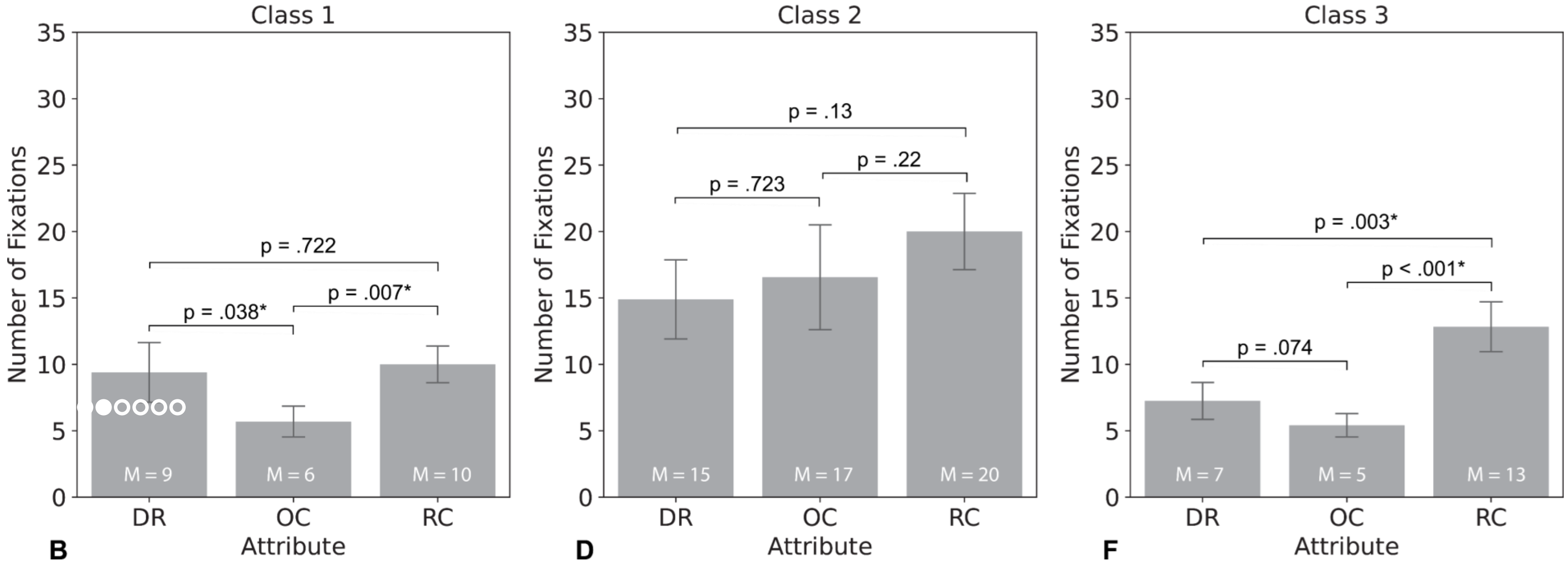
Characteristics	Lab Study	Online Study
Marital status		
Married	67.5%	63.0%
Single & Others	32.5%	37.0%
Experience as a ride-hailing/taxi driver		
Below 4 years	35.9%	9.9%
4–15 years	61.5%	77.5%
Above 15 years	2.6%	12.6%
Weekly working hours		
⩽ 70 hours	76.9%	82.1%
> 70 hours	23.1%	17.9%
Driving distance per day		
⩽ 350 km	85.4%	87.2%
> 350 km	14.6%	12.8%
Monthly net income		
⩽ S$3,000	38.5%	33.8%
> S$3,000	61.5%	66.2%

			
	Your conventional vehicle	**Electric vehicle (Model A)**	**Electric vehicle (Model B)**
Monthly renting cost	S$ 2700	S$ 2950	S$ 2900
Daily operating cost	S$ 49	S$ 20	S$ 20
Driving range with full fuel tank	600 km (refuel 15 times of times per month)	350 km (recharge 26 times per month)	200 km (recharge 45.5 times per month)

Heterogeneity in Attention across Clusters of Respondents



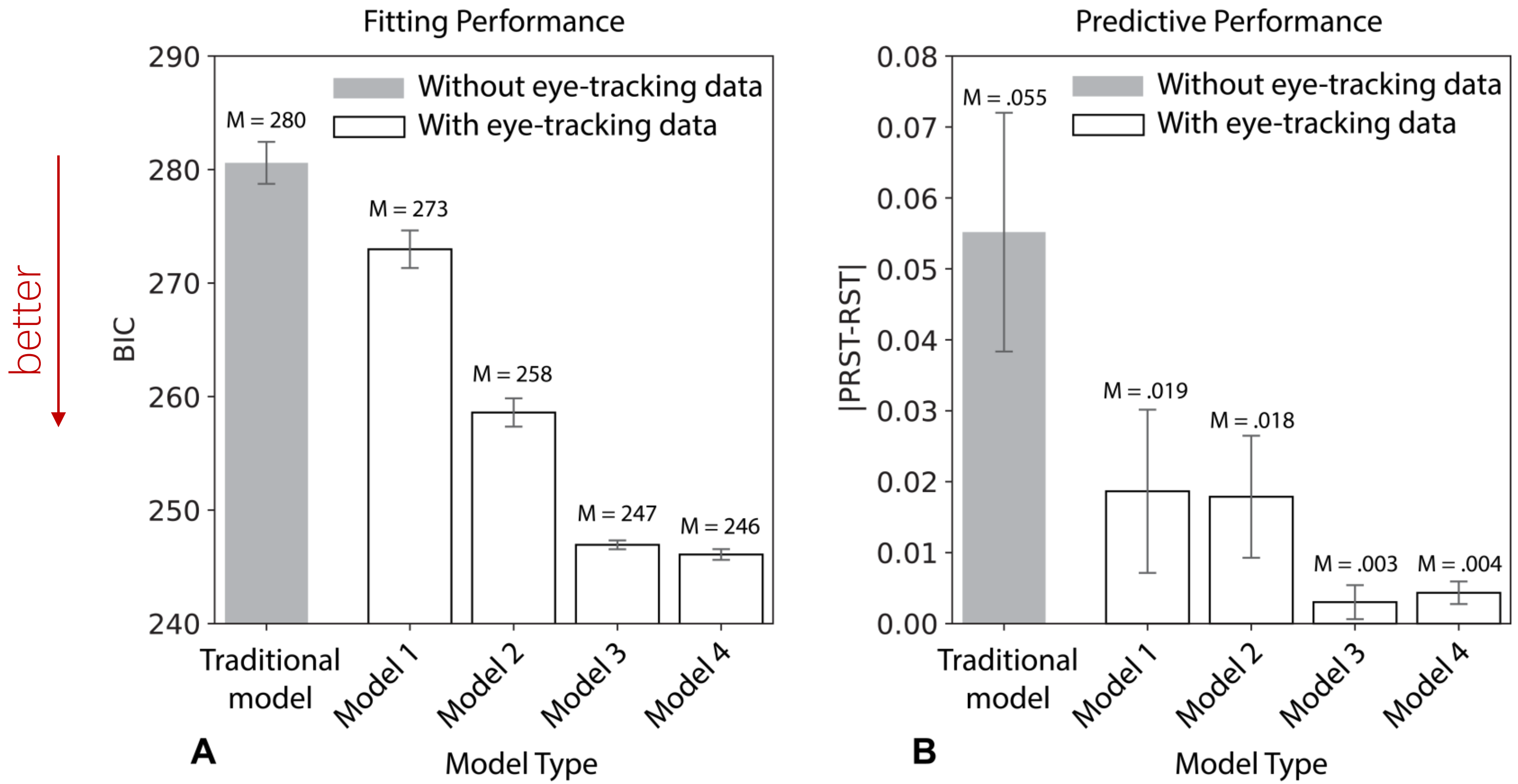
- Class 1: Longer working hours
- Class 2: Experienced drivers; longer response time
- Class 3: Shorter daily distance and working hours



- Class 1 (Longer working hours): Caring RC and DR
- Class 2 (**experienced drivers): Evaluating all attribute**
- Class 3 (shorter daily distance): Caring RC

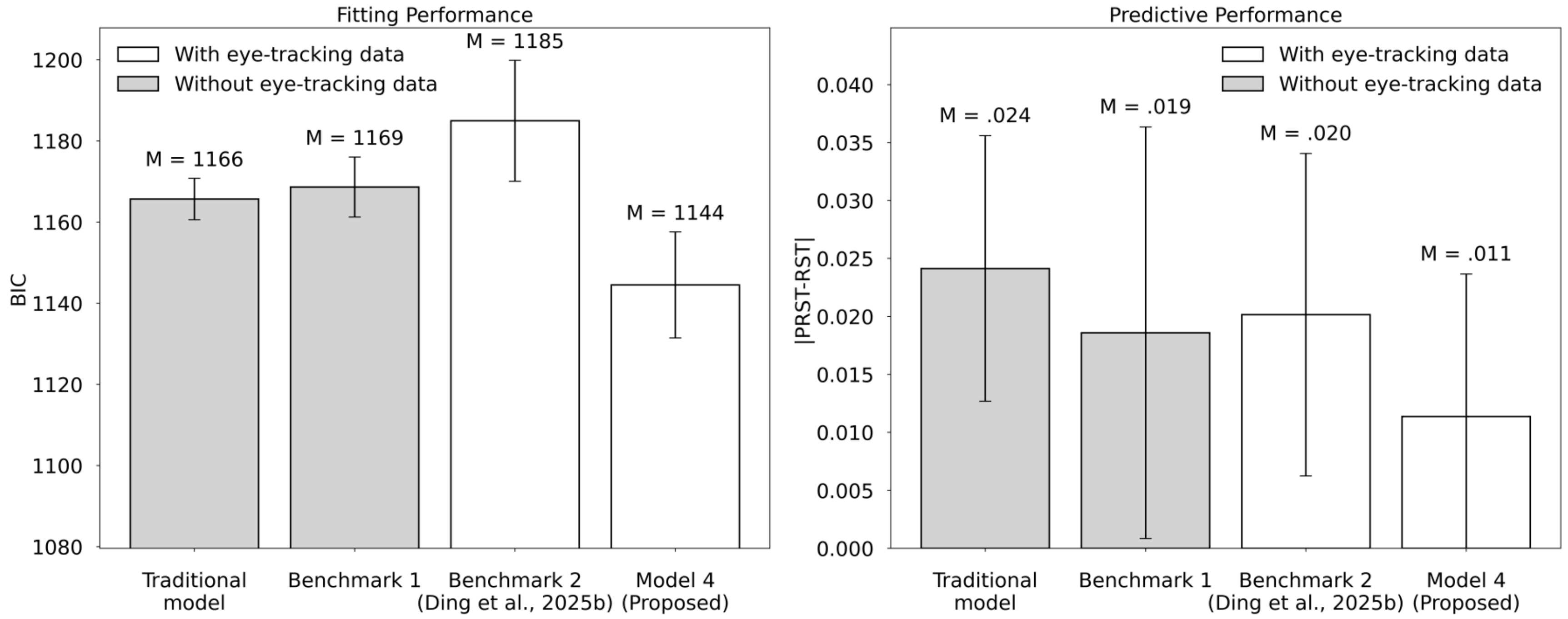
Improving Model Performance by Incorporating Gaze Data

- **Lower BIC value** (better model fitting) and **smaller |PRST-RST|** (better model predictive performance) are preferred
- **Fixation duration (model 2)** provides a more accurate reflection of respondents' attention patterns than fixation frequency (model 1)
- Models 3 and 4, which include **pairwise comparisons**, further enhanced model performance



Effectiveness of Transferring Framework

Improvement come from how attention is transferred, rather than from adding preference heterogeneity alone

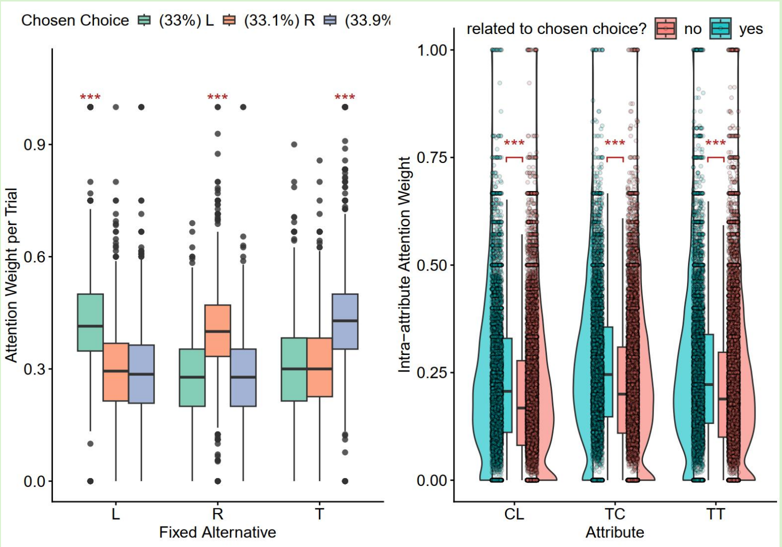


	Latent class on individual-level determinants			Attention specification (Model 4)		
	Preferences	Attention	None	Scenario-specific	Cluster-specific	None
Traditional model			√			√
Benchmark 1 (demographics only)	√					√
Benchmark 2 (scenario-level)	√			√		
Model 4 (Proposed)		√			√	

Top-down vs Bottom-up Attention

The Relation between Visual Attention and Attribute Evaluation
Motivations & Research Questions

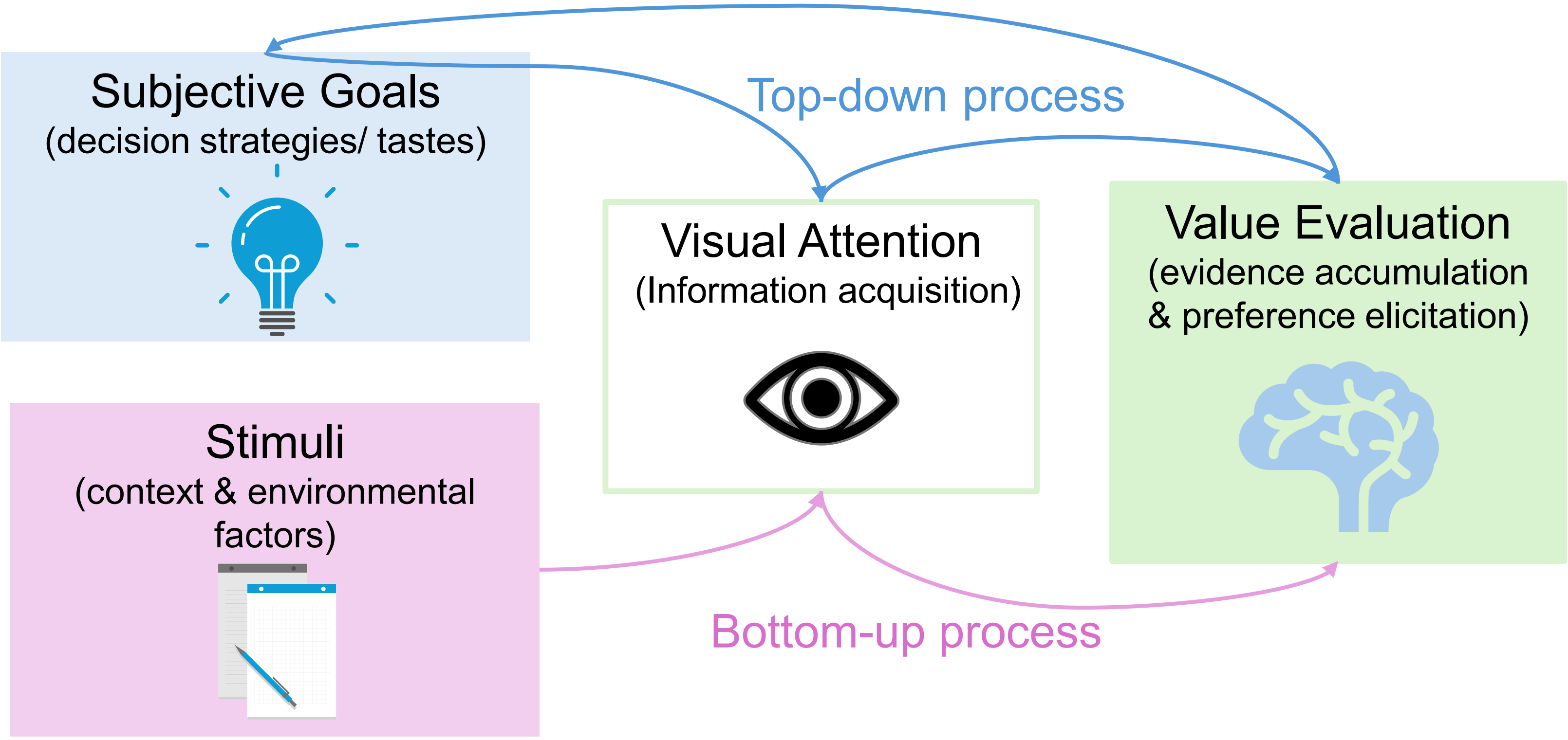
Revisit on **ET-choice Link**:



- (Left) More attention on the chosen choice;
- (Right) More intra-attribute comparisons relate to the chosen choice;

The *revealed ET data* can be used as an indicator to calibrate personal taste perturbation.

A reciprocal feedback loop! (Two processes)



Q: Can the preference & attribute evaluation be shifted by manipulating visual attention (through bottom-up process)? How?

Reference: Armel, Beaumel, and Rangel (2008), Glaholt, Wu, and Reingold (2010) and Shimojo et al. (2003)

The Relation between Visual Attention and Attribute Evaluation
Experiment Design

- **Manipulation on *Bottom-up process***
manipulate *attribute-wise* visual salience;

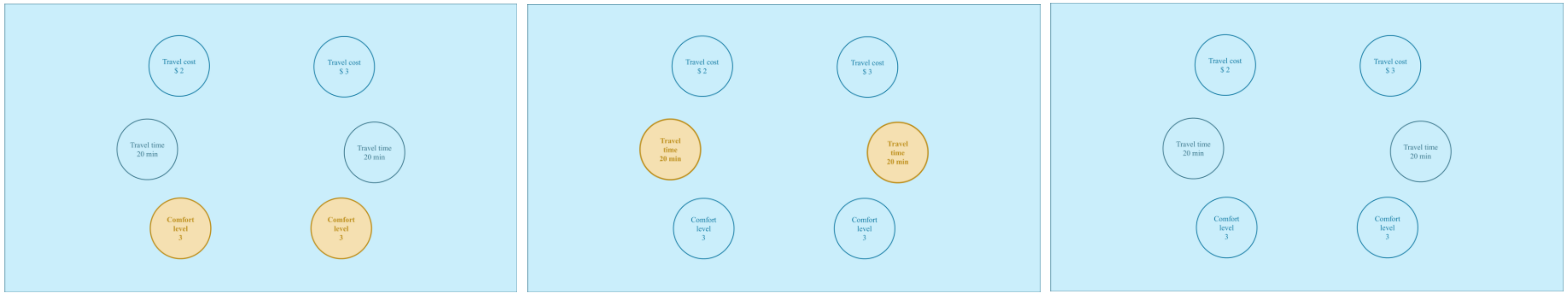
- ***Top-down process* control:**
self-reported attribute importance;

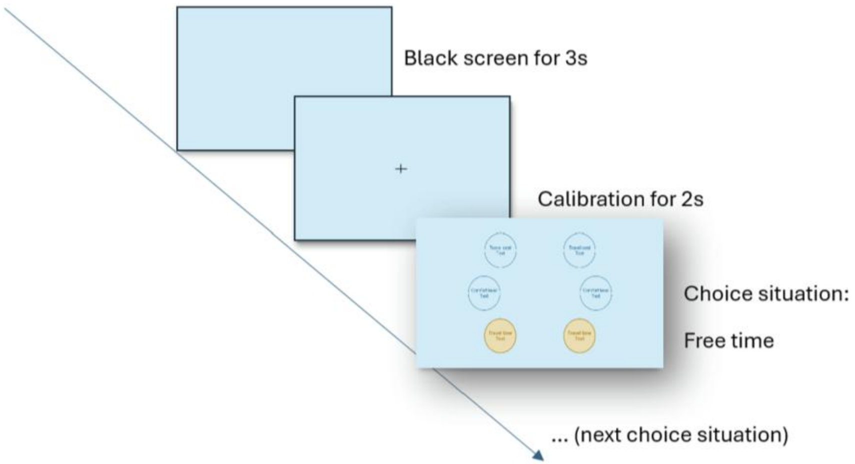
- **Choice tasks:**
Binary choice with 3 attributes, one is money, the other two is non-monetary features.

- **Choice context:**
3 different domains, including travel mode choice, risky inter-temporal choice, and tbc;

three coloring scenes

Control Scene





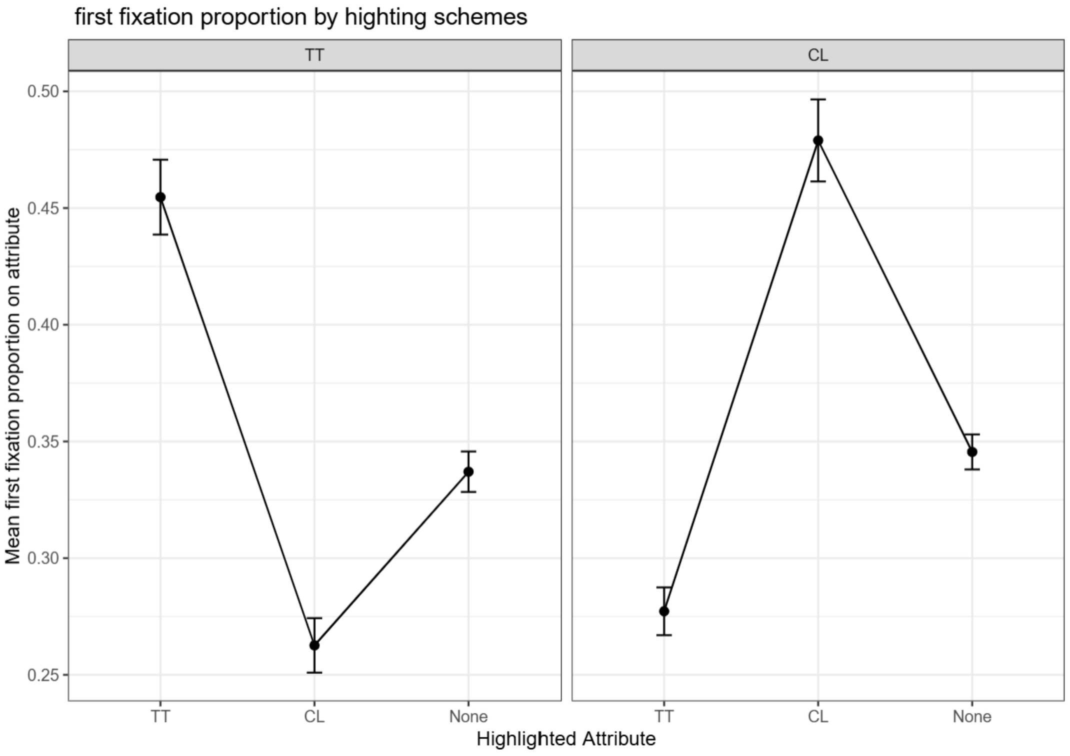
Each choice situation will be replicated under all *three coloring scenes*;

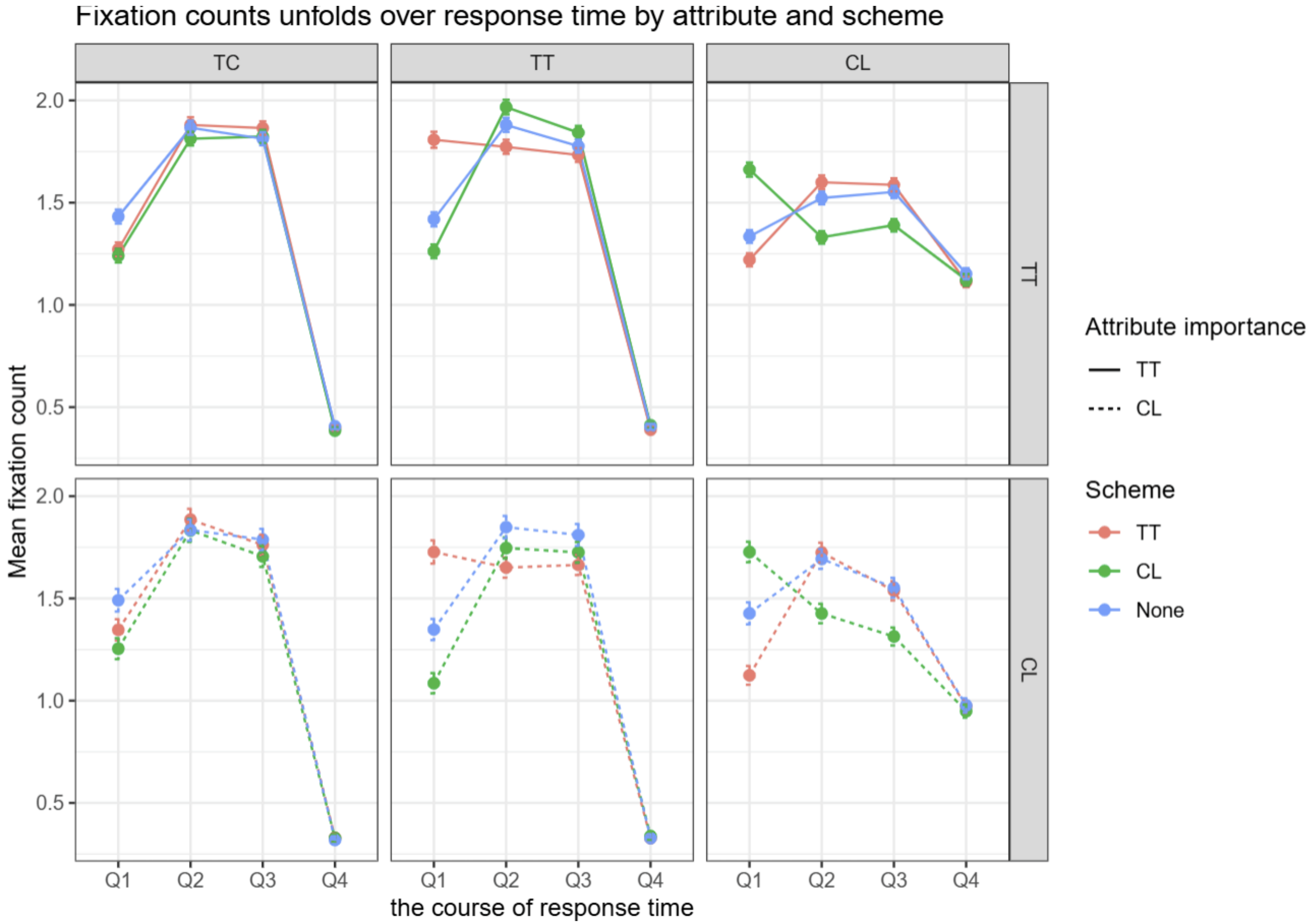
180 tasks in total, broken into 4 blocks for each subject.

The Relation between Visual Attention and Attribute Evaluation
Primary Results

Result 1:
Visual salience manipulation effectively trigger the **bottom-up process**.

Result 2:
Bottom-up process affects **top-down control** by intervening the attention allocation sequence and weights, whose magnitude is interacted with **subjective goals**.





Behavioural
COMPUTATIONAL SCIENCE LAB

NUS
National University of Singapore

LinkedIn Page

(Please follow for updates)

Advancing MDbS with Different Visual Attention Metrics

- Model 1: The more frequently it is fixated on, the more likely it is to be evaluated

$$p_n(\text{evaluate alternative } i \text{ on attribute } k) = \frac{\sum_{j=1}^{J_n} RS_{n,i,j,k} + exp\left(\gamma E_{n,i,k}^{Count}\right)}{\sum_{k=1}^{Q_n} \sum_{i=1}^{J_n} \left(\sum_{j=1}^{J_n} RS_{n,i,j,k} + exp\left(\gamma E_{n,i,k}^{Count}\right)\right)}$$

$E_{n,i,k}^{Count}$: relative fixation frequency of alterative i **on attribute** k for respondent n; $RS_{n,i,j,k}$: relative similarity between the attribute value of alterative i and j **on attribute** k for respondent n; γ: importance of the eye-tracking data

- Model 2: The longer the fixation duration it receives, the more likely it is to be evaluated

$$p_n(\text{evaluate alternative } i \text{ on attribute } k) = \frac{\sum_{j=1}^{J_n} RS_{n,i,j,k} + exp\left(\gamma E_{n,i,k}^{Duration}\right)}{\sum_{k=1}^{Q_n} \sum_{i=1}^{J_n} \left(\sum_{j=1}^{J_n} RS_{n,i,j,k} + exp\left(\gamma E_{n,i,k}^{Duration}\right)\right)}$$

$E_{n,i,k}^{Duration}$: relative fixation duration of alterative i **on attribute** k for respondent n.

- Model 3: The more frequently pair-wise comparison it receives, the more likely it is to be evaluated

$$p_n(\text{evaluate alternative } i \text{ on attribute } k) = \frac{\sum_{j=1}^{J_n} RS_{n,i,j,k} A_{n,i,j,k}}{\sum_{k=1}^{Q_n} \sum_{i=1}^{J_n} \left(\sum_{j=1}^{J_n} RS_{n,i,j,k} A_{n,i,j,k}\right)} \qquad A_{n,i,j,k} = \frac{exp(\gamma \tilde{p}_{n,i,j,k})}{\sum_{j=1}^{J_n} exp(\gamma \tilde{p}_{n,i,j,k})}$$

$\tilde{p}_{n,i,j,k}$: the probability of pair-wise comparison between alterative i and j on attribute k for respondent n; $A_{n,i,j,k}$: relative importance of pair-wise comparison between alterative i and j on attribute k for respondent n

- Model 4: The more frequent pairwise comparisons and the longer the fixation duration it receives, the more likely it is to be evaluated

$$p_n(\text{evaluate alternative } i \text{ on attribute } k) = \frac{\left(\sum_{j=1}^{J_n} RS_{n,i,j,k} A_{n,i,j,k}\right) + E_{n,i,k}^{Duration}}{\sum_{k=1}^{Q_n} \sum_{i=1}^{J_n} \left(\sum_{j=1}^{J_n} RS_{n,i,j,k} A_{n,i,j,k} + E_{n,i,k}^{Duration}\right)}$$