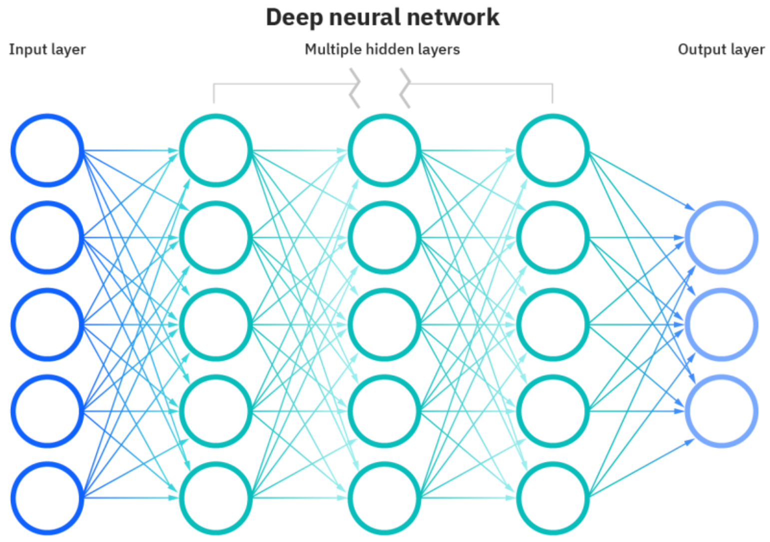
Behaviour Models, Machine Learning, and Psychology

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

UTILITY THEORY

+



Part 1: Machine Learning and Behavior Models

Motivation

Indirect Utility = Systematic Utility + Idiosyncratic Error Term

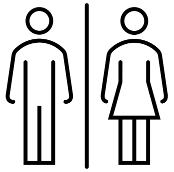
Flexible systematic utility?

(i) **Non-linear** effect of each alternative-specific attribute

(ii) **Interaction** effects of multiple alternative-specific attributes

(iii) **Interaction** effects of alternative- and individual-specific attributes (taste heterogeneity)

(iv) **Non-linear** effect of each individual-specific attribute and their interaction effect

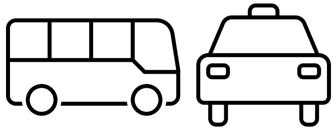


Individual attributes

Attributes	Individual
Income (INC)	3-5 million KRW
Full-time (FUL)	Yes

Alternative attributes

Attributes	Bus	Grab
Travel cost (TC)	\$1.5	\$12
Travel time (TT)	40 mins	20 mins
Waiting time (WT)	15 mins	5 mins



Interpretability vs Predictability

What exactly is 'interpretability'?

- The definition of interpretability is domain-dependent

Interpretability
(DCM-context)

Explainability

Ability to measure the **attribute-specific effect** on the utility (choice probability)

Post-analysis techniques
Individual conditional expectation (ICE)

Trustworthiness

Ability to impose **domain-specific constraints** on the attribute-specific effect

Inherent model structure
Monotonicity constraints

Monotonicity Constraint: Example

Imposing monotonicity constraints in linear function with first-order alternative-specific interactions:

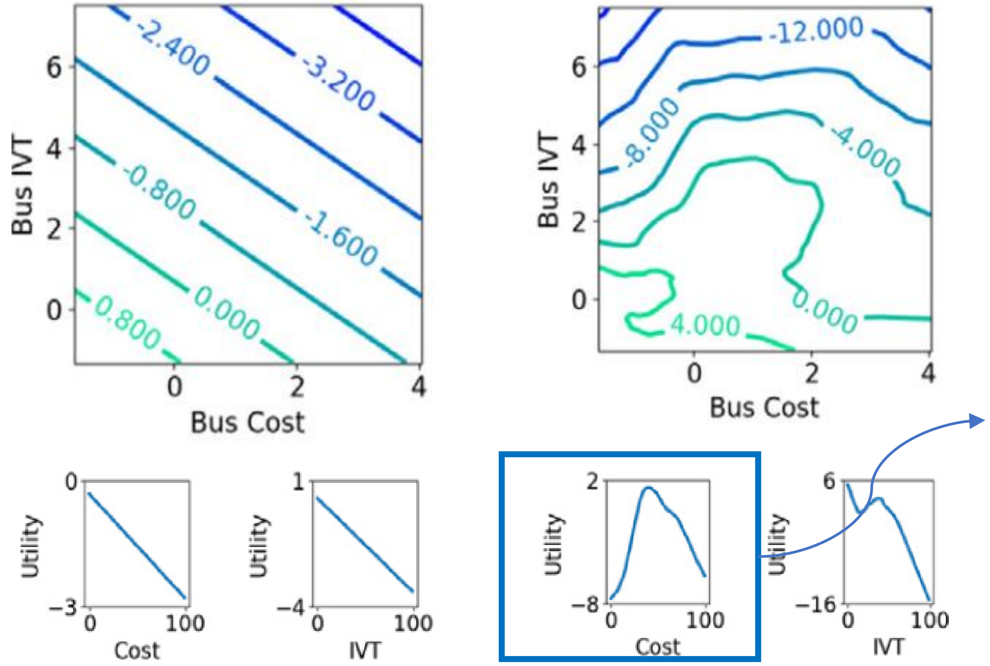
$$U(TT, TC) = \beta_0 + \beta_1 TT + \beta_2 TC + \beta_3 TT \times TC$$

$\beta_1 + \beta_3 TC < 0$; $\beta_2 + \beta_3 TT < 0$ for the entire domain of TT and TC.

Imagine the difficulty in case of non-linear function and multiple attributes.

Interpretability

Unreasonable attribute-specific effect of Discrete Choice Models with Deep Neural Network (DCM-DNN) at some attribute-level (Wang et al., 2021)



Incorrect positive effect of the travel cost on utility at some attribute levels

Lack of 'trustworthiness'

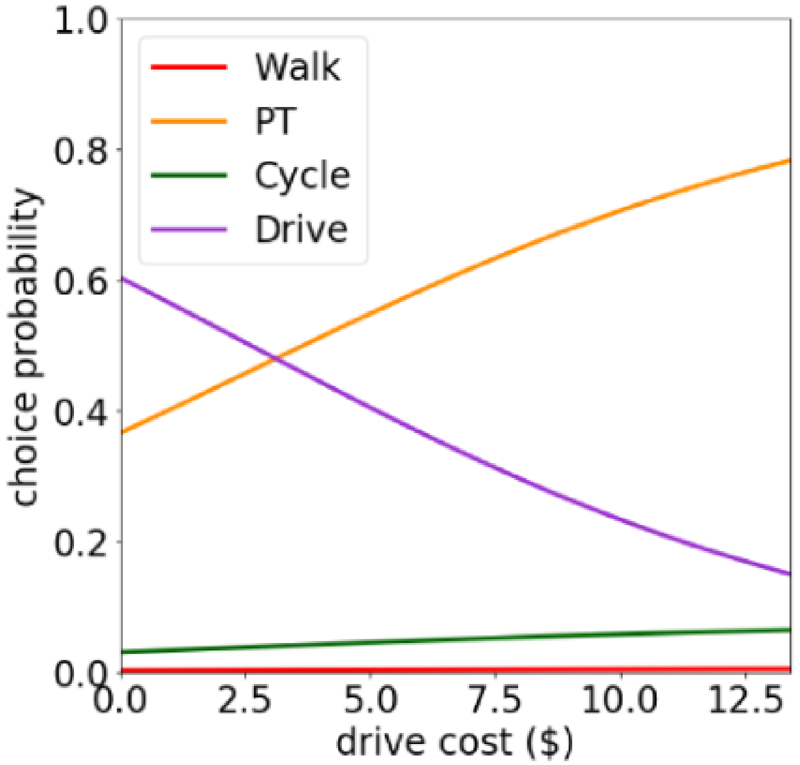
MNL
(50.6% accuracy)

DNN
(55.8% accuracy)

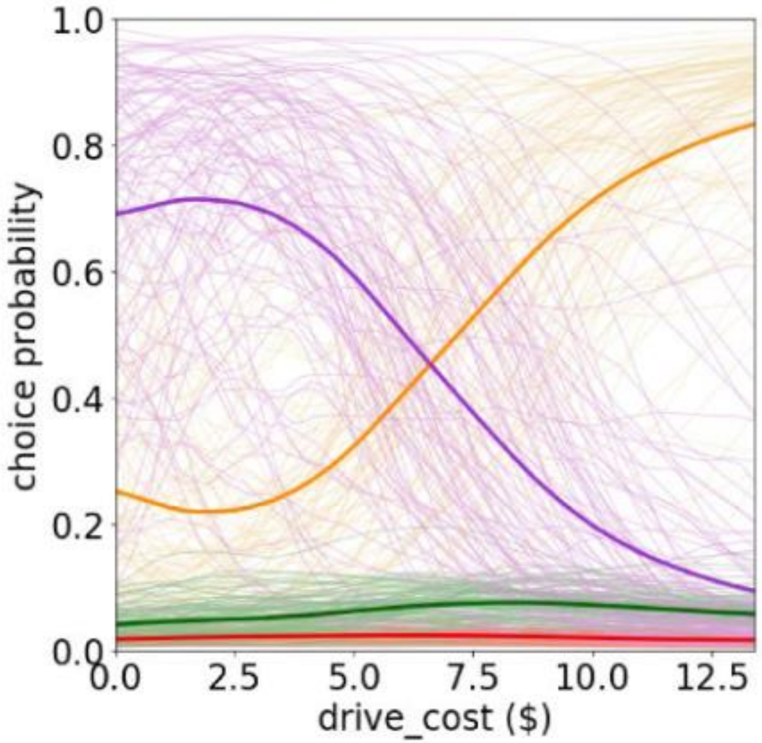
Wang, S., Mo, B., & Zhao, J. (2021). Theory-based residual neural networks: A synergy of discrete choice models and deep neural networks. *Transportation Research Part B: Methodological*, *146*, 333-358

Interpretability of DNN (Wang et al., 2020)

Unreasonable Individual-level attribute-specific effect of DCM-DNN for some individuals



Changes in choice probability according to drive cost (MNL)



Changes in choice probability according to drive cost (DCM-DNN)

<u>Over-estimated interaction effects</u>

<u>Unstable individual-level effects</u>

Lack of '<u>trustworthiness</u>'

Wang, S., Wang, Q., & Zhao, J. (2020). Deep neural networks for choice analysis: Extracting complete economic information for interpretation. Transportation Research Part C: Emerging Technologies

Interpretability of DNN (Han et al., 2022)

Activation function: Rectified linear or exponential

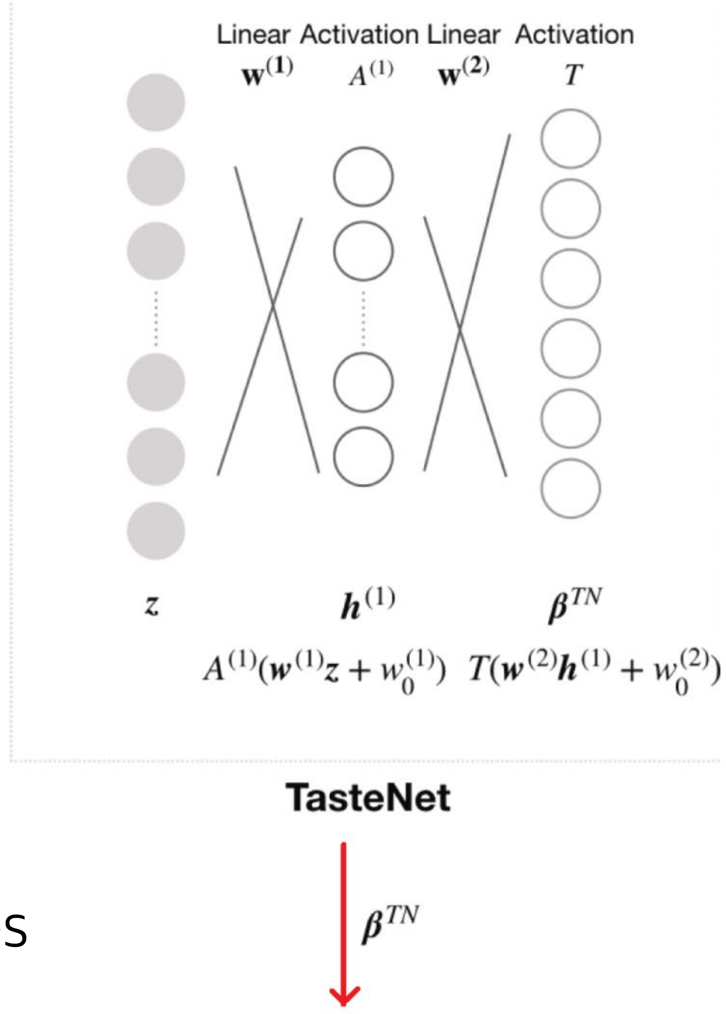
⬇

Ensures Monotonic Effect

What is Missing?

(i) **Non-linear** effect of each alternative-specific attribute

(ii) **Interaction** effects of multiple alternative-specific attributes



Systematic utility

$$V_i = \beta^{TN} x_i^{TN} + \beta^{MNL} x_i^{MNL}, \forall i \in C$$

Han, Y., Pereira, F.C., Ben-Akiva, M., Zegras, C., 2022. A neural-embedded discrete choice model: Learning taste representation with strengthened interpretability. Transportation Research Part B: Methodological 163, 166-186.

Summary of the Literature

Considerations in ideal systematic utility specification

(i) **Non-linear** effect of each alternative-specific attribute
(ii) **Interaction** effects of multiple alternative-specific attributes
(iii) **Interaction** effects of alternative- and individual-specific attributes (taste heterogeneity)
(iv) **Non-linear** effect of each individual-specific attribute and their interaction effect
(v) **Population** level trustworthiness of alternative-specific attributes
(vi) **Individual** level trustworthiness of alternative-specific attributes

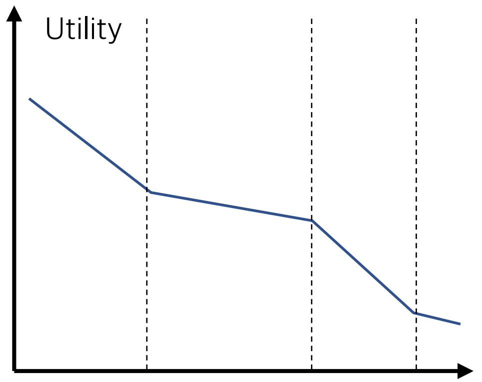
Authors	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Wang et al. (2020)	✓	✓	✓	✓		
Wang et al. (2021)	✓	✓	✓	✓		
Wong and Farooq (2021)	✓	✓	✓	✓		
Sifringer et al. (2020)			✓	✓	✓	
Han et al. (2022)			✓	✓	✓	✓
Kim and Bansal (2023)	✓	✓	✓	✓	✓	✓

Han, Y., Pereira, F.C., Ben-Akiva, M., Zegras, C., 2022. A neural-embedded discrete choice model: Learning taste representation with strengthened interpretability. Transportation Research Part B: Methodological 163, 166-186.
Sifringer, B., Lurkin, V., Alahi, A., 2020. Enhancing discrete choice models with representation learning. Transportation Research Part B: Methodological 140, 236-261.
Wang, S., Mo, B., Zhao, J., 2020. Deep neural networks for choice analysis: Architecture design with alternative-specific utility functions. Transp Res Part C Emerg Technol 112, 234-251.
Wang, S., Mo, B., Zhao, J., 2021. Theory-based residual neural networks: A synergy of discrete choice models and deep neural networks. Transportation Research Part B: Methodological 146, 333-358.
Wong, M., Farooq, B., 2021. ResLogit: A residual neural network logit model for data-driven choice modelling. Transp Res Part C Emerg Technol 126, 103050.

Objectives

Theory-constrained data-driven methods

- Discrete choice model with Lattice network (**DCM-LN**) (Gupta et al., 2016).
  - ✓ Efficiently implementing the **monotonic constraints** at individual-level
  - ✓ Capturing attribute-wise non-linear effect using piece-wise linear specification → non-linearity
  - ✓ Capturing complex interactions between the attributes → taste heterogeneity



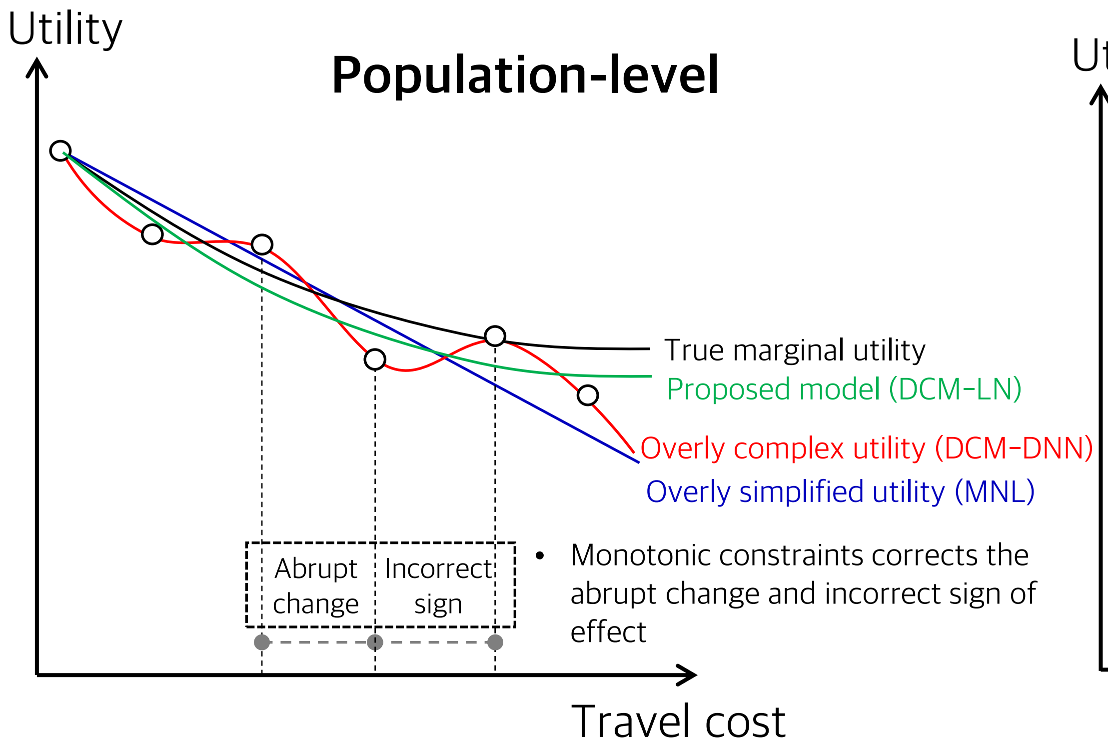
Example of non-linear effects

Gupta, M., Cotter, A., Pfeifer, J., Voevodski, K., Canini, K., Mangylov, A., Moczydlowski, W., & van Esbroeck, A. (2016). Monotonic Calibrated Interpolated Look-Up Tables. Journal of Machine Learning Research

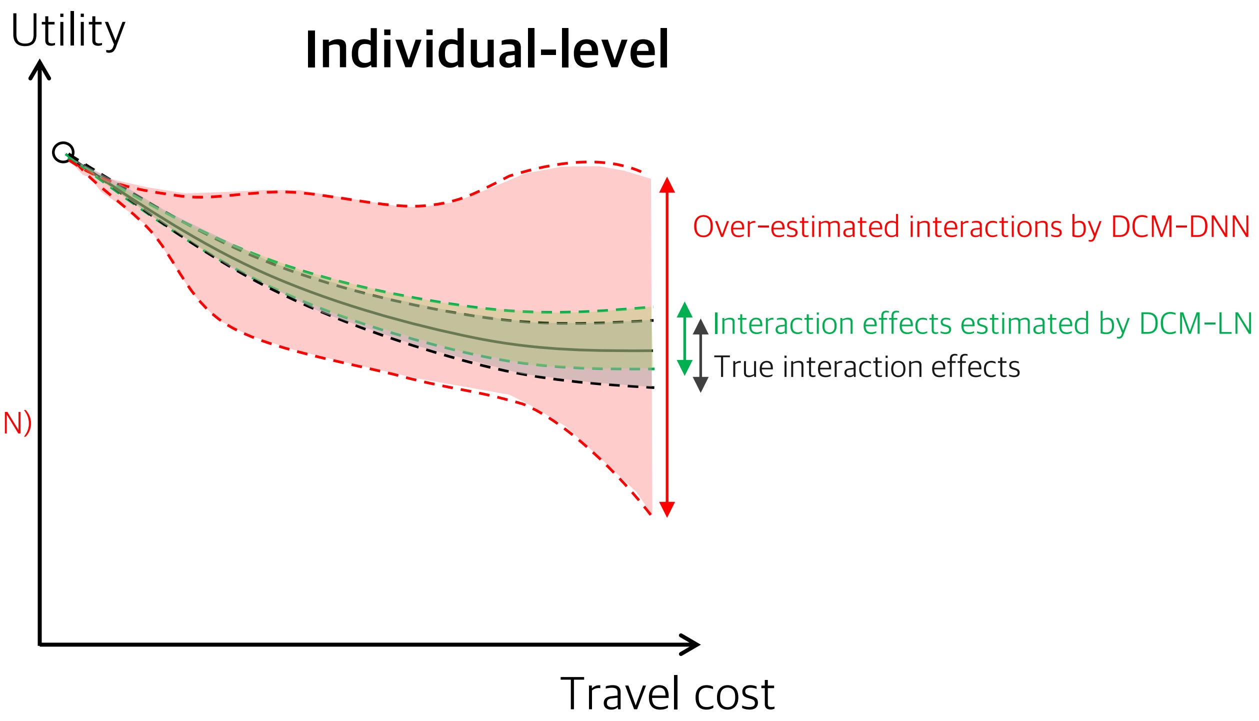
Objectives

‘Monotonicity’ is critical in data-driven learning of systematic utility

- **DCM-DNN** overfits to the data (overly complex) while the **MNL** underfits to the data (overly simplified)
- Monotonicity constraints correct the attribute-level abrupt changes and incorrect sign of effect
    - ✓ DCM-LN reduce the overfitting by theory-driven regularizations



- Monotonic constraints corrects the abrupt change and incorrect sign of effect



Methods

How can Lattice Network Impose 'Monotonicity' constraints?

Compute lattice function value for any (x_1, x_2)

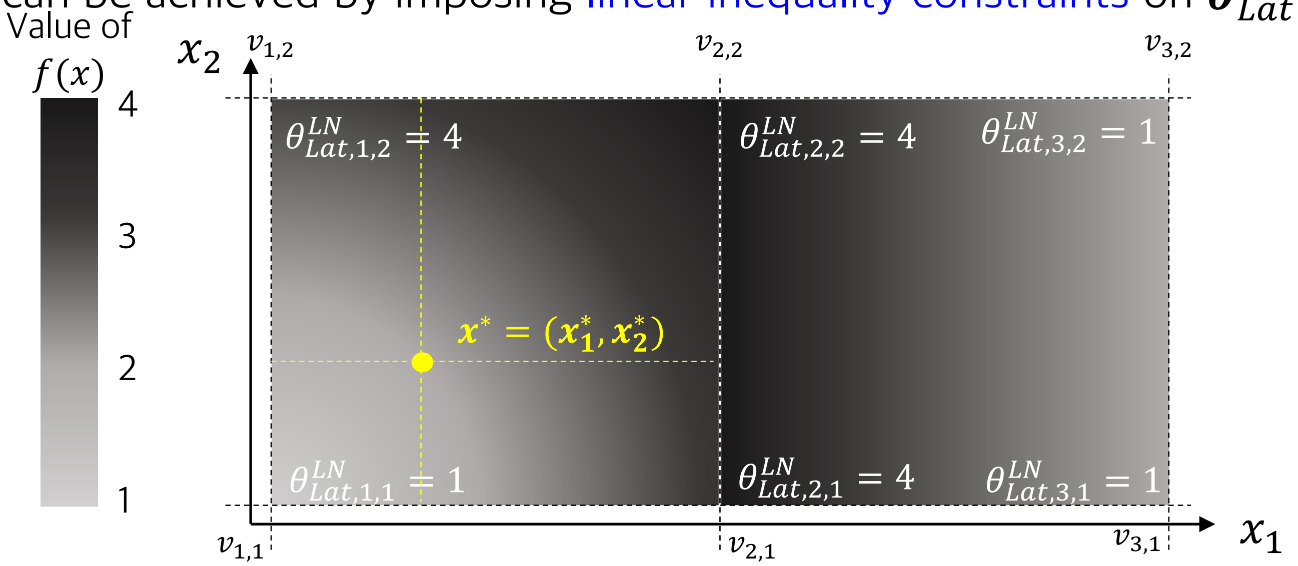
Consider 3×2 lattice layer (i.e., 3 vertices on x_1 and 2 vertices on x_2 dimension)

$\boldsymbol{\theta}_{Lat}^{LN} = \{\theta_{Lat,1,1}^{LN}, \theta_{Lat,1,2}^{LN}, \theta_{Lat,2,1}^{LN}, \theta_{Lat,2,1}^{LN}, \theta_{Lat,3,1}^{LN}, \theta_{Lat,3,1}^{LN}\}$ are model parameters.

$$f(\boldsymbol{x}^*) = \boldsymbol{\theta}_{Lat}^{LN} \cdot \psi(\boldsymbol{x}^*) = \theta_{Lat,1,1}^{LN}\psi_{1,1}(\boldsymbol{x}^*) + \theta_{Lat,2,1}^{LN}\psi_{2,1}(\boldsymbol{x}^*) + \theta_{Lat,1,2}^{LN}\psi_{1,2}(\boldsymbol{x}^*) + \theta_{Lat,2,2}^{LN}\psi_{2,2}(\boldsymbol{x}^*)$$

$\psi(\boldsymbol{x}^*)$ is a multi-linear interpolation weights for $\boldsymbol{x}^*$ which is function of corresponding vertex values $\boldsymbol{v}$. Note that, it is NOT a parameter.

Monotonicity constraints can be achieved by imposing linear inequality constraints on $\boldsymbol{\theta}_{Lat}^{LN}$.



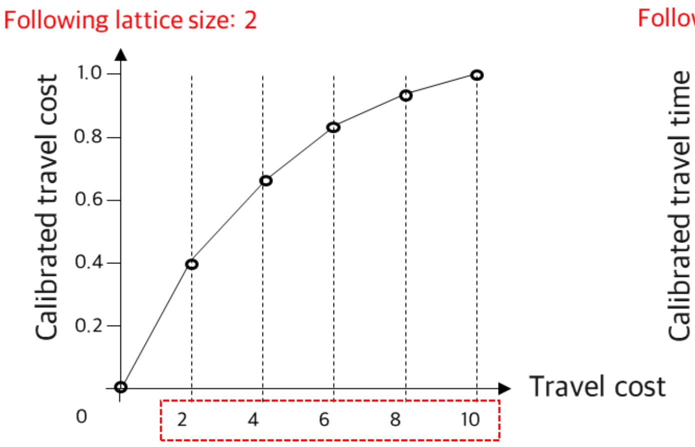
Methods

Lattice with monotonic constraints

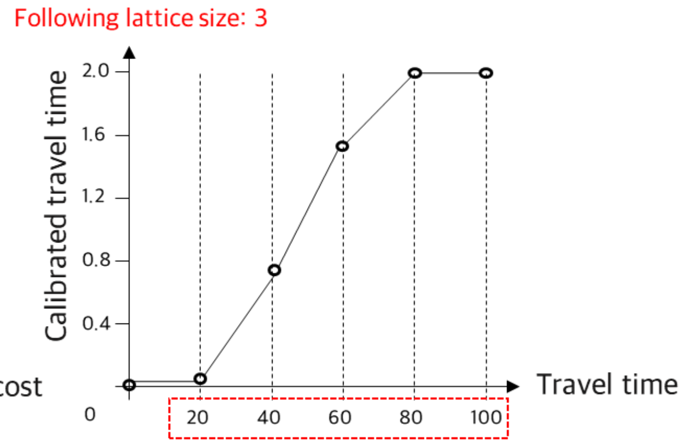
- Hyper-parameters: lattice size (model complexity)
- Model parameters : edge values

Calibrator: piecewise linear function with monotonic constraints

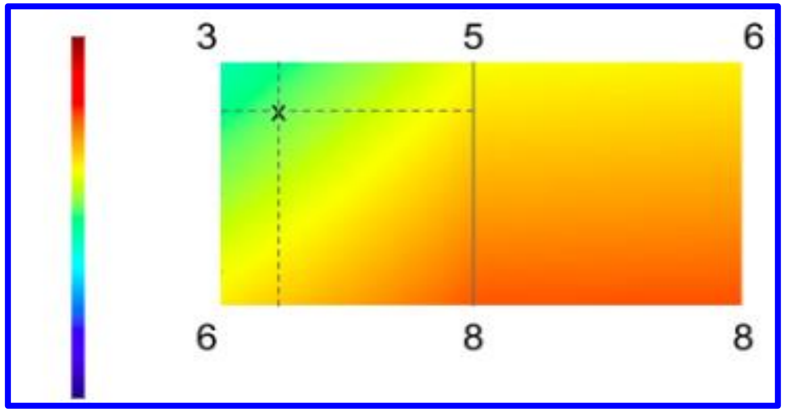
- Hyper-parameters: number of change points (Model complexity)
- Model parameters: slope of each intervals



(a)



(b)



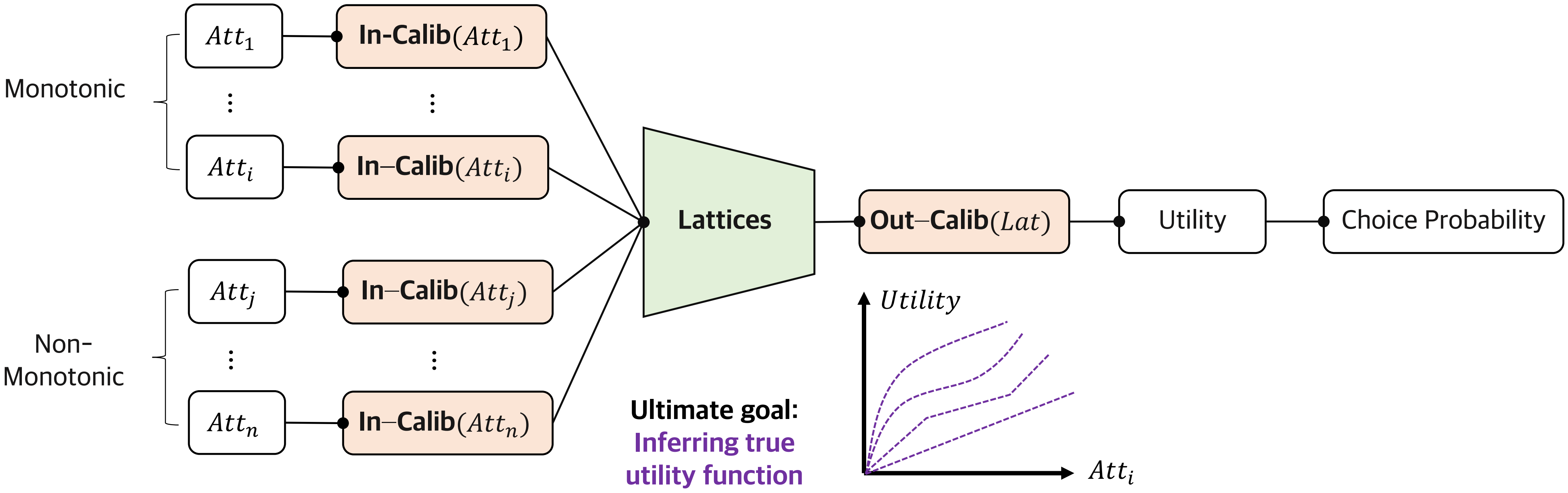
Lattice with 3×2 size

Input calibrator → **Lattices** → **Output calibrator**

Non-linearity | Interaction/non-linearities | Non-linearity

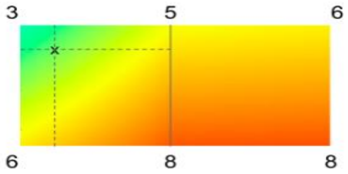
Methods

DCM-LN implemented by Lattice network



Hyper-parameters

- For each attribute
  - Number of change points
  - Convexity regularization
  - Smoothing regularization
- Lattices size



(e.g., 3x2 lattices)

- Number of change points
  - Non-linearity of output

Methods

DCM-LN Estimation

Input calibrator (*inCal*) → Lattice network (*Lat*) → Output calibrator (*outCal*) → Systematic Utility (*F*) → Choice probability (*P*)

$$F_{ij}^{LN}\left(\boldsymbol{X}_{ij};\boldsymbol{\theta}^{LN}\right) = \boldsymbol{G}_{\boldsymbol{outCal}}\left(\boldsymbol{G}_{Lat}\left(\boldsymbol{G}_{\boldsymbol{inCal}}\left(\boldsymbol{x}_{\boldsymbol{ij}};\boldsymbol{\theta}_{inCal}^{LN}\right);\boldsymbol{\theta}_{Lat}^{LN}\right);\boldsymbol{\theta}_{outCal}^{LN}\right)$$

Systematic Utility

$$P_{ij}^{LN} = e^{F_{ij}^{LN}\left(\boldsymbol{X}_{ij};\boldsymbol{\theta}^{LN}\right)} \Big/ \sum_{l=1}^{J} e^{F_{il}^{LN}\left(\boldsymbol{X}_{il};\boldsymbol{\theta}^{LN}\right)}$$

Choice probability

$$\underset{\boldsymbol{\theta}^{LN}}{\arg\min} \sum_{n=1}^{N} \mathcal{L}\left(y_i, P_{ij}^{LN}\right) + R\left(\boldsymbol{\theta}_{inCal}^{LN}\right)$$

Cross entropy +
Regularizer for calibration layer

$$s.t\ \boldsymbol{A}\boldsymbol{\theta}_{Lat}^{LN} \leq 0, \boldsymbol{B}\boldsymbol{\theta}_{inCal}^{LN} \leq 0, and\ \boldsymbol{C}\boldsymbol{\theta}_{outCal}^{LN} \leq 0$$

Monotonicity constraints

Regularization: Penalty for the changes in the second derivative of the output of input calibration layer (convexity) and the change in slopes of subsequent piece-wise linear functions (smoothing)

Bayesian optimization: Hyper-parameter tuning

Empirical risk minimization: Stochastic gradient descent with batching

Methods

Explainability: measuring attribute-wise effect (i.e., utility function)

- Partial dependence (PD) and Individual Conditional Expectation (ICE)

Remaining attributes in the set S

$$f_S = \mathbb{E}_{\boldsymbol{x}_C}\left[f(\boldsymbol{x}_S, \boldsymbol{x}_C)\right] = \int f(\boxed{\boldsymbol{x}_S}, \boldsymbol{x}_C)\mathrm{dP}(\boxed{\boldsymbol{x}_C})$$

Target attribute that varies over marginal distribution $dP(x_C)$

$$\boxed{\hat{f}_S} = \frac{1}{N}\sum_{i=1}^{N}\boxed{\hat{f}(\boldsymbol{x}_S, \boldsymbol{x}_{Ci})}$$

$\hat{f}$: the estimated models (e.g., DCM–LN, DCM–DNN, MNL)

PD ($\hat{f}_S$) **:** Changes in average utility according to attribute-level (i.e., population-level utility function)

ICE($\hat{f}(x_S, x_{Ci})$): individual-level PD (i.e., individual-level utility function)

Experiment and Result (Simulation data)

Simulation study

- True utility function is required to evaluate both interpretability and predictability
- Simulation choice data are generated to evaluate both utility function inference and choice prediction.
  - Alternative attributes: travel cost (TC), travel time (TT), waiting time (WT), and crowding (CR).
  - Individual attributes: income (INC), full-time (FUL), flexibility (FLX).

Non-linear effect of travel cost

Interaction between alternative attributes

$$V_{ij} = -0.1 - 8 \cdot \boxed{\sqrt{TC_{ij}}} - 2.0 \times CR_{ij} + \begin{pmatrix} -0.1 - 0.5 \times IN_i - 0.1 \times FUL_i + 0.05 \times FLX_i \\ -0.2 \times IN_i \times FUL_i + 0.05 \times IN_i \times FLX_i + 0.1 \times FUL_i \times FLX_i - 0.02 \times CR_{ij} \end{pmatrix} \times TT_{ij} + \boxed{\begin{pmatrix} -0.2 - 0.8 \times IN_i - 0.3 \times FUL_i + 0.1 \times FLX_i \\ -0.3 \times IN_i \times FUL_i + 0.08 \times IN_i \times FLX_i + 0.3 \times FUL_i \times FLX_i \end{pmatrix} \times WT_{ij}}$$

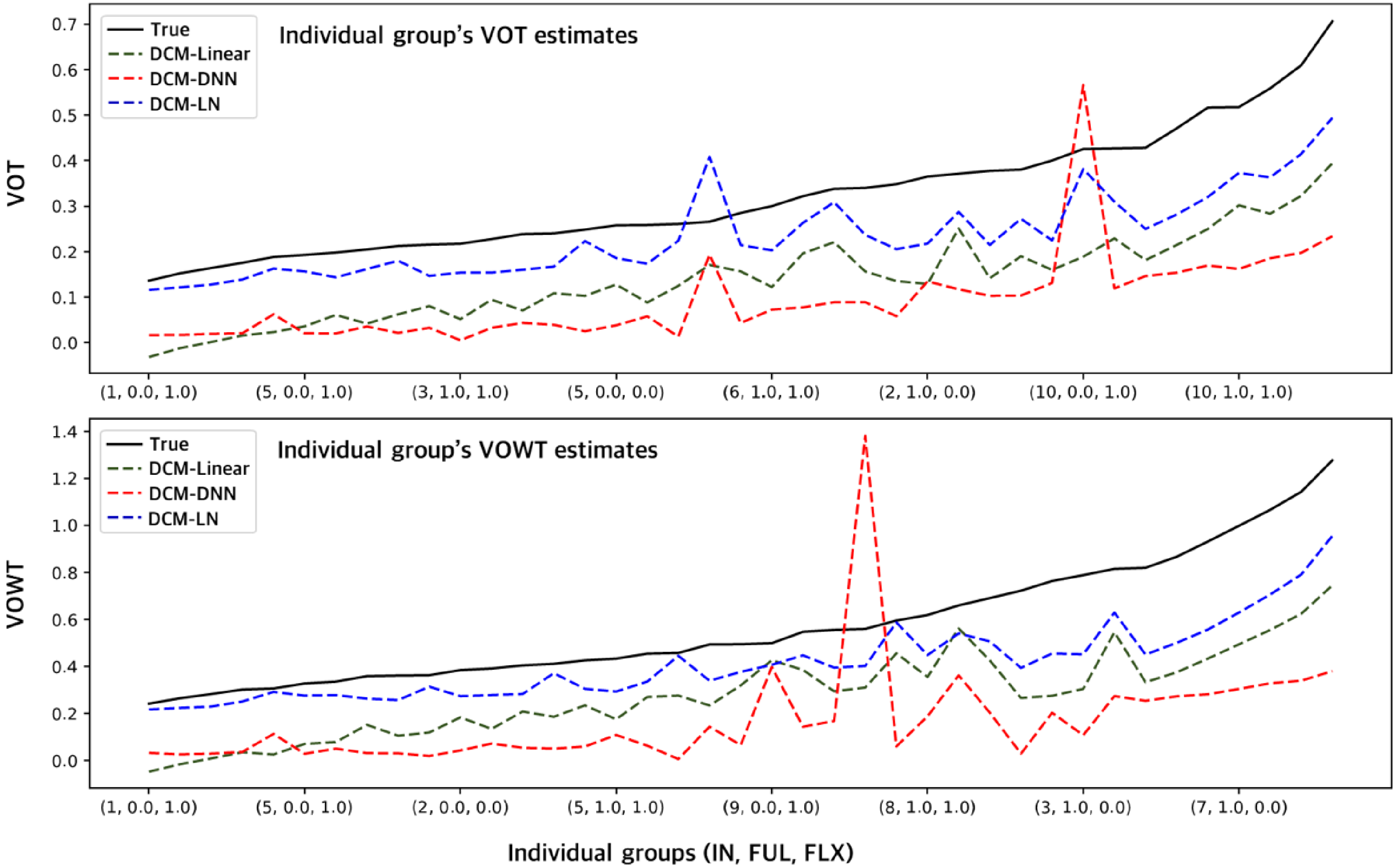
Interactions between individual and alternative attributes (i.e., individual taste heterogeneity)

Experiment and Result (Simulation data)

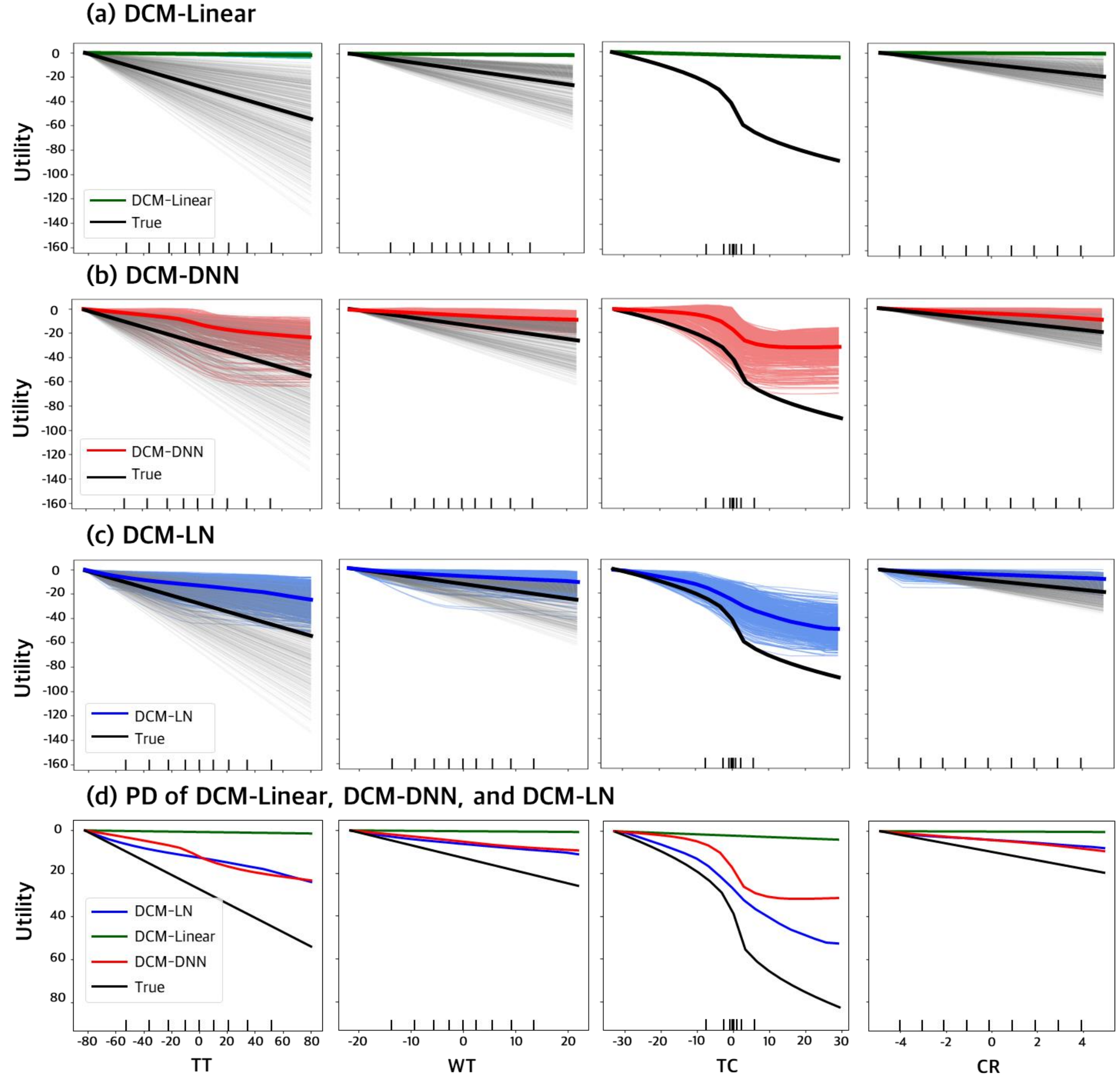
Parameter		True (50 trials)		MNL (50 trials)		DCM-DNN (50 trials)		DCM-LN (50 trials)	
		Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Interpretability									
Recovery of distribution	VOT (Median)	0.284	0.014	0.126	0.019	0.075	0.105	**0.188**	0.080
	VOT (1%)	0.142	0.010	-0.026	0.029	-0.012	0.281	**0.093**	0.063
	VOT (25%)	0.216	0.013	0.066	0.021	0.040	0.085	**0.135**	0.072
	VOWT (Median)	0.480	0.019	0.258	0.148	0.146	0.210	**0.322**	0.134
	VOWT (1%)	0.252	0.011	-0.068	0.124	-0.114	0.797	**0.153**	0.082
	VOWT (25%)	0.372	0.017	0.118	0.141	0.086	0.159	**0.244**	0.127
Recovery of individual groups' value	VOT (RMSE)			0.193	0.012	0.272	0.102	**0.129**	0.030
	VOWT (RMSE)			0.348	0.092	0.546	0.259	**0.243**	0.063
Predictability	Training accuracy			0.552	0.006	**0.775**	0.010	0.741	0.018
	Test accuracy			0.546	0.013	**0.716**	0.014	0.697	0.016

VOT: Value of Travel Time; **VOWT**: Value of Wait Time

Experiment and Result (Simulation data)



Experiment and Result (Simulation data)



Conclusions

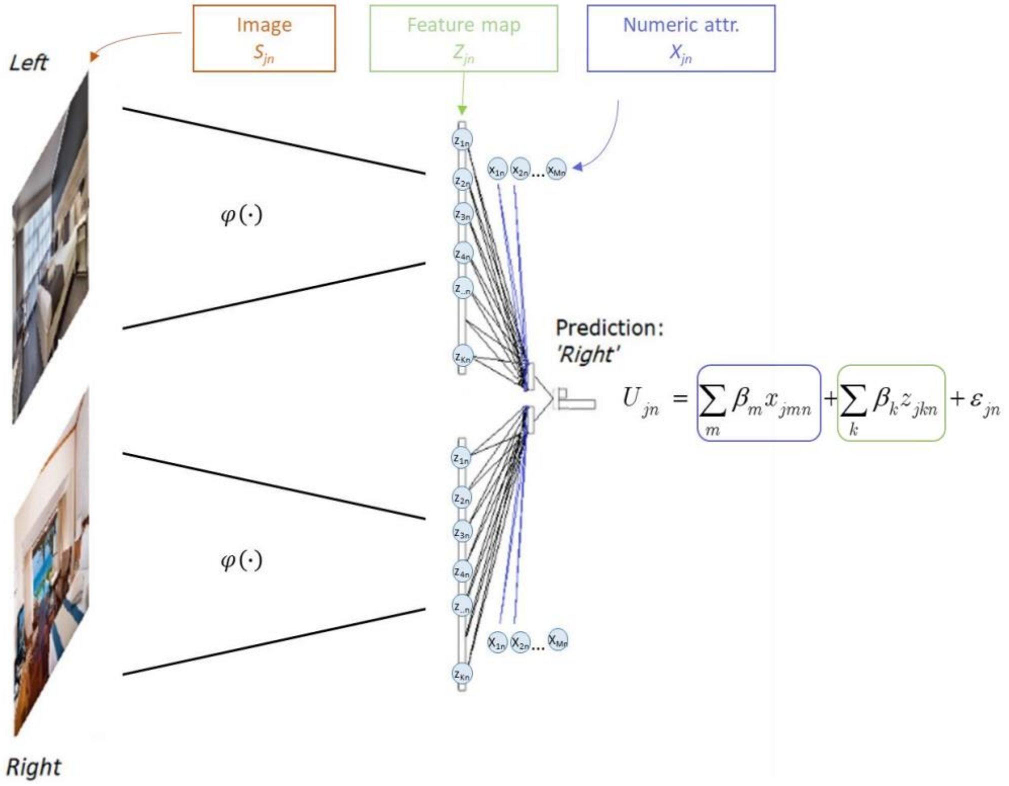
- The DCM-LN ensures interpretability.
  - DCM-LN infers underlying utility function better than theory-driven DCM (MNL).
  - Non-linearity and interactions are captured even with monotonic constraints.
  - Trade-off between interpretability and predictability is demonstrated.
  - **Monotonicity** significantly enhances the interpretability (trustworthiness).

- Lattice network can be used to model inflextion points in prospect theory and semi-compensatory choice models

Working Paper: Kim, E. J., & Bansal, P. (2023). A New Flexible and Partially Monotonic Discrete Choice Model. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4448172

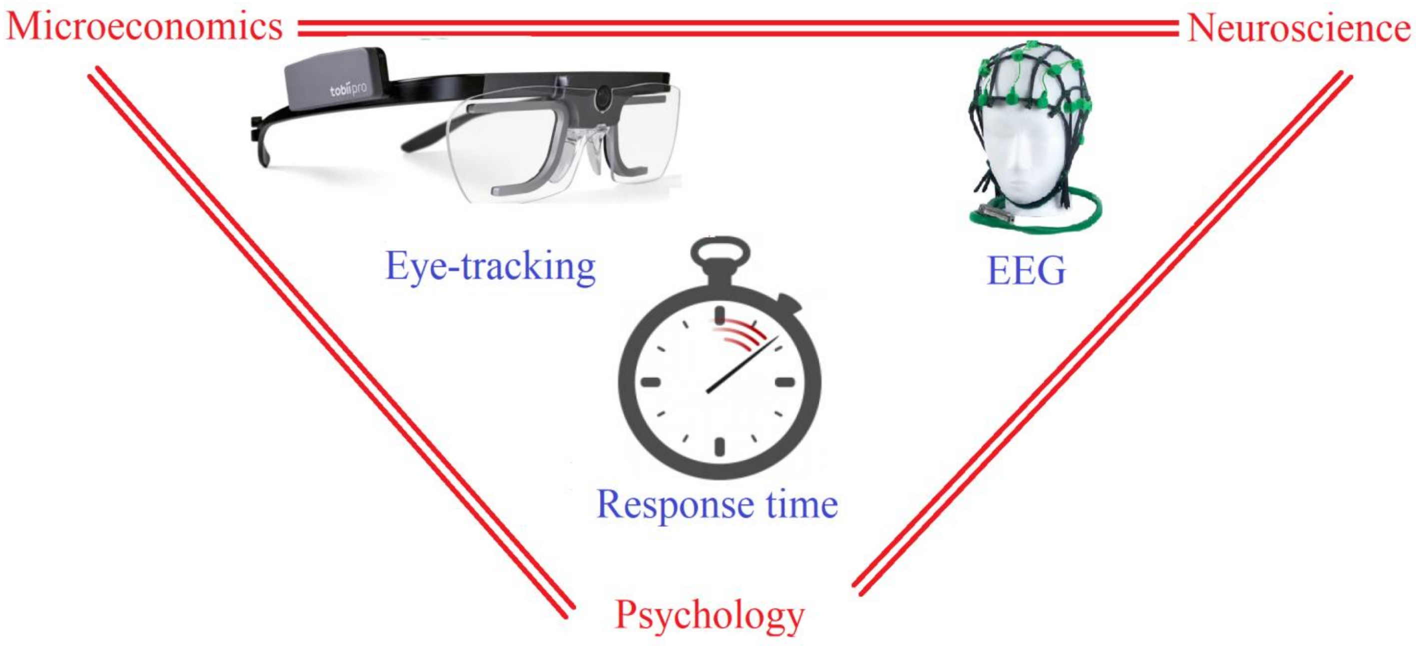
Future Work

- Lattice network can be used to model inflextion points in <u>prospect theory</u> and <u>semi-compensatory</u> choice models
- Incorporating computer-vision-based Choice Models to use image data.



van Cranenburgh & Garrido-Valenzuela (2023)

van Cranenburgh, S., & Garrido-Valenzuela, F. (2023). Computer vision-enriched discrete choice models, with an application to residential location choice. *arXiv preprint arXiv:2308.08276*.



Part 2: Behaviour Models, Psychology, and Process Data

Motivation

- Utility-based choice models are static.
- Difficult include process data and account for information acquisition process.
- Cannot handle decoy effect, i.e., violates regularity conditions

Adding a less attractive alternative (attraction decoy) can increase preference towards existing target alternative.

Which one do you prefer ?

S$ 5 S$ 9

Competitor (C) Target (T)

Which one do you prefer **now** ?

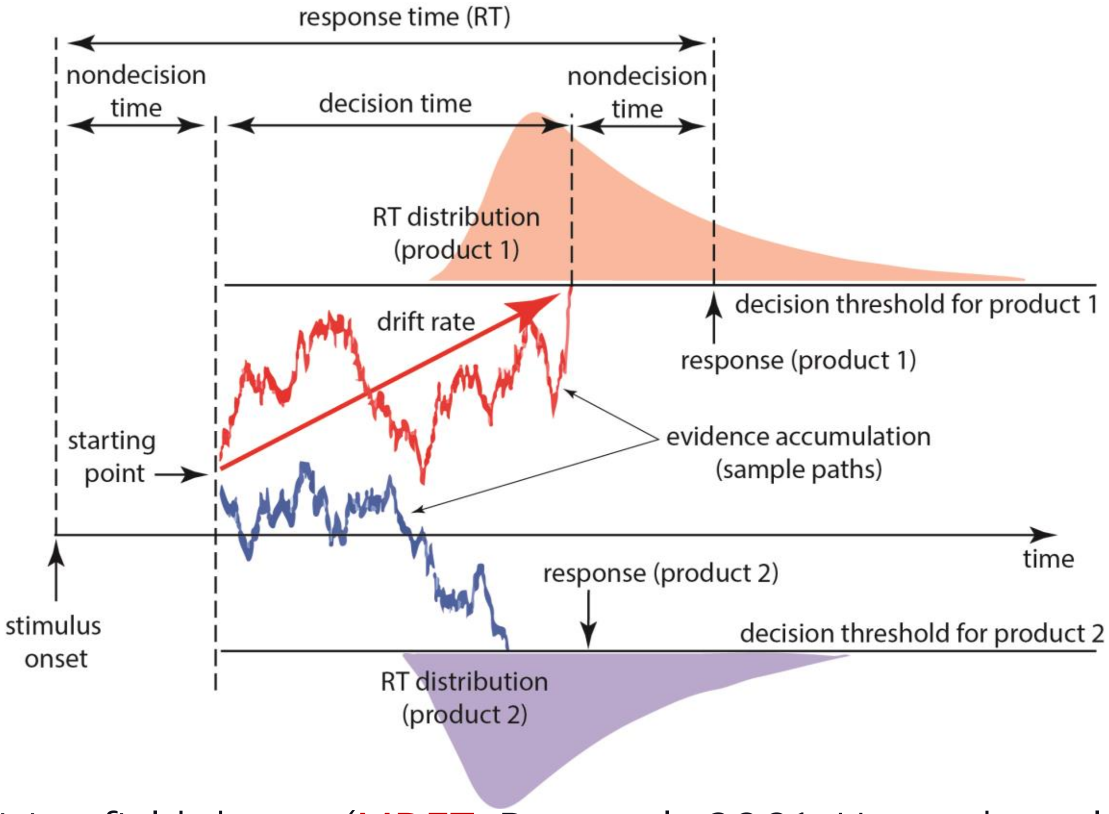
S$ 5 **S$ 8.5** S$ 9

Competitor (C) Decoy (D) Target (T)

Sequential Sampling Models (SSMs)

Parameters

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



Advantages

- Can handle process data

(e.g., eye-tracking & response time)

- Cognitive underpinning
- Can explain decoy effect better

Popular Models

- Multi-alternative decision field theory (MDFT, Roe et al., 2001; Hancock et al., 2021)
- Multiattribute linear ballistic accumulator model (MLBA, Trueblood et al., 2014)
- Multi-alternative decision by sampling (MdBS, Noguchi & Stewart, 2018)

Hancock, T. O., Hess, S., Marley, A. A., & Choudhury, C. F. (2021). An accumulation of preference: two alternative dynamic models for understanding transport choices. *Transportation Research Part B: Methodological.*
Noguchi, T., & Stewart, N. (2018). Multialternative decision by sampling: A model of decision making constrained by process data. *Psychological review, 125*(4), 512.
Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological review, 108*(2), 370.
Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological review, 121*(2), 179.

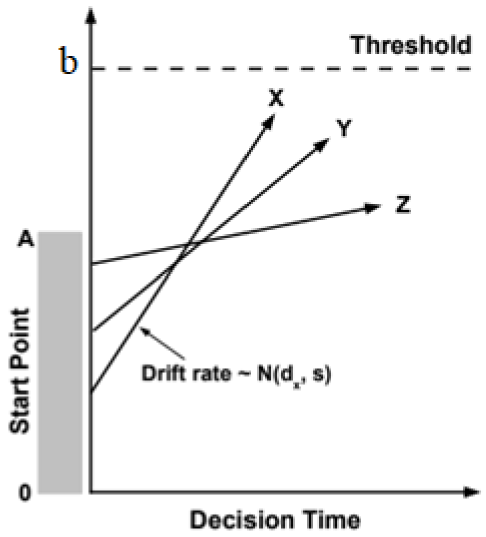
Model 1: Multiattribute linear ballistic accumulator (MLBA)

Closed-form probability expression of joint choice and response time

The joint probability of choice i and response time $t+\tau_0$

$$MLBA\_CRT(RC=i, RT=t+\tau_0) = f_i(t)\Pi_{j\neq i}(1-F_j(t))$$

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



$$f_i(t) = \frac{1}{A}[-d_i\Phi(\frac{b-A-td_i}{ts}) + s\phi(\frac{b-A-td_i}{ts}) + v_i\Phi(\frac{b-td_i}{ts}) - s\phi(\frac{b-td_i}{ts})]$$

$$d_i = max\{I_0 + \zeta_i + \sum_{j\in\mathscr{C}, j\neq i}\sum_{k=1}^{K}\omega_{ijk}\beta_k(X_{ik}-X_{jk}), 0\}$$

Where

$$\omega_{ijk} = \begin{cases} exp\{-\lambda_1|\beta_k(X_{ik}-X_{jk})|\} & \beta_k(X_{ik}-X_{jk}) \geq 0 \\ exp\{-\lambda_2|\beta_k(X_{ik}-X_{jk})|\} & \beta_k(X_{ik}-X_{jk}) < 0 \end{cases}$$

Where

$v_i \sim Normal(d_i, s^2)$: drift rate is

b: decision threshold

τ_0: non-decision time is

A: Start point upper bound

I_0: Drift rate mean constant

ζ_i: Alternative specific constant

Hancock, T. O., Hess, S., Marley, A. A., & Choudhury, C. F. (2021). An accumulation of preference: two alternative dynamic models for understanding transport choices. *Transportation Research Part B: Methodological*, *149*, 250-282.

Model 2: Multi-attribute decision by sampling (MDbS)

The accumulation process follows the random walk. Assumes pairwise comparison of alternatives on an attribute.

The probability of gaining 1 unit evidence in favour of alternative i in a time step by respondent n is:

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

$p(\text{evaluate alternative } i \text{ on attribute } k)$

$$w_{n,i,k} = \frac{\sum_{j=1}^{J_n} RS_{n,i,j,k}}{\sum_{i=1}^{J_n} \sum_{j=1}^{J_n} \sum_{k=1}^{Q_n} RS_{n,i,j,k}}$$

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

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

α: the larger α leads to stronger attraction effect

$p(\text{alternative } i \text{ win a comparision on attribute } k)$

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

$$p(\text{alternative } i \text{ is favored over alternative j })$$

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

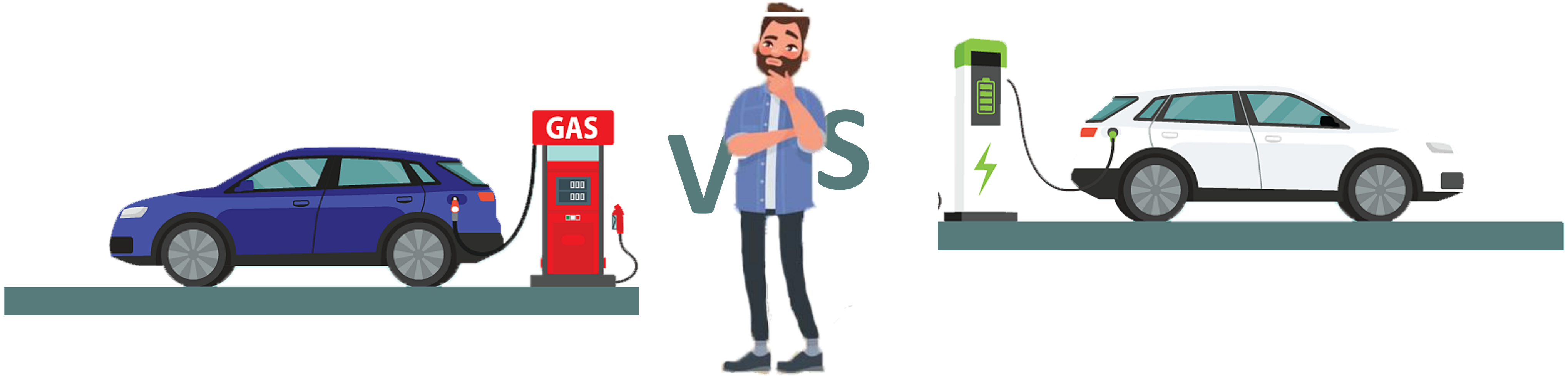
β_0: the minimal relative difference that can be identified
β_1: the maximum identifiable difference

SSM: Challenges

1. Decoy effect experiment have been conducted in lab-based settings
2. The value of response time is unclear
3. Model selection from behavioural perspective
4. Sensitivity to priors
5. Small sample size for lab-based studies studies

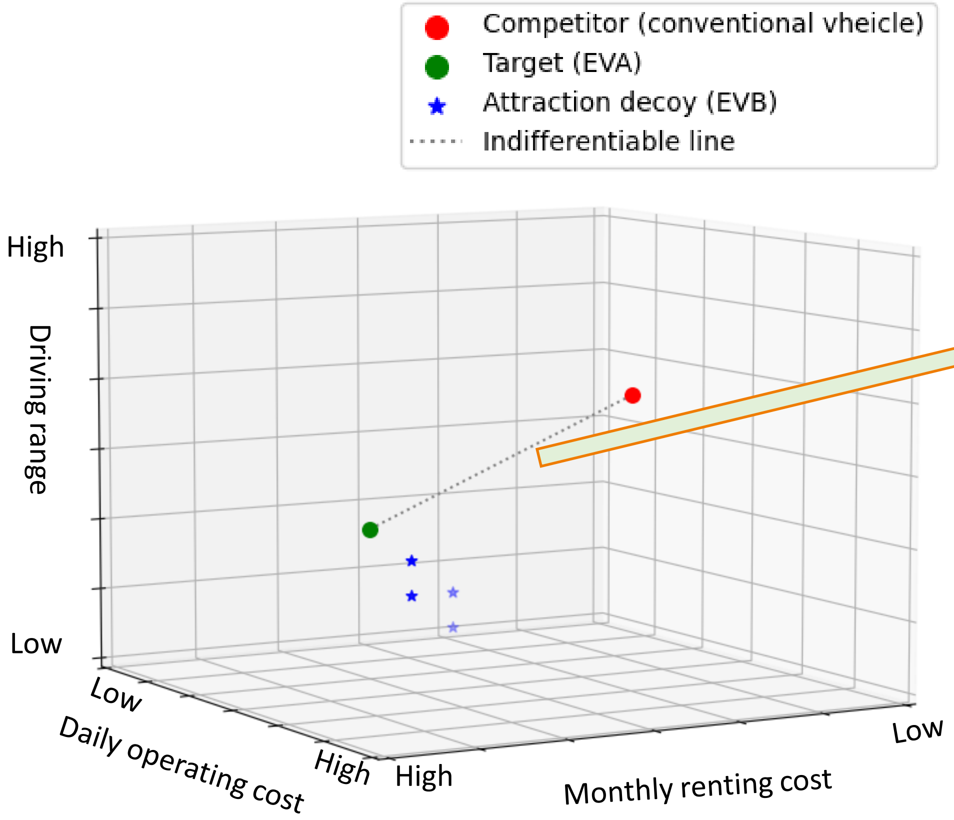
Eye-tracking + Online data

Challenge 1: Real-world Experiment Design



Experiment Design: Indifferentiable Line

- Alternatives on the indifferentiable line are **equally attractive** to respondents
- **Baseline** for decoy experiments design: mitigating the strong dislike or like toward the one alternative



		
	Your conventional vehicle	**Electric vehicle**
Monthly renting cost	S$ 2100	S$ 2750
Daily operating cost	S$ 45	S$ 23
Driving range with full fuel tank	750 km (refuel 10 times per month)	450 km (recharge 16.5 times per month)

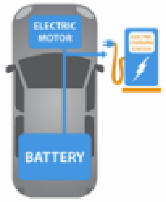
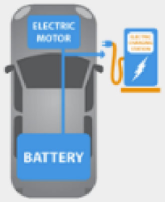
Which vehicle would you rent in the above situation?

- I will continue renting my **conventional vehicle.**
- I will switch to renting an **electric vehicle.**
- Electric vehicle and my conventional vehicle are **equally attractive** to me.

1. 1.3 times of the renting cost of convention car
2. 1.2 times
3. 1.1 times
4. Same
5. Lower than conventional car {0.9 times, 0.8 times, 0.7 times, 0.6 times or 0.5 times}

Experiment Design: Attraction Decoy

Attributes of conventional car & EV(model A):
Based on indifferentiable line experiment

			
	Your conventional vehicle	**Electric vehicle (Model A)**	**Electric vehicle (Model B)**
Monthly renting cost	S$ 2100	S$ 2750	S$ 2730
Daily operating cost	S$ 43	S$ 23	S$ 23
Driving range with full fuel tank	750 km (refuel 14 times of times per month)	450 km (recharge 23.5 times per month)	400 km (recharge 26.5 times per month)

Which vehicle would you rent in the above situation?

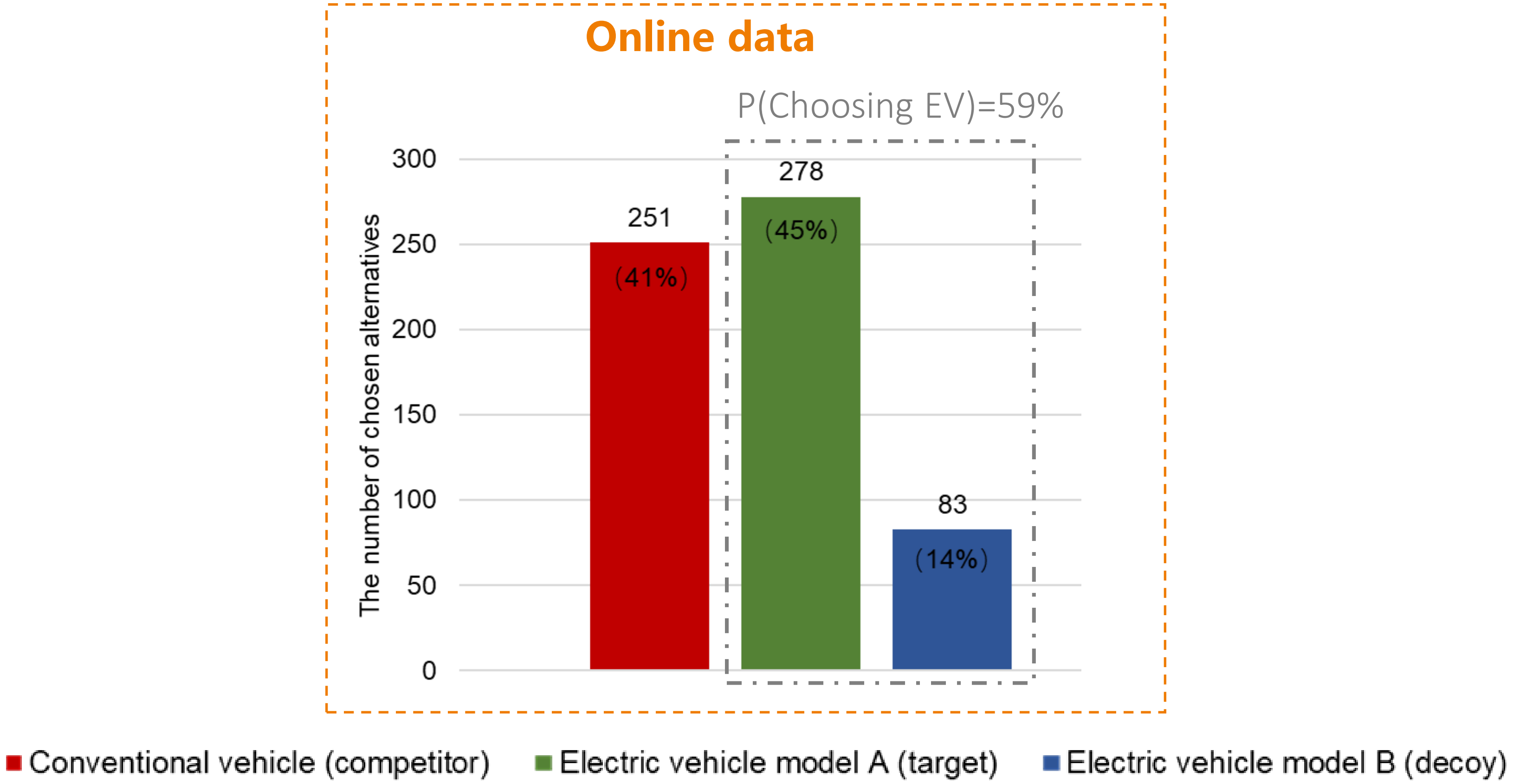
- ◯ Your conventional vehicle
- ◯ Electric vehicle (Model B)
- ◯ Electric vehicle (Model A)

Monthly renting cost of EV model B:
1. S$50 lower than EV model A
2. S$20 lower

Daily operating cost of EV model B:
Same operating cost as of EV model A

Driving range of EV model B:
1. 100 km lower than EV model A
2. 150 km lower than EV model A
3. 50 km lower than EV model A

Does Attraction Effect Exist in EV Rental Market?



Main Results

	MNL	MDFT	Original MDbS	Revised MDbS
In-sample estimation				
BIC	799.96	787.24	891.99	923.75
PRST -RST	0.01	0.01	-0.02	-0.06
Out-of-sample prediction				
BIC	286.31	371.39	327.28	337.05
PRST -RST	-0.09	-0.05	0.07	0

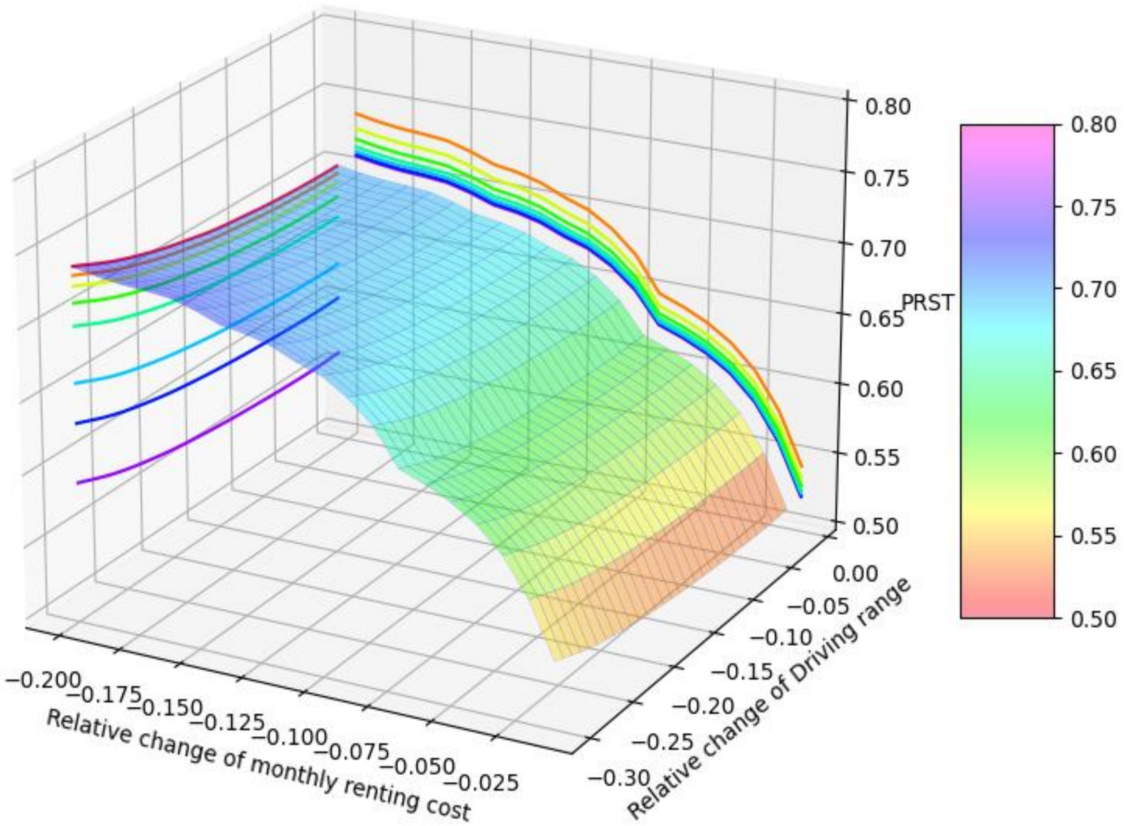
PRST: predicted relative choice share of the target: $\frac{P(T|T,C,D)}{P(T|T,C,D)+P(C|T,C,D)}$

Lower PRST-RST, the model is better in capturing substitution effect.

T: Target; C: Competitior; D: Decoy

Optimal deployment of attraction decoy

The optimal range of attribute levels of the decoy models is **10%–18% lower** in **monthly renting cost** and **17%–25% lower in driving range**.



- ✓ **(x-axis and y-axis) Relative change:** Proportional difference compared to the target EV.
- ✓ **(z-axis) Predicted relative choice share of the target (PRST):** higher PRST indicate stronger attraction effects: PRST=$\frac{P(T|T,C,D)}{P(T|T,C,D)+P(C|T,C,D)}$

Challenge 2: Value of Response Time

(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:

$$P\_CO_\theta(RC = i) = \int_0^\infty P\_CRT_\theta(RC = i, RT = t)dt, \quad i \in \mathscr{C}$$

Asymptotic Results: MLBA (Choice and Response Time)

$$\sqrt{n}(\tilde{\theta}_{CRT} - \theta_0) \xrightarrow{d} N(0, I_{CRT}(\theta_0)^{-1})$$

$$I_{CRT}(\theta) = -\sum_{i \in \mathscr{C}} \int_0^{\infty} \frac{\partial^2 \log MLBA\_CRT(RC = i, RT = t)}{\partial \theta \theta^T} MLBA\_CRT(RC = i, RT = t) dt$$

$$\sqrt{n}(\tilde{\theta}_{RTG} - \theta_0) \xrightarrow{d} N(0, I_{RTG}(\theta_0)^{-1})$$

$$I_{RTG}(\theta) = -\sum_{i \in \mathscr{C}} \int_0^{\infty} \frac{\partial^2 \log MLBA\_RTG(RC = i | RT = t)}{\partial \theta \theta^T} MLBA\_CRT(RC = iRT = t)] dt$$

$$\sqrt{n}(\tilde{\theta}_{CO} - \theta_0) \xrightarrow{d} N(0, I_{CO}(\theta_0)^{-1})$$

$$I_{CO}(\theta) = -\sum_{i \in \mathscr{C}} \int_0^{\infty} \frac{\partial^2 \log MLBA\_CO(RC = i)}{\partial \theta \theta^T} MLBA\_CRT(RC = i, RT = t) dt$$

Key Result: Lowest Asymptotic Variance of CRT

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

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

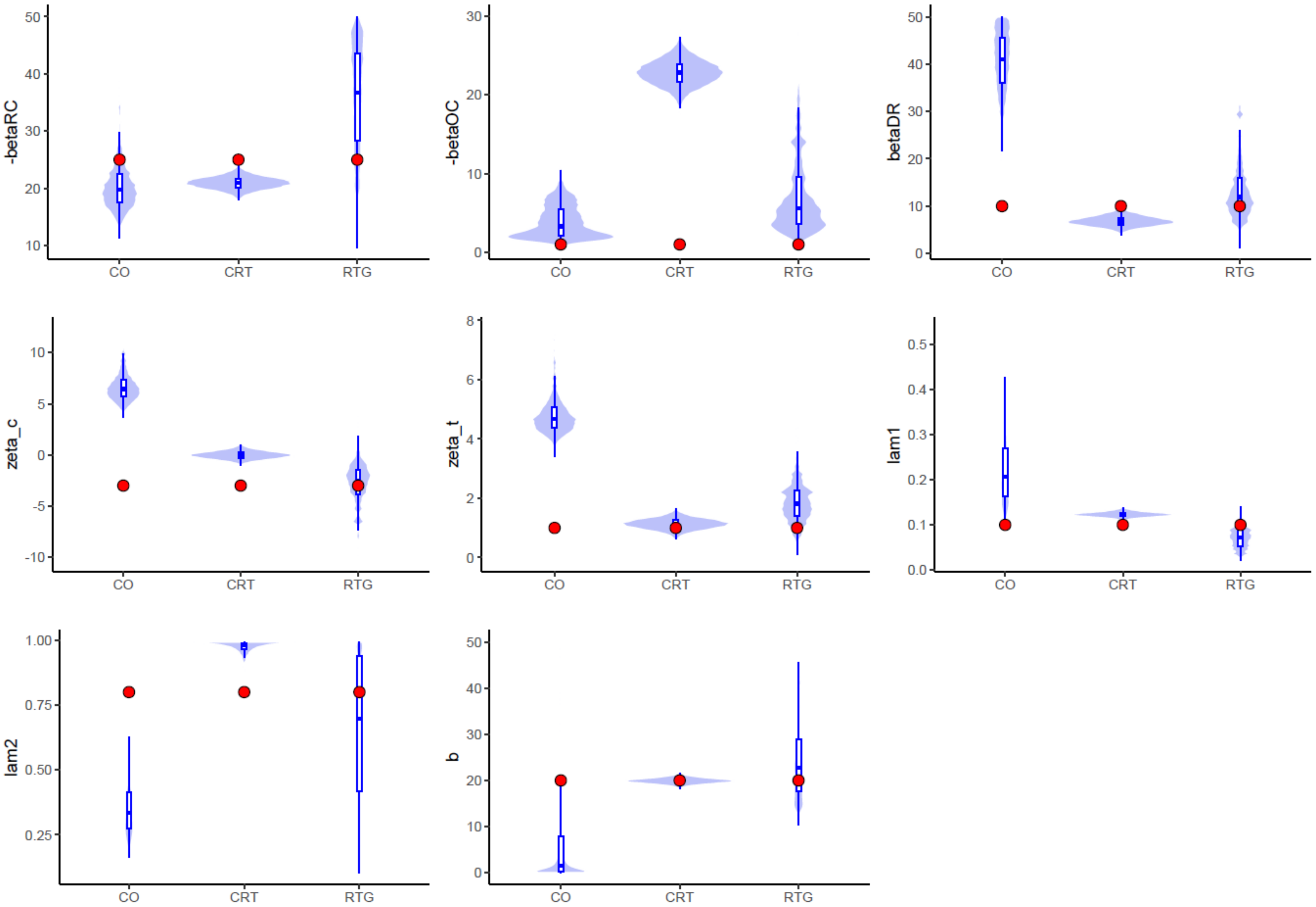
Hence, the Fisher Information matrix of three types of distribution follows:

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

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

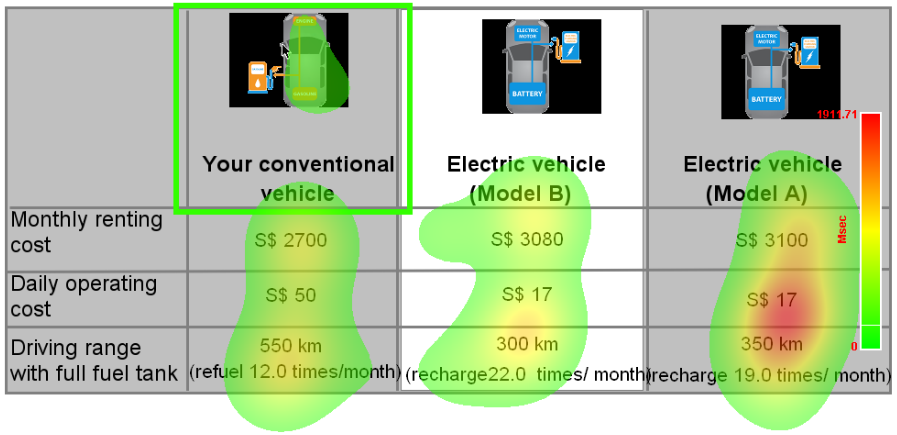
$$I_{CRT}(\theta_0) \geq I_{CO}(\theta_0)$$
$$I_{CRT}(\theta_0) \geq I_{RTG}(\theta_0)$$

Simulation: Validation of Asymptotic Result (MLBA)



Challenge 3: Model Selection from Behavioral Perspective

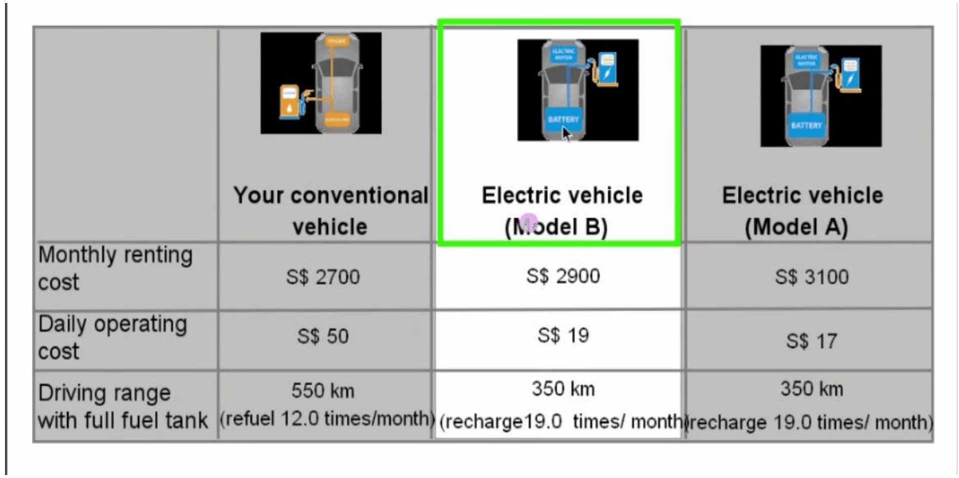
Fixation duration/count: attribute non-attendance



	Your conventional vehicle	Electric vehicle (Model B)	Electric vehicle (Model A)
Monthly renting cost	S$ 2700	S$ 3080	S$ 3100
Daily operating cost	S$ 50	S$ 17	S$ 17
Driving range with full fuel tank	550 km (refuel 12.0 times/month)	300 km (recharge22.0 times/ month)	350 km (recharge 19.0 times/ month)

Model selection

Eye-tracking Trajectory



	Your conventional vehicle	Electric vehicle (Model B)	Electric vehicle (Model A)
Monthly renting cost	S$ 2700	S$ 2900	S$ 3100
Daily operating cost	S$ 50	S$ 19	S$ 17
Driving range with full fuel tank	550 km (refuel 12.0 times/month)	350 km (recharge19.0 times/ month)	350 km (recharge 19.0 times/ month)

Challenge 4 & 5: Prior Sensitivity and Small Sample Size
(Fusing Lab & Online Data)

Model structure	Input & output	Dataset

Traditional MLBA:

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

$$\omega_{ijk} = \begin{cases} exp\{-\lambda_1 |\beta_k (X_{ik} - X_{jk})|\} & \beta_k (X_{ik} - X_{jk}) \geq 0 \\ exp\{-\lambda_2 |\beta_k (X_{ik} - X_{jk})|\} & \beta_k (X_{ik} - X_{jk}) < 0 \end{cases}$$

Non-informative priors $\xrightarrow{\theta \sim \pi_{lab}}$ Traditional MLBA

Traditional MLBA $\xrightarrow{\hat{\theta} \sim \hat{\psi}_{lab}}$ Posteriors

Choice + response time (lab) + Demographics (lab) → Traditional MLBA

Posteriors $\xrightarrow{\pi_{online} = \psi_{lab}}$ Informative priors

Traditional MLBA:

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

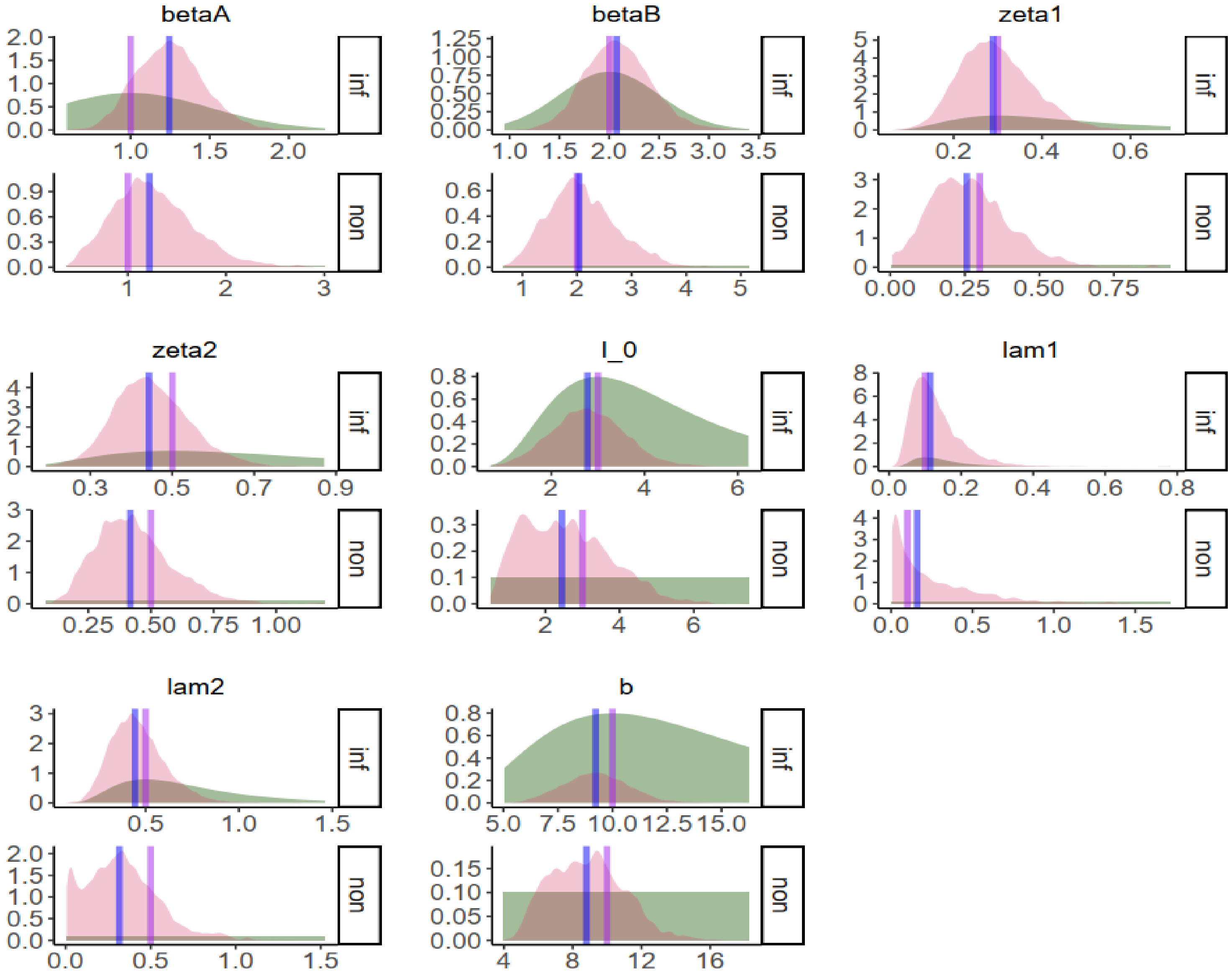
$$\omega_{ijk} = \begin{cases} exp\{-\lambda_1 |\beta_k (X_{ik} - X_{jk})|\} & \beta_k (X_{ik} - X_{jk}) \geq 0 \\ exp\{-\lambda_2 |\beta_k (X_{ik} - X_{jk})|\} & \beta_k (X_{ik} - X_{jk}) < 0 \end{cases}$$

Informative priors $\xrightarrow{\theta \sim \pi_{online}}$ Traditional MLBA

Traditional MLBA $\xrightarrow{\hat{\theta} \sim \hat{\psi}_{online}}$ **Posteriors**

Choice + response time (online) + Demographics (online) → Traditional MLBA

Fusing Lab & Online Data: Simulation Results



Posterior simulation result is sensitive to the prior.

A good prior leads a less biased, smaller variance posterior.

Pink area for posterior density;

Green area for prior density;

Blue line for posterior median;

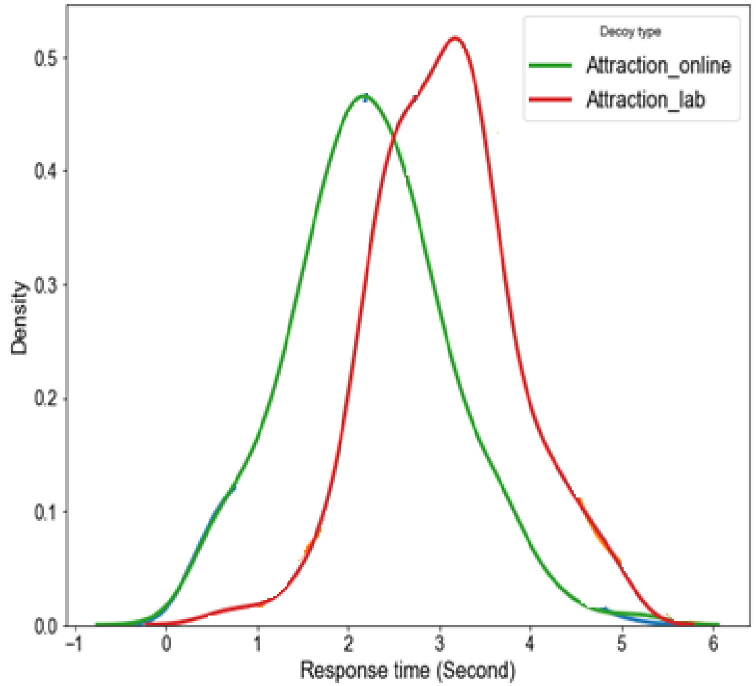
Purple line for true value;

Empirical Application: Similarity in Lab & Online Data

	The savings in operating costs due to renting electric vehicle are insufficient	
Lab data	Not a big concern	A big concern
Driving long distance	12 (80%)	3 (20%)
Driving short distance	13 (52%)	12 (48%)
Online data		
Driving long distance	66 (74%)	23 (26%)
Driving short distance	129 (60%)	87 (40%)

	The chargers in my neighbourhood areas are insufficient	
Lab data	Not a big concern	A big concern
working long hours	5 (22%)	18 (78%)
working short hours	8 (47%)	9 (53%)
Online data		
working long hours	64 (33%)	128 (67%)
working short hours	64 (57%)	49 (43%)

	The maintenance costs of electric vehicles are high	
Lab data	Not a big concern	A big concern
Old	13 (62%)	8 (38%)
Young	11 (58%)	8 (42%)
Online data		
Old	145 (55%)	118 (45%)
Young	24 (57%)	18 (43%)

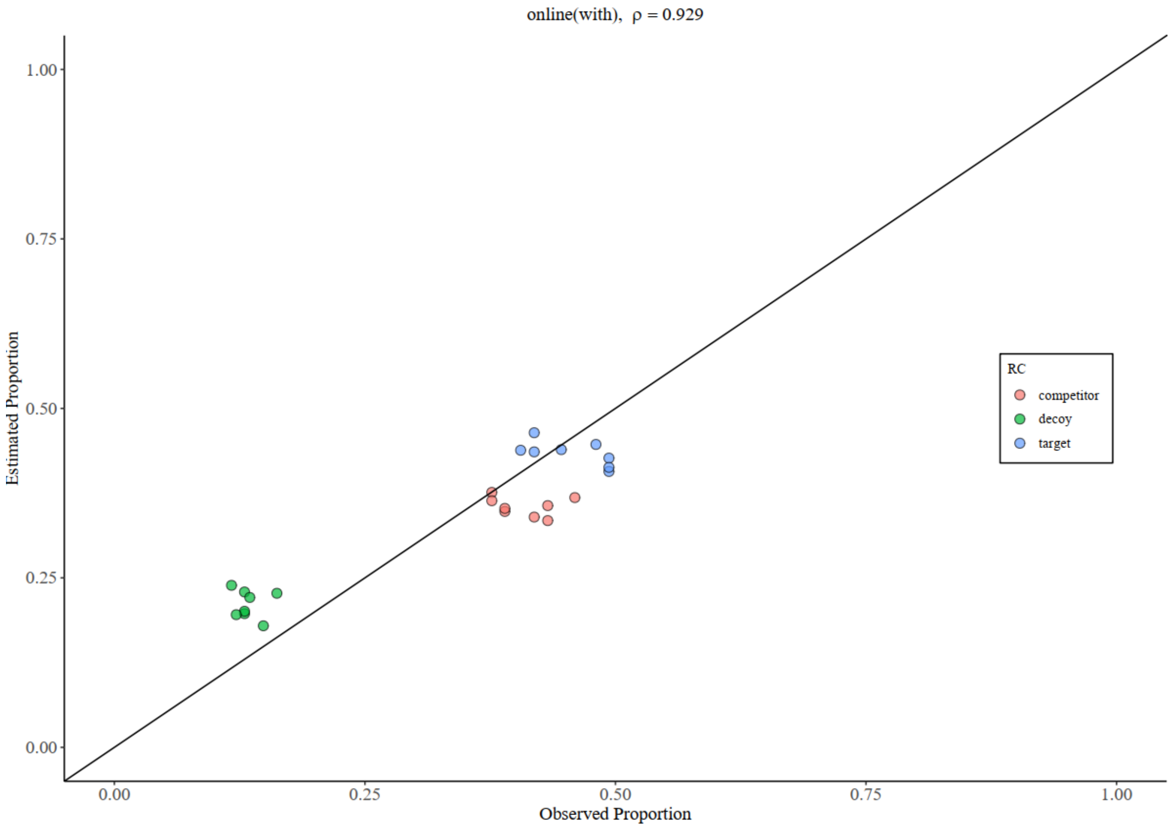


Fusing Lab & Online Data: Model Fit (Empirical Study)

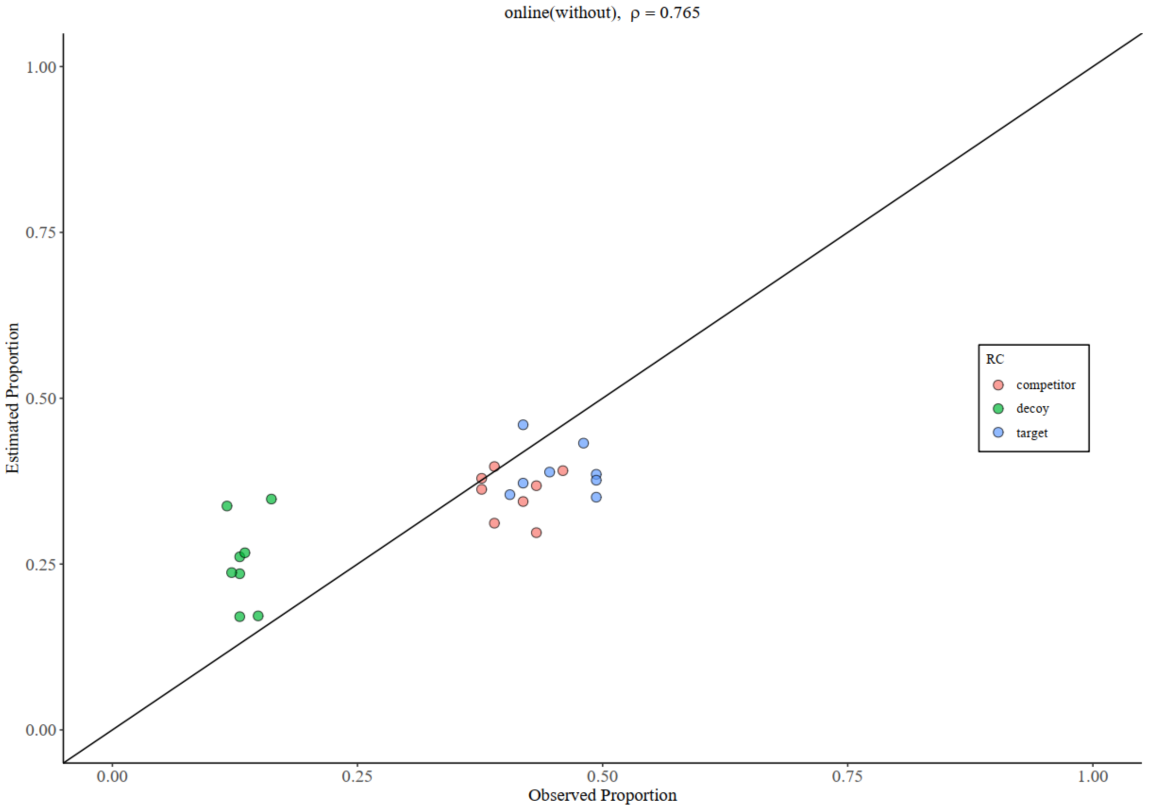
Better fitting performance with data fusion method

scenario-level predicted choice proportion-to-portion plot

WITH data fusion method



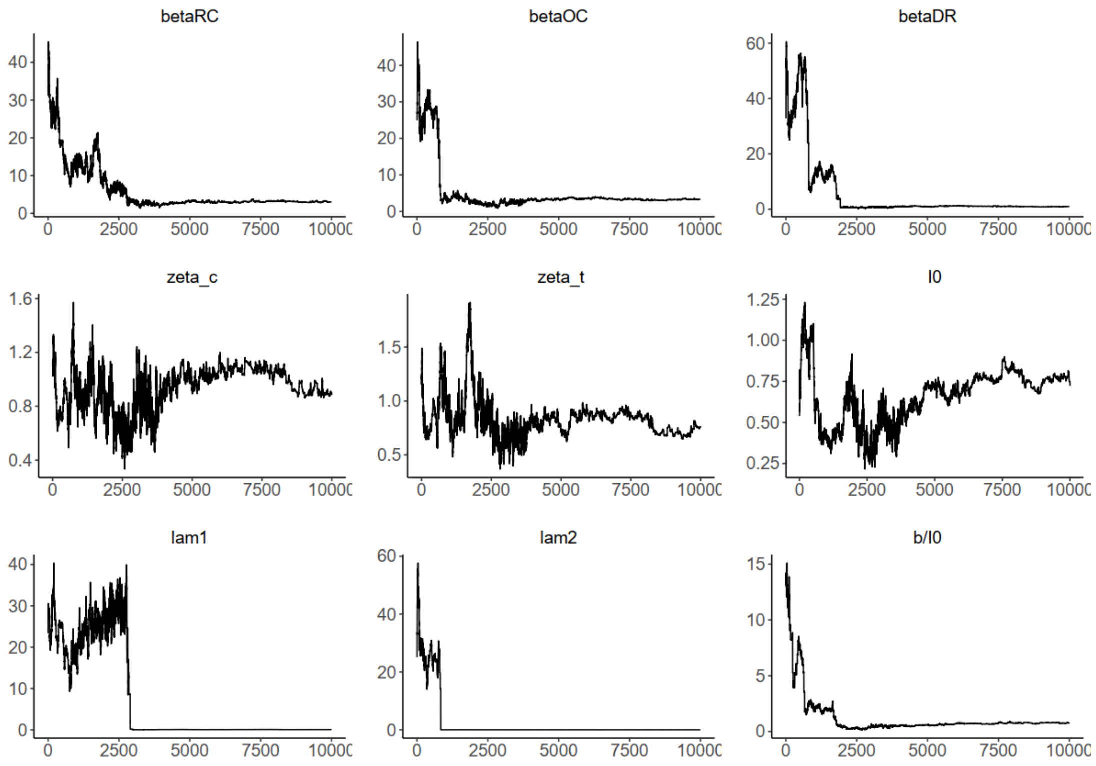
WITHOUT data fusion method



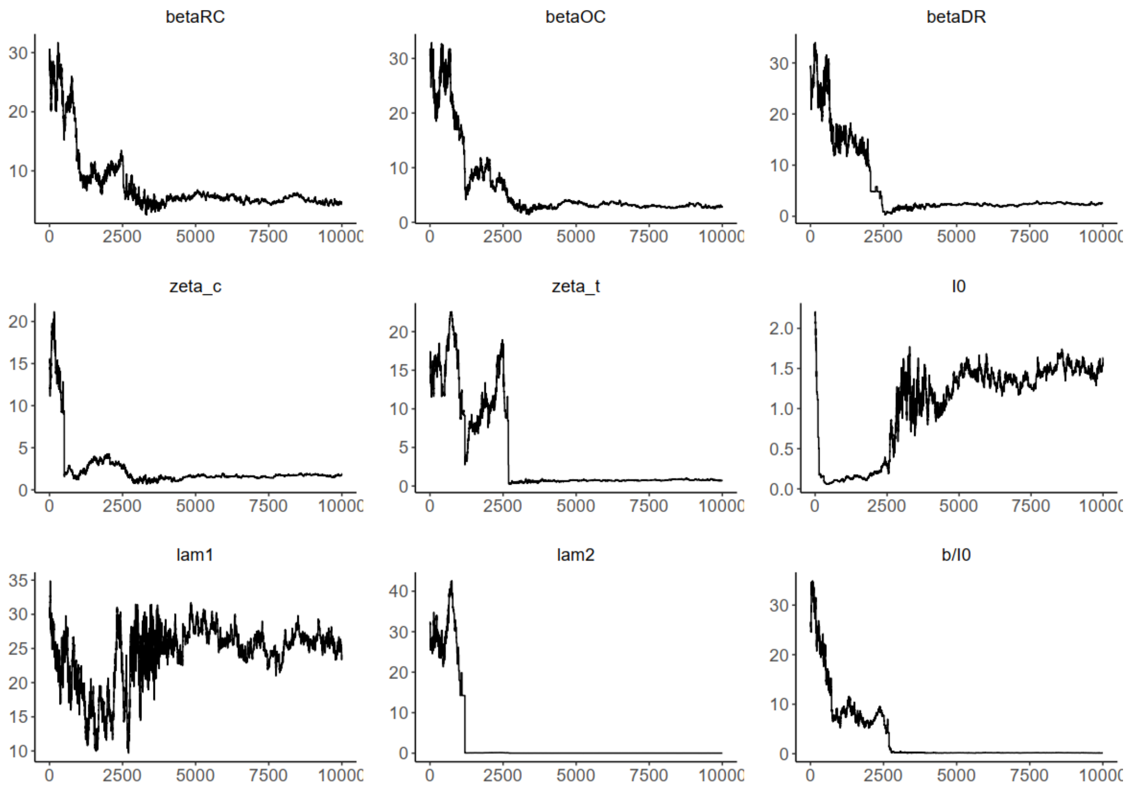
Fusing Lab & Online Data: Convergence (Empirical Study)

Faster convergence with data fusion method

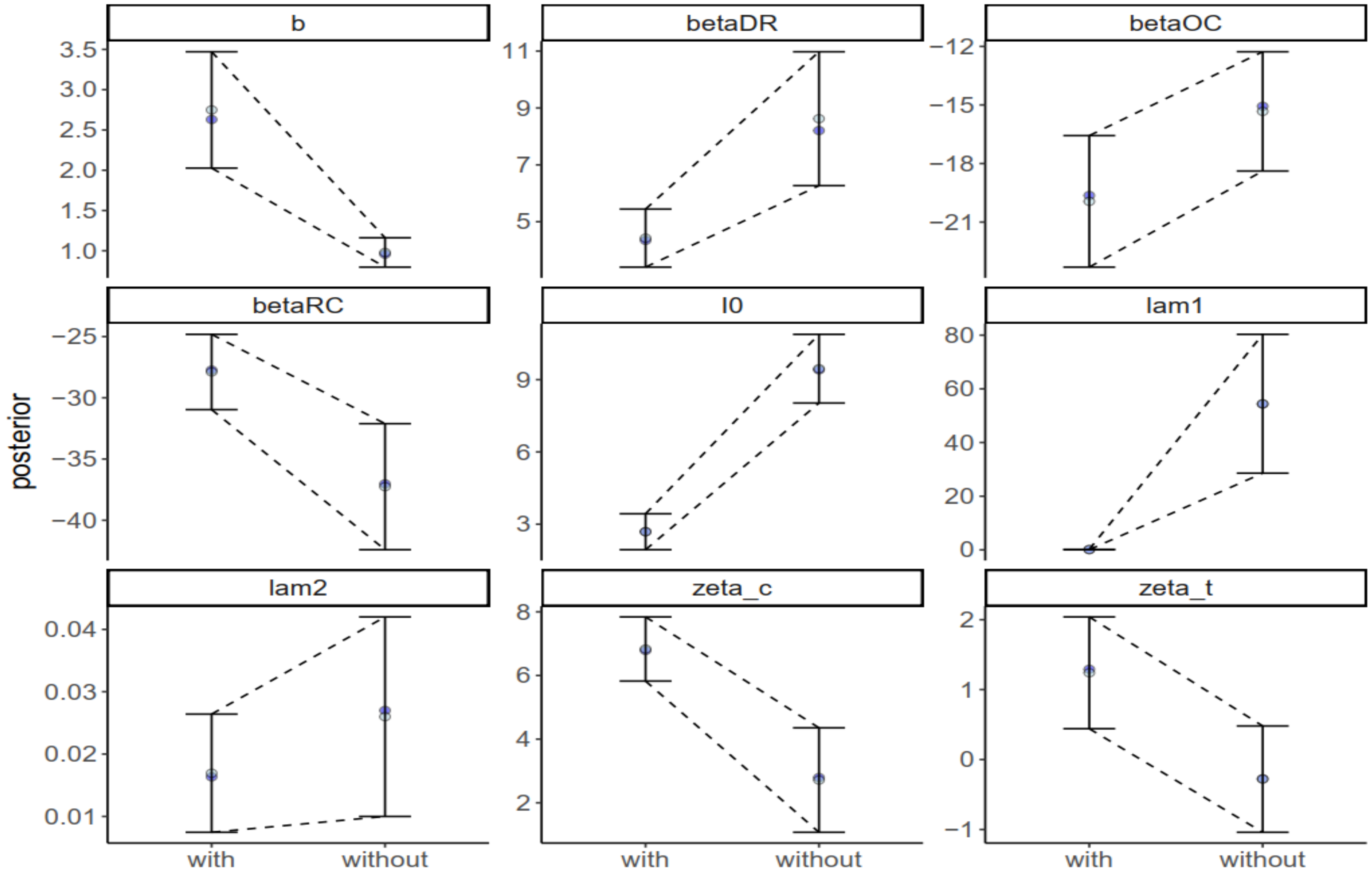
WITH data fusion method



WITHOUT data fusion method



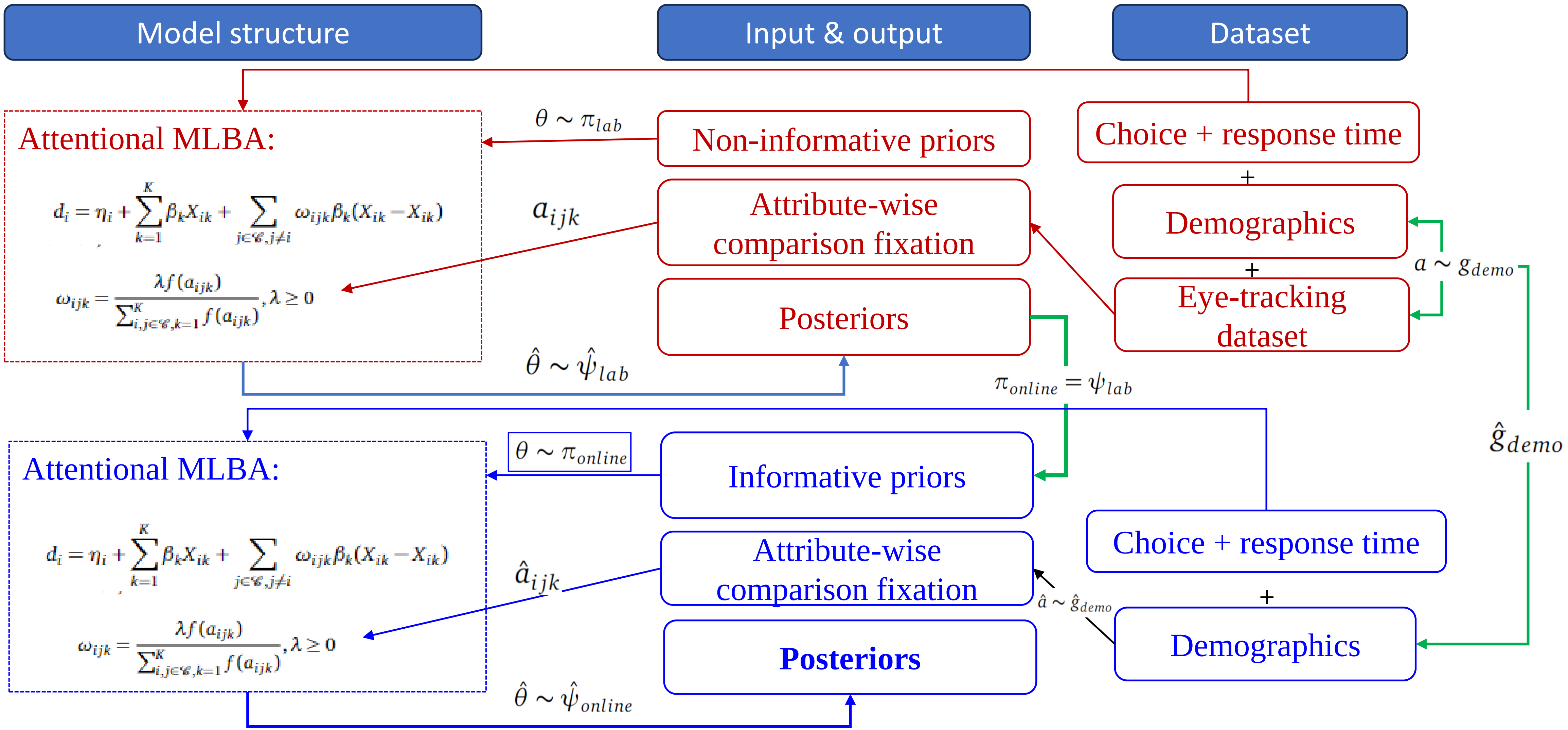
Fusing Lab & Online Data: Lower Std Error (Empirical Study)



Blue points are the mean and median of posterior

Half length of error bar is the std dev of posterior

Challenges 3,4,5: Model Selection, Sample Size, & Priors
(Fusing Choice-RT Data with and without Eye-tracking)



Conclusions

Sequential Sampling Models (SSMs) have a future (Bansal et al., 2023):

1. **Response time** is easily obtainable and should be utilized to improve statistical inference.

2. **Fusing** lab and online data is a way forward.

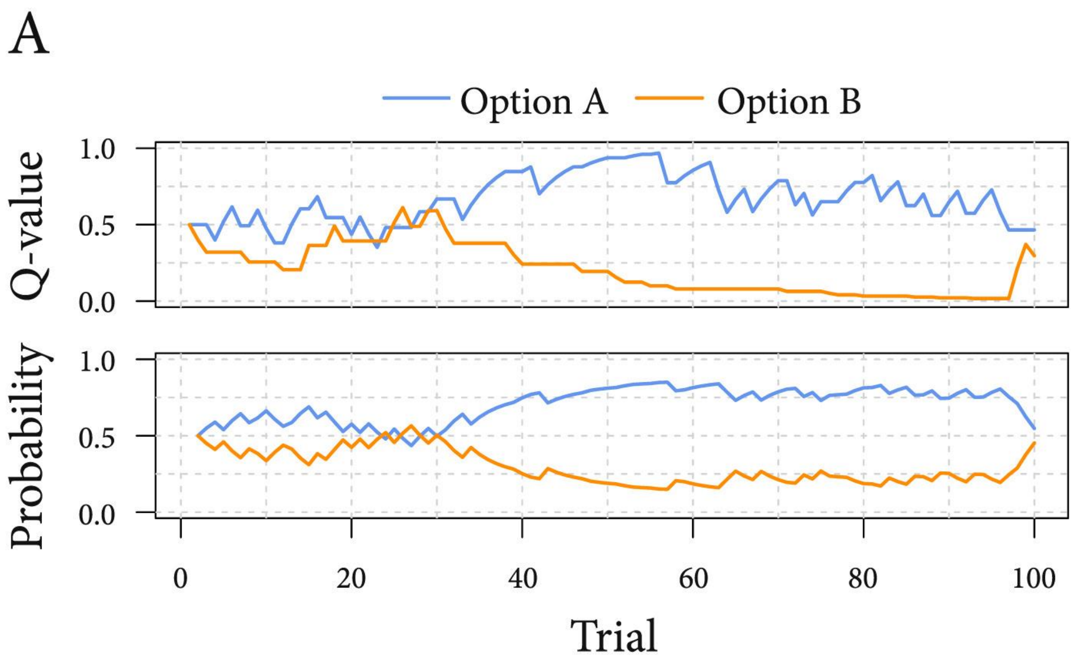
3. **Computationally-efficient** estimators need to be developed (e.g., variational inference).

4. There is potential of webcam-based eye-tracking, but still at early stages (Yang & Krajbich, 2021).

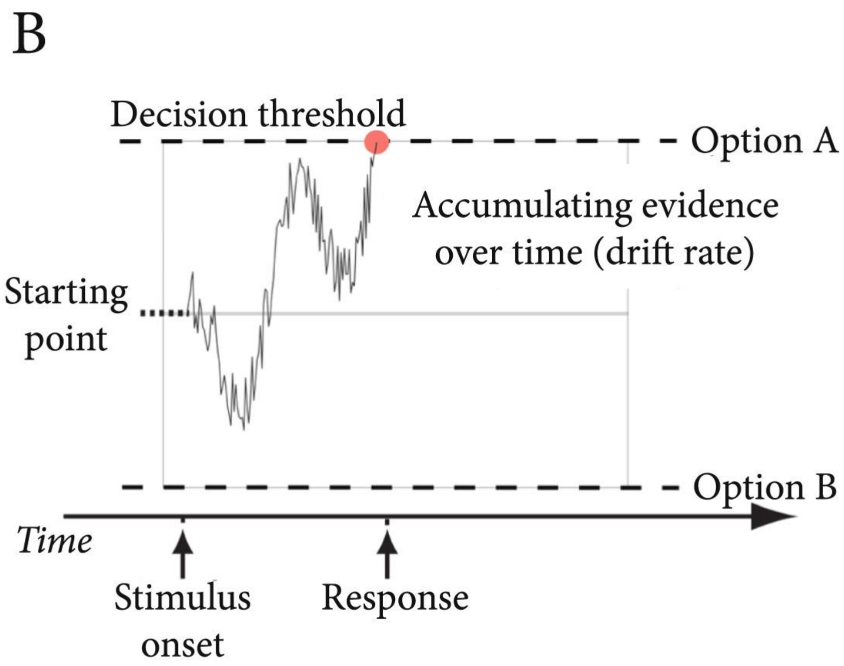
Bansal, P., Ozdemir, S., & Kim, E. J. (2023). Discrete Choice Experiments with Eye-tracking: How Far We Have Come and A Way Forward: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4324231
Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision making*, *16*(6), 1485-1505.

Conclusions

Mutual benefits of combining reinforcement learning with SSMs (Miletić et al., 2020).



$$Q_{i,t+1} = Q_{i,t} + \alpha(r_t - Q_{i,t})$$



$$dx/dt = vdt + sdW$$

Miletić, S., Boag, R. J., & Forstmann, B. U. (2020). Mutual benefits: Combining reinforcement learning with sequential sampling models. *Neuropsychologia*, *136*, 107261.

Behavioural
COGNITIVE SCIENCE LAB

NUS
National University of Singapore

Thank you !

prateekb@nus.edu.sg