Urban and Transportation planning lab.
Hiroshima University

Summer School
September 7, 2026 @University of Tokyo

AI for Behavioral Modeling: Interpretability by Design

Hiroshima University
Makoto Chikaraishi

Interpretability by Design


nature machine intelligence

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature › nature machine intelligence › perspectives › article

Perspective | Published: 13 May 2019

Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead

Cynthia Rudin

Naïve Application of Neural Network

$$Y = NN(R)$$

Behavioral outcomes

Raw data

Post-hoc explanation

Post-hoc explanation is not enough.
High-stakes applications require models that are interpretable **by design**.

Interpretability by Design at Two Stages

$$R \longrightarrow X \longrightarrow Y$$

R: Raw data

X: Human-interpretable measurements and concepts

Y: Behavioral outcomes

Measurement and Representation ($R \rightarrow X$)
Transform raw or multimodal data into human-interpretable measurements and concepts

Behavioral Modeling ($X \rightarrow Y$)
Map measurements and concepts to behavioral outcomes through theory-grounded structures

An example: MNL for Mode Choice

$$R \longrightarrow X \longrightarrow Y$$

Shortest-path-based measurement

$$r^* \in \arg\min_{r \in \mathcal{R}} TT_r$$

Extracted X:

$$x_{TT} = TT_{r^*}$$

$$x_{\text{cost}} = \text{Fare}_{r^*}$$

Structural constraints by MNL:

$$U_{ni} = V_{ni} + \varepsilon_{ni}, \qquad V_{ni} = \beta^\top x_{ni}$$

$$y_n = i \quad \Leftrightarrow \quad U_{ni} \geq U_{nj} \quad \forall j$$

$$P_{ni} = \Pr(y_n = i \mid X_n) = \frac{\exp(V_{ni})}{\sum_{j \in \mathcal{C}_n} \exp(V_{nj})}$$

The classical MNL can be viewed as replacing the unrestricted mapping $Y = NN(R)$ with two explicitly designed stages.

$$R \longrightarrow \boxed{X \longrightarrow Y}$$

Part I: Behavioral Mapping

A. NN represents the systematic utility

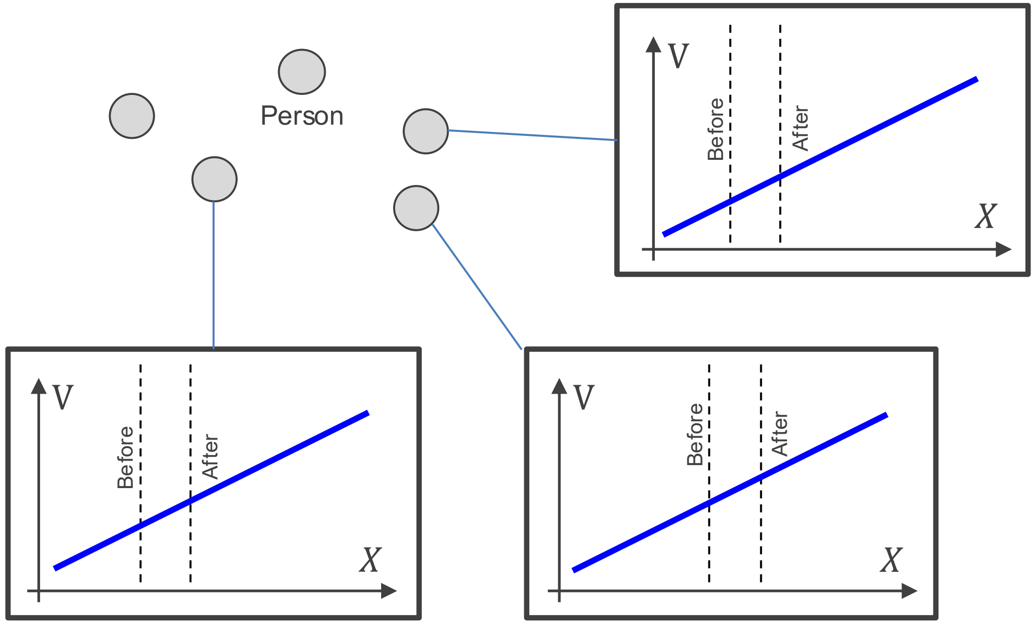
Model	What $NN(\cdot)$ represents	Utility	What is preserved
Classical MNL	no network	$V_{ni} = \beta^{\top} x_{ni}$	everything, by assumption
F-DNN	the entire utility	$V_n = NN(x_n, z_n)$	No explicit behavioral restrictions
ASU-DNN Wang et al. (2020)	alternative-specific utility	$V_{ni} = NN_i(x_{ni}, z_n)$	IIA
L-MNL Sifringer et al. (2020)	a utility residual	$V_{ni} = \beta^{\top} x_{ni}^{K} + NN_i(x_n^{L})$	linearity on x^K, so WTP for those attributes
TB-ResNet Wang et al. (2021)	a weighted mixture	$V_{ni} = (1 - \delta)\, V_{T,i} + \delta\, V_{DNN,i}$	Nothing guaranteed; δ measures the distance from the DCM
GAUNet Nishi & Hara (2026)	attribute-level utility shape	$V_{ni} = ASC_i + \sum_k NN_{ik}(x_{nik})$	additivity; the shape of each attribute's effect
DNN + gradient regularization Feng et al. (2024)	the entire utility	$V_n = NN(x_n, z_n)$ + Gradient penalty during training	monotonicity in designated attributes as a **soft** penalty
DCM-LN Kim & Bansal (2024)	constrained nonlinear utility	$V_{ni} = NN_i^{mono}(x_{ni})$	monotonicity in designated attributes, as a **hard** constraint

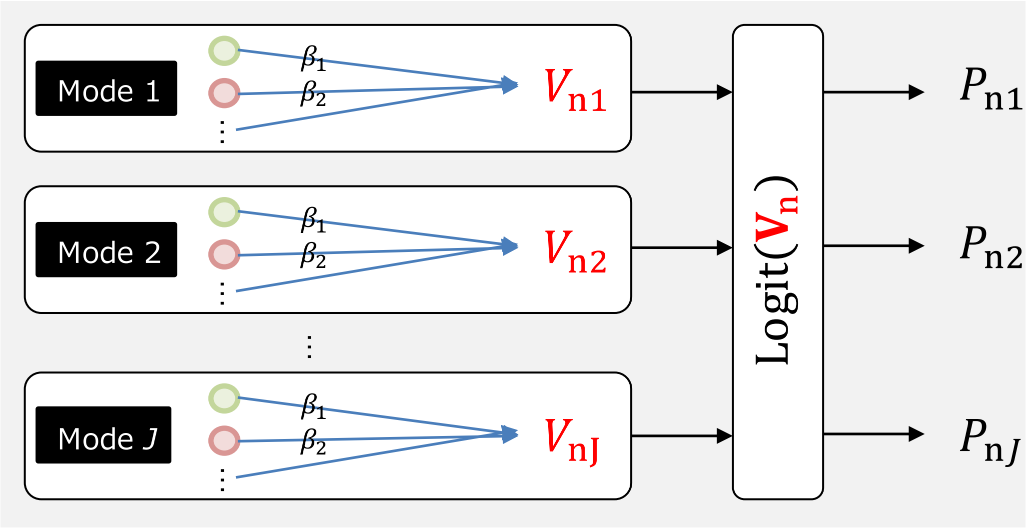
[0] MNL

$$U_{ni} = V_{ni} + \varepsilon_{ni}, \qquad V_{ni} = \beta^{\top} x_{ni}$$

$$y_n = i \quad \Leftrightarrow \quad U_{ni} \geq U_{nj} \quad \forall j$$

$$P_{ni} = \Pr(y_n = i \mid X_n) = \frac{\exp(V_{ni})}{\sum_{j \in \mathcal{C}_n} \exp(V_{nj})}$$





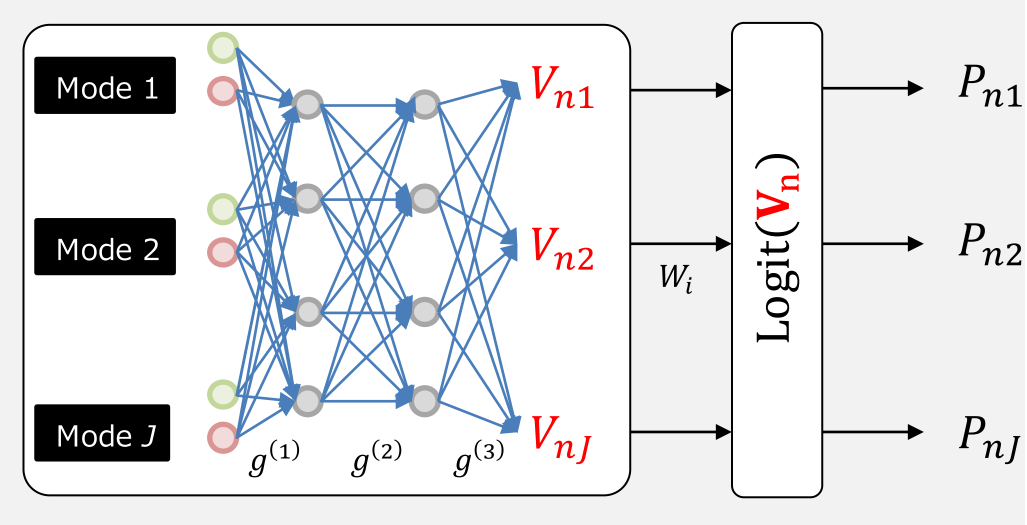
[1] Full DNN

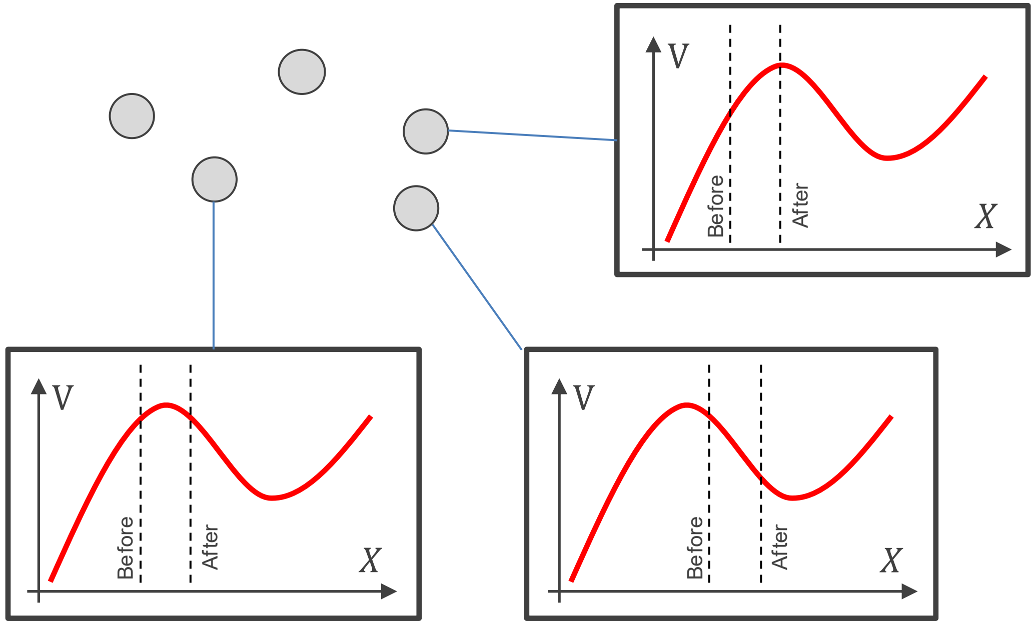
$$U_{ni} = V_{ni} + \varepsilon_{ni}$$

$$V_{ni} = NN(x_n, z_n)$$

$$y_n = i \quad \Leftrightarrow \quad U_{ni} \geq U_{nj} \quad \forall j$$

$$P_{ni} = \Pr(y_n = i \mid X_n) = \frac{\exp(V_{ni})}{\sum_{j \in \mathcal{C}_n} \exp(V_{nj})}$$





Each alternative's utility may depend arbitrarily on the attributes of all alternatives, obscuring stable alternative-specific preferences and substitution mechanisms.

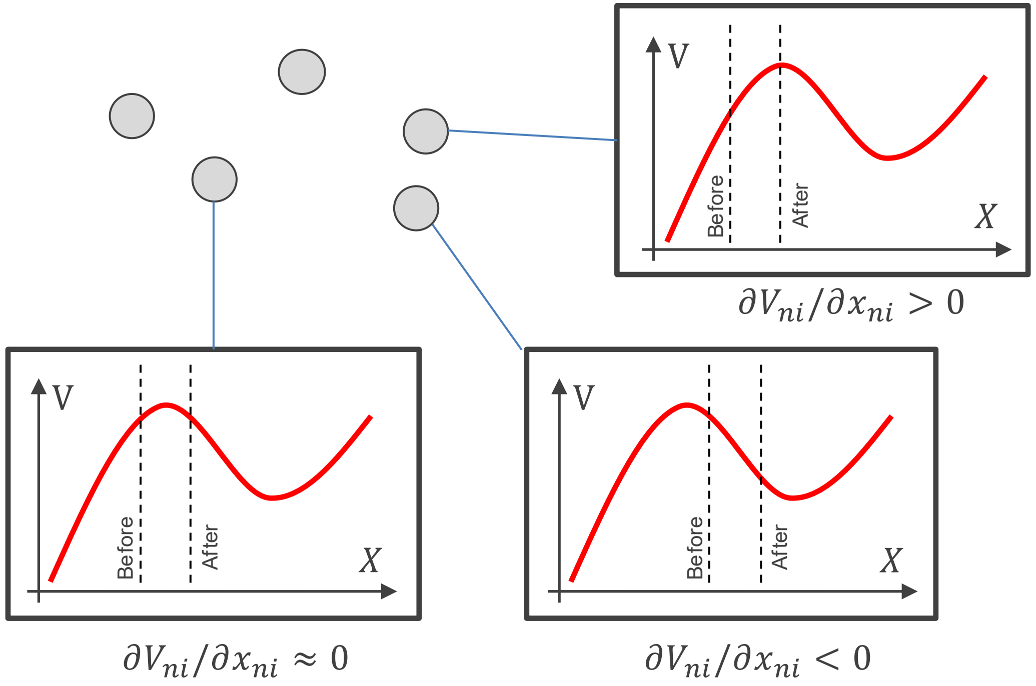
[2] ASU-DNN (Wang et al., 2020)

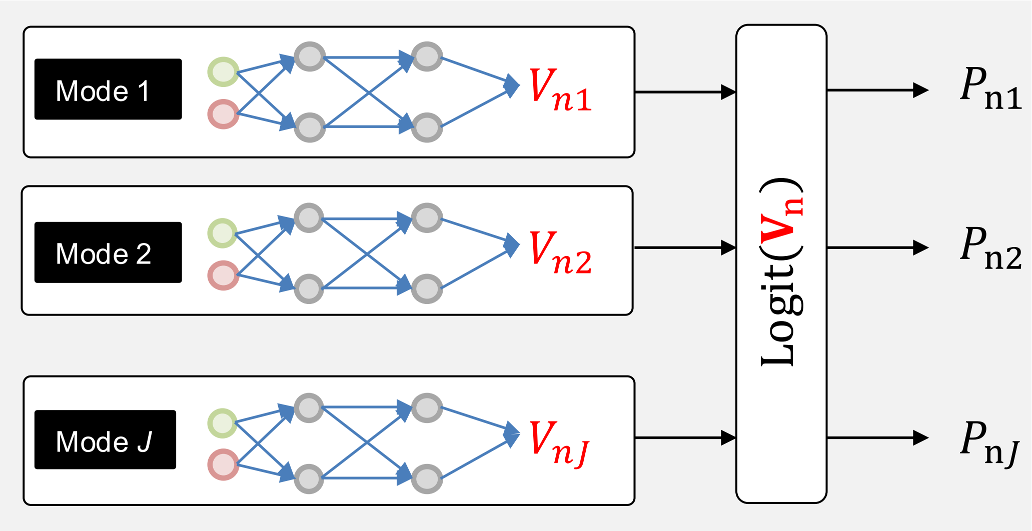
$$U_{ni} = V_{ni} + \varepsilon_{ni}$$

$$V_{ni} = NN_i(x_{ni}, z_n)$$

$$y_n = i \quad \Leftrightarrow \quad U_{ni} \geq U_{nj} \quad \forall j$$

$$P_{ni} = \Pr(y_n = i \mid X_n) = \frac{\exp(V_{ni})}{\sum_{j \in \mathcal{C}_n} \exp(V_{nj})}$$





ASU-DNN preserves alternative-specific scalar utility and therefore utility-based ranking. However, it does not guarantee monotonicity, sign consistency, or attribute-level interpretability.

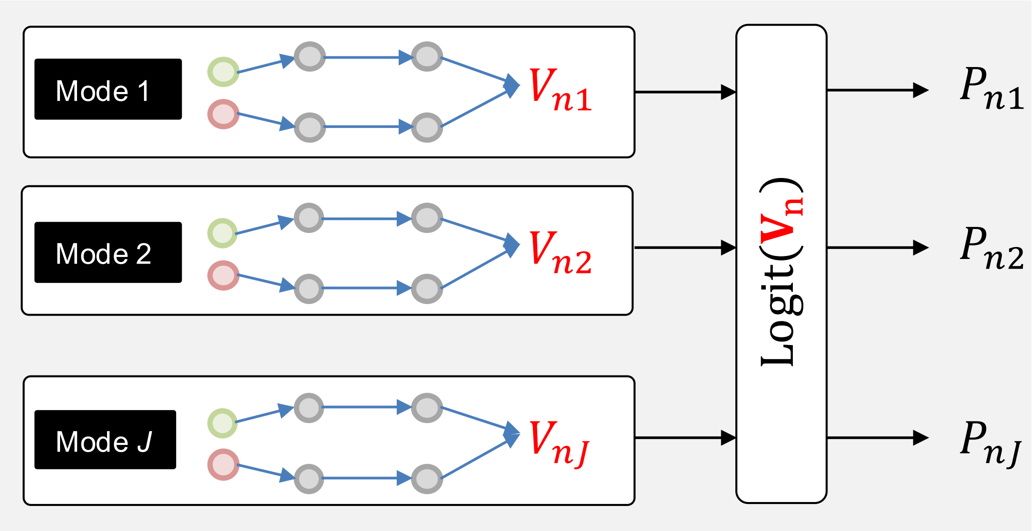
[3] GAUNet (Nishi & Hara, 2026)

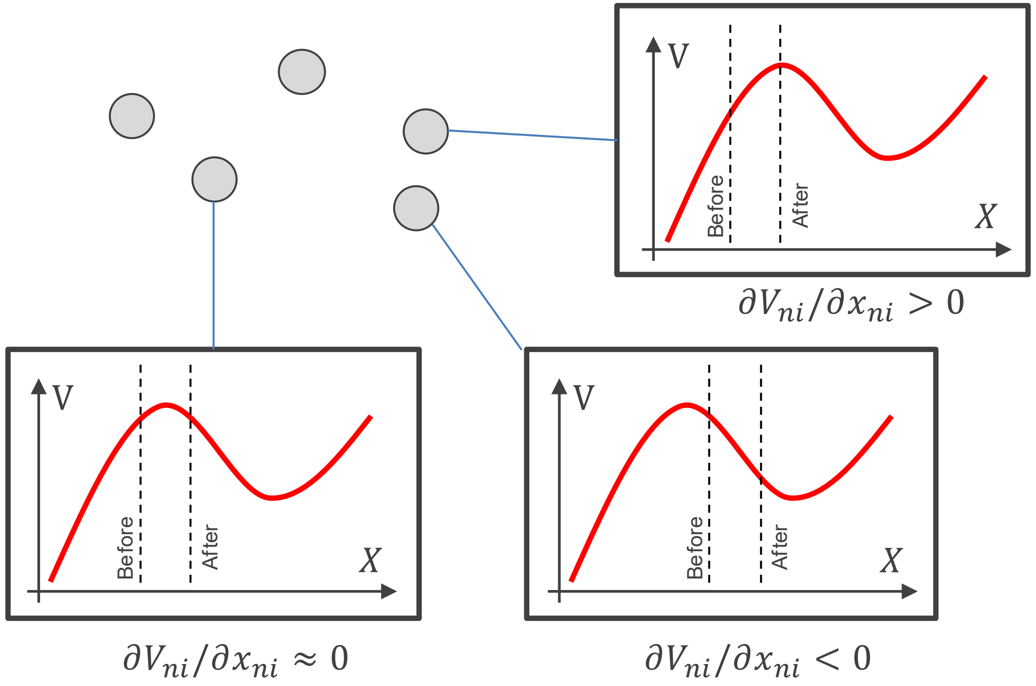
$$U_{ni} = V_{ni} + \varepsilon_{ni}$$

$$V_{ni} = ASC_i + \sum_k NN_{ik}\,(x_{nik})$$

$$y_n = i \quad \Leftrightarrow \quad U_{ni} \geq U_{nj} \quad \forall j$$

$$P_{ni} = \Pr(y_n = i \mid X_n) = \frac{\exp(V_{ni})}{\sum_{j \in \mathcal{C}_n} \exp(V_{nj})}$$





Behavioral constraints do not necessarily compromise predictive performance; they can act as effective regularization and improve generalization (Wang et al., 2020; Nishi and Hara, 2026).

However, it guarantees additivity and a readable per-attribute shape, but not monotonicity or sign.

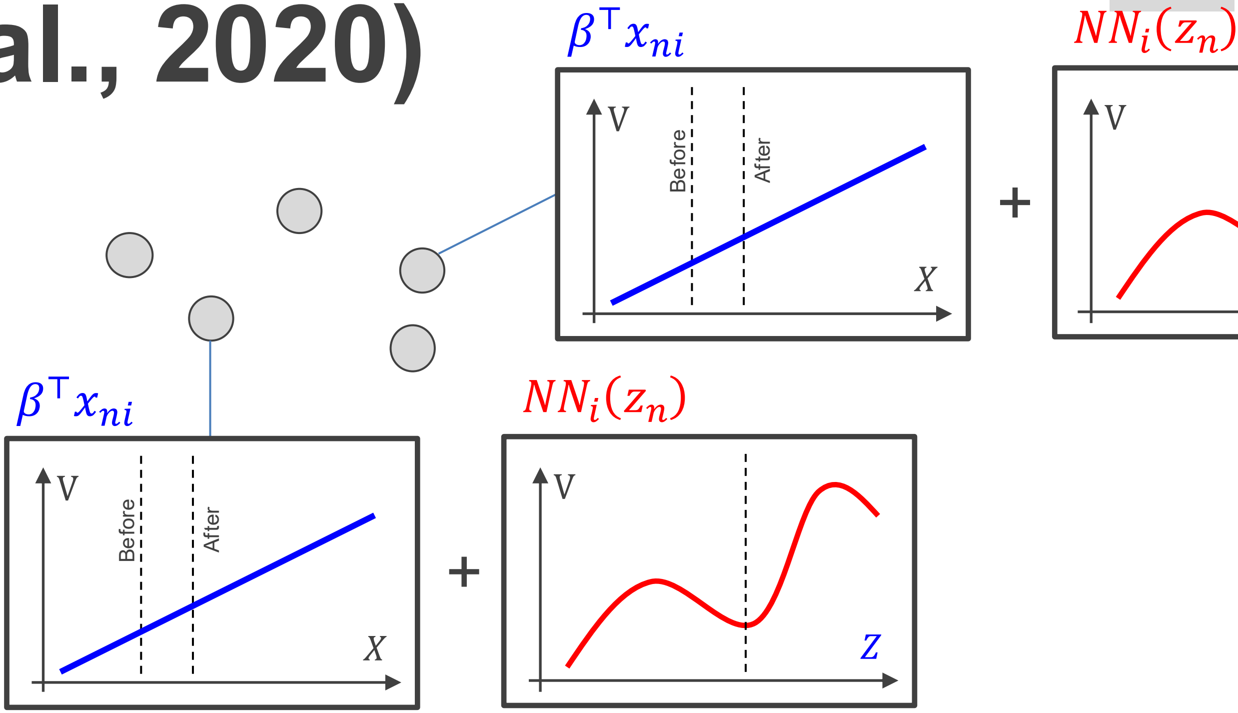
[4] L-MNL (Sifringer et al., 2020)

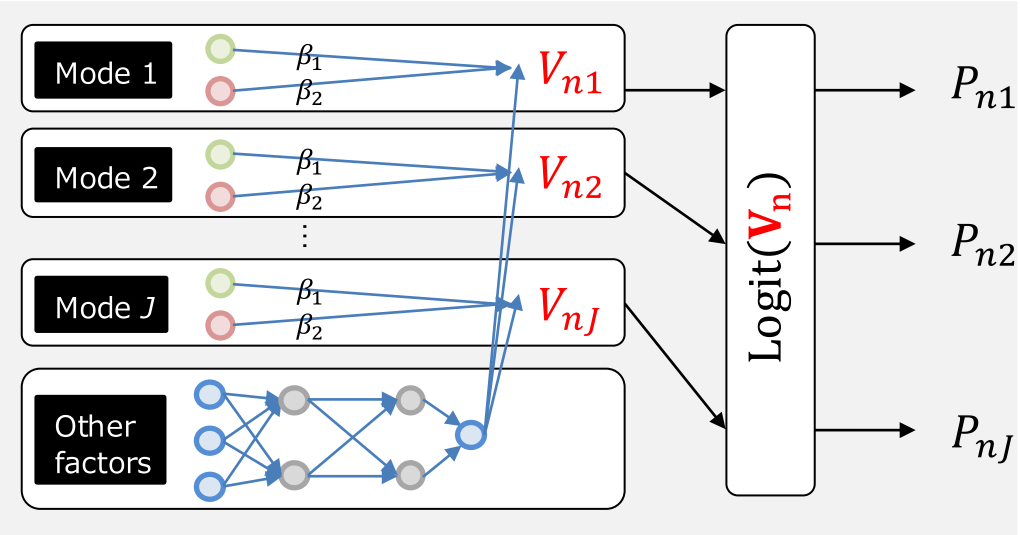
$$U_{ni} = V_{ni} + \varepsilon_{ni}$$

$$V_{ni} = \beta^{\top} x_{ni} + NN_i(z_n)$$

$$y_n = i \quad \Leftrightarrow \quad U_{ni} \geq U_{nj} \quad \forall j$$

$$P_{ni} = \Pr(y_n = i \mid X_n) = \frac{\exp(V_{ni})}{\sum_{j \in \mathcal{C}_n} \exp(V_{nj})}$$





Key attributes such as travel time and cost remain in an interpretable linear utility component, while the NN learns flexible utility corrections from the remaining variables.

Do not replace the entire behavioral model with an NN. Keep the part that must remain interpretable, and let the NN learn only the remaining utility component.

[5] Monotonicity: Hard Constraints vs. Soft Penalties

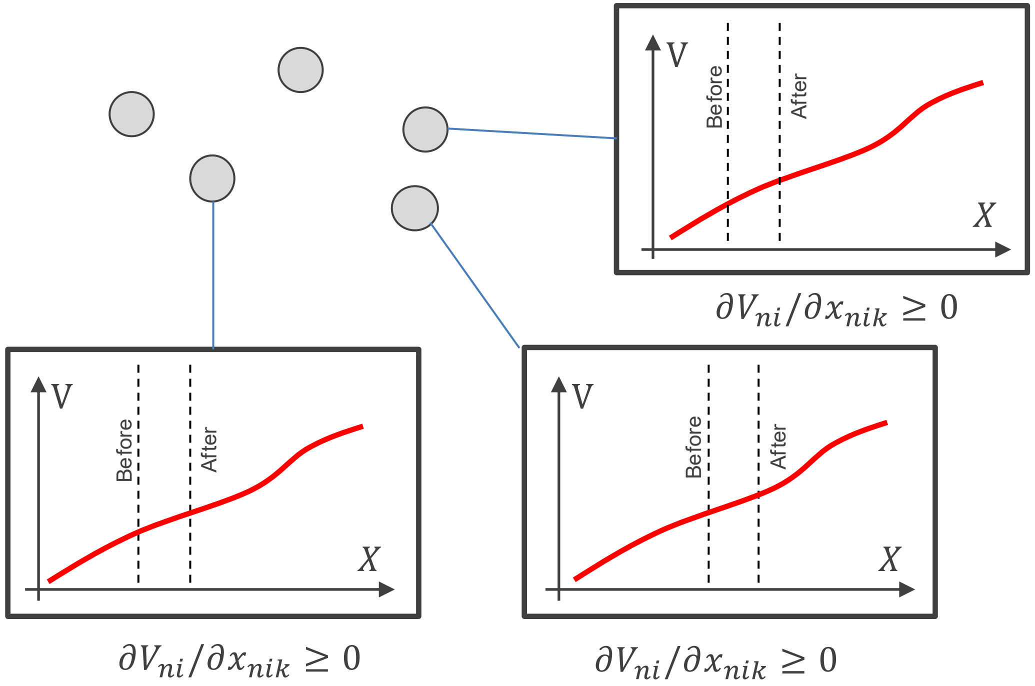
DCM-LN:

$$U_{ni} = V_{ni} + \varepsilon_{ni}$$

$$V_{ni} = NN_i^{mono}(x_{ni}) = LN_i(x_{ni})$$

$$y_n = i \quad \Leftrightarrow \quad U_{ni} \geq U_{nj} \quad \forall j$$

$$P_{ni} = \Pr(y_n = i \mid X_n) = \frac{\exp(V_{ni})}{\sum_{j \in \mathcal{C}_n} \exp(V_{nj})}$$



DCM-LN (Kim and Bansal, 2024)
Monotonicity of utility is guaranteed by design for selected attributes within the specified input domain (**Hard architectural constraint**)

DNN + gradient regularization (Feng et al., 2024)
Monotonicity of choice probabilities is encouraged through gradient regularization but is not guaranteed globally (**Soft gradient penalty**)

B. Learning taste heterogeneity

Model	How taste heterogeneity is represented	Formulation	What is preserved
Classical MNL	No heterogeneity: a common taste point	$\beta = \text{parameter}$ $V_{ni} = \beta^{\top} x_{ni}$	Linear-in-attributes utility; common and directly interpretable taste coefficients; closed-form logit probability
Entropy Tucker Ishii et al. (2022) * Non-neural, aggregate-level approach	Latent mobility-pattern-specific taste points	$c = (j_O, j_D, j_T, j_K)$ $\hat{X}_{odtk} = \sum_{c} w_{odtk,c} \exp(-\beta_c D_{od})$ $w_{odtk,c} = s_c\, a^{(O)}_{o,j_O} a^{(D)}_{d,j_D} a^{(T)}_{t,j_T} a^{(K)}_{k,j_K}$	Entropy/gravity structure; separation of latent attractiveness from travel impedance; interpretable pattern-specific impedance β_c
TasteNet Han et al. (2022)	An individual conditional taste point	$\beta_n = NN(z_n)$ $V_{ni} = \beta_n^{\top} x_{ni}$	Linear-in-attributes utility conditional on β_n; person-level WTP and elasticities; coefficient signs through output transformations
MDN-Logit Li et al. (2026)	A conditional taste distribution	$\{\pi_{nm}, \mu_{nm}, L_{nm}\}_{m=1}^{M} = NN(z_n)$ $\beta_n \sim \sum_{m=1}^{M} \pi_{nm}\, \mathcal{N}\left(\mu_{nm}, L_{nm} L_{nm}^{\top}\right)$ $V_{ni} = \beta_n^{\top} x_{ni}$	Linear-in-attributes utility conditional on β_n; RUM/logit kernel; interpretable taste distributions and flexible substitution; P_{ni}obtained by integrating over β_n

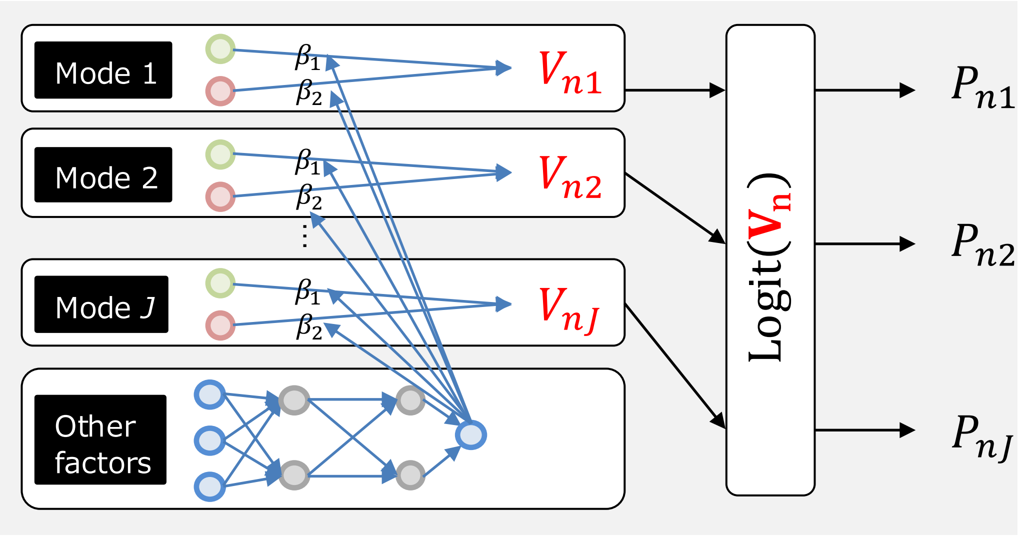
[6] TasteNet (Han et al., 2022)

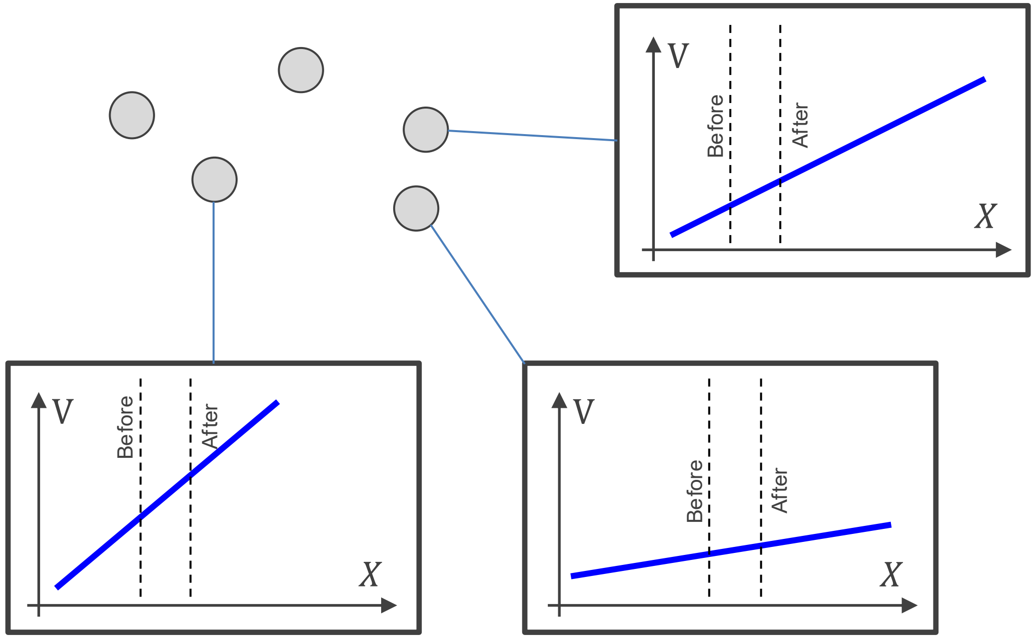
$$U_{ni} = V_{ni} + \varepsilon_{ni}$$

$$V_{ni} = \beta_n{}^{\top} x_{ni} \qquad \beta_n = NN_i(z_n)$$

$$y_n = i \quad \Leftrightarrow \quad U_{ni} \geq U_{nj} \quad \forall j$$

$$P_{ni} = \Pr(y_n = i \mid X_n) = \frac{\exp(V_{ni})}{\sum_{j \in \mathcal{C}_n} \exp(V_{nj})}$$





The NN maps individual characteristics z_n to individual-specific taste coefficients β_n. Utility remains linear in the alternative attributes x_{ni}, while its slopes vary flexibly across individuals.

Mixed Logit: Prespecified Taste Distribution

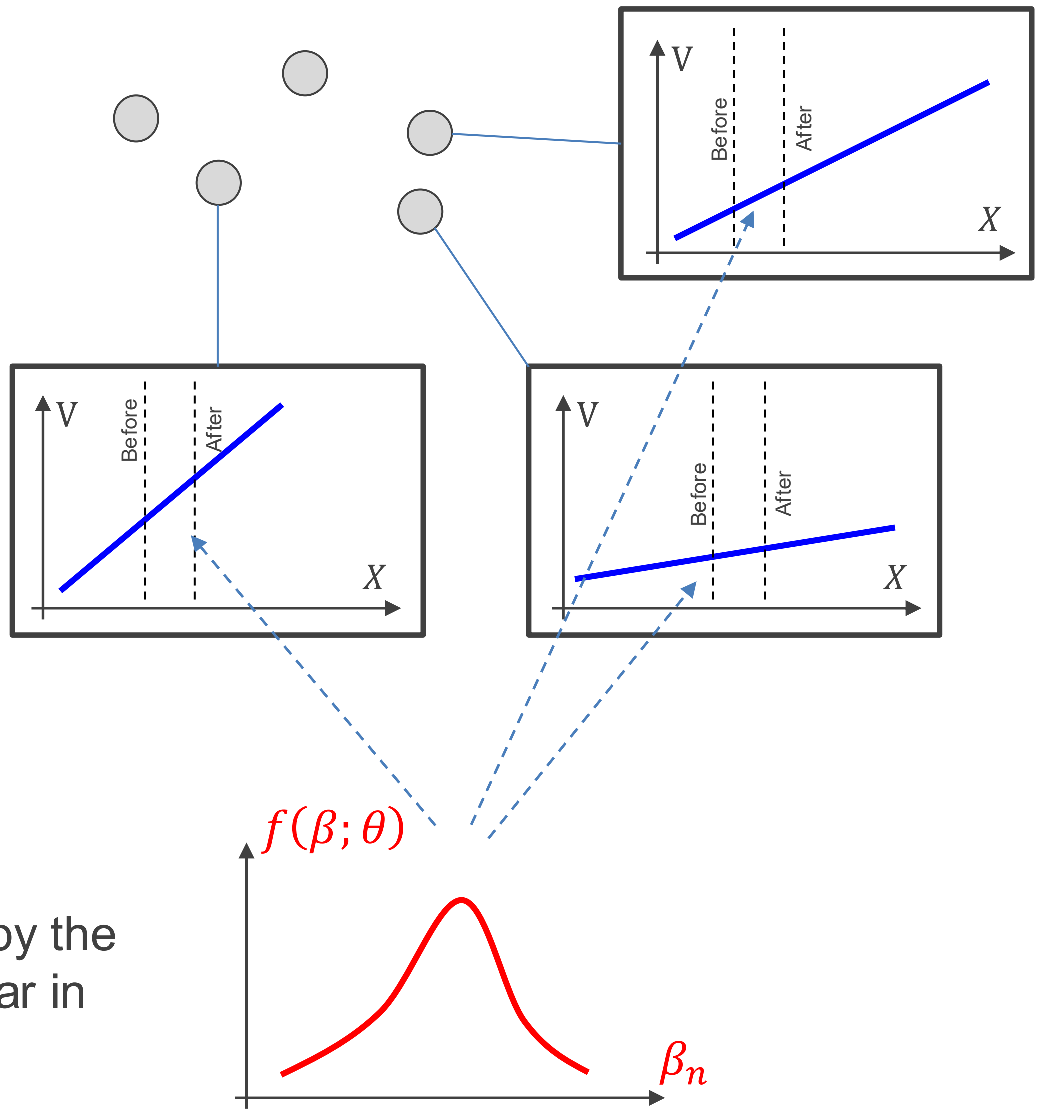
Mixed logit:

$$U_{ni} = V_{ni} + \varepsilon_{ni}, \qquad V_{ni} = \beta_n^\top x_{ni}$$

$$L_{ni}(\beta) = \frac{\exp\left(\beta^\top x_{ni}\right)}{\sum_{j \in \mathcal{C}_n} \exp\left(\beta^\top x_{nj}\right)}$$

$$P_{ni} = \int L_{ni}(\beta) f(\beta;\theta)\, d\beta \approx \frac{1}{R}\sum_{r=1}^{R} L_{ni}\left(\beta^{(r)}\right)$$

$$\beta_n \sim f(\beta;\theta)$$



The distributional family is prespecified by the analyst. Conditional utility remains linear in the alternative attributes x_{ni}.

[7] MDN-Logit (Li et al., 2026)

MDN-logit:

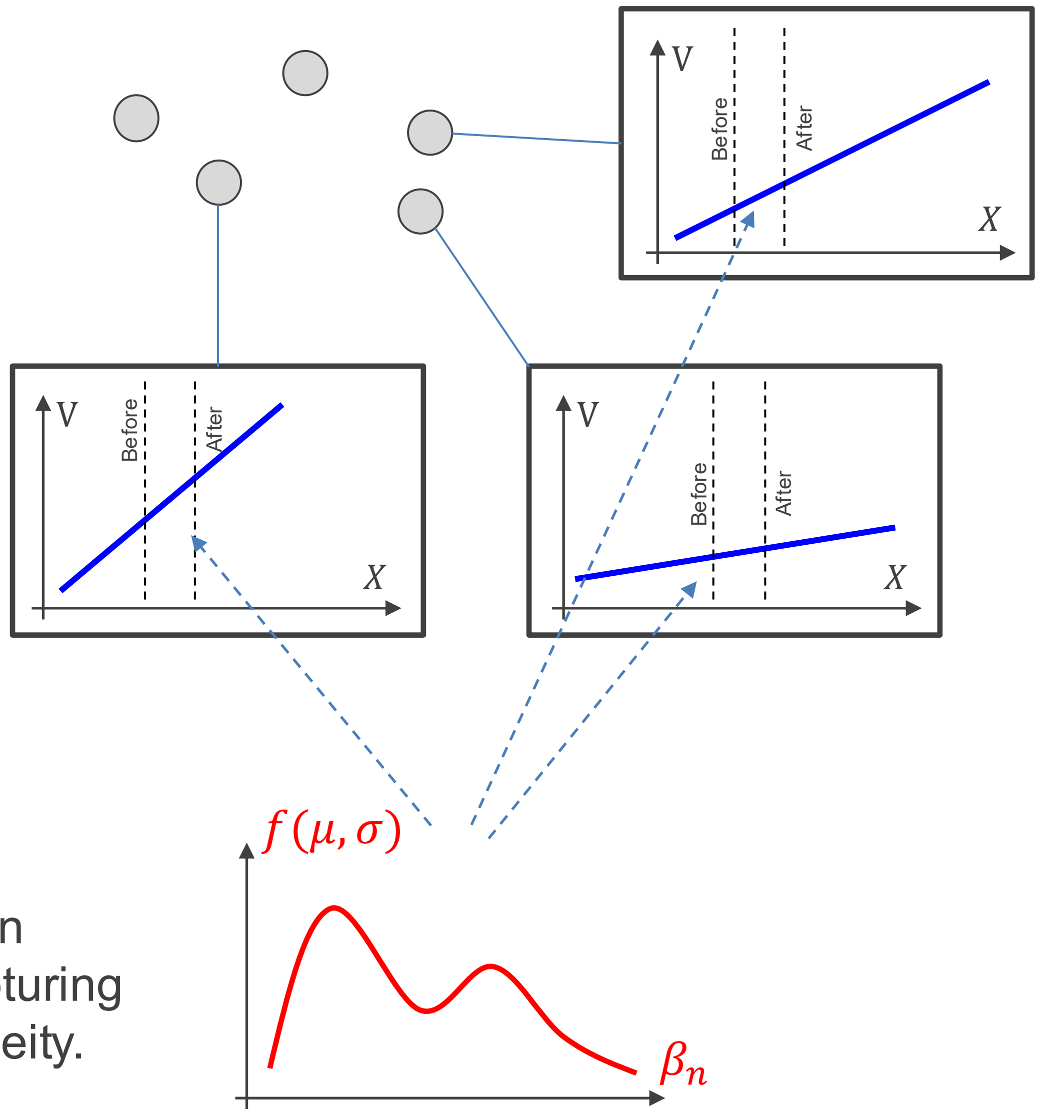
$$U_{ni} = V_{ni} + \varepsilon_{ni}, \qquad V_{ni} = \beta_n^\top x_{ni}$$

$$P_{ni} = \int L_{ni}(\beta) p_\theta(\beta \mid z_n)\, d\beta \approx \frac{1}{R} \sum_{r=1}^{R} L_{ni}\left(\beta^{(r)}\right)$$

Taste heterogeneity:

$$\{\pi_{nm}, \mu_{nm}, L_{nm}\}_{m=1}^{M} = NN(z_n)$$

$$\beta_n \sim p_\theta(\beta \mid z_n) = \underbrace{\sum_{m=1}^{M} \pi_{nm}\, \mathcal{N}\left(\mu_{nm}, L_{nm} L_{nm}^\top\right)}_{\text{Gaussian mixture}}$$



The MDN learns a flexible taste distribution conditional on observed characteristics z_n, capturing both systematic and residual taste heterogeneity. Conditional utility remains linear in x_{ni}.

Open modeling issue: How should individual taste heterogeneity be regularized?

Revisiting regularization

For heterogeneity modeling, regularization should do more than prevent overfitting: it should encode how much individual-level variation should be trusted.

$$\underbrace{\text{Fit to individual } n}_{\text{Data}} + \underbrace{\text{Pull toward the population}}_{\text{Penalty}}$$

A well-designed penalty should induce statistically meaningful shrinkage.

Li et al. (2026): $\mathcal{L} = -\mathrm{SLL} + \lambda(\|w\| + \|b\|)$, λ chosen by cross-validation

Does this regularization encode the desired shrinkage structure?

BLUP: Shrinkage as a Statistical Principle

Key idea: individual effects should not be estimated independently when individual data are limited.

A linear mixed-effects model: $\beta_n \sim \mathcal{N}(\mu, \Sigma)$

The BLUP (Best Linear Unbiased Predictor) of the individual parameter can be written schematically as

$\hat{\beta}_n^{\text{BLUP}} = W_n \tilde{\beta}_n + (I - W_n)\mu,$ where $\tilde{\beta}_n$ is the individual-specific estimate. $\quad W_n \sim \dfrac{\text{Ind. precision}}{\text{Ind. precision} + \text{Pop. precision}}$

More informative individual data: $W_n \to I \;\Rightarrow$ trust the individual estimate.

Less informative individual data: $W_n \to 0 \;\Rightarrow$ shrink toward the population mean.

The same estimator is obtained from penalized estimation:

$$\min_{\{\beta_n\}} \sum_n \Big[\underbrace{\mathcal{L}_n(\beta_n)}_{\text{Data}} + \underbrace{\frac{1}{2}(\beta_n - \mu)^\top \Sigma^{-1} (\beta_n - \mu)}_{\text{Penalty}} \Big]$$

Data

(individual information)

Gaussian negative log-likelihood

Penalty (Generalized L_2/ Mahalanobis penalty)

Panel MXL: BLUP-like Shrinkage in Distribution Space

Key idea: individual effects should not be estimated independently when individual data are limited.

A panel mixed logit model: $\beta_n \sim p_\eta(\beta), \qquad L_n(\beta_n) = \prod_{t=1}^{T_n} P\left(y_{nt} \mid \beta_n\right)$

$$\underbrace{p(\beta_n \mid D_n, \eta)}_{\text{posterior}} \propto \underbrace{L_n(\beta_n)}_{\text{Likelihood}}\, \underbrace{p_\eta(\beta_n)}_{\text{prior}}$$

$$-\log p\left(D_n \mid \eta\right) = \min_{q_n}\left\{\underbrace{-\mathbb{E}_{q_n}[\log L_n\left(\beta\right)]}_{\substack{\text{Data}\\\text{(individual information)}}} + \underbrace{\mathrm{KL}\left(q_n(\beta) \,\|\, p_\eta(\beta)\right)}_{\text{Penalty}}\right\}$$

$q_n^*(\beta) = p(\beta_n \mid D_n, \eta)$

q_n: any candidate distribution over individual taste β_n

Weak individual information keeps q_n close to the population distribution, while strong individual information allows q_n to concentrate elsewhere.

Proposed direction:

From global regularization to information-adaptive shrinkage

$\lambda(\|w\| + \|b\|) \rightarrow \mathrm{KL}\left(q_n(\beta) \,\|\, p_\theta(\beta \mid z_n)\right)$

Shrink only when individual evidence is weak

C. Generalizing the Stochastic Choice Structure

Model	What $NN(\cdot)$ represents	Formulation	What is specified / learned	What is preserved
RUM-NN Bagheri et al. (2025)	A differentiable simulator of a specified RUM; optionally, nonlinear systematic utility	$\eta_n^{(q)} \sim F_0, \varepsilon_n^{(q)} = L\eta_n^{(q)}$ $U_n^{(q)} = V_n + \varepsilon_n^{(q)}$ $\hat{P}_n = \frac{1}{Q}\sum_q \text{softmax}\left(U_n^{(q)}/\lambda\right)$	Error distribution and correlation	Implements a specified RUM as a differentiable simulator
RUMnets Aouad & Désir (2025)	The random utility function and sampled latent person- and alternative-side representations	$U_{ni}^{(k,\ell)}$ $= NN^U\left(x_i, N, N_k^{\varepsilon}, (x_i) z_n NN_\ell^{\nu}(z_n)\right)$ $P_{ni} = \frac{1}{K^2}\sum_{k,\ell} \text{softmax}_i\left(U_n^{(k,\ell)}\right)$	The entire latent random-utility representation	Learns a flexible distribution over RUM-consistent utilities

Part I summary: Where should NN enter?

A. Systematic utility

Learn nonlinear utility, while preserving selected behavioral constraints.

B. Taste heterogeneity

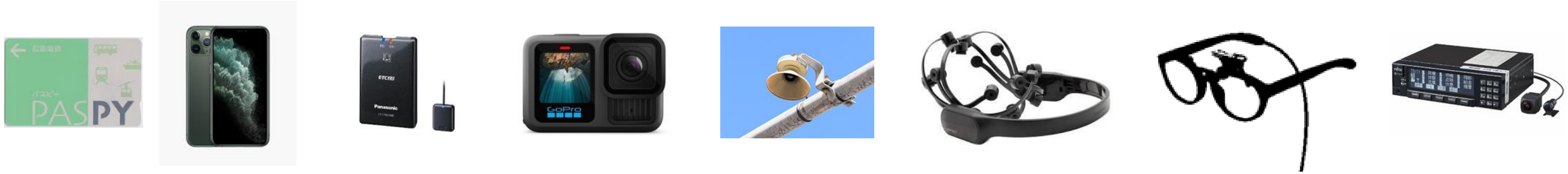
Learn individual tastes or taste distributions while preserving an interpretable utility structure.

C. Stochastic choice structure

Learn richer error and substitution patterns, while preserving RUM consistency.

Key message

- Ask where flexibility is needed, what should be preserved, and how it should be regularized.
- Flexible heterogeneity requires information-adaptive shrinkage.



$$R \longrightarrow X \longrightarrow Y$$

Part II: From Raw Data to Behaviorally Relevant Measurements

Common Digital Data Sources

Raw Data R	What It Records	Spatial Res.	Temporal Res.	Key Applications
Image / video	Visual scenes	Pixel-level	Frame rate	Counting, classification
GTFS	Routes, stops, timetables	Stop/route	Scheduled or real-time	Accessibility, service quality
GPS / smartphone	Lat, lon, timestamp, speed	Meter-level	Seconds–minutes	Trip detection, mode inference
Bus location	Vehicle position, speed	Meter-level	Seconds–minutes	Reliability, dwell time
ETC / probe vehicle	Vehicle trajectory	GPS/link	Seconds–minutes	Map matching, link travel time
Smart card (IC card)	Card ID, timestamp, station	Station/stop	Transaction	OD estimation, transfer analysis
Mobile phone (CDR)	Cell tower connections	100 m–km	Event-driven	Population flow, OD
Mobile spatial statistics	Aggregated population from mobile network signals	500 m mesh	Hourly	Population distribution, OD flow

From Raw Data to Concepts

- Digital data record scenes, signals, and traces.
- Behavioral models require meaningful attributes, perceptions, and states.
- Connecting the two requires measurement assumptions.

$$R \longrightarrow X \longrightarrow Y$$

Raw data R	Measurement / concept X	Behavioral outcome Y
Street-level images	Perceived safety	Cycling route choice
Eye movements	Attention allocation	Vehicle choice
EEG signals	Cognitive effort / inattention	Route switching

What does the data measure—and how should it enter a behavioral model?

Conventional Two-Stage Approach

$$R \longrightarrow X \longrightarrow Y$$

$$\hat{X} = g_\theta(R) \qquad\qquad P\left(Y \mid \hat{X}\right) = h_\beta\left(\hat{X}\right)$$

- **Semantic enrichment** has long been conducted in transportation field.
  - Map matching (source: GPS trajectory data) [e.g., Lou et al., 2009]
  - Travel mode (source: GPS trajectory data) [e.g., Feng & Timmermans, 2013; AlOlabi & Chikaraishi, 2026]
  - OD estimation (source: traffic count data) [e.g., Yang et al., 1992]
  - Trip purpose (source: smart card data) [e.g., Kusakabe & Asakura, 2014]
  - Comprehensive Package (source: GPS trajectory data) [Hara, 2017]

Measurement–Behavior Link: Three Questions

Conventional two-stage approach:

$$R \xrightarrow[\hat{X} = g_\theta(R)]{} X \xrightarrow[P(Y \mid \hat{X}) = h_\beta(\hat{X})]{} Y$$

Question	What should we examine?	Alternative formulations
Q1. Information used	What information is used to infer X?	$p(X \mid R)$ Raw data only or $p(X \mid R, Y)$ Raw data + observed behavior
Q2. Concept supervision	Is direct supervision used for X?	$L_X = d(g_\theta(R), X^{\text{label}})$ With concept supervision or No direct labels for X Without concept supervision
Q3. Uncertainty treatment	Is uncertainty in X retained downstream?	$f(\hat{X})$ Point estimate or $E_q[f(X)] = \int f(X) q(X)\, dX$ Distribution retained

$q(X)$: inferred distribution of X; $f(X)$: downstream prediction or evaluation.

[1] Predicting Named Concepts

Ito et al. (2026)

- ✓ A pretrained vision model predicts perceived safety.
- ✓ Safety scores enter a mixed logit model alongside other route attributes.
- ✓ The measurement model is fixed; preference parameters are estimated.

Street-level images — Perceived safety — Cycling route choice

$$R \xrightarrow[\hat{X}_i = g_{\hat{\theta}}(R_i)]{} X \xrightarrow[V_{ni} = \beta_n^\top a_{ni} + \gamma_n \hat{X}_i]{} Y$$

a: travel time, traffic lights, and segmentation shares

Question	This study
Q1. Information used	R only: $\hat{X}_i = g_{\hat{\theta}}(R_i)$
Q2. Concept supervision	Yes: external safety ratings
Q3. Uncertainty treatment	Point score; measurement uncertainty not explicitly retained

[1] Predicting Named Concepts

Ito et al. (2026)

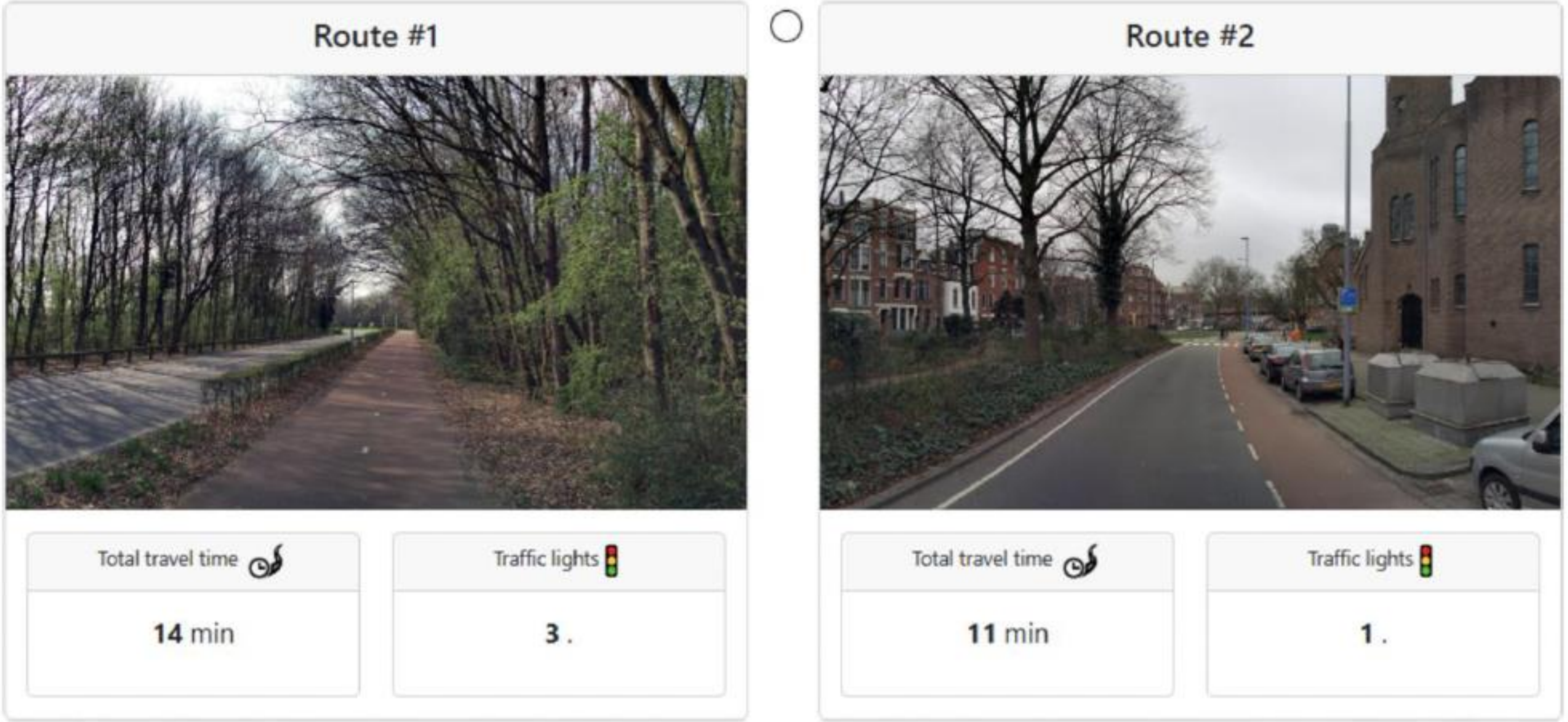


Fig. 2 Example choice task from the stated choice experiment of Terra et al. (2025). Each of the two route alternatives is presented as a street-level image with its travel time and number of traffic lights shown beneath; respondents selected their preferred route

[2] Learning from Choices

van Cranenburgh & Garrido-Valenzuela (2025)

- ✓ Image features & preference parameters are jointly learned from choice data.
- ✓ Numeric attributes and image features enter utility.
- ✓ Individual feature dimensions have no predefined meaning.

Street-level images — Learned image features — Residential location choice

$$R \xrightarrow{X_i = g_{\hat{\theta}}(R_i)} X \xrightarrow{V_{ni} = \hat{\beta}^\top a_{ni} + \hat{\omega}^\top X_i} Y$$

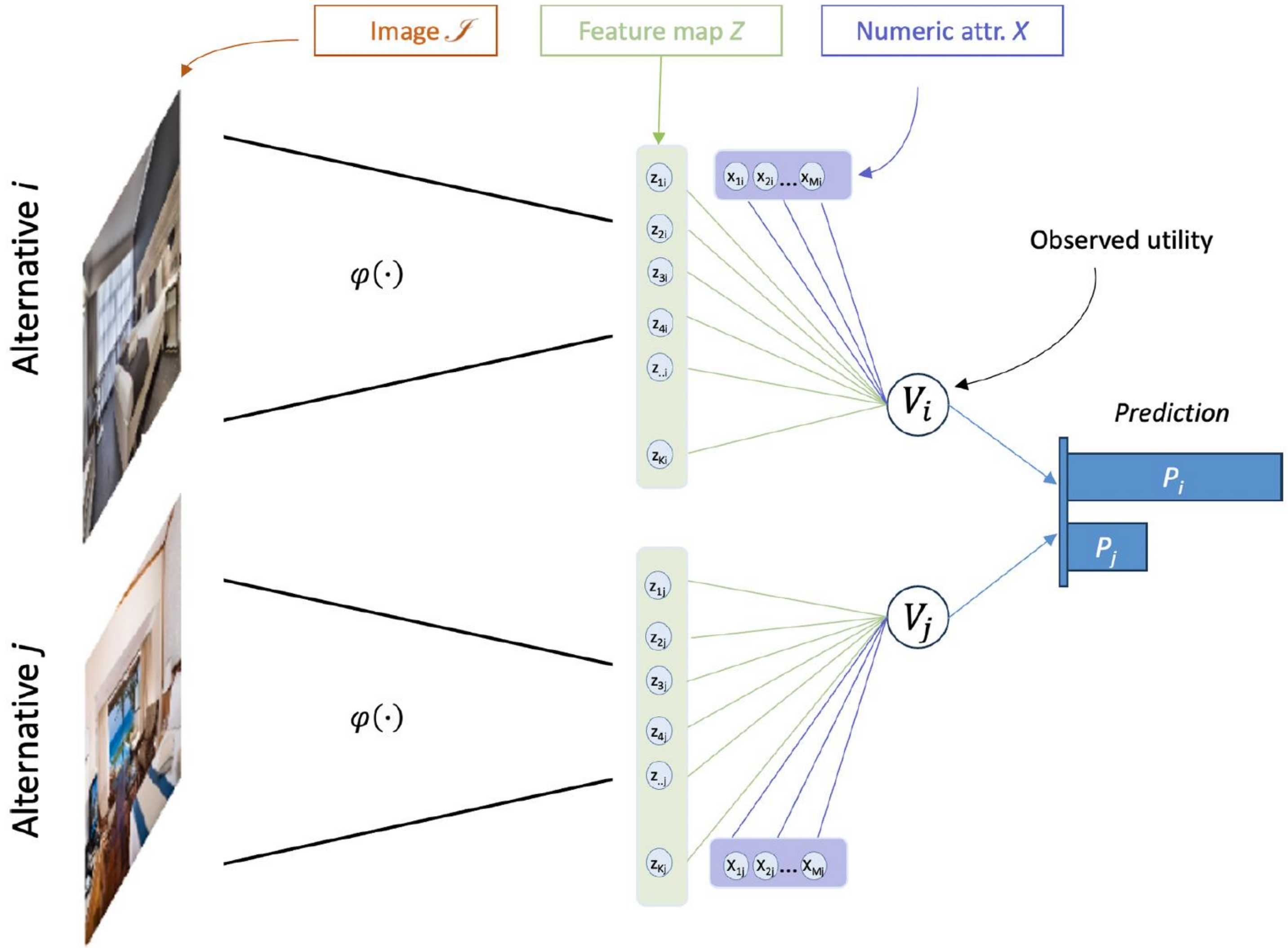
X: representation without predefined concept meanings

$$(\hat{\theta}, \hat{\beta}, \hat{\omega}) = \arg\min_{\theta,\beta,\omega} \mathcal{L}_{\text{choice}}(\theta, \beta, \omega)$$

Question	This study
Q1. Information used	R only when computing X_i; choices used during training
Q2. Concept supervision	No direct labels for the feature dimensions
Q3. Uncertainty treatment	Deterministic representation; no explicit concept-error model

[2] Learning from Choices

van Cranenburgh & Garrido-Valenzuela (2025)



[3] Modeling Latent Concepts

Agrawal & Peeta (2021)

- ✓ EEG-derived indicators measure latent cognitive states.
- ✓ A measurement model separates the indicators from the concepts.
- ✓ Latent cognitive states are linked to route switching.

EEG signals | Cognitive effort / inattention | Route switching

Measurement model: $X \rightarrow I$

$$R \longrightarrow X \longrightarrow Y$$

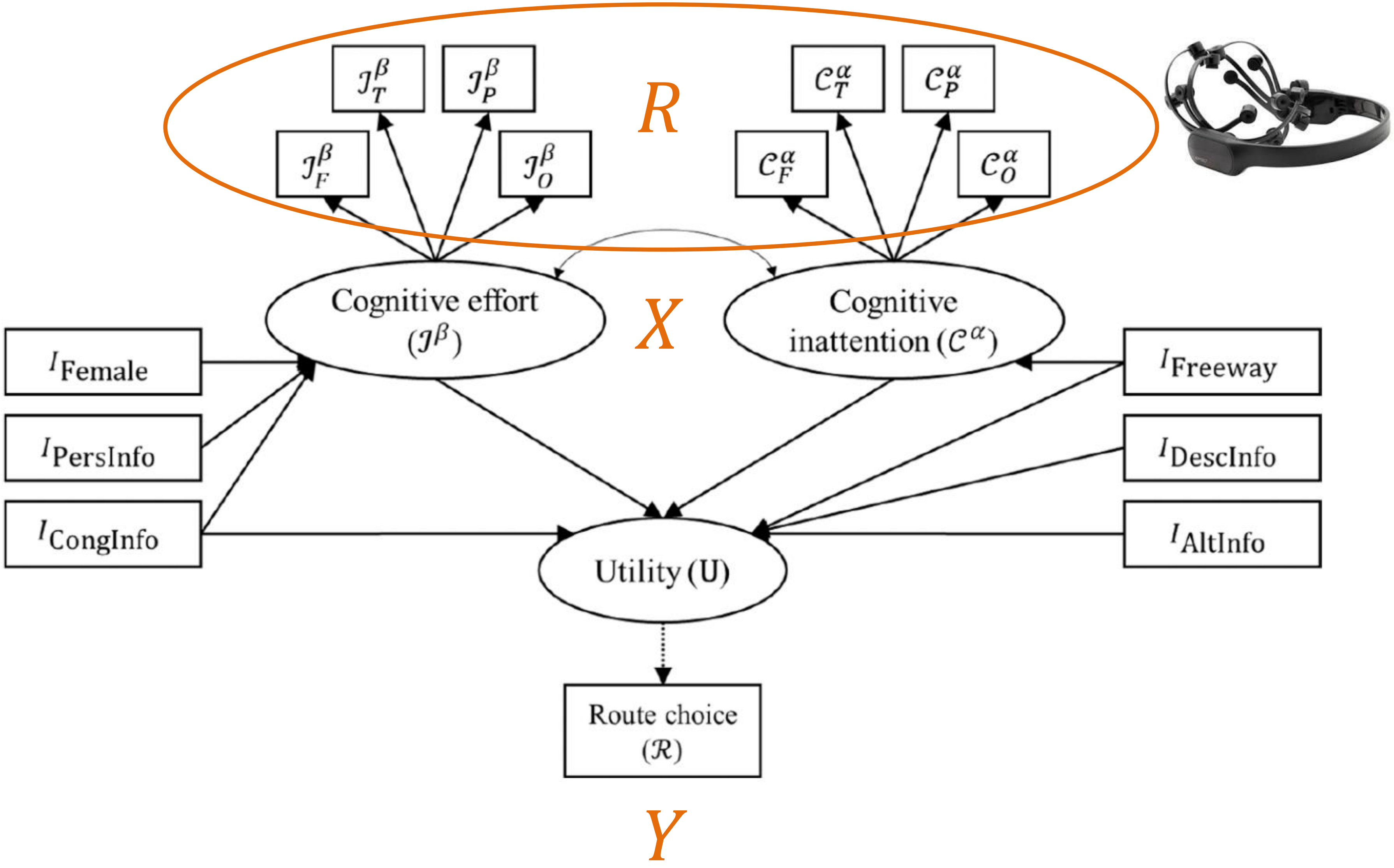
$$I_n = T(R_n), \quad I_n = \Lambda X_n + \varepsilon_n$$

$$P(Y_n = 1 \mid X_n, a_n) = \Phi(\beta^\top a_n + \gamma^\top X_n)$$

Question	This study
Q1. Information used	Indicators and choices are linked through a joint latent-variable model
Q2. Concept supervision	No direct concept labels; interpretation based on prior literature
Q3. Uncertainty treatment	Latent variation and indicator residuals are modeled

[3] Modeling Latent Concepts

Agrawal & Peeta (2021)



[4] Keeping Inference Uncertainty

Ali et al. (2026)

- ✓ Smartphone diaries contain trips with untagged purposes.
- ✓ Inferred purposes imply several possible daily trip patterns.
- ✓ These patterns can be collapsed to one or retained with probability weights.

Smartphone travel records — Trip purposes — Daily trip counts by purpose

$$R \longrightarrow X \xrightarrow[\text{Aggregation by purpose}]{} Y$$

Deterministic inference: $\hat{X} = \arg\max_X p_\theta(X \mid R)$ $\quad\hat{Y} = A(\hat{X})$

Probabilistic inference: $w_c = p_\theta(X_c \mid R)$ $\quad Y_c = A(X_c)$ $\quad \ell(\beta) = \sum_c w_c \log L_\beta(Y_c)$

Question	This study
Q1. Information used	$p(X \mid R)$; no feedback from the trip-generation model
Q2. Concept supervision	User-tagged trip purposes
Q3. Uncertainty treatment	Single assignment vs. probability-weighted candidate patterns

[5] Where Should Measurements Enter?

Hancock, Hess & Choudhury (2025)

- ✓ The same physiological features can enter different model components.
- ✓ They may affect utility, initial preference, or process noise.
- ✓ The contribution of the measurements depends on their behavioral role.

Heart rate, electrodermal activity, and time-pressure condition

Model-based stress index

Gap acceptance / rejection

$$R \longrightarrow X \longrightarrow Y$$

$$X_{ni} = \alpha_T T_n + \alpha_H HR_{ni}^{\text{norm}} + \alpha_S SCR_{ni}^{\text{norm}}$$

Decision Field Theory (DFT)

$$\mathbf{H}_{n,t} = S\mathbf{H}_{n,t-1} + \mathbf{v}_{n,t} + \boldsymbol{\varepsilon}_{n,t}$$

Question	This study
Q1. Information used	Computed from R; coefficients estimated using choices.
Q2. Concept supervision	No direct stress labels
Q3. Uncertainty treatment	No explicit measurement-error model

Entry point	Simplified specification
MNL utility	$V_{ni,\text{accept}} = V_{ni,\text{accept}}^0 + X_{ni}$
DFT initial preference	$H_{ni,\text{accept}}^0 = \bar{H}_{ni,\text{accept}}^0 + X_{ni}$
DFT process noise	$\sigma_{\varepsilon,ni} = \exp(X_{ni})$

Open Questions

Survey design with passive data **What should we infer, and what should we ask?**

Passive records R^P
Active reports R_d^A

Measurements / concepts

Behavioral outcomes

$$R \longrightarrow X \longrightarrow Y$$

LLM integration **Where should LLMs enter and what should they preserve?**

Passive records R^P
Active reports R_d^A

Measurements / concepts

Behavioral outcomes

$$R \xrightarrow[\text{LLM?}]{} X \xrightarrow[\text{LLM?}]{} Y$$

Take-home messages

1. Choose where flexibility belongs.

Let the network learn where flexibility is needed, and preserve the structure that interpretation requires: selected constraints on utility, a linear taste structure, RUM consistency.

2. Treat measurement as part of the model.

Say what X stands for, what information infers it (R alone, or R and Y), whether it is supervised, and where it enters the behavioral model.

3. Keep a distribution where the data cannot fix a point.

For tastes, let the amount of individual data set the shrinkage toward the population. For measurements, carry the uncertainty in X into Y instead of collapsing it to a point.

Design the whole $R \rightarrow X \rightarrow Y$ pipeline around the behavioral question!

Codes available

Paper	Model name	GitHub
Wang, Wang & Zhao (2020) TR-C	F-DNN	https://github.com/cjsyzwsh/dnn-for-economic-information
Wang, Mo & Zhao (2020) TR-C	ASU-DNN	https://github.com/cjsyzwsh/ASU-DNN
Sifringer, Lurkin & Alahi (2020) TR-B	L-MNL	https://github.com/BSifringer/EnhancedDCM
Han et al. (2022) TR-B	TasteNet	https://github.com/deborahmit/TasteNet-MNL
Wong & Farooq (2021) TR-C	ResLogit	https://github.com/LiTrans/reslogit-example
Feng et al. (2024) TR-C		https://github.com/siqi-feng/DNN-behavioral-regularity
Xia, Chen & Zimmermann (2023) TBS	RE-BNN	https://github.com/yutong-xia/RE-BNN
Li et al. (2026) TBS	BI-DNN（ICLV）	https://github.com/DeadliftData/BI-DNN-model
Bagheri, Ghasri & Barlow (2025) JOCM	RUM-NN	https://github.com/NioushaBagheri/RUM-NN
van Cranenburgh & Garrido-Valenzuela (2025) TR-A	CV-DCM	https://github.com/TUD-CityAI-Lab/Computer-vision-enriched-DCMs
Ito et al. (2026) Transportation		https://github.com/koito19960406/cycling_safety_perception
Ding, Yu & Bansal (2025) TR-D	DFT	https://github.com/jiaxuan-ding/uncertainty

References

- Agrawal, S., & Peeta, S. (2021). Hybrid route choice model incorporating latent cognitive effects of real-time travel information using physiological data. Transportation Research Part F: Traffic Psychology and Behaviour, 81, 223–239. https://doi.org/10.1016/j.trf.2021.05.021
- Ali, A., Flaata, E. H., Tørset, T., Hess, S., & Choudhury, C. F. (2026). Investigating the impact of inferring trip purposes in a daily trip generation model. Transportation Research Part C: Emerging Technologies, 186, 105619. https://doi.org/10.1016/j.trc.2026.105619
- Aouad, A., & Désir, A. (2026). Representing random utility choice models with neural networks. Management Science, 72(8), 6686–6701. https://doi.org/10.1287/mnsc.2023.02189
- Bansal, P., Kim, E.-J., & Ozdemir, S. (2024). Discrete choice experiments with eye-tracking: How far we have come and ways forward. Journal of Choice Modelling, 51, 100478. https://doi.org/10.1016/j.jocm.2024.100478
- Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., Morikawa, T., & Polydoropoulou, A. (2002). Integration of choice and latent variable models. In H. S. Mahmassani (Ed.), In Perpetual Motion: Travel Behavior Research Opportunities and Application Challenges (pp. 431–470). Pergamon.
- Ding, J., Yu, J., & Bansal, P. (2025). Preferences for electric vehicles under uncertain charging prices: An eye-tracking study. Transportation Research Part D: Transport and Environment, 140, 104608. https://doi.org/10.1016/j.trd.2025.104608
- Feng, S., Yao, R., Hess, S., Daziano, R. A., Brathwaite, T., Walker, J., & Wang, S. (2024). Deep neural networks for choice analysis: Enhancing behavioral regularity with gradient regularization. Transportation Research Part C: Emerging Technologies, 166, 104767. https://doi.org/10.1016/j.trc.2024.104767
- 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. https://doi.org/10.1016/j.trb.2022.07.001
- Hancock, T. O., Hess, S., & Choudhury, C. F. (2025). Beyond utility: Incorporating eye-tracking, skin conductance and heart rate data into cognitive and econometric travel behaviour models. arXiv preprint arXiv:2506.18068.
- Ito, K., Terra, R., Costa, M., Siebert, F. W., Garrido-Valenzuela, F., Lima Azevedo, C., Biljecki, F., & van Cranenburgh, S. (2026). Perceived traffic safety's impact on cycling route choice with street-level images. Transportation. https://doi.org/10.1007/s11116-026-10793-z
- Kim, E.-J., & Bansal, P. (2024). A new flexible and partially monotonic discrete choice model. Transportation Research Part B: Methodological, 183, 102947. https://doi.org/10.1016/j.trb.2024.102947
- Koh, P. W., Nguyen, T., Tang, Y. S., Mussmann, S., Pierson, E., Kim, B., & Liang, P. (2020). Concept bottleneck models. In Proceedings of the 37th International Conference on Machine Learning (PMLR 119, pp. 5338–5348).
- Li, X., Feng, T., Rasouli, S., Jia, P., & Kuang, H. (2026). Decomposing random taste heterogeneity in discrete choice modeling: A mixture density network-embedded logit model. Transportation Research Part E: Logistics and Transportation Review, 212, 104940. https://doi.org/10.1016/j.tre.2026.104940

References

- Li, X., Feng, T., Rasouli, S., Wong, M., & Kuang, H. (2026). A behaviorally-informed deep neural network for latent variable choice modeling. Travel Behaviour and Society, 45, 101332. https://doi.org/10.1016/j.tbs.2026.101332
- Maksimenko, V., Yuan, L., Kazemzadeh, K., & Bansal, P. (2026). Using brain response to measure pedestrian safety perception: Evidence from video and VR experiments. Safety Science, 199, 107202. https://doi.org/10.1016/j.ssci.2026.107202
- Nishi, T., Hara, Y. (2026) Discrete Choice Model with Generalized Additive Utility Network for Interpretable Policy Evaluation. Transportation Research Record, 03611981261461509.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x
- Sifringer, B., Lurkin, V., & Alahi, A. (2020). Enhancing discrete choice models with representation learning. Transportation Research Part B: Methodological, 140, 236–261. https://doi.org/10.1016/j.trb.2020.08.006
- van Cranenburgh, S., & Garrido-Valenzuela, F. (2025). Computer vision-enriched discrete choice models, with an application to residential location choice. Transportation Research Part A: Policy and Practice, 192, 104300. https://doi.org/10.1016/j.tra.2024.104300
- Wang, S., Mo, B., & Zhao, J. (2020). Deep neural networks for choice analysis: Architecture design with alternative-specific utility functions. Transportation Research Part C: Emerging Technologies, 112, 234–251. https://doi.org/10.1016/j.trc.2020.01.012
- 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. https://doi.org/10.1016/j.trb.2021.03.002
- Wang, S., Wang, Q., Bailey, N., & Zhao, J. (2021). Deep neural networks for choice analysis: A statistical learning theory perspective. Transportation Research Part B: Methodological, 148, 60–81. https://doi.org/10.1016/j.trb.2021.03.011
- 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, 118, 102701. https://doi.org/10.1016/j.trc.2020.102701
- Wong, M., & Farooq, B. (2021). ResLogit: A residual neural network logit model for data-driven choice modelling. Transportation Research Part C: Emerging Technologies, 126, 103050. https://doi.org/10.1016/j.trc.2021.103050
- Xia, Y., Chen, H., & Zimmermann, R. (2023). A random effect Bayesian neural network (RE-BNN) for travel mode choice analysis across multiple regions. Travel Behaviour and Society, 30, 118–134. https://doi.org/10.1016/j.tbs.2022.08.011
- Yu, J., Zhao, J., Miranda-Moreno, L., & Korp, M. (2025). Modular AI agents for transportation surveys and interviews: Advancing engagement, transparency, and cost efficiency. Communications in Transportation Research, 5, 100172. https://doi.org/10.1016/j.commtr.2025.100172
- Zhang, K., Song, Q., Ma, H., Qiu, W., Li, M., & Kim, I. (2025). Synergistic role of audio-visual perceptions in promoting bikeshare for active travel. Transportation Research Part D: Transport and Environment, 145, 104806. https://doi.org/10.1016/j.trd.2025.104806

THANK YOU

Makoto Chikaraishi, Hiroshima University
email: chikaraishim@hiroshima-u.ac.jp