Multiattribute Q-Learning: Identification, Economic Interpretation and Behavioral Testing in a Route Choice Experiment

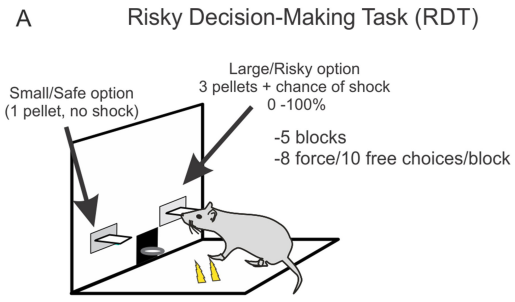
Bastian Henriquez-Jara
Universidad de Chile

with Camila Balbontin (PUC) and Omar D. Perez (U. Chile)

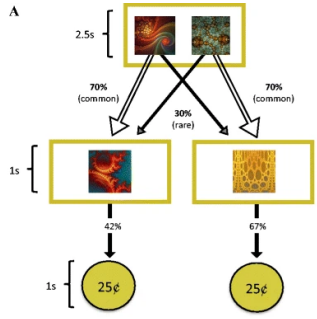
EPFL Jul 2026

UNIVERSIDAD DE CHILE

Learning problems in psychology



Animal learning experiment example (Chiavegatti and Floresco, 2024)



Human learning experiment example (Gillan et al., 2015)

Q-Learning

Standard Q-Learning (Sutton and Barto, 2018)

Reward prediction error (RPE):

$$\delta_{nit} = R_{nit} - Q_{nit}$$

Q-value update:

$$Q_{nit} = Q_{n,i,t-1} + \alpha\, \delta_{n,i,t-1}$$

Choice probability (logit):

$$P(y_{nit} = 1) = \frac{\exp(\mu\, Q_{nit})}{\sum_j \exp(\mu\, Q_{ntj})}$$

Limitation: R is a scalar — no attribute decomposition (no preference parameters!).

Q-Learning

Q-Learning (Sutton and Barto, 2018) is a well-stablished RL model in psychology & neuroscience, previously used in transport (e.g., Henriquez-Jara et al., 2025), but was designed for laboratory settings with:

- Unidimensional rewards (food, shock, binary outcomes)
- No subjective taste parameters — reward value assumed known to modeller
- No economic interpretation (no WTP, no MRS, no welfare)
- Complete feedback sequences — not typical field/survey data

Multiattribute Q-Learning (MAQL)

Multiattribute reward

$$R_{nit} = \sum_k \beta_k \, x_{nikt}$$

x_{nikt}: attribute k of alternative i at time t; β_k: satisfaction parameter. See Schultner et al. (2025) and Ng, Russell, et al. (2000) for other multi-attribute RL models.

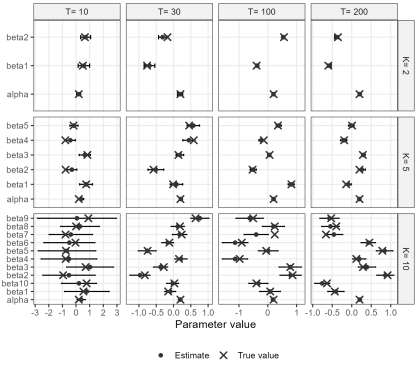
Q-value update

$$Q_{nit} = Q_{n,i,t-1} + \alpha \left(\sum_k \beta_k x_{nik,t-1} - Q_{n,i,t-1} \right)$$

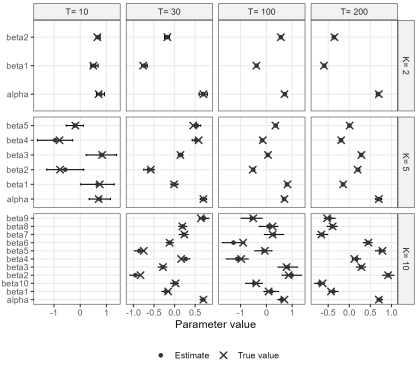
$$P(y_{nit} = 1) = \frac{\exp(\mu \, Q_{nit})}{\sum_j \exp(\mu \, Q_{ntj})} \text{ now with fixed } \mu = 1.$$

Identification: Monte Carlo Simulations

100 individuals · $K \in \{2, 5, 10\}$ attributes · $T \in \{10, 30, 100, 200\}$ trials

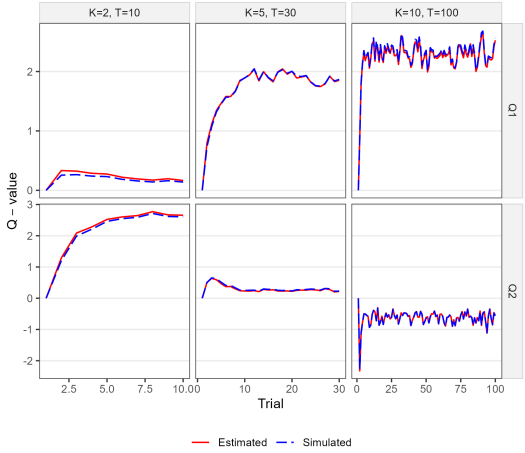


$\alpha = 0.2$



$\alpha = 0.8$

Identification: Monte Carlo Simulations



Economic interpretation

Marginal effect on next period's utility

$$Q_{nit} = Q_{n,i,t-1} + \alpha\left(\sum_k \beta_k x_{nik,t-1} - Q_{n,i,t-1}\right) \rightarrow \frac{\partial Q_{nit}}{\partial x_{nik,t-1}} = \alpha\beta_k$$

Immediate marginal rate of substitution

$$MRS_i(k,r) = \frac{\frac{\partial Q_{nit}}{\partial x_{nik,t-1}}}{\frac{\partial Q_{nit}}{\partial x_{nir,t-1}}} = \frac{\alpha\beta_k}{\alpha\beta_r} = \frac{\beta_k}{\beta_r}.$$

$$
\begin{aligned}
Q_{nit} &= Q_{n,i,t-1} + \alpha(R_{n,i,t-1} - Q_{n,i,t-1}) \\
&= Q_{n,i,t-1}(1-\alpha) + \alpha R_{n,i,t-1},
\end{aligned}
$$

$$\vdots$$

$$
Q_{nit} = \overbrace{(1-\alpha)^t Q_{ni0}}^{\text{Residual value of starting Q-value}} + \overbrace{\alpha \sum_{t'=1}^{t} (1-\alpha)^{t-t'} R_{nit'-1} y_{ni,t'-1}}^{\text{Discounted accumulated reward of alternative i}}.
$$

Economic interpretation

Overall marginal effects

Assuming each attribute k follows a generic distribution $f(\mu_k, \sigma_k)$, where μ_k and σ_k denote its expected value and standard deviation, respectively:

$$E(Q_{nit} \mid \mu) = (1-\alpha)^t Q_{ni0} + \alpha \sum_{t'=1}^{t} (1-\alpha)^{t-t'} \left(\sum_k \beta_k \mu_k \right) y_{ni,t'-1}$$

Economic interpretation

Overall marginal effects

Assuming each attribute k follows a generic distribution $f(\mu_k, \sigma_k)$, where μ_k and σ_k denote its expected value and standard deviation, respectively:

$$E(Q_{nit} \mid \mu) = (1-\alpha)^t Q_{ni0} + \alpha \sum_{t'=1}^{t} (1-\alpha)^{t-t'} \left(\sum_k \beta_k \mu_k \right) y_{ni,t'-1}$$

Rearranging,

$$E(Q_{nit} \mid \mu) = (1-\alpha)^t Q_{ni0} + \alpha \left(\sum_k \beta_k \mu_k \right) \sum_{t'=1}^{t} (1-\alpha)^{t-t'} y_{ni,t'-1}.$$

Economic interpretation

Then, a marginal variation in the expected value of attribute k produces a marginal variation in the expected Q-value given by:

$$\frac{\partial E(Q_{nit} \mid \mu)}{\partial \mu_k} = \alpha\beta_k \sum_{t'=1}^{t} (1-\alpha)^{t-t'} y_{ni,t'-1}.$$

Economic interpretation

Then, a marginal variation in the expected value of attribute k produces a marginal variation in the expected Q-value given by:

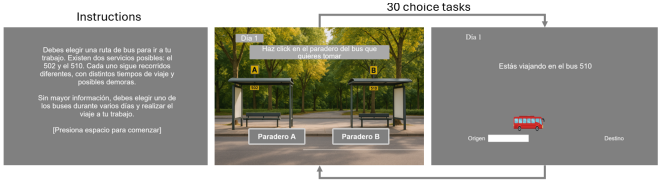
$$\frac{\partial E(Q_{nit} \mid \mu)}{\partial \mu_k} = \alpha \beta_k \sum_{t'=1}^{t} (1-\alpha)^{t-t'} y_{ni,t'-1}.$$

Overall marginal rate of substitution

The MRS between the expected values of two attributes, μ_k and μ_r, is

$$MRS_i(k,r) = \frac{\frac{\partial E(Q_{nit}|\mu)}{\partial \mu_k}}{\frac{\partial E(Q_{nit}|\mu)}{\partial \mu_r}} = \frac{\alpha \beta_k \sum_{t'=1}^{t} (1-\alpha)^{t-t'} y_{ni,t'-1}}{\alpha \beta_r \sum_{t'=1}^{t} (1-\alpha)^{t-t'} y_{ni,t'-1}} = \frac{\beta_k}{\beta_r}.$$

Case Study: Route Choice Under Travel Time Uncertainty



Experiment

- $N = 473$ individuals
- $T = 30$ choice tasks
- Chile, Nov. 2025
- Alt. A: fast & risky
  $\bar{t} = 28$ min, $\sigma = 10.5$
- Alt. B: slow & safe
  $\bar{t} = 30$ min, $\sigma = 2.9$

Case Study: Route Choice Under Travel Time Uncertainty

Reward function:

$$R_{nit} = \beta_i + \beta_{tt} \cdot tt_{nit} + \left(\beta_{var^+} + \beta_{var^-} \cdot \mathbf{1}_{\Delta tt_{nit}<0}\right) \cdot \Delta tt_{nit}$$

where $\Delta tt_{nit} = tt_{nit} - tt_{ni,\tau(n,i,t)}$ (change vs. last experience of alt. i).

Three Model Specifications

MAQL — static population-level α

$$\alpha = \frac{\exp(\alpha_0)}{1+\exp(\alpha_0)}$$

δ-MAQL — α depends on sign of RPE

$$\hat{\alpha}_{nit} = \alpha_0 + \alpha^- \cdot \mathbf{1}_{\delta_{n,i,t-1}<0}, \qquad \alpha_{nit} = \frac{\exp(\hat{\alpha}_{nit})}{1+\exp(\hat{\alpha}_{nit})}$$

Mix-MAQL — random individual learning rate

$$\hat{\hat{\alpha}}_n = \alpha_0 + \varepsilon_n, \quad \varepsilon_n \sim N(0, \sigma_\alpha), \qquad \alpha_n = \frac{\exp(\hat{\hat{\alpha}}_n)}{1+\exp(\hat{\hat{\alpha}}_n)}$$

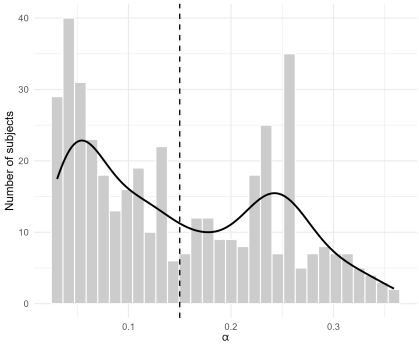
All models share the same Q-value update and reward function. Scale fixed at $\mu = 1$.

Estimation Results

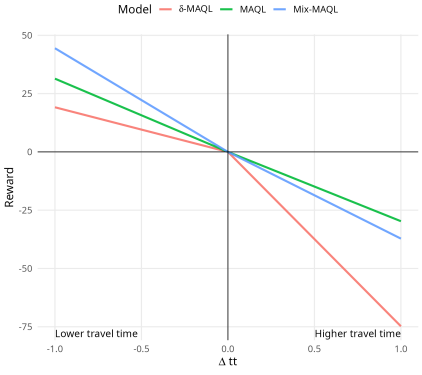
	MAQL		δ-MAQL		Mix-MAQL	
Parameter	Est.	t	Est.	t	Est.	t
α_0	−1.879	−16.3	−1.460	−22.0	−3.914	−17.3
α^-	—	—	−1.400	−58.1	—	—
Q_{01}	0	—	0	—	0	—
Q_{02}	0.166	3.66	0.082	1.46	−0.084	−1.58
β_{alt1}	7.447	9.64	6.378	31.7	11.532	7.58
β_{alt2}	8.177	11.0	7.357	110.	12.699	8.29
β_{tt}	−11.27	−8.28	−9.721	−195.	−19.88	−7.08
β_{var^+}	−29.73	−8.04	−74.80	−72.7	−37.20	−5.22
β_{var-}	−1.679	−0.87	55.66	36.1	−7.258	−2.21
σ_α	—	—	—	—	3.614	11.9
N; obs.	417; 12510		417; 12510		417; 12510	
LL(final)	−6470.1		−6446.5		−6321.7	
$\bar{\rho}^2$	0.253		0.256		0.270	

All t-ratios relative to $H_0 = 0$. μ fixed at 1 for identification. Mix-MAQL estimated with 500 Halton draws.

Learning Rate Heterogeneity



Posterior α_n distribution (Mix-MAQL): bimodal, right-skewed; mean ≈ 0.14



Δtt effect on reward by model

Key findings

- Population mean $\bar{\alpha} = 0.14$ masks bimodal heterogeneity (Mix-MAQL)
- δ-MAQL: $\alpha(\delta > 0) = 0.19 > \alpha(\delta < 0) = 0.05$ — faster updating for positive surprises
- $\beta_{tt} < 0$ and $\beta_{var^+} < 0$ consistently across all models

Conclusions and next steps

- The Q-learning model can be extended to capture multiattribute rewards.
- The Multiattribute Q-Learning (MAQL) model is interpretable, with marginal rates of substitution (MRS) depending only on the reward parameters.
- The model can be further extended to account for inter-individual heterogeneity.

Next steps

- Can we make simplification assumptions to avoid estimating a dynamic model?
- Can the Q-learning model be estimated even when the learning process is only partially observed (missing data)?
- Estimate the model with real life data

References I

Chiavegatti, G. L. and S. B. Floresco (2024). "Acute stress differentially alters reward-related decision making and inhibitory control under threat of punishment". In: *Neurobiology of Stress* 30, p. 100633.

Gillan, C. M., A. R. Otto, E. A. Phelps, and N. D. Daw (2015). "Model-based learning protects against forming habits". In: *Cognitive, Affective, & Behavioral Neuroscience* 15.3, pp. 523–536.

Henriquez-Jara, B., C. A. Guevara, M. Munizaga, and O. D. Perez (2025). "Habits and the subexploration of better transportation options: A dual-system approach". In: *Travel Behaviour and Society* 38, p. 100877.

Ng, A. Y., S. Russell, et al. (2000). "Algorithms for inverse reinforcement learning.". In: *Icml*. Vol. 1. 2, p. 2.

References II

Schultner, D., L. Molleman, and B. Lindström (2025). "Feature-based reward learning shapes human social learning strategies". In: *Nature Human Behaviour* 9.10, pp. 2183–2198.

Sutton, R. S. and A. G. Barto (2018). *Reinforcement learning: An introduction*. MIT press.