Human Behavior Models for Virtual Agents in Repeated Decision Making under Uncertainty

Ming Yin (Harvard) Yu-An Sun (PARC)

AAMAS-15, Istanbul, Turkey, May 4-8, 2015

Repeated Decision Making under Uncertainty

Horse-racing gambling

Financial investment

Choosing tariff schemes

The Environment Learning Problem

- N random variables, M options
- Each random variable X_i follows a stationary distribution.
- At the beginning of each period t, the DM chooses an option $Y_t = j$.
- At the end of each period t, the DM observes **all** x_i^t and obtains a utility of $U_t = f_j(x_1^t, x_2^t, \ldots, x_N^t)$
- Goal: Maximize $U_1 + U_2 + \ldots + U_T$

Research Questions

- Can we quantitatively model the actual human behavior in an environment learning problem?
- Does there exist a robust model to describe an average DM's behavior in various environments?
- How is the heterogeneity among individual DMs influenced by the environment?

Human Subject Experiment

Recruit MTurk workers to choose electricity tariff schemes repeatedly.

Today is March 1, 2015

Below is your electricity usage for February 2015:

You chose Option B for February 2015			
	Usage (kWh)	Unit Price ($)	Subtotal ($)
Day	101	0.2	20.20
Night	64	0.3	19.20
Total: 39.40			

Your current account balance is $918.50.

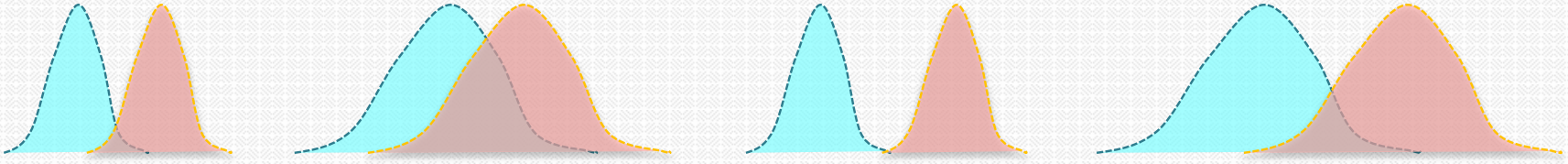
Now, please choose a tariff scheme for March 2015:

- A. Flat-rate: $0.25/kWh for electricity usage at any time throughtout the day.
- B. Cheaper in the day: $0.20/kWh for daytime usage; $0.30/kWh for night time usage.
- C. Cheaper in the night: $0.30/kWh for daytime usage; $0.20/kWh for night time usage.

- Review electricity bills of the previous month
- Choose a tariff scheme for the current month

Experiment Treatments

4 electricity usage conditions



Small Δμ
Small σ

Small Δμ
Large σ

Large Δμ
Small σ

Large Δμ
Large σ

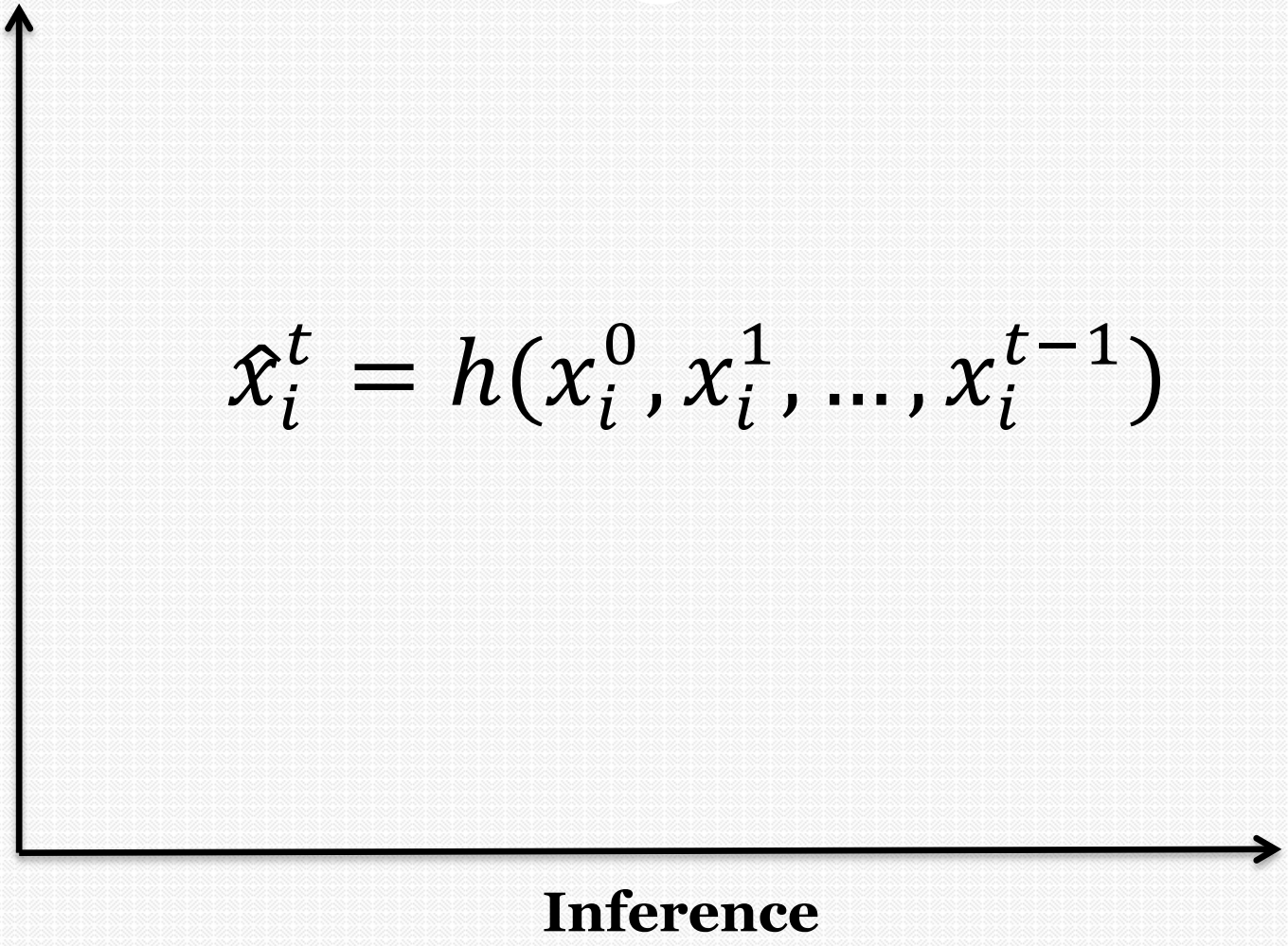
2 default option conditions: with/without

8 treatments

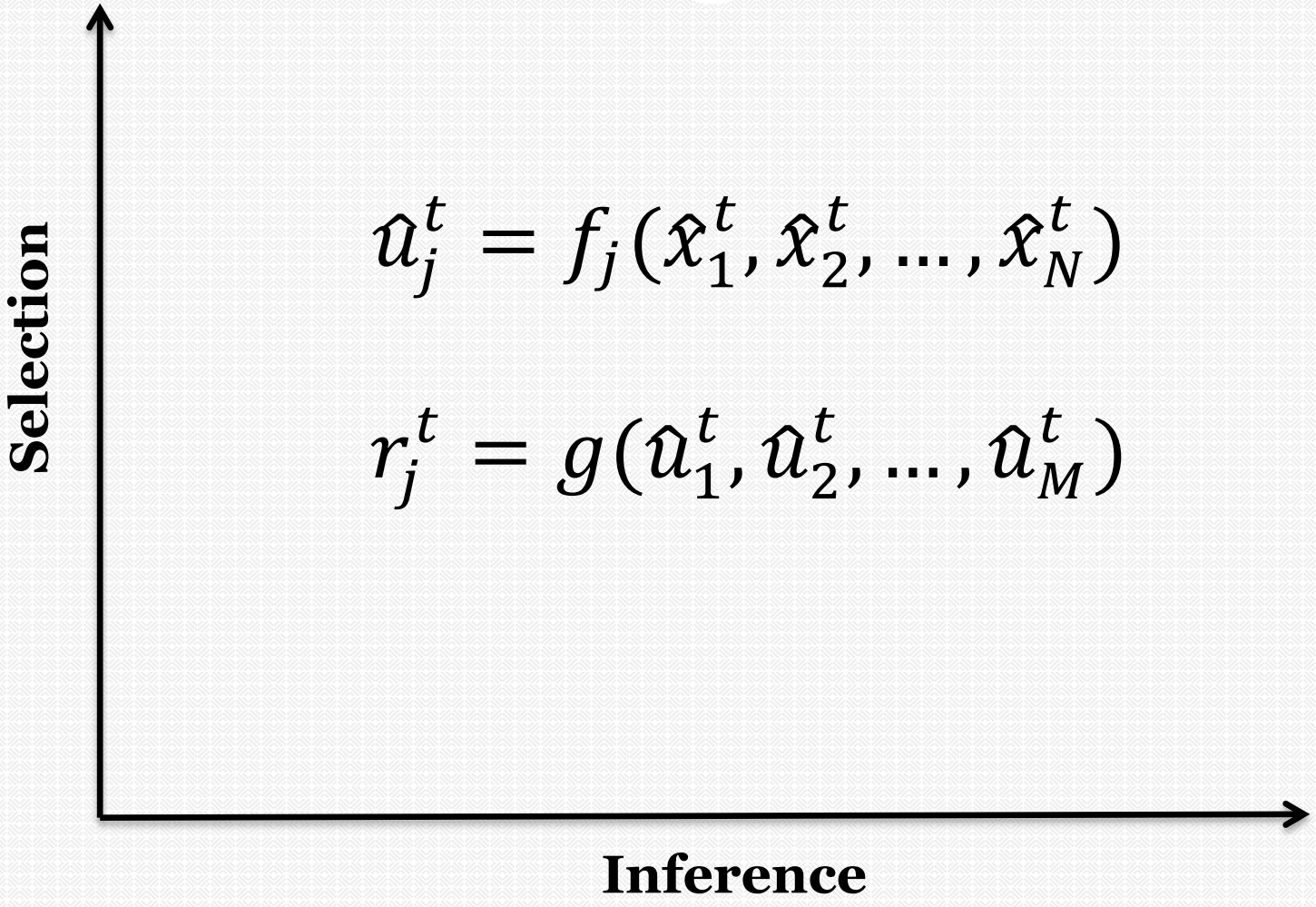
800 workers in total, **100** workers per treatment

Each worker makes **24** choices in a row

Two-Component Models



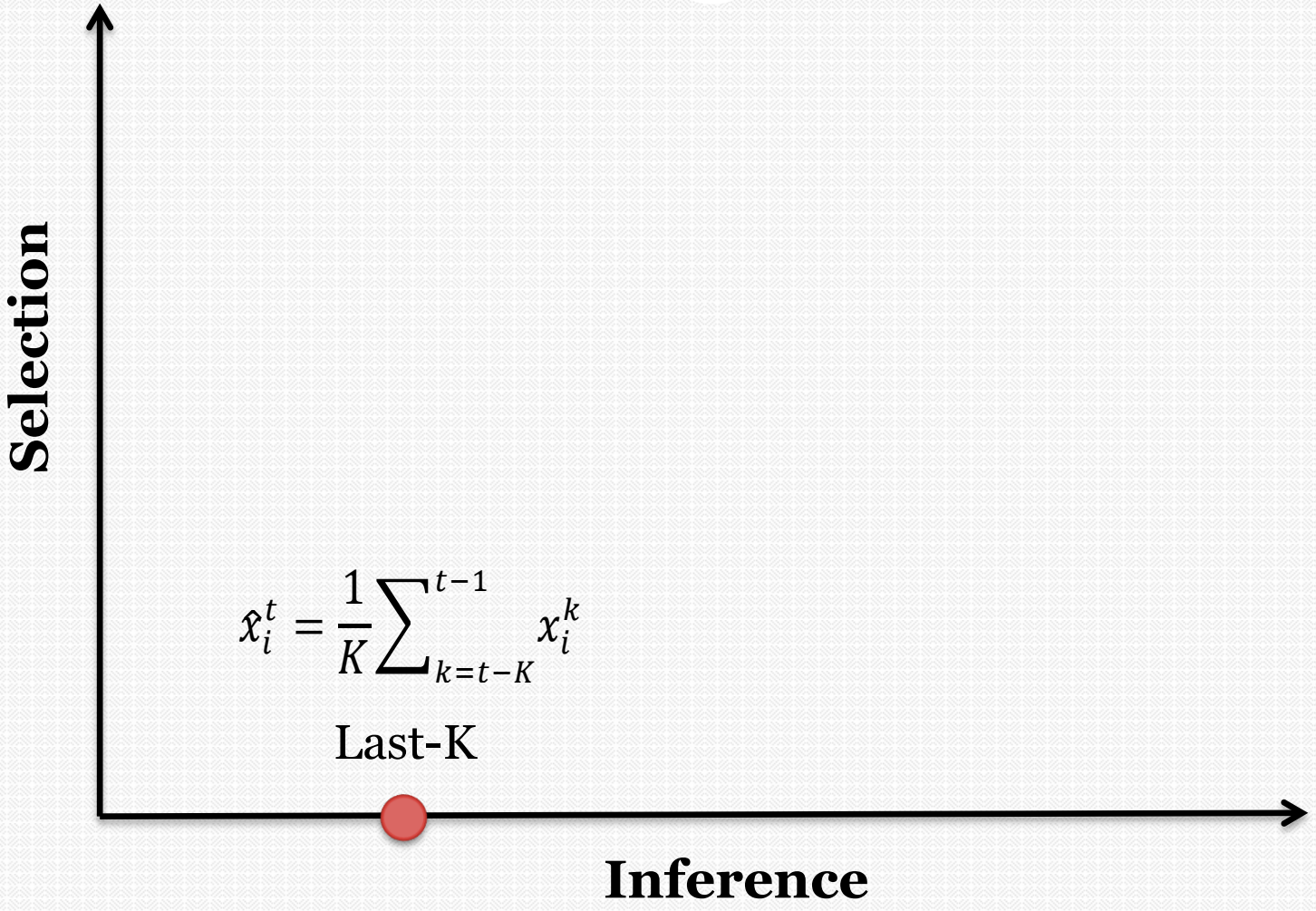
Two-Component Models



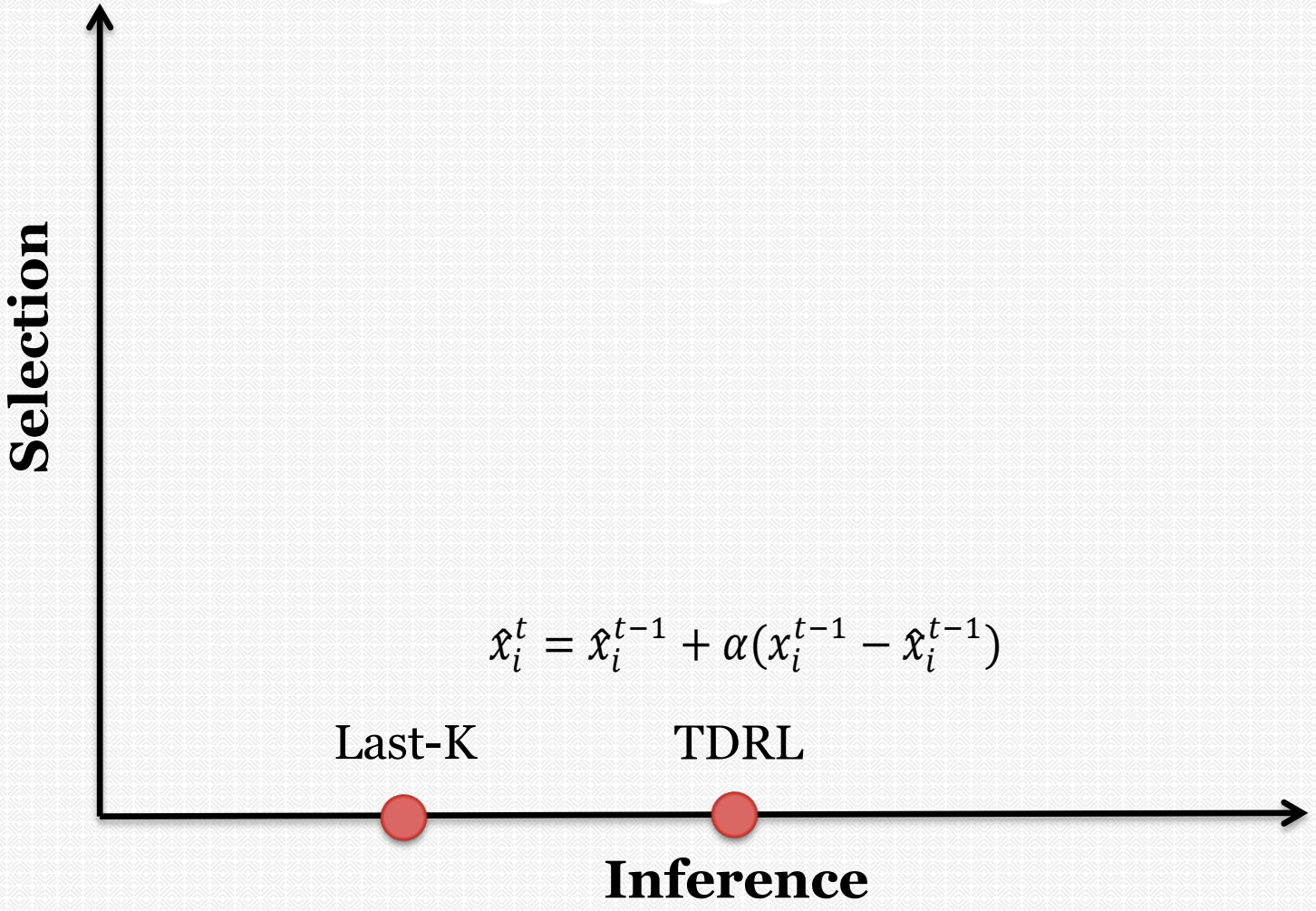
$$\hat{a}_j^t = f_j(\hat{x}_1^t, \hat{x}_2^t, \ldots, \hat{x}_N^t)$$

$$r_j^t = g(\hat{a}_1^t, \hat{a}_2^t, \ldots, \hat{a}_M^t)$$

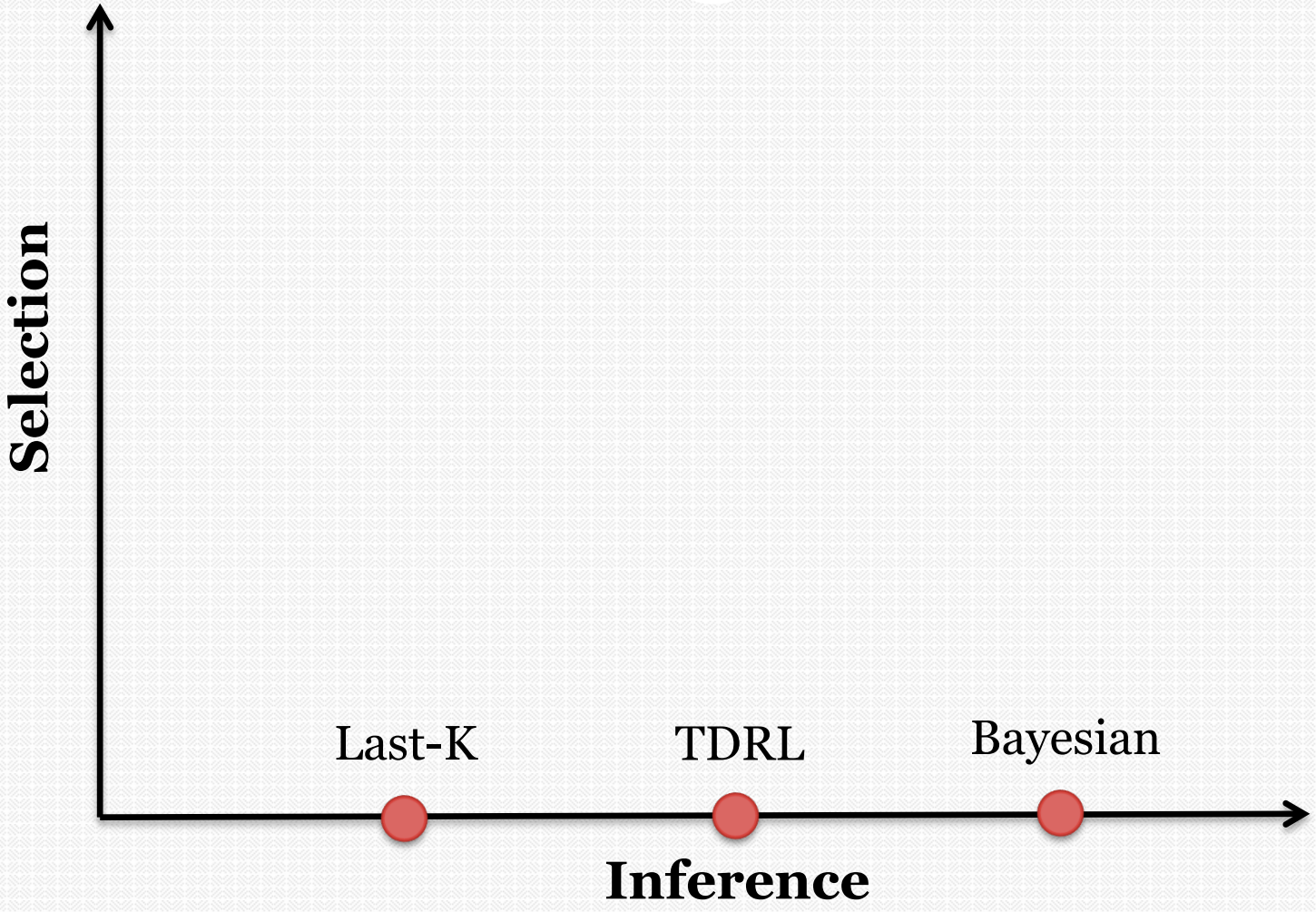
Two-Component Models



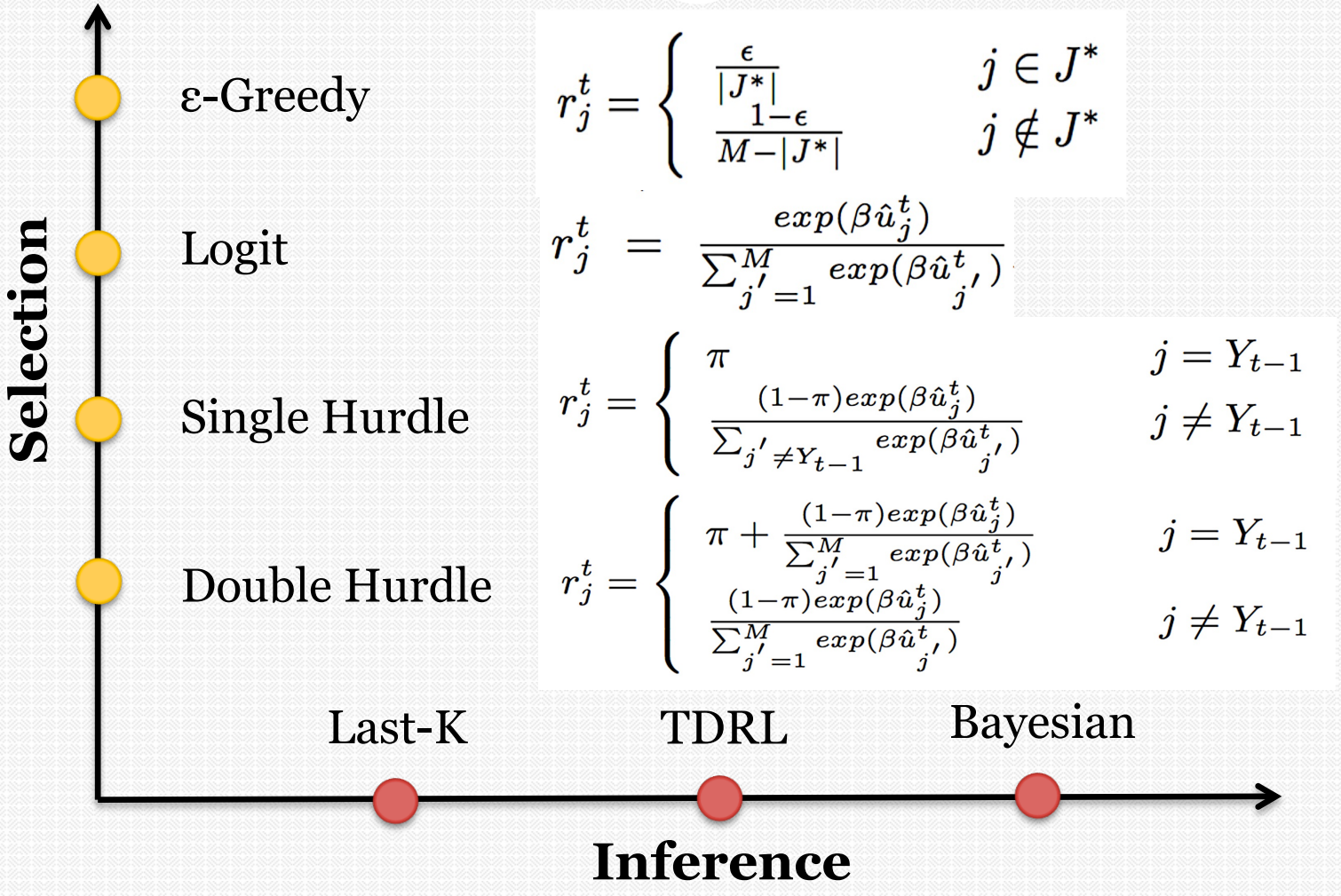
Two-Component Models



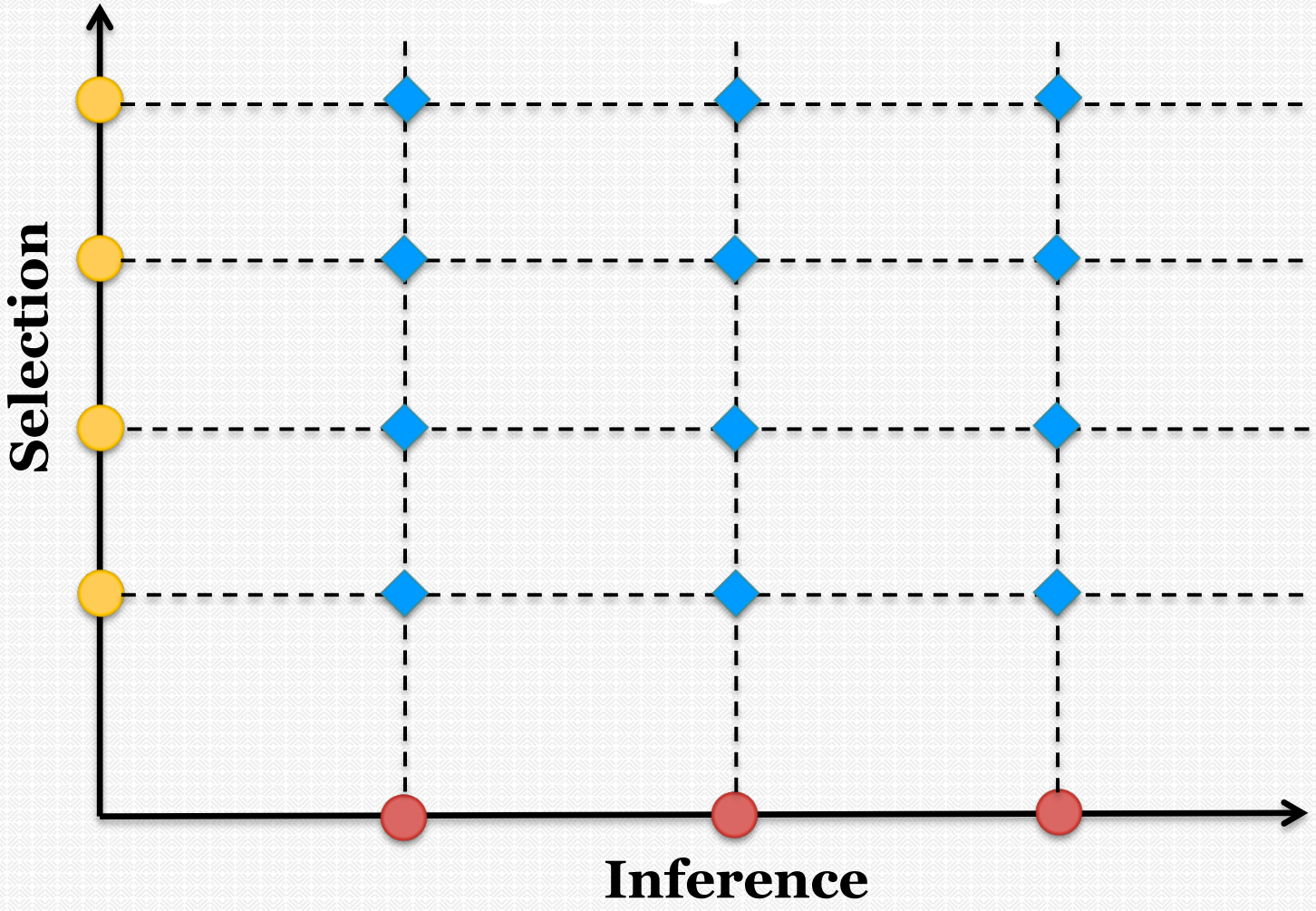
Two-Component Models



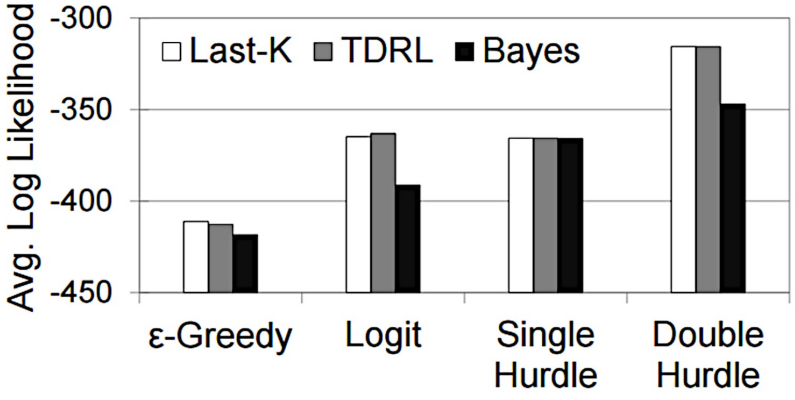
Two-Component Models



Two-Component Models



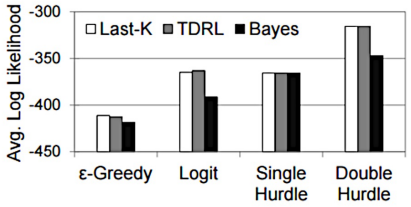
The Best Two-Component Model



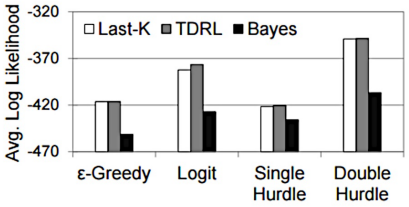
(a) T1: Small difference, small variance, no default

TDRL (or the best Last-K) + Double Hurdle best captures the average human DM's behavior

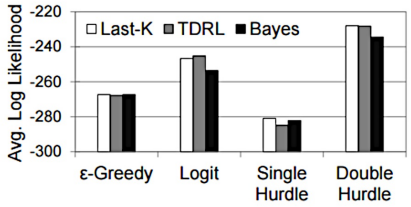
The Best Two-Component Model



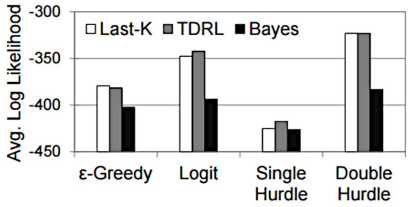
(a) T1: Small difference, small variance, no default



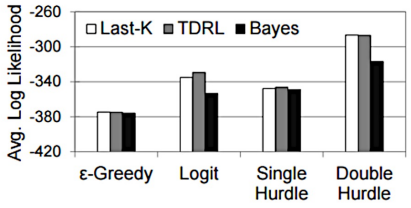
(b) T2: Small difference, large variance, no default



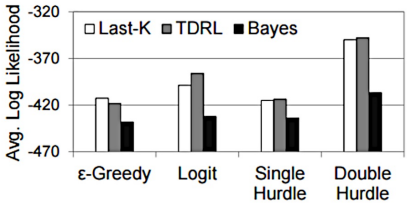
(c) T3: Large difference, small variance, no default



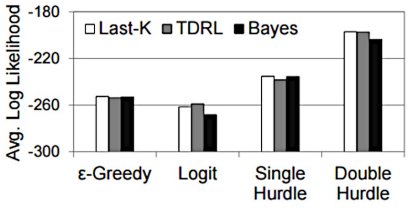
(d) T4: Large difference, large variance, no default



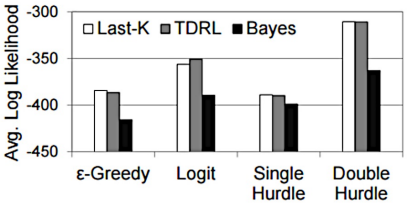
(e) T5: Small difference, small variance, default



(f) T6: Small difference, large variance, default



(g) T7: Large difference, small variance, default



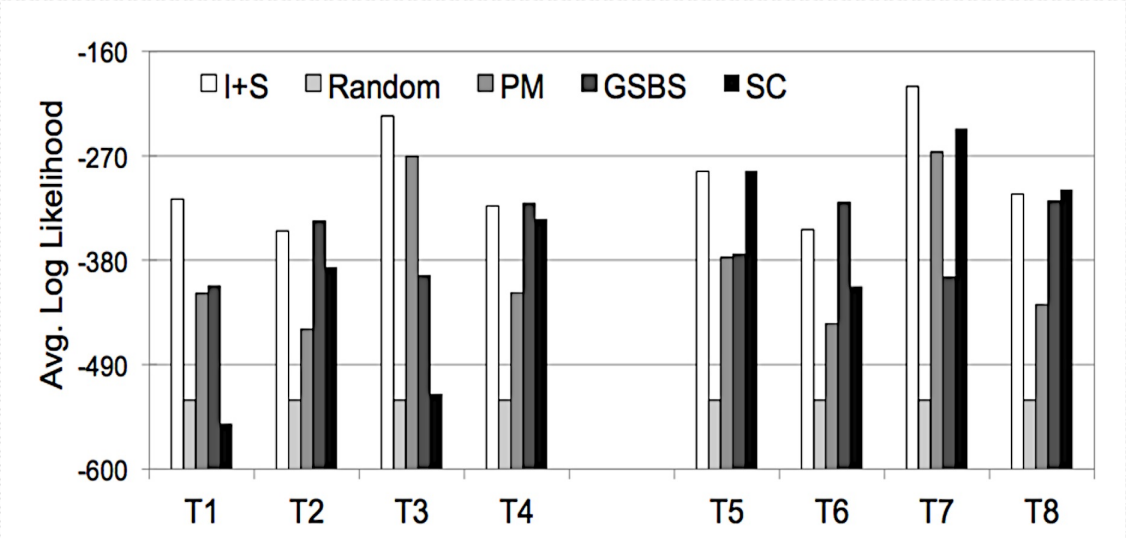
(h) T8: Large difference, large variance, default

TDRL (or the best Last-K) + Double Hurdle best captures the average human DM's behavior **in all environments**!

Human DMs display **recency bias** and **status-quo bias**!

Two-Component Model vs. Rules of Thumbs

- Random
- Probability matching
- Good-stay-bad-shift
- Safe choice



The activation of rules of thumbs is context-dependent!

TDRL + Double Hurdle is **robust** against various environments!

Summary

- We try to quantitatively model the actual human behavior in an environment learning problem in a principled manner.
- Our results show that an average DM's behavior can be robustly described by a two-component model (TDRL + Double Hurdle) across various environments.
- The average DM are also shown to be subject to recency bias and status-quo bias.