ICML
International Conference
On Machine Learning

Thompson Sampling for Robust Transfer in Multi-Task Bandits

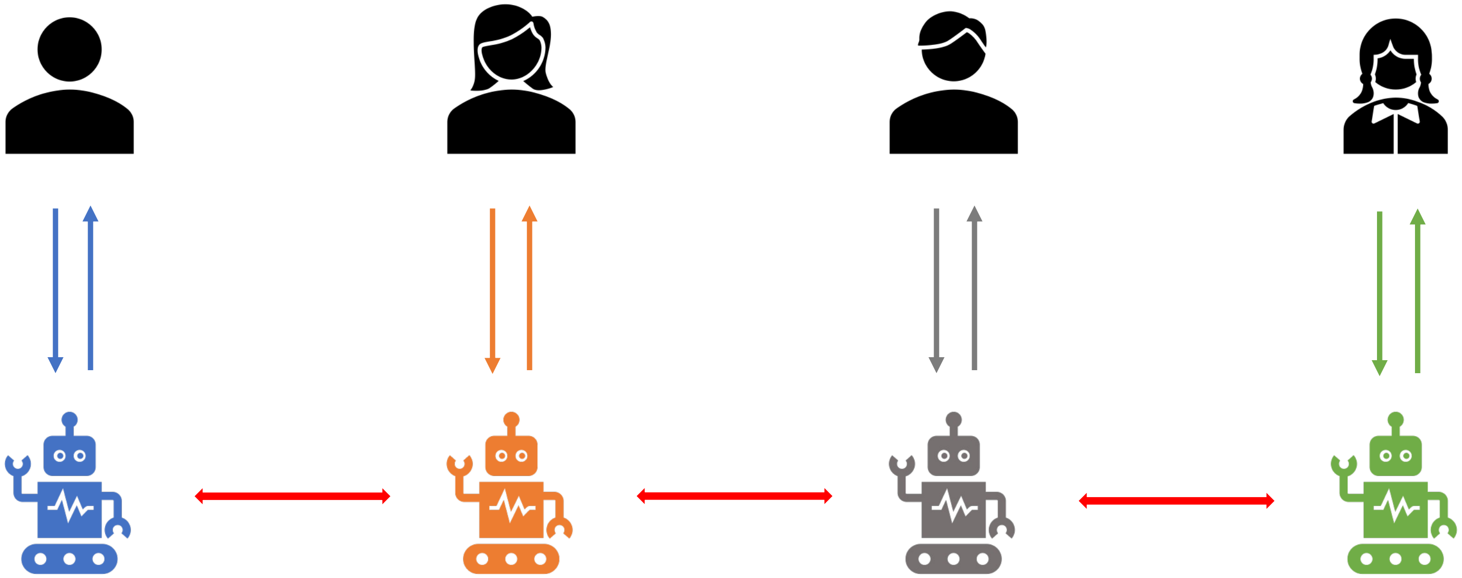
Zhi Wang[1], Chicheng Zhang[2], and Kamalika Chaudhuri[1]

[1] UC San Diego
JACOBS SCHOOL OF ENGINEERING

[2] THE UNIVERSITY OF ARIZONA

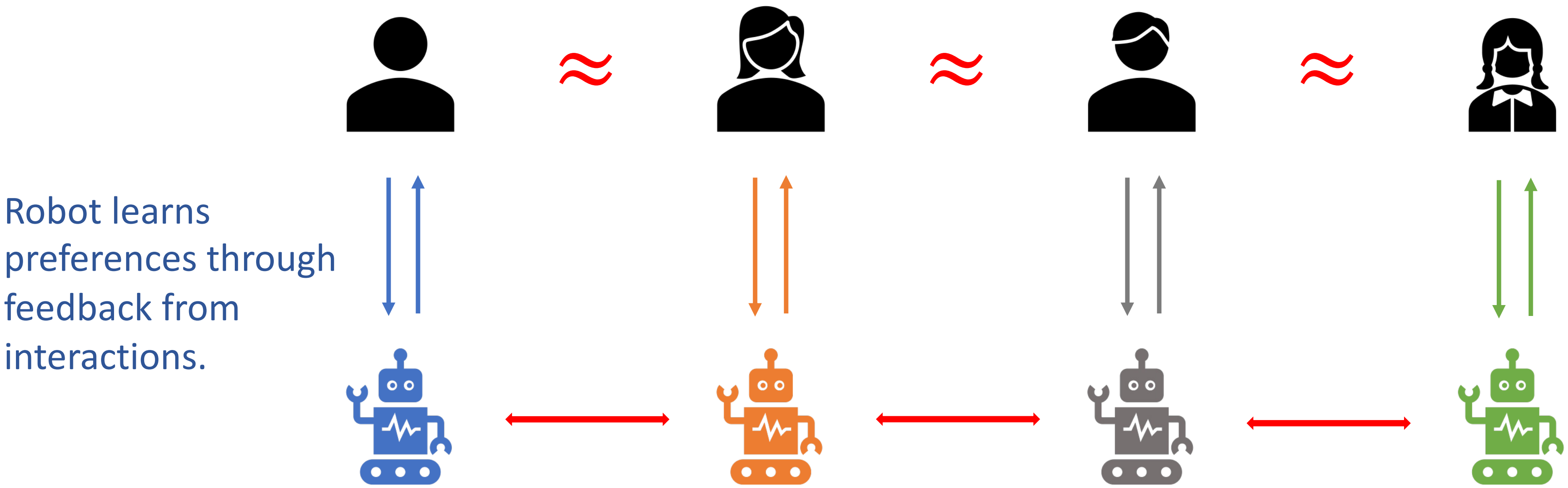
Transfer Learning in Multi-Task Bandits:
A Motivating Example (Wang et al., 2021)

Robot learns preferences through feedback from interactions.



- A group of assistive robots deployed to provide personalized healthcare services.

Transfer Learning in Multi-Task Bandits:
A Motivating Example (Wang et al., 2021)



- A group of assistive robots deployed to provide personalized healthcare services.
- Transfer learning: what can be done and what cannot when feedback is similar yet nonidentical?

The ε-MPMAB Problem (Wang et al., 2021)

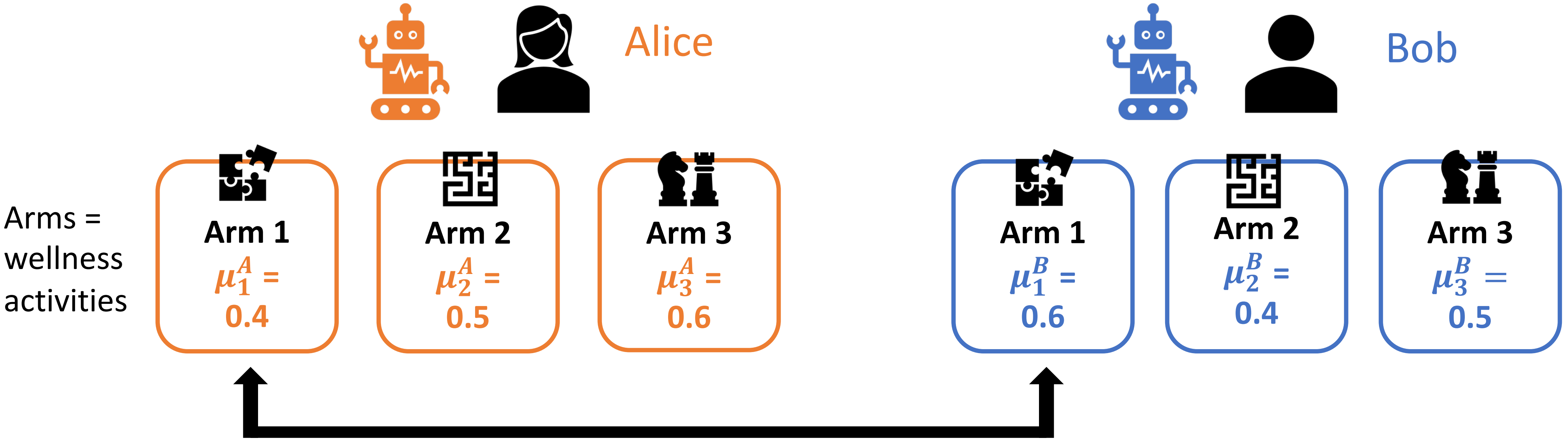
- A set of M players (robots) *interact* with K arms under a **generalized** protocol:
  - In each round t, a set of **active players** $\mathcal{P}_t$ is chosen and each pulls an arm (inspired by Hong et al., 2022).

The ε-MPMAB Problem (Wang et al., 2021)

- A set of M players (robots) *interact* with K arms under a **generalized** protocol:
  - In each round t, a set of **active players $\mathcal{P}_t$** is chosen and each pulls an arm (inspired by Hong et al., 2022).
  - When $\mathcal{P}_t = [M]$ → **concurrent** interaction (Wang et al., 2021)
  - When $|\mathcal{P}_t| = 1$ → **sequential** transfer (Azar et al., 2013; Cesa-Bianchi et al., 2013)

The ε-MPMAB Problem (Wang et al., 2021)

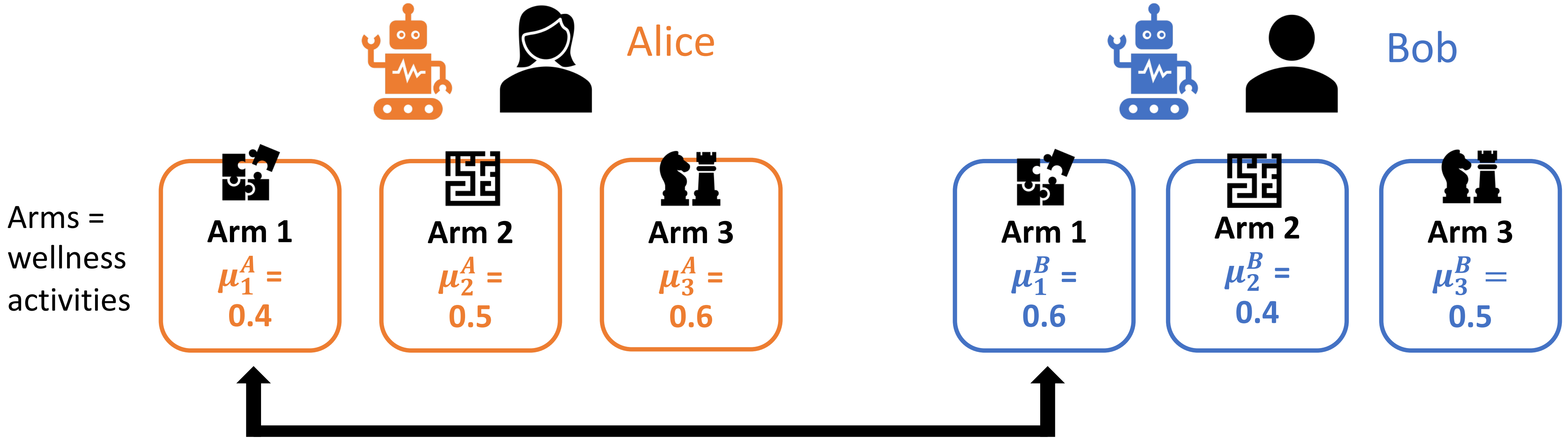
- A set of M players (robots) *interact* with K arms under a generalized protocol.



$\forall\, p, q \in [M], \max_{i \in [K]} \left|\mu_i^p - \mu_i^q\right| \leq \varepsilon \longrightarrow \varepsilon \in [0,1]$ dissimilarity parameter

The ε-MPMAB Problem (Wang et al., 2021)

- A set of M players (robots) *interact* with K arms under a generalized protocol.



$\forall\, p, q \in [M], \max_{i \in [K]} \left|\mu_i^p - \mu_i^q\right| \leq \varepsilon$ → $\varepsilon \in [0,1]$ dissimilarity parameter

- Goal: Minimize expected collective regret.

Known Results (Wang et al, 2021)

- When $\boldsymbol{\varepsilon}$ is *unknown*: not much can be done
- When $\boldsymbol{\varepsilon}$ is **known**:

Auxiliary data from transfer learning is **not** always helpful!

- Data aggregation is *only* provably beneficial on **$\mathcal{O}(\varepsilon)$-subpar arms**, defined as

$$\{i \in [K] : \; \exists p, \; \Delta_i^p > \mathcal{O}(\epsilon)\}.$$

Suboptimality gap
$\Delta_i^p = \max_{j \in [K]} \mu_j^p - \mu_i^p$

UCB-Based Algorithm (Wang et al, 2021)

- RobustAgg($\boldsymbol{\varepsilon}$):
  - UCB-based;
  - Near-optimal *gap-dependent* and gap-*independent* upper bounds on the collective regret;
  - Up to $\mathcal{O}(M)$ improvement for subpar arms compared with a UCB-based baseline without transfer.

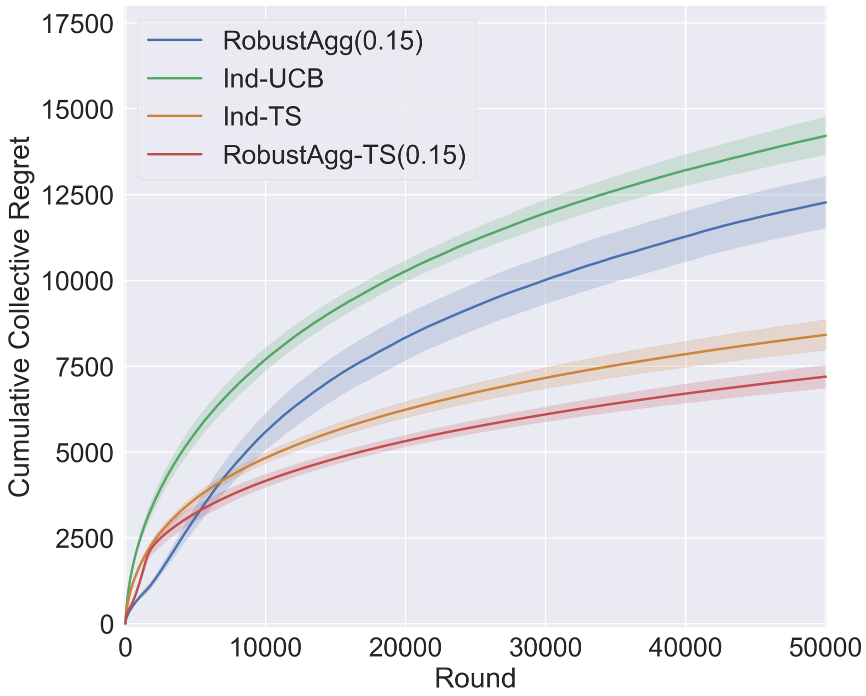
However, its empirical performance is **underwhelming**.

Thompson Sampling (TS)

- Superior empirically in comparison with UCB-based algorithms in standard single-task settings (Chapelle & Li, 2011).

- TS without transfer > RobustAgg($\boldsymbol{\varepsilon}$)

- Theoretical study of TS has lagged behind:
  - Frequentist analysis in multi-task setting



Our Contributions

- We design a TS-type algorithm, RobustAgg-TS(ε), that has *both*
  - Superior empirical performance, and
  - Strong, near-optimal theoretical guarantees.
- Balances bias-variance tradeoff
- Much harder to analyze
- Technical highlight:
  - A novel concentration inequality for multi-task data aggregation at random stopping times