Policy Fairness in Sequential Allocations under Bias Dynamics

Meirav Segal[1] , Anne-Marie George[1] , Christos Dimitrakakis[1,2]

[1]Department of Informatics, University of Oslo, [2]Computer Science Department, Université de Neuchâtel

Watch the video! →

Setting: Sequential Allocations

Allocation problems: allocating limited resources among a population to maximize an objective

Dynamic setting: sequential allocations with feedback, such that current decisions may change future population distribution in different ways

Past results: Static fairness constraints might lead to negative long-term effects

Use Case: College Admissions

Affirmative action: favoring individuals belonging to disadvantaged groups (e.g., by setting a lower acceptance threshold)

Possible feedback effects:

Positive: generate more role models (investment incentive)

Negative: reduce chances of graduation (reinforce stereotypes)

Model: Markov Decision Process (MDP)

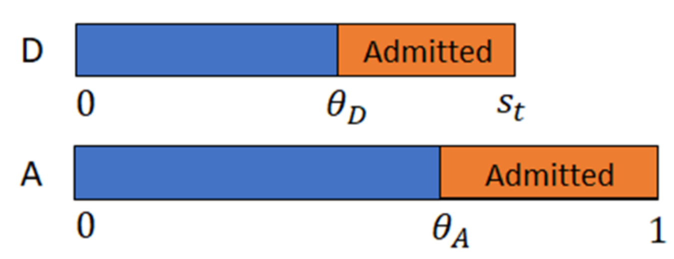
Population partitioned based on constant affiliation with constant group sizes

Advantaged group (A): success probability = ability

Disadvantaged group (D): success probability ≤ ability

State (s): current bias (upper bound of the success distribution of D)

Success distributions:



Decision maker (DM)

1. Takes action: sets admission thresholds θ_A, θ_D one for each group, admits a constant fraction of the population
2. Receives reward (R): fraction of successful students

Goal: maximizing utility (U, discounted sum of rewards) $U = \sum_t \gamma^t R_t$

Transition function: bias dynamics

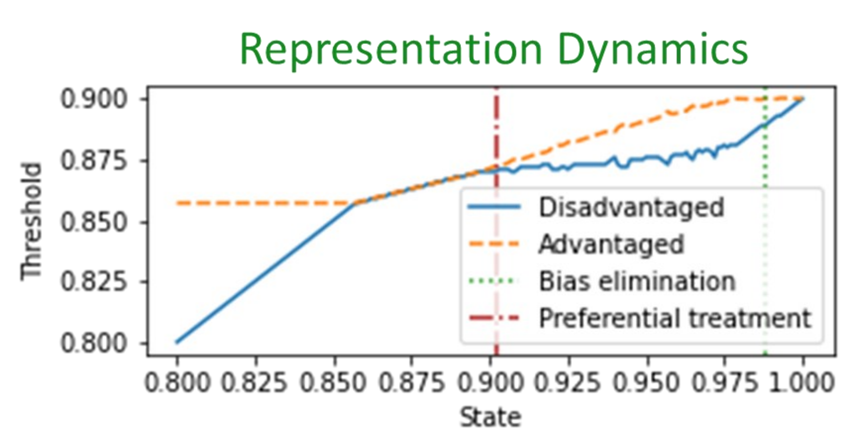
- **Representation dynamics** - fraction of selected students from D
- **Relative success dynamics** - success probability of admitted students from D

Utility Maximizing Policy

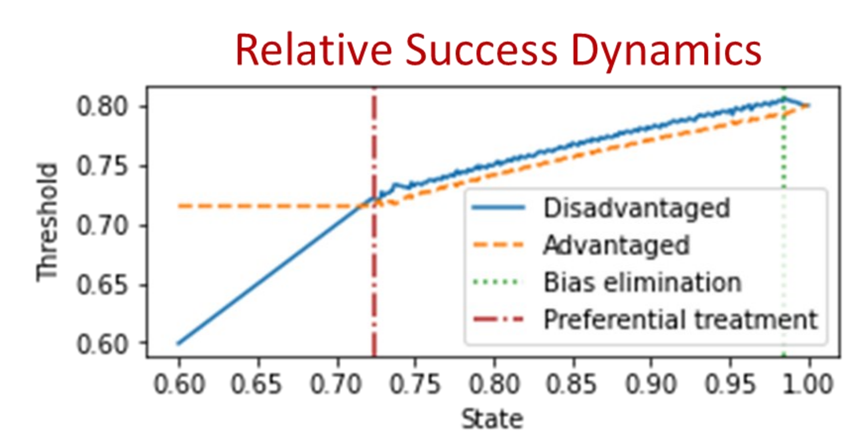
Optimization: Policy Iteration (infinite horizon, discrete state space)

No fairness constraints

- Both policies apply preferential treatment (PT), but in different ways
- **The effect of affirmative action on long-term fairness strongly depends on the dynamics**



PT = lower threshold for disadvantaged



PT = higher threshold for disadvantaged

Policy Fairness

$$F^\pi = \sum_t \gamma^t f(s_t)$$

Fairness of a policy: Discounted sum of state fairness

State fairness: The corresponding bias factor

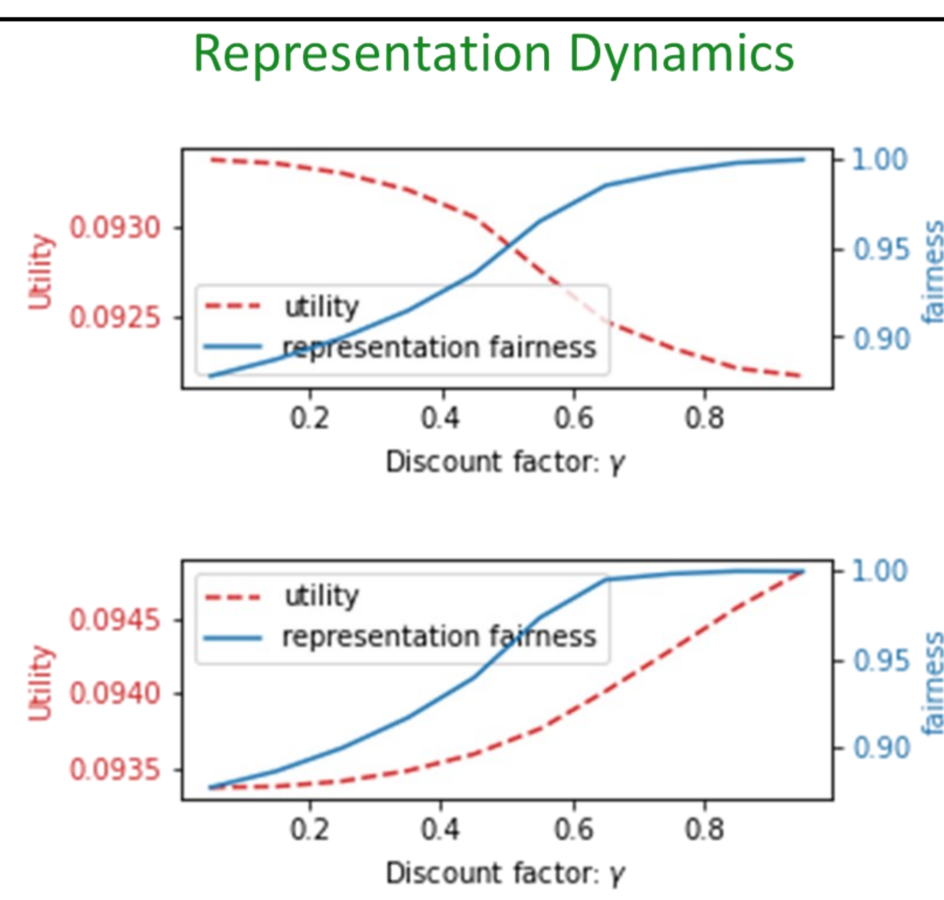
Fairness-Utility Trade-off

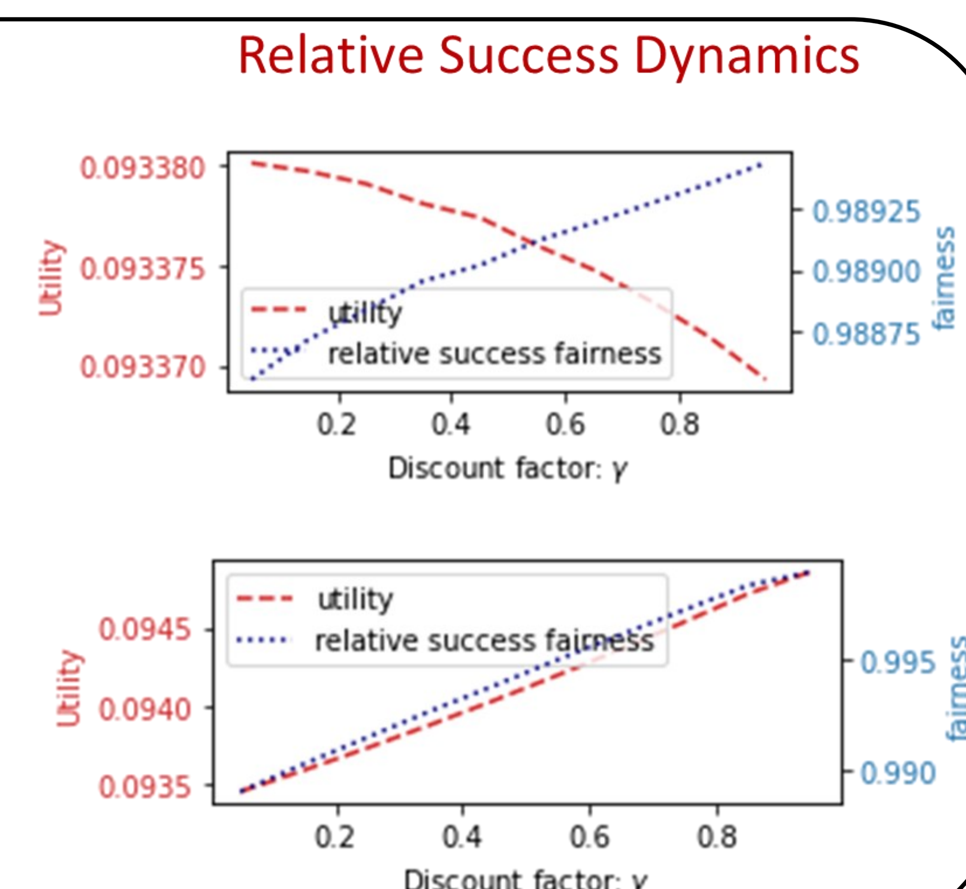
Discount factor:

larger values place more weight on the future

- When observing immediate reward and bias (after 1 step), it appears as if there is a fairness-utility trade-off, dependent on the discount factor.
- Yet, **fairness and utility are aligned in the long-term, the trade-off is between short-term and long-term rewards**

1 step / Long term





Future Work

Facing Uncertainty

In reality we do not know the bias dynamics

MDP with **unknown transition function**

⇨

Method

Learn transition function using Bayesian Linear Regression

Learn optimal policy under current belief using Policy Gradient

UNIVERSITY OF OSLO

unine Université de Neuchâtel