Local Justice and Machine Learning: Modeling and Inferring Dynamic Ethical Preferences around High-stakes Allocations

Violet (Xinying) Chen (vchen3@stevens.edu) [1], Joshua Williams [2], Hoda Heidari [2], Derek Leben [2]
[1]Stevens Institute of Technology [2] Carnegie Mellon University

EAAMO2022

Introduction and Motivation

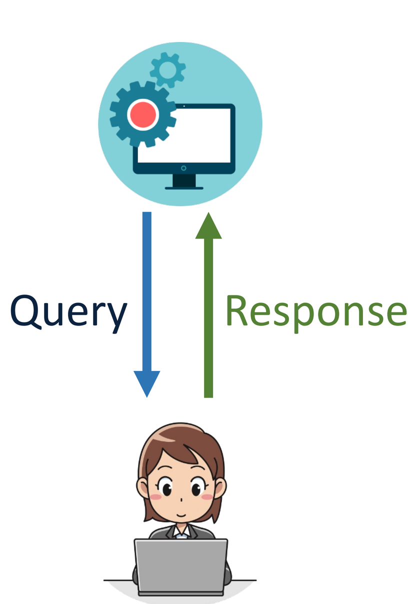
- **Sequential resource allocation decisions** in a high-stakes domain
  - **Evolving social contexts** as resources are allocated over time
  - **Dynamic moral judgments/ethical preferences**: **who should be prioritized** given histories and future implications?
  - **Long-term policy**: stir society towards long-term fairness
- This work: **a human-in-the-loop approach** to capture and infer dynamic ethical preferences toward allocation policies, i.e. **quantify how moral judgments evolve with decision-making contexts.**
  - Design a MDP model to represent sequential resource allocation: moral preferences captured in the MDP's reward function
  - Elicit moral judgment through active learning of reward

Markov Decision Process (MDP) Model

- MDP model: $\langle S, A, P, R\rangle$
  - **State** $s_t = (s_{t,1}, \dots, s_{t,n})$: time t's state of affairs on n groups.
  - **Action** $a_t = (a_{t,1}, \dots, a_{t,n})$: time t's allocation decision.
  - **Transition probability** $P(s_{t+1}|s_t, a_t)$: likelihood of transitioning to s_{t+1} from taking action a_t at state s_t.
  - **Reward** $R(s_t; \boldsymbol{\theta})$: cumulative state reward.
- An allocation policy → MDP trajectory, $\tau = (s_1, a_1, \dots, s_T, a_T, s_{T+1})$ → cumulative policy reward $R(\tau; \boldsymbol{\theta}) = \sum \gamma^{t-1} R(s_t; \boldsymbol{\theta})$
- **Moral judgments** regarding an allocation policy: how much **reward** the policy leads to on the MDP.
- Moral preferences captured in parameters $\boldsymbol{\theta}$ of reward function

Active Learning of Moral Preferences

- Bradley-Terry choice model for comparing policies:
  - Two policies lead to trajectories τ_1, τ_2
  - Likelihood of viewing τ_1 as more **morally desirable** than τ_2 is $P(\tau_1 \succ \tau_2|\boldsymbol{\theta}) = \exp R(\tau_1; \boldsymbol{\theta}) / (\exp R(\tau_1; \boldsymbol{\theta}) + \exp R(\tau_2; \boldsymbol{\theta}))$



- A user's true moral preference is $\boldsymbol{\theta}^*$
- Iterative interaction with the user
  1) Query to compare trajectories: $Q_t = \langle \tau_1, \tau_2 \rangle$
  2) User gives response w.r.t. unknown true reward $R(\tau; \boldsymbol{\theta}^*) : u_t \in \{\tau_1 \succ \tau_2, \tau_2 \succ \tau_1\}$
  3) Standard Bayesian update on estimate $\boldsymbol{\theta}$
     $P(\boldsymbol{\theta}|u_1, \dots, u_t; Q_1, \dots, Q_t)$
     $\propto P(u_1, \dots, u_t; Q_1, \dots, Q_t|\boldsymbol{\theta}) P(\boldsymbol{\theta})$

Example: Medical Resource Allocation

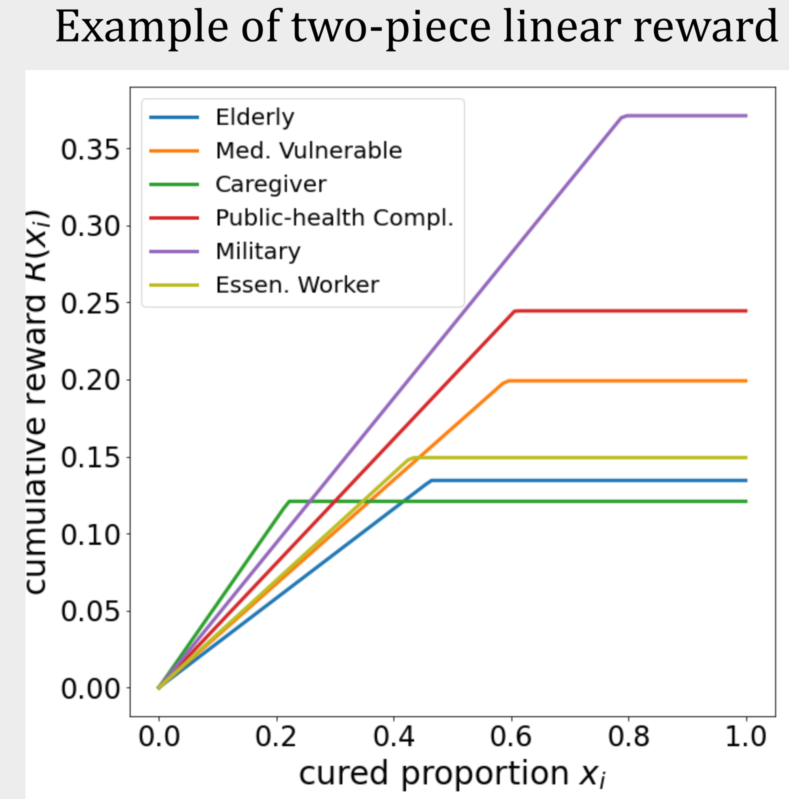
- In medical emergency: decision context shifts → relevant moral principles vary → moral preferences evolve
- Hypothetical viral epidemic: allocate a virus cure in phases
  - Susceptible ⇒ Cured (Immune)
  - Susceptible ⇒ Infected ⇒ Deceased
- Different moral principles → prioritizing different population groups

Prioritarian	Favors the *most vulnerable* members	G1. The elderly G2. The medically vulnerable
Distributive	Favor those with *instrumental values* for society and/or family	G3. Caregivers G4. Essential workers
Restorative	Favor those owed compensation due to their *past actions and efforts*	G5. People with current or previous military service G6. People compliant with public health recommendations

Specification on cure allocation example

S	$s_{t,i} = (x_i^t, v_i^t, d_i^t)$: in group i at time t, • x_i^t: the **cured** proportion (have received the resource) • v_i^t: the **susceptible** proportion (still require the resource) • d_i^t: the **deceased** proportion (have suffered negatively without the resource)
A	a_i^t: the proportion of time t's resources allocated to group i.
P	$P(s_{t+1}\|s_t, a_t) \in \{0,1\}$: deterministic transition
R	Piecewise reward: moral preferences shift between pieces. $R(x_1^t, \dots, x_n^t; \boldsymbol{w}^*, \boldsymbol{c}^*) = w_i^* \sum_i min\{x_i^t, c_i^*\}$

Example of two-piece linear reward

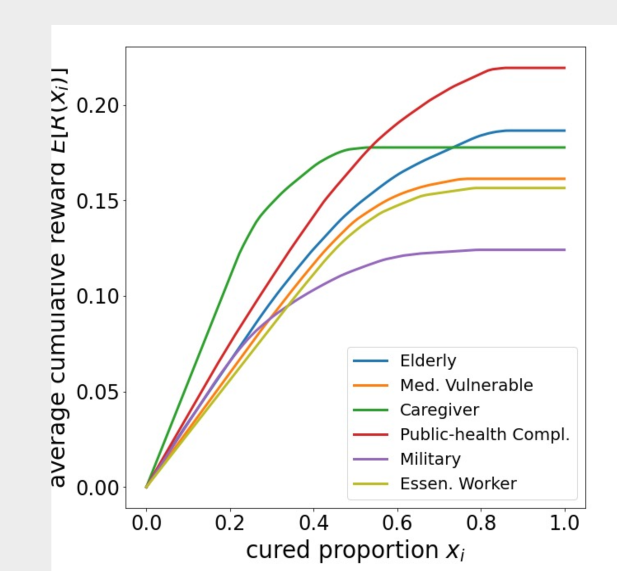


- Before a group is well-cured:
  - $x_i \in [0, c_i]$: cures given to group i rewarded linearly with **weight $\boldsymbol{w_i}$**
- After a group is well-cured:
  - $x_i \in (c_i, 1]$: more cures are not rewarded after group i is sufficiently cured (x_i exceeds **threshold $\boldsymbol{c_i}$**)

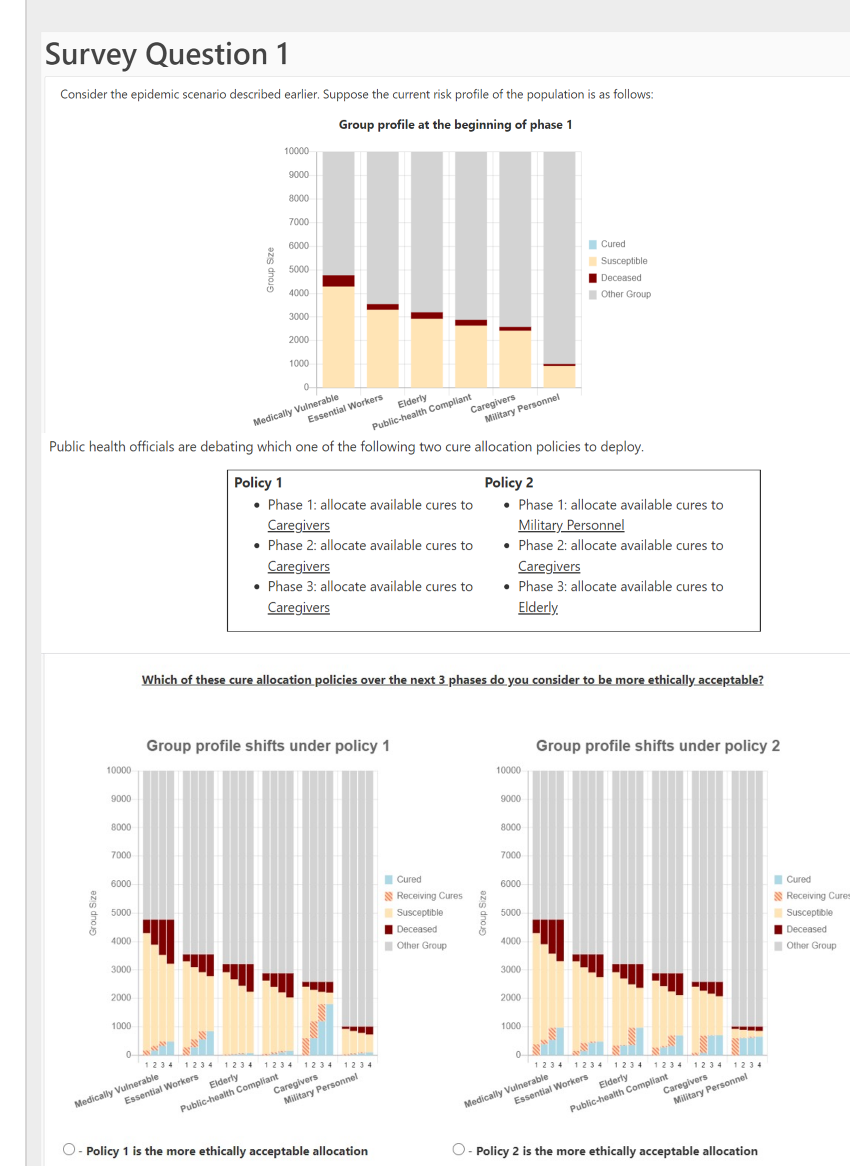
Experiment Design and Findings

- Synthetic population of 10000 people; 6 groups for prioritization.
- Survey run on Amazon Mechanical Turk: 33 responses collected.
- A participant answers 20 questions: each question is chosen to maximize information gain about $\boldsymbol{w}^*, \boldsymbol{c}^*$ based on current estimates
- $\boldsymbol{w}^*, \boldsymbol{c}^*$ are unavailable: use written justifications (respondents explain why a group should/should not be prioritized) as proxies
- **Key observations:**
  - ➢ The inferred rewards show good consistency with justifications.
  - ➢ Participants' moral judgments are highly diverse: they sometimes hold explicit opinions towards certain groups.
  - ➢ From averaging the inferred cumulative rewards, at relatively low cured levels, caregivers are the most prioritized.

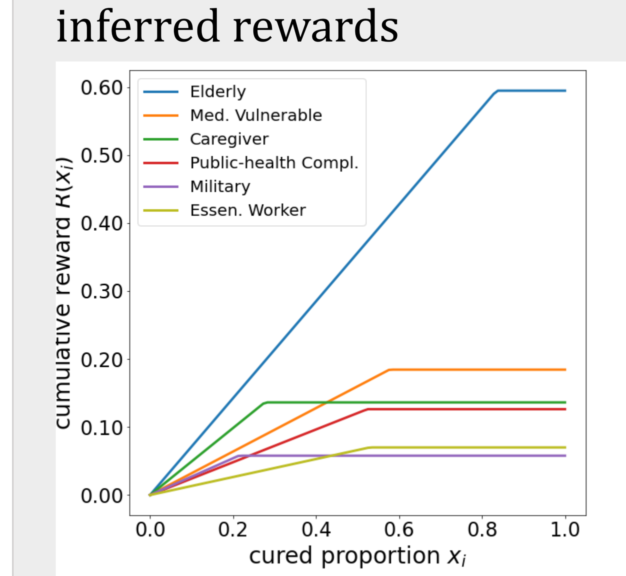
Average group rewards from all resp.



Sample survey question

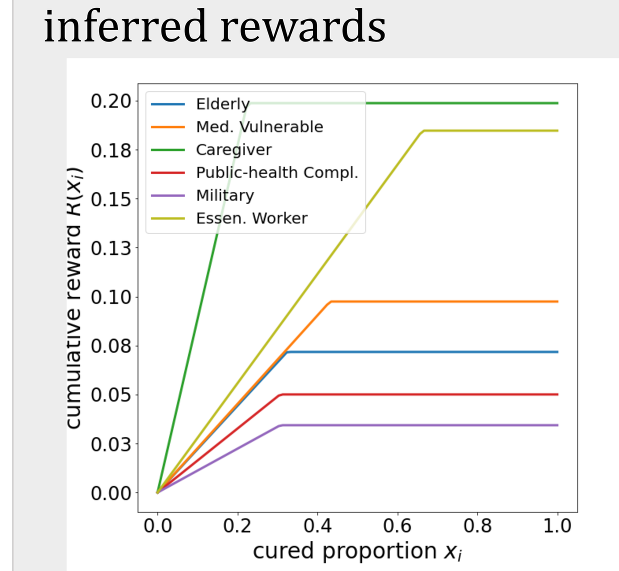


Example 1. individual resp. & inferred rewards



"Large number of **Elderly and Medically Vulnerable** are cured"

Example 2. individual resp. & inferred rewards



"while **the elderly group** has likely provided a great deal for society in the past, in the future they are likely to provide less than the **essential workers group** "