Local Justice and Machine Learning: Modeling and Inferring Dynamic Ethical Preferences toward Allocations

Violet (Xinying) Chen[1], **Joshua Williams**[2], **Derek Leben**[2], **Hoda Heidari**[2]
[1] Stevens Institute of Technology
[2] Carnegie Mellon University
vchen3@stevens.edu, {jnwillia, dleben}@andrew.cmu.edu, hheidari@cmu.edu

Abstract

We consider a setting in which a social planner has to make a *sequence* of decisions to allocate scarce resources in a high-stakes domain. Our goal is to understand stakeholders' *dynamic* moral preferences toward such allocational policies. In particular, we evaluate the sensitivity of moral preferences to the history of allocations and their perceived future impact on various socially salient groups. We propose a mathematical model to capture and infer such dynamic moral preferences. We illustrate our model through small-scale human-subject experiments focused on the allocation of scarce medical resource distributions during a hypothetical viral epidemic. We observe that participants' preferences are indeed history- and impact-dependent. Additionally, our preliminary experimental results reveal intriguing patterns specific to medical resources—a topic that is particularly salient against the backdrop of the global covid-19 pandemic.

Introduction

AI and ML tools are permeating society. These powerful technologies are applied in numerous policy domains to inform or make consequential decisions impacting people's lives. In particular, they increasingly inform or automate high-stakes *allocation* decisions in domains such as lending, employment, and healthcare. The past decade has witnessed an overwhelming body of evidence establishing the need for AI and ML to reflect collective values, such as justice and fairness. However, translating these principles into computationally tractable and verifiable forms has proven challenging. Majority of the efforts towards formulating fairness for ML have adopted a *static* point of view to capture fairness in terms of certain predictive parity condition across demographic groups. These notions are useful in guiding the design of *one-shot* algorithmic interventions, but as demonstrated in D'Amour et al. (2020); Liu et al. (2018), such interventions may be insufficient to attain long-term fairness and justice goals. Due to the dynamics in decision contexts and the *context-dependent* nature of moral ideals, incremental and evolving remedies are often needed to promote justice in the long run (Elster 1992).

A growing body of work has called on the AI-ethics community to bring stakeholders' judgments into the process of formulating values, such as fairness, for AI. Following this human-centric view, we first posit that moral judgments are seldom invariable across situations and contexts. In particular, the ethical principles they prioritize are often informed by numerous consideration, including the historical contexts and the future implications of policies. Even absent of disagreement regarding the right answer to the above questions, how they inform the choice of allocational policies could be a genuine point of normative disagreement. To further complicate the task, allocational policies are often sequential in nature, and their deployment may shift the above determinants of moral judgments, giving rise to a new context tomorrow. So an ethically-minded social planner has to understand and potentially reflect stakeholders' dynamic moral judgments. Understanding such dynamics allows the planner to design effective and acceptable interventions that stir society in the appropriate direction over time.

Along with the increasing recognition that moral judgments are context dependent, e.g. Sinnott-Armstrong (2008); van Baar, Chang, and Sanfey (2019); Andrejević et al. (2020), there has been rising interests in quantifying moral judgments, e.g. Armstrong and Skorburg (forthcoming); Awad et al. (2022). An important gap in literature is how to concretely capture moral judgments' evolution with the decision contexts. In this work, we consider a stylized setting in which a social planner or policymaker has to make a *sequence* of decisions regarding the *allocation* of scarce resources in a high-stakes domain. Stakeholders' moral preferences regarding such allocation policies are influenced by various ethical/moral principles. We aim to understand these preferences, in particular, to evaluate their sensitivity to the history of allocations and the expected future impacts on socially salient groups.

As a concrete example, consider a central agency in charge of allocating scarce medical resources (e.g. vaccines or hospital beds) to patients during a viral epidemic. The allocation decisions today influence the urgency and demand for the resources in question tomorrow. Such context shifts influence people's judgments over who should be prioritized for allocation, thus reflecting different normative moral principles as explained below.

Moral principles from bioethics. There is a robust history of debate in bioethics about normative principles for the distribution of scarce medical resources across a vari-

ety of contexts, e.g. Cohen, Schapire, and Singer (2009); Bayer et al. (2011); Wertheimer and Emanuel (2006). These contexts can range from non-emergency, such as organ donation and hospital triage, to emergency situations, such as natural disasters and pandemics (where the health of an entire population is impacted). One solution to this problem is to simply distribute resources randomly by either a lottery or other randomization policy (Peterson 2008). There is an Egalitarian justification for lotteries, and they are commonly implemented in non-emergency contexts under "first come, first served" schemes. However, in emergency contexts, the policies of hospitals and governments will almost always favor certain groups over others. As Savulescu, Persson, and Wilkinson (2020) declare: "*there are no egalitarians in a pandemic*." We focus on the normative principles typically used to justify these types of emergency policies. We group these normative principles into three broad categories (Further discussion in Appendix A):

- **Prioritarian:** This approach favors the most vulnerable members of a population, such as those who are the sickest, youngest, or oldest, regardless of how and why one is vulnerable. The concern is only about those who are most likely to be impacted severely by a lack of resources allocated to them (e.g., low probability of survival, given no treatment).
- **Distributive:** This approach attempts to maximize the overall benefits and minimize overall losses, which might mean favoring those who have "the most to lose" (e.g., those with high expected chance of survival given the resource), and those with instrumental values at a local (family) or global (society) level.
- **Restorative:** This approach favors those who are owed compensation because of their eligible actions or characteristics, usually some form of qualifying past behavior like lifestyle choices, effort, and social service.

Our work draws motivation from a key observation, that is, the collective judgments about which normative principles are relevant can change as the underlying situation shifts. In the famous case-study of Memorial Medical Center in New Orleans, when not all patients could be evacuated from the storm-ravaged hospital, the staff changed their normative principles over the course of several days from a Prioritarian to a Distributive approach. It is, therefore, crucial to understand the ways in which not only static but also dynamic features of a situation play a role in the development of policies for resource allocation.

The present work. We propose a mathematical model to capture and infer stakeholders' dynamic moral preferences. Our model utilizes a Markov Decision Process (MDP) to represent sequential resource allocation. We assume that the stakeholders' moral judgment regarding alternative allocational policies can be captured by comparing a so-called "reward" each policy leads to on the MDP. Using this moral preference model, we infer a stakeholder's dynamic preferences by learning the reward function they associate to each state-action pair. Utilizing an active preference-based reward learning framework, we design an interactive process to elicit stakeholders' preferences through a series of pairwise comparison questions. We illustrate our framework through small-scale human-subject experiments, designed to elicit crowd workers' moral judgments regarding the allocation of scarce medical resources during a hypothetical viral epidemic. We observe that participants' preferences are indeed context-dependent, and the qualitative justifications they offer for their choice closely match our model's predictions. While we cannot establish statistical significance due to our small sample size, our preliminary results reveal intriguing patterns specific to medical resource allocation.

Related Works

Eliciting ethical judgements for moral AI. While AI literature has a long history of studying preferences, the recent wave of moral AI has inspired questions related to preference modeling, learning and elicitation. Rossi (2016) propose moral preference as one concrete way to approach moral AI. These preferences characterize how people's actions and the implications of such, encode their personal moral/ethical beliefs. Recent works, e.g. Hiranandani, Narasimhan, and Koyejo (2020); Jung et al. (2019); Zhang, Bellamy, and Varshney (2020), study the elicitation of static ethical judgement under different setups. Yaghini, Krause, and Heidari (2021) propose methods of learning context features that could influence ethical judgments. Different from their work, we use ethical principles to select features and focus on inferring moral preferences over these features.

Preference learning via inverse reinforcement learning. One conventional approach to modeling an agent's preference is to use numeric utility functions, e.g. Luce (2012), which associate greater utilities to more preferable options. When the preference is partially or fully unknown, there lacks information to state its utility definition. Preference learning aims to infer the missing information. For example, Kim, Menzefricke, and Feinberg (2007) study spline utility functions and propose a Bayesian approach to infer the spline knots and shapes of individual pieces. Rothkopf and Dimitrakakis (2011) formalize inverse reinforcement learning (IRL), a reinforcement learning (RL) inspired framework aiming to learn the reward functions from feedback signals on optimal or near-optimal policies, as a framework for preference learning by viewing the reward function as a preference-capturing utility function.

Active preference-based reward learning. Rothkopf and Dimitrakakis also highlight the use of active learning to design IRL-based interactive preference elicitation frameworks, where each query is selected based on the previously collected information from an agent. In recent literature, Sadigh et al. (2017) propose an active volume removal method for selecting queries; intuitively, the algorithm selects queries to remove uncertainty in the belief distributions about the reward function parameters. The follow-up work Bıyık et al. (2019) propose an alternative query selection method that aims to optimize the mutual information between belief distributions and query responses.

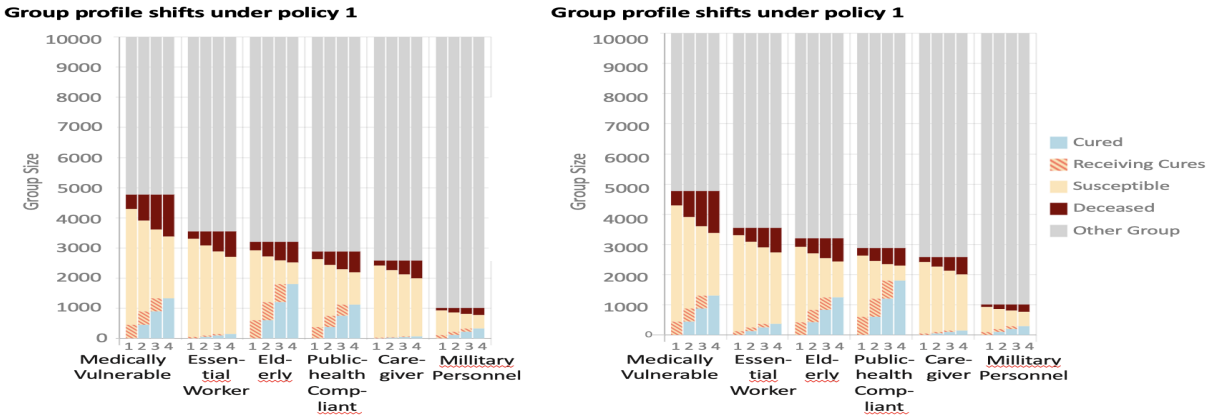


Figure 1: A query of trajectories of three future phases starting from the same initial state; Policy 1 (Left) prioritizes the elderly, while Policy 2 (Right) prioritizes the public-health compliant.

A Mathematical Framework for Dynamic Moral Preferences

We consider a Markov Decision Process (MDP) model representing the sequential allocation of scarce resources. For simplicity, we work with a single type of resource, such as, hospital beds during an epidemic, admission decisions to college applicants. We suppose the allocation policy proceeds in phases: in each phase, the policy prioritizes people with certain features, such as specific age ranges, occupations, etc. We select features based on the relevant ethical principles, so that stakeholders' moral preferences over these principles are reflected via their opinions on who should be prioritized next. As the allocation unfolds over time, stakeholders' moral preferences may shift in line with the evolving societal context.

MDP Model

A standard MDP is defined with a tuple, $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, P, R \rangle$, where $\mathcal{S}$ is a set of states, $\mathcal{A}$ is a set of actions, $P : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$ is the state transition probability function, with $P(s'|s, a)$ denoting the probability of action a transitioning state s to state s', $R : \mathcal{S} \to \mathbb{R}$ is a state-based reward function. We use the medical resource allocation setting adopted in our human-subject experiments as the running example, but we note that the MDP model is generally applicable in other scenarios.

A state $s_t \in \mathcal{S}$ in our running example describes the current state of affairs. We consider n ethically relevant features, and uses these features to assign people to groups. We define s_t with the group features: $s_t = (s_{t,1}, \ldots, s_{t,n})$, where $s_{t,i}$ is a vector representing the profile of group i at step t. Note that we do not require mutually exclusive features, namely, a person may have multiple features, thus belonging to multiple groups simultaneously. In our running example, $s_{t,i} = (x_i^t, v_i^t, d_i^t)$, which respectively are the proportion of group i that, at step t, have received the resource (a cure to a hypothetical virus); require the resource (are susceptible to viral infection); and have suffered negative outcomes due to not receiving the resource (have passed away from contracting the virus). We skip the superscript t when time step is not specified. As we are not aiming to use the most efficient state representations, it may be possible to reduce the size of $\mathcal{S}$ with alternative state definitions.

An action $a_t \in \mathcal{A}$ represents the current time step's resource allocation decision. Similar to the state notation, we can define a_t with the group based action features: $a_t = (a_1^t, \ldots, a_n^t)$, where a_i^t is the *proportion* of the current step's available resources allocated to group i.

Next, to define the transition probabilities, we suppose the MDP model is deterministic, namely $\mathbb{P}(s'|s, a) \in \{0, 1\}$ for all s', s and a. The deterministic assumption fits the resource allocation setting when the population under consideration is large. Because, in this case the statistical effect of the allocation can be estimated with accuracy. Under this assumption, taking action a_t at state s_t gives a deterministic transition to the new state s_{t+1}. Through possible group overlaps, the resources given to one group can spread to other groups. We show an example from our experiment to illustrate further. Figure 1 demonstrates the state shifts from two policies where each phase allocates a fixed number of resources to one group. Since all groups are overlapping, we see that each phase's allocation impacts all groups.

Lastly, we assume the reward function captures a stakeholder's dynamic moral preferences on the resource distribution contexts. A stakeholder will associate higher cumulative rewards to states that s/he views as more ethically acceptable. In literature, such as Bıyık et al. (2019); Sadigh et al. (2017); Wirth et al. (2017), the cumulative reward of a state is commonly defined as a linear combination of the relevant state features. In our running example, a possible linear cumulative reward function is $R(s; w) \coloneqq R(x_1, \ldots, x_n; w_1, \ldots, w_n) = \sum_{i=1}^{n} w_i x_i$, where the weights reflect the degrees of prioritization given to each group. When an action shifts state s to s', the immediate reward is gained at constant rates w, namely, $r(s', s; w) \coloneqq R(s'; w) - R(s; w) = \sum_{i=1}^{n} w_i (x_i' - x_i)$. We note that $R(s; w)$ and $r(s', s; w)$ capture a fixed moral preference, where the importance ranking among groups re-

main unchanged throughout state shifts. Therefore, a linear reward is insufficient for modeling changing priorities across groups over time.

A more flexible alternative is to use a spline reward function (i.e. a piece-wise polynomial function) to model changes in priorities resulted from the shifting contexts in states. Suppose a spline reward function R consists of m pieces $R_1, \ldots, R_m$. For $k = 1, \ldots, m$, each piece represents a different type of moral preference, characterized by a slope vector $w^{(k)}$, in its corresponding domain, defined as a n-dimensional box $[c_1^{(k-1)}, c_1^{(k)}] \times \ldots \times [c_n^{(k-1)}, c_n^{(k)}]$. We can interpret $w^{(k)}$ as the weights used to prioritize the groups. The domain boundary points indicate where the weights change, namely the preferences shift.

Our primary focus in this work is the expression of piecewise linear/constant reward functions. In high-stakes settings such as the allocation of cures during an epidemic, one natural expression of the reward for providing cures is,

$$R(s; w, c) \coloneqq R(x_1, \ldots, x_n; w, c) \coloneqq \sum_{i=1}^{n} w_i \min\{x_i, c_i\}. \quad (1)$$

In this reward function, we skip the piece index in $w^{(k)}, c^{(k)}$ to simplify notation. Throughout the allocation process, group i's feature x_i increases as the group receives more resources. In the beginning, when $x_i \leq c_i$, allocations to group i are rewarded linearly with the weight of w_i. After x_i increases to exceed c_i, the cumulative reward will stay fixed at $w_i c_i$, namely, further allocations to group i gain 0 additional rewards. c_i can be viewed as the resource level that is considered sufficiently high for group i so that further allocation to the group will not be rewarded. Using the two-piece reward function, we can model changes in the importance ranking among groups.

Based on the MDP model, we use a trajectory $\tau = (s_1, a_1, \ldots, s_T, a_T, s_{T+1})$ to denote a sequence of allocation decisions and the resulting state changes. The trajectory reward is a discounted sum of the immediate reward gained at each state of the trajectory, namely, $R(\tau; w, c) = \sum_{t=1}^{T} \gamma^t r(s_{t+1}, s_t; w, c) = \sum_{t=1}^{T} \gamma^t (R(s_{t+1}; w, c) - R(s_t; w, c))$ with $\gamma > 0$ as the discount factor. $R(\tau; w, c)$ represents a stakeholder's perceived gain from shifting the societal context from s_1 to s_{T+1} through a sequence of allocation policies $a_1, \ldots, a_T$ overtime.

Moral Preference Model. We take the perspective of a policy planner who wishes to infer a stakeholder's preference by learning her/his reward function. We follow the framework of preference-based reward learning studied in literature, e.g. Bıyık et al. (2019); Sadigh et al. (2017). The reward learning involves an interactive process where the planner asks a stakeholder to answer a sequence of queries, and uses the query answers to iteratively update the estimates of w and c. Each query is a comparison between two trajectories both starting from the same initial state and of equal length/number of phases. A sample query in the setting of our running example is shown in Fig. 1.

We use the Bradley-Terry model, a standard human choice model (Luce 2012), to represent a stakeholder's moral preference. Suppose a query asks to compare trajectories τ_1 and τ_2. If a stakeholder prefers τ_1 to τ_2, we denote it with $\tau_1 \succ \tau_2$. A stakeholder with reward function $R(s; w, c)$ will choose τ_1 as more preferable, namely, more ethically acceptable with probability:

$$P(\tau_1 \succ \tau_2 | w, c) = \frac{\exp R(\tau_1; w, c)}{\exp R(\tau_1; w, c) + \exp R(\tau_2; w, c)}. \quad (2)$$

By choosing a trajectory, the stakeholder indicates that they consider it morally more acceptable compared to the unchosen alternative trajectory.

An Active Learning Scheme to Learn Moral Preferences

Suppose a stakeholder's true reward function is parameterized by weight w^* and threshold c^*, then our learning goal is to estimate w^* and c^* using the preference queries. Let W, C denote the random variables representing the inferred beliefs about w^*, c^* based on query responses. We begin with prior probability distributions on W and C, then update the posterior distributions with Bayesian inference. Using this Bayesian setup, we adopt the active preference-based reward learning method introduced in Sadigh et al. (2017). We next provide a brief summary of the learning framework and give further details, e.g. our prior choices, in Appendix C.

With the initiated priors of W and C, each iteration of the active learning process has two steps: generate a new preference query, then use the query response to compute the posteriors on W, C. Each query is a comparison between two trajectories. We define $U_{w,c}(Q)$ as a Bernoulli random variable representing a user's response to a given pairwise query $Q = \langle \tau_1, \tau_2 \rangle$ when their preference is parametrized by w and c. In iteration t, let $Q_t \coloneqq \langle \tau_1^t, \tau_2^t \rangle$ denote the selected query, and $u_t \sim U_{w^*,c^*}(Q_t)$ denote the user's response, then we apply the standard Bayesian update $P(w, c | u_1, \ldots, u_t; Q_1, \ldots, Q_t) \propto P(u_1, \ldots, u_t; Q_1, \ldots, Q_t | w, c) P(w, c)$.

Mutual Information-based Query Selection. For efficient learning, we seek to use a small number of queries to obtain accurate estimates. In literature on reward learning, various query selection methods have been proposed for the common linear reward. We extend the approach from Bıyık et al. (2019) which selects queries to optimize the information gain about W. Specifically, we choose queries via maximizing the mutual information between the joint distribution on W, C and the query response distribution $U_{\bar{w},\bar{c}}$. Similar to Bıyık et al. (2019), we implement the selection via sample-based approximations based on a sample Ω drawn from the current belief distribution on W, C and a pre-generated set $\mathcal{Q}$ of pairwise queries to select from. Details on the selection step are provided in Appendix C. We note that the choice of $\mathcal{Q}$ and the generation of Ω both involve trade-offs between computational costs and learning performances. Using a larger $\mathcal{Q}$ or advanced sampling methods to generate Ω could improve the inference accuracy of w^*, c^*, but each iteration's query selection problem would become computationally more expensive. As the active learning process depends on repeated interactions between the learner and stakeholders as the respondents, it is impractical to have respondents wait an extended amount of time for each new

query. In our experiments, as discussed in the next section, we utilize heuristics to reduce the query selection time.

Experimental Design

We evaluate the effectiveness of our framework through human-subject experiments on Amazon Mechanical Turk. Figure 2 provides a brief description of the experimental goals as presented to survey respondents. Each participant is introduced to a hypothetical pandemic scenario in which they choose how to allocate a limited number of cures based on their personal moral assessments. Over a series of 20 questions, each participant is shown two potential cure allocations and a description of how many people from different groups (e.g., the elderly) will be cured, will die, or will remain susceptible depending to their choice (Figure 1). Once a participant chooses an allocation of cures, they are then required to briefly justify their choice in order to allow us to validate our framework against an individual's stated moral preferences. We obtain 33 responses satisfying our worker and acceptance criteria stated in Appendix D. For more detail, we include a full set of instructions, and example questions in Appendix D as well. Before performing this study, we received approval to run this survey from our university's Institutional Review Board (IRB).

A note on our choice of groups. Our group choices are driven by the moral principles that we introduced in previous sections. Prioritizing the elderly and/or the medically vulnerable reflects *Prioritarianism*. The past-oriented *Restorative* principle is reflected in two ways: prioritizing people compliant with public health recommendations rewards their past responsible behaviors; prioritizing military personnel represents reciprocity for their past services. Lastly, prioritizing caregivers and essential workers respectively promotes instrumental value for local communities (e.g. family) and global communities (e.g. society), which reflect the future-oriented *Distributive* principle.

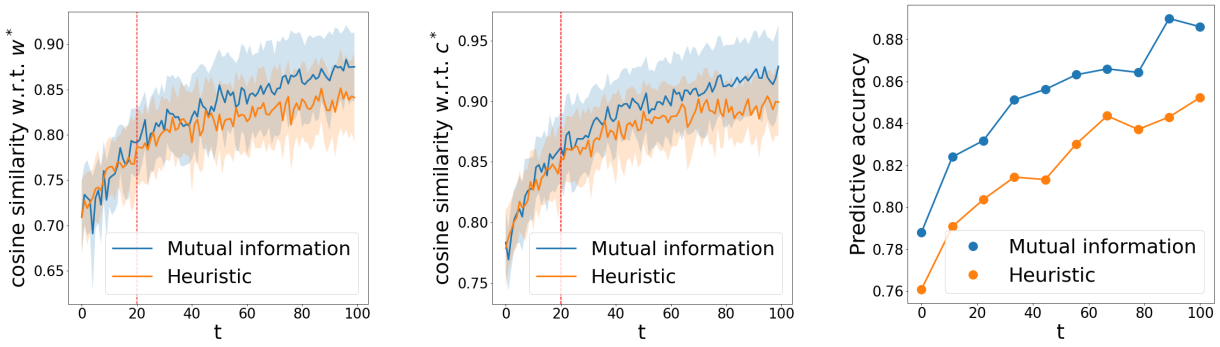
Data Generation. In order to ensure that our scenario is realistic, we constructed a synthetic population of 10000 adults using three real-world data sources: (1) 2020 National Health Interview Survey, Adult interview results; (2) 2020 Labor Force Statistics, Employed persons by industry and age; (3) United States Census Data. We utilized this synthetic population as the basis of all our queries. To determine the state shifts from a given cure allocation policy, we simulate the synthetic population at the individual level. The state shifts include two types of changes in a group's profile: (1) a susceptible individual contracts the virus and succumbs to it (2) a not-yet-immune individual (either susceptible or infected) receives the cure, thus becoming fully immune. In Appendix D, we describe in detail each data source, and how we generate the population and simulate state shifts.

Heuristics for Query Generation and Selection. Recall from previous sections, we utilize heuristics for generating and selecting queries to minimize delays incurred by respondents during the survey. For query generation, we generate $\mathcal{Q}$ to include a mix of start states and allocation policies over the next three phases (leading to trajectories of length 4). For query selection, we pre-generate the possible query sequence scenarios up to 10 iterations from a fixed starting query as a query tree, then use the tree to select queries in our experiments. Further details on both generation and selection heuristics are provided in Appendix D.

Background and Task Description The goal of this survey is to understand your moral judgments regarding the sequential allocation of scarce medical resources.
Hypothetical Scenario Imagine a viral epidemic that has infected millions of people around the world leading to a disease with a very high mortality rate. There is currently only a single highly effective **cure** for the disease—those who receive the cure will fully recover (if currently infected) and become immune to the virus in the future. Unfortunately, the number of cure doses that can be produced and administered every month is limited, so public health officials need to decide which groups should be prioritized at any given time. In the questionnaire that follows, we will present you with additional information about several possible states of the epidemic and ask you to choose your preferred allocational policy between two cure allocation policies.
Taking numerous considerations into account, public health officials have decided to adopt a **phased** cure distribution program among adults. Each phase consists of allocating the small number of available cure dosages to one of the following demographic groups. (Note that groups can be overlapping. For example an individual may be a member of the elderly group as well as the medically vulnerable).
G1 (Elderly): people who are 65 years old or older;
G2 (Medically Vulnerable): people with pre-existing medical conditions that increase their susceptibility to contracting the virus. (Conditions include cancer, heart conditions, chronic lung/liver/kidney diseases, diabetes, immunosuppression, and pregnancy);
G3 (Caregivers): people who have dependents (e.g. young children);
G4 (Public-health Compliant): people who are compliant with public health recommendations (which include, for example, reducing unnecessary commutes and convenings);
G5 (Military Personnel): people who have previously or currently served in the military;
G6 (Essential Workers): people who are essential workers across industries like agriculture, manufacturing, transportation and utilities, education services, and health services.
Officials are now debating which group should be prioritized for receiving the cure at each phase. In the questions that follow, you will be presented with the initial state of the epidemic (broken down to group-dependent **risk profiles**), two alternative **allocational policies** along with their projected **impact** on the population. You will then be asked to indicate which policy you believe is more acceptable from a moral standpoint. We will provide further details on the above key terms next.

Figure 2: Task Description Screen Shown to Participants

Establishing Viability through Simulations. As a proof of concept, we use simulations to demonstrate the performance of the active learning framework at inferring the true parameters w^* and c^* underlying a participant's reward function. In addition, we aim to justify two design choices: asking 20 questions to each participant, and the query selection heuristic. Suppose w^*, c^* capture a partici-



(a) Cosine similarity between w^* and estimates w over 100 iterations

(b) Cosine similarity between c^* and estimates c over 100 iterations

(c) Predictive accuracy on 1000 random queries every 20 iterations

Figure 3: Means from 20 instances, the shaded region is 95% confidence interval for the means. Dashed line indicates the values after 20 iterations. Predictive accuracy improves from running more iterations. The gaps between heuristic and mutual information are no more than 0.05.

pant's true preference in the reward function $R(\tau; w^*, c^*)$, and $\{w^t, c^t\}_{t=0,1,\ldots,T}$ is the sequence of point estimates $\bar{w}, \bar{c}$, sample means of each iteration's inferred belief distributions. We are interested in two types of performance measures. One is the *estimation accuracy*: we use the expectations of cosine similarity to measure how well the inferred belief distributions on W, C characterize w^* and c^*. The other one is the *prediction accuracy* on new queries, that is, how well an estimate $\hat{w}, \hat{c}$ can predict one's preferences over new pairs of trajectories. We use the public Github repository APReL (Bıyık 2020) as the basis to implement the simulations. Additional details are discussed in Appendix D.

Figures 3 show the simulation results. From these plots, we observe that 20 iterations are sufficient to achieve reasonably high cosine similarities and predictive accuracy. While running more iterations leads to further improvement in estimate and prediction performances, the rate of improvement begins to decline around iteration 20 and appears to stabilize around iteration 40. In human subject experiments, it is desirable to ask a smaller number of questions due to people's limited attention span, so we restrict the learning process to 20 iterations. By comparing results from our query selection heuristic and the full mutual information based selection, as we expect, the heuristic improves computational efficiency at the cost of learning performances. Since the drop in performances is relatively minor, we argue that the heuristic is an effective choice.

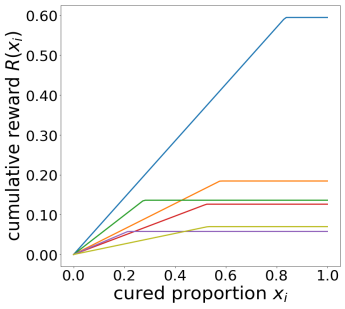
Experimental Findings and Discussion

On the responses collected from 33 surveys, we examine the preference estimates w, c and the written justifications to understand participants' ethical judgments, and provide insights on the practicality of our model. Additional validation tests are discussed in Appendix E. Since participants' true rewards are not accessible, we cannot compare the inferred estimates against the ground truth as in the previous simulation. Instead, we rely on the written justifications as a proxy for participants' true preferences.

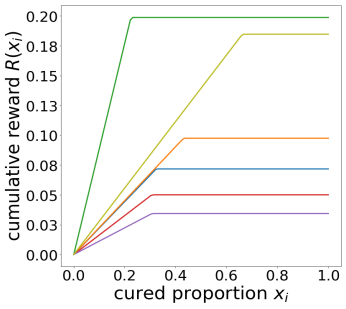
An overview of individual participants' responses. We begin by checking individual surveys separately to evaluate whether the learned preference is consistent with the justifications given by participants. Given the small number of questions, we individually examine each participant's top priority group(s) and/or ethical position by reading their free-form justifications. In Figure 4, we show examples of two participants' inferred preferences and justifications. Overall, we observe that participants' ethical judgments are highly diverse; they often hold explicit opinions towards specific groups and are overt with their preferences when writing justifications, which provides a convenient mechanism for verifying their inferred rewards. In Appendix E, we discuss in detail the notable patterns in the responses and additional evidence for the inference validity.

An overview of collective responses. Next we look at the aggregate preference from all respondents to offer a better sense of the data as a whole. It is important to note that that we don't endorse simple averaging as an an appropriate mechanism of aggregating preferences. Rather, due to space constraints we cannot detail every participant's preferences; we use it as one heuristic approach to offer a glimpse of the entirety of the responses. Fig. 4c displays the average cumulative reward associated with each group.The question of how such multi-dimensional preferences should be aggregated appropriately is outside the scope of the current paper. We observe that each average reward has a concave shape, which fits the intuition that marginal benefits of cure allocation decrease as more people receive the cure.

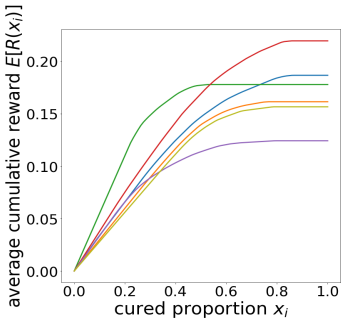
Based on these average rewards, we observed that when all groups are at relatively low cured levels, caregivers, corresponding to the Distributive principle, is the most prioritized group. As cure allocation continues, the public-health compliant and the elderly, respectively corresponding to Restorative and Prioritarian principles, become more important to our participants. In other words, public-health compliant and elderly groups will gain greater priority once caregivers are viewed as sufficiently cured. The medically vulnerable and essential workers have similar reward trends as



(a) 18-25 years old, male, undergraduate, asian, liberal



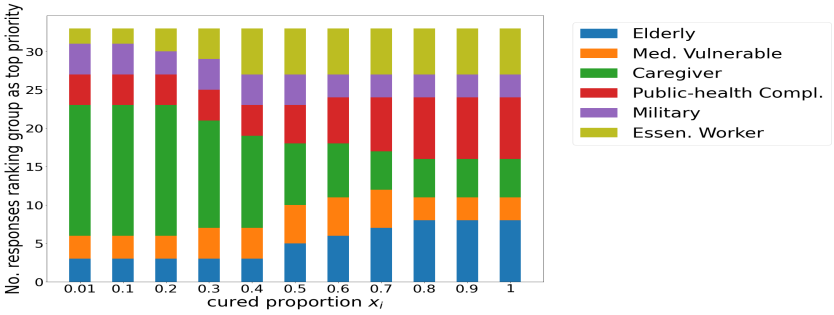
(b) 41-60 years old, male, graduate, white, other



(c) Average group rewards based on aggregate survey responses

(a): *"Large number of elderly and medically vulnerable are cured"*.
(b): *"I don't feel justified in spending a precious limited resource (medicine) on a group that has the shorter lifespan ahead of them. Also, 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"*.

(d) Sample justifications from respondents (a) and (b) shown above



(e) Group ranking distribution based on aggregate survey responses

Figure 4: (a,b) Samples of survey responses and corresponding preference estimates. Groups' reward rankings reflect their priority orderings in a respondent's ethical judgments. (c) Average reward for each group over all participants. (e) Distribution of participants' highest ranked preferences as the proportion of the cured population increases; color legends apply to all figures.

the elderly. Lastly, the military personnel group has slightly higher rewards than the medically vulnerable and essential workers in the beginning, but drops to the least prioritized for the rest of the allocation process, reflecting a belief that Restorative principles were of greater importance when many individuals were vulnerable. Fig. 4e displays such dynamics from an alternative view: each stacked bar shows the distribution of the top priority groups when all groups are cured to the same level. Caregivers are the most common top priority at lower cured levels. As more people are cured, it becomes increasingly desirable to prioritize essential workers, public-health compliant and elderly groups.

Conclusion and Future Directions

We study the modeling and inference of stakeholders' moral preferences regarding the sequential allocation of scarce resources. We propose a human-in-the-loop approach to quantify the dynamic shifts in people's moral judgments as the decision contexts evolve. Our approach provides a useful middle ground between the primarily qualitative social science aspects of moral judgment and quantitative modeling of moral preferences. As the running example, we consider stakeholders' moral judgments towards allocating medical resource during a public health emergency driven by the Prioritarian, Distributive and Restorative principles.

We apply our framework to a small-scale human subject experiment on Amazon Mechanical Turk. Our key finding is that the inferred ethical preferences are consistent with the respondents' reported ethical reasoning. The responses demonstrate that people's ethical judgments are context dependent and evolve with resource distribution shifts.

The limitations of our work suggest several directions for future work. First, the proposed model has methodological limitations due to our modeling assumptions. For example, out of computational considerations, we focus on one of the simplest spline formats to define reward function and rely on heuristics to streamline the query selection in our human-subject experiments. A natural extension is to use more expressive reward formats. On a related note, it can be useful to study whether and how the selected queries affect the learned preferences. The findings could lead to more efficient query selection. Second, the practice of abstracting moral preferences into mathematical models comes with inherent limitations as the abstract model is unable to capture all subtleties and nuances of moral judgments. Future research is needed to better understand the potentials and limits of moral preference modeling. Lastly, our reliance on crowdworkers to illustrate our framework should not be misinterpreted as us advocating for the crowdsourcing of high-stakes moral judgments. Rather, we hope our illustration of the nuanced dynamic nature of stakeholders' moral preferences serves as evidence that policy planners or decision makers should seek better understandings of these preferences before designing allocation policies and interventions in socially consequential domains.

Acknowledgments

H. Heidari acknowledges support from NSF (IIS2040929), J.P.Morgan, CyLab, Meta, and PwC. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not reflect the views of the National Science Foundation and other funding agencies.

References

Akrour, R.; Schoenauer, M.; and Sebag, M. 2012. April: Active preference learning-based reinforcement learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 116–131. Springer.

Andrejević, M.; Feuerriegel, D.; Turner, W.; Laham, S.; and Bode, S. 2020. Moral judgements of fairness-related actions are flexibly updated to account for contextual information. *Scientific reports*, 10(1): 1–17.

Armstrong, W. S.; and Skorburg, J. A. forthcoming. How AI can AID bioethics. *Journal of Practical Ethics*.

Arora, S.; and Doshi, P. 2021. A survey of inverse reinforcement learning: Challenges, methods and progress. *Artificial Intelligence*, 297: 103500.

Awad, E.; Levine, S.; Anderson, M.; Anderson, S. L.; Conitzer, V.; Crockett, M.; Everett, J. A.; Evgeniou, T.; Gopnik, A.; Jamison, J. C.; et al. 2022. Computational ethics. *Trends in Cognitive Sciences*.

Basu, C.; Bıyık, E.; He, Z.; Singhal, M.; and Sadigh, D. 2019. Active learning of reward dynamics from hierarchical queries. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 120–127. IEEE.

Bayer, R.; Bernheim, R.; Crawley, L.; Daniels, N.; Goodman, K.; Kass, N.; Lo, B.; Rosenbaum, S.; Ruger, J.; Sankar, P.; Wheeler, M.; and Wolf, L. 2011. Ethical Considerations for Decision Making Regarding Allocation of Mechanical Ventilators during a Severe Influenza Pandemic or Other Public Health Emergency. In *Centers for Disease Control and Prevention*.

Bechavod, Y.; Jung, C.; and Wu, S. Z. 2020. Metric-free individual fairness in online learning. *Advances in neural information processing systems*, 33: 11214–11225.

Bıyık, E. 2020. APReL: A Library for Active Preference-based Reward Learning Algorithms. https://github.com/Stanford-ILIAD/APReL. Accessed: 2021-11-01.

Bıyık, E.; Palan, M.; Landolfi, N. C.; Losey, D. P.; and Sadigh, D. 2019. Asking easy questions: A user-friendly approach to active reward learning. *arXiv preprint arXiv:1910.04365*.

Christiano, P.; Leike, J.; Brown, T. B.; Martic, M.; Legg, S.; and Amodei, D. 2017. Deep reinforcement learning from human preferences. *arXiv preprint arXiv:1706.03741*.

Cohen, W. W.; Schapire, R. E.; and Singer, Y. 2009. Principles for allocation of scarce medical interventions. *Lancet*, 373: 423–431.

D'Amour, A.; Srinivasan, H.; Atwood, J.; Baljekar, P.; Sculley, D.; and Halpern, Y. 2020. Fairness is not static: deeper understanding of long term fairness via simulation studies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 525–534.

Elster, J. 1992. *Local justice: How institutions allocate scarce goods and necessary burdens*. Russell Sage Foundation.

Emanuel, E.; Persad, G.; Upshur, R.; Thome, B.; Parker, M.; Glickman, A.; Zhang, C.; Boyle, C.; Smith, M.; and Phillips, J. 2020. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *New England Journal of Medicine*, 382: 2049–2055.

Gillen, S.; Jung, C.; Kearns, M.; and Roth, A. 2018. Online learning with an unknown fairness metric. *Advances in neural information processing systems*, 31.

Hiranandani, G.; Narasimhan, H.; and Koyejo, S. 2020. Fair performance metric elicitation. *Advances in Neural Information Processing Systems*, 33: 11083–11095.

Jabbari, S.; Joseph, M.; Kearns, M.; Morgenstern, J.; and Roth, A. 2017. Fairness in reinforcement learning. In *International conference on machine learning*, 1617–1626. PMLR.

Joseph, M.; Kearns, M.; Morgenstern, J.; Neel, S.; and Roth, A. 2018. Meritocratic fairness for infinite and contextual bandits. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 158–163.

Joseph, M.; Kearns, M.; Morgenstern, J. H.; and Roth, A. 2016. Fairness in learning: Classic and contextual bandits. *Advances in neural information processing systems*, 29.

Jung, C.; Kearns, M.; Neel, S.; Roth, A.; Stapleton, L.; and Wu, Z. S. 2019. An algorithmic framework for fairness elicitation. *arXiv preprint arXiv:1905.10660*.

Kim, J. G.; Menzefricke, U.; and Feinberg, F. M. 2007. Capturing flexible heterogeneous utility curves: A Bayesian spline approach. *Management Science*, 53(2): 340–354.

Liu, L. T.; Dean, S.; Rolf, E.; Simchowitz, M.; and Hardt, M. 2018. Delayed impact of fair machine learning. In *International Conference on Machine Learning*, 3150–3158. PMLR.

Luce, R. D. 2012. *Individual choice behavior: A theoretical analysis*. Courier Corporation.

Marcus, R. 2021. Doctors should be allowed to give priority to vaccinated patients when resources are scarce. In *Washington Post*.

Mukherjee, D.; Yurochkin, M.; Banerjee, M.; and Sun, Y. 2020. Two simple ways to learn individual fairness metrics from data. In *International Conference on Machine Learning*, 7097–7107. PMLR.

Ng, A. Y.; Russell, S. J.; et al. 2000. Algorithms for inverse reinforcement learning. In *Icml*, volume 1, 2.

Palan, M.; Landolfi, N. C.; Shevchuk, G.; and Sadigh, D. 2019. Learning reward functions by integrating human demonstrations and preferences. *arXiv preprint arXiv:1906.08928*.

Peterson, M. 2008. The moral importance of selecting people randomly. *Bioethics*, 22: 321–327.

Ramachandran, D.; and Amir, E. 2007. Bayesian Inverse Reinforcement Learning. In *IJCAI*, volume 7, 2586–2591.

Rosenbaum, L. 2020. Facing Covid-19 in Italy — Ethics, Logistics, and Therapeutics on the Epidemic's Front Line. *New England Journal of Medicine*, 382: 1873–1875.

Rossi, F. 2016. Moral preferences. In *The 10th Workshop on Advances in Preference Handling (MPREF)*.

Rothkopf, C. A.; and Dimitrakakis, C. 2011. Preference elicitation and inverse reinforcement learning. In *Joint European conference on machine learning and knowledge discovery in databases*, 34–48. Springer.

Sadigh, D.; Dragan, A. D.; Sastry, S.; and Seshia, S. A. 2017. Active Preference-Based Learning of Reward Functions. In *Robotics: Science and Systems*.

Savulescu, J.; Persson, I.; and Wilkinson, D. 2020. Utilitarianism and the pandemic. *Bioethics*, 34: 620–632.

Singer, P. 2022. Victims of the Unvaccinated. In *Project Syndicate*.

Sinnott-Armstrong, W. 2008. Framing moral intuitions.

Tokars, J. I.; Olsen, S. J.; and Reed, C. 2018. Seasonal incidence of symptomatic influenza in the United States. *Clinical Infectious Diseases*, 66(10): 1511–1518.

van Baar, J. M.; Chang, L. J.; and Sanfey, A. G. 2019. The computational and neural substrates of moral strategies in social decision-making. *Nature communications*, 10(1): 1–14.

Wertheimer, A.; and Emanuel, E. 2006. Who should get influenza vaccine when not all can? *Science*, 5775: 854–855.

Wirth, C.; Akrour, R.; Neumann, G.; Fürnkranz, J.; et al. 2017. A survey of preference-based reinforcement learning methods. *Journal of Machine Learning Research*, 18(136): 1–46.

Wirth, C.; Fürnkranz, J.; and Neumann, G. 2016. Model-free preference-based reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.

Yaghini, M.; Krause, A.; and Heidari, H. 2021. A Human-in-the-loop Framework to Construct Context-aware Mathematical Notions of Outcome Fairness. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 1023–1033.

Zhang, Y.; Bellamy, R.; and Varshney, K. 2020. Joint optimization of AI fairness and utility: A human-centered approach. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 400–406.

Ziebart, B. D.; Maas, A. L.; Bagnell, J. A.; Dey, A. K.; et al. 2008. Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, 1433–1438. Chicago, IL, USA.