Participatory Objective Design via Preference Elicitation

Ali Shirali[*]
shirali_ali@berkeley.edu
UC Berkeley
Berkeley, CA, USA

Jessie Finocchiaro[*]
finocch@bc.edu
Harvard University CRCS
Cambridge, MA, USA

Rediet Abebe
Harvard Society of Fellows
Cambridge, MA, USA

ABSTRACT

In standard resource allocation problems, the designer sets the objective function, which captures the central allocation goal, in a top-down manner. The agents primarily participate in the allocation mechanism by reporting their preferences over the items; they cannot influence the objective once the designer sets it. Implicitly, this approach presumes that standard ways of eliciting the agents' preferences adequately represent their true preferences—an assumption which does not hold if agents have preferences not just over the *items* they receive but also over the *objective* being optimized. For instance, agents may also have social preferences, such as inequality-aversion, altruism, or similar other-regarding behavior. We cannot express such preferences through standard cardinal utilities or ordinal rankings over the items the designer would typically elicit from the agents.

This work examines how we can use this bottom-up preference elicitation stage to enable participants to express preferences over the objectives. We present a versatile framework that elicits agents' preferences over a possible set of objectives and then minimally alters the underlying optimization problem to solve for a new objective that combines both the standard benchmark objective and the agents' preferences for other objectives. We show how to evaluate this new participatory approach against the standard approach, using our notions of loss and gain in social welfare as well as individual tradeoffs.

We illustrate the potency of this framework using a well-studied fair division problem where the designer aims to allocate m divisible items to n agents. In the standard setting, the designer optimizes for utilitarian social welfare, i.e., the sum of the agents' cardinal utilities. We assume that some agents are also inequality-averse and may, therefore, have preferences for objectives that minimize inequality. Using the popular Fehr and Schmidt [31] model, we demonstrate how to map this fair division question to our framework, where the participatory approach optimizes both the standard utilitarian social welfare objective and the agents' heterogeneous preferences over the level of inequality. We examine this problem theoretically to show that there can be large gains in social welfare if the designer uses this participatory approach. Further, we show that the loss in social welfare is linear in the level of inequality aversion and independent of the number of agents. We present a tighter bound in both cases under further natural assumptions on the preferences. We also examine the worst-case cost an individual agent might incur.

Our results indicate that the loss in social welfare (measured by the standard objective) and gain in social welfare (measured by the participatory one) can favor the participatory approach in several natural settings. Throughout the work, we highlight various promising avenues for examining this participatory approach in the specific case study tackled in this paper and a broader range of resource allocation problems.

CCS CONCEPTS

• **Social and professional topics** → *Computing / technology policy*; • **Theory of computation** → *Algorithmic game theory and mechanism design*; *Mathematical optimization*; • **Applied computing** → Decision analysis.

KEYWORDS

Participatory Design, Preference Elicitation, Resource Allocation, Social Preferences, Mechanism Design

ACM Reference Format:
Ali Shirali, Jessie Finocchiaro, and Rediet Abebe. 2024. Participatory Objective Design via Preference Elicitation. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 26 pages. https://doi.org/10.1145/3630106.3658994

1 INTRODUCTION

In standard use of algorithms and mechanism design for resource allocation, a central planner determines various aspects of the mechanism, including the central objective function we optimize. On the other hand, the participating agents primarily engage by contributing their preferences over the items. Underlying this setup is an assumption that standard ways of eliciting preferences—which often entail reporting their cardinal utilities over the items or their ranking of the items—can adequately represent the agents' true preferences. Notably, agents cannot influence the overall objective once the designer sets it. Naturally, we may assume that enabling agents to participatorily design the central objective is costly, difficult to implement, and challenging to study theoretically.

In this work, we identify a possible participatory design framework that balances these competing needs. We present a versatile framework that minimally alters the underlying optimization problem in resource allocation to incorporate the agents' preferences over the central objective. This framework leverages the natural bottom-up preference elicitation stage to capture agents' preferences not only over the items but also over the set of possible objectives. We further define notions of loss and gain in social welfare as well as individual tradeoffs incurred by the worst-off agent

[*] Both authors contributed equally to this research.

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0450-5/24/06
https://doi.org/10.1145/3630106.3658994

to evaluate how this participatory approach stacks up against the standard approach.

We then illustrate and stress-test this participatory approach using a well-studied fair division problem where the designer wants to allocate m divisible items to n agents. Under the standard approach, these multi-objective agents would only report their cardinal utilities u_i over the items and the designer optimizes for utilitarian social welfare, $\sum_i u_i$. For our case study, we assume that some of these agents are inequality-averse, as modeled by Fehr and Schmidt [31]. This popular behavioral economics model is one of the canonical social preference models, which generally study other-regarding behavior, including altruism, certain fairness concerns, and inequality aversion [5, 19, 30].

We then study the loss to standard social welfare from the designers' perspective and the gain to social welfare from the agents' perspective when we move from the standard to the participatory approach. We study the gain and loss in various general settings, finding that the relative loss can, at most, grow linearly in the level of inequality aversion and is independent of the number of participants. We also find that the ratio of gain-to-loss can be unbounded in some natural settings, highlighting potential significant gains. We provide tighter bounds under further natural assumptions on the (dis)similarity of the agents' preferences. We also examine the worst-case tradeoff any individual may suffer and find that individual tradeoffs can be linear in the number of participants. Finally, we address questions of strategy-proofness, by discussing possible designs to elicit the agents' true preferences over the objectives.

Our analyses suggest that the participatory approach, which elicits agents' preferences for inequality aversion, comes only at a small cost to efficiency, measured by standard notions of utilitarian welfare. Moreover, it can yield significant gains, measured by the participatorily designed objective. This suggests that empowering algorithm participants to contribute to shaping the objective function can drastically improve community-level outcomes. We contextualize our contribution within the broader research literature in Appendix B and discuss possible avenues for research exploration in Section 6.

2 PROBLEM FORMULATION

We begin by introducing our broader framework, specifically in the context of resource allocation. We then illustrate this framework's potency using a well-studied fair division problem. While we present our key technical contributions via this case study, the framework applies more broadly. We discuss generalizations in Section 6, where we highlight additional research avenues, and in Appendix A, where we demonstrate how this framework captures other existing studies of resource allocation.

Consider a resource allocation problem where a designer wants to allocate m items to n agents. Let $x = [x_{ij}]$ be the allocation matrix in the set of feasible allocations $\mathcal{X} \subseteq \mathbb{R}_+^{n\times m}$. For the case of divisible items, x_{ij} is the proportion of item j allocated to agent i. We begin with a standard formulation, where agents have linear utilities

$$u_i(x) = \sum_{j\in[m]} a_{ij}\, x_{ij}\,. \tag{1}$$

Here, u_i is the utility of agent i and $a = [a_{ij}]$ denotes *utility coefficients* that parameterize the utility function. We use $\boldsymbol{u}(x)$ to denote the utility profile over the n agents: $\boldsymbol{u}(x) = (u_1(x), u_2(x), \dots, u_n(x))$. The utility function above can be rewritten more concisely as $u_i(x) = \langle \boldsymbol{a}_i, \boldsymbol{x}_i \rangle$, where $\boldsymbol{a}_i$ and $\boldsymbol{x}_i$ denote the i^{th} rows of a and x, respectively.[1]

In standard resource allocation problems, the designer determines various aspects of the allocation mechanism—such as the objective function, resource availability, or fairness constraints—in a top-down manner. Each agent's participation is primarily limited to reporting their utility coefficients $\boldsymbol{a}_i$. The standard approach, therefore, implicitly assumes that such utilities adequately reflect the agents' preferences for allocative outcomes.

Agents may, however, have preferences beyond their utility over the items. In our case study, we consider the setting where agents have social preferences, i.e., they care not only about their utility for items they receive but also about others' utility for their respective allocations. Examples include preferences over the level of inequality imposed by the allocation, altruism towards other agents, and various fairness concerns. Put more simply, the agents' preferred *objective* can be distinct from one another and the objective set by the designer, and might be more complex than their preference for their allocation.

Definition 2.1 (The set of objectives). We assume there is a set of possible objectives $\mathcal{H}$, where each objective $h \in \mathcal{H} : \mathcal{X} \times [n] \to \mathbb{R}$ maps an allocation x into a real value for the respective agent. The objectives depend on the utility coefficients a, which we drop from the notation for brevity.[2]

The set of possible objectives can be general. It may include, for instance, each agent's utility for their allocation ($h(x; i) = \langle \boldsymbol{a}_i, \boldsymbol{x}_i \rangle$), their utility for other agents' allocations, or the minimum utility over all the agents. In standard resource allocation problems, the designer selects one of these objectives from this set $\mathcal{H}$. We refer to this objective as the *benchmark objective* and denote it by h^*. Intuitively, we can think of h^* as the objective defining u_i.

By contrast, we assume that each agent may have a different preferred objective, aggregating the objectives in $\mathcal{H}$ into a single one. We will assume that the aggregation function belongs to a specific function class parameterized by θ. For example, the aggregation function may be a weighted linear combination of the objectives in $\mathcal{H}$, with weights determined by the preference $\theta \in \mathbb{R}^{|\mathcal{H}|}$.

Definition 2.2 (Multi-objective agents). A multi-objective agent has a preference θ within the space of valid preferences Θ that determines how to reconcile conflicting objectives in $\mathcal{H}$. More precisely, there exists a function $v : \mathbb{R}^{|\mathcal{H}|} \times \Theta \to \mathbb{R}$ that aggregates values for all the objectives in $\mathcal{H}$ into a scalar value based on the agent's preference θ. We use the shorthand $v_i(x)$ to denote agent i's aggregated value: $v_i(x) = v(\{h(x; i)\}_{h\in\mathcal{H}}; \theta_i)$. We also denote the value profile $(v_1(x), v_2(x), \dots, v_n(x))$ by $\boldsymbol{v}(x)$.

[1] Note, for the sake of clarity, we consider the case of divisible items and linear utilities, but it is straightforward to map our formalization to the setting where items are indivisible, or other assumptions hold on the agents' utilities.

[2] We assume that we can elicit the utility coefficients truthfully, though we will revisit this assumption in Appendix C.

Standard resource allocation problems unavoidably elicit the utility coefficients a. Although our multi-objective agent formalization captures far greater complexity in the agents' preferences, it only requires additional elicitation of θ to determine how an agent aggregates the objectives.

After associating each agent with a single objective—whether this is a benchmark objective h^* chosen by the designer or the agent's preferred objective v_i—the next step is for the designer to define an optimization problem. We assume that the designer has a *social welfare function* $f : \mathbb{R}^n \to \mathbb{R}$ that, along with the allocation constraints, defines the optimization problem. To concretely illustrate our framework, for the rest of this paper, we will use the utilitarian social welfare, which sums all the individual values with equal weight.

Using the above notions, we can now define our participatory approach, which enables agents to express preferences not only over the items but also over the objectives.

Definition 2.3 (Participatory objective design). In the participatory approach to resource allocation, a designer first associates each multi-objective individual i with a single objective function v_i by eliciting their preference θ_i over the set of possible objectives $\mathcal{H}$. The designer then maximizes $f(\boldsymbol{v}(x)) := \sum_{i=1}^{n} v_i(x)$ subject to $x \in \mathcal{X}$.

This participatory approach is in contrast to the *standard approach*, where the designer would optimize over $\sum_{i=1}^{n} h^*(x; i)$.

2.1 Loss and Gain from Participatory Objective Design

We now introduce notions of loss, gain, and individual tradeoffs incurred by moving from the standard approach, where the designer selects an objective in a top-down manner, to the participatory approach, where agents influence the overall objective in a bottom-up fashion.

Specifically, we study loss in social welfare, as measured by the benchmark objective, and gain in social welfare, as measured by preferences elicited from the participatory approach. We also look at individual tradeoffs, which consider the maximum cost to utility incurred by a single agent.

Central to these notions is the comparison of social welfare maximizing allocations under the standard and participatory approaches. Formally, let $\boldsymbol{h}^*(x)$ denote the profile of benchmark objectives $(h^*(x; 1), \dots, h^*(x; n))$, and $\boldsymbol{v}$ denote the profile of aggregated objectives of individuals $(v_1(x), \dots, v_m(x))$. We define these two optimal allocations as

$$x^* := \arg\max_{x \in \mathcal{X}} f(\boldsymbol{h}^*(x)), \tag{2}$$

$$x^\theta := \arg\max_{x \in \mathcal{X}} f(\boldsymbol{v}(x)). \tag{3}$$

We first consider the notion of loss, which pessimistically measures the potential reduction in social welfare as measured by the benchmark objective if we move to the participatory approach.

Definition 2.4 (Loss in social welfare). Suppose $h^* \in \mathcal{H}$ is the benchmark objective and agents may have multi-objective preferences. We define the loss in social welfare measured by the benchmark objective as

$$loss := f(\boldsymbol{h}^*(x^*)) - f(\boldsymbol{h}^*(x^\theta)), \tag{4}$$

where $f(\boldsymbol{h}^*(x^*))$ is the optimal social welfare, as measured by the benchmark objective, and $f(\boldsymbol{h}^*(x^\theta))$ measures the same notion of social welfare under the participatory approach.

We similarly define the gain in social welfare, which measures the potential improvement in social welfare, as measured by the participatory approach.

Definition 2.5 (Gain in social welfare). Suppose $h^* \in \mathcal{H}$ is the benchmark objective and agents may have multi-objective preferences. We define the gain in social welfare measured using the elicited preferences as

$$gain := f(\boldsymbol{v}(x^\theta)) - f(\boldsymbol{v}(x^*)). \tag{5}$$

Overall, small loss and large gain values indicate that it is favorable to move to the participatory approach as doing so does not incur much loss in social welfare, *even by the benchmark objective*, and it can result in an increase in social welfare, as measured by the participatory approach.

Our notions of loss and gain are similar to the notion of price of fairness in resource allocation. Price of fairness typically compares optimal allocations abiding by certain fairness constraints with optimal allocations determined without such constraints. In particular, the loss in this setting is closely related to existing notions of price of fairness (PoF) when $h^*(x; i) = u_i(x)$. For instance:

$$\text{Bertsimas et al. [6]'s PoF} \;=\; loss/f(\boldsymbol{u}(x^*)), \tag{6}$$

$$\text{Caragiannis et al. [16]'s PoF} \;=\; 1 + loss/f(\boldsymbol{u}(x^\theta)). \tag{7}$$

So far, we have looked at loss and gain, when we look at the overall social welfare. We may additionally consider the cost incurred by a single individual. For a benchmark objective $h^* \in \mathcal{H}$, individual tradeoffs capture the maximum relative decrease in h^* that any single individual would incur when we move to a participatory approach.

Definition 2.6 (Individual tradeoffs). For a benchmark objective $h^* \in \mathcal{H}$, the individual tradeoffs (IT) is

$$IT(\theta) := \max_i \frac{h^*(x^*; i)}{h^*(x^\theta; i)}. \tag{8}$$

A small IT suggests that no single agent experiences a significant drop in social welfare, as measured by the benchmark objective, when we move to a participatory approach, A large IT indicates that there exists an individual who may experience such a cost. Individual tradeoffs might diverge significantly from what we observe in aggregate, giving us a distinct notion of loss to consider.

2.2 Inequality-Averse Preferences as a Case Study

The above framework provides a backbone for studying participatory objective design and the tradeoffs incurred when moving from a standard, single-objective approach to a participatory, multi-objective approach. We now illustrate the potency of this framework by turning to a well-studied resource allocation problem with inequality-averse agents.

Specifically, we consider a fair division problem where agents have utilities over the items they receive, as well as the overall inequality of a given allocation. This study of inequality-averse agents is a special case of other-regarding behavior and is well-studied empirically and theoretically in behavioral economics [10, 19, 31]. The model of inequality-aversion we study draws on work by Fehr and Schmidt [31].

Let $u_i(x)$ be agent i's utility for the items, which we take as the benchmark objective h^*. Suppose the agents are additionally inequality-averse. Concretely, we let $\mathcal{H}$ be the set containing u_is, the inequality imposed by being "more wealthy" than others (advantageous inequality), and inequality incurred by being "less wealthy" than others (disadvantageous inequality).

Let $I_i^+(x)$ and $I_i^-(x)$ be agent i's perceptions of disadvantageous and advantageous inequality, respectively. For simplicity, we assume that objectives are aggregated linearly by $\theta_i := (\alpha_i, \beta_i)$, for some $\alpha_i, \beta_i \geq 0$.

$$\begin{aligned} v_i(x) &= v\left(\{u_i(x), I_i^+(x), I_i^-(x)\}; \theta_i = (\alpha_i, \beta_i)\right) \\ &= u_i(x) - \alpha_i I_i^+(x) - \beta_i I_i^-(x)\,. \end{aligned} \tag{9}$$

We follow the formulation of Fehr and Schmidt [31] for I_i^+ and I_i^-, which serve as a first-order approximation of general inequality metrics that are: (1) agent-centric, meaning they only depend on $|u_i(x) - u_{i'}(x)|$ terms for each agent i and other agents i', and (2) anonymous, meaning agent i perceives similar inequality from other agents i' and i'', if $u_{i'}(x) = u_{i''}(x)$. This social preference model by Fehr and Schmidt [31] is well-motivated and broadly-studied [27, 62].[3]

$$I_i^+(x) = \frac{1}{n-1} \sum_{i': u_{i'}(x) \geq u_i(x)} (u_{i'}(x) - u_i(x)) \tag{10}$$

$$I_i^-(x) = \frac{1}{n-1} \sum_{i': u_{i'}(x) < u_i(x)} (u_i(x) - u_{i'}(x)) \tag{11}$$

We denote the agents' profile of α_i and β_i by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. As is common in the literature, we assume that $\alpha_i \geq \beta_i > 0$, and that β_i and α_i are of the same order [48]. We examine the changes in allocation when agents exhibit slight inequality aversion, i.e., the parameters α_i and β_i are less than $1/2$ and are typically small, to capture settings where the agents' true preferred objectives do not deviate significantly from the benchmark.

Historically, social welfare calculation in resource allocation has relied on an inequality-agnostic objective for each agent, i.e., $h^*(x; i) = u_i(x)$, for all $i \in [n]$. Different mechanisms have been developed based on the particular structure of the social welfare function f to elicit agents' preferences for items and efficiently allocate resources. Nevertheless, models that do not account for other-regarding behaviors, like inequality aversion, may fail to capture agents' true preferences. Our framework presents one way to capture these preferences, so that they can influence the overall allocations. It also presents a way to evaluate how this approach compares to the standard approach using the benchmark objective.

In the rest of the paper, we consider the utilitarian social welfare and our measures of loss, gain, and individual tradeoffs when the designer's benchmark objective of $f(\boldsymbol{h}^*(x)) = f(\boldsymbol{u}(x))$ is replaced by the participatory variant which takes inequality-aversion into account.

3 LOSS IN INEQUALITY-AGNOSTIC SOCIAL WELFARE

We first study the loss in social welfare as measured by the benchmark objective when we move from the standard approach to the participatory approach. For this analysis, we first consider the case of two agents and show that the worst-case loss linearly scales with the level of inequality aversion. We further show that we can obtain tighter bounds by imposing structures on the agents' preferences.

We then consider the case of n agents, where we show that the worst-case loss is independent of the number of agents, thereby inheriting the above linear relationship with the level of inequality aversion. We then consider the case of clustered agents with similar preferences and independent agents whose utility coefficients for the items are drawn independently from the same distribution. For the case of clustered agents, we provide a possibly tighter bound when the clusters are sufficiently distinguishable. For the case of independent agents, we prove an improved upper bound that grows quadratically with the level of inequality aversion.

We provide a summary of our results in Table 1.

3.1 Two-Agent Setting

For the case of two agents, we plug in $I_i^+(x)$ and $I_i^-(x)$ into agent i's aggregated value and obtain

$$v_i(x) = u_i(x) - \alpha_i \cdot \max\{u_{-i}(x) - u_i(x), 0\} - \beta_i \cdot \max\{u_i(x) - u_{-i}(x), 0\}\,,$$

Here, $-i$ refers to the agent other than agent i. The utilitarian social welfare in this setting is

$$\begin{aligned} f(\boldsymbol{v}(x)) =& u_1(x) + u_2(x) \\ &- (\beta_1 + \alpha_2) \cdot \max\{u_1(x) - u_2(x), 0\} \\ &- (\alpha_1 + \beta_2) \cdot \max\{u_2(x) - u_1(x), 0\}\,. \end{aligned} \tag{12}$$

To begin, we characterize $x^{\alpha,\beta} = \arg\max_x f(\boldsymbol{v}(x))$ as the maximizer of social welfare under the participatory approach.[4] In essence, $x^{\alpha,\beta}$ assigns all of an item j to an agent i if her utility for the item is significantly higher than that of the other agent. What constitutes a "significant" difference is determined by a function of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and the total demand on the item, given by $a_{ij} + a_{-ij}$.

Lemma 3.1 (Two-agent solution characterization). *The social welfare-maximizing allocation for the sake of two inequality-averse agents follows*

$$x_{ij}^{\alpha,\beta} = \begin{cases} 1\,, & \frac{a_{ij} - a_{-ij}}{a_{ij} + a_{-ij}} > \Delta_i\,, \\ 0\,, & \frac{a_{ij} - a_{-ij}}{a_{ij} + a_{-ij}} < \Delta_i\,, \\ \in [0,1]\,, & \frac{a_{ij} - a_{-ij}}{a_{ij} + a_{-ij}} = \Delta_i\,, \end{cases} \tag{13}$$

[3] Under certain axioms, such as weak separability, neutrality, scale invariance, and minimal increase in payoff, the Fehr and Schmidt model and its variants are the defacto acceptable choices for modeling social preferences [58].

[4] For ease of notation, we denote this solution by $x^{\alpha,\beta}$ instead of x^{θ}.

Setting	Restriction	Upper bound on *loss* or $\mathbb{E}[loss]$
Two-agent	Unrestricted	$m\, c(\boldsymbol{\alpha}, \boldsymbol{\beta})$
	δ-similar agents	δ
	γ-dissimilar agents	Worst-case: $m\, c(\boldsymbol{\alpha}, \boldsymbol{\beta})\ \sqrt{\gamma m}$ Uniform: $m\, c(\boldsymbol{\alpha}, \boldsymbol{\beta})\ (\gamma m)^{4/3}$ [Corollary D.2]
	Independent agents	$O\Big(c^2(\boldsymbol{\alpha}, \boldsymbol{\beta})\ \sum_{j\in[m]} \kappa_j\Big)$ [Corollary E.3]
Multi-agent	Unrestricted	$\sum_{j\in[m]} c_j(\boldsymbol{\alpha}, \boldsymbol{\beta})$ [Theorem 3.4]
	K-cluster	$O\Big(K\delta \min\{n, m\}\, b + (K-1)(\gamma m)^{1/2} \sum_{j\in[m]} c_j(\boldsymbol{\alpha}, \boldsymbol{\beta})\Big)$ [Theorem 3.5]
	Independent agents	$O\Big(n\, b\ \sum_{j\in[m]} c_j^2(\boldsymbol{\alpha}, \boldsymbol{\beta})\, \kappa_j^2\Big)$ [Theorem 3.6]

Table 1: Summary of upper bounds on *loss* or $\mathbb{E}[loss]$ of allocating m items to n agents. Two-agent: $c(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is $O(\alpha_1 + \alpha_2 + \beta_1 + \beta_2)$. For γ-dissimilar agents ($\gamma m \leq 1$), we provide both a general worst-case bound and a tighter bound assuming utility coefficients are distributed uniformly. For independent agents, we introduce κ_j as a measure of how well the underlying distribution of item j's utility coefficients is spread. The bound shows a quadratic improvement in terms of $c(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Multi-agent: Roughly speaking, $c_j(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is $O(\max_i \alpha_i + \beta_i)$. For a more accurate definition of $c_j(\boldsymbol{\alpha}, \boldsymbol{\beta})$ refer to Theorem 3.4. All the provided bounds in the multi-agent setting assume no agent can get more than a b proportion of an item.

where $\Delta_i = -\Delta_{-i} \in [-(\alpha_i + \beta_{-i}), \beta_i + \alpha_{-i}]$ is a bounded slack variable. When $u_i(x^{\alpha,\beta}) \geq u_{-i}(x^{\alpha,\beta})$, we have $\Delta_i \geq 0$. In particular, if $u_i(x^{\alpha,\beta}) > u_{-i}(x^{\alpha,\beta})$, then $\Delta_i = -\Delta_{-i} = \alpha_{-i} + \beta_i$.

In general, $x_{ij}^{\alpha,\beta}$ depends on the whole matrix of utility coefficients a. However, Eq. (13) significantly simplifies the problem by characterizing $x_{ij}^{\alpha,\beta}$ as a function of utility coefficients only for item j, i.e., a_{1j} and a_{2j}, and a common bounded variable Δ_1. This characterization enables us to study the worst-case allocation for an item in isolation from the other items. As a direct application of Lemma 3.1, we next provide a worst-case bound on $loss := f(\boldsymbol{u}(x^*)) - f(\boldsymbol{u}(x^{\alpha,\beta}))$.

Theorem 3.2 (Upperbounded loss in an unrestricted two-agent setting). *Without loss of generality, suppose that $u_1(x^*) \geq u_2(x^*)$. We can upperbound loss as a sum over the terms per item: $loss \leq \sum_{j:a_{1j}>a_{2j}} loss_j$, where*

$$loss_j := (a_{1j} - a_{2j}) \cdot \mathbb{1}\left\{a_{2j} \geq r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}\right\}, \tag{14}$$

and $r_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) := (1 - \beta_i - \alpha_{-i})/(1 + \beta_i + \alpha_{-i})$. For a fixed $\boldsymbol{a}_1$, $loss_j$ will be maximized when $a_{2j} = r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}$. Without any restriction on $\boldsymbol{a}_2$, this gives us a worst-case upper bound of

$$loss \leq c_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) \sum_{j:a_{1j}>a_{2j}} a_{1j} = O(m(\beta_1 + \alpha_2)), \tag{15}$$

where $c_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) := 1 - r_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) = 2(\beta_i + \alpha_{-i})/(1 + \beta_i + \alpha_{-i}) = \Theta(\beta_i + \alpha_{-i})$.

For a fixed a_{1j}, Fig. 1 shows $loss_j$ as a function of a_{2j}. When $a_{2j} \geq a_{1j}$, the item goes to agent 2 and $j \notin \mathcal{J}_1$. We can therefore assume that $loss_j = 0$. For $a_{2j} < a_{1j}$, as long as $\Delta a_j \leq (\beta_1 + \alpha_2)(a_{1j} + a_{2j})$ or equivalently $a_{2j} \geq r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})$, $loss_j$ scales linearly with a_{2j}. The maximum of $loss_j$ occurs at $a_{2j} = r(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}$. At this point, $loss_j = c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}$, where $c_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) := 1 - r_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$. For $a_{2j} < r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}$, inequality aversion is not strong enough to change the allocation. Note, for small α_is and β_is, we have $r_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \Theta(1)$ and $c_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \Theta(\beta_i + \alpha_{-i})$.

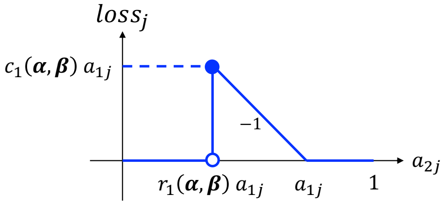

Figure 1: $loss_j$ of Eq. (14) as a function of a_{2j} when a_{1j} is fixed.

The worst-case upper bound of Eq. (15), or equivalently $loss \leq c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, \|\boldsymbol{a}_1\|_1$, is observed when the agents have aligned preferences and $\boldsymbol{a}_2$ is a down-scaled version of $\boldsymbol{a}_1$. Next, we investigate whether we can avoid this worst-case scenario and attain better guarantees by imposing further restrictions on the agents' preferences. In particular, we consider three cases: similar agents, dissimilar agents, and independent agents. We briefly discuss these cases in the following and refer the reader to Appendices D and E for the complete analysis.

Similar Agents. Since our measure of inequality depends on the absolute difference of utilities, a natural choice to impose similarity is to bound $\|\boldsymbol{a}_1 - \boldsymbol{a}_2\|_1$. We say that agents are δ-similar if $\|\boldsymbol{a}_1 - \boldsymbol{a}_2\|_1 \leq \delta$. For δ-similar agents, an immediate result of Eq. (14) is $loss \leq \sum_j \Delta a_j \leq \delta$. This bound is tight up to a factor of 2 (Proposition F.1).

Dissimilar Agents. We say agents are γ-dissimilar if $\langle \boldsymbol{a}_1, \boldsymbol{a}_2 \rangle \leq \gamma \|\boldsymbol{a}_1\|_1 \|\boldsymbol{a}_2\|_1$. For agents that are γ-dissimilar, the maximal loss, which corresponds to $\boldsymbol{a}_2$ aligning with $\boldsymbol{a}_1$, occurs only if $\gamma \geq \|\boldsymbol{a}_1\|_2^2 / \|\boldsymbol{a}_1\|_1^2$. This ratio is lowerbounded by $1/m$, which we obtain using Jensen's inequality. Hence, for γ-dissimilar agents, when $\gamma m \ll 1$, we anticipate a significantly smaller loss compared to the worst-case scenario. Intuitively, the dissimilarity constraint prevents the alignment of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ for many items with large a_{1j}, resulting in little competition for the items between the agents.

We upperbound $loss_j$ for the general case of γ-dissimilar agents in Theorem D.1. As the theorem states, an item will not contribute more than $O\big(c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\cdot\max\left\{(\gamma m)^{1-\tau},(\gamma m)^{\tau}\right\}\big)$ to the loss, where $\tau\in[0,1]$ is an arbitrary constant. Without any assumption on how utility coefficients are distributed, for the choice of $\tau=0.5$, an immediate result of this theorem is $loss=O\big(m\,c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\sqrt{\gamma m}\big)$, which is an informative bound only if $\gamma m<1$. The theorem further states that the loss will be mainly realized from items with $a_{1j}<O\big((\gamma m)^{\tau}\big)$. Therefore, introducing a prior distribution over the a_{1j}s can improve our bounds. For example, assuming $a_{1j}\sim unif(0,1)$, for $\tau=0.5$, no more than $O\big(\sqrt{\gamma m}\big)$ proportion of items will contribute to the loss in the worst-case scenario, resulting in $loss=O\big(m\,c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,\gamma m\big)$. By carefully selecting our τ, we show an improved bound of $O\big(m\,c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,(\gamma m)^{4/3}\big)$ in Corollary D.2.

Independent Agents. Let a_{1j} and a_{2j} be independently and identically distributed according to distribution g_j. Here, g_j is the density function with the corresponding cumulative distribution function G_j. We do not make any assumption on the independence of items. Looking at Fig. 1, recall that $loss_j\le c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}$, and a positive $loss_j$ occurs only when a_{2j} falls within the interval $[r_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j},a_{1j}]$. Since agents are independent, for a well-spread distribution g_j, we can argue

$$\Pr(a_{2j}\in[r_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j},a_{1j}])=O(c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j})\,. \tag{16}$$

Hence, we expect $\mathbb{E}[loss_j]=O\big(c_1^2(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}^2\big)$, and consequently $\mathbb{E}[loss]=O\big(c_1^2(\boldsymbol{\alpha},\boldsymbol{\beta})\|\mathbf{a}_1\|_2^2\big)=O\big(m\,c_1^2(\boldsymbol{\alpha},\boldsymbol{\beta})\big)$. Since we do not know which agent is doing better a priori when preferences are random, we can use $c_m(\boldsymbol{\alpha},\boldsymbol{\beta})=\max_i c_i(\boldsymbol{\alpha},\boldsymbol{\beta})$ in our bounds: $\mathbb{E}[loss]=O\big(m\,c_m^2(\boldsymbol{\alpha},\boldsymbol{\beta})\big)$. This quadratic bound is a significant improvement over the worst-case $O(m\,c_m(\boldsymbol{\alpha},\boldsymbol{\beta}))$ bound, though it only holds if g_j is well-spread. Refer to Proposition F.2 for a counterexample.

In Appendix E, we provide bounds for the loss with general distributions. Of particular interest is Corollary E.3 where we introduce $\kappa_j:=\sup_{0\le\bar a\le 1}\bar a\,g_j(\bar a)/G_j(\bar a)$ as a measure of how well-spread distribution g_j is. We show that having bounded κ_j for every item j is sufficient to bound the expected loss quadratically in $c_m(\boldsymbol{\alpha},\boldsymbol{\beta})$. For example, this holds for the uniform distribution, which has a κ value of 1.

3.2 Multi-Agent Setting

We now consider the general case of n agents. Similar to the two-agent case, we begin with a characterization of the optimal allocation $x^{\alpha,\beta}$, allowing us to analyze items independently. In this setting, we add the constraint that no agent can get more than a b_j portion of item j. An example of such constraint is assigning students to a class of size $1/b_j$, where no single student can occupy more than one seat. For $n=2$ and $b_j=1$, this is equivalent to the problem we studied in the two-agent setting.

Lemma 3.3 (Multi-agent solution characterization). *Suppose there are m items and n inequality-averse agents, where we would like to maximize social welfare as measured by the multi-objective preferences ($x^{\alpha,\beta}$), subject to the constraint that the share of each agent from any item j does not exceed b_j.*

For any pair of agents i and k, if agent i has not received her maximum share from item j (i.e., $x_{ij}^{\alpha,\beta}<b_j$), then agent k can only get a share of j (i.e., $x_{kj}^{\alpha,\beta}>0$) if $a_{kj}\ge r_{ik}(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{ij}$, where

$$r_{ik}(\boldsymbol{\alpha},\boldsymbol{\beta})=\frac{1-\beta_i-(\|\boldsymbol{\alpha}\|_1-\alpha_i)/(n-1)}{1+\alpha_k+(\|\boldsymbol{\beta}\|_1-\beta_k)/(n-1)}\,. \tag{17}$$

If, in the limit of many agents, $\frac{1}{n}\sum_i\alpha_i\to\bar\alpha$ and $\frac{1}{n}\sum_i\beta_i\to\bar\beta$, then $r_{ik}(\boldsymbol{\alpha},\boldsymbol{\beta})\to\frac{1-\beta_i-\bar\alpha}{1+\alpha_k+\bar\beta}$.

Eq. (17) indicates that in the reallocation of an item from agent i to k, both the society view regarding inequality represented by $\|\boldsymbol{\alpha}\|_1$ and $\|\boldsymbol{\beta}\|_1$, and inequality aversion of the individuals involved play a role. For instance, a society moderately averse to (disadvantageous) inequality (moderate $\|\boldsymbol{\alpha}\|_1$) facilitates reallocation even when the better-off agent is not averse to (advantageous) inequality (small β_i). As an immediate result of Lemma 3.3 we can upperbound the loss in the most general case:

Theorem 3.4 (Upperbounded loss in an unrestricted multi-agent setting). *Suppose $1/b_j\in\mathbb{N}$ and let top_j be the set of $1/b_j$ agents with the highest a_{ij}. For the top_j agents, define average utility coefficient $a_{top_j}:=b_j\sum_{i\in top_j}a_{ij}$, and maximum level of advantageous inequality aversion $\beta_{m,top_j}:=\max_{i\in top_j}\beta_i$. Further, denote the maximum level of disadvantageous inequality aversion by $\alpha_m:=\max_i\alpha_i$.*

Defining

$$c_j(\boldsymbol{\alpha},\boldsymbol{\beta}):=\frac{\beta_{m,top_j}+\alpha_m+(\|\boldsymbol{\alpha}\|_1+\|\boldsymbol{\beta}\|_1)/(n-1)}{1+\alpha_m+\|\boldsymbol{\beta}\|_1/(n-1)}\,, \tag{18}$$

we can upperbound the loss as a sum over terms per item:

$$loss\le\sum_j loss_j:=\sum_j c_j(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{top_j}. \tag{19}$$

If, in the limit of many agents, $\frac{1}{n}\sum_i\alpha_i\to\bar\alpha$ and $\frac{1}{n}\sum_i\beta_i\to\bar\beta$, then

$$loss\le\sum_j\frac{\beta_{m,top_j}+\alpha_m+\bar\alpha+\bar\beta}{1+\alpha_m+\bar\beta}a_{top_j}\,. \tag{20}$$

This result is in stark contrast with similar studies of the price of fairness. For instance, Caragiannis et al. [16] show in the allocation of divisible goods enforcing proportionality or envy-freeness, the price of fairness grows at least with $\sqrt{n}$. This is equivalent to a relative loss of $\Omega(1-1/\sqrt{n})$ and approaches 1 asymptotically, implying that the fair allocation becomes inefficient. In contrast, we bound the relative loss in our setting by $\max_j c_j(\boldsymbol{\alpha},\boldsymbol{\beta})$, which is a constant.

In simple terms, a loss as severe as Eq. (19) can occur if, for each item j, there is a group o_j consisting of at least $|top_j|$ worse-off agents with closely aligned down-scaled interests to those of top_j. Like the two-agent setting, we ask whether having further structure on agents' utility coefficients can help avoid the worst-case loss. We consider two cases: clustered agents and independent agents.

Clustered Agents. Suppose that our agents are in K clusters. We denote the set of agents within cluster $q\in[K]$ by C_q. For each cluster q, define the cluster's upper and lower representative coefficients: $\bar a_{qj}:=\max_{i\in C_q}a_{ij}$, $\underline{a}_{qj}:=\min_{i\in C_q}a_{ij}$. We assume that clusters are easily distinguishable, i.e., agents within a cluster are similar to one another and dissimilar agents in other clusters. More precisely,

(1) Within each cluster q, suppose $\|\bar{\boldsymbol{a}}_q - \underline{\boldsymbol{a}}_q\|_1 \leq \delta$. We call δ the radius of the cluster.
(2) Between distinct clusters q and q', suppose $\bar{\boldsymbol{a}}_q$ and $\boldsymbol{a}_{i'}$ are γ-dissimilar for every $i' \in C_{q'}$, i.e., $\langle \bar{\boldsymbol{a}}_q, \boldsymbol{a}_{i'} \rangle \leq \gamma \|\bar{\boldsymbol{a}}_q\|_1 \|\boldsymbol{a}_{i'}\|_1$.

For small values of δ and γ, this structure enables us to directly apply findings from the two-agent setting and derive tighter bounds.

Theorem 3.5 (Upperbounded loss in clustered agents setting). *Suppose each agent i belongs to one of K distinct clusters. Clusters have a radius of δ and between clusters dissimilarity of γ, where $\gamma m \leq 1$. Each agent's share from an item is bounded by b, where we assume that $1/b \in \mathbb{N}$ for the sake of simplicity. The expected loss is bounded by*

$$\mathbb{E}[loss] = O\Big(\delta\, b\, K\, \min\{n, m\} + (K-1)\,(\gamma m)^{\frac{1}{2}} \sum_j c_j(\boldsymbol{\alpha}, \boldsymbol{\beta})\Big). \tag{21}$$

If, for $i \in top_j$, the utility coefficients a_{ij} are best explained by $Beta(s, 1)$, the exponent of the (γm) term improves to $(s+1)^2/(s+2)$.

For a small δ, the latter term of Eq. (21) is dominant. This only improves our bound beyond the unrestricted bound of Theorem 3.4 if the number of clusters is small and clusters are sufficiently distinguishable.

Independent Agents. Consider $a_{ij} \sim g_j$ independently for each agent i, but note that preferences may not be independent across items. The following theorem demonstrates that if b_j is sufficiently small, with the number of winners for item j (i.e., $1/b_j$) being comparable to n, the loss can be quadratically bounded in the level of inequality aversion, irrespective of the number of agents.

Theorem 3.6. *Suppose each agent i's utility coefficients a_{ij} are drawn independently from the distribution g_j, with a corresponding cumulative distribution G_j. Suppose further that $n \longrightarrow \infty$, but $n\, b_j$ is bounded. For $G_j \in C^1$, define $\kappa_j := \max_{\bar{a}>0} \bar{a}\, g_j(\bar{a})/G_j(\bar{a})$.*

If $\kappa_j \leq 1/(b_j\, n\, c_j(\boldsymbol{\alpha}, \boldsymbol{\beta}))$ for every j, then

$$\mathbb{E}[loss] = O\Big(\sum_j c_j^2(\boldsymbol{\alpha}, \boldsymbol{\beta})\, \kappa_j^2\, b_j n\Big). \tag{22}$$

4 GAIN IN INEQUALITY-AVERSE SOCIAL WELFARE

In this section, we study the gain in social welfare, as measured using elicited preferences over objectives, from moving to a participatory approach. We do so by examining the relationship between loss and gain using the gain-to-loss ratio. A high gain-to-loss ratio indicates a relatively higher benefit to moving to a participatory approach than the loss in social welfare, as measured by the benchmark objective.

Specifically, we ask: Is $gain/loss$ bounded in general? If not, could we bound this ratio using further assumptions on agents' preferences?

4.1 Two-Agent Setting

We start from a two-agent setting and present a lower bound on the gain. The ratio of this lowerbounded gain over the upperbounded loss from the previous section will provide a lower bound on $gain/loss$.

Proposition 4.1. *Without loss of generality, suppose $u_1(x^*) \geq u_2(x^*)$. We can lowerbound gain as a sum over terms per item: $gain \geq \sum_{j: a_{1j} > a_{2j}} gain_j$, where*

$$gain_j := [(1+\beta_1+\alpha_2)a_{2j} - (1-\beta_1-\alpha_2)a_{1j}] \cdot \mathbb{1}\{a_{2j} > r_1(\Delta_1)\, a_{1j}\}, \tag{23}$$

and $r_1(\Delta_1) := (1-\Delta_1)/(1+\Delta_1)$ for $\Delta_1 \in [0, \beta_1 + \alpha_2]$ as introduced in Lemma 3.1.

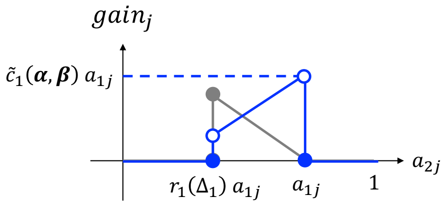

Figure 2: $gain_j$ from Eq. (23), as a function of a_{2j} with a_{1j} kept fixed, is plotted in blue. $loss_j$ is also depicted in gray.

We now examine $gain_j$ in Eq. (23) as a function of a_{2j} (Fig. 2). For a fixed a_{1j}, the gain is increasing in a_{2j} as long as $a_{2j} < a_{1j}$. The maximum gain of $2(\beta_1+\alpha_2)\, a_{1j}$ will be realized as $a_{2j} \rightarrow^{-} a_{1j}$. Defining

$$\tilde{c}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) := 2(\beta_1 + \alpha_2), \tag{24}$$

the maximum gain can be written as $\tilde{c}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}$. In Fig. 2, we have depicted $gain_j$ along with an upper bound on the loss in gray (duplicating Fig. 1). There are especially two interesting regimes in this figure:

(1) For $a_{2j} \rightarrow^{-} a_{1j}$, $gain_j/loss_j$ takes very large values. In this case, we can also expect large values for $gain/loss$.
(2) For $a_{2j} \rightarrow^{+} r_1(\Delta_1)\, a_{1j}$, the ratio of $gain_j/loss_j$ has small values. Specifically, when $\Delta_1 \rightarrow \beta_1 + \alpha_2$, one can verify $gain_j$ in Eq. (23) goes to zero. But for Δ_1 far from $\beta_1 + \alpha_2$, $gain_j/loss_j$ can be lowerbounded meaningfully above 0.

The next two propositions formally state the above observations.

Proposition 4.2. *Suppose the agents are δ-similar, i.e., $\|\boldsymbol{a}_1 - \boldsymbol{a}_2\|_1 \leq \delta$, and initially $u_1(x^*) > u_2(x^*)$ with no tie for any item. Then*

$$\frac{gain}{loss} \geq \frac{\tilde{c}_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, f(\boldsymbol{u}(x^*))}{2\delta} - \frac{\tilde{c}_1(\alpha, \beta)}{c_1(\Delta_1)} = \Omega\Big(\frac{m\, \tilde{c}_1(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\delta}\Big). \tag{25}$$

Proposition 4.3. *Suppose $u_1(x^*) > u_2(x^*)$ and $loss > 0$. Then, $gain/loss \geq (\beta_1 + \alpha_2)/\Delta_1 - 1$, where $\Delta_1 \leq \beta_1 + \alpha_2$.*

Intuitively, even when the gain-to-loss ratio is small, it can still be meaningfully above 0.

4.2 Multi-Agent Setting

The large gain-to-loss ratio is not limited to the two-agent setting. Consider the allocation of n goods to n agents with $b_j = 1$ and the following utility coefficients:

$$a_{ij} = \begin{cases} 1\,, & i = 1\,, \\ 1-\epsilon\,, & i = j > 1\,, \\ r_{1i}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \epsilon & \text{o.w.} \end{cases} \tag{26}$$

As $\epsilon \to^{+} 0$, it is straightforward to see $x^*_{ij} = \mathbb{1}\{i = 1\}$ and $x^{\alpha,\beta}_{ij} = \mathbb{1}\{i = j\}$. Thus,

$$\mathit{loss} = (n-1)\epsilon\,, \quad \mathit{gain} = -(n-1)\epsilon + (2-\epsilon)\sum_{i>1}(\beta_1 + \alpha_i)\,. \tag{27}$$

In the limit of $\epsilon \to^{+} 0$, we have $\mathit{loss} \to 0$ and $\mathit{gain} \to 2\sum_{i>1}(\beta_1 + \alpha_i)$. Therefore, $\mathit{gain}/\mathit{loss} \to \infty$.

Note, although a dissimilarity constraint can potentially upperbound the gain, it cannot upperbound the gain-to-loss ratio. The following proposition shows that, under weak conditions, even extreme dissimilarity cannot guarantee a bounded gain-to-loss ratio.

Proposition 4.4. *For any $\gamma > 0$, suppose there exist n agents who are pairwise γ-dissimilar. If there exists an agent i for whom either $\beta_i > 0$ or $\alpha_k > 0$ for some $k \neq i$, then $\mathit{gain}/\mathit{loss} \to \infty$.*

5 INDIVIDUAL COSTS OF INEQUALITY AVERSION

The above notions of loss and gain consider aggregate outcomes, but we may also care about the worst-off individual. To this end, we wish to bound the worst-case *individual tradeoffs*

$$IT(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \max_i \frac{u_i(x^*)}{u_i(x^{\alpha,\beta})}\,. \tag{28}$$

In this section, we show that $IT(\boldsymbol{\alpha}, \boldsymbol{\beta})$ can approach n, even for small levels of inequality aversion. This high individual tradeoff often stems from competition for items when agents' preferences are similar, suggesting this may result from the benchmark allocation rather than the ill-suitedness of the inequality-averse allocation. We give an example where $IT(\boldsymbol{\alpha}, \boldsymbol{\beta}) \approx n$, yet $f(\boldsymbol{u}(x^*)) \approx f(\boldsymbol{u}(x^{\alpha,\beta}))$. However, we can give slightly more optimistic bounds on $IT(\boldsymbol{\alpha}, \boldsymbol{\beta})$ under mild assumptions on which agent gives up the most utility.

We start by examining individual tradeoffs in the two-agent, two-good setting. If utilities are normalized. i.e., $\sum_j a_{ij} = 1$ for all i, we can write the utility profile with coefficients

$$a = \begin{bmatrix} a_1 & 1-a_1 \\ a_2 & 1-a_2 \end{bmatrix}\,, \tag{29}$$

where a_1 (resp. a_2) is how much agent 1 (resp. agent 2) values good 1 and $1-a_1$ (resp. $1-a_2$) is how much agent 1 (resp. agent 2) values good 2. We additionally assume $a_{ij} > 0$ for all i, j, so that there exists a complete allocation x, i.e., $\sum_i x_{ij} = 1$ for all $j \in [m]$, with no inequality, $I^+_i(x) = I^-_i(x) = 0$ for all i.

Without loss of generality, suppose $a_1 > a_2$ and $a_1 > 1 - a_2$. In this setting, x^* is the identity, giving all of good 1 to agent 1 and all of good 2 to agent 2, yet $u_1(x^*) > u_2(x^*)$. Moreover, by the characterization given in Lemma 3.1, assuming inequality aversion is sufficiently strong, i.e., $\beta_1 + \alpha_2 \geq (a_1 - a_2)/(a_1 + a_2)$, then we have

$$x^{\alpha,\beta} = x^* + \left(\frac{a_1 - (1-a_2)}{a_1 + a_2}\right)\begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}\,. \tag{30}$$

This allocation gives enough of good 1 to agent 2 to equalize the agents' utilities. As agent 1 will be giving up some of her allocations and agent 2 will be receiving goods from agent 1 relative to x^*, agent 1 incurs the maximum cost and

$$IT(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \max_i \frac{u_i(x^*)}{u_i(x^{\alpha,\beta})} = \frac{u_1(x^*)}{u_1(x^{\alpha,\beta})}\,, \tag{31}$$

with

$$a = \begin{bmatrix} 1-\epsilon_1 & \epsilon_1 \\ 1-\epsilon_2 & \epsilon_2 \end{bmatrix}\,,$$

for $0 < \epsilon_1 < \epsilon_2$ sufficiently small. In particular, the closed form of $IT(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is

$$\max_i \frac{u_i(x^*)}{u_i(x^{\alpha,\beta})} = \frac{a_1}{a_1 - \frac{a_1 - (1-a_2)}{a_1 + a_2}} = \frac{2 - \epsilon_1 - \epsilon_2 + \epsilon_1^2 + \epsilon_1\epsilon_2 - 2\epsilon_1}{1 + \epsilon_1^2 + \epsilon_1\epsilon_2 - 2\epsilon_1}\,,$$

which approaches 2 as $\epsilon_1, \epsilon_2 \to 0$.

We now move to the multi-agent setting. In Proposition 5.1, we show that in the worst case, individual tradeoff scales linearly with the number of agents n.

Proposition 5.1. *There exist n agents with normalized utility coefficients for whom $IT(\boldsymbol{\alpha}, \boldsymbol{\beta}) \to n$.*

The above bound gets arbitrarily close to n when all agents have similar preferences. However, allocating "almost all of the utility" to a single agent is optimal. Inequality aversion, in this case, leads to a significant drop for the single agent despite having a small loss in social welfare. This leads us to ask what happens if agents are sufficiently dissimilar. We bound $\max_i \frac{u_i(x^*)}{u_i(x^{\alpha,\beta})}$ by bounding $u_i(x^{\alpha,\beta})$ by a 0-inequality allocation x^e. This bound enables us to understand $IT(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in terms of the agent who receives the highest and lowest utilities under x^* (Lemma G.1).

Proposition 5.2. *Suppose $\boldsymbol{\alpha} = \boldsymbol{\beta} = \alpha\mathbf{1}$ and let x^e be an allocation such that $\sum_i I^+_i(x^e) + I^-_i(x^e) = 0$. Moreover, let $i \in \arg\max_{i'} \frac{u_{i'}(x^*)}{u_{i'}(x^{\alpha,\beta})}$ and $\mathcal{J}^\alpha := \{i' : u_{i'}(x^{\alpha,\beta}) \geq u_i(x^{\alpha,\beta})\}$. If $\sum_i I^+_i(x^{\alpha,\beta}) + I^-_i(x^{\alpha,\beta}) \geq \sum_{i' \in \mathcal{J}^\alpha}(u_{i'}(x^{\alpha,\beta}) - u_i(x^{\alpha,\beta}))$. Then*

$$\frac{u_i(x^*)}{u_i(x^{\alpha,\beta})} \leq \max_{i'} \frac{u_{i'}(x^*)}{u_{i'}(x^e)} \leq \frac{\max_{i'} u_{i'}(x^*)}{\min_{i'} u_{i'}(x^*)}. \tag{32}$$

This bound lends itself to a more straightforward interpretation, as we now have a bound characterized by the best- and worst-off agents according to $\boldsymbol{u}(x^*)$. Consequently, if the worst-off person in the x^* allocation has positive utility, we can bound individual tradeoffs based on x^*.

6 DISCUSSION AND CONCLUSION

In this paper, we study the impact of eliciting inequality preferences from inequality-averse agents on resource allocation. We upperbound the loss to inequality-agnostic welfare the principal incurs by eliciting inequality aversion, and upperbound the gain to inequality-averse welfare by eliciting such preferences. In general, these bounds are linear in the inequality aversion α, though the bounds can be tightened with certain assumptions on the structure of agents' preferences. Moreover, we show the largest tradeoff that any one agent might incur to their inequality-agnostic utility from inequality aversion can be arbitrarily bad even under stronger assumptions on preferences, growing linearly in the number of agents.

In our work, we hope to encourage further exploration of preference elicitation can be used to inform bottom-up approaches to resource allocation. We assume in inequality-averse allocation the agents are able to communicate their inequality aversion preferences. There may be cases where a principal can elicit more granular information such as a partial ranking from agents to estimate α_i values in settings where agents cannot communicate them exactly. Extending our results to the allocation of indivisible goods and other resource allocation problems will require careful consideration. Finally, the tradeoffs noted in this work, which we explore from a largely theoretical lens, may point to and provide a tool for understanding empirical behavior in resource allocation settings. For instance, in resource allocation settings that result in surprising or undesirable outcomes in practice, we wish to see if we can evaluate if these outcomes are caused by the planner failing to incorporate inequality-aversion or other social preferences in the utility model.

ETHICAL CONSIDERATIONS, POSITIONALITY, AND REFLECTIONS ON ADVERSE IMPACTS

This work is primarily a theoretical proof-of-concept to understand the potential for preference elicitation to enable a bottom-up approach to objective design. This is concerned with questions related to the distribution of power and, in particular, how we can design objectives through a participatory approach.

The motivation for modeling inequality-averse agents comes from the behavioral economics literature [10, 31]. Of course, the true modeling of participant preferences is far more nuanced than what this model captures. Consequently, trade-offs of how granularly participants can and should express community-level preferences should be studied in future work. We believe the results of this paper—which show potential inefficiencies in standard approaches compared to our participatory approach—provide further motivation for this line of work.

While we hope this work leads to a deeper examination of bottom-up approaches for designing objectives, we also acknowledge that such approaches might be susceptible to adversarial participants, such as spoofing attacks. We encourage future research to scrutinize this potential. Finally, the participatory approach we present in this work is one of many such approaches and may not be appropriate or efficient, depending on the underlying sets of objectives. We encourage future work to examine this broader space of theoretical possibilities and empirically evaluate these frameworks in practice.

ACKNOWLEDGMENTS

RA was supported by a grant from the Andrew Carnegie Fellowship Program. JF was supported by a grant by the National Science Foundation under Award No. 2202898. The authors would like to thank Andreas Bjerre-Nielson, Yiling Chen, Richard Cole, Moritz Hardt, Max Kasy, and Evi Micha for extensive feedback and discussions on earlier versions of this work.

REFERENCES

[1] Rediet Abebe, Jon Kleinberg, and David C Parkes. 2017. Fair Division via Social Comparison. Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems. , 281–289 pages.

[2] Jonas Agell and Per Lundborg. 1995. Fair wages in the open economy. , 335–351 pages.

[3] James Alm, Isabel Sanchez, Ana De Juan, et al. 1995. Economic and noneconomic factors in tax compliance. *KYKLOS-BERNE-* 48 (1995), 3–3.

[4] Mark Bedaywi, Bailey Flanigan, Mohamad Latifian, and Nisarg Shah. 2023. The distortion of public-spirited participatory budgeting. https://drive.google.com/file/d/1wOQ377OW-jJOnjigBMnHcs0jmzk7Ckxa/view.

[5] Joyce Berg, John Dickhaut, and Kevin McCabe. 1995. Trust, reciprocity, and social history. *Games and economic behavior* 10, 1 (1995), 122–142.

[6] Dimitris Bertsimas, Vivek F Farias, and Nikolaos Trichakis. 2011. The price of fairness. *Operations research* 59, 1 (2011), 17–31.

[7] Felix Bierbrauer and Nick Netzer. 2016. Mechanism design and intentions. *Journal of Economic Theory* 163 (2016), 557–603.

[8] Felix Bierbrauer, Axel Ockenfels, Andreas Pollak, and Désirée Rückert. 2017. Robust mechanism design and social preferences. *Journal of Public Economics* 149 (2017), 59–80. https://doi.org/10.1016/j.jpubeco.2017.03.003

[9] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the People? Opportunities and Challenges for Participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Arlington, VA, USA) *(EAAMO '22)*. Association for Computing Machinery, New York, NY, USA, Article 6, 8 pages. https://doi.org/10.1145/3551624.3555290

[10] Gary E Bolton and Axel Ockenfels. 2000. ERC: A theory of equity, reciprocity, and competition. *American economic review* 90, 1 (2000), 166–193.

[11] C. Boutilier. 2012. Eliciting Forecasts from Self-interested Experts: Scoring Rules for Decision Makers. http://www.cs.toronto.edu/~cebly/Papers/Boutilier_aamas12.pdf

[12] Steven J Brams, Michael A Jones, and Christian Klamler. 2007. Better ways to cut a cake-revisited.

[13] Steven J Brams, Michael A Jones, Christian Klamler, et al. 2006. Better ways to cut a cake. *Notices of the AMS* 53, 11 (2006), 1314–1321.

[14] Steven J Brams and Alan D Taylor. 1995. An envy-free cake division protocol. *The American Mathematical Monthly* 102, 1 (1995), 9–18.

[15] Eric Budish. 2011. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy* 119, 6 (2011), 1061–1103.

[16] Ioannis Caragiannis, Christos Kaklamanis, Panagiotis Kanellopoulos, and Maria Kyropoulou. 2012. The efficiency of fair division. *Theory of Computing Systems* 50 (2012), 589–610.

[17] Ioannis Caragiannis, David Kurokawa, Hervé Moulin, Ariel D. Procaccia, Nisarg Shah, and Junxing Wang. 2019. The Unreasonable Fairness of Maximum Nash Welfare. *ACM Trans. Econ. Comput.* 7, 3, Article 12 (sep 2019), 32 pages. https://doi.org/10.1145/3355902

[18] Ioannis Caragiannis, Swaprava Nath, Ariel D Procaccia, and Nisarg Shah. 2017. Subset selection via implicit utilitarian voting. *Journal of Artificial Intelligence Research* 58 (2017), 123–152.

[19] Gary Charness and Matthew Rabin. 2002. Understanding social preferences with simple tests. *The quarterly journal of economics* 117, 3 (2002), 817–869.

[20] Po-An Chen and David Kempe. 2008. Altruism, selfishness, and spite in traffic routing. In *Proceedings of the 9th ACM Conference on Electronic Commerce* (Chicago, Il, USA) *(EC '08)*. Association for Computing Machinery, New York, NY, USA, 140–149. https://doi.org/10.1145/1386790.1386816

[21] Yann Chevaleyre, Paul E Dunne, Ulle Endriss, Jérôme Lang, Michel Lemaitre, Nicolas Maudet, Julian Padget, Steve Phelps, Juan A Rodrígues-Aguilar, and Paulo Sousa. 2005. Issues in multiagent resource allocation. Informatica.

[22] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences.

[23] Yuga Cohler, John Lai, David Parkes, and Ariel Procaccia. 2011. Optimal envy-free cake cutting. , 626–631 pages.

[24] Richard Cole and Yixin Tao. 2021. On the existence of Pareto efficient and envy-free allocations. *Journal of Economic Theory* 193 (2021), 105207.

[25] Dinky Daruvala. 2010. Would the right social preference model please stand up! *Journal of Economic Behavior & Organization* 73, 2 (2010), 199–208.

[26] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Boston, MA, USA) *(EAAMO '23)*. Association for Computing Machinery, New York, NY, USA, Article 37, 23 pages. https://doi.org/10.1145/3617694.3623261

[27] Dirk Engelmann and Martin Strobel. 2004. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American economic review* 94, 4 (2004), 857–869.

[28] Ernst Fehr and Urs Fischbacher. 2002. Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *The economic journal* 112, 478 (2002), C1–C33.

[29] Ernst Fehr and Urs Fischbacher. 2003. The nature of human altruism. *Nature* 425, 6960 (2003), 785–791.

[30] Ernst Fehr and Urs Fischbacher. 2004. Third-party punishment and social norms. *Evolution and human behavior* 25, 2 (2004), 63–87.

[31] Ernst Fehr and Klaus M Schmidt. 1999. A theory of fairness, competition, and cooperation. *The quarterly journal of economics* 114, 3 (1999), 817–868.

[32] Ernst Fehr and Klaus M Schmidt. 2006. The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity* 1 (2006), 615–691.

[33] Jessie Finocchiaro, Roland Maio, Faidra Monachou, Gourab K Patro, Manish Raghavan, Ana-Andreea Stoica, and Stratis Tsirtsis. 2021. Bridging machine learning and mechanism design towards algorithmic fairness. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. , 489–503 pages.

[34] Bailey Flanigan, Ariel D Procaccia, and Sven Wang. 2022. A Bit of Altruism Can Curb Distortion. http://procaccia.info/wp-content/uploads/2022/08/altruism.pdf

[35] Rafael M. Frongillo and Ian A. Kash. 2019. General Truthfulness Characterizations Via Convex Analysis. arXiv:1211.3043 [cs.GT]

[36] Ali Ghodsi, Matei Zaharia, Benjamin Hindman, Andy Konwinski, Scott Shenker, and Ion Stoica. 2011. Dominant resource fairness: Fair allocation of multiple resource types.

[37] Jason D Hartline. 2013. Mechanism design and approximation. Book draft.

[38] Zoë Hitzig. 2020. The normative gap: mechanism design and ideal theories of justice. *Economics & Philosophy* 36, 3 (2020), 407–434.

[39] Borja Ibarz, Jan Leike, Tobias Pohlen, Geoffrey Irving, Shane Legg, and Dario Amodei. 2018. Reward learning from human preferences and demonstrations in atari. Advances in neural information processing systems.

[40] Hailey Joren, Chirag Nagpal, Katherine A Heller, and Berk Ustun. 2023. Participatory systems for personalized prediction. Neural Information Processing Systems.

[41] Ralph L Keeney, Howard Raiffa, and Richard F Meyer. 1993. Decisions with multiple objectives: preferences and value trade-offs. Cambridge university press.

[42] Frank Kelly. 1997. Charging and rate control for elastic traffic. *European transactions on Telecommunications* 8, 1 (1997), 33–37.

[43] Bogdan Kulynych, David Madras, Smitha Milli, Inioluwa Deborah Raji, Angela Zhou, and Richard Zemel. 2020. Participatory approaches to machine learning. International Conference on Machine Learning Workshop.

[44] Nicolas S. Lambert and Yoav Shoham. 2009. Eliciting truthful answers to multiple-choice questions. Proceedings of the 10th ACM conference on Electronic commerce. , 109–118 pages.

[45] John Ledyard. 1997. Public Goods: A Survey of Experimental Research.

[46] Shengwu Li. 2017. Obviously strategy-proof mechanisms. *American Economic Review* 107, 11 (2017), 3257–3287.

[47] Deirdre K. Mulligan, Joshua A. Kroll, Nitin Kohli, and Richmond Y. Wong. 2019. This Thing Called Fairness: Disciplinary Confusion Realizing a Value in Technology. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 119 (nov 2019), 36 pages. https://doi.org/10.1145/3359221

[48] Salvatore Nunnari and Massimiliano Pozzi. 2022. Meta-analysis of inequality aversion estimates. CESifo Working Paper.

[49] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems* 35 (2022), 27730–27744.

[50] David C Parkes. 2005. Auction design with costly preference elicitation. *Annals of Mathematics and Artificial Intelligence* 44 (2005), 269–302.

[51] Ariel D Procaccia and Jeffrey S Rosenschein. 2006. The distortion of cardinal preferences in voting. Cooperative Information Agents X: 10th International Workshop, CIA 2006 Edinburgh, UK, September 11-13, 2006 Proceedings 10. , 317–331 pages.

[52] Ariel D Procaccia and Moshe Tennenholtz. 2013. Approximate mechanism design without money. *ACM Transactions on Economics and Computation (TEAC)* 1, 4 (2013), 1–26.

[53] Ariel D Procaccia and Junxing Wang. 2017. A lower bound for equitable cake cutting. Proceedings of the 2017 ACM Conference on Economics and Computation.

[54] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems* 36 (2024).

[55] Jack Robertson and William Webb. 1998. *Cake-cutting algorithms: Be fair if you can.* AK Peters/CRC Press.

[56] Samantha Robertson, Tonya Nguyen, Cathy Hu, Catherine Albiston, Afshin Nikzad, and Niloufar Salehi. 2023. Expressiveness, Cost, and Collectivism: How the Design of Preference Languages Shapes Participation in Algorithmic Decision-Making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing*

Systems (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 600, 16 pages. https://doi.org/10.1145/3544548.3580996

[57] Samantha Robertson and Niloufar Salehi. 2020. What If I Don't Like Any Of The Choices? The Limits of Preference Elicitation for Participatory Algorithm Design. Workshop on Participatory Approaches to Machine Learning at ICML 2020.

[58] Martin Eiliv Sandbu. 2008. Axiomatic foundations for fairness-motivated preferences. *Social Choice and Welfare* 31, 4 (2008), 589–619.

[59] Leonard J Savage. 1971. Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* 66, 336 (1971), 783–801.

[60] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2022. Participation is not a design fix for machine learning. Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization. , 6 pages.

[61] Rodrigo A Velez. 2016. Fairness and externalities. *Theoretical Economics* 11, 1 (2016), 381–410.

[62] Yang Yang, Sander Onderstal, and Arthur Schram. 2016. Inequity aversion revisited. *Journal of Economic Psychology* 54 (2016), 1–16.

[63] Sixie Yu, David Kempe, and Yevgeniy Vorobeychik. 2021. Altruism design in networked public goods games. IJCAI.

[64] J Zhang, S Branzei, and A Procaccia. 2013. Externalities in cake cutting. International Joint Conferences on Artificial Intelligence.

A BROADER APPLICATIONS OF MULTI-OBJECTIVE FRAMEWORK

In Section 2, we motivate and discuss a general framework of bottom-up resource allocation with multi-objective agents, then proceed to use inequality-averse agents as one case study of this model. We now demonstrate some ways in which this framework holds with greater generality by mapping some existing works on resource allocation where agents' utilities cannot solely be captured by classical utility measures.

While this paper examines inequality aversion, this is just one of many well-studied formulations of social preferences [32] . Other formulations of social preferences, such as altruism [45], can be neatly modeled here. Chen and Kempe [20] model altruistic agents (in a traffic routing setting) as having cost for their latency plus θ times the latency they create. In a similar vein, Flanigan et al. [34] model altruistic voters as having utility for their preferences plus θ times utility for the "public preference" represented by the population mean, with valuation $v_i(x) = u_i(x) + \frac{\theta}{n} \sum_{i'} u_{i'}(x)$.

There's also a body of literature that looks at price of fairness, including price of envy-freeness [14, 16, 23], in which fairness is imposed top-down by the mechanism designer. In reality, this fairness might be valued by multi-objective agents in fair division problems, and might be resolved bottom-up. Instead of treating envy-freeness as a constraint, we can model multi-objective agents as valuing envy-freeness as part of the objective. Here, multi-objective agents might have valuations $v_i(x) = u_i(x) - \theta \max_{i'} \langle a_i, \boldsymbol{x}_{i'} - \boldsymbol{x}_i \rangle$, with weight θ on the maximum envy they have for another agents' allocation. Moreover, this framework can also capture more recent variations of envy-freeness that incorporate, for example, positions on social networks [1] by modifying the terms in the summand above.

Finally, we note that our framework extends beyond just social preferences. For example, externalities are well-studied, whereby the utility individuals have for items they receive might be positively or negatively impacted by whether other agents have those items. One common example of externalities is the widespread use of cell phones: a user only has high utility for a cell phone if their friends and family also have one so they can communicate. In this case of externalities, we can model multi-objective agents $v_i(\boldsymbol{x}) = u_i(x) + \theta \sum_{i' \neq i} \mathbb{1}\{x_{i'} = 1\}$, where utility increases with the number of other people who have the item.

B RELATED WORK

Participation in Algorithm Design. While research on *participatory machine learning* has surged in recent years [43], there is little consensus about how to operationalize participation [9, 26]. In light of this increasing spotlight on participation, Sloane et al. [60] warn about the temptation to "participation wash," where algorithm designers boast ingenuine, deceptive, or nonconsensual participation. Similarly, Hitzig [38] observes the widening of a *normative gap* between the normative theory of the mechanism design literature and the normative goals of policymakers, which creates barriers to participation in mechanisms and algorithms in the way policymakers want.

Our work uses inequality-aversion as a case study, which can be contrasted with the fairness literature and work on equity in the resource allocation literature. In light of this, Mulligan et al. [47] discuss disciplinary-specific conceptualizations of fairness in our era, which can obscure discussions about values in technology. They call for interdisciplinary discussions and collaborations around the concept of fairness. Similarly, Finocchiaro et al. [33] emphasize that the approach to fairness has always been restricted to what can be reduced to the field's scope. In particular, machine learning traditionally approaches fairness by incorporating a pre-defined metric into optimization, treating people as data points with no agency. On the other hand, mechanism design, while considering potential strategic behavior, often tends to measure utility as a proxy for equality. Our framework can be seen as a step towards bridging these views; our proposal is to avoid a fixed objective function through participatory objective design that can incorporate diverse views on, for example, fairness or inequality aversion. We also deviate from the conventional notion of selfish agents and utilities, allowing for richer preferences over overall outcomes.

Other-Regarding, or Social, Preferences. The economics literature has studied models of agents who have other-regarding preferences [10, 25, 27, 31], but most often in the context of analyzing equilibrium strategies in various games, rather than evaluating the allocation produced by a fixed mechanism or game. These social preferences have been empirically validated, whether because of social norms [30], a desire for reciprocity [5], or a desire to be fair [29], and have been observed in contexts ranging from tax compliance [3] and fair wages [2] to general games [19]. Fehr and Fischbacher [28] argue that it is impossible to fully understand the effects of market outcomes and competition without considering social preferences, and Fehr and Fischbacher [29] show that even a few altruists or egoists can drastically affect market outcomes. Recent work has additionally shown that social preferences can curb distortion in voting [34] and participatory budgeting [4].

Fair Division as a Case Study. As a case study of our framework, we study a classical resource allocation problem of allocating m divisible goods among n agents, often studied by the fair division literature, which most often studies a top-down approach with different conceptualizations of fairness. Implicitly much of this literature studies mechanisms that do not directly optimize utilitarian social welfare (because of its implicit unfairness), but benchmark proposed mechanisms against this metric. For example, the proportionally fair allocation mechanism of Kelly [42] (equivalent to the Nash Bargaining Solution and CEEI) has become the one of the most widely-used mechanisms for allocating bandwidth rates on networks, as it maximizes Nash social welfare (products of utilities) for all of the agents, which Caragiannis et al. [17] observes is envy-free and efficient. Other notions of fairness yield slightly different fair division mechanisms, such as those of Ghodsi et al. [36], Robertson and Webb [55]. Moreover, Cole and Tao [24] leverages randomization to guarantee ex-ante envy-freeness of efficient resource allocation mechanisms. Our work diverges from these since much of the fair division literature proposes mechanisms satisfying some fairness constraints, and possibly benchmarks quality against social welfare. In Section 2.2, we introduce inequality-averse as our case study for the rest of the paper. The notion of inequality-aversion we adopt is in line with *equitable* cake-cutting [12, 13]. A division of cake is deemed *equitable* if every agent receiving cake has the same utility for their allocation. Equitable cake cuts cannot be exactly computed, and even approximations are expensive [53].

An additional line of work on fairness when reporting agents have externalities affecting their utilities has emerged [61, 64], which is one possible application of our general framework (see Appendix A for more discussion). However, Velez [61] uses money to address externalities in the allocation of indivisible goods. In contrast, Zhang et al. [64] study cake-cutting ($m = 1$) with externalities and proves the existence of (generalized) envy-free and proportional allocations when agents have utilities for each others' allocations. However, these preferences are not about community-level outcomes; rather, they are about others' individual outcomes, which aligns more with the altruism literature [20, 45, 63].

Preference Elicitation. Our work also intersects with the literature on *preference elicitation.* In the canonical principal-agent problem, informational asymmetries often require a principal to elicit some information from agents — often about their preferences [11, 18, 21, 41] or predictions about future events [35, 44, 59]. The principal then uses the elicited information to make decisions, such as allocating resources [15, 46, 50] or making decisions about public goods (e.g., determining the winner of an election, facility placement) [51, 52]. We defer discussing strategyproofness to § C, and primarily focus more on the question of what bits of information are being collected from agents, rather than the question of whether or not agents are being truthful. However, Bierbrauer and Netzer [7], Bierbrauer et al. [8] discuss mechanism design that is (non-)robust to social preferences, albeit in different settings, and only through analyzing the existence of a dominant strategy equilibrium. Recently, Joren et al. [40] also study participation in machine learning by presenting a system in which decision subjects can choose what data they provide to a model or ensemble of models in a way that improves expected performance of the *applied* model.

In the context of student assignment to schools, our framework aligns with the suggestion of Robertson and Salehi [57] that individual preferences can serve as valuable signals for promoting justice if they are expanded and made more expressive. This includes offering more avenues for expressing preferences regarding desirable social outcomes. Moreover, Robertson et al. [56] found that a more expressive preference language can encourage greater participation by students. These findings reinforce our proposal that preference elicitation across a broad spectrum of objectives, incorporating fairness considerations, can effectively advance social and distributive justice with meaningful participation from stakeholders.

Alignment. Our work relates to aligning reinforcement learning agents or large language models using human preferences over sampled outputs [22, 39, 49]. When directly specifying objectives is difficult for the designer or the human objective is hard to formulate, one approach is to collect human preferences over a set of model outcomes and learn a reward function to maximize. Recent works have also shown that alignment of large language models can be achieved indirectly by fine-tuning on collected preferences without explicitly modeling a human reward function [54]. Our framework follows a similar procedure, allowing agents to specify their objectives as a function of possible objectives. While alignment discussions mainly focus on single reinforcement learning agents, our work addresses allocating goods to multiple agents, where a central planner makes final decisions. Unlike common alignment approaches in machine learning that use arbitrary reward functions, our approach limits options to, for example, those supported by behavioral economics. This restriction enables tractable optimal allocation and provides theoretical bounds on potential improvement or loss due to preference elicitation. It also limits what aspect of the problem agents can change, for instance, they can be involved in determining the inequality-efficiency tradeoff.

C ELICITING PREFERENCES OF STRATEGIC AGENTS

Throughout our study, we assumed the planner has access to the agents' true preferences a, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$. Next, we briefly discuss possible strategic manipulations and mechanism design in the presence of inequality-averse agents. For demonstration, we assume $\boldsymbol{\beta} = \boldsymbol{\alpha}$ so there is only a single deviation from the standard setting of mechanism design.

First of all, as long as payments are permitted and agents are quasi-linear, a well-known result is that externality pricing is dominant-strategy incentive compatible and maximizes social welfare [37, Chapter 8]. A quasi-linear agent i cares for $v_i(x) - p_i$ where p_i is the payment made by i. Note that v_i can be a non-linear multi-dimensional function here. For reported utility coefficients a' and inequality-aversion levels $\boldsymbol{\alpha'}$, let $v_i(x; a', \boldsymbol{\alpha'})$ represent agent i's valuation of allocation x under the reported preferences. The inequality-averse allocation under the reported preferences is $x(a', \boldsymbol{\alpha'}) = \arg\max_x f(v(x; a', \boldsymbol{\alpha'}))$. Externality pricing defines the payment rule

$$p_i = v_{-i}(x(a'_{-i}, \boldsymbol{\alpha'}_{-i}); a'_{-i}, \boldsymbol{\alpha'}_{-i}) - v_{-i}(x(a', \boldsymbol{\alpha'}); a', \boldsymbol{\alpha'}), \tag{33}$$

where v_{-i} is the sum of valuations for every agent other than i. Let us examine this payment rule in the two-agent setting. In the absence of agent 1, $x_{2j}(a'_2, \alpha'_2) = 1$ and $v_2(1; a'_2, \alpha'_2) = \sum_j a'_{2j}$. Thus,

$$\begin{aligned} p_1 = &\sum_j a'_{2j}[1 - x_{2j}(a', \boldsymbol{\alpha'})] \\ &+ \alpha_2 \Big| \sum_j a'_{2j} x_{2j}(a', \boldsymbol{\alpha'}) - a'_{1j} x_{1j}(a', \boldsymbol{\alpha'}) \Big| . \end{aligned} \tag{34}$$

For inequality-agnostic agents (i.e., $\boldsymbol{\alpha} = \mathbf{0}$), the above payment rule is equivalent to a second-price auction per item. But in general, allocation and payments depend on all elements of a and $\boldsymbol{\alpha}$. Going beyond truthful mechanisms, it turns out there exists a simple modification of the standard second-price auction such that significant deviation from truthful reporting cannot be justified.

Proposition C.1. *Assume there is a cap of α_m on the maximum level of inequality aversion one is allowed to report. Given a' and $\boldsymbol{\alpha}'$ are reported by the two agents, define payment rule $p_{ij} = \mathbb{1}\{a'_{ij} \geq r(\boldsymbol{\alpha}')\, a'_{-ij}\}$ for item j, where $r(\boldsymbol{\alpha}') := (1 - \alpha'_1 - \alpha'_2)/(1 + \alpha'_1 + \alpha'_2)$. Then no agent i has any incentive to report a'_{ij} outside of $[r^2(\alpha_i, \alpha_m)\, a_{ij}, r^{-2}(\alpha_i, \alpha_m)\, a_{ij}]$.*

Proof. Let a'_{2j} and α'_2 be the agent 1's belief about agent 2's report. Also, let a_{1j} and α_1 be the true preference of agent 1. There are three possibilities:

(1) If $a_{1j} < r(\alpha_1, \alpha'_2)\, a'_{2j}$, the allocation under truthful report will be $x_{1j} = 0$. Any deviation from truthful reporting that results in $x_{1j} > 0$ requires $a'_{1j} \geq r(\alpha'_1, \alpha'_2)\, a'_{2j}$. So, agent 1 needs to pay an extra a'_{2j} and optimistically her overall change of utility will be

$$\begin{aligned} -a'_{2j} + a_{1j} + \alpha_1(a_{1j} + a'_{2j}) &= a_{1j}(1 + \alpha_1) - a'_{2j}(1 - \alpha_1) \\ &< a'_{2j}(1 + \alpha_1)[-r(\alpha_1, 0) + r(\alpha_1, \alpha_2)] \leq 0\,. \end{aligned} \tag{35}$$

Therefore, agent 1 has no beneficial deviation in this case.

(2) If $a_{1j} > a'_{2j}/r(\alpha_1, \alpha'_2)$, the allocation under truthful report is $x_{1j} = 1$ and agent 1 has to pay a'_{2j}. To avoid this payment, agent 1 should report $a'_{1j} < r(\alpha'_1, \alpha'_2)$. In this case, agent 1 will lose all of item j and optimistically her overall change of utility will be

$$\begin{aligned} -a_{1j} + a'_{2j} + \alpha_1(a_{1j} + a'_{2j}) &= -a_{1j}(1 - \alpha_1) + a'_{2j}(1 + \alpha_1) \\ &< a_{1j}(1 + \alpha_1)[-r(\alpha_1, 0) + r(\alpha_1, \alpha_2)] \leq 0\,. \end{aligned} \tag{36}$$

Again, agent 1 has no incentive to deviate.

(3) If $a_{1j} \in [r(\alpha_1, \alpha'_2)\, a'_{2j}, a'_{2j}/r(\alpha_1, \alpha'_2)]$, for $\alpha'_1 = \alpha_1$, deviation of a'_{1j} from a_{1j} might be justified up to a scale of $r^{\pm 2}(\alpha_1, \alpha'_2)$.

Fig. 3 shows the summary of possibilities. Regardless of agent 1's belief about agent 2, she has no incentive to report $a'_{1j} \notin [r(\alpha_1, \alpha_m)\, a_{1j}, a_{1j}/r(\alpha_1, \alpha_m)]$.

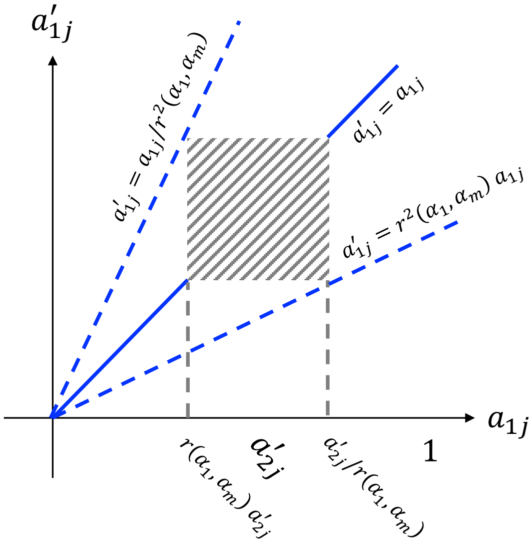
a'_{1j}
$a'_{1j} = a_{1j}/r^2(\alpha_1, \alpha_m)$
$a'_{1j} = a_{1j}$
$a'_{1j} = r^2(\alpha_1, \alpha_m)\, a_{1j}$
a_{1j}
$r(\alpha_1, \alpha_m)\, a'_{2j}$
a'_{2j}
$a'_{2j}/r(\alpha_1, \alpha_m)$
1

Figure 3: Bounding deviation from truthfulness.

□

Define $c(\boldsymbol{\alpha}) := 1 - r(\boldsymbol{\alpha})$. Since $r^2(\boldsymbol{\alpha}) = 1 - O(c(\boldsymbol{\alpha}))$, we don't expect any bound estimated based on strategically manipulated data to be *relatively* wrong by more than $O(\max_i c(\alpha_i, \alpha_m))$. Further, this proposition immediately implies there exists a Bayes Nash equilibrium with such a property.

D THE LOSS OF INCORPORATING INEQUALITY AVERSION: TWO γ-DISSIMILAR AGENTS

The next theorem shows how the knowledge of dissimilarity can be helpful in bounding the loss.

THEOREM D.1 (UPPERBOUNDED LOSS FOR TWO DISSIMILAR AGENTS). *For two γ-dissimilar agents and any $\tau \in [0, 1]$, the loss can be bounded by $loss \le \sum_j loss_j$, where*

(1) If $\gamma\|\boldsymbol{a}_1\|_1 > 1$:

$$loss_j \le c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}\,. \tag{37}$$

(2) If $\gamma\|\boldsymbol{a}_1\|_1 \le 1$:

$$loss_j \le \begin{cases} c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\big(1 + \frac{2}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})}\big)\gamma^{1-\tau}\|\boldsymbol{a}_1\|_1^{1-\tau}\,, & \gamma\|\boldsymbol{a}_1\|_1 > a_{1j}\,,\\ 2c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\gamma^{\tau}\|\boldsymbol{a}_1\|_1^{\tau}\,, & \gamma\|\boldsymbol{a}_1\|_1 \le a_{1j} < \gamma\|\boldsymbol{a}_1\|_1 + \gamma^{\tau}\|\boldsymbol{a}_1\|_1^{\tau},\\ 0\,, & \gamma\|\boldsymbol{a}_1\|_1 + \gamma^{\tau}\|\boldsymbol{a}_1\|_1^{\tau} \le a_{1j}\,. \end{cases} \tag{38}$$

PROOF. For notational convenience, we denote a_{2j} by x_j in this proof. For fixed $\boldsymbol{a}_1$, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$, we can rewrite $loss_j$ in Eq. (14) as

$$loss_j(x_j) = \mathbb{1}\{r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j} \le x_j \le a_{1j}\}\,(a_{1j} - x_j)\,. \tag{39}$$

We will explicitly show the dependence on x throughout the proof. Since the $loss_j$ is discontinuous in x_j and hard to analyze, we first upperbound it with two quadratic terms (Fig. 4). Specifically, we require the upperbound to be zero at $x_j = 0$ and $x_j = 1$, and have a zero derivative at $x_j = r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}$. These constraints uniquely determine the upper bound:

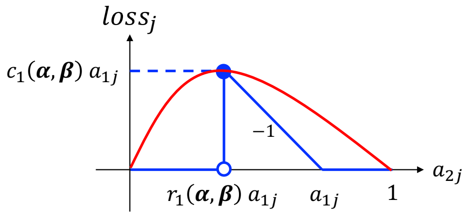

Figure 4: $loss_j$ with its quadratic upperbound (Eq. (40)) in red.

$$loss_j(x_j) \le \overline{loss}_j(x_j) := \begin{cases} -x_j \frac{c_1(\boldsymbol{\alpha},\boldsymbol{\beta})}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})}\big(\frac{x_j}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})\, a_{1j}} - 2\big)\,, & x_j < r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}\,,\\ \frac{c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\, a_{1j}}{(1 - r_1(\boldsymbol{\alpha},\boldsymbol{\beta})\, a_{1j})^2}(1 - x_j)(x_j + 1 - 2r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j})\,, & x_j \ge r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}\,. \end{cases} \tag{40}$$

To find the worst-case upperbounded loss, we need to solve:

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & \overline{loss}(\boldsymbol{x}) := \sum_j \overline{loss}_j(x_j) \\ \text{s.t.} \quad & c(\boldsymbol{x}) = \gamma\|\boldsymbol{a}_1\|_1 \sum_j x_j - \sum_j a_{1j}x_j \ge 0\,, \\ & \underline{c}_j(\boldsymbol{x}) = x_j \ge 0\,,\ \forall j \\ & \overline{c}_j(\boldsymbol{x}) = 1 - x_j \ge 0\,,\ \forall j\,. \end{aligned} \tag{41}$$

The Lagrangian function of the above optimization problem is

$$\mathcal{L}(\boldsymbol{x}; \lambda, \overline{\boldsymbol{\lambda}}, \underline{\boldsymbol{\lambda}}) = \overline{loss}(\boldsymbol{x}) + \lambda c(\boldsymbol{x}) + \sum_j \underline{\lambda}_j \underline{c}_j(\boldsymbol{x}) + \sum_j \overline{\lambda}_j \overline{c}_j(\boldsymbol{x}) \tag{42}$$

$$= \sum_j \overline{loss}_j(x_j) + \lambda(\gamma\|\boldsymbol{a}_1\|_1 - a_{1j})x_j + \underline{\lambda}_j x_j + \overline{\lambda}_j(1 - x_j)\,, \tag{43}$$

where λ, $\overline{\boldsymbol{\lambda}}$, and $\underline{\boldsymbol{\lambda}}$ are non-negative Lagrange multipliers. We know for every valid value of the multipliers

$$\max_{\boldsymbol{x}} \mathcal{L} \ge \max_{\boldsymbol{x}} \min_{\lambda, \overline{\boldsymbol{\lambda}}, \underline{\boldsymbol{\lambda}} \ge 0} \mathcal{L} = \max_{\substack{\boldsymbol{x} \\ \text{s.t. constraints}}} \overline{loss}\,. \tag{44}$$

So, in the following, we first solve $\max_{\boldsymbol{x}} \mathcal{L}$ and then choose appropriate values for multipliers to obtain a good bound.

As the definition of $\mathcal{L}$ suggests, $\mathcal{L}$ is separable over items: $\mathcal{L} = \sum_j \mathcal{L}_j$. Defining

$$\delta_j := \underline{\lambda}_j - \overline{\lambda}_j + \lambda(\gamma\|\boldsymbol{a}_1\|_1 - a_{1j})\,, \tag{45}$$

we can write $\mathcal{L}_j(x_j; \lambda, \overline{\lambda}_j, \underline{\lambda}_j) \coloneqq \overline{loss}_j(x_j) + \delta_j x_j + \overline{\lambda}_j$. The separability over items allows us to write $\max_{\boldsymbol{x}} \mathcal{L} = \sum_j \max_{x_j} \mathcal{L}_j$. As $\mathcal{L}_j$ is a concave quadratic function of x_j, it is necessary and sufficient for the maximizer to satisfy the first-order condition

$$\frac{d\mathcal{L}_j}{dx_j} = \frac{d}{dx_j}\overline{loss}_j + \delta_j = 0\,. \tag{46}$$

The derivative of $\overline{loss}_j$ is positive if and only if $x_j < r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}$, so, there are two possibilities based on the sign of δ_j:

(1) $\delta_j < 0$: The first-order condition requires $x_j^* = \arg\max_{x_j} \mathcal{L}_j < r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}$. Plugging derivative of $\overline{loss}_j$ into Eq. (46) and solving for x_j gives

$$x_j^* = r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j} + \delta_j \frac{r_1^2(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}}{2c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})}\,. \tag{47}$$

By evaluating $\mathcal{L}_j$ at x_j^* and simplifying equations, we obtain

$$\mathcal{L}_j^*(\lambda, \overline{\lambda}_j, \underline{\lambda}_j) \coloneqq \max_{x_j} \mathcal{L}_j = \mathcal{L}_j(x_j^*; \lambda, \overline{\lambda}_j, \underline{\lambda}_j) = c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j} \big(1 + \frac{r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})}{2c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})}\delta_j\big)^2 + \overline{\lambda}_j\,. \tag{48}$$

(2) $\delta_j \geq 0$: The first-order condition in Eq. (46) requires $x_j^* = \arg\max_{x_j} \mathcal{L}_j \geq r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}$. Plugging derivative of $\overline{loss}_j$ and solving for x_j gives

$$x_j^* = r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j} + \delta_j \frac{(1 - r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j})^2}{2c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}}\,. \tag{49}$$

By evaluating $\mathcal{L}_j$ at x_j^* and simplifying equations, we obtain

$$\mathcal{L}_j^*(\lambda, \overline{\lambda}_j, \underline{\lambda}_j) = \mathcal{L}_j(x_j^*; \lambda, \overline{\lambda}_j, \underline{\lambda}_j) = c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j} \Big(1 + \frac{r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})}{c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})}\delta_j + \frac{(1 - r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j})^2}{4c_1^2(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}^2}\delta_j^2\Big) + \overline{\lambda}_j\,. \tag{50}$$

Now that we have found $\mathcal{L}^*(\lambda, \overline{\boldsymbol{\lambda}}, \underline{\boldsymbol{\lambda}}) \coloneqq \max_{\boldsymbol{x}} \mathcal{L} = \sum_j \mathcal{L}_j^*(\lambda, \overline{\lambda}_j, \underline{\lambda}_j)$, the next step is to find appropriate λ, $\overline{\boldsymbol{\lambda}}$, and $\underline{\boldsymbol{\lambda}}$ that minimize $\mathcal{L}^*$ or make it sufficiently small. Again we consider two cases:

(1) $\delta_j < 0$: In this case,

$$\frac{\partial \mathcal{L}_j^*}{\partial \overline{\lambda}_j} = -r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j} \big(1 + \frac{r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})}{2c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})}\delta_j\big) + 1 > 1 - r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j} \geq 0\,. \tag{51}$$

So, $\mathcal{L}_j^*$ is monotone increasing in $\overline{\lambda}_j$. The optimal values of $\overline{\lambda}_j$ and $\underline{\lambda}_j$ depend on the sign of $\gamma\|\boldsymbol{a}_1\|_1 - a_{1j}$:

(a) $\gamma\|\boldsymbol{a}_1\|_1 > a_{1j}$: We set $\underline{\lambda}_j = 0$ and $\overline{\lambda}_j \to \lambda(\gamma\|\boldsymbol{a}_1\|_1 - a_{1j})^+$. For these values, we have $\delta_j \to 0^-$ and

$$\mathcal{L}_j^*(\lambda, \overline{\lambda}_j \to \lambda(\gamma\|\boldsymbol{a}_1\|_1 - a_{1j})^+, \underline{\lambda}_j = 0) \to \big[c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j} + \lambda(\gamma\|\boldsymbol{a}_1\|_1 - a_{1j})\big]^+\,. \tag{52}$$

(b) $\gamma\|\boldsymbol{a}_1\|_1 \leq a_{1j}$: We set $\overline{\lambda}_j = 0$ and

$$\underline{\lambda}_j = \begin{cases} 0\,, & \lambda(a_{1j} - \gamma\|\boldsymbol{a}_1\|_1) < \frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})}\,, \\ \lambda(a_{1j} - \gamma\|\boldsymbol{a}_1\|_1) - \frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})} & \text{o.w.} \end{cases} \tag{53}$$

For such a choice,

$$\mathcal{L}_j^*(\lambda, \overline{\lambda}_j = 0, \underline{\lambda}_j) = \begin{cases} c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j} \big(1 - \frac{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})}{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})}\lambda(a_{1j} - \gamma\|\boldsymbol{a}_1\|_1)\big)^2\,, & \lambda(a_{1j} - \gamma\|\boldsymbol{a}_1\|_1) < \frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})}\,, \\ 0 & \text{o.w.} \end{cases} \tag{54}$$

(2) $\delta_j \geq 0$: In this case,

$$\frac{\partial \mathcal{L}_j^*}{\partial \underline{\lambda}_j} = c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j} \Big(\frac{r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})}{c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})} + 2\frac{(1 - r_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j})^2}{4c_1^2(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}^2}\delta_j\Big) \geq 0\,. \tag{55}$$

So, $\mathcal{L}_j^*$ is monotone increasing in $\underline{\lambda}_j$. Again, the optimal values of $\overline{\lambda}_j$ and $\underline{\lambda}_j$ depend on the sign of $\gamma\|\boldsymbol{a}_1\|_1 - a_{1j}$:

(a) $\gamma\|\boldsymbol{a}_1\|_1 \leq a_{1j}$: We set $\overline{\lambda}_j = 0$ and $\underline{\lambda}_j = \lambda(a_{1j} - \gamma\|\boldsymbol{a}_1\|_1)$. For these values,

$$\mathcal{L}(\lambda, \overline{\lambda}_j = 0, \underline{\lambda}_j = \lambda(a_{1j} - \gamma\|\boldsymbol{a}_1\|_1)) = c_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}\,. \tag{56}$$

(b) $\gamma\|\boldsymbol{a}_1\|_1 > a_{1j}$: We set $\underline{\lambda}_j = 0$ and

$$\overline{\lambda}_j = \begin{cases} 0\,, & \lambda(\gamma\|\boldsymbol{a}_1\|_1 - a_{1j}) < \frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}}{1-r_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}}\,, \\ \lambda(\gamma\|\boldsymbol{a}_1\|_1 - a_{1j}) - \frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}}{1-r_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}} & \text{o.w.} \end{cases} \tag{57}$$

Evaluating $\mathcal{L}_j^*$ at these values and simplifying equations,

$$\mathcal{L}_j^*(\lambda, \overline{\lambda}_j, \underline{\lambda}_j = 0) \leq \begin{cases} \frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}}{1-r_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}}\,, & \lambda(\gamma\|\boldsymbol{a}_1\|_1 - a_{1j}) < \frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}}{1-r_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}}\,, \\ \lambda(\gamma\|\boldsymbol{a}_1\|_1 - a_{1j}) & \text{o.w.} \end{cases} \tag{58}$$

$$= \max\left\{\lambda(\gamma\|\boldsymbol{a}_1\|_1 - a_{1j}), \frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}}{1-r_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}}\right\} \tag{59}$$

Now for every a_{1j} we can choose between $\delta_j < 0$ and $\delta_j \geq 0$ cases and choose $\overline{\lambda}_j$ and $\underline{\lambda}_j$ such that $\mathcal{L}_j^*$ is minimized:

(1) $\gamma\|\boldsymbol{a}_1\|_1 > a_{1j}$: We choose $\underline{\lambda}_j = 0$ and $\overline{\lambda}_j \to \lambda(\gamma\|\boldsymbol{a}_1\|_1 - a_{1j})^+$.

(2) $\gamma\|\boldsymbol{a}_1\|_1 \leq a_{1j}$: We set $\overline{\lambda}_j = 0$ and $\underline{\lambda}_j$ as Eq. (53).

For such a choice:

$$\mathcal{L}_j^* \to \begin{cases} c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j} + \lambda(\gamma\|\boldsymbol{a}_1\|_1 - a_{1j})\,, & \gamma\|\boldsymbol{a}_1\|_1 > a_{1j}\,, \\ c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}(1 - \frac{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})}{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})}\lambda(a_{1j} - \gamma\|\boldsymbol{a}_1\|_1))^2\,, & \gamma\|\boldsymbol{a}_1\|_1 \leq a_{1j} < \gamma\|\boldsymbol{a}_1\|_1 + \frac{1}{\lambda}\frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})}\,, \\ 0\,, & \gamma\|\boldsymbol{a}_1\|_1 + \frac{1}{\lambda}\frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})} \leq a_{1j}\,. \end{cases} \tag{60}$$

Let

$$\lambda = \begin{cases} 0\,, & \gamma\|\boldsymbol{a}_1\|_1 \geq 1\,, \\ \frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})}\frac{1}{\gamma^\tau\|\boldsymbol{a}_1\|_1^\tau} & \text{o.w.}\,, \end{cases} \tag{61}$$

where $1 \geq \tau \geq 0$. Then

(1) $\gamma\|\boldsymbol{a}_1\|_1 \geq 1$:

$$\mathcal{L}_j^* \to c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,a_{1j}\,. \tag{62}$$

(2) $\gamma\|\boldsymbol{a}_1\|_1 < 1$:

$$\mathcal{L}_j^* \leq \begin{cases} \gamma\|\boldsymbol{a}_1\|_1(c_1(\boldsymbol{\alpha},\boldsymbol{\beta}) + \lambda)\,, & \gamma\|\boldsymbol{a}_1\|_1 > a_{1j}\,, \\ c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\big(\gamma\|\boldsymbol{a}_1\|_1 + \frac{1}{\lambda}\frac{2c(\boldsymbol{\alpha},\boldsymbol{\beta})}{r(\boldsymbol{\alpha},\boldsymbol{\beta})}\big)\,, & \gamma\|\boldsymbol{a}_1\|_1 \leq a_{1j} < \gamma\|\boldsymbol{a}_1\|_1 + \frac{1}{\lambda}\frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})}\,, \\ 0\,, & \gamma\|\boldsymbol{a}_1\|_1 + \frac{1}{\lambda}\frac{2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})} \leq a_{1j} \end{cases} \tag{63}$$

$$\leq \begin{cases} c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\big[\gamma\|\boldsymbol{a}_1\|_1 + \frac{2}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})}\gamma^{1-\tau}\|\boldsymbol{a}_1\|_1^{1-\tau}\big]\,, & \gamma\|\boldsymbol{a}_1\|_1 > a_{1j}\,, \\ c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\big[\gamma\|\boldsymbol{a}_1\|_1 + \gamma^\tau\|\boldsymbol{a}_1\|_1^\tau\big]\,, & \gamma\|\boldsymbol{a}_1\|_1 \leq a_{1j} < \gamma\|\boldsymbol{a}_1\|_1 + \gamma^\tau\|\boldsymbol{a}_1\|_1^\tau\,, \\ 0\,, & \gamma\|\boldsymbol{a}_1\|_1 + \gamma^\tau\|\boldsymbol{a}_1\|_1^\tau \leq a_{1j} \end{cases} \tag{64}$$

$$\leq \begin{cases} c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\big(1 + \frac{2}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})}\big)\gamma^{1-\tau}\|\boldsymbol{a}_1\|_1^{1-\tau}\,, & \gamma\|\boldsymbol{a}_1\|_1 > a_{1j}\,, \\ 2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\gamma^\tau\|\boldsymbol{a}_1\|_1^\tau\,, & \gamma\|\boldsymbol{a}_1\|_1 \leq a_{1j} < \gamma\|\boldsymbol{a}_1\|_1 + \gamma^\tau\|\boldsymbol{a}_1\|_1^\tau\,, \\ 0\,, & \gamma\|\boldsymbol{a}_1\|_1 + \gamma^\tau\|\boldsymbol{a}_1\|_1^\tau \leq a_{1j}\,. \end{cases} \tag{65}$$

□

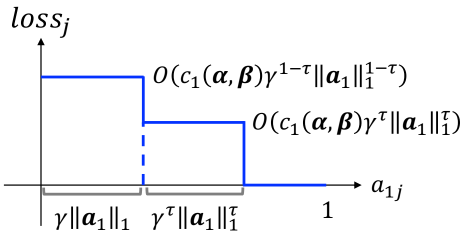

Figure 5: Upperbounded $loss_j$ for γ-dissimilar agents ($\gamma\|a_1\|_1 \leq 1$) based on Theorem D.1.

Fig. 5 shows the upperbounded $loss_j$ for sufficiently small γ. The parameter τ gives us the flexibility to penalize extreme values of a_{1j}. Without any further knowledge on a_{1j}, we cannot do better than $loss_j = O\big(c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\sqrt{\gamma\|\boldsymbol{a}_1\|_1}\big)$ which can be achieved by choosing $\tau = 0.5$. This bound is more informative than $loss \le c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\|\boldsymbol{a}_1\|_1$ only if $\gamma < \|\boldsymbol{a}_1\|_1/m^2$. Having a prior over a_{1j} can result in strictly better bounds. The next corollary gives an example of a_{1j} being drawn from a uniform distribution, even without any assumptions on the independence of items.

Corollary D.2. *Assume $\gamma \le 1/m$. For $a_{1j} \sim unif(0,1)$ for all $j \in [m]$, we have*

$$\mathbb{E}[loss] = O\big(m\,c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,(\gamma m)^{\frac{4}{3}}\big)\,. \tag{66}$$

Proof. Using Theorem D.1, for $\tau \ge 0.5$ we have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{a}_1}[loss_j] &\le \mathbb{E}_{\boldsymbol{a}_1}\Bigg[\begin{cases} c_1(\boldsymbol{\alpha},\boldsymbol{\beta})(1+\frac{2}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})})(\gamma m)^{1-\tau}\,, & a_{1j} < \gamma m\,, \\ 2c_1(\boldsymbol{\alpha},\boldsymbol{\beta})(\gamma m)^{\tau}\,, & \gamma m \le a_{1j} < \gamma m + (\gamma m)^{\tau}\,, \\ 0\,, & \gamma m + (\gamma m)^{\tau} \le a_{1j} \end{cases} \Big|\; a_{1j}\Bigg] \\ &= c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\Big[(1+\frac{2}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})})(\gamma m)^{2-\tau} + 2(\gamma m)^{2\tau}\Big]\,. \end{aligned} \tag{67}$$

A good choice of τ would be a value such that all the terms have the same exponent: $2-\tau = 2\tau \Rightarrow \tau = 2/3$. For $\tau = 2/3$,

$$\mathbb{E}[loss] \le \sum_{j=1}^{m}\mathbb{E}[loss_j] \le 2m\,c_1(\boldsymbol{\alpha},\boldsymbol{\beta})\,(1+\frac{1}{r_1(\boldsymbol{\alpha},\boldsymbol{\beta})})\,(\gamma\,m)^{\frac{4}{3}}\,. \tag{68}$$

□

Since $\mathbb{E}[\|\boldsymbol{a}_1\|_1] = \Theta(m)$, this is a tighter bound than the distribution-free bound.

E THE LOSS OF INCORPORATING INEQUALITY AVERSION: TWO INDEPENDENT AGENTS

The following Lemma bounds $loss_j$ for a general two independent agents setting.

Lemma E.1. *For $\bar{a} = \max\{a_{1j}, a_{2j}\}$ with cumulative distribution $G_j^2(\cdot)$, if G_j is continuous,*

$$\mathbb{E}[loss_j] = \mathbb{E}_{\bar{a}\sim G_j^2}\Big[\bar{a}\int_0^{c_m(\boldsymbol{\alpha},\boldsymbol{\beta})}\big[\hat{G}_j(1-\hat{l}|\bar{a}) - \hat{G}_j(1-c_m(\boldsymbol{\alpha},\boldsymbol{\beta})|\bar{a})\big]d\hat{l}\Big]\,, \tag{69}$$

where $\hat{G}_j(\cdot|\bar{a})$ is defined as $\hat{G}_j(\hat{a}|\bar{a}) \coloneqq G_j(\hat{a}\,\bar{a})/G_j(\bar{a})$, and $c_m(\boldsymbol{\alpha},\boldsymbol{\beta}) = \max_i c_i(\boldsymbol{\alpha},\boldsymbol{\beta})$.

Proof. Let $\bar{a} = \max\{a_{1j}, a_{2j}\}$, $\underline{a} = \min\{a_{1j}, a_{2j}\}$, $r_m(\boldsymbol{\alpha},\boldsymbol{\beta}) = \min_i r_i(\boldsymbol{\alpha},\boldsymbol{\beta})$, and $c_m(\boldsymbol{\alpha},\boldsymbol{\beta}) = \max_i c_i(\boldsymbol{\alpha},\boldsymbol{\beta})$. For notational convenience, we drop the dependence on j from g_j and G_j as it is clear from the context. Given $\bar{a}$, the distribution of $\underline{a}$ follows $g(\underline{a}|\bar{a}) = \mathbb{1}\{\underline{a} \le \bar{a}\}\,g(\underline{a})/G(\bar{a})$. Define a new variable $\hat{a} \coloneqq \underline{a}/\bar{a}$. Given $\bar{a}$, the cumulative distribution of $\hat{a}$ is

$$\hat{G}(\hat{a}|\bar{a}) = G(\hat{a}\,\bar{a}|\bar{a}) = \frac{G(\hat{a}\,\bar{a})}{G(\bar{a})}\,. \tag{70}$$

Conditioned on $\bar{a}$, we have

$$loss_j|\bar{a} = \mathbb{1}\{\underline{a} \ge r_m(\boldsymbol{\alpha},\boldsymbol{\beta})\,\bar{a}\}(\bar{a}-\underline{a}) = \mathbb{1}\{\hat{a} \ge r_m(\boldsymbol{\alpha},\boldsymbol{\beta})\}(1-\hat{a})\bar{a}\,. \tag{71}$$

Let L be the cumulative distribution of $loss_j$. Define $\widehat{loss}_j|\bar{a} \coloneqq (loss_j/\bar{a})|\bar{a}$ with the cumulative distribution of $\hat{L}(\hat{l}|\bar{a}) = L(\hat{l}\,\bar{a}|\bar{a})$. For any $\hat{l} \in [0, c_m(\boldsymbol{\alpha},\boldsymbol{\beta})]$, we have $\widehat{loss}_j|\bar{a} \le \hat{l}$ if and only if $\hat{a} < r_m(\boldsymbol{\alpha},\boldsymbol{\beta})$ or $\hat{a} \ge 1-\hat{l}$. Therefore, for a continuous G, we have

$$\hat{L}(\hat{l}|\bar{a}) = \hat{G}(r_m(\boldsymbol{\alpha},\boldsymbol{\beta})|\bar{a}) + 1 - \hat{G}(1-\hat{l}|\bar{a})\,. \tag{72}$$

Now we can calculate the expected loss by

$$\mathbb{E}[loss_j] = \mathbb{E}_{\bar{a}\sim G_j^2}\big[\bar{a}\,\mathbb{E}_{\hat{L}}[\widehat{loss}_j|\bar{a}]\big] = \mathbb{E}_{\bar{a}\sim G_j^2}\Big[\bar{a}\int_0^{c_m(\boldsymbol{\alpha},\boldsymbol{\beta})}\big[\hat{G}(1-\hat{l}|\bar{a}) - \hat{G}(r_m(\boldsymbol{\alpha},\boldsymbol{\beta})|\bar{a})\big]d\hat{l}\Big]\,, \tag{73}$$

where we used tower property and integration by parts. □

For some classes of distributions, such as uniform and beta distributions, $\hat{G}_j(\cdot|\bar{a})$ in Lemma E.1 has no dependence on $\bar{a}$ and Eq. (69) can be significantly simplified. The next corollary shows this is generally true for $g_j = Beta(s_j, 1)$.

Corollary E.2. *For $g_j = Beta(s_j, 1)$ with $s_j \ge 1$, we have*

$$\mathbb{E}[loss] \le c_m^2(\boldsymbol{\alpha},\boldsymbol{\beta})\,\big(\sum_j s_j\big)\,. \tag{74}$$

Proof. Plugging $G_j(a) = a^{s_j}$ into $\hat{G}_j$ definition, we have $\hat{G}_j(\hat{a}) = \hat{a}^{s_j} = G_j(\hat{a})$. Also, note that G_j^2 will be the cumulative distribution of $Beta(2s_j, 1)$. So, we can simplify and upperbound $\mathbb{E}[loss_j]$ from Eq. (69):

$$\begin{aligned}
\mathbb{E}[loss_j] &= \mathbb{E}_{G_j^2}[\bar{a}] \int_0^{c_m(\boldsymbol{\alpha},\boldsymbol{\beta})} \big[(1-\hat{l})^{s_j} - (1-c_m(\boldsymbol{\alpha},\boldsymbol{\beta}))^{s_j}\big] d\hat{l} && (75)\\
&= \frac{2s_j}{2s_j+1}\Big(\frac{1}{s_j+1}[1-(1-c_m(\boldsymbol{\alpha},\boldsymbol{\beta}))^{s_j+1}] - c_m(\boldsymbol{\alpha},\boldsymbol{\beta})(1-c_m(\boldsymbol{\alpha},\boldsymbol{\beta}))^{s_j}\Big) && (76)\\
&= \frac{s_j}{(s_j+1/2)(s_j+1)}\Big(1-(1+c_m(\boldsymbol{\alpha},\boldsymbol{\beta})s_j)(1-c_m(\boldsymbol{\alpha},\boldsymbol{\beta}))^{s_j}\Big) && (77)\\
&\leq \frac{s_j^3}{(s_j+1/2)(s_j+1)} c_m^2(\boldsymbol{\alpha},\boldsymbol{\beta})\,. && (78)
\end{aligned}$$

We applied $(1-c)^s \geq 1 - s\,c$ for $c \leq 1$ and $s \geq 1$ to obtain the last inequality. This bound is tight for $s = 1$ (uniform distribution). The rest of the proof is straightforward. □

For general distributions, we can approximate Eq. (69) for small $c_m(\boldsymbol{\alpha},\boldsymbol{\beta})$ and find a sufficient condition to get a $O(c_m^2(\boldsymbol{\alpha},\boldsymbol{\beta}))$ bound on loss:

Corollary E.3. *For $G_j \in C^1$ define $\kappa_j := \sup_{\bar{a}>0} \bar{a}\, g_j(\bar{a})/G_j(\bar{a})$. If κ_j is finite for every j, then*

$$\mathbb{E}[loss] \leq c_m^2(\boldsymbol{\alpha},\boldsymbol{\beta})\,\Big(\sum_j \kappa_j/2\Big) + o(c_m^2(\boldsymbol{\alpha},\boldsymbol{\beta}))\,. \qquad (79)$$

Proof. For a small $c_m(\boldsymbol{\alpha},\boldsymbol{\beta})$, we can use the first-order Taylor expansion of $\hat{G}$ around 1 in calculating Eq. (69):

$$\begin{aligned}
\int_0^{c_m(\boldsymbol{\alpha},\boldsymbol{\beta})} \big[\hat{G}_j(1-\hat{l}|\bar{a}) - \hat{G}_j(1-c_m(\boldsymbol{\alpha},\boldsymbol{\beta})|\bar{a})\big] d\hat{l} &= \int_0^{c_m(\boldsymbol{\alpha},\boldsymbol{\beta})} \big[\hat{g}_j(1|\bar{a})\,(c_m(\boldsymbol{\alpha},\boldsymbol{\beta}) - \hat{l}) + o(c_m(\boldsymbol{\alpha},\boldsymbol{\beta}))\big] d\hat{l} && (80)\\
&= \frac{1}{2}\hat{g}_j(1|\bar{a})\, c_m^2(\boldsymbol{\alpha},\boldsymbol{\beta}) + o(c_m^2(\boldsymbol{\alpha},\boldsymbol{\beta}))\,. && (81)
\end{aligned}$$

Here $\hat{g}_j(\hat{a}|\bar{a}) = \hat{G}_j'(\hat{a}|\bar{a}) = \bar{a}\, g_j(\hat{a}\,\bar{a})/G_j(\bar{a})$. Define $\kappa_j := \max_{\bar{a}>0} \bar{a}\, g_j(\bar{a})/G_j(\bar{a})$. We have

$$\mathbb{E}[loss_j] \leq \mathbb{E}_{G_j^2}[\bar{a}]\,\Big[\frac{1}{2}\kappa_j\, c^2(\boldsymbol{\alpha}) + o(c_m^2(\boldsymbol{\alpha},\boldsymbol{\beta}))\Big] \leq \frac{1}{2}\kappa_j\, c_m^2(\boldsymbol{\alpha},\boldsymbol{\beta}) + o(c_m^2(\boldsymbol{\alpha},\boldsymbol{\beta}))\,. \qquad (82)$$

The rest of the proof is straightforward. □

Note that for $g_j = Beta(s_j, 1)$, we have $\kappa_j = s_j$ and for small $c_m(\boldsymbol{\alpha},\boldsymbol{\beta})$, Corollary E.3 gives a slightly tighter bound than Corollary E.2. It is also worth noting κ_j is not necessarily finite for all distributions even if $g_j \in C^\infty$. For example, for any $s_j' \in (0,1)$ and $g_j = Beta(s_j, s_j')$, we have $\bar{a}\, g_j(\bar{a})/G_j(\bar{a}) \longrightarrow \infty$ as $\bar{a} \longrightarrow 1^-$.

F ADDITIONAL STATEMENTS

Proposition F.1. *There exist two δ-similar agents for whom $loss \geq \delta/2$.*

Proof. The proof is constructive: For an even number of items m, consider $a_{1j} = 1$ and $a_{2j} = 1 - \delta/m + \epsilon\chi\{j \text{ is odd}\}$, where $\epsilon \longrightarrow 0^+$. Under x^*, all items go to agent 1. But for values of $\boldsymbol{\alpha}$ such that $|r_1(\boldsymbol{\alpha},\boldsymbol{\beta}) - (1-\delta/m)| < \epsilon$, one can verify $x_{1j}^{\alpha,\beta} = \mathbb{1}\{j \text{ is even}\}$. So, in this example, $\|\boldsymbol{a}_1 - \boldsymbol{a}_2\|_1 = \delta$ and $loss = \delta/2$. □

Proposition F.2. *There exist two independent agents for whom $\mathbb{E}[loss_j] = \Theta(c_1(\boldsymbol{\alpha},\boldsymbol{\beta}))$.*

Proof. The proof is constructive. Define g_j to be concentrated around two points: $g_j(a_{ij}) := 0.5\,\delta(a_{ij} - 1) + 0.5\,\delta(a_{ij} - r_1(\boldsymbol{\alpha},\boldsymbol{\beta}))$. Here $\delta(\cdot)$ denotes the Dirac delta function. In this example, with the probability of 25% we will have $a_{1j} = 1$ and $a_{2j} = r_1(\boldsymbol{\alpha},\boldsymbol{\beta})$ for which the maximum of $loss_j$ will be realized. □

G MISSING PROOFS

G.1 Deferred Proofs from Section 3

Proof of Lemma 3.1. First of all, observe that $\max_x f(\boldsymbol{v}(x))$ can be formulated as a linear program

$$\begin{aligned} \max_{x,z} \quad & u_1(x) + u_2(x) - z \\ \text{s.t.} \quad & c_{ij}(x) = x_{ij} \geq 0\,,\ \forall i, j\,, \\ & c_j(x) = 1 - \sum_i x_{ij} = 0\,,\ \forall j\,, \\ & \tilde{c}_1(x,z) = z - 2\bar{\alpha}_1(u_1(x) - u_2(x)) \geq 0\,, \\ & \tilde{c}_2(x,z) = z - 2\bar{\alpha}_2(u_2(x) - u_1(x)) \geq 0\,, \end{aligned}$$

where $\bar{\alpha}_1 := (\beta_1 + \alpha_2)/2$ and $\bar{\alpha}_2 := (\alpha_1 + \beta_2)/2$. In the case of linear programming, KKT conditions are necessary and sufficient to characterize the optimum. In order to write KKT conditions, we first find the gradients of the objective function $h(x,z) := -u_1(x) - u_2(x) + z$ and constraints w.r.t. x_{ij} and z:

$$\begin{aligned} &\nabla_{ij} h = -a_{ij}\,, && \nabla_z h = 1\,, \\ &\nabla_{ij} c_{kl} = \mathbb{1}\{ij = kl\}, && \nabla_z c_{kl} = 0\,, \\ &\nabla_{ij} c_l = -\mathbb{1}\{j = l\}\,, && \nabla_z c_l = 0\,, \\ &\nabla_{ij} \tilde{c}_k = -2\chi\{i = k\}\bar{\alpha}_k a_{ij}\,, && \nabla_z \tilde{c}_k = 1\,. \end{aligned}$$

KKT conditions require

$$\nabla_{ij} h = -a_{ij} = \sum_{k,l} \lambda_{kl} \nabla_{ij} c_{kl} + \sum_l \lambda_l \nabla_{ij} c_l + \sum_k \tilde{\lambda}_k \nabla_{ij} \tilde{c}_k = \lambda_{ij} - \lambda_j - 2(\bar{\alpha}_i \tilde{\lambda}_i - \bar{\alpha}_{-i} \tilde{\lambda}_{-i}) a_{ij}\,,\ \forall i, j\,, \tag{83}$$

$$\nabla_z h = 1 = \tilde{\lambda}_1 + \tilde{\lambda}_2\,, \tag{84}$$

$$\lambda_{ij}, \tilde{\lambda}_i \geq 0\,,\ \forall i, j\,, \tag{85}$$

$$\sum_i x_{ij} = 1\,,\ \forall j\,, \tag{86}$$

$$\lambda_{ij} c_{ij} = 0\,,\ \forall i, j\,, \tag{87}$$

$$\tilde{\lambda}_i \tilde{c}_i = 0\,,\ \forall i\,. \tag{88}$$

Let us define $\Delta_i := 2\bar{\alpha}_i \tilde{\lambda}_i - 2\bar{\alpha}_{-i} \tilde{\lambda}_{-i} = -\Delta_{-i}$. Eq. (84) and nonnegativity of $\tilde{\lambda}_i$ and $\tilde{\lambda}_{-i}$ ensure $\Delta_i \in [-2\bar{\alpha}_{-i}, 2\bar{\alpha}_i]$.

Using Eq. (83) to find $\nabla_{-ij} h - \nabla_{ij} h$, we have $a_{-ij}(\Delta_{-i} - 1) + a_{ij}(1 - \Delta_i) = \lambda_{-ij} - \lambda_{ij}$. If $x_{ij} > 0$, we need $\lambda_{ij} = 0$. Then $\lambda_{-ij} \geq 0$ requires $\Delta_i \leq \frac{a_{ij} - a_{-ij}}{a_{ij} + a_{-ij}}$. Further, if $1 > x_{ij} > 0$, then $\lambda_{-ij} = 0$ and $\Delta_i = \frac{a_{ij} - a_{-ij}}{a_{ij} + a_{-ij}}$. This proves Eq. (13).

Finally, if $\tilde{\lambda}_1, \tilde{\lambda}_2 > 0$, we need $\tilde{c}_1 = \tilde{c}_2 = 0$. This only happens if $u_1(x) = u_2(x)$. Otherwise, one of $\tilde{\lambda}_1$ or $\tilde{\lambda}_2$ will be zero and Δ_i will be either $2\bar{\alpha}_i$ or $-2\bar{\alpha}_{-i}$. Specifically, if $u_i(x) > u_{-i}(x)$, we will have $\tilde{c}_i(x) = 0$ and $\tilde{c}_{-i}(x) > 0$, which gives $\tilde{\lambda}_i > 0$ and $\tilde{\lambda}_{-i} = 0$. So, in this case $\Delta_i = 2\bar{\alpha}_i$. □

Proof of Theorem 3.2. Define $\Delta a_j := a_{1j} - a_{2j}$ and $\mathcal{J}_1 := \{j \mid a_{1j} > a_{2j}\}$. The allocation $x^{\alpha,\beta}$ only reallocates items allocated to agent 1 under x^*, i.e., $\mathcal{J}_1$. Lemma 3.1 shows among items in $\mathcal{J}_1$ only $\tilde{\mathcal{J}}_1(\Delta_1) = \{j \mid j \in \mathcal{J}_1, \Delta_1 \geq \Delta a_j/(a_{1j} + a_{2j})\}$ are subject to reallocation. The loss due to allocating item $j \in \mathcal{J}_1$ to agent 2 is Δa_j. If all items in $\tilde{\mathcal{J}}_1(\Delta_1)$ are allocated to agent 2, the loss will be bounded by $\sum_{j \in \tilde{\mathcal{J}}_1(\Delta_1)} \Delta a_j$. Since $\tilde{\mathcal{J}}_1(\cdot)$ is a monotone increasing set and $\Delta_1 \leq \beta_1 + \alpha_2$, the loss can be bounded by $\sum_{j \in \tilde{\mathcal{J}}_1(\beta_1 + \alpha_2)} \Delta a_j = \sum_{j \in \mathcal{J}_1} \Delta a_j \, \mathbb{1}\{\beta_1 + \alpha_2 \geq \Delta a_j/(a_{1j} + a_{2j})\}$. Rearranging the terms in $\mathbb{1}\{\cdot\}$ completes the proof of Eq. (14). Setting $a_{2j} = r(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}$, the rest of the proof is straightforward. □

Proof of Lemma 3.3. Plugging the $I_i^+(x)$ and I_i^- from Eqs. (10) and (11) into agent i's aggregated value gives

$$v_i(x) = u_i(x) - \frac{1}{n-1} \alpha_i \sum_{k \neq i} \max\{u_k(x) - u_i(x), 0\} - \frac{1}{n-1} \beta_i \sum_{k \neq i} \max\{u_i(x) - u_k(x), 0\}\,. \tag{89}$$

Maximization of $f(\boldsymbol{v}(x)) = \sum_i v_i(x)$ is then equivalent to solving the linear program

$$\begin{aligned}
\max_{x,z} \quad & -h(x,z) := \sum_i \Big(u_i(x) - \frac{1}{n-1} \sum_{k \neq i} z_{ik} \Big) \\
\text{s.t.} \quad & \underline{c}_{ij}(x) = x_{ij} \geq 0\,,\ \forall i, j\,, \\
& \bar{c}_{ij}(x) = b_j - x_{ij} \geq 0\,,\ \forall i, j\,, \\
& c_j(x) = 1 - \sum_i x_{ij} \geq 0\,,\ \forall j\,, \\
& \underline{c}^z_{ik}(x,z) = z_{ik} - \beta_i(u_i(x) - u_k(x)) \geq 0\,,\ \forall k \neq i\,, \\
& \bar{c}^z_{ik}(x,z) = z_{ik} - \alpha_i(u_k(x) - u_i(x)) \geq 0\,,\ \forall k \neq i\,.
\end{aligned}$$

In the case of linear programming, KKT conditions are necessary and sufficient to characterize the optimum. Let $\underline{\lambda}_{ij}$, $\bar{\lambda}_{ij}$, γ_j, $\underline{\xi}_{ik}$, and $\bar{\xi}_{ik}$ be corresponding KKT multipliers of $\underline{c}_{ij}$, $\bar{c}_{ij}$, c_j, $\underline{c}^z_{ik}$, and $\bar{c}^z_{ik}$, respectively. KKT conditions require

$$\begin{aligned}
\nabla_{x_{ij}} h = -a_{ij} &= \sum_{k,l} \underline{\lambda}_{kl} \nabla_{x_{ij}} \underline{c}_{kl} + \sum_{k,l} \bar{\lambda}_{kl} \nabla_{x_{ij}} \bar{c}_{kl} + \sum_l \gamma_l \nabla_{x_{ij}} c_l + \sum_{i' \neq k'} \underline{\xi}_{i'k'} \nabla_{x_{ij}} \underline{c}^z_{i'k'} + \sum_{i' \neq k'} \bar{\xi}_{i'k'} \nabla_{x_{ij}} \bar{c}^z_{i'k'} \\
&= \underline{\lambda}_{ij} - \bar{\lambda}_{ij} - \gamma_j + a_{ij} \sum_{k \neq i} (-\beta_i) \underline{\xi}_{ik} + a_{ij} \sum_{k \neq i} \beta_k \underline{\xi}_{ki} + a_{ij} \sum_{k \neq i} \alpha_i \bar{\xi}_{ik} + a_{ij} \sum_{k \neq i} (-\alpha_k) \bar{\xi}_{ki}\,, \qquad (90)
\end{aligned}$$

$$\nabla_{z_{ik}} h = \frac{1}{n-1} = \sum_{i' \neq k'} \underline{\xi}_{i'k'} \nabla_{z_{ik}} \underline{c}^z_{i'k'} + \sum_{i' \neq k'} \bar{\xi}_{i'k'} \nabla_{z_{ik}} \bar{c}^z_{i'k'} = \underline{\xi}_{ik} + \bar{\xi}_{ik}\,, \qquad (91)$$

$$\underline{\lambda}, \bar{\lambda}, \gamma, \underline{\xi}, \bar{\xi} \geq 0\,, \qquad (92)$$

$$\underline{\lambda}_{ij} \underline{c}_{ij}(x) = 0\,,\ \bar{\lambda}_{ij} \bar{c}_{ij}(x) = 0\,,\ \forall i, j\,, \qquad (93)$$

$$\gamma_j c_j(x) = 0\,,\ \forall j\,, \qquad (94)$$

$$\underline{\xi}_{ik} \underline{c}^z_{ik}(x,z) = 0,\ \bar{\xi}_{ik} \bar{c}^z_{ik}(x,z) = 0\,,\ \forall k \neq i\,. \qquad (95)$$

Define $\xi_{ik} := \alpha_i \bar{\xi}_{ik} - \beta_i \underline{\xi}_{ik}$ and denote its matrix form by ξ. Since $\underline{\xi}_{ik}$ and $\bar{\xi}_{ik}$ are nonnegative, Eq. (91) requires $\xi_{ik} \in [-\frac{\beta_i}{n-1}, \frac{\alpha_i}{n-1}]$. Define $\boldsymbol{\sigma} := (\xi - \xi^{\mathbb{T}})\mathbf{1}$, where $\mathbf{1}$ is a vector of all ones. The range on ξ requires

$$\sigma_i \in \Big[-\beta_i - \frac{1}{n-1} \sum_{k \neq i} \alpha_k,\ \alpha_i + \frac{1}{n-1} \sum_{k \neq i} \beta_k \Big]\,. \qquad (96)$$

Note that the inequality aversion levels are less than 1/2, so, $|\sigma_i| < 1/2$. Using the new definitions, we can rewrite Eq. (90) as

$$-a_{ij} = \underline{\lambda}_{ij} - \bar{\lambda}_{ij} - \gamma_j + \sigma_i a_{ij}\,. \qquad (97)$$

If $x_{kj} > 0$, then $\underline{\lambda}_{kj} = 0$ and Eq. (97) for agent k and item j gives $\gamma_j = (1 + \sigma_k) a_{kj} - \bar{\lambda}_{kj}$. Further, if $x_{ij} < b_j$, we have $\bar{\lambda}_{ij} = 0$. Then plugging γ_j into Eq. (97) gives

$$a_{kj}(1 + \sigma_k) - a_{ij}(1 + \sigma_i) = \underline{\lambda}_{ij} + \bar{\lambda}_{kj}\,. \qquad (98)$$

Since both $\underline{\lambda}$ and $\bar{\lambda}$ only take non-negative values and $\sigma_i > -1, \sigma_k > -1$, we can write

$$a_{kj} \geq a_{ij} \frac{1 + \sigma_i}{1 + \sigma_k}\,. \qquad (99)$$

Then using Eq. (96) and choosing extreme values for σ_i and σ_k, we have

$$a_{kj} \geq a_{ij} \frac{1 - \beta_i - \frac{1}{n-1} \sum_{i' \neq i} \alpha_{i'}}{1 + \alpha_k + \frac{1}{n-1} \sum_{k' \neq k} \beta_{k'}} = r_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{ij}\,. \qquad (100)$$

□

Proof of Theorem 3.4. Assume under x^*, agent i receives her full b_j share from item j. If $x^{\alpha,\beta}$ reallocates agent i's share of item j to agent k, a loss of $b_j(a_{ij} - a_{kj})$ will be incurred. Lemma 3.3 requires $a_{kj} \geq r_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{ij}$ for this reallocation to be possible. Hence, the loss is bounded by $b_j(1 - r_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta}))\, a_{ij}$. Defining

$$c_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta}) := 1 - r_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\beta_i + \alpha_k + (\|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\beta}\|_1 - \alpha_i - \beta_k)/(n-1)}{1 + \alpha_k + (\|\boldsymbol{\beta}\|_1 - \beta_k)/(n-1)}\,, \qquad (101)$$

we can rewrite the loss of reallocating agent i's share to agent k as $b_j\, c_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{ij}$. Since inequality aversion levels are less than $1/2$, $r_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta}) > 0$ and $c_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta}) < 1$. Then a simple argument shows

$$c_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \le \bar{c}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \frac{\beta_i + \alpha_m + (\|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\beta}\|_1)/(n-1)}{1 + \alpha_m + \|\boldsymbol{\beta}\|_1/(n-1)} \,. \tag{102}$$

Since initially only top_j agents have a share from item j, the loss only incurs if we get some share of agent $i \in top_j$ and give it to agent $k \neq j$. Hence, we can upperbound the loss from reallocation of item j by

$$loss_j \le \sum_{i \in top_j} b_j\, \bar{c}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{ij} \le \big(\max_{i \in top_j} \bar{c}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\big)\,\big(\sum_{i \in top_j} b_j a_{ij}\big) \tag{103}$$

$$\le \frac{\beta_{m,top_j} + \alpha_m + (\|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\beta}\|_1)/(n-1)}{1 + \alpha_m + \|\boldsymbol{\beta}\|_1/(n-1)} a_{top_j} \,. \tag{104}$$

□

Proof of Theorem 3.5. Based on Lemma 3.3, agent $i \in top_j$ might give up her share from item j to agent $k \neq i$ only if $a_{kj} \ge r_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{ij}$. Therefore, we can bound the possible loss due to a transfer from i to k by $\phi(a_{ij}; a_{kj}) := b\mathbb{1}\{a_{ij} \ge a_{kj} \ge r_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{ij}\}(a_{ij} - a_{kj})$. Let $\mathcal{J}(i)$ be the items that agent i has a share from. We can upperbound the overall loss by

$$loss \le \sum_i \sum_{j \in \mathcal{J}(i)} \max_k \phi(a_{ij}; a_{kj}) \,. \tag{105}$$

The max over agents can be upperbounded by the sum of max over agents in each cluster:

$$loss \le \sum_i \sum_{j \in \mathcal{J}(i)} \sum_{q \in [K]} \max_{k \in C_q} \phi(a_{ij}; a_{kj}) \,. \tag{106}$$

- If $i \in C_q$, for every $k \in C_q$, we have

$$\sum_{j \in \mathcal{J}(i)} \max_{k \in C_q} \phi(a_{ij}; a_{kj}) \le \sum_{j \in \mathcal{J}(i)} \max_{k \in C_q} |a_{ij} - a_{kj}| \le \sum_{j \in \mathcal{J}(i)} |\bar{a}_{qj} - \underline{a}_{qj}| \le \delta\, b\, \min\{1, |\mathcal{J}(i)|\} \,. \tag{107}$$

- If $i \notin C_q$, we break $\phi(a_{ij}; a_{kj})$ into two terms. Define $\bar{\phi}(a_{ij}; a_{kj}) := b\mathbb{1}\{a_{ij} \ge a_{kj}\}(a_{ij} - a_{kj})$ as an upper bound for $\phi(a_{ij}; a_{kj})$. Then, for every $k \in C_q$, one can verify

$$\phi(a_{ij}; a_{kj}) \le \phi(a_{ij}; \bar{a}_{qj}) + \bar{\phi}(\bar{a}_{qj}; a_{kj}) \,. \tag{108}$$

  The proof as follows: If $\bar{a}_{qj} \ge a_{ij}$, then $\phi(a_{ij}; a_{kj}) \le \bar{\phi}(\bar{a}_{qj}; a_{kj})$. If $a_{ij} > \bar{a}_{qj} \ge r_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{ij}$, then $a_{kj} \le \bar{a}_{qj}$ implies Eq. (108). Finally, if $r_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{ij} > \bar{a}_{qj}$, then $a_{kj} \le \bar{a}_{qj}$ implies $\phi(a_{ij}; a_{kj}) = 0$, so, Eq. (108) is obvious.
  Since cluster q has a radius of δ, we have

$$\sum_{j \in \mathcal{J}(i)} \max_{k \in C_q} \bar{\phi}(\bar{a}_{qj}; a_{kj}) \le \sum_{j \in \mathcal{J}(i)} \max_{k \in C_q} |\bar{a}_{qj} - a_{kj}| \le \sum_{j \in \mathcal{J}(i)} |\bar{a}_{qj} - \underline{a}_{qj}| \le \delta\, b\, \min\{1, |\mathcal{J}(i)|\} \,. \tag{109}$$

  In order to bound $\sum_{j \in \mathcal{J}(i)} \max_{k \in C_q} \phi(a_{ij}; \bar{a}_{qj}) = \sum_{j \in \mathcal{J}(i)} \phi(a_{ij}; \bar{a}_{qj})$, we apply results from the two-agent setting: Since $i \notin C_q$, agent i and cluster q's upper representative are γ-dissimilar. Applying Theorem D.1 from the two-agent setting requires defining a single $c_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$ for agent i. Observe from Eq. (102) that $c_{ik}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \le \bar{c}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Using a strategy similar to the proof of Corollary D.2 and setting $\beta = (s+1)/(s+2)$, we obtain:

$$\mathbb{E}\Big[\sum_{j \in \mathcal{J}(i)} \phi(a_{ij}; \bar{a}_{qj})\Big] = O\Big(b\,(\gamma m)^{\frac{(s+1)^2}{s+2}}\, |\mathcal{J}(i)|\, \bar{c}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\Big) \,. \tag{110}$$

  Note that without any distributional assumption, we cannot find a bound better than

$$O\Big(b\,(\gamma m)^{\frac{1}{2}}\, |\mathcal{J}(i)|\, \bar{c}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\Big) \,. \tag{111}$$

Putting these all together, we have

$$\mathbb{E}[loss] \le \sum_i \Big[\delta\, b\, K\, \min\{1, |\mathcal{J}(i)|\} + O\Big(b\,(K-1)\,(\gamma m)^{\frac{(s+1)^2}{s+2}}\, |\mathcal{J}(i)|\, \bar{c}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\Big)\Big] \tag{112}$$

$$= O\Big(\delta\, b\, K\, \min\{n, m\} + b\,(K-1)\,(\gamma m)^{\frac{(s+1)^2}{s+2}} \sum_i |\mathcal{J}(i)|\, \bar{c}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\Big) \,. \tag{113}$$

Suppose that in the case of a tie when finding x^*, a random agent gets the complete share, so that all elements of x^* are either 0 or b. Note that this assumption has no effect on $f(\boldsymbol{u}(x^*))$ and only simplifies the analysis by ensuring $\sum_i \mathbb{1}\{j \in \mathcal{J}(i)\} \leq 1/b$ for every item j. Then using a similar argument as in Eq. (103), we can rewrite the sum over agents in the above equation as

$$\sum_i |\mathcal{J}(i)|\, \bar{c}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_j \sum_{i \in top_j} \bar{c}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leq \sum_j \frac{1}{b} c_j(\boldsymbol{\alpha}, \boldsymbol{\beta})\,. \tag{114}$$

Plugging this into Eq. (113) completes the proof. □

Proof of Theorem 3.6. We start with the proof intuition and then present the formal proof.

Proof Intuition. The idea behind the proof is as follows. Roughly speaking, a loss of at most $b_j\, c_j(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{top_j}$ incurs if an agent $i \in top_j$ losses her share to an agent $k \notin top_j$, where $a_{ij} \geq a_{kj} \geq (1 - c_j(\boldsymbol{\alpha}, \boldsymbol{\beta}))a_{ij}$. If there were only two agents ($b_j = 1, a_{kj} = a_{-ij}$), we could argue that if agents are independent and g_j is smooth, the probability that a_{kj} lies in this narrow band is $p_j = O(c_j(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{ij})$, so the expected loss from the reallocation of item j will be $p_j\, c_j(\boldsymbol{\alpha}, \boldsymbol{\beta}) = O(c_j^2(\boldsymbol{\alpha}, \boldsymbol{\beta}))$. However, in an n-agent setting with large n, this probability is roughly $1 - (1 - c_j(\boldsymbol{\alpha}, \boldsymbol{\beta}))^{n-1} \approx 1$. So, in general, for $1/b_j$ winners of item j, we cannot hope to get a bound smaller than $c_j(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{top_j}$ for this item.

But there is another effect neglected: Let $a_{(i)j}$ be the i^{th} largest utility coefficient for item j. Lemma 3.3 implies that for $i \leq 1/b_j$, an agent with utility coefficient $a_{(i)j}$ can lose her share to an agent with utility coefficient $a_{(k)j}$ ($k > 1/b_j$) only if $a_{(k)j} \geq (1 - c_j(\boldsymbol{\alpha}, \boldsymbol{\beta}))a_{(i)j}$. Since $k > 1/b_j$, this requires $a_{(1/b_j)j} \geq (1 - c_j(\boldsymbol{\alpha}, \boldsymbol{\beta}))a_{(i)j}$. Therefore, for example, if all $1/b_j$ agents in the top_j are going to lose their share, then $a_{(1)j}, \ldots, a_{(1/b)j}$ should all lie in a narrow band with length $c_j(\boldsymbol{\alpha}, \boldsymbol{\beta})$. The number of agents falling in this band follows $Z \sim Binomial(n, p_j)$. Then using Markov's inequality, the probability to have $1/b_j$ or more positive draws is bounded by $\Pr\left(Z \geq 1/b_j\right) \leq b_j\, \mathbb{E}[Z] = np_j b_j$. So, roughly speaking, $loss_j$ in this case is $O(np_j b_j c_j(\boldsymbol{\alpha}, \boldsymbol{\beta}))$ which is $O(c_j^2(\boldsymbol{\alpha}, \boldsymbol{\beta}))$ if nb_j is a bounded constant. For brevity, we drop the index j from $c_j(\boldsymbol{\alpha}, \boldsymbol{\beta})$, $r_j(\boldsymbol{\alpha}, \boldsymbol{\beta}) \coloneqq 1 - c_j(\boldsymbol{\alpha}, \boldsymbol{\beta})$, and b_j, when it is clear from the context, and use the shorthands c, r, and b instead. Define $a_{(i)j}$ as the i^{th} largest value among all $\{a_{kj}\}_k$.

Formal Proof. Since x^* already allocates item j to top_j, we do not need to consider the reallocation of j between top_j. Lemma 3.3 implies that the agent with utility coefficient $a_{(i)j}$ ($i \leq 1/b$) can lose her share to an agent with utility coefficient $a_{(k)j}$ ($k > 1/b$) only if $a_{(k)j} \geq ra_{(i)j}$. Since $k > 1/b$, this requires $a_{(1/b)j} \geq ra_{(i)j}$. Consider $a_{(1)j}, a_{(2)j}, \ldots, a_{(1/b-1)j}$ to be fixed. Then the value of $a_{(1/b)j}$ determines the highest rank agent among top_j that might lose her share to an agent not it top_j: $t = \min\{i :\ a_{(1/b)j} \geq ra_{(i)j}\}$. For instance, if $a_{(1/b)j} < ra_{(3)}$ but $a_{(1/b)j} \geq ra_{(4)j}$, we have $t = 4$, so agents 1 to 3 will not lose their share to any agent not in top_j (though they might exchange goods between themselves) and agent 4 is the highest rank agent that might do so.

Now suppose agent $i \in top_j$ has lost her share to agent $k \notin top_j$. The loss of this reallocation is $a_{ij} - a_{kj}$ which will be less than $a_{(t)j} - a_{kj}$. Let $loss_{kj}$ be the random variable describing the loss an agent $k \notin top_j$ imposes. Using a similar notation as Lemma E.1, we can write $loss_{kj} \leq ba_{(t)j}(\widehat{loss}_{kj}|a_{(t)j})$. Define $loss_{(i)j}$ to be the i^{th} largest value of $\{loss_{kj}\}_{k \notin top_j}$. Putting these all together, given $a_{(1)j}, a_{(2)j}, \ldots, a_{(1/b-1)j}$, we can bound $loss_j$ by

$$loss_j|a_{(1)j}, \cdots, a_{(1/b-1)j} = \mathbb{1}\{a_{(1/b)j} \geq ra_{(1)j}\}[ba_{(1)j} \sum_{i=1}^{1/b} \mathbb{E}[\widehat{loss}_{(i)j}|a_{(1)j}]] \tag{115}$$

$$+ \mathbb{1}\{ra_{(1)j} > a_{(1/b)j} \geq ra_{(2)j}\}[ba_{(2)j} \sum_{i=1}^{1/b-1} \mathbb{E}[\widehat{loss}_{(i)j}|a_{(2)j}]] \tag{116}$$

$$+ \cdots + \mathbb{1}\{ra_{(1/b-1)j} > a_{(1/b)j}\}[ba_{(1/b)j}\mathbb{E}[\widehat{loss}_{(1)j}|a_{(1/b)j}]] \tag{117}$$

$$\leq b\mathbb{1}\{a_{(1/b)j} \geq ra_{(1)j}\}a_{(1)j}\frac{1}{b}\hat{M} \tag{118}$$

$$+ b \sum_{k=2}^{1/b-1} \mathbb{1}\{ra_{(k-1)j} > a_{(1/b)j} \geq ra_{(k)j}\}a_{(k)j}(\frac{1}{b} - k + 1)\hat{M} \tag{119}$$

$$+ b\mathbb{1}\{ra_{(1/b-1)j} > a_{(1/b)j}\}a_{(1b)j}\hat{M}\,. \tag{120}$$

Here, $\hat{M} \coloneqq \max_{\bar{a}} \mathbb{E}[\widehat{loss}_{(1)j}|\bar{a}] = O(c)$ (refer to Lemma E.1). We start by approximating the expectation of Eq. (118):

$$\mathbb{E}[\mathbb{1}\{a_{(1/b)j} \geq ra_{(1)j}\}a_{(1)j}\hat{M}] \leq \hat{M}\mathbb{E}[1 - \hat{G}_j^{(1/b-1)}(r|a_{(1)j})]\,. \tag{121}$$

Here, $\hat{G}_j^{(1/b-1)}$ can be expanded as

$$\hat{G}_j^{(1/b-1)}(r|a_{(1)j}) = \sum_{t=0}^{1/b-2} \binom{n-1}{t} \hat{G}_j^{n-1-t}(r|a_{(1)j})\big(1 - \hat{G}_j(r|a_{(1)j})\big)^t\,. \tag{122}$$

It is straightforward to show for $1 - \hat{G}_j(r|a_{(1)j}) \leq \kappa_j c \leq 1/(bn)$, the above sum has almost all of the important terms of a binomial expansion and $\hat{G}_j^{(1/b-1)}(r|a_{(1)j}) = 1 - o(1)$. Therefore Eq. (118) can be bounded by $o(\hat{M})$. Next, we approximate expectation of the k^{th} term of Eq. (119):

$$\begin{aligned}&\mathbb{E}[b\mathbb{1}\{ra_{(k-1)j} > a_{(1/b)j} \geq ra_{(k)j}\}a_{(k)j}(\frac{1}{b} - k + 1)\hat{M}] = \\ &\qquad b\hat{M}(\frac{1}{b} - k + 1)\mathbb{E}\Big[a_{(k)j}(G_j^{(1/b-k)}(ra_{(k-1)j}|a_{(k)j}) - G_j^{(1/b-k)}(ra_{(k)j}|a_{(k)j}))\Big] .\end{aligned} \tag{123}$$

Here, the difference of G terms can be approximated by

$$G_j^{(1/b-k)}(ra_{(k-1)j}|a_{(k)j}) - G_j^{(1/b-k)}(ra_{(k)j}|a_{(k)j}) \approx g_j^{(1/b-k)}(ra_{(k)j}|a_{(k)j})r(a_{(k-1)j} - a_{(k)j}) . \tag{124}$$

Plugging the definition of $g_j^{(1/b-k)}$, using $ra_{(k)j}g(ra_{(k)j}|a_{(k)j}) \leq \kappa_j$, and summing up the terms for $k = 2$ to $1/b$, one can obtain the following upper bound on Eq. (119):

$$b\hat{M}n\kappa_j \sum_{t=0}^{1/b-3} (t+2)\binom{n}{t}\mathbb{E}\Big[\hat{G}_j^{n-1-t}(r|a_{(k)j})\big(1 - \hat{G}_j(r|a_{(k)j})\big)^t(a_{(k-1)j} - a_{(k)j})\Big] . \tag{125}$$

For a large enough n, the probability that $a_{(k-1)j} - a_{(k)j}$ takes a value much larger than $1/n$ goes to zero. The $(t+2)$ factor inside the sum can be upperbounded by $1/b$. But, if $1/b > n\kappa_j c$, then the terms corresponding to large ts will be negligible and it suffices to sum up only the first $n\kappa_j c$ terms. Upperbounding $(t+2)$ factor, the remaining terms can be upperbounded by a binomial expansion of 1. So, putting these all together, the expectation of Eq. (119) can be bounded with high probability by

$$O(\kappa_j \hat{M} \min\{1, nb\kappa_j c\}) . \tag{126}$$

Finally, Eq. (120) is obviously bounded by $b\hat{M}$. Since $n \longrightarrow \infty$ and the ratio of n and $1/b$ is constant, $1/b$ also goes to infinity and Eq. (120) becomes negligible. This completes the proof.

□

G.2 Deferred Proofs from Section 4

Proof of Lemma 3.1. Consider an item j such that $a_{1j} > a_{2j}$. There are three possibilities: 1) If $a_{2j} > r_1(\Delta_1)\, a_{1j}$, an immediate result of Lemma 3.1 is $x_{2j} = 1$. In this case, the loss of overall utility is $a_{1j} - a_{2j}$ but the inequality is also reduced by $a_{1j} + a_{2j}$. So, the social welfare based on true valuations has increased by $(\beta_1 + \alpha_2)(a_{1j} + a_{2j}) - (a_{1j} - a_{2j})$, which is reflected in Eq. (23). A simple calculation shows this term is non-negative for any $\Delta_1 \leq \beta_1 + \alpha_2$. 2) In the case of $a_{2j} = r_1(\Delta_1)\, a_{1j}$, for any arbitrary value of x_{2j}, the resulting gain is x_{2j} times the gain of full reallocation which is non-negative. 3) The allocation does not change if $a_{2j} < r_1(\Delta_1)\, a_{1j}$ and gain is zero in this case. So, overall, Eq. (23) gives a lower bound for the gain that can be realized from reallocation of item j. □

Proof of Proposition 4.2. Without loss of generality we assumed agent 1 is better off under x^*, so $\sum_{j:a_{1j}>a_{2j}} a_{1j} > f(\boldsymbol{u}(x^*))/2$, and $\Delta_1 > 0$. Agent 2 can be seen as an adversary with budget δ minimizing gain. Starting at the point where $a_{2j} \rightarrow^- a_{1j}$, $gain_j$ is as large as $\tilde{c}_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}$. To reduce this gain, agent 2 might spend $c_1(\Delta_1)\, a_{1j} := (1 - r_1(\Delta_1))a_{1j}$ from her dissimilarity budget to make $gain_j$ zero. The return rate of agent 2's investment or equivalently reduction rate in $gain_j$ is $\tilde{c}_1(\boldsymbol{\alpha}, \boldsymbol{\beta})/c_1(\Delta_1)$. Hence, agent 2's best strategy is to greedily spend her money on items with the smallest a_{1j} and make a_{2j} sufficiently different on those axes. Ideally, this will reduce the total gain by $\delta\, \tilde{c}_1(\boldsymbol{\alpha}, \boldsymbol{\beta})/c_1(\Delta_1)$, resulting in $gain \geq (\sum_{j:a_{1j}>a_{2j}} \tilde{c}_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\, a_{1j}) - \delta\, \tilde{c}_1(\boldsymbol{\alpha}, \boldsymbol{\beta})/c_1(\Delta_1)$. On the other hand, we already know the loss in the case of δ-similar agents is upperbounded by δ. Putting these together and treating $\tilde{c}_1(\boldsymbol{\alpha}, \boldsymbol{\beta})/c_1(\Delta_1)$ as a constant completes the proof. □

Proof of Proposition 4.3. For any item j, there are three possibilities: 1) If $a_{2j} > r_1(\Delta_1)\, a_{1j}$

$$\min_{a_{2j}} \frac{gain_j}{loss_j} = \lim_{a_{2j} \to^+ r_1(\Delta_1)\, a_{1j}} \frac{[(1 + \beta_1 + \alpha_2)a_{2j} - (1 - \beta_1 - \alpha_2)a_{1j}]}{a_{1j} - a_{2j}} \tag{127}$$

$$= \frac{(\beta_1 + \alpha_2)(1 + r_1(\Delta_1)) - (1 - r_1(\Delta_1))}{1 - r_1(\Delta_1)} = \frac{\beta_1 + \alpha_2}{\Delta_1} - 1 . \tag{128}$$

Note that this ratio does not depend on j. 2) For $a_{2j} = r_1(\Delta_1)\, a_{1j}$, for any value of x_{2j}, we have $gain_j = x_{2j}[(1+\beta_1+\alpha_2)a_{2j} - (1-\beta_1-\alpha_2)a_{1j}]$ and $loss_j = x_{2j}[a_{1j} - a_{2j}]$. Therefore, if $x_{2j} > 0$, a similar calculation as above shows $gain_j/loss_j = (\beta_1+\alpha_2)/\Delta_1 - 1$. 3) For $a_{2j} < r_1(\Delta_1)\, a_{1j}$, $x_{2j} = 0$ and both $loss_j$ and $gain_j$ are zero. Putting these all together, as long as $loss_j$ is non-zero for at least one item, $gain/loss \geq (\beta_1 + \alpha_2)/\Delta_1 - 1$. □

Proof of Proposition 4.4. We provide a constructive proof. Suppose there exists an agent i for which the proposition's conditions are met. Without loss of generality, assume $i = 1$. Consider the allocation of $2n - 1$ goods to n agents with $b_j = 1$ and the following utility coefficient:

$$a = \begin{bmatrix} 1+\epsilon & 0 & 0 & \dots & 0 & \epsilon & \epsilon & \dots & \epsilon \\ 0 & 1 & 0 & \dots & 0 & \epsilon(1-\epsilon) & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & \epsilon(1-\epsilon) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & \epsilon(1-\epsilon) \end{bmatrix}. \tag{129}$$

For sufficiently small $\epsilon > 0$, we have $x^*_{ij} = \mathbb{1}\{i = j\} + \mathbb{1}\{j > n, i = 1\}$, which gives the first n items to each agent, and the last $n-1$ items to agent 1. Under an inequality-averse allocation, one can verify $x^{\alpha,\beta}_{ij} = \mathbb{1}\{i = j\} + \mathbb{1}\{j > n, i = j - n + 1\}$, which divides the final $n-1$ goods between the final $n-1$ agents. Thus,

$$gain = -(n-1)\epsilon^2 + \epsilon(2-\epsilon)\sum_{i>1}(\beta_1 + \alpha_i)\,, \tag{130}$$

$$loss = (n-1)\epsilon^2\,. \tag{131}$$

Then in the limit of $\epsilon \rightarrow^+ 0$, although gain and loss both go to zero, their ratio goes infinity. This is happening despite every two agents being γ-dissimilar with $\gamma \rightarrow^+ 0$. Therefore, dissimilarity constraints are not helpful in upperbounding $gain/loss$. □

G.3 Deferred Proofs from Section 5

Proof of Proposition 5.1. For $0 < \epsilon_1 < \epsilon_2 \leq (m-1)(1-\epsilon_1)$, consider the utility profile

$$a = \begin{bmatrix} 1-\epsilon_1 & 1-\epsilon_2 & 1-\epsilon_2 & \dots & 1-\epsilon_2 \\ \frac{\epsilon_1}{m-1} & \frac{\epsilon_2}{m-1} & \frac{\epsilon_2}{m-1} & \dots & \frac{\epsilon_2}{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\epsilon_1}{m-1} & \frac{\epsilon_2}{m-1} & \frac{\epsilon_2}{m-1} & \dots & \frac{\epsilon_2}{m-1} \end{bmatrix}.$$

Then the allocation

$$x^{\alpha,\beta} = \begin{bmatrix} 1-y & \frac{y}{n-1} & \frac{y}{n-1} & \dots & \frac{y}{n-1} \\ 0 & \frac{1}{m-1} & \frac{1}{m-1} & \dots & \frac{1}{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{1}{m-1} & \frac{1}{m-1} & \dots & \frac{1}{m-1} \end{bmatrix}$$

improves $f(v(\cdot))$ for some value of y (which we will return to), if $\frac{2\alpha}{n-1}(1-\epsilon_1-\epsilon_2) \geq f(u(x^*) - f(u(x^{\alpha,\beta}))$, where $\alpha = \frac{\sum_i I_i^+(x^*)+I_i^-(x^*)}{2(1-\epsilon_1-\epsilon_2)}$. Observe that as $\epsilon_1, \epsilon_2 \rightarrow 0$, this inequality is true for shrinking $\alpha_i = \beta_i$ values.

Given this utility profile, $\{1\} = \arg\max_i \frac{u_i(x^*)}{u_i(x^{\alpha,\beta})}$, and $\epsilon_1 \rightarrow 0 \implies u_i(x^*) \rightarrow 1$. Moreover, we claim that as $\epsilon_1, \epsilon_2 \rightarrow 0$, that $u_i(x^{\alpha,\beta}) \rightarrow \frac{1}{n}$, which gives the result.

To see that as $\epsilon_1, \epsilon_2 \rightarrow 0 \implies u_i(x^{\alpha,\beta}) \rightarrow \frac{1}{n}$, consider that $\alpha > \frac{n-1}{2}(f(u(x^*) - f(u(x^{\alpha,\beta})))$ implies $\sum_i I_i^+(x^{\alpha,\beta}) + I_i^-(x^{\alpha,\beta}) = 0$. In order for $I(x^{\alpha,\beta}) = 0$, one must have all utilities be the same, and therefore one must have

$$\begin{aligned} & u_1(x^{\alpha,\beta}) = u_2(x^{\alpha,\beta}) \\ \iff & (1-y)(1-\epsilon_1) = \frac{y}{n-1}(1-\epsilon_2) + \frac{\epsilon_2}{m-1} \\ \iff & 1 - \epsilon_1 - y + y\epsilon_1 = \frac{y}{n-1} - \frac{y\epsilon_2}{n-1} + \frac{\epsilon_2}{m-1} \\ \iff & (n-1)(1-\epsilon_1 - \frac{\epsilon_2}{m-1}) = y(n - \epsilon_2 - \epsilon_1 n + \epsilon_1) \\ \iff & y = \frac{(n-1)(1-\epsilon_1-\frac{\epsilon_2}{m-1})}{n-\epsilon_2-n\epsilon_1+\epsilon_1}\,, \end{aligned}$$

which tends to $\frac{n-1}{n}$ as $\epsilon_1, \epsilon_2 \rightarrow 0$. Therefore, $u_i(x^{\alpha,\beta}) = (1-\epsilon_1)(1 - \frac{(n-1)(1-\epsilon_1-\frac{\epsilon_2}{m-1})}{n-\epsilon_2-n\epsilon_1+\epsilon_1}) \rightarrow \frac{1}{n}$, yielding the result. □

Lemma G.1. *Suppose* $\boldsymbol{\alpha} = \boldsymbol{\beta} = \alpha\mathbf{1}$ *for some* $\alpha \geq 0$. *Let* x^e *be an allocation such that* $\sum_i I_i^+(x^e)+I_i^-(x^e) = 0$. *Moreover, let* $i \in \arg\max_{i'} \frac{u_{i'}(x^*)}{u_{i'}(x^{\alpha,\beta})}$. *If* $\frac{1}{n}\sum_j I_j(x^{\alpha,\beta}) \geq \sum_{i':u_{i'}(x^{\alpha,\beta})\geq u_i(x^{\alpha,\beta})}(u_{i'}(x^{\alpha,\beta}) - u_i(x^{\alpha,\beta}))$, *then* $u_i(x^{\alpha,\beta}) \geq u_i(x^e)$.

Proof. Let $I(x) = \sum_i I_i^+(x) + I_i^-(x)$ and $\mathcal{J}^\alpha = \{i' : u_{i'}(x^{\alpha,\beta}) \geq u_i(x^{\alpha,\beta})\}$. Let us consider the contrapositive: if $u_i(x^{\alpha,\beta}) < u_i(x^e)$, then $\frac{1}{n} I(x^{\alpha,\beta}) < \sum_{i' \in \mathcal{J}^\alpha}(u_{i'}(x^{\alpha,\beta}) - u_i(x^{\alpha,\beta}))$. Observe that if

$$f(v^\alpha(x^e)) = n(u_i(x^e)) > n(u_i(x^{\alpha,\beta})) \geq f(v^\alpha(x^{\alpha,\beta}))\,, \tag{132}$$

we contradict the $f(\boldsymbol{v}(\cdot))$-optimality of $x^{\alpha,\beta}$. The first equality follows as $I(x^e) = 0$, and therefore each agent receives the same utility, so the social welfare is n times that utility. Moreover, the strict inequality from the hypothesis of the contrapositive. It remains to show the last inequality:

$$\begin{aligned}
f(\boldsymbol{v}(x^{\alpha,\beta})) &= \sum_k \left(u_k(x^{\alpha,\beta}) - \frac{\alpha}{n(n-1)} \sum_{i' \neq k} |u_j(x^{\alpha,\beta}) - u_{i'}(x^{\alpha,\beta})|\right) && (133)\\
&= \sum_k u_k(x^{\alpha,\beta}) - \frac{1}{n} I(x^{\alpha,\beta}) && (134)\\
&= \sum_{k \in \mathcal{J}^\alpha} u_k(x^{\alpha,\beta}) + \sum_{k \notin \mathcal{J}^\alpha} u_j(x^{\alpha,\beta}) - \frac{1}{n} I(x^{\alpha,\beta}) && (135)\\
&\leq \sum_{k \in \mathcal{J}^\alpha} u_k(x^{\alpha,\beta}) + |\mathcal{J}^\alpha| u_i(x^{\alpha,\beta}) - \frac{1}{n} I(x^{\alpha,\beta}) && (136)\\
&= n u_i(x^{\alpha,\beta}) + \sum_{k \in \mathcal{J}^\alpha} \left(u_k(x^{\alpha,\beta}) - u_i(x^{\alpha,\beta})\right) - \frac{1}{n} I(x^{\alpha,\beta}). && (137)
\end{aligned}$$

Notably, if $\sum_{k \in \mathcal{J}^\alpha} \left(u_k(x^{\alpha,\beta}) - u_i(x^{\alpha,\beta})\right) \leq \frac{1}{n} I(x^{\alpha,\beta})$, then the last line is less than or equal to $n u_i(x^{\alpha,\beta})$, and we have shown the last inequality in the chain, yielding a contradiction on the optimality of $x^{\alpha,\beta}$. Therefore, if $\sum_{k \in \mathcal{J}^\alpha} \left(u_k(x^{\alpha,\beta}) - u_i(x^{\alpha,\beta})\right) \leq \frac{1}{n(n-1)} I(x^{\alpha,\beta})$, then $u_i(x^{\alpha,\beta})) \geq u_i(x^e)$. □

Proof of Proposition 5.2. By Lemma G.1, we have $u_i(x^{\alpha,\beta}) \geq u_i(x^e)$. Therefore,

$$IT(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{u_i(x^*)}{u_i(x^{\alpha,\beta})} \leq \frac{u_i(x^*)}{u_i(x^e)} \leq \max_{i'} \frac{u_{i'}(x^*)}{u_i(x^e)} = \max_{i'} \frac{u_{i'}(x^*)}{u_{i'}(x^e)} \leq \frac{\max_{i'} u_{i'}(x^*)}{\min_{i'} u_{i'}(x^*)}\,. \tag{138}$$

The last inequality follows as x^e is an efficient allocation; any re-allocation of goods from x^* to x^e must result in $u_{i'}(x^*) \leq u_{i'}(x^e)$. □