The Computational Rise and Fall of Fairness

John P. Dickerson
Carnegie Mellon University
dickerson@cs.cmu.edu

Jonathan Goldman
Carnegie Mellon University
jagoldma@andrew.cmu.edu

Jeremy Karp
Carnegie Mellon University
jkarp@andrew.cmu.edu

Ariel D. Procaccia
Carnegie Mellon University
arielpro@cs.cmu.edu

Tuomas Sandholm
Carnegie Mellon University
sandholm@cs.cmu.edu

Abstract

The fair division of indivisible goods has long been an important topic in economics and, more recently, computer science. We investigate the existence of *envy-free* allocations of indivisible goods, that is, allocations where each player values her own allocated set of goods at least as highly as any other player's allocated set of goods. Under additive valuations, we show that even when the number of goods is larger than the number of agents by a linear fraction, envy-free allocations are unlikely to exist. We then show that when the number of goods is larger by a logarithmic factor, such allocations exist with high probability. We support these results experimentally and show that the asymptotic behavior of the theory holds even when the number of goods and agents is quite small. We demonstrate that there is a sharp phase transition from nonexistence to existence of envy-free allocations, and that on average the computational problem is hardest at that transition.

Introduction

The allocation of goods to interested agents is a central tenet of society. Some goods, like land, are *divisible*: a mechanism can split a single good amongst multiple agents. Others, like the houses or cars in an estate sale or divorce proceedings, are *indivisible*: a mechanism must allocate each good to exactly one agent. A chief concern in the assignment of divisible and indivisible goods to agents—and in the employment of divorce lawyers—concerns defining and guaranteeing the *fairness* of the final allocation.

One formal notion of fairness is *envy-freeness*. An allocation of goods is *envy free (EF)* if each player values her own allocated set of goods at least as highly as any other player's allocated set of goods. While EF divisions exist for any number of players in the divisible goods case (see, e.g., (Procaccia 2013), and the references therein), it is not guaranteed that such fair allocations exist when indivisible goods are considered. Indeed, consider the simple case of a single good and two agents, both of which have positive value for the good. Allocating the good to either agent will result in envy from its empty-handed partner.

In this paper, we investigate the conditions under which EF allocations of indivisible goods exist, when agents' values of goods are drawn at random. Under additive valuations and rather general conditions on the distributions over values of individual goods, we characterize conditions for nonexistence, showing that even when the number of goods is larger than the number of agents by a linear fraction, an EF allocation is unlikely to exist (Theorem 1). We then show that when the number of goods is larger by a logarithmic factor than the number of agents, an EF allocation exists with high probability (Theorem 2). Thus, these asymptotic existence results are almost tight.

We support our theoretical results, which apply asymptotically, with an empirical exploration of the EF allocation problem on different distributions over valuations and different objectives over EF allocations using two integer programming models. The theory applies to each of these experiments even when the number of agents and goods is quite small. We also uncover a phenomenon common to many problems in artificial intelligence: that the hardest computational EF allocation problems on average occur during the (sharp) transition from nonexistence to existence.

Related Work

Fair division occupies an important place in AI research; see, e.g., Chevaleyre, Endriss, and Maudet (2007), Bouveret and Lang (2008), Chen et al. (2010), Cohler et al. (2011) and the survey by Chevaleyre et al. (2006). Among the many AI papers that study the EF allocation of indivisible goods, the work of Bouveret and Lang (2008) is of particular interest. They showed that determining the existence of an EF allocation is computationally hard. In contrast, we focus on *typical* instances, and show that EF allocations exist, or do not exist, with high probability.

Similarly, the phase transition phenomenon is a staple of AI research (Cheeseman, Kanefsky, and Taylor 1991; Hogg, Huberman, and Williams 1996). In a nutshell, constraint satisfaction problems (CSPs) typically have the curious property that as the problem becomes more constrained, the probability of the existence of a feasible solution sharply drops from 1 to 0. Around the same point where this phase transition occurs (known as *the critical value* of the *order parameter*), search algorithms experience a sharp spike in running time—a steep computational rise, then an equally steep fall as the problem becomes more constrained (so ruling out the existence of a solution becomes easier). We show that a similar phenomenon occurs in the context of the existence and computation of EF allocations.

The problem of coalitional manipulation in elections is another popular topic in computational social choice that has run the gauntlet from (i) worst-case complexity (Conitzer, Sandholm, and Lang 2007), through (ii) probabilistic existence and nonexistence results (Conitzer and Sandholm 2006; Procaccia and Rosenschein 2007; Xia and Conitzer 2008), to (iii) investigations of the phase transition at the threshold between nonexistence and existence (Walsh 2011; Mossel, Procaccia, and Rácz 2013). From the technical and conceptual viewpoints, though, our problem is completely different. Note that we tackle (ii) and (iii) simultaneously, and for the first time (in the context of fair division).

Brams, Kilgour, and Klamler (2014) design a mechanism for the EF allocation of indivisible goods, which also satisfies other desirable properties. While their scheme guarantees envy-freeness, it may not allocate all the goods. To ameliorate this shortcoming, they show that when the (ordinal) preferences of agents over goods are drawn uniformly at random, their scheme will allocate all goods with high probability, as the number of goods goes to infinity. Our existence result, Theorem 2, is significantly stronger in several ways: (i) it gives an exact relation between the number of agents and number of goods, instead of assuming that one is constant and the other goes to infinity, (ii) it holds under far milder assumptions on the probability distribution over instances, and (iii) it relies on an intuitively desirable allocation mechanism that gives each good to the agent that wants it the most, thereby maximizing social welfare (as we discuss below). Brams and Fishburn (2000) also work in a probabilistic model, but with only two agents; our results hold for any number of agents.

Our Model

Denote the set of agents by $N = \{1, \dots, n\}$, and the set of goods by G, where $|G| = m$. Agent i has utility $u_i(g) \in [0, 1]$ for good g; note that constraining the utilities to an interval is without loss of generality. We make the very common assumption that utilities are *additive*, that is, for a subset of goods $G' \subseteq G$ and agent $i \in N$, it holds that $u_i(G') = \sum_{g\in G'} u_i(g)$.

An *allocation* is a partition $A = (A_1, \dots, A_n)$ of the goods, where A_i is the bundle of goods allocated to agent $i \in N$. The allocation A is said to be *envy free (EF)* if and only if for any two agents $i, j \in N$, $u_i(A_i) \geq u_i(A_j)$, that is, each agent weakly prefers its own bundle to the bundle allocated to any other agent.

Distributions Over Utilities

For every agent i and good $g \in G$, the utilities $u_1(g), \dots, u_n(g)$ are drawn from a joint, *non-atomic* distribution $\mathcal{D}_n$ over $[0, 1]^n$, that is, for every $x \in [0, 1]$, $\Pr[u_i(g) = x] = 0$. Let us state two assumptions on $\mathcal{D}_n$, which hold for every $g \in G$; Theorem 1 will require the first, and Theorem 2 will require the second:

[A1] For all $i, j \in N$ such that $i \neq j$, $u_i(g)$ and $u_j(g)$ are independent and identically distributed.

[A2] For all $i, j \in N$,
$$\Pr[\arg\max_{k\in N} u_k(g) = \{i\}] = 1/n,$$
and there exist constants μ, μ^* such that
$$0 < \mathbb{E}[u_i(g) \mid \arg\max_{k\in N} u_k(g) = \{j\}] \leq \mu$$
$$< \mu^* \leq \mathbb{E}[u_i(g) \mid \arg\max_{k\in N} u_k(g) = \{i\}].$$

Let us illustrate these assumptions using two natural distributions that will be featured in our empirical results:

- UNIFORM(x, y): For each agent $i \in N$ and good $g \in G$, draw $u_i(g) \sim \mathcal{U}[x, y]$, where $\mathcal{U}$ is the uniform distribution.
- CORRELATED(x, y): Independently assign each good g an intrinsic base value $\mu_g \sim \mathcal{U}[x, y]$. Then, for each agent i, draw $u_i(g) \sim \mathcal{N}(\mu_g, \sigma_g)$, where $\mathcal{N}$ is the (truncated) normal distribution and $\sigma_g \propto \mu_g$.

First, consider UNIFORM. Clearly it satisfies [A1]. Assumption [A2] seems technical, but is actually quite mild. To be concrete, take UNIFORM$(0, 1)$, so the utilities are drawn uniformly at random in $[0, 1]$. The first part of [A2] holds due to symmetry. Moreover, in this case, $\mathbb{E}[u_i(g)] = \frac{1}{2}$, and $\mathbb{E}[\max_{k\in N} u_k(g)] = \frac{n}{n+1} \geq \frac{2}{3}$ (see, e.g., (Boutilier et al. 2012, Corollary 4.5)). Clearly
$$\mathbb{E}[u_i(g) \mid \arg\max_{k\in N} u_k(g) = \{j\}] \leq \mathbb{E}[u_i(g)],$$
and due to symmetry
$$\mathbb{E}[u_i(g) \mid \arg\max_{k\in N} u_k(g) = \{i\}] = \mathbb{E}[\max_{k\in N} u_k(g)],$$
so we can set $\mu^* = 2/3$ and $\mu = 1/2$. Assumption [A2] still holds if the utilities are drawn from an interval $[x, y] \subseteq [0, 1]$ (by scaling and shifting μ and μ^*).

Similarly, CORRELATED(x, x) satisfies both assumptions for any $x \in (0, 1)$—utilities are simply drawn i.i.d. from the same normal distribution. But when $x < y$, CORRELATED(x, y) only satisfies assumption [A2]. This distribution does not satisfy [A1], because for a fixed $g \in G$, $u_i(g)$ and $u_j(g)$ are not independent.

A Small Number of Goods

If there are fewer goods than agents, i.e., $m < n$, then clearly no EF allocation is possible—there will be an agent with no goods. Conceivably, though, it could be that if m is slightly larger than n—say, $m = n + \sqrt{n}$—then an EF allocation is likely to exist. In this section we show that this is not the case: the number of "extra" goods must be linear in n.

Recall that our distributions over utilities are non-atomic. Therefore, for $i \in N$ and two goods $g \neq g'$, $\Pr[u_i(g) = u_i(g')] = 0$. So, we can safely assume that each agent has a unique favorite good, and define a function $f : N \to G$ that maps each agent to its favorite good, that is, $f(i) = \text{argmax}_{g\in G} u_i(g)$.

We are now ready to state our first result.

Theorem 1. *Assume that [A1] holds. Let $\delta \in \left(0, \frac{1}{2} - \frac{1}{2\sqrt{e}}\right)$ be a constant. If the probability that there exists an EF allocation is at least $1-\delta$ then $m \geq (1+c(\delta))n$, where $c(\delta) > 0$ is a constant that depends only on δ.*

We require the following lemma, which gives a necessary condition for envy-freeness that depends on the number of collisions of the function f.

Lemma 1. *Let $u_1, \ldots, u_n$ be utility functions for the n agents such that $u_i(g) \neq u_i(g')$ for all $g \neq g'$. For each good g, let $X_g = f^{-1}(g)$ be the set of agents whose favorite good is g. If there is an EF allocation then $m \geq n + \sum_{g \in G} \max\{|X_g| - 1, 0\}$.*

Proof of Lemma 1. Fix an allocation A, and let $i \in N$ such that $g \in A_i$. In order to avoid envying i, every agent in $X_g \setminus \{i\}$ must receive at least two goods. Hence,

$$\begin{aligned} m &\geq \sum_{g \in G:\ |X_g| > 0} (2|X_g| - 1) \\ &= \sum_{g \in G:\ |X_g| > 0} |X_g| + \sum_{g \in G:\ |X_g| > 0} (|X_g| - 1) \\ &= \sum_{g \in G} |X_g| + \sum_{g \in G} \max\{|X_g| - 1, 0\} \\ &= n + \sum_{g \in G} \max\{|X_g| - 1, 0\}. \end{aligned}$$

□

Proof of Theorem 1. Let C be a random variable that counts the number of *collisions* between agents' top preferences. Specifically, the value of C is determined as follows. Starting from $C = 0$, for each $i = 1, \ldots, n$, if there exists $j < i$ such that $f(i) = f(j)$ then increment C by 1. Using the notations of Lemma 1, it is easy to see that $C = \sum_{g \in G} \max\{|X_g| - 1, 0\}$.

Our first goal is to compute $\mathbb{E}[C]$. Let Y_{ij} be a Bernoulli random variable that takes the value 1 if $f(i) = f(j)$, and 0 otherwise. Then for all $i \neq j$, $\Pr[Y_{ij} = 1] = 1/m$, due to assumption [A1].

Let Z_i be another Bernoulli random variable that takes the value 1 if $f(i) = f(j)$ for some $j < i$, and 0 otherwise. $Z_i = 1$ if and only if there exists $j < i$ such that $Y_{ij} = 1$. Furthermore, for a fixed i the variables Y_{ij} are independent due to assumption [A1]. Therefore

$$\mathbb{E}[Z_i] = 1 - \left(1 - \frac{1}{m}\right)^{i-1}.$$

Now we can simply write $C = \sum_{i \in N} Z_i$. Using the linearity of expectation:

$$\begin{aligned} \mathbb{E}[C] &= \mathbb{E}\left[\sum_{i \in N} Z_i\right] = \sum_{i=1}^{n} \left(1 - \left(1 - \frac{1}{m}\right)^{i-1}\right) \\ &\geq \sum_{i=1}^{n} \left(1 - e^{-\frac{i-1}{m}}\right) \geq \sum_{i=\frac{n}{2}+1}^{n} \left(1 - e^{-\frac{i-1}{m}}\right) \\ &\geq \frac{n}{2}\left(1 - e^{-\frac{n}{2m}}\right), \end{aligned}$$

where the third transition follows from the well-known fact that $(1-x) \leq e^{-x}$ for all $x \in \mathbb{R}$, and the fourth transition assumes that n is even purely for ease of exposition.

Now, suppose that $C \leq k$ with probability $1 - \delta$. Also using the fact that $C \leq n$, we get

$$(1-\delta)k + \delta n \geq \mathbb{E}[C] \geq \frac{n}{2}\left(1 - e^{-\frac{n}{2m}}\right),$$

and therefore

$$k \geq \frac{n}{1-\delta}\left(\frac{1 - e^{-\frac{n}{2m}}}{2} - \delta\right). \tag{1}$$

We want k to be lower-bounded by a constant fraction of n, which is true if and only if $e^{-\frac{n}{2m}} < 1 - 2\delta$. Denoting $m = \alpha n$, we can write $e^{-\frac{1}{2\alpha}} < 1 - 2\delta$; equivalently, $-\frac{1}{2\alpha} < \ln(1 - 2\delta)$, and by rearranging we get

$$\alpha < \frac{1}{2\ln\left(\frac{1}{1-2\delta}\right)}. \tag{2}$$

Using our assumption that $\delta < \frac{1}{2} - \frac{1}{2\sqrt{e}}$, we see that the right hand side of Equation (2) is a constant greater than 1. Let us therefore set

$$\beta(\delta) = \frac{1 + \frac{1}{2\ln\left(\frac{1}{1-2\delta}\right)}}{2},$$

then β is a constant greater than 1 that depends only on δ.

We now have all the ingredients in place to complete the theorem's proof. On the one hand, if $m > \beta(\delta)n$ then we are done. On the other hand, if $m \leq \beta(\delta)n$ then Equation (1) gives us a constant lower bound on k that depends only on δ, say $\gamma(\delta)n$, which was derived under the assumption that $\Pr[C \leq k] \geq 1 - \delta$. Next, set $\gamma'(\delta) = \gamma(\delta)/2$. So, if it holds that $\gamma'(\delta)n \leq k < \gamma(\delta)n$, i.e., k is a constant fraction of n that is strictly smaller than the lower bound, then the assumption does not hold, i.e., $\Pr[C \geq k] > \delta$.

By Lemma 1, in those cases where $C \geq k$, there is an EF allocation only if the number of goods is at least $n + k \geq n(1 + \gamma'(\delta))$. In other words, if $m < n(1 + \gamma'(\delta))$, an EF allocation would not exist with probability $1 - \delta$ for the preceding choices of parameters. We conclude that it must be the case that $m \geq \min\{\beta(\delta), 1 + \gamma'(\delta)\} \cdot n$. Setting $c(\delta) = \min\{\beta(\delta) - 1, \gamma'(\delta)\}$ completes the proof. □

A Large Number of Goods

Next, we examine the case where the number of goods is significantly larger than the number of agents—by a logarithmic factor, to be precise. In this case, an EF allocation exists with high probability.

Theorem 2. *Assume that [A2] holds. Let $n = O\left(\frac{m}{\ln m}\right)$. Then an EF allocation exists with probability 1 as $m \to \infty$.*

Before proving the theorem, two comments are in order. First, why are we writing $n = O\left(\frac{m}{\ln m}\right)$ instead of the more intuitive $m = \Omega(n \ln n)$? The reason is that we want to emphasize that only the number of goods has to go to infinity; the number of agents can stay small, even constant. The theorem holds *even* if the number of agents goes to infinity, as long as this happens at most at the specified rate compared to the number of goods.

Second, Theorem 2 states that there *exists* an EF allocation, but the proof shows something stronger: that this allocation can be obtained by giving each good to the agent that values it the most, i.e., to $\arg\max_{i\in N} u_i(g)$. This is, in fact, the allocation that maximizes the (utilitarian) *social welfare*, which is the sum of utilities. So, an alternative formulation is that, under the theorem's condition, the social-welfare-maximizing allocation is EF with high probability.

Turning to the theorem's proof, we require the following well-known result.

Lemma 2 (Chernoff). *Let $X^1, \ldots, X^m$ be independent random variables in $[0,1]$. Denote $X = \sum_{i=1}^m X^i$. Then for all $\epsilon \in [0,1]$,*

1. $\Pr[X \geq (1+\epsilon)\mathbb{E}[X]] \leq \exp\left(-\frac{\epsilon^2}{3}\mathbb{E}[X]\right)$.
2. $\Pr[X \leq (1-\epsilon)\mathbb{E}[X]] \leq \exp\left(-\frac{\epsilon^2}{2}\mathbb{E}[X]\right)$.

Proof of Theorem 2. We explicitly construct an allocation by giving each good $g \in G$ to the agent that likes it most, that is, to $\arg\max_{i\in N} u_i(g)$. This "algorithm" induces an allocation $A = (A_1, \ldots, A_n)$, where each A_i can be formally thought of as a random variable that takes values in 2^G. We prove the allocation A is EF with high probability.

Let X_i^g be a random variable that takes the value $u_i(g)$ if $\{i\} = \arg\max_{k\in N} u_k(g)$, and 0 otherwise. It holds that $u_i(A_i) = \sum_{g\in G} X_i^g$; we will use this observation to calculate $\mathbb{E}[u_i(A_i)]$. Using [A2] (twice), for all $i \in N$ and $g \in G$ it holds that $\mathbb{E}[X_i^g]$

$$= \Pr\left[\{i\} = \arg\max_{k\in N} u_k(g)\right] \cdot \mathbb{E}\left[u_i(g) \middle| \{i\} = \arg\max_{k\in N} u_k(g)\right]$$
$$= \frac{1}{n} \cdot \mathbb{E}\left[u_i(g) \;\middle|\; \{i\} = \arg\max_{k\in N} u_k(g)\right] \geq \frac{\mu^*}{n}.$$

Therefore, using the linearity of expectation,

$$\mathbb{E}[u_i(A_i)] = \sum_{g\in G} \mathbb{E}[X_i^g] \geq \mu^* \frac{m}{n}.$$

Next, for all $i \neq j$ and $g \in G$ let Y_{ij}^g be random variables that take the value $u_i(g)$ if $\{j\} = \arg\max_{k\in N} u_k(g)$, and 0 otherwise. It holds that $u_i(A_j) = \sum_{g\in G} Y_{ij}^g$. Furthermore

$$\mathbb{E}[Y_{ij}^g] = \frac{1}{n} \cdot \mathbb{E}\left[u_i(g) \;\middle|\; \{j\} = \arg\max_{k\in N} u_k(g)\right] \leq \frac{\mu}{n},$$

by assumption [A2]. However, a technicality is that our assumptions do not provide a lower bound for

$$\mathbb{E}\left[u_i(g) \;\middle|\; \{j\} = \arg\max_{k\in N} u_k(g)\right]$$

(which is required to use Lemma 2). We therefore define variables Z_{ij}^g such that $\mathbb{E}[Z_{ij}^g] = \mu/n$, $Z_{ij}^g \in [0,1]$, and Z_{ij}^g stochastically dominates Y_{ij}^g. In particular, $\mathbb{E}[\sum_{g\in G} Z_{ij}^g] = \mu\frac{m}{n}$, and due to stochastic dominance, for all $x \in \mathbb{R}^+$,

$$\Pr\left[\sum_{g\in G} Z_{ij}^g \geq x\right] \geq \Pr\left[\sum_{g\in G} Y_{ij}^g \geq x\right].$$

We can therefore use the Z_{ij}^g variables to reason about $u_i(A_j)$.

Let E_{ij} be the event that agent i envies agent j. For E_{ij} to happen, it must be the case that $\sum_{g\in G} Y_{ij}^g > \sum_{g\in G} X_i^g$, which happens only if

$$\sum_{g\in G} X_i^g \leq \mu^* \frac{m}{n} - \frac{\mu^*-\mu}{2}\frac{m}{n} = \left(1 - \frac{\mu^*-\mu}{2\mu^*}\right)\mu^* \frac{m}{n}$$
$$= \left(1 - \frac{\mu^*-\mu}{2\mu^*}\right)\mathbb{E}\left[\sum_{g\in G} X_i^g\right],$$

or

$$\sum_{g\in G} Y_{ij}^g \geq \mu^* \frac{m}{n} - \frac{\mu^*-\mu}{2}\frac{m}{n} = \mu\frac{m}{n} + \frac{\mu^*-\mu}{2}\frac{m}{n}$$
$$= \left(1 + \frac{\mu^*-\mu}{2\mu}\right)\mu\frac{m}{n}$$
$$= \left(1 + \frac{\mu^*-\mu}{2\mu}\right)\mathbb{E}\left[\sum_{g\in G} Z_{ij}^g\right],$$

Let us set

$$\epsilon = \min\left\{1, \frac{\mu^*-\mu}{2\mu^*}\right\}.$$

Because $\mu < \mu^*$, it also holds that $\epsilon \leq \frac{\mu^*-\mu}{2\mu}$.

The variables X_i^g and $X_i^{g'}$ are independent for $g \neq g'$, and similarly Z_{ij}^g and $Z_{ij}^{g'}$ are independent. Using Lemma 2, we have that

$$\Pr\left[\sum_{g\in G} X_i^g \leq (1-\epsilon)\mathbb{E}\left[\sum_{g\in G} X_i^g\right]\right] \leq \exp\left(-\frac{\epsilon^2}{2}\mu^*\frac{m}{n}\right),$$

and

$$\Pr\left[\sum_{g\in G} Y_{ij}^g \geq (1+\epsilon)\mathbb{E}\left[\sum_{g\in G} Z_{ij}\right]\right]$$
$$\leq \Pr\left[\sum_{g\in G} Z_{ij}^g \geq (1+\epsilon)\mathbb{E}\left[\sum_{g\in G} Z_{ij}\right]\right] \leq \exp\left(-\frac{\epsilon^2}{3}\mu\frac{m}{n}\right).$$

Setting

$$n \leq \frac{\epsilon^2\mu}{3} \cdot \frac{m}{\ln(2m^3)}$$

and using the union bound, we conclude that

$$\Pr[E_{ij}] \leq \exp\left(-\frac{\epsilon^2}{2}\mu^*\frac{m}{n}\right) + \exp\left(-\frac{\epsilon^2}{3}\mu\frac{m}{n}\right) \leq 2 \cdot \frac{1}{2m^3} = \frac{1}{m^3}. \tag{3}$$

The allocation A is EF if and only if E_{ij} does not occur for all $i \neq j$. Using Equation (3) and the union bound over $\binom{n}{2}$ pairs of agents, the probability that A is *not* EF is at most

$$\Pr\left[\bigvee_{i\neq j} E_{ij}\right] \leq \sum_{i\neq j} \Pr[E_{ij}] \leq \binom{n}{2}\frac{1}{m^3} \leq \frac{1}{m}.$$

Thus, the probability that A is not EF goes to zero as m grows. □

In Between: A Phase Transition

In this section, we support our theoretical results with an empirical exploration of the transition from nonexistence to existence of envy-free allocations as a function of the number of goods and agents. We find that the most difficult allocation problems occur during the sharp *phase transition* from nonexistence to existence. We show that this behavior, which is common to many discrete feasibility problems, holds under both of two natural optimization models (one with and one without an objective function) and under different distributions over agents' utility values.

Experimental Setup

We generate instances with n agents and m goods as follows by sampling valuations for each agent and each good from a given distribution over utility functions. In our experimental setup, we draw from two distributions—CORRELATED$(0.4, 0.6)$ and UNIFORM$(0, 1)$—defined earlier. Intuitively, the UNIFORM distribution randomly assigns a value to each good for each agent, while the CORRELATED distribution first draws an intrinsic value for each good, then assigns a random value to each agent drawn from a (truncated nonnegative normal) distribution around that intrinsic value. UNIFORM satisfies both distributional assumptions and thus aligns with both Theorems 1 and 2, while our instantiation of CORRELATED only satisfies assumption [A2], or the assumption needed for Theorem 2. Still, we will show that both theoretical results hold experimentally for both distributions, even when the number of agents and goods is quite small.

Given an instance as generated above, we search for an envy-free allocation using one of two mixed integer programs (MIPs). Both formulations use $n \times m$ binary variables x_{ig} that are activated if and only if agent i is allocated good g. Model #1, a feasibility problem, is defined as follows:

$$
\begin{array}{lll}
\text{find} & x_{ig} \quad \forall i \in N, g \in G & \\
\text{s.t.} & \sum_{i \in N} x_{ig} = 1 & \forall g \in G \\
& \sum_{g \in G} v_{ig} x_{i'g} - \sum_{g \in G} v_{ig} x_{ig} \leq 0 & \forall i \neq i' \in N \\
& x_{ig} \in \{0, 1\} & \forall i \in N, g \in G
\end{array}
$$

Intuitively, the first set of constraints ensures that each good is allocated to exactly one agent, while the second set of constraints ensures that each agent values its allocation at least as highly as any other agent's allocation. For this feasibility problem, no explicit objective function is necessary; indeed, the feasible region defined by the constraints is exactly the space of all envy-free allocations.

We now define Model #2, an optimization version of the envy-free allocation problem, as follows:

$$
\begin{array}{lll}
\text{min} & e & \\
\text{s.t.} & \sum_{i \in N} x_{ig} = 1 & \forall g \in G \\
& \sum_{g \in G} v_{ig} x_{i'g} - \sum_{g \in G} v_{ig} x_{ig} \leq e & \forall i \neq i' \in N \\
& x_{ig} \in \{0, 1\} & \forall i \in N, g \in G \\
& e \in \mathbb{R}^{\text{nonneg}} &
\end{array}
$$

This second MIP model minimizes a real-valued nonnegative variable e representing the maximum envy between any two agents; thus, an EF allocation exists if and only if the objective value is zero at the optimum. This is an integer programming-based implementation of the envy minimization problem described by Lipton et al. (2004).

Model #1 may seem like the more general model since it is amenable to the addition of various objective functions. For example, adding an objective function that maximizes $\sum_{i \in N} \sum_{g \in G} v_{ig} x_{ig}$ would produce an envy-free allocation that also maximizes social welfare. It is not obvious how to adapt Model #2 to include arbitrary objective functions. Still, there is some evidence that relaxing the feasible region and then re-casting the feasibility problem as an optimization problem may result in better runtime performance. For example, Sandholm, Gilpin, and Conitzer (2005) saw speedups using an optimization model instead of a feasibility model in specific problem classes when exploring various MIP models for finding Nash equilibria in two-player games (although they did not see an overall speedup). We compare the performance of both models in the coming section.

All experiments were performed in Python using IBM ILOG CPLEX 12.6[1] in single-threaded mode under its default configuration.[2] Runs were conducted on Blacklight,[3] a ccNUMA supercomputer with 8GB of RAM per core; each experiment was run at least 160 times with a time limit of 12 hours per run. For solve time comparison, runs that timed out were conservatively considered to have completed in 12 hours. When timeouts were ignored or penalized heavily (e.g., counted as a $10 \times 12 = 120$ hour run), our experiments exhibited the same qualitative behavior.

Phase Transitions

We now explore the existence of phase transitions in various instantiations of the envy-free allocation problem.

Figure 1 shows an example phase transition for the existence of, and hardness of finding, an envy-free allocation in a problem with $n = 10$ agents valuing $m \in \{10, \ldots, 30\}$ goods. Results are presented for both the UNIFORM and CORRELATED distributions over utility functions using Model #1 without and with a social welfare maximizing objective function. The thick red line (corresponding to the left y-axis) plots the fraction of instances with m goods and n agents such that an envy-free allocation existed.

Aligning with Theorem 1, Figure 1 shows that the probability of an EF allocation existing is small when the number of goods is not much larger than the number of agents. Similarly, aligning with Theorem 2, when the number of goods is more (but not necessarily substantially more), the probability of an EF allocation existing is essentially one. Figure 2 explores this transition quantitatively for increasing numbers of agents n by plotting the minimum value m where at least 99% of the generated instances were feasible. Fitting an $m/\ln(m)$ function for either UNIFORM or CORRELATED shows that the asymptotically-stated Theorem 2 holds even when the number of goods and agents is quite small.

Figure 1 also plots runtime as a function of the number of goods m. The thick dashed line (corresponding to the right y-axis) plots the median runtime to either prove the nonexistence of a solution or find and prove the optimality of a feasible solution. The two dotted lines (also corresponding to

[1] ibm.com/software/commerce/optimization/cplex-optimizer/

[2] Source code & data: https://github.com/JohnDickerson/EnvyFree

[3] blacklight.psc.edu

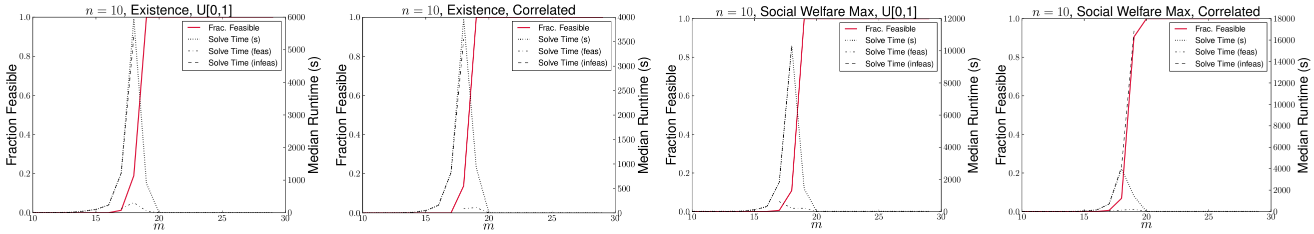

Figure 1: Phase transition for $n = 10$ under either UNIFORM or CORRELATED, with or without maximizing social welfare.

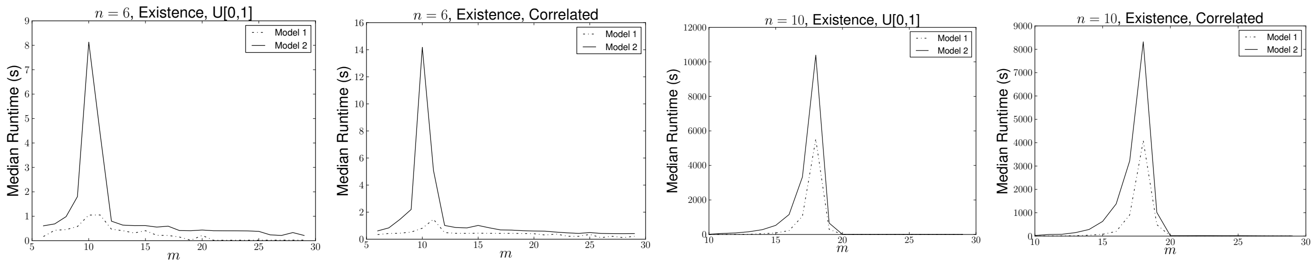

Figure 3: Runtime comparison of Model #1 (feasibility) and Model #2 (optimization) for $n = 6$ and $n = 10$ agents.

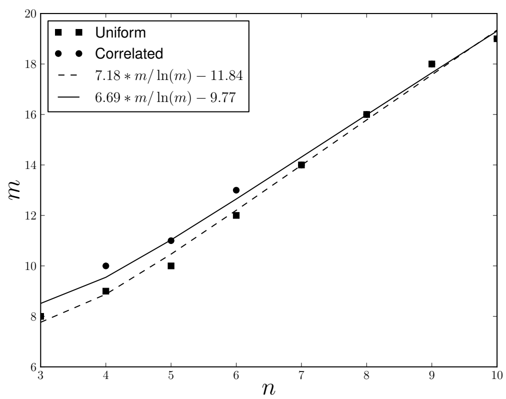

Figure 2: The minimum value of m where at least 99% of the instances were feasible as n increases.

the right y-axis) plot the median runtimes for only the feasible and infeasible instances, respectively. We see a classical "hardness bump" around the phase transition, with median solution time being much higher when the probability of a feasible instance is small but not trivial. Here, proving infeasibility takes significant computational effort.

Figure 3 shows that this hardness behavior is not just an artifact of the feasibility Model #1; indeed, the optimization problem defined by Model #2 exhibits an even more stark hardness bump around the phase transition. This roughly aligns with the experiences of Sandholm, Gilpin, and Conitzer (2005), who found that relaxing the feasible region while moving some constraints into the objective did not result in an overall speedup.

Discussion & Future Research

In this paper, we theoretically and empirically investigated the existence of envy-free allocations of indivisible goods. Under additive valuations and general assumptions on the distributions over values of individual goods, we theoretically characterized the conditions for nonexistence and existence of envy-free allocations. We supported these asymptotic results with experiments on two value distributions using two MIP models and found, empirically, that the theoretical conditions for (non)existence of envy-free allocations apply even when the number of agents and goods is quite small. Furthermore, we discovered that the hardest computational problems in this space on average exist during the phase transition between nonexistence and existence.

In typical phase transition work, what is increased on the "x-axis" is the number of constraints while keeping the number of variables constant. Our phase transition is, in that sense, different because as we increase the number of goods (while keeping the number of agents fixed), both the number of variables and constraints increases. Our phase transition is nevertheless similar to prior ones in that (i) there is a sharp transition from infeasibility to feasibility, (ii) the complexity peak occurs at that transition, (iii) the complexity peak is driven mainly by infeasible instances, and (iv) the infeasible instances get harder—and rarer—as we move to the side of the phase transition where instances are typically feasible.

While the theoretical results we presented are essentially tight, it would be useful to completely characterize the phase transition between nonexistence and existence of an envy-free allocation. We showed experimentally that this phase transition is quite sharp, but either proving that the logarithmic factor in Theorem 2 is necessary or further whittling down this bound toward Theorem 1 would be helpful. Results of this nature are actively being pursued with random 3-SAT problems (Kaporis, Kirousis, and Lalas 2006; Maneva and Sinclair 2008). Furthermore, relaxing the distributional assumptions (especially on Theorem 1) would, if possible, be useful toward this end.

Along the lines of enhanced MIP techniques, it would be interesting to try to "flatten the hardness bump" we saw in the experiments through the use of custom branching and fathoming rules, variable prioritization schemes, and other heuristics that maintain search completeness.

Acknowledgments

This material was funded by NSF grants IIS-1320620, CCF-1101668, CCF-1215883, and IIS-0964579, by an NDSEG fellowship, and used the Pittsburgh Supercomputing Center in partnership with the XSEDE, which is supported by NSF grant OCI-1053575. We thank Intel Corporation for machine gifts.

References

Boutilier, C.; Caragiannis, I.; Haber, S.; Lu, T.; Procaccia, A. D.; and Sheffet, O. 2012. Optimal social choice functions: A utilitarian view. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC)*, 197–214.

Bouveret, S., and Lang, J. 2008. Efficiency and envy-freeness in fair division of indivisible goods: logical representation and complexity. *Journal of Artificial Intelligence Research* 32:525–564.

Brams, S. J., and Fishburn, P. C. 2000. Fair division of indivisible items between two people with identical preferences: Envy-freeness, Pareto-optimality, and equity. *Social Choice and Welfare* 17:247–267.

Brams, S. J.; Kilgour, M.; and Klamler, C. 2014. Two-person fair division of indivisible items: An efficient, envy-free algorithm. *Notices of the AMS* 61(2):130–141.

Cheeseman, P.; Kanefsky, B.; and Taylor, W. 1991. Where the really hard problems are. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJ-CAI)*, 331–337.

Chen, Y.; Lai, J. K.; Parkes, D. C.; and Procaccia, A. D. 2010. Truth, justice, and cake cutting. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, 756–761.

Chevaleyre, Y.; Dunne, P. E.; Endriss, U.; Lang, J.; Lemaître, M.; Maudet, N.; Padget, J.; Phelps, S.; Rodríguez-Aguilar, J. A.; and Sousa, P. 2006. Issues in multiagent resource allocation. *Informatica* 30:3–31.

Chevaleyre, Y.; Endriss, U.; and Maudet, N. 2007. Allocating goods on a graph to eliminate envy. In *Proceedings of the 22nd AAAI Conference on Artificial Intelligence (AAAI)*, 700–705.

Cohler, Y. J.; Lai, J. K.; Parkes, D. C.; and Procaccia, A. D. 2011. Optimal envy-free cake cutting. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI)*, 626–631.

Conitzer, V., and Sandholm, T. 2006. Nonexistence of voting rules that are usually hard to manipulate. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*.

Conitzer, V.; Sandholm, T.; and Lang, J. 2007. When are elections with few candidates hard to manipulate? *Journal of the ACM* 54(3):1–33.

Hogg, T.; Huberman, B. A.; and Williams, C. P. 1996. Phase transitions and the search problem. *Artificial Intelligence* 81(1–2):1–15.

Kaporis, A. C.; Kirousis, L. M.; and Lalas, E. G. 2006. The probabilistic analysis of a greedy satisfiability algorithm. *Random Structures & Algorithms* 28(4):444–480.

Lipton, R.; Markakis, E.; Mossel, E.; and Saberi, A. 2004. On approximately fair allocations of indivisible goods. In *ACM Conference on Electronic Commerce (EC)*, 125–131.

Maneva, E., and Sinclair, A. 2008. On the satisfiability threshold and clustering of solutions of random 3-sat formulas. *Theoretical Computer Science* 407(1):359–369.

Mossel, E.; Procaccia, A. D.; and Rácz, M. Z. 2013. A smooth transition from powerlessness to absolute power. *Journal of Artificial Intelligence Research* 48:923–951.

Procaccia, A. D., and Rosenschein, J. S. 2007. Average-case tractability of manipulation in elections via the fraction of manipulators. In *Proceedings of the 6th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 718–720.

Procaccia, A. D. 2013. Cake cutting: Not just child's play. *Communications of the ACM* 56(7):78–87.

Sandholm, T.; Gilpin, A.; and Conitzer, V. 2005. Mixed-integer programming methods for finding Nash equilibria. In *Proceedings of the 20th AAAI Conference on Artificial Intelligence (AAAI)*, 495–501.

Walsh, T. 2011. Where are the hard manipulation problems? *Journal of Artificial Intelligence Research* 42:1–29.

Xia, L., and Conitzer, V. 2008. Generalized scoring rules and the frequency of coalitional manipulability. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC)*, 109–118.