Selfish Knapsack

Itai Feigenbaum and **Matthew P. Johnson**
Lehman College and the Graduate Center, City University of New York

Abstract

We consider a strategic variant of the knapsack problem: the items are owned by agents, and agents can misrepresent their sets of items—either by hiding items (understating), or by reporting fake ones (overstating). Each agent's utility equals the total value of her items included in the knapsack. We wish to maximize social welfare, and attempt to design mechanisms that lead to small worst-case approximation ratios at equilibrium. We provide a randomized mechanism with attractive strategic properties: it has a price of anarchy of 2 for Bayes-Nash and coarse correlated equilibria. For overstating-only agents, it becomes strategyproof, and has a matching lower bound. For the case of two understating-only agents, we provide a specialized randomized strategyproof $\frac{5+4\sqrt{2}}{7} \approx 1.522$-approximate mechanism, and a lower bound of $\frac{5\sqrt{5}-9}{2} \approx 1.09$. When all agents but one are honest, we provide a deterministic strategyproof $\frac{1+\sqrt{5}}{2} \approx 1.618$-approximate mechanism with a matching lower bound. The latter two mechanisms are also useful in problems beyond the one in consideration.

1 Introduction

We study a strategic variant of the knapsack problem, in which there are n agents, each owning a set of items, where each item has a value and size. A social planner must design a mechanism to choose which items to include in a knapsack of a certain capacity, where the total size of the chosen items cannot exceed the capacity. Each agent gets a utility equal to the total value of her own items included in the knapsack, while the designer wishes to maximize social welfare (the sum of the utilities of the agents, which amounts to the total value of the items in the knapsack). However, the set of items each agent owns is private information, and an agent may choose not to disclose all of her items (may report any subset of them). We call this the *understating* model (UM). Revealing all items might not be in an agent's best interest:

Example 1. *Assume the knapsack's capacity is* 1. *Consider a mechanism which always chooses an optimal (social welfare maximizing) solution based on the reported items. Assume agent* 1*'s true item set is* $\{a, b\}$ *and agent* 2*'s true item set is* $\{c\}$*, where* a*,* b *and* c *have values* 1*,* $\frac{3}{4}$ *and* $\frac{3}{4}$ *and sizes* 1*,* $\frac{1}{2}$ *and* $\frac{1}{2}$ *respectively. If the agents report truthfully, the mechanism chooses* $\{b, c\}$ *as the solution; however, if agent 1 hides item* b *and reports her set of items to be* $\{a\}$*, the chosen solution becomes* $\{a\}$*, increasing agent* 1*'s utility from* $\frac{3}{4}$ *to* 1 *while decreasing social welfare from* $\frac{3}{2}$ *to* 1*.*

To incentivize truthful reporting, we look for *strategyproof* (SP) mechanisms, where truth-telling is a dominant strategy equilibrium (no agent can benefit from misreporting). As Example 1 suggests, such mechanisms cannot always achieve optimality, so we seek mechanisms that approximate optimality well. Specifically, we try to design SP mechanisms with small worst-case approximation ratios. A mechanism is α-approximate if it provides, on every instance, social welfare value of at least $\frac{1}{\alpha}$ times the optimal welfare, as long as the agents are truthful; the worst-case approximation ratio of a mechanism is the smallest such α.[1] While we aspire to design SP mechanisms, we note that one of our mechanisms actually fails to be SP, giving agents an incentive to misreport. Nevertheless, as long as agents' (usually non-truthful) reports follow a Bayes-Nash equilibrium (BNE) or a coarse correlated equilibrium (CCE), our mechanism still generates a high fraction of the optimal welfare on every instance.

We emphasize that agents can misreport the *existence* of items, but not their *properties*—their size and value; that is, the planner has the power to verify the size and value of the reported items. One example of such a scenario is the allocation of a scientific resource, like time on a particle accelerator or NSF funding. Scientists submit research proposals, each requesting a certain amount of resource, which would provide a certain expected scientific value. This expected scientific value is evaluated/confirmed by an impartial expert. Since scientists can avoid submitting some of their proposals, the problem of choosing which proposals to accept in order to maximize total expected scientific value falls within our model.[2] Another example is the allocation of advertising time on a video-sharing website, using a sim-

[1] The optimal welfare on an instance is the value of an optimal solution of the knapsack problem defined by the agents' true sets of items.

[2] We assume that scientists are not allowed to make partial use of the resource allocated to an accepted proposal: they cannot use a fraction of the resource and produce a corresponding fraction of the value.

ple pay-per-sale model, where the website charges a (fixed and advertiser-independent) constant fraction of the profit generated directly by the displayed commercial. We assume that there is some agreed upon formula for estimating the expected profit generated by a commercial. In this example, the capacity is the (estimated) time a user is willing to watch commercials, the agents are advertisers, and the items are commercials, with size equal to the duration of the commercial and value being the expected profit it generates. As advertisers are free to avoid providing some of their commercials to the website, the problem of profit-maximization by the website falls within our model.

While our primary focus is on UM, we also consider the *overstating* model (OM) where an agent is allowed to report fake items, which she does not actually own (report supersets of her true item set); the planner is assumed to be unable to differentiate between real and fake items.[3] Despite the fact that the agent derives no value from the inclusion of fake items in the knapsack, she can use them to indirectly increase her utility:

Example 2. *Consider Example 1, only now agent 1's true item set is $\{a\}$ and agent 2's true item set is $\{c\}$. If the agents report truthfully, the mechanism chooses $\{a\}$ as the solution. However, if agent 2 reports $\{c, d\}$, where d is a fake item of value $\frac{3}{4}$ and size $\frac{1}{2}$, the chosen solution becomes $\{c, d\}$. Agent 2 does not derive any benefit from the inclusion of d in the knapsack, but she does benefit from the inclusion of c; thus this manipulation increases her utility from 0 to $\frac{3}{4}$, while decreasing social welfare from 1 to $\frac{3}{4}$.*

Note that since agents only derive utility from real items, social welfare (the designer's objective) amounts to the total value the *real* items in the knapsack. In the allocation of scientific resources, fake items are proposals that the scientist has no intention to seriously pursue. Finally, in addition to UM and OM, we also consider their joint generalization, the *full* model (FM), where agents can simultaneously hide items and report fake ones.

Our paper is part of a growing literature on the subject of approximate mechanism design without money (Procaccia and Tennenholtz 2013) (as well as the price of anarchy (Roughgarden 2015b)). This approach has been applied to many types of problems, such as matching (Dughmi and Ghosh 2010), facility location (Alon et al. 2010; Feldman and Wilf 2013; Feigenbaum, Sethuraman, and Ye 2013), and kidney exchange (Ashlagi et al. 2013). The most relevant paper we could find is (Chen, Gravin, and Lu 2011), which (among other results) provides a randomized SP mechanism for UM with a large constant approximation ratio; there is no overlap between our results and theirs. Also related is the "Funding Games" model of (Bar-Noy et al. 2012), where agents wish to maximize the size of their chosen items. In addition, there is other work that considers manipulation involving the existence of objects rather than their properties. In the context of exchange markets, such manipulation is considered in (Atlamaz and Klaus 2007; Postlewaite 1979); in connection with approximation, similar manipulation is considered in (Ashlagi et al. 2013; Dughmi and Ghosh 2010; Chen, Gravin, and Lu 2011). OM bears similarity to the notion of "slot destruction" in (Schummer and Vohra 2013), where airlines withhold information regarding cancellation of flights (equivalent to reporting fake flights) in order to manipulate a mechanism assigning landing times. Finally, examples of price of anarchy analysis for non-SP mechanisms can be found in (Caragiannis et al. 2015; Bhawalkar and Roughgarden 2011).

We note that UM and the problems in (Ashlagi et al. 2013; Dughmi and Ghosh 2010; Chen, Gravin, and Lu 2011) can be viewed as private cases of a general class of problems, which we shall call the hiding class. Let F be a set of feasible solutions. Each agent derives some utility from every solution in F, and has the ability to hide some subsets of solutions in F from the designer, who wishes to maximize social welfare.[4] Some of the mechanisms we design are able to tackle problems other than knapsack in this class.

Our Contributions. Our main contribution is a systematic analysis of a randomized mechanism we call HALF-GREEDY, which we show enjoys strong strategic properties:

- In the overstating model: It is strategyproof and provides a 2-approximation, the best achievable approximation ratio under strategyproofness.
- In the understating and full models: Although not strategyproof, every *Bayes-Nash equilibrium* and *coarse correlated equilibrium* induced is 2-approximate (under a mild assumption in the case of the full model).

We also study two specialized settings in the understating model:

1. Duopoly ($n = 2$ agents): we design a randomized strategyproof $\frac{5+4\sqrt{2}}{7} \approx 1.522$-approximate mechanism, and provide a lower bound of $\frac{5\sqrt{5}-9}{2} \approx 1.09$ on achievable approximation ratios under strategyproofness.
2. One-bad-apple, with only one manipulative agent among an otherwise honest population: we design a deterministic strategyproof $\frac{1+\sqrt{5}}{2} \approx 1.618$-approximate mechanism, and a matching lower bound.

These last two mechanisms can be applied to other problems in the hiding class as well. Due to space constraints, all proofs, and additional results, appear in the online appendix at http://www.itaifeigenbaum.com/research/selfish-knapsack-aaai17-appendix.pdf.

2 Model

The knapsack's capacity is (w.l.o.g.) 1. $N = \{1, 2, \ldots, n\}$ denotes the set of agents, $n \geq 2$. We define G to be the universe of all possible items: every item $a \in G$ has size $s(a) \in (0, 1]$ and value $v(a) \in (0, \infty)$, and we assume that for every $s \in (0, 1]$, $v \in (0, \infty)$, G contains infinitely many

[3] The distinction we make between manipulating properties and existence of real items is meaningless for fake items, as long as agents are free to report fake items with any properties in any amount they wish.

[4] For example, in UM, if agent i owns items a and b, she may hide subsets of the form "all solutions including a", "all solutions including b" or "all solutions including a or b".

items of size s and value v.[5] We denote $\mathcal{X}$ to be the collection of all finite subsets of G. For $A \in \mathcal{X}$, we define $s(A) = \sum_{a \in A} s(a)$ and $v(A) = \sum_{a \in A} v(a)$.

Each agent i owns a finite set of items $X_i \in \mathcal{X}$; $\mathbf{X}$ denotes the profile $(X_1, \dots, X_n)$. The report space of each agent i with item set X_i is denoted by $\mathcal{R}(X_i) \subseteq \mathcal{X}$; $\mathcal{R}(\mathbf{X})$ denotes $(\mathcal{R}(X_1), ..., \mathcal{R}(X_n))$. In the *understating* (UM), *overstating* (OM) and *full* (FM) models, $\mathcal{R}(X_i)$ equals 2^{X_i}, $\{A \in \mathcal{X} : X_i \subseteq A\}$ and $\mathcal{X}$ respectively. Each agent i reports some $R_i \in \mathcal{R}(X_i)$, and $\mathbf{R}$ denotes the profile $(R_1, \dots, R_n)$.

We disallow all forms of joint ownership: we do not allow an item to be owned/reported by more than one agent (this also excludes an item owned by one agent and reported by another). This assumption is formalized as follows: we assume that G is partitioned into $G_1, \dots, G_n$, each G_i contains infinitely many items of every possible size-value combination, and each agent i is assumed to only own and report items from G_i. To simplify notation, we generally avoid explicitly mentioning the partition; instead, throughout this paper, any set of items that is indexed by i (such as X_i and R_i) is assumed implicitly to be a subset of G_i.

A deterministic mechanism is a function $f : \mathcal{X}^n \to \mathcal{X}$, mapping the agents' reports to a set of items to include in the knapsack; a randomized mechanism is a function from $\mathcal{X}^n$ to all random variables over $\mathcal{X}$. We restrict our attention to *feasible* mechanisms; a deterministic (resp. randomized) mechanism f is feasible iff, for all $\mathbf{R} \in \mathcal{X}^n$:

1. f only uses the reported items: $f(\mathbf{R}) \subseteq \cup_{i \in N} R_i$ (resp. surely, meaning with probability 1).
2. f doesn't violate the knapsack's capacity: $s(f(\mathbf{R})) \leq 1$ (resp. surely).

In general, we define $\mathbf{A}_{-i}$ to be the tuple $(A_1, \dots, A_{i-1}, A_{i+1}, \dots, A_n)$. We slightly abuse notation and also define $(\mathbf{A}_{-i}, B_i) = (A_1, \dots, A_{i-1}, B_i, A_{i+1}, \dots, A_n)$.[6] The utility that agent i derives from a chosen solution $S \in \mathcal{X}$ is defined as $u(X_i, S) = v(X_i \cap S)$. A mechanism is *strategyproof* (SP) if truthfulness is a dominant strategy equilibrium. For a deterministic (resp. randomized) mechanism f, this means that for all $i \in N$, $\mathbf{X} \in \mathcal{X}^n$, $R_i \in \mathcal{R}(X_i)$, we have $u(X_i, f(\mathbf{X})) \geq u(X_i, f(\mathbf{X}_{-i}, R_i))$ (resp. $\mathbb{E}[u(X_i, f(\mathbf{X}))] \geq \mathbb{E}[u(X_i, f(\mathbf{X}_{-i}, R_i))]$). We emphasize that for randomized mechanisms, the requirement is that truthful reporting maximizes an agent's utility *in expectation*.

Informally speaking, the planner wants to choose a solution S which maximizes social welfare: $\sum_i u(X_i, S)$. However, as we saw in Examples 1 and 2, SP mechanisms cannot always choose the optimal solution. Thus, we settle for an approximation to optimality: we attempt to design SP mechanisms with small *worst-case approximation ratios*. The worst-case approximation ratio of a deterministic (resp. randomized) mechanism f is $\max_{\mathbf{X} \in \mathcal{X}^n} \frac{\sum_{i=1}^n u(X_i, OPT(\cup_{i \in N} X_i))}{\sum_{i=1}^n u(X_i, f(\mathbf{X}))}$ (resp. $\max_{\mathbf{X} \in \mathcal{X}^n} \frac{\sum_{i=1}^n u(X_i, OPT(\cup_{i \in N} X_i))}{\sum_{i=1}^n \mathbb{E}[u(X_i, f(\mathbf{X}))]}$), where $OPT(A)$ is an optimal solution to the knapsack problem when the set of available items is A.[7]

Finally, one of our mechanisms fails to be SP in UM and FM, but still has attractive strategic properties in terms of Bayes-Nash equilibrium (BNE), suitable when agents have distributional knowledge of each other's items, and coarse correlated equilibrium (CCE), suitable when that knowledge is exact. We briefly remind the reader of the relevant definitions; for a complete discussion, see (Roughgarden 2015a; 2015b).

1. Let $\dot{\mathbf{X}}$ be a random variable over $\mathcal{X}^n$ with distribution $\mathcal{F}$. A strategy S_i is a function mapping $X_i \in \mathcal{X}$ to a random variable over $\mathcal{R}(X_i)$; we also define $\dot{S}_i = S_i(\dot{X}_i)$. A strategy profile $\mathbf{S}$ is a BNE w.r.t. mechanism f and distribution $\mathcal{F}$ iff, for every $i \in N$ and strategy S'_i, $\mathbb{E}[u(\dot{X}_i, f(\dot{\mathbf{S}}))] \geq \mathbb{E}[u(\dot{X}_i, f(\dot{\mathbf{S}}_{-i}, \dot{S}'_i))]$. $\mathbf{S}$ is α-approximate iff $\frac{\sum_{i=1}^n \mathbb{E}[u(\dot{X}_i, OPT(\cup_{i \in N} \dot{X}_i))]}{\sum_{i=1}^n \mathbb{E}[u(\dot{X}_i, f(\dot{\mathbf{S}}))]} \leq \alpha$.
2. For a given $\mathbf{X} \in \mathcal{X}^n$, a random variable $\dot{\mathbf{R}}$ over $\mathcal{R}(\mathbf{X})$ is a CCE under mechanism f if for every $i \in N$, $R'_i \in \mathcal{R}(X_i)$, we have $\mathbb{E}[u(X_i, f(\dot{\mathbf{R}})] \geq \mathbb{E}[u(X_i, f(\dot{\mathbf{R}}_{-i}, R'_i)]$. $\dot{\mathbf{R}}$ is α-approximate iff $\frac{\sum_{i=1}^n u(X_i, OPT(\cup_{i \in N} X_i))}{\sum_{i=1}^n \mathbb{E}[u(X_i, f(\dot{\mathbf{R}}))]} \leq \alpha$.

3 The Half-Greedy Mechanism

In this section, we analyze the strategic properties of a randomized mechanism we call HALF-GREEDY. In OM, we show that HALF-GREEDY is SP and 2-approximate. We also show that no randomized SP mechanism can beat this approximation guarantee, and no deterministic SP mechanism can provide a constant worst-case approximation ratio. In UM, we show that while HALF-GREEDY is not SP, every BNE and CCE it induces is 2-approximate; in FM, we can preserve this result, under a mild additional assumption. We note that a slightly modified version of HALF-GREEDY is useful in a different model, where each agent owns a single item and can misreport its properties (size and value) in a restricted fashion; see appendix for details.

To define HALF-GREEDY, we need two auxiliary mechanisms. The first is GREEDY (GR), which adds items to the knapsack by decreasing value-to-size ratio, breaking ties arbitrarily but *consistently*. The consistent tie-breaking is best represented by assuming the existence of a total order $\succeq$ on G. We may assume w.l.o.g. that $\succeq$ satisfies that, for all $a, b \in G$, if $\frac{v(a)}{s(a)} > \frac{v(b)}{s(b)}$, then $a \succ b$; this can be effectively

[5] The fact that our formulation allows us to distinguish between items with identical size, value and owner is a mere convenience. All of our results translate to a model where such items are indistinguishable. In addition, the assumption that G contains infinitely many items of each size-value pair is needed for our lower bounds, but not needed for any of our positive results.

[6] The notational abuse stems from the fact that, according to our definition of $\mathbf{A}_{-i}$, it should be the case that $(\mathbf{A}_{-i}, B_i) = ((A_1, \dots, A_{i-1}, A_{i+1}, \dots, A_n), B_i)$, that is a 2-tuple where the first element is in $\mathcal{X}^{n-1}$. However, throughout this paper we are never interested in such tuples, so no confusion arises.

[7] Note that the only source of randomization is f, so there is no need to take expectation of the numerator in the randomized case.

ALGORITHM 1: GREEDY

```
Input: Sets of reported items $\mathbf{R} \in \mathcal{X}^n$
$S \leftarrow \emptyset, T \leftarrow \cup_{i \in N} R_i$
while $T \neq \emptyset$ do
    $next \leftarrow \max_{\succeq} T$
    $T \leftarrow T \backslash \{next\}$
    if $s(S \cup \{next\}) \leq 1$ then
        $S \leftarrow S \cup \{next\}$
    else
        break
return $S$
```

ALGORITHM 2: MAXIMUM-VALUE

```
Input: Sets of reported items $\mathbf{R} \in \mathcal{X}^n$
if $\cup_{i \in N} R_i = \emptyset$ then
    return $\emptyset$
return
  $\max_{\succeq} \{a \in \cup_{i \in N} R_i : v(a) \geq v(b) \;\; \forall b \in \cup_{i \in N} R_i\}$
```

imposed by sorting the (finite) set of items reported to the mechanism. GREEDY is defined as Algorithm 1.

The second auxiliary mechanism is MAXIMUM-VALUE (MV, see Algorithm 2), which returns a single item with the maximum value possible, breaking ties according to $\succeq$.[8]

Now we can define HALF-GREEDY, which is well known to be 2-approximate (via a simple adaptation of Theorem 18.5 in (Burke and Kendall 2005)):[9]

Definition 1. *The HALF-GREEDY mechanism HG runs GR with probability $\frac{1}{2}$ and MV with probability $\frac{1}{2}$ (probabilities chosen independently of the input).*

Overstating Model

HALF-GREEDY's strategyproofness in OM follows from a very simple fact: under both GR and MV, every real item that is included in the knapsack when agent i reports R_i, remains in the knapsack when agent i reports $R_i \cap X_i$, namely avoids reporting the fake items within R_i:

Lemma 1. *Let $i \in N$, $\mathbf{X} \in \mathcal{X}^n$ and $R_i \in \mathcal{X}$. Then $X_i \cap GR(\mathbf{X}_{-i}, R_i) \subseteq GR(\mathbf{X}_{-i}, R_i \cap X_i)$ and $X_i \cap MV(\mathbf{X}_{-i}, R_i) \subseteq MV(\mathbf{X}_{-i}, R_i \cap X_i)$.*

In other words, an agent never loses from restricting her report to the real items within that report. This immediately implies SP in OM, since there avoiding reporting fake items amounts to truthfulness.

Corollary 1. *In OM, HALF-GREEDY is strategyproof and 2-approximate.*

It is important to note that once GREEDY first fails to add an item to the knapsack (namely, it attempts to pick up an item that does not fit in the remaining space), it stops and returns the items currently in the knapsack; it does *not* try to add the next item that fits in the remaining space. This seemingly trivial choice is actually crucial for maintaining SP, as the following example shows.

Example 3. *Consider the BAD-GREEDY mechanism BG, obtained from GREEDY by removing the "else break" statement, applied to this $n = 2$ case: $X_1 = \{a\}$, $X_2 = \{b\}$, $v(a) = s(a) = 1$, $v(b) = \frac{1}{4}$, $s(b) = \frac{1}{2}$. Here, $BG(\mathbf{X}) = \{a\}$ and the utility of agent 2 is $v(X_2 \cap \{a\}) = 0$. However, if agent 2 reports $R_2 = \{b, c\}$, where $v(c) = 1$, $s(c) = \frac{1}{2}$ (that is, agent 2 reports a fake item c in addition to her true item b), then $BG(X_1, R_2) = \{b, c\}$, and agent 2's utility is $v(X_2 \cap \{b, c\}) = \frac{1}{4}$. Thus, BAD-GREEDY is not SP in OM.*

We also provide matching lower bounds, which complete the picture for OM. They show that HALF-GREEDY is best possible in OM among randomized SP mechanisms, and that randomization is necessary for good approximation.

Theorem 1. *In OM, there is no randomized SP mechanism with a worst-case approximation ratio strictly smaller than 2. Also, there is no deterministic SP mechanism with a constant worst-case approximation ratio.*

Understating and Full Models

In UM, HALF-GREEDY is no longer SP—it is sometimes beneficial for an agent to hide items.

Example 4. *Consider this $n = 2$ case: $X_1 = \{a, b\}$, $X_2 = \{c, d\}$, $v(a) = 2$, $v(c) = 2 - \epsilon$, $s(a) = s(c) = \frac{1}{4} + \epsilon$, $v(b) = 3 + \epsilon$, $v(d) = 3$, $s(b) = s(d) = \frac{1}{2}$, and $\epsilon > 0$ is very small. It is easy to check that there are no dominant strategies in this case. For agent 1: $\emptyset$ is strictly dominated by $\{a\}$, which is worse than $\{b\}$, which is worse than $\{a, b\}$ when agent 2 reports $\emptyset$, and $\{a, b\}$ is worse than $\{b\}$ when agent 2 reports X_2. For agent 2: $\emptyset$ is strictly dominated by $\{c\}$, which is a worse response than $\{d\}$, which is worse than $\{c, d\}$ when agent 1 reports $\emptyset$, and $\{c, d\}$ is a worse response than $\{d\}$ when agent 1 reports $\{b\}$.*

However, that is not necessarily bad news. Both GR and MV satisfy the following property: fix some agent $i \in N$. For all *other* agents $j \neq i$, every item of agent j which is included in the knapsack when agent i reports R_i, remains in the knapsack when agent i reports a subset of R_i.

Lemma 2. *For an agent i, let $R_i, R'_i \in \mathcal{X}$, $R'_i \subseteq R_i$, and $\mathbf{R}_{-i}, \mathbf{X}_{-i} \in \mathcal{X}^{n-1}$. Then, for every $j \neq i$, $X_j \cap GR(\mathbf{R}) \subseteq GR(\mathbf{R}_{-i}, R'_i)$ and $X_j \cap MV(\mathbf{R}) \subseteq MV(\mathbf{R}_{-i}, R'_i)$.*

Thus, when an agent hides items, all *other* agents weakly benefit. Therefore, when an agent benefits from hiding items, social welfare increases, since then all other agents benefit as well. Following this observation, it is intuitive (but not trivial) to expect that the agents' manipulations would not hurt social welfare. To model how agents hide items, we consider two options regarding the knowledge agents have

[8] Any total order on G can be used for consistent tie-breaking in MV; it need not be the same as the one used in GR.

[9] HALF-GREEDY might leave much of the knapsack unused. This is needed due to the potential existence of items with very large sizes. Consider the case where such items do not exist, meaning that for some $x << 1$, $s(a) < x$ for all $a \in G$. In that case, our analysis gives that replacing HALF-GREEDY with GREEDY preserves our positive results (Corollary 1 and Theorems 2 and 3), with an approximation ratio of $\frac{1}{1-x}$ instead of 2, while wasting at most x of the knapsack's capacity. Thus, we wrote our proofs for HALF-GREEDY due to theoretical concerns, while in practice we believe that GREEDY is an appropriate choice for many cases.

of each other's items. The first option is when that knowledge is distributional (a.k.a. Bayesian game): agents know a joint distribution from which the real items are drawn. For this option, we assume that agents behave according to a BNE. The second option is when the knowledge is exact (a.k.a complete information game): agents can see their peers' items. For this option, we merely assume that agents behave according to a CCE. Note that HALF-GREEDY is prior independent, namely we do not assume that the planner has any knowledge (apart from the reports) of the agents' items, distributional or exact. We use a smoothness-based argument (Roughgarden 2015b) to show that every BNE and CCE under HALF-GREEDY results in a weakly greater social welfare than when the agents are truthful; since truthfulness results in 2-approximation, this implies:[10]

Theorem 2. *In UM, for every prior $\mathcal{F}$ over $\mathcal{X}^n$, every BNE w.r.t. HALF-GREEDY and $\mathcal{F}$ is* 2*-approximate. Similarly, for every $\mathbf{X} \in \mathcal{X}^n$, every CCE w.r.t. HALF-GREEDY and* $\mathbf{X}$ *is* 2*-approximate.*

Next, we consider FM. In FM, Theorem 2 almost holds. The reason we say "almost" is indifference. As Lemma 1 shows, no agent can benefit from reporting fake items. However, an agent might report fake items in a way that does not change her utility, but decreases other agents' utilities. Let us give an example of such a Nash equilibrium, which is a special case of both BNE and CCE:

Example 5. *Consider this $n = 2$ case: $X_1 = \{a\}$, $X_2 = \{b\}$, where $v(a) = 1$, $s(a) = \frac{1}{M}$, $v(b) = M - 2$, $s(b) = \frac{M-1}{M}$, where M is some large integer, $M >> 2$. Truthful reporting is a Nash equilibrium. Note that when agents report truthfully, HALF-GREEDY chooses a with probability $\frac{1}{2}$ and b with probability* 1*. However, if agent 1 reports $R_1 = \{a, c\}$ where $v(c) = M - 1$ and $s(c) = \frac{M-1}{M}$, and agent* 2 *reports truthfully, we still get a Nash equilibrium, in which HALF-GREEDY still chooses a with probability $\frac{1}{2}$, but b is chosen with probability* 0 *(c is chosen with probability* 1*, but since it is a fake item, it does not add to the agents' utilities). In the latter Nash equilibrium, the approximation ratio is $2M - 2$.*

We show that problematic equilibria such as the one above cannot occur if agents are not deliberately malicious.

Definition 2. *$R_i \in \mathcal{X}$ is a* malicious report *for agent i when her true set of items is $X_i \in \mathcal{X}$, if there exists $R_i' \in \mathcal{X}$ where for all $\mathbf{X}_{-i}$, $\mathbf{R}_{-i} \in \mathcal{X}^{n-1}$, and all $j \in N$, $\mathbb{E}[u(X_j, HG(\mathbf{R}))] \leq \mathbb{E}[u(X_j, HG(\mathbf{R}_{-i}, R_i'))]$, with the inequality being strict for at least one agent in at least one instance.*

In other words, a malicious report is a report that can never benefit any agent (including the agent reporting it), and can sometimes hurt an agent. Thus, if the agents are even very mildly altruistic, they would not report maliciously. Also, in a Bayesian game, we say that a strategy S_i is malicious if $S_i(X_i)$ is malicious w.r.t. X_i, for some $X_i \in \mathcal{X}$, with positive probability.[11] Non-malicious reports satisfy an important property—fake items included in those reports have no impact on the real items included in the solution:

Lemma 3. *If $R_i \in \mathcal{X}$ is not malicious for agent i when her true set of items is $X_i \in \mathcal{X}$, then for every choice of $\mathbf{X}_{-i}$, $\mathbf{R}_{-i} \in \mathcal{X}^{n-1}$, and for every $j \in N$, we have that $X_j \cap GR(\mathbf{R}) = X_j \cap GR(\mathbf{R}_{-i}, R_i \cap X_i)$ and $X_j \cap MV(\mathbf{R}) = X_j \cap MV(\mathbf{R}_{-i}, R_i \cap X_i)$.*

Therefore, non-maliciousness rules out equilibria like the one in Example 5. Fake items might be reported at equilibria, but they would have no impact on welfare. This, in addition to Theorem 2, leads to the following result:

Theorem 3. *In FM, for every prior $\mathcal{F}$ over $\mathcal{X}^n$, every BNE w.r.t. HALF-GREEDY and $\mathcal{F}$ in which no malicious strategy is used is* 2*-approximate. Similarly, for every $\mathbf{X} \in \mathcal{X}^n$, every CCE w.r.t. HALF-GREEDY and $\mathbf{X}$ in which no malicious report is used with positive probability is* 2*-approximate.*

4 The Equal-Utility Mechanism

In this section, we consider the special case of $n = 2$ in UM. We design a specialized randomized mechanism, called EQUAL-UTILITY for this environment, which is SP and $\frac{5+4\sqrt{2}}{7} \approx 1.522$-approximate. In the next section, we show that no deterministic SP mechanism can beat EQUAL-UTILITY's approximation ratio, thus randomization leads to strict improvement. We also provide a lower bound of $\frac{5\sqrt{5}-9}{2} \approx 1.09$ on the approximation ratio attainable by randomized SP mechanisms, showing the necessity of some approximation gap. We note that our analysis of EQUAL-UTILITY holds for every problem in the hiding class for which the following assumption is satisfied: for distinct agents i and j, there exists at least one utility maximizing solution for i in F that cannot be hidden by j.

The idea behind EQUAL-UTILITY (see Algorithm 3) is to solve the knapsack problem optimally, with one additional constraint: that the agents' utilities are *exactly* equal. Since in general, apart from $\emptyset$, there might not be a deterministic solution that satisfies this additional constraint, we allow for randomized solutions instead. Formally, we want to solve the following mathematical program (PROGRAM), where A is a random decision variable (set of items):[12] maximize $\mathbb{E}[v(A)]$ subject to: (1) $A \subseteq X_1 \cup X_2$ and $s(A) \leq 1$ surely and (2) $\mathbb{E}[v(A \cap X_1)] = \mathbb{E}[v(A \cap X_2)]$. As every agent gets exactly half of the expected welfare, it is in each agent's best interest to maximize that welfare. Therefore agents have no

[10] Theorem 2 implies that the game induced by HALF-GREEDY has a price of anarchy of 2 w.r.t. these equilibria concepts.

[11] Assuming non-maliciousness is weaker and easier to justify than assuming no fake items are reported. Assuming no fake items are reported is similar in spirit to the assumption of "no overbidding" in generalized second price auctions (Caragiannis et al. 2015). Reporting fake items in our mechanism, much like overbidding in GSP, is a weakly dominated strategy; however, both can lead to unreasonable and bad equilibria due to indifference, and are thus ruled out by assumption.

[12] PROGRAM can be stated as a linear programming problem with exponentially many variables. Let $T = \{S \subseteq X_1 \cup X_2 : s(S) \leq 1\}$. Then PROGRAM can be stated as: maximize $\sum_{S \in T} v(S) p_S$ subject to $\sum_{S \in T} v(S \cap X_1) p_S = \sum_{S \in T} v(S \cap X_2) p_S$, $\sum_{S \in T} p_S = 1$ and $p_S \geq 0$ for all $S \in T$ (where the p_S's are our decision variables).

ALGORITHM 3: EQUAL-UTILITY

```
Input: Sets of items $X_1, X_2 \in \mathcal{X}$; parameter $\alpha \in [1, 2)$
$Z_1 \leftarrow OPT(X_1)$, $Z_2 \leftarrow OPT(X_2)$
if $v(Z_i) \geq (1/\alpha)(v(Z_1) + v(Z_2))$ for some $i \in \{1, 2\}$ then
    return $Z_i$ (option 1)
return optimal solution to PROGRAM with input $\mathbf{X}$ (option 2)
```

incentive to restrict the feasible region of PROGRAM by hiding items, and thus PROGRAM alone is a SP mechanism.

However, PROGRAM does not always lead to good approximation. It fails to do so on instances where one agent's items are much superior to the other's. For example, if one agent has one item of value M, the other agent has one item of value ϵ, and $M >> \epsilon$, the equal utility constraint dictates that the M-valued item is almost never chosen. Thus, we add a preliminary check meant to catch such instances. The preliminary check is as follows: say we wish for our mechanism to be α-approximate. Consider $OPT(X_1)$ and $OPT(X_2)$, namely the optimal solutions using just a single agent's items. If $OPT(X_i)$ is significantly bigger than $OPT(X_j)$, to the extent where $OPT(X_i)$ is guaranteed to be an α-approximation on its own to the optimal value, then we simply return $OPT(X_i)$. This is checked via the condition $v(OPT(X_i)) \geq \frac{1}{\alpha}(v(OPT(X_1)) + v(OPT(X_2)))$.[13] If neither agent satisfies this condition, we turn to PROGRAM.

Theorem 4. *In UM, for $\alpha \geq \frac{5+4\sqrt{2}}{7} \approx 1.522$, EQUAL-UTILITY is strategyproof and α-approximate.*

There exist instances on which the approximation ratio of EQUAL-UTILITY is arbitrarily close to $\frac{5+4\sqrt{2}}{7}$ (see appendix). We note that EQUAL-UTILITY requires solving NP-hard problems: computing $OPT(X_1)$ and $OPT(X_2)$ means solving the knapsack problem, which is known to be NP-hard. In the appendix, we show that solving PROGRAM is NP-hard as well. We refer the reader to the appendix for a brief discussion regarding managing this running-time issue. Finally, we note that no randomized SP mechanism can be arbitrarily close to optimality—some separation is required:

Theorem 5. *In UM, no randomized SP mechanism can provide a worst-case approximation ratio strictly better than $\frac{5\sqrt{5}-9}{2} \approx 1.09$ (even when $n = 2$).*

5 The Pacify-the-Liar Mechanism

We continue exploring UM. We now allow for a general number of agents n, however we restrict ourselves to an environment where there is only one bad apple—specifically, $n - 1$ agents are assumed to be honest. We assume without loss of generality that agent 1 is the manipulative agent (note that our results hold for free even if the honesty of an agent—whether or not that agent has the ability to be manipulative—is private information of that agent, since we can simply say that if all agents report to be honest, we include nothing in the knapsack).[14] For this environment, we will provide a ϕ-approximate deterministic strategyproof mechanism ($\phi = \frac{1+\sqrt{5}}{2} \approx 1.618$ is the golden ratio), along with a matching lower bound. We note that the analysis of our mechanism holds for every problem in the hiding class under the assumption that there is at least one solution, which cannot be hidden by the manipulative agent, and which maximizes social welfare for all agents but the manipulative one.

ALGORITHM 4: PACIFY-THE-LIAR

```
Input: Sets of items $\mathbf{X} \in \mathcal{X}^n$; parameter $\alpha \geq 1$
$Z_1 \leftarrow OPT(X_1)$, $Z_2 \leftarrow OPT(\cup_{i \in N \setminus \{1\}} X_i)$
if $v(Z_1) \geq (1/\alpha)(v(Z_1) + v(Z_2))$ then
    return $Z_1$ (option 1)
if $v(Z_2) \geq (1/\alpha)(OPT(\cup_{i \in N} X_i))$ then
    return $Z_2$ (option 2)
$S \leftarrow \{A \subseteq \cup_{i \in N} X_i : v(A) > \alpha v(Z_2)\}$
return $\text{argmax}_{A \in S}\, v(A \cap X_1)$ (option 3)
```

Our deterministic mechanism, called PACIFY-THE-LIAR (Algorithm 4), begins with a preliminary test which checks if agent 1 can guarantee an α-approximation on her own (option 1), or if agents 2 through n can guarantee an α-approximation together, without agent 1 (option 2). In the former case, we return $OPT(X_1)$, and in the latter case we return $OPT(\cup_{i \in N \setminus \{1\}} X_i)$. Note that, unlike in EQUAL-UTILITY, in option 2 the benchmark used is the optimal solution $v(OPT(\cup_{i \in N} X_i))$ rather than the upper bound $v(Z_1) + v(Z_2)$. This is crucial for maintaining a ϕ approximation ratio, and does not violate SP due to the honesty of all agents other than 1. If the preliminary test fails, we move to option 3, where we attempt to "pacify" agent 1 by choosing her favorite solution among a collection of solutions that guarantee α-approximation.

Note that if we reach option 3, S is nonempty since we did not stop at option 2 (thus S includes the optimal solution). By hiding items, agent 1 can only make S smaller; however, since the mechanism returns agent 1's favorite solution in S, she has no incentive to make S smaller, and thus no incentive to misreport. Furthermore, since we did not stop at option 1, $\frac{1}{\alpha-1} v(Z_2) > v(Z_1)$, hence $\alpha v(Z_2) > (\alpha - 1)(v(Z_1) + v(Z_2)) \geq (\alpha - 1) v(OPT(\cup_{i \in N} X_i))$. Thus, every solution in S leads to α-approximation as long as $\alpha - 1 \geq \frac{1}{\alpha}$, which is satisfied as long as $\alpha \geq \phi$.

Theorem 6. *In UM, PACIFY-THE-LIAR is SP and α-approximate for $\alpha \geq \phi$.*

Finally, there is no better deterministic SP mechanism:

Theorem 7. *In UM, no deterministic SP mechanism can provide a worst-case approximation ratio strictly better than ϕ (even when $n = 2$).*

[13] We use $v(OPT(X_1)) + v(OPT(X_2))$ instead of $v(OPT(X_1 \cup X_2))$ to maintain strategyproofness.

[14] If we naturally extend our definition of mechanism to allow reporting of all private data, including honesty. This observation relies on the honest agents reporting their honesty correctly.

6 Future Research

There are several natural directions for the continuation of our research. First, all of our lower bounds hold even when there are only two agents and only one is manipulative. It will be interesting to know if having more manipulative agents necessarily increases the attainable worst-case approximation ratio under strategyproofness. Second, we did not provide a strategyproof mechanism for the full model, and whether one with a constant worst-case approximation ratio exists is an open problem. Third, we did not provide a strategyproof mechanism for the general understating model, only for special cases of it; a randomized strategyproof mechanism with a very large worst-case approximation ratio is given in (Chen, Gravin, and Lu 2011), but it is unclear whether a smaller worst-case approximation ratio is attainable. Finally, it will be interesting to design more general techniques for handling problems in the hiding class.

Acknowledgements

This work was supported in part by CUNY Collaborative Incentive Research Grant (CIRG 21) 2153 and PSC-CUNY Research Award 67665-00 45. The authors also thank Jay Sethuraman and Garud Iyengar for helpful comments, as well as the anonymous reviewers.

References

Alon, N.; Feldman, M.; Procaccia, A. D.; and Tennenholtz, M. 2010. Strategyproof approximation of the minimax on networks. *Math. Oper. Res.* 35(3):513–526.

Ashlagi, I.; Fischer, F.; Kash, I. A.; and Procaccia, A. D. 2013. Mix and match: A strategyproof mechanism for multi-hospital kidney exchange. *Games and Economic Behavior*.

Atlamaz, M., and Klaus, B. 2007. Manipulation via endowments in exchange markets with indivisible goods. *Social Choice and Welfare* 28(1):1–18.

Bar-Noy, A.; Gai, Y.; Johnson, M. P.; Krishnamachari, B.; and Rabanca, G. 2012. Funding games: the truth but not the whole truth. In *Internet and Network Economics*, 128–141. Springer.

Bhawalkar, K., and Roughgarden, T. 2011. Welfare guarantees for combinatorial auctions with item bidding. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, 700–709. SIAM.

Burke, E. K., and Kendall, G. 2005. *Search methodologies*. Springer.

Caragiannis, I.; Kaklamanis, C.; Kanellopoulos, P.; Kyropoulou, M.; Lucier, B.; Leme, R. P.; and Tardos, É. 2015. Bounding the inefficiency of outcomes in generalized second price auctions. *Journal of Economic Theory* 156:343–388.

Chen, N.; Gravin, N.; and Lu, P. 2011. Mechanism design without money via stable matching. *CoRR* abs/1104.2872.

Dughmi, S., and Ghosh, A. 2010. Truthful assignment without money. In *11th ACM conference on Electronic Commerce*, 325–334. ACM.

Feigenbaum, I.; Sethuraman, J.; and Ye, C. 2013. Approximately optimal mechanisms for strategyproof facility location: Minimizing l_p norm of costs. *CoRR* abs/1305.2446.

Feldman, M., and Wilf, Y. 2013. Strategyproof facility location and the least squares objective. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, 873–890. ACM.

Postlewaite, A. 1979. Manipulation via endowments. *The review of economic studies* 46(2):255–262.

Procaccia, A. D., and Tennenholtz, M. 2013. Approximate mechanism design without money. *ACM Trans. Economics and Comput.* 1(4):18.

Roughgarden, T. 2015a. Intrinsic robustness of the price of anarchy. *Journal of the ACM (JACM)* 62(5):32.

Roughgarden, T. 2015b. The price of anarchy in games of incomplete information. *ACM Transactions on Economics and Computation* 3(1):6.

Schummer, J., and Vohra, R. V. 2013. Assignment of arrival slots. *American Economic Journal: Microeconomics* 5(2):164–185.