Allocating Indivisible Goods to Strategic Agents: Pure Nash Equilibria and Fairness*

Georgios Amanatidis[1], Georgios Birmpas[2], Federico Fusco[2], Philip Lazos[2,3], Stefano Leonardi[2], and Rebecca Reiffenhäuser[2]

[1]Department of Mathematical Sciences; University of Essex.
georgios.amanatidis@essex.ac.uk

[2]Department of Computer, Control, and Management Engineering; Sapienza University of Rome.
{birbas, fuscof, lazos, leonardi, rebeccar}@diag.uniroma1.it

[3]IOHK
philip.lazos@iohk.io

Abstract

We consider the problem of fairly allocating a set of indivisible goods to a set of *strategic* agents with additive valuation functions. We assume no monetary transfers and, therefore, a *mechanism* in our setting is an algorithm that takes as input the reported—rather than the true—values of the agents. Our main goal is to explore whether there exist mechanisms that have pure Nash equilibria for every instance and, at the same time, provide fairness guarantees for the allocations that correspond to these equilibria. We focus on two relaxations of envy-freeness, namely *envy-freeness up to one good* (EF1), and *envy-freeness up to any good* (EFX), and we positively answer the above question. In particular, we study two algorithms that are known to produce such allocations in the non-strategic setting: Round-Robin (EF1 allocations for any number of agents) and a cut-and-choose algorithm of Plaut and Roughgarden [42] (EFX allocations for two agents). For Round-Robin we show that all of its pure Nash equilibria induce allocations that are EF1 with respect to the underlying true values, while for the algorithm of Plaut and Roughgarden we show that the corresponding allocations not only are EFX but also satisfy *maximin share fairness*, something that is not true for this algorithm in the non-strategic setting! Further, we show that a weaker version of the latter result holds for any mechanism for two agents that always has pure Nash equilibria which all induce EFX allocations.

1 Introduction

Fair division refers to the problem of distributing a set of resources among a set of agents in such a way that everyone is "happy" with the overall allocation. Capturing this "happiness" can be elusive, as it may be determined by complicated underlying social dynamics; however, two well-motivated (and mathematically conducive) interpretations are those of *envy-freeness* [28, 27, 45] and *proportionality* [44]. When an allocation is envy-free, each agent values the set of resources that she receives at least as much as the set of any other agent, while when an allocation is proportional, each agent receives at least $1/n$ of her total

* This work was supported by the ERC Advanced Grant 788893 AMDROMA "Algorithmic and Mechanism Design Research in Online Markets", the MIUR PRIN project ALGADIMAR "Algorithms, Games, and Digital Markets", and the NWO Veni project No. VI.Veni.192.153.

value for all the goods, assuming there are n agents. Since the first mathematically formal treatment of fair division by Banach, Knaster and Steinhaus [44], the multifaceted questions that arise for the different variants of the problem have been studied in a diverse group of fields, including mathematics, economics, and political science. As many of these questions are inherently algorithmic, fair division questions, especially the ones related to the existence, computation, and approximation of different fairness notions have been very actively studied by computer scientists during the last two decades (see, e.g., [43, 14, 39] for surveys of recent results).

In the standard discrete fair division setting that we study here, the resources are indivisible goods and the agents have additive valuation functions over them. Typically, there is also the additional assumption that all the goods need to be allocated. This discrete setting poses a significant conceptual challenge, as the classic notions of fairness originally introduced for *divisible* goods, such as envy-freeness and proportionality, are impossible to satisfy. The example that illustrates this situation needs only two agents and just one positively valued good. Whoever does not receive the good will not consider the result to be either envy-free or proportional. However, this should not necessarily be considered an *unfair* outcome, as it is done out of necessity, not malice: the only other (deterministic) option would be to deprive both agents of the good, which seems wasteful. To define what is *fair* in this context, a number of weaker fairness notions have been proposed. Among the most prevalent of those are *envy-freeness up to one good* (EF1), *envy-freeness up to any good* (EFX), and *maximin share fairness* (MMS). The notions of EF1 and EFX were introduced by Lipton et al. [37], Budish [17], and Gourvès et al. [33], Caragiannis et al. [20] respectively, and they can be seen as additive relaxations of envy-freeness. Both of them are based on the following rationale: an agent may envy another agent but only by the value of the most (for EF1) or the least (for EFX) desirable good in the other agent's bundle. It is straightforward that EF1 is weaker than EFX, and indeed this is reflected to the known results for the two notions. The concept of the *maximin share* of an agent was introduced by Budish [17] as a relaxation to the proportionality benchmark. The corresponding fairness notion, *maximin share fairness*, requires that each agent receives the maximum value that this agent would obtain if she was allowed to partition the goods into n bundles and then take the worst of them (see Section 2 for a more detailed description and a formal definition).

From an algorithmic point of view, there are many results regarding the existence and the computation of these notions (see our Related Work). Here, however, we are interested in exploring the problem from a game theoretic perspective. In particular, we assume that the agents are *strategic*, which means that it is possible for an agent to intentionally misreport her values for (some of) the goods in order to end up with a bundle of higher total value. We see this as a very natural direction, as it captures what may happen in practice in many real-life scenarios where fair division solutions can be applied, e.g., in a divorce settlement. It should be noted here that, in accordance to the existing literature on truthful allocation mechanisms [26, 34, 40, 41, 1, 2, 18], we assume there are *no monetary transfers*. Therefore, a *mechanism* in our setting is just an algorithm that takes as input the, possibly misreported, values that the agents declare. The existence of *truthful* mechanisms, i.e., mechanisms where no agent ever has an incentive to lie, was studied in the same setting by Amanatidis et al. [2] who showed that, even for two agents, truthfulness and fairness are incompatible by providing impossibility results for several fairness notions. As a consequence, the next natural question to ask is:

Is it possible to have non-truthful mechanisms that have equilibria which define fair *allocations?*

So, our main quest is to investigate whether there exist mechanisms that have *pure Nash equilibria* for every instance and each allocation corresponding to an equilibrium provides fairness guarantees with respect to the *true* valuation functions of the agents. The stability notion of a pure Nash equilibrium, on which we focus here, describes a state where each agent plays a deterministic strategy (namely, reports her value for each good) and no agent can attain higher value by deviating to a different strategy.

1.1 Our Contributions

To the best of our knowledge, our work is the first to consider the above question. The results we provide are mostly positive, as we show that the class of mechanisms that are implementable in polynomial time, have pure Nash equilibria for every instance, and provide some fairness guarantee at the allocations they produce in their equilibria is non-empty. Specifically, in Section 3, we study a mechanism adaptation of the Round-Robin algorithm which is known to produce EF1 allocations in the non-strategic setting [20]. Also, under some mild assumptions which we show that can be lifted, Aziz et al. [8] showed that the Round-Robin mechanism always has pure Nash equilibria. Further, in Section 4, we consider the stronger fairness notion of EFX. We focus on the case of two agents and study a mechanism adaptation of the algorithm of Plaut and Roughgarden [42], Mod-Cut&Choose, which is known to always produce EFX allocations in the non-strategic setting. Our main contributions can be summarized as follows:

- Round-Robin has pure Nash equilibria for every instance and these equilibria induce allocations that are always EF1 with respect to the underlying true values (Theorems 3.2 and A.3). That is, Round-Robin retains its fairness properties at its equilibria, even when the input is given by strategic agents! To show this, we combine well-known properties of Round-Robin with a novel recursive construction of "nicely structured" bid profiles. We consider this as the main technical result of our paper.
- Mod-Cut&Choose has pure Nash equilibria for every instance with two agents and these equilibria induce allocations that are always EFX and MMS with respect to the underlying true values (Theorem 4.4). Notice that for the case of two agents MMS allocations are always EFX allocations, i.e., MMS fairness is stronger. It should be also noted that in the non-strategic setting the allocations returned by Mod-Cut&Choose are not necessarily MMS! (see discussion before Proposition 4.1).
- We generalize a weaker version of the latter. All mechanisms that have pure Nash equilibria for every instance with two agents and these equilibria induce allocations that are always EFX provide stronger MMS guarantees in these allocations than generic EFX allocations do (Theorems 4.6 and 4.7). This provides a very interesting separation between the strategic and non-strategic settings.

1.2 Further Related Work

The non-strategic version of the problem of fairly allocating goods to additive agents has been studied extensively. We provide a summary of indicative results mostly for the notions that we consider. In particular, EF1 allocations always exist and can be computed in polynomial time [37, 39, 20]. For the stronger notion of EFX, the picture is not that clear. It is known that such allocations always exist when there are 2 or 3 agents [20, 33, 21], and in the former case they can be efficiently computed using Mod-Cut&Choose [42]. The existence of complete EFX allocations for 4 or more agents remains one of the most intriguing open problems in fair division. There are, however, positive results for any number of agents if the valuation functions are restricted [6, 38, 29], if it is allowed to discard some of the goods [19, 22, 23, 12], or if one considers approximate EFX allocations [5]. Finally, regarding the notion of MMS, allocations that provide this guarantee always exist when there are only 2 agents, although computing them is an NP-hard problem [46]. Even worse, for three or more agents, such allocations do not always exist [36]. However, there are algorithms that run in polynomial time and produce constant factor approximation guarantees [36, 3, 9, 32, 31], with $\frac{3}{4} + \frac{1}{12n}$ being the current state of the art [30].

The works of Caragiannis et al. [18] and Amanatidis et al. [1, 2] are very relevant to ours in the sense that they all studied the exact same strategic discrete fair division setting. As we mentioned earlier, though, their focus was different as they were only interested in truthful mechanisms. Amanatidis et al. [2] provided strong impossibility results in this direction: for instances with two agents, no truthful mechanism can consistently produce EF1 (and thus EFX) allocations when there are more than 4 goods, while the best possible approximation with respect to MMS declines linearly with the number of goods.

Aziz et al. [8] studied the existence of pure Nash equilibria of Round-Robin and showed that when no agent values any two goods equally, there always exists a pure Nash equilibrium. In addition, they provided a linear time algorithm that computes the preference rankings (i.e., the orderings of the goods that correspond to the reported values) that leads to this equilibrium, thus giving a constructive solution. Aziz et al. [7] showed that computing best responses for Round-Robin, and for *sequential mechanisms* more generally, is NP-hard, fixing an error in the work of Bouveret and Lang [13] on the same topic.

We conclude by pointing out that in contrast to the case of indivisible goods, the problem of fairly allocating a set of divisible goods to a set of strategic agents has been repeatedly studied. For some indicative papers in this line of work, we refer the reader to [25, 24, 15, 11, 16] and references therein.

2 Preliminaries

We consider the problem of allocating a set of indivisible goods to a set of agents in a fair manner under the presence of incentives. For $a \in \mathbb{N}$ we use $[a]$ to denote the set $\{1, 2, \ldots, a\}$. An instance to our problem is an ordered triple $(N, M, \mathbf{v})$, where $N = [n]$ is a set of n agents, $M = \{g_1, \ldots, g_m\}$ is a set of m goods, and $\mathbf{v} = (v_1, \ldots, v_n)$ is a vector of the agents' additive valuation functions. In particular, each agent i has a non-negative value $v_i(\{g\})$ (or simply $v_i(g)$) for each good $g \in M$, and for every $S, T \subseteq M$ with $S \cap T = \emptyset$ we have $v_i(S \cup T) = v_i(S) + v_i(T)$. Equivalently, the value of an agent is simply the sum of the values of the goods that she got. We assume there is no free disposal, which means that all the goods must be allocated. Thus, an allocation $(A_1, \ldots, A_n)$, where A_i is the *bundle* of agent i, is a partition of M. It is often useful to refer to the order of preference an agent has over the goods. We say that a valuation function v_i *induces a preference ranking* $\succeq_i$ if $g \succeq_i g' \Leftrightarrow v_i(g) \geq v_i(g')$ for all $g, g' \in M$. We use $\succ_i$ if the corresponding preference ranking is *strict*, i.e., when $g \succeq_i g' \,\wedge\, g' \succeq_i g \;\Rightarrow\; g = g'$, for all $g, g' \in M$.

2.1 Fairness Notions

There is a significant number of different notions one can use to determine which allocations are "fair". The most prominent such notions are *envy-freeness* (EF) [28, 27, 45] and *proportionality* (PROP) [44], and, in the discrete setting we study here, their relaxations, namely *envy-freeness up to one good* (EF1) [17], *envy-freeness up to any good* (EFX) [20], and *maximin share fairness* (MMS) [17]. Particularly for additive valuation functions, we have that EF $\Rightarrow$ EFX $\Rightarrow$ EF1 and EF $\Rightarrow$ PROP $\Rightarrow$ MMS, where $X \Rightarrow Y$ means that any allocation that satisfies fairness criterion X always satisfies fairness criterion Y as well.

Definition 2.1. An allocation $(A_1, \ldots, A_n)$ is

- *envy-free* (EF), if for every $i, j \in N$, $v_i(A_i) \geq v_i(A_j)$.
- *envy-free up to one good* (EF1), if for every pair of agents $i, j \in N$, with $A_j \neq \emptyset$, there exists a good $g \in A_j$, such that $v_i(A_i) \geq v_i(A_j \setminus \{g\})$.
- *envy-free up to any good* (EFX), if for every pair $i, j \in N$, with $A_j \neq \emptyset$ and every good $g \in A_j$ with $v_i(g) > 0$, it holds that $v_i(A_i) \geq v_i(A_j \setminus \{g\})$.

While these notions rely on comparisons among the agents, proportionality focuses on everyone receiving at least a $1/n$ fraction of the total value.

Definition 2.2. An allocation $(A_1, \ldots, A_n)$ is *proportional* (PROP), if for every $i \in N$, $v_i(A_i) \geq v_i(M)/n$.

In the same direction, but adjusted for indivisible goods, a number of fairness notions have been based on the notion of *maximin shares* [17]. Imagine that agent i is asked to partition the goods into n bundles, under the condition that she will receive the worst bundle among those. If the resources were divisible, then she would clearly split everything evenly into n bundles of value $v_i(M)/n$ each, thus capturing the benchmark required for proportionality. However, now that the goods are indivisible, agent i would like to create a partition maximizing the minimum value of a bundle. This value is her maximin share.

Definition 2.3. Given a subset $S \subseteq M$ of goods, the n-*maximin share* of agent i with respect to S is

$$\boldsymbol{\mu}_i(n, S) = \max_{\mathcal{A} \in \Pi_n(S)} \min_{A_j \in \mathcal{A}} v_i(A_j)\,,$$

where $\Pi_n(S)$ is the set of all partitions of S into n bundles.

From the definition and the preceding discussion, we have that $n \cdot \boldsymbol{\mu}_i(n, S) \leq v_i(S)$. When $S = M$, we call $\boldsymbol{\mu}_i(n, M)$ the *maximin share* of agent i and denote it by $\boldsymbol{\mu}_i$ as long as it is clear what n and M are.

Definition 2.4. An allocation $\mathcal{A} = (A_1, \dots, A_n)$ is called an α-*maximin share fair* (α-MMS) allocation if $v_i(A_i) \geq \alpha \cdot \boldsymbol{\mu}_i$, for every $i \in N$. When $\alpha = 1$ we just say that $\mathcal{A}$ is an MMS allocation.

Besides MMS, there exist other fairness criteria based on the notion of maximin shares, like *pairwise maximin share fairness* (PMMS) [20] and *groupwise maximin share fairness* (GMMS) [10]. While we are not going into more details about them, it should be noted that PMMS $\Rightarrow$ EFX [20] and that for $n = 2$, MMS, PMMS, and GMMS coincide. In particular, we need the following result of Caragiannis et al. [20].

Theorem 2.5 (Follows from Theorem 4.6 of [20])**.** *For $n = 2$, any MMS allocation is also an EFX allocation.*

In addition to the implications mentioned so far, one can consider how the approximate versions of EF1, EFX and MMS relate to each other (see [4]). Here we need the following result about the worst case MMS guarantee of an EFX allocation for the case of two agents.

Theorem 2.6 (Follows from Proposition 3.3 of [4])**.** *For $n = 2$, any EFX allocation is also a $\frac{2}{3}$-MMS allocation. This guarantee is tight, in the sense that for every $\delta > 0$ there exists an EFX allocation that is not a $\left(\frac{2}{3}+\delta\right)$-MMS allocation, for any $m \geq 4$.*

2.2 Mechanisms and Equilibria

We are interested in *mechanisms* that produce allocations with fairness guarantees. In our setting, where there are *no payments*, an allocation mechanism $\mathcal{M}$ is essentially just an algorithm that takes its input from the agents and allocates all the goods to them. We use this distinction in terminology to highlight that this reported input may differ from the actual valuation functions. In particular, we assume that each agent i reports a *bid vector* $\boldsymbol{b}_i = (b_{i1}, b_{i2}, \dots, b_{im})$, where $b_{ij} \geq 0$ is the value agent i claims to have for good $g_j \in M$. A mechanism $\mathcal{M}$ takes as input a *bid profile* $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_n)$ of bid vectors and outputs an allocation $\mathcal{M}(\mathbf{b})$. In our setting we assume that the agents are *strategic*, i.e., an agent may misreport her true values if this results to a better allocation from her point of view. Hence, in general, $\boldsymbol{b}_i \neq (v_i(g_1), v_i(g_2), \dots, v_i(g_m))$. While $\boldsymbol{b}_i$ is defined as a vector, for a generic good $h \in M$ it is often convenient to use the function notation $\boldsymbol{b}_i(h)$ to denote the bid value $b_{i\ell}$, where ℓ is such that $h = g_\ell$; extending this we may write $\boldsymbol{b}_i(S)$ for $\sum_{h \in S} \boldsymbol{b}_i(h)$. Like above, we say that a bid vector $\boldsymbol{b}_i$ induces a preference ranking $\succeq_i$ if $g \succeq_i g' \Leftrightarrow \boldsymbol{b}_i(g) \geq \boldsymbol{b}_i(g')$ for all $g, g' \in M$, and use $\succ_i$ for strict rankings.

We focus on the fairness guarantees of the (pure) equilibria of the mechanisms we study. As is common, given a profile $\mathbf{b} = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$, we write $\mathbf{b}_{-i}$ to denote $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_{i-1}, \boldsymbol{b}_{i+1}, \dots, \boldsymbol{b}_n)$ and, given a bid vector $\boldsymbol{b}'_i$, we use $(\boldsymbol{b}'_i, \mathbf{b}_{-i})$ to denote the profile $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_{i-1}, \boldsymbol{b}'_i, \boldsymbol{b}_{i+1}, \dots, \boldsymbol{b}_n)$. For the next definition we abuse the notation slightly: given an allocation $\mathcal{A} = (A_1, \dots, A_n)$, we write $v_i(\mathcal{A})$ to denote $v_i(A_i)$.

Definition 2.7. Let $\mathcal{M}$ be an allocation mechanism and consider a profile $\mathbf{b} = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$. We say that $\boldsymbol{b}_i$ is a *best response* to $\mathbf{b}_{-i}$ if for every $\boldsymbol{b}'_i \in \mathbb{R}^m_{\geq 0}$, we have $v_i(\mathcal{M}(\boldsymbol{b}'_i, \mathbf{b}_{-i})) \leq v_i(\mathcal{M}(\mathbf{b}))$. The profile $\mathbf{b}$ is a *pure Nash equilibrium* (PNE) if, for each $i \in N$, $\boldsymbol{b}_i$ is a best response to $\mathbf{b}_{-i}$.

When $\mathbf{b}$ is a PNE and the allocation $\mathcal{M}(\mathbf{b})$ has a fairness guarantee, e.g., $\mathcal{M}(\mathbf{b})$ is EF1, we will atribute the same guarantee to the profile itself, i.e., we will say that $\mathbf{b}$ is EF1.

Remark 2.8. The mechanisms we consider in this work run in polynomial time. However there are computational complexity questions that go beyond the mechanisms themselves. For instance, how does an agent compute a best response or how do all the agents reach an equilibrium? While we consider such questions interesting directions for future work, we do not study them here and we only focus on the fairness properties of PNE. It should be noted, however, that such problems are typically hard. For instance, computing a best response for Round-Robin is NP-hard in general [7] (although for fixed n it can be done in polynomial time [47]), and we show that the same is true for Mod-Cut&Choose (Proposition 4.1).

Remark 2.9. An easy observation on the main question of this work is that *any* PNE of *any* α-approximation mechanism for computing MMS allocations is an α-MMS allocation. Indeed, this is true, not only for MMS but for any fairness notion that depends on agents achieving specific value benchmarks that depend on their own valuation function, e.g., it is also true for PROP. While this is definitely interesting to note, nothing is known on the existence of PNE of any constant factor approximation algorithm for computing MMS allocations in the literature. Even for a very simple $1/2$-approximation algorithm that only slightly differs from Round-Robin [3], showing that PNE always exist seems very challenging. Clearly, an existence result for any such algorithm [36, 3, 9, 32, 31, 30] would imply an analogue of Theorem 3.2 for approximate MMS. We see this as another promising direction in line with the research agenda we initiate here.

3 Fairness of Nash Equilibria of Round-Robin

In this section we focus on one of the simplest and most well-studied allocation algorithms, Round-Robin, a draft algorithm where the agents take turns and in each turn the active agent receives her most preferred available (i.e., unallocated) good. Below we state Round-Robin as a mechanism (Mechanism 1) that takes as input a bid profile rather than the valuation functions of the agents. In its full generality, Round-Robin should also take a permutation N as an input to determine the priority of the agents. Here, for the sake of presentation, we assume that the agents in each *round* (lines 3–6) are always considered according to their "name", i.e., agent 1 is considered first, agent 2 second, and so on. This is without loss of generality, as it only requires renaming the agents accordingly.

Mechanism 1 Round-Robin$(\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ // For $i \in N$, $\boldsymbol{b}_i = (b_{i1}, \dots, b_{im})$ is the bid of agent i.

1: $S = M$; $(A_1, \dots, A_n) = (\emptyset, \dots, \emptyset)$; $k = \lceil m/n \rceil$
2: **for** $r = 1, \dots, k$ **do** // Each value of r determines the corresponding *round*.
3:     **for** $i = 1, \dots, n$ **do**
4:         $g = \arg\max_{h \in S} \boldsymbol{b}_i(h)$ // Break ties lexicographically (hence we use "$=$" instead of the correct "$\in$").
5:         $A_i = A_i \cup \{g\}$ // The current agent receives (what appears to be) her favorite available good.
6:         $S = S \setminus \{g\}$ // The good is no longer available.
7: **return** $\mathcal{A} = (A_1, \dots, A_n)$

While it is long known that truth-telling is generally not a PNE in sequential allocation mechanisms (a special case of which is Round-Robin) [35], we present here a minimal example that illustrates the mechanics of manipulation. Let $N = \{1, 2\}$ and $M = \{a, b, c\}$ with the valuation functions being as shown in the table on the left. The circles show the allocation returned by Round-Robin when the agents bid their true values, while the superscripts indicate in which order were the goods assigned. Given that agent 2 is not particularly interested to good a, agent 1 can manipulate the mechanism into giving her $\{a, b\}$ instead $\{a, c\}$ by claiming that these are her top goods as in the table on the right.

	a	b	c
v_1 :	$\textcircled{6}^1$	5	$\textcircled{4}^3$
v_2 :	4	$\textcircled{6}^2$	5

	a	b	c
$\boldsymbol{b}_1$:	$\textcircled{5}^3$	$\textcircled{6}^1$	4
v_2 :	4	6	$\textcircled{5}^2$

Thus, bidding according to v_1, v_2 is not a PNE. The example is minimal, in the sense that with just 1 agent or less than 3 goods truth-telling is a PNE of Round-Robin almost trivially.

Before moving to the main technical part of this section, we discuss some assumptions that again are without loss of generality, and give an easy proof for the case of two agents. As we have mentioned in the Introduction, it is known that, as an algorithm, Round-Robin outputs EF1 allocations when all agents have additive valuation functions [39, 20]. Also Round-Robin as a mechanism is known to have PNE for any instance where *no agent values two goods exactly the same*, and at least some such equilibria (namely, the ones consistent with the so-called *bluff profile*) are easy to compute [8]. From a technical point of view, this assumption that all the valuation functions induce strict preference rankings is convenient, as it greatly reduces the number of corner cases one has to deal with. However, as we show in Theorem A.3 in Appendix A, the result of Aziz et al. [8] extends to general additive valuation functions. On a different but related note, we assume, for the remainder of this section, that all the bid vectors induce strict preference rankings (but not necessarily consistent with the preference rankings induced by the corresponding valuation functions). This is without loss of generality, because even if a bid vector contains some bids that are equal to each other, a strict preference ranking is imposed by the lexicographic tie-breaking of the mechanism itself. So, formally, when we abuse the notation and write $g \succ_i h$ we mean that either $\boldsymbol{b}_i(g) > \boldsymbol{b}_i(h)$, or $\boldsymbol{b}_i(g) = \boldsymbol{b}_i(h)$ and g has a lower index than h in the standard naming of goods as $g_1, g_2, \dots, g_m$.

Next we show that for only two agents all PNE of Round-Robin are EF1 with respect to the real valuation functions. To appreciate this easy result, one should compare it to the involved general proof of Theorem 3.2 in the next section, the full complexity of which seems to be necessary even for $n = 3$. The straightforward but crucial observation that makes things work here is that envy-freeness and proportionality are equivalent when there are only two agents.

Theorem 3.1. *For any fair division instance $\mathcal{I} = (\{1, 2\}, M, \mathbf{v})$, every PNE of the Round-Robin mechanism is EF1 with respect to the valuation functions v_1, v_2.*

Proof. Suppose towards a contradiction that this is not the case. That is, there exists a PNE $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ such that in the allocation (A_1, A_2) returned by Round-Robin($\mathbf{b}$) at least one of the agents envies the other, even after removing the most valuable good from her bundle. We will examine each agent separately.

If agent 1 does not see the allocation as EF1, then this means that she does not see it as EF either. Since envy-freeness and proportionality are equivalent for $n = 2$, we get that $v_1(A_1) < v_1(M)/2$. It is known that, no matter what others bid, if the agent with the highest priority (here agent 1) reports her true values to Round-Robin, the resulting allocation is EF from her perspective (see, e.g., the proof of Theorem 12.2 in [39]). So, if (A'_1, A'_2) is the allocation after agent 1 deviates to her true values, it is EF from the point of view of the agent 1, which in turn implies that $v_1(A'_1) \geq v_1(M)/2 > v_1(A_1)$. This contradicts the fact that $\mathbf{b}$ is a PNE.

If agent 2 does not see the allocation as EF1, then let h_1 be the good that agent 1 takes during the first round of round-robin, and $g^* \in \arg\max_{h \in A_1} v_2(h)$ be the highest valued good in A_1 according to agent 2. Since agent 2 does not consider (A_1, A_2) to be EF1, we have that $v_2(A_2) < v_2(A_1 \setminus \{g^*\}) \leq v_2(A_1 \setminus \{h_1\})$. This implies that the partition $(A_1 \setminus \{h_1\}, A_2)$ of $M \setminus \{h_1\}$ is not an EF allocation with respect to agent 2. Now we may use a similar argument as in the previous case. First, since envy-freeness and proportionality are equivalent when $n = 2$, we get that $v_2(A_2) < v_2(M \setminus \{h_1\})/2$. Then suppose agent 2 deviates to reporting her true values and let (A_1', A_2') be the resulting allocation. Notice that the allocation of good h_1 is not affected by the deviation; it is still given to agent 1 during the first step of Round-Robin. From that point forward, the execution of the mechanism would be exactly the same as it would be if the input was the restrictions of $\boldsymbol{b}_1, v_2$ on $M \setminus \{h_1\}$ and agent 2 had higher priority than agent 1. The latter would result in an EF allocation with respect to agent 2 and, in particular, to the allocation $(A_1' \setminus \{h_1\}, A_2')$. That is, we have $v_2(A_2') \geq v_2(A_1' \setminus \{h_1\})$ and, therefore, $v_2(A_2) \geq v_2(M \setminus \{h_1\})/2 > v_2(A_2)$. Like before, this contradicts the fact that $\mathbf{b}$ is a PNE. ☐

Moving to the case of general $n \geq 3$, the above simple argument no longer works. When an agent i does not consider an allocation EF1 because of an agent i', this does not imply that i got value less than $1/n$ of her value for the reduced bundle $M \setminus \{g^*\}$, where g^* is her best good in $A_{i'}$. The reason for this is that PROP $\not\Rightarrow$ EF anymore.

3.1 Nash Equilibria of Round-Robin for Any Number of Agents

Here we state and prove the main result of our work. Despite its proof being rather involved, the intuition behind it is simple. On one hand, whenever an agent bids truthfully, she sees the resulting allocation as being EF1. On the other hand, no matter what an agent bids, we show it is possible to "replace" her with an imaginary version of herself who does not affect the allocation, and not only bids truthfully, but she considers the bundles of the allocation to be as valuable as the original agent thought they were. The rather elaborate formal argument relies on the recursive construction of auxiliary valuation functions and bids, and on the fact that small changes in a single preference ranking minimally change the "history" of available goods during the execution of the mechanism.

Theorem 3.2. *For any fair division instance $\mathcal{I} = (N, M, \mathbf{v})$, every PNE of the Round-Robin mechanism is EF1 with respect to the valuation functions $v_1, \ldots, v_n$.*

As we will see shortly, proving Theorem 3.2 reduces to showing that the agent who "picks first" in the Round-Robin mechanism views the final allocation as envy-free, as long as she bids a best response to other agents' bids. While Theorem 3.3 sounds very much like the standard statement about the value of the first agent in the algorithmic setting, its proof relies on a technical lemma that carefully builds a "nice" instance which is equivalent, in some sense, to the original. Recall that we have assumed that the agents' priority is indicated by their indices.

Theorem 3.3. *For any fair division instance $\mathcal{I} = (N, M, \mathbf{v})$, if the reported bid vector $\boldsymbol{b}_1$ of agent 1 is a best response to the (fixed) bid vectors $\boldsymbol{b}_2, \ldots, \boldsymbol{b}_n$ of all other players, then agent 1 does not envy (with respect to v_1) any bundle in the allocation outputted by Round-Robin$(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$.*

Note that since we are interested in PNE, it is always the case that each agent's bid is a best response to other agents' bids. As mentioned above, Theorem 3.3 is essentially a corollary to Lemma 3.4. The lemma shows the existence of an alternative version of agent 1 who is truthful, her presence does not affect the original allocation, and, as long as the allocation is the same, she shares the same values with the original agent 1. While its proof is rather involved, the high level idea is that we recursively construct a sequence of bids and valuation functions, each pair of which preserves the original allocation and the view of agent 1 for it, while being closer to being truthful. To achieve this we occasionally move value between the goods originally allocated to agent 1 and update the bid accordingly.

Lemma 3.4. *Suppose that the valuation function v_1 induces a strict preference ranking on the goods, i.e., for any $g, h \in M$, $v_1(g) = v_1(h) \Rightarrow g = h$. Let $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n)$ be such that $\boldsymbol{b}_1$ is a best response of agent 1 to $\mathbf{b}_{-i} = (\boldsymbol{b}_2, \ldots, \boldsymbol{b}_n)$. Then there exists a valuation function v_1^* with the following properties:*

- *If $\boldsymbol{b}^*_1 = (v^*_1(g_1), v^*_1(g_2), \dots, v^*_1(g_m))$, i.e., $\boldsymbol{b}^*_1$ is the truthful bid for v^*_1, then Round-Robin$(\mathbf{b})$ and Round-Robin$(\boldsymbol{b}^*_1, \mathbf{b}_{-1})$ produce the same allocation $(A_1, \dots, A_n)$.*
- $v^*_1(A_1) = v_1(A_1)$.
- *For every good $g \subseteq M \setminus A_1$, it holds that $v^*_1(g) = v_1(g)$.*

For the sake of presentation, we defer the proof of the lemma to the end of this section (as it needs an additional technical lemma that is itself quite long) and move to the proofs of Theorems 3.2 and 3.3. In fact, given Lemma 3.4, the two theorems are not hard to prove.

Proof of Theorem 3.3. Consider an arbitrary instance $\mathcal{I} = (N, M, \mathbf{v})$ and assume that the input of Round-Robin is $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_n)$, where $\boldsymbol{b}_1$ is a best response of agent 1 to $\mathbf{b}_{-i} = (\boldsymbol{b}_2, \dots, \boldsymbol{b}_n)$ according to her valuation function v_1. Let $(A_1, \dots, A_n)$ be the output of Round-Robin$(\mathbf{b})$. In order to apply Lemma 3.4, we need v_1 to induce a strict preference ranking over the goods. For the sake of presentation, we assume here that this is indeed the case, and we treat the general case formally in Appendix A. So, we now consider the hypothetical scenario implied by Lemma 3.4 in this case: keeping agents 2 through n fixed, suppose that the valuation function of agent 1 is the function v^*_1 given by the lemma, and her bid $\boldsymbol{b}^*_1$ is the truthful bid for v^*_1. The first part of Lemma 3.4 guarantees that the output of Round-Robin$(\boldsymbol{b}^*_1, \mathbf{b}_{-i})$ remains $(A_1, \dots, A_n)$.

It is known that, no matter what others bid, if the agent with the highest priority (here agent 1 with v^*_1) reports her true values to Round-Robin, the resulting allocation is EF from her perspective (see, e.g., the proof of Theorem 12.2 in [39]). In our hypothetical scenario this is the case for agent 1 and it translates into having $v^*_1(A_1) \geq v^*_1(A_i)$ for all $i \in N$. Then the second and third parts of the lemma imply that $v_1(A_1) \geq v_1(A_i)$ for all $i \in N$, i.e., agent 1 does not envy any bundle in the original instance. □

Having shown Theorem 3.3, the proof of Theorem 3.2 is of similar flavour to the proof on Round-Robin producing EF1 allocations in the non-strategic setting [39].

Proof of Theorem 3.2. Let $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_n)$ be a PNE of the Round-Robin mechanism for the instance $\mathcal{I}$. By Theorem 3.3, it is clear that the allocation returned by Round-Robin$(\mathbf{b})$ is EF, and hence EF1, from the point of view of agent 1. We fix an agent ℓ, where $\ell \geq 2$. For $i \in [\ell - 1]$, let h_i be the good that agent i *claims to be* her favourite among the goods that are available when it is her turn in the first round, i.e., $h_i = \arg\max_{h \in M \setminus \{h_1 \dots, h_{i-1}\}} b_i(h)$. Right before agent ℓ is first assigned a good, all goods in $H = \{h_1, \dots, h_{\ell-1}\}$ have already been allocated. We are going to consider the instance $\mathcal{I}' = (N', M', \mathbf{v}')$ in which all goods in H are missing. That is, $N' = N$, $M' = M \setminus H$, and $\mathbf{v}' = (v'_1, \dots, v'_n)$ where $v'_i = v_i|_{M'}$, for $i \in [n]$, is the restriction of the function v_i on M'. Similarly define $\boldsymbol{b}'_i = \boldsymbol{b}_i|_{M'}$, for $i \in [n]$, the restrictions of the bids to the available goods, and $\mathbf{b}' = (\boldsymbol{b}'_1, \dots, \boldsymbol{b}'_n)$. Finally, we consider the version of Round-Robin, call it Round-Robin$_\ell$, that starts with agent ℓ and then follows the indices in increasing order.

We claim that for Round-Robin$_\ell$ the bid $\boldsymbol{b}'_\ell$ is a best response for agent ℓ assuming that the restricted bid vectors of all the other agents are fixed. To see this, notice that for any $\boldsymbol{c}_\ell = (c_{\ell 1}, c_{\ell 2}, \dots, c_{\ell m})$, the bundles given to agent ℓ by Round-Robin$(\boldsymbol{c}_\ell, \mathbf{b}_{-\ell})$ and Round-Robin$_\ell(\boldsymbol{c}_\ell|_{M'}, \mathbf{b}'_{-\ell})$ are the same! In fact, the execution of Round-Robin$_\ell(\boldsymbol{c}_\ell|_{M'}, \mathbf{b}'_{-\ell})$ is identical to the execution of Round-Robin$(\boldsymbol{c}_\ell, \mathbf{b}_{-\ell})$ from its ℓth step onward. So, if $\boldsymbol{b}'_\ell$ was not a best response in the restricted instance, then there would be a profitable deviation for agent ℓ, say $\boldsymbol{b}^*_\ell$, so that ℓ would prefer her bundle in Round-Robin$_\ell(\boldsymbol{b}^*_\ell, \mathbf{b}'_{-\ell})$ to her bundle in Round-Robin$_\ell(\mathbf{b}')$. This would imply that any extension of $\boldsymbol{b}^*_\ell$ to a bid vector for all goods in M (by arbitrarily assigning numbers to goods in H) would be a profitable deviation for agent ℓ in the profile $\mathbf{b}$ for Round-Robin, contradicting the fact that $\mathbf{b}$ is a PNE.

Now we may apply Theorem 3.3 for Round-Robin$_\ell$ (where agent ℓ plays the role of agent 1 of the theorem's statement) for instance $\mathcal{I}'$ and bid profile $\mathbf{b}'$. The theorem implies that agent ℓ does not envy any bundle in the allocation $(A_1, \dots, A_n)$ outputted by Round-Robin$_\ell(\mathbf{b}')$, i.e., $v'_\ell(A_\ell) \geq v'_\ell(A_i)$, for all $i \in [n]$. Using the observation made above about the execution of Round-Robin$_\ell(\mathbf{b}')$ being identical to the execution of Round-Robin$(\mathbf{b})$ after $\ell - 1$ goods have been allocated, we have that Round-Robin$(\mathbf{b})$ returns the allocation $(A_1 \cup \{h_1\}, \dots, A_{\ell-1} \cup \{h_{\ell-1}\}, A_\ell, \dots, A_n)$. So, for any $i < \ell$ we have $v_\ell(A_\ell) = v'_\ell(A_\ell) \geq v'_\ell(A_i) = v_\ell(A_i) = v_\ell(A_i \cup \{h_i\}) - v_\ell(h_i)$, while for $i > \ell$ we simply have $v_\ell(A_\ell) = v'_\ell(A_\ell) \geq v'_\ell(A_i) = v_\ell(A_i)$. Thus, the allocation returned by Round-Robin$(\mathbf{b})$ is EF1 from the point of view of agent ℓ. □

Before we move on to the proof of Lemma 3.4, we state another technical lemma. Suppose an agent changes her bid so that in her preference ranking a single good is moved down the ranking, and then—keeping everything else fixed—we run Round-Robin on the new instance. Surprisingly, Lemma 3.6 states that, in any step, the set of available goods differs by at most one good from the corresponding set in the original run of Round-Robin. To formalize this, we need some additional notation and terminology.

Definition 3.5. Let $\succ$ and $\succ'$ be two strict preference rankings on M and $\{q_1, q_2, \dots, q_m\}$ be a renaming of the goods according to $\succ$, i.e., $q_1 \succ q_2 \succ \dots \succ q_m$. We say that $\succ$ and $\succ'$ are *within a partial slide of each other* if there exist $x, y \in [m]$, $x < y$, such that

$$q_1 \succ' \dots \succ' q_{x-1} \succ' q_{x+1} \succ' \dots \succ' q_y \succ' q_x \succ' q_{y+1} \succ' \dots \succ' q_m \,.$$

Also, given a profile $\mathbf{b} = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$, let $M_t(\mathbf{b})$ denote the set of available goods right after $t-1$ goods have been allocated in a run of Round-Robin$(\mathbf{b})$.

Lemma 3.6. *Let $\mathbf{b} = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ and $\mathbf{b}' = (\boldsymbol{b}'_i, \mathbf{b}_{-i})$ be two profiles such that the corresponding induced preference rankings $\succ_i$ and $\succ'_i$ of agent i are within a partial slide of each other. Then $|M_t(\mathbf{b}) \setminus M_t(\mathbf{b}')| = |M_t(\mathbf{b}') \setminus M_t(\mathbf{b})| \leq 1$ for all $t \in [m+1]$.*

Proof. Clearly, for $t \leq i$ we have $M_t(\mathbf{b}) = M_t(\mathbf{b}')$ as the runs of Round-Robin$(\mathbf{b})$ and Round-Robin$(\mathbf{b}')$ are identical at least up to the allocation of the first $i-1$ goods. We are going to prove the statement by induction on t using this observation as our base case. Assume that for some $t \geq i$, $|M_t(\mathbf{b}) \setminus M_t(\mathbf{b}')| = |M_t(\mathbf{b}') \setminus M_t(\mathbf{b})| \leq 1$. Up to this point, $t-1$ goods have been allocated already. Let j be the next agent to get a good and let g (resp. g') be this good in Round-Robin$(\mathbf{b})$ (resp. in Round-Robin$(\mathbf{b}')$).

Case 1 ($M_t(\mathbf{b}) = M_t(\mathbf{b}')$). No matter who j is and what g and g' are, it is straightforward to see that either $M_{t+1}(\mathbf{b}) \setminus M_{t+1}(\mathbf{b}') = M_{t+1}(\mathbf{b}') \setminus M_{t+1}(\mathbf{b}) = \emptyset$ (when $g = g'$), or $M_{t+1}(\mathbf{b}) \setminus M_{t+1}(\mathbf{b}') = \{g'\}$ and $M_{t+1}(\mathbf{b}') \setminus M_{t+1}(\mathbf{b}) = \{g\}$ (when $g \neq g'$). Thus, $|M_{t+1}(\mathbf{b}) \setminus M_{t+1}(\mathbf{b}')| = |M_{t+1}(\mathbf{b}') \setminus M_{t+1}(\mathbf{b})| \leq 1$.

Before we move to Case 2, it is important to take a better look on how can we move away from Case 1 for the very first time. That is, we want to focus on the first time step t_* when $g \neq g'$, if such a t_* exists for the specific profiles. Notice that this cannot happen if $j \neq i$ at this point. Indeed, if it was $j \neq i$, since $M_\ell(\mathbf{b}) = M_\ell(\mathbf{b}')$ for all $\ell \in [t_*]$ and the induced preference ranking of j in this case is the same in both $\mathbf{b}$ and $\mathbf{b}'$, we have that the two runs of Round-Robin should make the same choice for j in the time step t_*; that would contradict the choice of t_* itself. So, after the same $t_* - 1$ goods have been allocated by Round-Robin$(\mathbf{b})$ and Round-Robin$(\mathbf{b}')$, agent i is about to be given g and g' respectively in the two runs from the set $M_{t_*} = M_{t_*}(\mathbf{b}) = M_{t_*}(\mathbf{b}')$ of available goods. We claim that these two goods cannot be arbitrary. In particular, let $s \in M$ be the good that goes from a better position in $\succ_i$ to a worse position in $\succ'_i$. We are going to show that $g = s$. Recall that g is the best good in M_{t_*} with respect to $\succ_i$; similarly for g' and $\succ'_i$. First, notice that $\succ_i$ and $\succ'_i$ are identical on $M_{t_*} \setminus \{s\}$ and, thus, for g and g' to be distinct at least one of them must be s. Since $g \neq g'$, either $g = s$ or $g' = s$ but not both. Assume for a contradiction that $g' = s$ and $g = x \neq s$. Since $s \in M_{t_*}$, it is available to both. The fact that $x \neq s$ implies that $x \succ_i s$. However, this also mean $x \succ'_i s$ which, given the availability of x, contradicts the choice of g'. We conclude that

$g = s$ and $g' \neq s$. This observation will be crucial for showing that the last subcase of Case 2 below cannot happen.

Case 2 ($M_t(\mathbf{b}) \setminus M_t(\mathbf{b}') = \{h\}$ and $M_t(\mathbf{b}') \setminus M_t(\mathbf{b}) = \{h'\}$). If $g = h$ and $g' = h'$, then we immediately get $M_{t+1}(\mathbf{b}) = M_{t+1}(\mathbf{b}')$. Moreover, if $g = h$ and $g' \neq h'$, then we have that

$$M_{t+1}(\mathbf{b}) \setminus M_{t+1}(\mathbf{b}') = (M_t(\mathbf{b}) \setminus \{h\}) \setminus (M_t(\mathbf{b}') \setminus \{g'\}) = ((M_t(\mathbf{b}) \setminus M_t(\mathbf{b}')) \setminus \{h\}) \cup \{g'\} = \{g'\}\,,$$

where the second equality holds because $g' \in M_t(\mathbf{b}) \cap M_t(\mathbf{b}')$ in this case, and

$$M_{t+1}(\mathbf{b}') \setminus M_{t+1}(\mathbf{b}) = (M_t(\mathbf{b}') \setminus \{g'\}) \setminus (M_t(\mathbf{b}) \setminus \{h\}) = M_t(\mathbf{b}') \setminus M_t(\mathbf{b}) = \{h'\}\,,$$

where here the second equality holds because $g' \in M_t(\mathbf{b})$ and $h \notin M_t(\mathbf{b}')$. The subcase where $g \neq h$ and $g' = h'$ is symmetric and we similarly get

$$M_{t+1}(\mathbf{b}) \setminus M_{t+1}(\mathbf{b}') = \{h\} \quad \text{and} \quad M_{t+1}(\mathbf{b}') \setminus M_{t+1}(\mathbf{b}) = \{g\}\,.$$

It remains to deal with the subcase where $g \neq h$ and $g' \neq h'$. If $g = g'$, then we immediately get $M_{t+1}(\mathbf{b}) \setminus M_{t+1}(\mathbf{b}') = M_t(\mathbf{b}) \setminus M_t(\mathbf{b}') = \{h\}$ and $M_{t+1}(\mathbf{b}') \setminus M_{t+1}(\mathbf{b}) = M_t(\mathbf{b}') \setminus M_t(\mathbf{b}) = \{h'\}$. So, we may assume that $h \neq g \neq g' \neq h'$. We are going to show that this cannot actually happen, as it would lead to a contradiction. Notice that $h \neq g \neq g' \neq h'$ implies $g, g' \in M_t(\mathbf{b}) \cap M_t(\mathbf{b}')$. If agent j is different than agent i, this would mean that $g \succ_j g'$ and $g' \succ_j g$ because of the corresponding choices of the algorithm when the input is $\mathbf{b}$ and $\mathbf{b}'$ respectively (recall that the bid, and thus the induced preference ranking, of j is the same in both profiles); that would be a contradiction. Therefore, it must be the case that $j = i$. Since we are in Case 2, a scenario leading to Case 2 for the first time (as described after Case 1) must have already happened. Consequently, s is not available at this point in $M_t(\mathbf{b})$ and hence $s \notin M_t(\mathbf{b}) \cap M_t(\mathbf{b}')$. This means that $\{g, g'\} \subseteq M \setminus \{s\}$ and, therefore, g and g' have the same ordering in both preference rankings of agent i. That is, $g \succ_i g'$ implies $g \succ_i' g'$, contradicting the optimality of g' in $M_t(\mathbf{b}')$ with respect to $\succ_i'$.

We conclude that in any possible case, $|M_{t+1}(\mathbf{b}) \setminus M_{t+1}(\mathbf{b}')| = |M_{t+1}(\mathbf{b}') \setminus M_{t+1}(\mathbf{b})| \leq 1$. This concludes the induction. ☐

We are now ready to prove Lemma 3.4. As it was noted before the lemma's statement, we will occasionally move value among the goods allocated to agent 1. This is when Lemma 3.6 is crucial. It allows us to guarantee that there is sufficient value for satisfying all the desired properties of the intermediate valuation functions we define.

Proof of Lemma 3.4. Recall that $k = \lceil m/n \rceil$, i.e., we have k total rounds. Let $\succ_1$ be the preference ranking induced by $\boldsymbol{b}_1$ and consider all the goods according to this ranking: $h_1 \succ_1 h_2 \succ_1 \ldots \succ_1 h_m$. Let $n_1 = 1 < n_2 < \cdots < n_k$ be the indices in this ordering of the goods assigned to agent 1 by Round-Robin($\mathbf{b}$), i.e., in round r agent 1 receives good h_{n_r}. This means that $A_1 = \{h_{n_1}, \ldots, h_{n_k}\}$.

We will recursively construct v_1^* from v_1, over the rounds of Round-Robin. In particular, we are going to define a sequence of intermediate bid vectors $\boldsymbol{b}_1^r$ and valuation functions v_1^r, one for each round r starting from the last round k, so that $v_1^* = v_1^1$ and $\boldsymbol{b}_1^* = \boldsymbol{b}_1^1$. For defining each $\boldsymbol{b}_1^r$ we typically use a number of auxiliary bid vectors to break down and better present the construction. Also, for any round r, we are going to maintain that

(i) $v_1^r(A_1) = v_1(A_1)$.

(ii) $v_1^r(g) = v_1(g)$, for any $g \subseteq M \setminus A_1$.

(iii) $\boldsymbol{b}_1^r$ is *truthful from round* r with respect to v_1^r, meaning that for every good that is no better than h_{n_r}, according to the preference ranking $\succ_1^r$ induced by $\boldsymbol{b}_1^r$, we have that its bid matches its value; formally, $g \not\succ_1^r h_{n_r} \Rightarrow \boldsymbol{b}_1^r(g) = v_1^r(g)$.

(iv) The preference ranking $\succ_1^r$ (induced by $\boldsymbol{b}_1^r$) is identical to $\succ_1$ (induced by $\boldsymbol{b}_1$) up to good $h_{n_{r-1}}$.

(v) $\min_{g,h \in M, g \neq h} |v_1^r(g) - v_1^r(h)| > 0$.

Let us focus on round k, i.e., the last round. Let λ be the most valuable (according to v_1) available good at the very beginning of the round. It is easy to see that $h_{n_k} = \lambda$; if not, then by increasing her bid for λ to be slightly above her bid for h_{n_k} agent 1 would end up with the bundle $\{h_{n_1}, \ldots, h_{n_{k-1}}, \lambda\}$ which is a strict improvement over A_1 and would contradict the fact that $\boldsymbol{b}_1$ is a best response of agent 1. We construct the auxiliary bid $\bar{\boldsymbol{b}}_1^k$ by "moving up" in $\succ_1$ every good that is more valuable than λ but comes after it in $\succ_1$. Formally, $\lambda \succ_1 g \wedge v_1(g) > v_1(\lambda) \Rightarrow \boldsymbol{b}_1(h_{n_k}) < \bar{\boldsymbol{b}}_1^k(g) < \boldsymbol{b}_1(h_{n_k-1})$, where these bids are chosen arbitrarily, as long as they are distinct from each other. Note that this small modification does not affect the allocation at all. Indeed, every good the bid of which was improved is still worse than h_{n_k-1} in the preference ranking $\bar{\succ}_1^k$ induced by $\bar{\boldsymbol{b}}_1^k$, so no decision in rounds $1, \ldots, k-1$ is affected and, by the definition of λ, these goods were not actually available for agent 1 in the beginning of round k, so the decisions in round k are not affected either. Next we define $\boldsymbol{b}_1^k$ by replacing the bids with the actual values for every good that is no better than h_{n_k} in $\bar{\succ}_1^k$, as well as by scaling the bids of all other goods to remain larger than $\boldsymbol{b}_1^k(h_{n_k})$, if necessary. While the latter can be done in several ways, we can simply multiply bids by $\boldsymbol{b}_1^k(h_{n_k})/\bar{\boldsymbol{b}}_1^k(h_{n_k})$. Formally, $\boldsymbol{b}_1^k$ is defined by

$$g \not\bar{\succ}_1^k h_{n_k} \Rightarrow \boldsymbol{b}_1^k(g) = v_1(g) \quad \text{and} \quad g \mathrel{\bar{\succ}_1^k} h_{n_k} \Rightarrow \boldsymbol{b}_1^k(g) = \bar{\boldsymbol{b}}_1^k(g) \cdot \boldsymbol{b}_1^k(h_{n_k})/\bar{\boldsymbol{b}}_1^k(h_{n_k})\,.$$

Note that the preference ranking $\succ_1^k$ induced by $\boldsymbol{b}_1^k$ is identical to $\bar{\succ}_1^k$ up to good h_{n_k-1}, and that λ is the good with the highest bid in $\boldsymbol{b}_1^k$ among the goods that are available in the last round. Hence, Round-Robin$(\boldsymbol{b}_1^k, \mathbf{b}_{-1})$ still produces the allocation $(A_1, \ldots, A_n)$. Also, recall that $\bar{\succ}_1^k$ is identical to $\succ_1$ up to good h_{n_k-1} and, thus, up to at least good $h_{n_{k-1}}$, implying that $\succ_1^k$ satisfies property (iv) above. Finally, by setting $v_1^k = v_1$, it is clear that $\boldsymbol{b}_1^k$ is truthful from round k with respect to v_1^k, but also that $\min_{g,h \in M, g \neq h} |v_1^k(g) - v_1^k(h)| > 0$, $v_1^k(A_1) = v_1(A_1)$, and $v_1^k(g) = v_1(g)$, for all $g \subseteq M \setminus A_1$.

Moving to an arbitrary round $r < k$ we are going to follow a similar, albeit a bit more complicated, approach, where now it will be necessary to move value among the goods of A_1. So, assume that $\boldsymbol{b}_1^{r+1}$ and v_1^{r+1} have already been constructed and have the desired properties (i)–(v) mentioned above, and let $\succ_1^{r+1}$ be the preference ranking induced by $\boldsymbol{b}_1^{r+1}$. Consider the execution of Round-Robin$(\boldsymbol{b}_1^{r+1}, \mathbf{b}_{-1})$. For $i \geq r$, let λ_i be the most valuable available good with respect to v_1^{r+1} at the very beginning of round i and ℓ_i be the most valuable good with respect to v_1^{r+1} (or equivalently with respect to v_1 as $\ell_i \in M \setminus A_1$) that is allocated to some *other* agent during round i. By property (iii) of $\boldsymbol{b}_1^{r+1}$ and v_1^{r+1} we know that in future rounds agent 1 will have $\lambda_{r+1} = h_{n_{r+1}}, \lambda_{r+2} = h_{n_{r+2}}, \ldots, \lambda_k = h_{n_k}$ allocated to her. By property (iv) of $\boldsymbol{b}_1^{r+1}$ we further know that in the current round agent 1 is going to get good h_{n_r}. Unlike what happened for round k, however, here h_{n_r} may be different from λ_r. We will consider two cases depending on this.

First, though, similarly to what we did before, we define the auxiliary bid $\bar{\bar{\boldsymbol{b}}}_1^r$ by setting $\boldsymbol{b}_1^{r+1}(h_{n_r}) < \bar{\bar{\boldsymbol{b}}}_1^r(g) < \boldsymbol{b}_1^{r+1}(h_{n_r-1})$ for all goods g such that $h_{n_r} \succ_1^{r+1} g$ and $v_1^{r+1}(g) > v_1(\lambda_r)$; these $\bar{\bar{\boldsymbol{b}}}_1^r$ entries are arbitrary, as long as they satisfy the inequalities and are distinct from each other. By now it should be clear that moving from $\boldsymbol{b}_1^{r+1}$ to $\bar{\bar{\boldsymbol{b}}}_1^r$ does not affect the allocation since every good that had its bid improved is still worse than h_{n_r-1} in the preference ranking $\bar{\bar{\succ}}_1^r$ induced by $\bar{\bar{\boldsymbol{b}}}_1^r$ and, by the definition of λ_r, these goods were already not available in the beginning of round r. That is, Round-Robin$\big(\bar{\bar{\boldsymbol{b}}}_1^r, \mathbf{b}_{-1}\big)$ returns $(A_1, \ldots, A_n)$.

<u>*Case 1* $(h_{n_r} = \lambda_r)$.</u> This case is similar to what we did for round k. We go straight from $\bar{\bar{\boldsymbol{b}}}_1^r$ to $\boldsymbol{b}_1^r$ by replacing the bids with the corresponding v_1^{r+1} values for all goods that are no better than h_{n_r} in $\bar{\bar{\succ}}_1^r$, and by scaling

the bids of all other goods to remain larger than $v_1^{r+1}(\lambda_r) = \boldsymbol{b}_1^r(\lambda_r) = \boldsymbol{b}_1^r(h_{n_r})$. Formally, we have

$$g \not\mathrel{\bar{\bar{\succ}}}_1^r h_{n_r} \Rightarrow \boldsymbol{b}_1^r(g) = v_1^{r+1}(g) \quad \text{and} \quad g \mathrel{\bar{\bar{\succ}}}_1^r h_{n_r} \Rightarrow \boldsymbol{b}_1^r(g) = \bar{\bar{\boldsymbol{b}}}_1^r(g) \cdot v_1^{r+1}(\lambda_r)/\bar{\bar{\boldsymbol{b}}}_1^r(h_{n_r})\,.$$

The preference ranking $\succ_1^r$ induced by $\boldsymbol{b}_1^r$ is identical to $\bar{\bar{\succ}}_1^r$ up to good h_{n_r}, so Round-Robin$(\boldsymbol{b}_1^r, \mathbf{b}_{-1})$ up to the beginning of round r still allocates $\{h_{n_1}, \ldots, h_{n_{r-1}}\}$ to agent 1 in that order. Also, from good h_{n_r} onward, $\succ_1^r$ is defined in such a way that the best available good in the beginning of round $i \geq r$ with respect to $\succ_1^r$ is h_{n_i}. Therefore, the final bundle for agent 1 is still A_1 and the overall allocation is still $(A_1, \ldots, A_n)$ as $\mathbf{b}_{-1}$ is fixed and goods in A_1 are allocated in the exact same order. Moreover, recall that $\bar{\bar{\succ}}_1^r$ is identical to $\succ_1$ up to good $h_{n_{r-1}}$ (in fact, up to good h_{n_r-1}), implying that $\succ_1^r$ satisfies property (iv). Given that no changes to values were necessary and that we made the relevant (for rounds $r, \ldots, k$) entries of $\boldsymbol{b}_1^r$ equal to the corresponding v_1^{r+1} values, we may set $v_1^r = v_1^{r+1}$ to get that $\boldsymbol{b}_1^r$ is truthful from round r with respect to v_1^r, but also that $\min_{g,h\in M, g\neq h} |v_1^r(g) - v_1^r(h)| > 0$, $v_1^r(A_1) = v_1(A_1)$, and $v_1^r(g) = v_1(g)$, for all $g \subseteq M \setminus A_1$.

Case 2 ($h_{n_r} \neq \lambda_r$). Here we are going to move value from goods $\lambda_{r+1}, \lambda_{r+2}, \ldots, \lambda_k$ to h_{n_r} while defining v_1^r. The main idea is that we would like h_{n_r} to become the most valuable available good at the beginning of round r with respect to v_1^r, although this is not the case for v_1^{r+1} as $v_1^{r+1}(h_{n_r}) < v_1^{r+1}(\lambda_r)$. The constraints we need to satisfy make this task rather tricky: properties (i) and (ii) must hold, so value can only be transferred between goods of A_1, but this should happen in a way that ensures that in future rounds the goods given to agent 1 remain $h_{n_{r+1}}, \ldots, h_{n_k}$ in that order.

We begin with a rather benign modification of $\bar{\bar{\boldsymbol{b}}}_1^r$, which is almost identical to what we did in Case 1, except that we do not update the bid of h_{n_r} with its v_1^{r+1} value. We do this to make sure that h_{n_r} still seems like the most attractive good of round r and the overall allocation remains the same. Specifically, we define the auxiliary bid $\bar{\boldsymbol{b}}_1^r$ by

$$h_{n_r} \mathrel{\bar{\bar{\succ}}}_1^r g \Rightarrow \bar{\boldsymbol{b}}_1^r(g) = v_1^{r+1}(g) \quad \text{and} \quad h_{n_r} \not\mathrel{\bar{\bar{\succ}}}_1^r g \Rightarrow \bar{\boldsymbol{b}}_1^r(g) = \bar{\bar{\boldsymbol{b}}}_1^r(g) \cdot (v_1^{r+1}(\lambda_r) + \delta/2)/\bar{\bar{\boldsymbol{b}}}_1^r(h_{n_r})\,, \quad (1)$$

where $\delta = \min_{g,h\in M, g\neq h} |v_1^{r+1}(g) - v_1^{r+1}(h)| > 0$. It is easy to check that Round-Robin$(\bar{\boldsymbol{b}}_1^r, \mathbf{b}_{-1})$ returns $(A_1, \ldots, A_n)$. Indeed, the preference ranking $\bar{\succ}_1^r$ induced by $\bar{\boldsymbol{b}}_1^r$ is identical to $\bar{\bar{\succ}}_1^r$ up to good h_{n_r}, so agent 1 receives $h_{n_1}, \ldots, h_{n_r}$ in the first r rounds, while any bid that was higher than $v_1^{r+1}(h_{n_{r+1}})$ and has been updated to its v_1^{r+1} value is not available at the beginning of round $r+1$ anyway. The latter is true, because otherwise such a good would have been chosen by Round-Robin$(\bar{\boldsymbol{b}}_1^r, \mathbf{b}_{-1})$ and Round-Robin$(\boldsymbol{b}_1^r, \mathbf{b}_{-1})$ instead of $h_{n_{r+1}}$.

Having $\bar{\boldsymbol{b}}_1^r$ as a point of reference, we now take a closer look to what happens if, starting at round r, agent 1 would receive her goods according to v_1^{r+1}. Note that this would be the same as just changing $\bar{\boldsymbol{b}}_1^r(h_{n_r})$ to $v_1^{r+1}(h_{n_r})$. We call this new auxiliary bid $\hat{\boldsymbol{b}}_1^r$; note that this is the first time we introduce a bid that *does not* preserve the original allocation. Similarly to our definition of the λ_is, we consider the execution of Round-Robin$(\hat{\boldsymbol{b}}_1^r, \mathbf{b}_{-1})$ and define $\hat{\lambda}_i$ to be the most valuable available good with respect to v_1^{r+1} at the very beginning of round i, for $i \geq r$. While we know that $\hat{\lambda}_r = \lambda_r$, in general we have no reason to expect that $\hat{\lambda}_r$ and λ_r are the same. Actually, the fact that $\boldsymbol{b}_1$—and thus $\bar{\bar{\boldsymbol{b}}}_1^r$—is a best response, combined with $h_{n_r} \neq \lambda_r$, imply that

$$v_1^{r+1}(h_{n_r}) + \sum_{i=r+1}^{k} v_1^{r+1}(\lambda_i) > \sum_{i=r}^{k} v_1^{r+1}(\hat{\lambda}_i)\,. \quad (2)$$

Coming back to the challenge of moving value from $v_1^{r+1}(\lambda_{r+1}), \ldots, v_1^{r+1}(\lambda_k)$ to $v_1^{r+1}(h_{n_r})$ (and equally so from $\hat{\boldsymbol{b}}_1^{r+1}(\lambda_{r+1}), \ldots, \hat{\boldsymbol{b}}_1^{r+1}(\lambda_k)$ to $\hat{\boldsymbol{b}}_1^{r+1}(h_{n_r})$), we want to make sure that enough value can be moved to eventually get $v_1^{r+1}(h_{n_r})$ slightly above $v_1^{r+1}(\lambda_r)$ while each h_{n_i} maintains more value than any other available good in round $i > r$. Notice that the preference rankings $\bar{\succ}_1^r$ and $\hat{\succ}_1^r$, induced by $\bar{\boldsymbol{b}}_1^r$ and $\hat{\boldsymbol{b}}_1^r$ respectively, are within by a partial slide from each other! Indeed, h_{n_r} is moved to a worst position in $\hat{\succ}_1^r$ compared to $\bar{\succ}_1^r$, but otherwise the two preference rankings are the same. Besides the easy observation that Round-Robin$(\hat{\boldsymbol{b}}_1^r, \mathbf{b}_{-1})$ and Round-Robin$(\bar{\boldsymbol{b}}_1^r, \mathbf{b}_{-1})$ run identically for $i-1$ rounds, this also means that Lemma 3.6 applies. That is, we have that in each round i of Round-Robin$(\hat{\boldsymbol{b}}_1^r, \mathbf{b}_{-1})$, for $i \geq r$, there is at most one good that is unavailable despite being available in round i of Round-Robin$(\bar{\boldsymbol{b}}_1^r, \mathbf{b}_{-1})$. In particular, in round i of Round-Robin$(\hat{\boldsymbol{b}}_1^r, \mathbf{b}_{-1})$, for $i \geq r$, at least two goods from $\{h_{n_i}, h_{n_{i+1}}, \ell_i\}$ are available, where conventionally we define $h_{n_{k+1}}$ to be the second best available good at the beginning of round k in Round-Robin$(\bar{\boldsymbol{b}}_1^r, \mathbf{b}_{-1})$; recall that ℓ_i is the most valuable good with respect to v_1^{r+1} that is allocated to some agent other than 1 during round i. By the definition of the $\hat{\lambda}_i$s, this observation implies

$$v_1^{r+1}(\hat{\lambda}_i) \geq \max\left\{v_1^{r+1}(h_{n_{i+1}}), v_1^{r+1}(\ell_i)\right\}, \text{ for all } r < i \leq k\,.$$

Given that $\bar{\boldsymbol{b}}_1^r$ trivially remains truthful from round $r+1$ with respect to v_1^{r+1}, the above maximum captures what is the mechanism's view of the second best good for agent 1 in round i *after* λ_i, for $i > r$, when the input is $(\bar{\boldsymbol{b}}_1^r, \mathbf{b}_{-1})$. Plugging this bound into inequality (2) along with $\hat{\lambda}_r = \lambda_r$, and rearranging the terms of round r, it yields

$$\sum_{i=r+1}^{k} \left(v_1^{r+1}(\lambda_i) - \max\left\{v_1^{r+1}(h_{n_{i+1}}), v_1^{r+1}(\ell_i)\right\}\right) = v_1^{r+1}(\lambda_r) - v_1^{r+1}(h_{n_r}) + \varepsilon\,,$$

for some $\varepsilon > 0$. We also define $\varepsilon_i = v_1^{r+1}(\lambda_i) - \max\left\{v_1^{r+1}(h_{n_{i+1}}), v_1^{r+1}(\ell_i)\right\}$ for $i > r$.

We can now define v_1^r by appropriately changing some of the values of v_1^{r+1}:

- For each $r+1 \leq i \leq k$, we set $v_1^r(h_{n_i}) = \max\left\{v_1^{r+1}(h_{n_{i+1}}), v_1^{r+1}(\ell_i)\right\} + \alpha \cdot \varepsilon_i$.
- We also set $v_1^r(h_{n_r}) = v_1^{r+1}(\lambda_r) + \varepsilon - \alpha \sum_{i=r+1}^{k} \varepsilon_i$, where α is chosen so that $\varepsilon - \alpha \sum_{i=r+1}^{k} \varepsilon_i = \min\{\varepsilon, \delta\}/3$; recall that $\delta = \min_{g,h \in M, g \neq h} |v_1^{r+1}(g) - v_1^{r+1}(h)|$. (Choosing such an α is always possible as $f(\alpha) = \varepsilon - \alpha \sum_{i=r+1}^{k} \varepsilon_i$ with $\alpha \in (0,1)$ is a continuous function with values in the interval $(v_1^{r+1}(h_{n_r}) - v_1^{r+1}(\lambda_r), \varepsilon)$.)
- For any other good g, we set $v_1^r(g) = v_1^{r+1}(g)$.

We also define $\boldsymbol{b}_1^r$ from $\bar{\boldsymbol{b}}_1^r$ by replacing the bids with the corresponding v_1^r values for all goods that are no better than h_{n_r} in $\bar{\succ}_1^r$, i.e, we have

$$g \not\bar{\succ}_1^r h_{n_r} \Rightarrow \boldsymbol{b}_1^r(g) = v_1^r(g) \quad \text{and} \quad g \,\bar{\succ}_1^r\, h_{n_r} \Rightarrow \boldsymbol{b}_1^r(g) = \bar{\boldsymbol{b}}_1^r(g)\,.$$

By this point, it should be straightforward to verify properties (i)-(iv). For (v), notice that any α giving $\varepsilon - \alpha \sum_{i=r+1}^{k} \varepsilon_i \in (0, \min\{\varepsilon, \delta\}/2)$ would work for consistently defining $\boldsymbol{b}_1^r$. If the above definition of v_1^r happens to give one of the finitely many values already in the range of v_1^{r+1}, then we may change α slightly to make all of the newly introduced values unique. □

4 Towards EFX Equilibria: The Case of Two Agents

As we saw, Round-Robin has PNE for every instance, and the corresponding allocations are always EF1. The natural next question is *can we have a similar guarantee for a stronger fairness notion?* In particular, we want to explore whether an analogous result is possible when we consider envy-freeness up to *any* good. When the agents are not strategic, it is known that EFX allocations exist when we have at most 3 agents [20, 21]. It should be noted that for the case of 3 agents no polynomial time algorithm is known, and it is unclear whether the constructive procedure of Chaudhury et al. [21] has any PNE. For $n \geq 4$, the existence of EFX allocations remains a major open problem. Therefore, we turn our attention to the case of two agents.

4.1 A Mechanism with EFX Nash Equilibria

A polynomial-time algorithm that outputs EFX allocations when we have two agents is given by Plaut and Roughgarden [42]. This is a *cut-and-choose* algorithm where the cut (lines 3–5) is produced using a variant of the *envy-cycle-elimination* algorithm of Lipton et al. [37] on two copies of agent 1, and then agent 2 "chooses" the best bundle among the two (line 6). We state it as mechanism Mod-Cut&Choose below (recall the notation $\boldsymbol{b}_i(S)$ for $\sum_{h \in S} \boldsymbol{b}_i(h)$). We should point out that this mechanism is not truthful, since there is no truthful mechanism for two agents that produces EF1 (or EFX for that matter) allocations for more than four goods [2]. Interestingly, we show that although not truthful, Mod-Cut&Choose always has at least one PNE for any instance, while all its equilibria are MMS and, by Theorem 2.5, EFX.

Mechanism 2 Mod-Cut&Choose($\boldsymbol{b}_1, \boldsymbol{b}_2$) // For $i \in \{1, 2\}$, $\boldsymbol{b}_i = (b_{i1}, \dots, b_{im})$ is the bid of agent i.

1: $(E_1, E_2) = (\emptyset, \emptyset)$
2: $(h_1, h_2, \dots, h_m)$ is M sorted in decreasing order with respect to v_1 // Break ties lexicographically.
3: **for** $i = 1, \dots, m$ **do**
4: $\quad j = \arg\min_{k \in [2]} \boldsymbol{b}_1(E_k)$ // Identify the worst bundle according to $\boldsymbol{b}_1$; break ties in favor of E_1.
5: $\quad E_j = E_j \cup \{h_i\}$ // Add the next good to that bundle.
6: $\ell = \arg\max_{k \in [2]} \boldsymbol{b}_2(E_k)$ // Identify the best bundle according to $\boldsymbol{b}_2$; break ties in favor of E_1.
7: **return** $\mathcal{A} = (M \setminus E_\ell, E_\ell)$ // Give this bundle to agent 2 and the remaining bundle to agent 1.

By Theorem 2.6, there is no reason to expect that the mechanism would guarantee more than $2\boldsymbol{\mu}_i/3$ to each agent. Indeed, seen as an algorithm, it does not always produce MMS allocations unless $\mathrm{P} = \mathrm{NP}$! To see this, first notice that when the agents are identical, that would mean that we can run the algorithm to exactly find their maximin share. This is equivalent to having an oracle for the classic PARTITION problem. As the algorithm's running time is polynomial, that would imply that $\mathrm{P} = \mathrm{NP}$. The same simple argument shows that it is NP-hard to compute a best response bid vector for agent 1.

Proposition 4.1. *Computing a best response $\boldsymbol{b}_1$ of agent 1 for Mod-Cut&Choose, given $\boldsymbol{b}_2$, is* NP*-hard.*

We begin with the following lemma on the "cut" part of Mod-Cut&Choose, stating that agent 1 may create any desirable partition of the goods (up to the ordering of the two sets). This is a necessary component of the proof of the main result of this section.

Lemma 4.2. *Let (X_1, X_2) be a partition of M. Agent 1, by bidding accordingly, can force Mod-Cut&Choose to construct E_1, E_2 in lines 3–5, such that $\{E_1, E_2\} = \{X_1, X_2\}$.*

Proof. We consider different cases depending on the cardinality of the sets X_1, X_2. Each case describes a bid that agent 1 can report in order to create the desired partition (X_1, X_2) or its permutation (X_2, X_1). Note that only the first case is relevant when $m = 1$, and only the first two cases are relevant when $2 \leq m \leq 3$.

Case 1 (one set has all the goods). Agent 1 declares zero value for all the goods. According to these values, j in line 4 is always 1, so every good goes to E_1, and we have the desired partition.

Case 2 (one set has $m-1$ goods). Agent 1 declares value 1 for the good that is contained in the set with cardinality 1, and for every good that is contained in the set with cardinality $m-1$ she declares a value equal to $\frac{1}{m-1}$. The first good is added in E_1, so E_2 is going to get chosen next. Actually, according to these values, E_2 must get all the remaining goods. Thus, the desired partition is produced.

Case 3 (the two sets have cardinalities $k \geq 2$ and $m - k \geq 2$). Agent 1 declares a value of 1 for one of the goods that are contained in the set with cardinality k. For every good that is contained in the set with cardinality $m - k$ she declares a value equal to $\frac{1+\varepsilon}{m-k}$, where $0 < \varepsilon < \frac{1}{m-k-1}$. Finally, for the rest of the goods that are contained in the set with cardinality k she declares a value of $\frac{\varepsilon}{k-1}$. E_1 gets the first good, so it appears to be more valuable than E_2. According to these values, E_2 ceases to appear to be the worst of the two when it gets every good of the set with cardinality $m - k$. This is the point where E_1 becomes worse than E_2, and continues to be worse until it contains every good of the set of cardinality k. Thus, again, the desired partition is produced. □

In particular, agent 1 can force the mechanism to construct E_1, E_2, such that $\min\{v_1(E_1), v_1(E_2)\} = \boldsymbol{\mu}_1$. Such a pair (E_1, E_2) is called a $\boldsymbol{\mu}_1$*-partition*. At least one $\boldsymbol{\mu}_1$-partition exists, by the definition of $\boldsymbol{\mu}_1$.

Corollary 4.3. *Agent 1 can force Mod-Cut&Choose to construct a $\boldsymbol{\mu}_1$-partition in lines 3–5.*

We can now proceed to the main theorem of this section on the existence and fairness properties of the PNE of Mod-Cut&Choose.

Theorem 4.4. *For any instance $\mathcal{I} = (\{1,2\}, M, \mathbf{v})$, the Mod-Cut&Choose mechanism has at least one PNE. Moreover, every PNE of the mechanism is MMS and EFX with respect to the valuation functions v_1, v_2.*

Proof. Given a partition $\mathcal{X} = (X_1, X_2)$ we are going to slightly abuse the notation—as we do in our pseudocode—and consider $\arg\min_{X \in \mathcal{X}} v_2(X)$ to be a single set in $\mathcal{X}$ rather than a subset of $\{X_1, X_2\}$. To do so, we assume that ties are broken in favor of the highest indexed set (here X_2) and tie-breaking is applied by the $\arg\min$ operator.

We will define a profile $(\boldsymbol{b}_1, \boldsymbol{b}_2)$ and show that it is a PNE. First, let $\boldsymbol{b}_2 = (v_2(g_1), v_2(g_2), \ldots, v_2(g_m))$ be the truthful bid of agent 2. Next $\boldsymbol{b}_1$ is the bid vector (as defined within the proof of Lemma 4.2) that results in Mod-Cut&Choose constructing a partition in

$$\arg\max_{\mathcal{X} \in \Pi_2(M)} v_1\big(\arg\max_{X \in \mathcal{X}} v_2(X)\big)\,.$$

To see that there exists such $\boldsymbol{b}_1$, notice that the set $\Pi_2(M)$ of all possible partitions is finite, and , by Lemma 4.2, every possible partition can be produced by Mod-Cut&Choose given the appropriate bid vector of agent 1. So, agent 1 forces the partition that maximizes, according to v_1, the value of the least desirable bundle according to v_2. Now it is easy to see that given the bidding strategy of agent 2, i.e., playing truthfully, there is no deviation for agent 1 that is profitable (by definition). Moreover, agent 2 gets the best of the two bundles according to her valuation function (regardless of the partition, truth telling is a dominant strategy for her), thus there is no profitable deviation for her either. Therefore, $(\boldsymbol{b}_1, \boldsymbol{b}_2)$ is a PNE for $\mathcal{I}$.

Regarding the second part of the statement, suppose for a contradiction that there is a PNE $\mathbf{b}$, where an agent i does not achieve her $\boldsymbol{\mu}_i$ in the allocation returned by Mod-Cut&Choose($\mathbf{b}$). If this agent is agent 1, then according to Corollary 4.3, there is a bid vector $\boldsymbol{b}_1'$ she can report, so that the algorithm will produce a $\boldsymbol{\mu}_1$-partition. By deviating to $\boldsymbol{b}_1'$, regardless of the set given to agent 2, agent 1 will end up with a bundle she values at least $\boldsymbol{\mu}_1$. As this would be a strict improvement over what she currently gets, it would contradict the fact that $\mathbf{b}$ is a PNE. So, it must be the case where agent 2 gets a bundle she values strictly less than $\boldsymbol{\mu}_2$. Notice that, regardless of the partition which Mod-Cut&Choose to constructs in lines 3–5, by declaring her truthful bid, agent 2 gets a bundle of value at least $v_2(M)/2$. By Definition 2.3, it is immediate to see that this value is at least $\boldsymbol{\mu}_2$, i.e., deviating to her truthful bid is a strict improvement over what she currently gets by Mod-Cut&Choose($\mathbf{b}$), which is a contradiction.

It remains to show that the allocation returned by Mod-Cut&Choose($\mathbf{b}$) is also EFX. However, since here $n = 2$, this directly follows from Theorem 2.5. □

4.2 The Enhanced Fairness of EFX Nash Equilibria

As it was discussed before Proposition 4.1, it is surprising that the EFX equilibria of Mod-Cut&Choose impose stronger fairness guarantees compared to generic EFX allocations or even EFX allocations produced by Mod-Cut&Choose itself in the non-strategic setting. In this section we explore whether something similar holds for *every* mechanism with EFX equilibria. Specifically, we consider the (obviously non-empty) class of mechanisms that have PNE for every instance and these equilibria always lead to EFX allocations. Our goal is to determine if these allocations have better fairness guarantees (with respect to the underlying true valuation functions) than EFX allocations in general. To this end, we start by examining instances of two agents and 4 goods and we prove that for every mechanism of this class, all allocations at a PNE are MMS allocations. The reason we start from this restricted set of instances is that it already provides a clear separation with the non-strategic setting. Recall from Theorem 2.6 that there are instances with just 4 goods where an EFX allocation may not be a $\left(\frac{2}{3}+\delta\right)$-MMS allocation, for any $\delta > 0$.

We begin by showing the following lemma which regards some very simple cases of such instances, and then we proceed to the proof of the statement.

Lemma 4.5. *Consider an instance with 2 agents and 4 goods. If agent $i \in [2]$ has positive value for three or less goods, then in every allocation which is EFX from her point of view, agent i has value at least $\boldsymbol{\mu}_i$.*

Proof. Suppose agent i has positive value for at most three goods. The statement is trivial when there is *at most one* positively valued good as in this case $\boldsymbol{\mu}_i = 0$ and agent i always gets $\boldsymbol{\mu}_i$ no matter the bundle that she gets. When she has a positive value for *two* goods, in order to consider the allocation as EFX she must get at least one of them. In this case she also achieves her $\boldsymbol{\mu}_i$ as it is equal to the smaller of the two positive values. Finally, suppose agent i has positive value for *three* goods. Notice that $\boldsymbol{\mu}_i$ in this case is either equal to the largest of the three values or to the sum of the two smallest values; whichever is smaller. So, if agent i gets two goods, then she always derives a value of at least gets $\boldsymbol{\mu}_i$. If she gets just one good, then this good must have the highest value, otherwise the she would not consider the allocation as EFX. So, in this case too, she gets value at least $\boldsymbol{\mu}_i$. ☐

We are now ready for the general result.

Theorem 4.6. *Let $\mathcal{M}$ be a mechanism that has PNE for any instance $(\{1, 2\}, M, (v_1, v_2))$ with $|M| = 4$, and all these equilibria lead to EFX allocations with respect to v_1, v_2. Then each such EFX allocation is also an MMS allocation.*

Proof. Suppose for contradiction that this is not the case. This means that there exists a valuation instance $\mathbf{v} = (v_1, v_2)$, for which there is a PNE $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ that produces an EFX allocation (A_1, A_2), where, without loss of generality, $v_1(A_1) < \boldsymbol{\mu}_1$. Rename, if necessary, the goods to $\{h_1, h_2, h_3, h_4\}$, so that $v_1(h_1) \geq v_1(h_2) \geq v_1(h_3) \geq v_1(h_4)$. Finally, notice that from Lemma 4.5 we know that $v_1(h_j) > 0$ for any $j \in \{1, 2, 3, 4\}$. Before we proceed, we let (B_1, B_2) to be a $\boldsymbol{\mu}_1$-partition of agent 1.

Case 1: $|A_1| = 1$. Since (A_1, A_2) is EFX and all the goods have positive value for agent 1 according to v_1, it is easy to see that $\{h_1\} = A_1$. Moreover, the value of h_1 is at least as much as the value of any set of cardinality 2 that contains goods from $\{h_2, h_3, h_4\}$. Since $v_1(h_1) < \boldsymbol{\mu}_1$ and h_1 is the highest valued good, we get $|B_1| = |B_2| = 2$. This implies that good h_1 needs to be placed with another good to reach the MMS value. However, this means that either B_1 or B_2 must contain two goods from $\{h_2, h_3, h_4\}$ and provides a value of at least $\boldsymbol{\mu}_1$. This is not possible as h_1 has at least as much value as any set of cardinality 2 that contains goods from $\{h_2, h_3, h_4\}$ and $v_1(h_1) < \boldsymbol{\mu}_1$. Thus, if $|A_1| = 1$, then $v_1(A_1) \geq \boldsymbol{\mu}_1$.

Case 2: $|A_1| = 3$ or $|A_1| = 4$. In both cases it is easy to see that agent 1 achieves a value of at least $\boldsymbol{\mu}_1$.

Case 3: $|A_1| = 2$, where $A_1 = \{h_1, h_2\}$, or $A_1 = \{h_1, h_3\}$. Notice that these bundles achieve at least the proportional value of agent 1, so in this case she gets a value of at least $\boldsymbol{\mu}_1$.

Case 4: $|A_1| = 2$, where $A_1 = \{h_1, h_4\}$. Initially notice that $|B_1| \neq 1$ and $|B_1| \neq 3$, as the opposite would imply that $v_1(g_1) \geq \boldsymbol{\mu}_i \Rightarrow v_1(A_1) > \boldsymbol{\mu}_1$, a contradiction. Thus, we get that $|B_1| = |B_2| = 2$. Since by assumption $v_1(\{h_1, h_4\}) < \boldsymbol{\mu}_1$, we have that good h_1 must be placed with either good h_2, or good h_3 in a $\boldsymbol{\mu}_1$-partition of agent 1. So, a $\boldsymbol{\mu}_1$-partition can be either $(\{h_1, h_2\}, \{h_3, h_4\})$, or $(\{h_1, h_3\}, \{h_2, h_4\})$. However, $v_1(\{h_3, h_4\}) \leq v_1(\{h_1, h_4\}) < \boldsymbol{\mu}_1$ and $v_1(\{h_2, h_4\}) \leq v_1(\{h_1, h_4\}) < \boldsymbol{\mu}_1$. A contradiction.

Case 5: $|A_1| = 2$, where $A_1 = \{h_2, h_3\}$. Initially notice that since (A_1, A_2) is an EFX allocation, we have that $v_1(A_1) \geq v_1(h_1)$. This implies that in a $\boldsymbol{\mu}_1$-partition of agent 1., one of the sets has cardinality 1, and the maximum possible value that this set can have is $v_1(h_1) \leq v_1(A_1) < \boldsymbol{\mu}_1$, a contradiction. Thus, we get that $|B_1| = |B_2| = 2$. Since by assumption $v_1(\{h_2, h_3\}) < \boldsymbol{\mu}_1$, we have that good h_2 must be placed with either good h_1 in a $\boldsymbol{\mu}_1$-partition of agent 1. So, we conclude that a $\boldsymbol{\mu}_1$-partition is $(\{h_1, h_2\}, \{h_3, h_4\})$. However, $v_1(\{h_3, h_4\}) \leq v_1(\{h_2, h_3\}) < \boldsymbol{\mu}_1$. A contradiction.

Case 6: $|A_1| = 2$, where $A_1 = \{h_2, h_4\}$, or $A_1 = \{h_3, h_4\}$. Notice that so far we have not use the fact that $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ is a PNE under the valuation instance $\mathbf{v} = (v_1, v_2)$. However, we need this property for these two last cases. Consider a different valuation profile $\mathbf{v}^* = (v^*, v^*)$, where the agents have identical values over the goods. The valuation instance is defined as follows:

$$v^*(h_j) = \begin{cases} 1.2 & j = 1 \\ 1 & j \in \{2, 3\} \\ \varepsilon & j = 4 \end{cases}$$

where $0 < \varepsilon < 0.2$. It is easy to see that for this valuation instance there are only two EFX allocations, i.e., $(\{h_1, h_4\}, \{h_2, h_3\})$ and its symmetric $(\{h_2, h_3\}, \{h_1, h_4\})$. According to our assumption, there must be a bid vector $\mathbf{b}^* = (\boldsymbol{b}_1^*, \boldsymbol{b}_2^*)$ that is a PNE for this valuation instance, and since we require the PNE to be also EFX, it must produce one of these allocations as they are the only ones that are EFX. Moreover, observe that the value that agent 2 derives in these allocations is 2 and $1.2 + \varepsilon < 1.4$ respectively.

For now assume that $\boldsymbol{b}_1 \neq \boldsymbol{b}_1^*$ and $\boldsymbol{b}_2 \neq \boldsymbol{b}_2^*$. Recall that $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ is by assumption a PNE under valuation instance $\mathbf{v} = (v_1, v_2)$, and in addition, $v_1(h_j) > 0$ for every $j \in \{1, 2, 3, 4\}$. Let us now examine what each agent can get if agent 1 deviates to $\mathbf{b}' = (\boldsymbol{b}_1^*, \boldsymbol{b}_2)$:

- In case the bundle of agent 1 contains just one good, then agent 2 gets a bundle of cardinality 3. This contradicts the fact that $\mathbf{b}^* = (\boldsymbol{b}_1^*, \boldsymbol{b}_2^*)$ is a PNE under valuation instance $\mathbf{v}^* = (v^*, v^*)$, as any such set gives agent 2 a value that is strictly higher than 2.

- In case the bundle of agent 1 has cardinality 3, this contradicts the fact that $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ is a PNE under valuation instance $\mathbf{v} = (v_1, v_2)$, as the least valuable such set is $\{h_2, h_3, h_4\}$ and it has strictly higher value than $v_1(A_1)$, since $v_1(h_j) > 0$ for every $j \in \{1, 2, 3, 4\}$.

- In case the bundle of agent 1 is one of the following:$\{h_1, h_2\}$, $\{h_1, h_3\}$, $\{h_1, h_4\}$, $\{h_2, h_3\}$, then this implies that A_1 has a value equal to the value of one of these bundles. By using the same arguments as in Cases 3, 4, 5, we can show that $v_1(A_1) \geq \boldsymbol{\mu}_1$, a contradiction.

- In case the bundle of agent 1 is one of the following: $\{h_2, h_4\}$, $\{h_3, h_4\}$, then agent 2 gets either $\{h_1, h_2\}$ or $\{h_1, h_3\}$. This contradicts the fact that $\mathbf{b}^* = (\boldsymbol{b}_1^*, \boldsymbol{b}_2^*)$ is a PNE under valuation instance $\mathbf{v}^* = (v^*, v^*)$, as any such set gives agent 2 a value that is strictly higher than 2.

The remaining corner cases are straightforward to deal with. To begin with, it is not possible to have $\boldsymbol{b}_1 = \boldsymbol{b}_1^*$ and $\boldsymbol{b}_2 = \boldsymbol{b}_2^*$, as the produced allocations from these bid profiles are different.

Next, assume that $\boldsymbol{b}_1 = \boldsymbol{b}_1^*$ and $\boldsymbol{b}_2 \neq \boldsymbol{b}_2^*$. This directly contradicts the fact that $(\boldsymbol{b}_1^*, \boldsymbol{b}_2^*)$ is a PNE under $\mathbf{v}^* = (v^*, v^*)$. To see this, start from $(\boldsymbol{b}_1^*, \boldsymbol{b}_2^*)$ and let agent 2 to deviate to $\boldsymbol{b}_2$. She gets set A_2 which is either $\{h_1, h_2\}$ or $\{h_1, h_3\}$ and achieves a value $2.2 > 2$.

Finally, assume that $\boldsymbol{b}_1 \neq \boldsymbol{b}_1^*$ and $\boldsymbol{b}_2 = \boldsymbol{b}_2^*$. Start from $(\boldsymbol{b}_1, \boldsymbol{b}_2)$ and let agent 1 to deviate to $\boldsymbol{b}_1^*$. She gets either set $\{h_1, h_4\}$, or set $\{h_2, h_3\}$, both of which provide her with a value of at least $v_1(A_1)$. In case this value is strictly higher than $v_1(A_1)$, then this directly contradicts the fact that $(\boldsymbol{b}_1, \boldsymbol{b}_2)$ is a PNE under $\mathbf{v} = (v_1, v_2)$. Finally, if this value is equal to $v_1(A_1)$, by using the same arguments as in cases 4, 5, we can show that $v_1(A_1) \geq \boldsymbol{\mu}_1$, a contradiction.

Since every possible case leads to a contradiction, we conclude that every allocation that corresponds to a PNE of a mechanism in the class of interest, guarantees to each agent her maximin share. □

The proof of Theorem 4.6 relies on extensive case analysis, where in each case assuming that the allocation is EFX but not MMS eventually contradicts the fact that the current profile is a PNE. When we consider instances with 5 or more goods, this approach is not fruitful anymore. The reason for that is not solely the increased number of cases one has to handle, but rather the fact that now some of the cases do not seem to lead to a contradiction at all.

Although we suspect that the theorem is no longer true for more than 4 goods, we are able prove a somewhat weaker property that still separates the EFX allocations in PNE from generic EFX allocations in the non-strategic setting. In particular, for general mechanisms that have PNE for every instance and these equilibria are always EFX, we show that the corresponding allocations always guarantee an approximation to MMS that is *strictly better* than $2/3$.

Theorem 4.7. *Let $\mathcal{M}$ be a mechanism that has PNE for any instance $(\{1,2\}, M, (v_1, v_2))$, and all these equilibria lead to EFX allocations with respect to v_1, v_2. Then each such EFX allocation is also an α-MMS allocation for some $\alpha > 2/3$.*

Proof. Suppose for a contradiction that this is not the case. This means that there exists such a mechanism $\mathcal{M}$ and an instance $(\{1,2\}, M, (v_1, v_2))$, for which there is a PNE $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ that results in an EFX allocation $\mathcal{A} = (A_1, A_2)$, where $v_i(A_i) \leq 2\boldsymbol{\mu}_i/3$ for at least one $i \in [2]$. Without loss of generality, assume $v_1(A_1) \leq 2\boldsymbol{\mu}_1/3$ and notice that this means that $v_1(A_1) = 2\boldsymbol{\mu}_i/3$, as $v_1(A_1)$ cannot be smaller than $2\boldsymbol{\mu}_i/3$, by Theorem 2.6. This implies that $v_1(A_2) \geq 4\boldsymbol{\mu}_i/3$, since $v_1(M) \geq 2\boldsymbol{\mu}_1$ by Definition 2.3.

Initially, we will restrict the number of the goods with positive value (according to v_1) in A_2. Let $S \subseteq A_2$ be the set of such goods, i.e, $S = \{g \in A_2 \,|\, v_1(g) > 0\}$. Let $|S| = k$ and notice that k cannot be 0 or 1 since otherwise $v_1(A_1) \geq \boldsymbol{\mu}_1$. Finally, let $x \in \arg\min_{g \in S} v_1(g)$ be a minimum valued good for agent 1 in S. We have

$$\frac{2}{3}\boldsymbol{\mu}_1 = v_1(A_1) \geq v_1(S \setminus \{x\}) \geq v_1(S) - \frac{v_1(S)}{k} = \frac{(k-1)}{k} v_1(A_2) \geq \frac{(k-1)}{k}\frac{4}{3}\boldsymbol{\mu}_1 \,,$$

where the first inequality follows from (A_1, A_2) being EFX. Given our observation that $k \leq 2$, the above implies that $k = 2$. Name h_1 and h_2 the goods of S, and observe that if $v_1(A_2) = v_1(\{h_1, h_2\}) > 4\boldsymbol{\mu}_1/3$, then (A_1, A_2) cannot be EFX from the perspective of agent 1. Thus, we get that $v_1(A_2) = 4\boldsymbol{\mu}_1/3$, which in conjunction with EFX implies $v_1(h_1) = v_1(h_2) = 2\boldsymbol{\mu}_1/3$.

Next we argue that A_1 contains at least 2 goods that have positive value for agent 1. Indeed, if all the goods in A_1 had zero value, then we would have $v_1(A_1) = 0 < 2\boldsymbol{\mu}_1/3$ as A_2 contains two positively valued goods, while if there was just one positively valued good in A_1, this would imply that only three goods have positive value for agent 1, and each one of them has value $2\boldsymbol{\mu}_1/3$. The latter would make the existence of a $\boldsymbol{\mu}_1$-partition impossible, which is a contradiction. So, since there are at least two positively valued goods in A_1 for agent 1, we arbitrarily choose two of them, and we name them h_3 and h_4. We arbitrarily name the remaining goods $h_5, h_6, \ldots, h_m$.

Consider now a different valuation instance $\mathbf{v}^* = (v^*, v^*)$ where the agents have identical values over the goods. The valuation function is defined as

$$v^*(h_j) = \begin{cases} 1.2 & j = 1 \\ 1 & j \in \{2,3\} \\ \varepsilon & j \in \{4,\dots,m\} \end{cases}$$

where $\varepsilon > 0$ and $(m-3)\cdot\varepsilon < 0.2$. It is easy to see that for this valuation instance there are only two EFX allocations, namely, $\mathcal{X} = (\{h_1, h_4, \dots, h_m\}, \{h_2, h_3\})$, and its symmetric $\mathcal{Y} = (\{h_2, h_3\}, \{h_1, h_4, \dots, h_m\})$. According to our assumption, there must be a bidding vector $\mathbf{b}^* = (\boldsymbol{b}_1^*, \boldsymbol{b}_2^*)$ that is a PNE of $\mathcal{M}$ for the instance $(\{1,2\}, M, \mathbf{v}^*)$, and since all PNE of $\mathcal{M}$ are also EFX, $\mathcal{M}(\mathbf{b}^*)$ must output one of $\mathcal{X}$ and $\mathcal{Y}$. Moreover, observe that the value agent 2 receives (with respect to V^*) in these allocations is 2 and $1.2+(m-3)\varepsilon < 1.4$ respectively.

For now assume that $\boldsymbol{b}_1 \neq \boldsymbol{b}_1^*$ and $\boldsymbol{b}_2 \neq \boldsymbol{b}_2^*$. We will show that, in this case, running $\mathcal{M}$ with input $\mathbf{b}' = (\boldsymbol{b}_1^*, \boldsymbol{b}_2)$ results to agent 2 receiving a bundle of value strictly better than 2 according to v^*. This contradicts the fact that $\mathbf{b}^* = (\boldsymbol{b}_1^*, \boldsymbol{b}_2^*)$ is a PNE under $\mathbf{v}^* = (v^*, v^*)$. Recall that $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ is a PNE under $\mathbf{v} = (v_1, v_2)$, that $v_1(h_1) = v_1(h_2) = v_1(A_1) = 2\boldsymbol{\mu}_1/3$, and that $v_1(h_3), v_1(h_4)$ are strictly positive. So, let us examine what each agent may get if agent 1 deviates from $\mathbf{b}$ to $\mathbf{b}' = (\boldsymbol{b}_1^*, \boldsymbol{b}_2)$:

- In case the bundle of agent 1 contains good h_1, it cannot contain any good from $\{h_2, h_3, h_4\}$; otherwise $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ would not be a PNE under $\mathbf{v} = (v_1, v_2)$. Thus, $\{h_2, h_3, h_4\}$ is part of the bundle of agent 2.

- In case the bundle of agent 1 contains good h_2, it cannot contain any good from $\{h_1, h_3, h_4\}$; otherwise $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ would not be a PNE under $\mathbf{v} = (v_1, v_2)$. Thus, $\{h_1, h_3, h_4\}$ is part the bundle of agent 2.

- In case the bundle of agent 1 does not contain any of h_1 and h_2, then it is possible for her to get any subset $T \subseteq \{h_3, h_4, \dots, h_m\}$. However, $\{h_1, h_2\}$ is part the bundle of agent 2.

Thus, in the allocation returned by $\mathcal{M}(\mathbf{b}')$, agent 2 gets a bundle that contains $\{h_2, h_3, h_4\}$ or $\{h_1, h_3, h_4\}$ or $\{h_1, h_2\}$. Consider the value of these sets according to v^*:

$$v^*(\{h_2, h_3, h_4\}) = 2 + \varepsilon\,, \qquad v^*(\{h_1, h_3, h_4\}) = 2.2 + \varepsilon\,, \qquad v^*(\{h_1, h_2\}) = 2.2\,.$$

That is, in every single case the value agent 2 derives under $\mathbf{v}^* = (v^*, v^*)$ when the profile $\mathbf{b}' = (\boldsymbol{b}_1^*, \boldsymbol{b}_2)$ is played is strictly better than 2. However, 2 is the maximum possible value that agent 2 could derive under $\mathbf{v}^*$ when the profile $\mathbf{b}^*$ is played. This contradicts the fact that $\mathbf{b}^*$ is a PNE under $\mathbf{v}^*$, as $\boldsymbol{b}_2$ is a profitable deviation for agent 2.

The remaining corner cases are straightforward to deal with. To begin with, it is not possible to have $\boldsymbol{b}_1 = \boldsymbol{b}_1^*$ and $\boldsymbol{b}_2 = \boldsymbol{b}_2^*$, as $\mathcal{X} \neq \mathcal{A}$ and $\mathcal{Y} \neq \mathcal{A}$.

Next, assume that $\boldsymbol{b}_1 = \boldsymbol{b}_1^*$ and $\boldsymbol{b}_2 \neq \boldsymbol{b}_2^*$. This directly contradicts the fact that $\mathbf{b}^*$ is a PNE under $\mathbf{v}^* = (v^*, v^*)$. To see this, starting from $\mathbf{b}^*$ let agent 2 deviate to $\boldsymbol{b}_2$. She then gets A_2 which contains h_1, h_2 and has value for her $v^*(A_2) \geq 2.2 > 2$.

Finally, assume that $\boldsymbol{b}_1 \neq \boldsymbol{b}_1^*$ and $\boldsymbol{b}_2 = \boldsymbol{b}_2^*$. This directly contradicts the fact that $\mathbf{b}$ is a PNE under $\mathbf{v} = (v_1, v_2)$. To see this, starting from $\mathbf{b}$ let agent 1 deviate to $\boldsymbol{b}_1^*$. She either gets $\{h_1, h_4, \dots, h_m\}$ of value at least $v_1(h_1)+v_1(h_4) > 2\boldsymbol{\mu}_1/3 = v_1(A_1)$ or she gets $\{h_2, h_3\}$ of value $v_1(h_2)+v_1(h_3) > 2\boldsymbol{\mu}_1/3 = v_1(A_1)$.

Since every possible case leads to a contradiction, we conclude that every allocation that corresponds to a PNE of a mechanism in the class of interest, guarantees to each agent i value that is strictly better than $2\boldsymbol{\mu}_i/3$, for $i \in [2]$. ☐

5 Discussion

In this work we studied the problem of fair allocating a set of indivisible goods, to a set of strategic agents. Somewhat surprising—given the existing strong impossibilities for truthful mechanisms—our results are mostly positive. In particular, we showed that there exist mechanisms that have PNE for every instance, and at the same time the allocations that correspond to PNE have strong fairness guarantees with respect to the true valuation functions.

We believe that there are several interesting directions for future work that follow our research agenda. For instance, it would be interesting to explore how algorithms that compute EF1 allocations for richer valuation function domains (e.g., the Envy-Cycle-Elimination algorithm [37]) behave in the strategic setting we study in this work. Here the question is twofold. On one hand, it is unclear whether such algorithms have PNE for every valuation instance, while on the other, it would be important to determine if they maintain their fairness properties at their equilibria or not. The existence of PNE for algorithms that compute approximate MMS allocation is on a similar direction and, as we mentioned in Section 2, in this case we get the MMS guarantee on the equilibria for free.

Theorems 4.6 and 4.7 leave an open question on the MMS guarantee that the equilibria of mechanisms that always have PNE and these are EFX. Although we suspect that the corresponding allocations are not always MMS, such a result would immediately imply that for every such mechanism which runs in polynomial time, finding a best response of an agent is a computationally hard problem. Going beyond the case of two agents here seems to be a highly nontrivial problem as it is not very plausible that the current state of the art for the non-strategic setting could be analysed under incentives.

Finally, while we did not really focus on complexity questions, it is clear that computing best responses is generally hard. However, when they are not, for instance when the number of agents in Round-Robin is fixed [47], we would like to know if best response dynamics always converge to a PNE or there might be cyclic behavior (as it happens with better response dynamics [8]).

A Dealing With Ties Among the Values

We begin with a lemma that is used twice: once to show that Round-Robin has PNE for any instance (Theorem A.2), and then again in the complete proof of Theorem 3.3. Both proofs are presented in this appendix.

Lemma A.1. *For any fair division instance $\mathcal{I} = (N, M, \mathbf{v})$, then there exists a valuation function v_i' with the following properties:*

- *v_i' induces a strict preference ranking over M, which is consistent with the preference ranking induced by v_i;*
- *if a bid vector $\boldsymbol{b}_i$ is a best response of agent i with respect to v_i to the (fixed) bid vectors $\mathbf{b}_{-i}$ of all other players in Round-Robin, then $\boldsymbol{b}_i$ is still a best response to $\mathbf{b}_{-i}$ with respect to v_i';*
- *$v_i(T) \leq v_i'(T) \leq v_i(T) + \varepsilon/3$, for any $T \subseteq M$, where ε is the smallest positive difference between the values of two goods with respect to v_1, i.e., $\varepsilon = \min\{|v_i(g) - v_i(h)| \,:\, g, h \in M,\, v_i(g) \neq v_i(h)\}$, or, if there is no positive difference, $\varepsilon = 1$.*

Proof. If v_i already induces a strict preference ranking over the goods, then clearly $v_i' = v_i$ has all these properties. So, suppose that there are goods with exactly the same v_1 value and let $S = \{g \in M \,:\, \exists h \in M$ such that $h \neq g$ and $v_i(h) = v_i(g)\}$ be the set of all goods that do not have a unique v_1 value. Also, for $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_n)$ as in the second bullet of the statement, let $(A_1, A_2, \dots, A_n)$ be the allocation returned by Round-Robin($\mathbf{b}$). Then we define v_i' on $M = \{g_1, \dots, g_m\}$ as follows

$$v_i'(g_j) = \begin{cases} v_i(g_j) + \frac{j \cdot \varepsilon}{3m^2} & g_j \in S \cap A_i \\ v_i(g_j) + \frac{j \cdot \varepsilon}{6m^5} & g_j \in S \setminus A_i \\ v_i(g_j) & g_j \in M \setminus S \end{cases}$$

It is straightforward to verify that ties are broken without introducing any new ties and without violating the preference ranking induced by v_i. Also, the added quantities sum up to a value smaller than $\varepsilon/3$. So the first and third properties hold for v_i'. To see that $\boldsymbol{b}_i$ is still a best response to $\mathbf{b}_{-1}$ with respect to v_i', suppose for a contradiction that this is not the case. That is, there is some bid vector $\boldsymbol{b}_i'$, such that in the allocation $(A_1', A_2', \dots, A_n')$ returned by Round-Robin($\boldsymbol{b}_i', \mathbf{b}_{-i}$) we have $v_i'(A_i') > v_i'(A_i)$. The latter implies that $A_i' \neq A_i$. Given that $v_i(A_i') \leq v_i(A_i)$, we distinguish two cases.

First, suppose $v_i(A_i') < v_i(A_i)$. By the definition of ε we have $v_i(A_i') \leq v_i(A_i) - \varepsilon$. This, however, implies $v_i'(A_i') \leq v_i(A_i') + \varepsilon/3 \leq v_i(A_i) - 2\varepsilon/3 \leq v_i'(A_i) - 2\varepsilon/3$, which contradicts $v_i'(A_i') > v_i'(A_i)$.

So, it must be the case that $v_i(A_i') = v_i(A_i)$. Then the difference $v_i'(A_i') - v_i'(A_i)$ must be due to the small terms we added to the values of some goods. Note that if $A_i \not\subseteq A_i'$, then the value added to $v_i(A_i')$ is at most the total value added to $S \setminus A_i$ plus *almost* the total value added to $S \cap A_i$, as we should exclude at least $\frac{\varepsilon}{3m^2}$. But as $\sum_{j=1}^{m} \frac{j \cdot \varepsilon}{6m^5} < \frac{\varepsilon}{3m^2}$, we have that this difference $v_i'(A_i') - v_i'(A_i)$ must be negative, and this again contradicts $v_i'(A_i') > v_i'(A_i)$. So it must be the case where $A_i \subseteq A_i'$. However, independently of the bid profile, we know that agent i will receive *exactly* the same number of goods in the two executions of Round-Robin (i.e., either $\lceil m/n \rceil$ in both or $\lfloor m/n \rfloor$ in both). But then $A_i = A_i'$ again contradicting $v_i'(A_i') > v_i'(A_i)$.

We conclude that $\boldsymbol{b}_i$ is still a best response to $\mathbf{b}_{-1}$ with respect to v_i'. ☐

Recall that Aziz et al. [8] showed is that as long as all the valuation functions in an instance induce strict preference rankings and all the values are positive, then there is a way to construct PNE. In the terminology of [8] these are all the bid profiles that are consistent with the so-called *bluff profile* defined therein. Here we do not need to define what the bluff profile is explicitly. We are going to use the following result which essentially is a corollary of [8].

Theorem A.2 (Follows from [8]). *For any instance $\mathcal{I} = (N, M, \mathbf{v})$, where all goods have positive values for all agents and all the valuation functions induce strict preference rankings, Round-Robin has at least one PNE.*

Using Theorem A.2 and Lemma A.1, we will show that Round-Robin has PNE in every single instance with additive valuation functions.

Theorem A.3. *For any instance $\mathcal{I} = (N, M, \mathbf{v})$ Round-Robin has at least one PNE.*

Proof. For each one of $v_1, v_2, \dots, v_n$ we apply Lemma A.1 to get $\mathbf{v}' = (v'_1, v'_2, \dots, v'_n)$. When we apply it for v_i, let ε_i be the corresponding constant of the third bullet of the lemma; the second bullet of the lemma is irrelevant here. Clearly, for all $i \in N$, v'_i induces a strict preference ranking, so for Theorem A.2 to apply we only need that all values are positive. This may not always be the case. If v_i assigned value 0 to multiple goods, then all the 0 are taken care of during the definition of v'_i. If, however there was a single good g such that $v_i(g) = 0$, then $v'_i(g) = 0$ as well. We can resolve this by setting $v'_i(g) = \varepsilon_i/3$. This does not affect the induced preference ranking of v'_i, while the property of the third bullet of Lemma A.1 becomes $v_i(T) \leq v'_i(T) \leq v_i(T) + 2\varepsilon/3$ instead.

Now we may apply Theorem A.2. So, for the instance $\mathcal{I}' = (N, M, \mathbf{v}')$ Round-Robin has at least one PNE; suppose the profile $\mathbf{d} = (\boldsymbol{d}_1, \boldsymbol{d}_2, \dots, \boldsymbol{d}_n)$ is such a PNE and let $(A_1, A_2, \dots, A_m)$ be the allocation returned by Round-Robin($\mathbf{d}$). We claim that $\mathbf{d}$ is also an equilibrium of the original instance $\mathcal{I}$.

Suppose it is not, for a contradiction. This means that in $\mathcal{I}$ there is an agent, say agent k, who can deviate to a bid profile $\boldsymbol{b}_k$, so that the allocation returned by Round-Robin($\boldsymbol{b}_k, \mathbf{d}_{-k}$) is $(B_1, B_2, \dots, B_m)$ with $v_k(B_k) > v_k(A_k)$. By the definition of ε_k, we have $v_k(A_k) \leq v_k(B_k) - \varepsilon_k$. This implies $v'_k(A_k) \leq v_k(A_k) + 2\varepsilon_k/3 \leq v_k(B_k) - \varepsilon_k/3 \leq v'_k(B_k) - \varepsilon/3 < v'_k(B_k)$, which contradicts the fact that $\mathbf{d}$ is a PNE in $\mathcal{I}'$. Therefore, $\mathbf{d}$ is a PNE in $\mathcal{I}$ as well. ☐

Finally, we can present a complete proof of Theorem 3.3, without any assumptions on the valuation function of agent 1.

Complete Proof of Theorem 3.3. Consider an arbitrary instance $\mathcal{I} = (N, M, \mathbf{v})$ and assume that the input of Round-Robin is $\mathbf{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_n)$, where $\boldsymbol{b}_1$ is a best response of agent 1 to $\mathbf{b}_{-i} = (\boldsymbol{b}_2, \dots, \boldsymbol{b}_n)$ according to her valuation function v_1. Let $(A_1, \dots, A_n)$ be the output of Round-Robin($\mathbf{b}$). In order to apply Lemma 3.4, we need v_1 to induce a strict preference ranking over the goods.

Instead, we are going to use Lemma A.1 first to get a valuation function v'_1 having all the properties stated therein. Note that by the second bullet of Lemma A.1, $\boldsymbol{b}_1$ is still a best response of agent 1 to $\mathbf{b}_{-i}$ in the instance $\mathcal{I}' = (N, M, (v'_1, \mathbf{v}_{-1}))$. So, we apply Lemma 3.4 here. That is, we consider the hypothetical scenario implied by the lemma: keeping agents 2 through n fixed, suppose that the valuation function of agent 1 is the function v_1^* given by the lemma, and her bid $\boldsymbol{b}_1^*$ is the truthful bid for v_1^*. The first part of Lemma 3.4 guarantees that the output of Round-Robin($\boldsymbol{b}_1^*, \mathbf{b}_{-i}$) remains $(A_1, \dots, A_n)$.

As we have mentioned a couple of times already, no matter what others bid, if the agent with the highest priority (here agent 1 with valuation function v_1^*) reports her true values to Round-Robin, the resulting allocation is EF from her perspective (see the proof of Theorem 12.2 in [39]). In our hypothetical scenario this is the case for agent 1 and it translates into having $v_1^*(A_1) \geq v_1^*(A_i)$ for all $i \in N$. Then the second and third parts of the lemma imply that $v'_1(A_1) \geq v'_1(A_i)$ for all $i \in N$.

Suppose for a contradiction that there is a $j \in N$, such that $v_1(A_1) < v_1(A_j)$. By the definition of ε in the statement of Lemma A.1 we have $v_1(A_1) \leq v_1(A_j) - \varepsilon$. This implies $v'_1(A_1) \leq v_1(A_1) + \varepsilon/3 \leq v_1(A_j) - 2\varepsilon/3 \leq v'_1(A_j) - 2\varepsilon/3 < v'_1(A_j)$, which contradicts $v'_1(A_1) \geq v'_1(A_j)$ that we showed above. We conclude that agent 1 does not envy (with respect to v_1 any bundle in the original instance. ☐

References

[1] G. Amanatidis, G. Birmpas, and E. Markakis. On truthful mechanisms for maximin share allocations. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence, IJCAI '16*, pages 31–37. IJCAI/AAAI Press, 2016.

[2] G. Amanatidis, G. Birmpas, G. Christodoulou, and E. Markakis. Truthful allocation mechanisms without payments: Characterization and implications on fairness. In *Proceedings of the 2017 ACM Conference on Economics and Computation, EC' 17*, pages 545–562. ACM, 2017.

[3] G. Amanatidis, E. Markakis, A. Nikzad, and A. Saberi. Approximation algorithms for computing maximin share allocations. *ACM Trans. Algorithms*, 13(4):52:1–52:28, 2017.

[4] G. Amanatidis, G. Birmpas, and E. Markakis. Comparing approximate relaxations of envy-freeness. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI '18*, pages 42–48. ijcai.org, 2018.

[5] G. Amanatidis, E. Markakis, and A. Ntokos. Multiple birds with one stone: Beating 1/2 for EFX and GMMS via envy cycle elimination. *Theor. Comput. Sci.*, 841:94–109, 2020.

[6] G. Amanatidis, G. Birmpas, A. Filos-Ratsikas, A. Hollender, and A. A. Voudouris. Maximum nash welfare and other stories about EFX. *Theor. Comput. Sci.*, 863:69–85, 2021.

[7] H. Aziz, S. Bouveret, J. Lang, and S. Mackenzie. Complexity of manipulating sequential allocation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI '17*, pages 328–334. AAAI Press, 2017.

[8] H. Aziz, P. Goldberg, and T. Walsh. Equilibria in sequential allocation. In *Proceedings of the 5th International Conference on Algorithmic Decision Theory, ADT '17*, volume 10576 of *LNCS*, pages 270–283. Springer, 2017.

[9] S. Barman and S. K. Krishnamurthy. Approximation algorithms for maximin fair division. *ACM Trans. Economics and Comput.*, 8(1):5:1–5:28, 2020.

[10] S. Barman, A. Biswas, S. K. K. Murthy, and Y. Narahari. Groupwise maximin fair allocation of indivisible goods. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence, AAAI '18*, pages 917–924. AAAI Press, 2018.

[11] X. Bei, N. Chen, G. Huzhang, B. Tao, and J. Wu. Cake cutting: Envy and truth. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI '17*, pages 3625–3631. ijcai.org, 2017.

[12] B. Berger, A. Cohen, M. Feldman, and A. Fiat. (almost full) EFX exists for four agents (and beyond). *CoRR*, abs/2102.10654, 2021.

[13] S. Bouveret and J. Lang. Manipulating picking sequences. In *Proceedings of the 21st European Conference on Artificial Intelligence - ECAI 2014*, volume 263, pages 141–146. IOS Press, 2014.

[14] S. Bouveret, Y. Chevaleyre, and N. Maudet. Fair allocation of indivisible goods. In *Handbook of Computational Social Choice*, pages 284–310. Cambridge University Press, 2016.

[15] S. Brânzei, I. Caragiannis, D. Kurokawa, and A. D. Procaccia. An algorithmic framework for strategic fair division. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence AAAI '16*, pages 418–424. AAAI Press, 2016.

[16] S. Brânzei, V. Gkatzelis, and R. Mehta. Nash social welfare approximation for strategic agents. In *Proceedings of the 2017 ACM Conference on Economics and Computation, EC '17*, pages 611–628. ACM, 2017.

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

[18] I. Caragiannis, C. Kaklamanis, P. Kanellopoulos, and M. Kyropoulou. On low-envy truthful allocations. In *First International Conference on Algorithmic Decision Theory, ADT '09*, pages 111–119, 2009.

[19] I. Caragiannis, N. Gravin, and X. Huang. Envy-freeness up to any item with high nash welfare: The virtue of donating items. In *Proceedings of the 2019 ACM Conference on Economics and Computation, EC '19*, pages 527–545. ACM, 2019.

[20] I. Caragiannis, D. Kurokawa, H. Moulin, A. D. Procaccia, N. Shah, and J. Wang. The unreasonable fairness of maximum Nash welfare. *ACM Trans. Economics and Comput.*, 7(3):12:1–12:32, 2019.

[21] B. R. Chaudhury, J. Garg, and K. Mehlhorn. EFX exists for three agents. In *Proceedings of the 2020 ACM Conference on Economics and Computation, EC '20*, pages 1–19. ACM, 2020.

[22] B. R. Chaudhury, T. Kavitha, K. Mehlhorn, and A. Sgouritsa. A little charity guarantees almost envy-freeness. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA '20*, pages 2658–2672. SIAM, 2020.

[23] B. R. Chaudhury, J. Garg, K. Mehlhorn, R. Mehta, and P. Misra. Improving EFX guarantees through rainbow cycle number. In *Proceedings of the 22nd ACM Conference on Economics and Computation, EC '21*, pages 310–311. ACM, 2021.

[24] Y. Chen, J. K. Lai, D. C. Parkes, and A. D. Procaccia. Truth, justice, and cake cutting. *Games Econ. Behav.*, 77(1):284–297, 2013.

[25] R. Cole, V. Gkatzelis, and G. Goel. Mechanism design for fair division: allocating divisible items without payments. In *Proceedings of the 14th ACM Conference on Electronic Commerce, EC '13*, pages 251–268. ACM, 2013.

[26] L. Ehlers and B. Klaus. Coalitional strategy-proof and resource-monotonic solutions for multiple assignment problems. *Soc. Choice Welf.*, 21(2):265–280, 2003.

[27] D. K. Foley. Resource allocation and the public sector. *Yale Economics Essays*, 7:45–98, 1967.

[28] G. Gamow and M. Stern. *Puzzle-Math.* Viking press, 1958.

[29] J. Garg and A. Murhekar. Computing fair and efficient allocations with few utility values. In *Proceedings of the 14th International Symposium on Algorithmic Game Theory, SAGT 2021*, volume 12885 of *LNCS*, pages 345–359. Springer, 2021.

[30] J. Garg and S. Taki. An improved approximation algorithm for maximin shares. In *Proceedings of the 2020 ACM Conference on Economics and Computation, EC '20*, pages 379–380. ACM, 2020.

[31] J. Garg, P. McGlaughlin, and S. Taki. Approximating maximin share allocations. In *Proceedings of the 2nd Symposium on Simplicity in Algorithms, SOSA@SODA 2019*, volume 69 of *OASICS*, pages 20:1–20:11. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

[32] M. Ghodsi, M. T. Hajiaghayi, M. Seddighin, S. Seddighin, and H. Yami. Fair allocation of indivisible goods: Improvements and generalizations. *Math. Oper. Res.*, 46(3):1038–1053, 2021.

[33] L. Gourvès, J. Monnot, and L. Tlilane. Near fairness in matroids. In *Proceedings of the 21st European Conference on Artificial Intelligence - ECAI 2014*, volume 263, pages 393–398. IOS Press, 2014.

[34] B. Klaus and E. Miyagawa. Strategy-proofness, solidarity, and consistency for multiple assignment problems. *Int. J. Game Theory*, 30(3):421–435, 2002.

[35] D. A. Kohler and R. Chandrasekaran. A class of sequential games. *Oper. Res.*, 19(2):270–277, 1971.

[36] D. Kurokawa, A. D. Procaccia, and J. Wang. Fair enough: Guaranteeing approximate maximin shares. *J. ACM*, 65(2):8:1–8:27, 2018.

[37] R. J. Lipton, E. Markakis, E. Mossel, and A. Saberi. On approximately fair allocations of indivisible goods. In *Proceedings of the 5th ACM Conference on Electronic Commerce, EC '04*, pages 125–131. ACM, 2004.

[38] R. Mahara. Existence of EFX for two additive valuations. *CoRR*, abs/2008.08798, 2020.

[39] E. Markakis. Approximation algorithms and hardness results for fair division with indivisible goods. In *Trends in Computational Social Choice*, chapter 12. AI Access, 2017.

[40] S. Papai. Strategyproof multiple assignment using quotas. *Review of Economic Design*, 5(1):91–105, 2000.

[41] S. Papai. Strategyproof and nonbossy multiple assignments. *Journal of Public Economic Theory*, 3(3): 257–271, 2001.

[42] B. Plaut and T. Roughgarden. Almost envy-freeness with general valuations. *SIAM J. Discret. Math.*, 34(2):1039–1068, 2020.

[43] A. D. Procaccia. Cake cutting algorithms. In *Handbook of Computational Social Choice*, pages 311–330. Cambridge University Press, 2016.

[44] H. Steinhaus. Sur la division pragmatique. *Econometrica*, 17 (Supplement):315–319, 1949.

[45] H. R. Varian. Equity, envy and efficiency. *Journal of Economic Theory*, 9:63–91, 1974.

[46] G. J. Woeginger. A polynomial-time approximation scheme for maximizing the minimum machine completion time. *Oper. Res. Lett.*, 20(4):149–154, 1997.

[47] M. Xiao and J. Ling. Algorithms for manipulating sequential allocation. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence, AAAI '20*, pages 2302–2309. AAAI Press, 2020.