ONE ITEM OF SLACK: SEPARATING EF1 FROM EFX UNDER NOISY OBSERVATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Envy-freeness up to one item (EF1) asks that an agent's envy of another bundle vanish once *some* single item is dropped from it; EFX asks this for *every* item. We study what that quantifier costs when valuations are learned from noisy observations rather than given. Our setting repairs envy in a fixed allocation from a pool of spare goods, with standard fair division the case in which nothing is allocated in advance and every good must be handed out. In it we exhibit two agents with coverage valuations, a standard submodular class, over m pool goods and two fixed items, on which one fixed extension, consulting nothing at all, is ε-EF1, while ε-EFX and ε-EF need $\Omega(m^2)$ observations for every $\varepsilon \leq 1/128$. For ε-EFX the exponent is tight: $O(m^2)$ observations suffice on the same family. Every instance admits an exactly envy-free extension, so the cost is not infeasibility in disguise. Neither ingredient forces the quadratic cost alone. Letting the algorithm split the two fixed items makes ε-EFX free on these valuations, and with no fixed allocation two agents reach ε-EFX from $\widetilde{O}(m/\varepsilon^3)$ observations under arbitrary monotone valuations, where $\widetilde{O}$ hides logarithmic factors; exact EFX there still takes exponentially many queries. A transfer principle turns any deterministic oracle algorithm tolerating small errors into a learning algorithm, giving $\widetilde{O}(nm/\varepsilon^2)$ there for ε-EF1 with n agents. One item of slack lets an allocation stop tracking valuations at the scale of single items. A fixed pair that cannot be reallocated separately puts that scale back.

1 INTRODUCTION

A division of goods among agents is *envy-free* (EF) if no agent prefers another's bundle to their own (Varian, 1974). When the goods are indivisible, such a division often does not exist: a single car cannot be split between two people who both want it (Bouveret et al., 2016). Two relaxations dominate. *Envy-freeness up to one item* (EF1) (Budish, 2010; Lipton et al., 2004) asks that an agent's envy of another bundle vanish once *some* single item is dropped from that bundle. *Envy-freeness up to any item* (EFX) (Caragiannis et al., 2019b; Plaut & Roughgarden, 2020b) asks the same once *any* single item is dropped, whichever it is. They differ in one quantifier, some against every. That difference is already stark when valuations are known exactly: an EF1 division always exists and is found in polynomial time, whereas EFX is not known to exist beyond three agents even for additive valuations, and finding one for two agents with submodular valuations is complete for the class of local-search problems (Goldberg et al., 2025).

Valuations are rarely known exactly. They are elicited through queries (Bouveret & Lang, 2011), inferred from partial feedback (Benadè et al., 2022), or estimated from behavior, which is the case we study. For the first the picture is sharp, with n agents and m goods. Deciding whether an envy-free division exists takes communication exponential in m for every $n > 2$, even for submodular valuations and even for randomized protocols (Plaut & Roughgarden, 2020a). Against that, EF1 for two agents with arbitrary monotone valuations, meaning only that extra goods never hurt, needs just $\Theta(\log m)$ queries for the value of a bundle, the lower half of that bound holding for deterministic algorithms. For any n it needs $O(nm)$, achieved by envy-cycle elimination, which repeatedly gives a good to an unenvied agent and rotates bundles around envy cycles (Oh et al., 2021). Asking for exact envy-freeness is expensive. Asking for EF1 is cheap.

That settles what it costs to *ask*, not what it costs to *estimate*. A query returns the value; an observation returns a random quantity whose mean is that value. An estimate can be arbitrarily close to the truth

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

and still be on the wrong side of it, and every fairness condition compares two bundle values, so an error far smaller than either value can flip the verdict. Recent work has approached this from several directions. Li et al. (2025) keep the queries recoverable, returning the exact answer with constant probability, so that repetition and a majority vote restore it at a total cost of $\Theta(nm \log(1/\delta))$ for failure probability δ, with no dependence on any tolerance. Li et al. (2026) add Gaussian noise, which no amount of repetition removes, and price exact envy-freeness for two agents with additive valuations at $\widetilde{\Theta}(m/\Delta^2)$ when the best division clears envy by a margin Δ. Two further lines fix the perturbation instead of paying to reduce it: Dang et al. (2026) bound the envy left by a single perturbation of each report, and Peters et al. (2022) maximize the radius of reports around which a division stays envy-free. Procaccia et al. (2024) study online arrival under bandit feedback. None of them asks which fairness notions remain *certifiable* when every value is only ever known to within a tolerance, nor whether the quantifier that separates EF1 from EFX is what decides the answer. That is the question of this paper, and the quantifier decides it once part of the allocation is fixed.

Why additive tolerance. Under bounded noisy observations, distinguishing bundle values separated by a gap Δ can require $\Theta(\Delta^{-2})$ observations in the worst case at constant confidence. We therefore study ε-EF1 and ε-EFX, in which each comparison is slackened by an additive ε, and count the value observations an algorithm must draw. One might expect the learning theory of submodular functions to close this off immediately: approximating such a function multiplicatively at every point needs a factor $\Omega(\sqrt{m}/\log m)$ on a ground set of m elements (Goemans et al., 2009); learning one in the probably-mostly-approximately-correct sense needs a factor $\widetilde{\Omega}(m^{1/3})$ (Balcan & Harvey, 2011); and reaching additive ℓ_1-error ε needs $2^{\Omega(\varepsilon^{-2/3})}$ queries even under the uniform distribution and even with exact answers (Feldman et al., 2013). Every one of these is a barrier to recovering the valuation. We never recover a valuation. We estimate finitely many bundle values and ask whether a fairness condition can be certified from them, a strictly easier task on which the two notions part company.

The setting. A division is already in place and envy is repaired from a pool of spare goods, the framework of Bentert et al. (2025). Ordinary fair division is the special case in which nothing has been allocated in advance and every pool good must be assigned, so everything proved here applies there too. The part already in place is what makes the setting a useful instrument: it lets us vary how much of the division an algorithm may touch, from none of it to all of it, while holding the valuations and the fairness notion fixed. Filling freed course seats after an initial round is one such case.

Contributions. We study fair division from noisy value observations and separate the two notions.

1. **A quadratic barrier, and it is tight.** Two agents have *coverage* valuations, one of them holds two items that cannot be reallocated, and a pool of m goods with a budget of $m/2$ is available to repair the envy between them. On this family, with Bernoulli observations, every randomized adaptive algorithm that returns an ε-EFX or ε-EF extension with constant probability must draw $\Omega(m^2)$ observations, for every $\varepsilon \leq 1/128$ (Thm. 3.1). An exactly envy-free extension exists on every instance, and one fixed pool item, handed over without any observation, is *exactly* EF1 throughout the family, so the barrier is not one of existence. Nor is it an artifact of our proof: a single observation of a bundle's value is a noisy linear measurement of every hidden bit, and an estimator exploiting this reaches ε-EFX from $O(m^2)$ observations on the same family (Prop. 3.4).
2. **A transfer principle, instantiated for EF1.** A transfer principle converts any deterministic oracle algorithm reaching EF1 while consulting N bundle values, tolerating small errors in them, into a learning algorithm using $O(N\varepsilon'^{-2}\log(N/\delta))$ observations at accuracy ε' (Thm. 4.1). Robustness is a real hypothesis, not a formality: envy-cycle elimination certifies the giver's side by monotonicity, which estimates lack. A margin variant restores it, giving $O(nm\varepsilon^{-2}\log(nm/\delta))$ observations for ε-EF1 in the standard setting with arbitrary monotone valuations (Thm. 4.2).
3. **Neither the quantifier nor the fixed allocation acts alone.** With two agents and no fixed allocation, ε-EFX costs $O(m\varepsilon^{-3}\log(m/\varepsilon\delta))$ observations for arbitrary monotone valuations (Thm. 4.3), although exact EFX there takes exponentially many queries (Plaut & Roughgarden, 2020b). On the hard family itself, letting the algorithm split the two fixed items makes ε-EFX free (Prop. 4.4). What costs is holding two items an algorithm cannot take apart.
4. **Where the separation comes from.** Estimation error transfers to EF1 and to EFX identically (Rem. B.1), so the separation is not a difference in robustness to noise but one of resolution. EF1 may drop the *most* valuable item of the envied bundle, so one large item absorbs the entire envy

gap; EFX must survive dropping the *least*, which leaves it no cushion. Telling apart values that differ by $1/m$ from observations bounded in $[0, 1]$ costs $\Theta(m^2)$ samples.

2 THE MODEL

Allocations. A set A of n agents is to be given items. Some items, the set P, possibly empty, are already allocated: a fixed map σ assigns each agent a subset $\sigma(a)$, with the $\sigma(a)$ pairwise disjoint and $\biguplus_a \sigma(a) = P$, and these cannot be moved. The remaining items form a finite *pool* R of m goods. An *extension* ρ assigns each agent a subset $\rho(a) \subseteq R$, with the assigned subsets pairwise disjoint and $\sum_a |\rho(a)| \leq k$ for a *budget* k. Agent a ends with the bundle $B_a = \sigma(a) \cup \rho(a)$ and values it at $v_a(B_a)$, where each $v_a : 2^{P \cup R} \to [0, 1]$ is monotone. We call this *Envy Elimination by Adding Goods* (EEAG). When $P = \emptyset$ and every pool good is required to be allocated, so that $\biguplus_a \rho(a) = R$, the model specializes to ordinary fair division; we call this the *standard setting*, with budget $k = m$.

Three notions, one inequality. Agent a *envies* a' if $v_a(B_{a'}) > v_a(B_a)$, and an extension is *envy-free* (EF) if no agent envies another. The two relaxations we study ask instead that envy be absorbed by dropping a single item from the envied bundle. They differ only in how that item is quantified: *envy-freeness up to one item* (EF1) drops some item, *envy-freeness up to any item* (EFX) must survive dropping each one. Slackened by a tolerance $\varepsilon \in [0, 1]$, all three read as a single inequality, with the envied bundle required to be nonempty so that there is an item to drop. An extension ρ is

$$\varepsilon\text{-EF1},\ \varepsilon\text{-EFX},\ \text{or}\ \varepsilon\text{-EF} \quad \text{if} \quad v_a(B_a) \geq v_a(B_{a'} \setminus \{g\}) - \varepsilon \quad \text{for all } a, a' \in A \text{ with } B_{a'} \neq \emptyset, \tag{1}$$

where g ranges over *some* item of $B_{a'}$ for ε-EF1, over *every* item of $B_{a'}$ for ε-EFX, and over *no* item at all, meaning that g is dropped from the statement, for ε-EF. Taking $\varepsilon = 0$ recovers the exact notions. Monotonicity makes the notions nested: ε-EF implies ε-EFX implies ε-EF1, and every extension is trivially 1-EF1. The first implication costs nothing to learn; this paper is about the second. In the standard setting EF1 always exists for monotone valuations, whereas EFX need not: it fails already for three agents with coverage valuations, the class our lower bound uses (Mackenzie & Suzuki, 2026). So that our results are about the cost of finding an object rather than about its absence, every instance in our lower bound admits an exactly envy-free extension, so an EFX one exists a fortiori.

Coverage valuations. A valuation is *coverage* if there is a universe U with non-negative weights w, and a set $C_r \subseteq U$ for each item r, such that $v(S) = \sum_{u \in \bigcup_{r \in S} C_r} w(u)$: each item covers part of the universe and a bundle is worth the weight it covers. Coverage valuations are monotone and submodular, and the inclusions Unit-Demand $\subsetneq$ Coverage $\subsetneq$ Submodular are strict, as is Additive $\subsetneq$ Submodular, where unit-demand means valuing a bundle by its best item (Lem. A.4).

Observations. An algorithm does not see v_a. A single *value observation* of the pair (a, S) returns an independent random variable in $[0, 1]$ whose mean is $v_a(S)$; its distribution is part of the instance and unknown, and the algorithm may choose (a, S) adaptively. Writing $\hat{v}_a(S)$ for the mean of T repeated observations of (a, S), the algorithm must output an extension that is ε-EF1, or ε-EFX, or ε-EF, with probability at least $1 - \delta$ over both its own randomness and the observations. The *sample complexity* of a notion, on a class of instances and for given $n, m, \varepsilon, \delta$, is the least number of observations for which some algorithm achieves this on every instance in the class; where the class fixes a budget k and an initial allocation σ, the bounds below may depend on k and $|P|$ as well. The noiseless query model of Oh et al. (2021) is the special case in which the observation is deterministic.

Consultations and robustness. An oracle algorithm accesses valuations only by requesting values $v_a(S)$. An *answer table* is a function $\hat{v}$ on agent-set pairs; a request for (a, S) receives $\hat{v}_a(S)$, so repeated requests agree, and the table is ε'-accurate if $|\hat{v}_a(S) - v_a(S)| \leq \varepsilon'$ for every pair requested; its *consultation number* N bounds the distinct pairs (a, S) it requests under any answer table accurate to ε'. Such an algorithm is (ε', s)-*robust* for a notion $\mathcal{F} \in \{\mathrm{EF1}, \mathrm{EFX}\}$ on a class of valuations if, whenever every answer it receives is within ε' of the true value, its output is s-$\mathcal{F}$ under the true valuations. Robustness is not automatic. An algorithm correct on exact values may fail on perturbed ones, because the perturbed answers need not be the values of any valuation in the class: estimation error does not preserve monotonicity. A margin wide enough to absorb the error restores it (§4.1). We assume throughout that the instance is *EF1-feasible*: some extension within budget is EF1 under the true valuations. Without it the goal can be unreachable for reasons unrelated to observation noise.

3 A QUADRATIC BARRIER FOR EFX AND EF

The quadratic barrier persists under small additive tolerance.

Theorem 3.1 (Quadratic Barrier)**.** *Let $m \geq 2$ be even and $\varepsilon \leq 1/128$. There is a family $\{\mathcal{I}_m(\theta)\}_{\theta\in\{0,1\}^{m/2}}$ of instances with two agents, two initial items, a pool of m goods, budget $k = m/2$, coverage valuations, and Bernoulli value observations, such that every instance admits an exactly envy-free extension within budget, and yet every randomized adaptive algorithm that outputs an ε-EFX extension within budget with probability at least $2/3$ on every θ must draw more than $m^2/32$ observations. The same bound holds for ε-EF.*

Read against Thm. 4.2, which prices ε-EF1 at $O(nm\varepsilon^{-2}\log(nm/\delta))$, for every fixed $\varepsilon \in (0, 1/128]$ the EFX lower bound is quadratic in m whereas the EF1 upper bound is near-linear in m.

The hard family. Fix m even and put $\eta = 2^{-10}$, the value of the item EFX is forced to drop. Agent a_2 starts with both initial items, $\sigma(a_2) = P = \{g^*, g^\circ\}$, and a_1 starts with nothing. The pool is $R = \{r_1, \ldots, r_m\}$ and the budget is $k = m/2$. Group the pool into $m/2$ pairs $\{r_{2i-1}, r_{2i}\}$; a hidden parameter $\theta \in \{0,1\}^{m/2}$ picks one item from each pair, forming

$$S_\theta \;=\; \{\, r_{2i-1} : \theta_i = 0 \,\} \cup \{\, r_{2i} : \theta_i = 1 \,\}, \qquad |S_\theta| = m/2. \tag{2}$$

Agent a_2 values only its own two items, $v_2(S) = \frac{1}{2}\mathbf{1}[g^* \in S] + \frac{1}{2}\mathbf{1}[g^\circ \in S]$, and so never envies. Agent a_1 values

$$v_1^\theta(S) = \tfrac{3}{16}\,\mathbf{1}[S \cap R \neq \emptyset] + \tfrac{3}{16}\min\Big(1, \tfrac{2|S\cap S_\theta|}{m}\Big) + \big(\tfrac{3}{8} - \eta\big)\mathbf{1}[g^* \in S] + \eta\,\mathbf{1}[g^\circ \in S], \tag{3}$$

so that g^* is worth almost everything to them and g° almost nothing. Any nonempty set of pool items is worth $\frac{3}{16}$ on its own, and each item of the hidden set S_θ adds $\frac{3}{8m}$ more, subject to a cap that keeps the function submodular.

The design has one purpose. Agent a_1 envies the pair $\{g^*, g^\circ\}$, worth $\frac{3}{8}$ to them, and the only repair is to hand them pool items. Dropping g° from that pair barely helps, since g° is worth η, so ε-EFX still requires $v_1^\theta(\rho(a_1)) \geq \frac{3}{8} - \eta - \varepsilon$, bringing them almost to the value of g^*. A nonempty extension supplies $\frac{3}{16}$ of that at once, and the remainder can come only from the hidden term, at $\frac{3}{8m}$ per item of S_θ. A budget of $m/2$ items is therefore enough exactly when almost all of them lie in S_θ. EF1 may drop g^* instead, leaving a bundle worth η, which any single pool item beats. That is why one fixed item, chosen without observations, is EF1 throughout, and the fixed pair denies EFX the same escape.

The two lemmas that carry the argument are proved in Appx. L; both were checked exhaustively in exact rational arithmetic for small m, and the constants of the Assouad step symbolically (Appx. K).

Lemma 3.2 (Well-Formedness and Feasibility)**.** *For every θ: (i) v_1^θ and v_2 are coverage valuations, with v_1^θ taking values in $[0, \frac{3}{4}]$ and v_2 in $[0,1]$; (ii) the extension $\rho^*(a_1) = S_\theta$ has size k and is exactly envy-free, with $v_1^\theta(S_\theta) = v_1^\theta(\{g^*, g^\circ\}) = \frac{3}{8}$; (iii) the extension $\rho(a_1) = \{r_1\}$ is exactly EF1, and the empty extension is η-EF1.*

Lemma 3.3 (Recovery)**.** *Let ρ be an extension within budget that is ε-EFX on $\mathcal{I}_m(\theta)$, and write $S = \rho(a_1) \cap R$. Then $d_H(S, S_\theta) \leq \beta m$, where d_H counts the items on which S and S_θ differ and $\beta = \frac{16}{3}(\eta + \varepsilon)$, at most $\frac{3}{64}$ for $\varepsilon \leq 1/128$. The same holds for ε-EF.*

At most $3/32$ of the $m/2$ hidden coordinates may be wrong, so an algorithm that certifies ε-EFX has recovered nearly all of them, and the cost of that is information-theoretic. Flipping one bit of θ moves $v_1^\theta(S)$ by at most $\frac{3}{8m}$ on every S, and since Bernoulli divergence is quadratic in the mean gap, one observation carries $O(1/m^2)$ bits about that coordinate. With $T \leq m^2/32$ observations the transcript laws under θ and under θ with coordinate i flipped stay at total variation distance at most $\frac{1}{8}$. Assouad's lemma, which turns a family of pairwise tests into one estimation bound, converts this per-coordinate indistinguishability into a bound on expected Hamming error across all $m/2$ coordinates, contradicting Lem. 3.3 for $\varepsilon \leq 1/128$; Appx. L gives the constants.

3.1 THE BOUND IS TIGHT ON THIS FAMILY

The barrier is not an artifact of the argument. A single observation of $v_1^\theta(S)$ is a noisy linear measurement of *all* of θ at once, and $O(m^2)$ observations suffice on this family for an algorithm that exploits this to recover enough of θ, matching Thm. 3.1.

Proposition 3.4 (Matching Upper Bound). *On the family of Thm. 3.1, with Bernoulli observations and any $\varepsilon \le 1/128$, drawing $T = \lceil 700\, m^2 \rceil$ observations suffices to output an ε-EFX extension within budget with probability at least $3/4$.*

The constant is not optimized; it absorbs the Hoeffding and union-bound factors. With Thm. 3.1 this makes the sample complexity of ε-EFX on $\{\mathcal{I}_m(\theta)\}$ exactly $\Theta(m^2)$ at constant success probability. The estimator is one line. Over random half-sets of R, the average of the observation restricted to draws containing r_{2i} but not r_{2i-1} exceeds the average over the complementary draws by exactly $\frac{3}{8m}(2\theta_i - 1)$ in expectation, because every other pair contributes identically to both; the sign recovers θ_i, and Appx. L.6 supplies the concentration. A scheme that compares the two singletons of each pair instead learns one bit per $\Theta(m^2)$ observations and pays $\Theta(m^3)$: sharing observations across coordinates is worth a factor of m.

3.2 Where the Barrier Stops

The relaxation must be additive. Under the multiplicative form, which asks $v_a(B_a) \ge c \cdot v_a(B_{a'} \setminus \{g\})$, the fixed one-item extension $\rho(a_1) = \{r_1\}$ already qualifies on this family for every $c \le \frac{1}{2}$, without any observations, so the mechanism behind our lower bound is gone already at $c = \frac{1}{2}$. An additive ε rescales nothing, which is why it preserves the fine structure the bound exploits.

The tolerance must be small. For $\varepsilon > \frac{3}{32} - \eta \approx 0.0928$ a uniformly random extension of size $m/2$ is ε-EFX with probability $1 - e^{-\Omega(m)}$, again with no observations (Prop. L.5). Thm. 3.1 establishes the quadratic lower bound for every $\varepsilon \le 1/128$; the intervening regime is not resolved by our results.

A budget of one item is not enough, and the valuations are already simple. At $k = 1$ the conclusion fails on EF-feasible instances, since Prop. E.1 then gives $o(m^2)$ observations, so $k = m/2$ is not incidental (Cor. E.2). Here ε-EF1 is free and the valuations are coverage; for unit-demand, the class just below it, ε-EFX and ε-EF coincide once the envied bundle has two items (Prop. F.1), so any barrier there is purely informational; Prop. E.4 gives a near-linear bound in $q = |P| + m$ whenever an s-EFX extension exists for a fixed $s < \varepsilon$, since the nq singleton values determine every bundle.

4 Learning Fair Allocations

EF1 is learnable, and for two agents, with the allocation released, so is EFX. Both follow two steps: bound how many bundle values an oracle algorithm consults, then pay a concentration factor for each.

4.1 From Oracle Algorithms to Learning Algorithms

Estimation error transfers to a fairness condition without regard to the quantifier. If every value appearing in a certificate is accurate to ε', then a certificate valid under the estimates is valid under the truth up to $2\varepsilon'$; the argument is identical for EF1 and for EFX (Lem. B.2 and Rem. B.1). What separates the two notions is therefore not this step.

The step that does the work is simulation. Run an oracle algorithm, and whenever it requests a value not requested before, draw T_0 observations and answer with their mean; repeated requests return the stored answer. All N distinct requests are accurate with probability $1 - \delta$ by Hoeffding and a union bound, unaffected by adaptivity because each term conditions on the request's history.

Theorem 4.1 (Transfer Principle). *Let $\varepsilon', \delta \in (0,1)$ and $s \ge 0$, fix a notion $\mathcal{F} \in \{\mathrm{EF1}, \mathrm{EFX}\}$, and let $\mathcal{B}$ be a deterministic (ε', s)-robust oracle algorithm for $\mathcal{F}$ that terminates on every answer table and has consultation number N at accuracy ε'. If $N = 0$ no observations are needed; for $N \ge 1$, simulating $\mathcal{B}$ with $T_0 = \lceil \ln(2N/\delta)/(2\varepsilon'^2) \rceil$ observations per distinct request outputs an s-$\mathcal{F}$ extension within budget with probability at least $1 - \delta$, at a total cost of $NT_0 = O(N\varepsilon'^{-2} \log(N/\delta))$.*

An $(\varepsilon/4, \varepsilon)$-robust algorithm for EF1 therefore yields ε-EF1 from $N\lceil 8\varepsilon^{-2} \ln(2N/\delta) \rceil$ observations. The theorem is deliberately modest: it prices the learning setting at $O(\varepsilon'^{-2} \log(N/\delta))$ per consulted value and puts the rest into robustness. That hypothesis is not free. Envy-cycle elimination certifies the giver's side of each EF1 constraint by the monotonicity of the profile it is fed, and estimates are not monotone. The cut-and-choose below needs a separate argument, and Appx. M.1 shows that the

obvious greedy alternative fails even with exact values, so robustness is not what rules it out. In every case the consultation count is what enters the bound, and keeping it small is what the algorithms do.

4.2 EF1 IN THE STANDARD SETTING

The standard setting is EEAG with $\sigma \equiv \emptyset$, budget $k = m$, and every good allocated (§2). Envy-cycle elimination allocates goods one at a time to an agent nobody envies, restoring that invariant by rotating bundles along directed cycles of the envy graph; the item allocated last to a bundle is the EF1 witness against it (Lipton et al., 2004). Driven by estimates, a rotation that looks improving under $\hat{v}$ need not improve an agent's true value, so the true-value monotonicity the EF1 proof uses can fail. We therefore rotate only along cycles whose estimated envy exceeds a margin τ; under an ε'-accurate answer table the margin makes every such rotation truly improving, and restores robustness.

Theorem 4.2 (EF1, Standard Setting)**.** *For $\varepsilon, \delta \in (0, 1)$, in the standard setting with arbitrary monotone valuations, margin envy-cycle elimination with $\varepsilon' = \varepsilon/8$ and $\tau = 3\varepsilon'$, simulated as in Thm. 4.1 with $\mathcal{F} =$ EF1, outputs an ε-EF1 allocation with probability at least $1 - \delta$ using $O\big(nm\varepsilon^{-2}\log(nm/\delta)\big)$ value observations.*

Two counts give the bound. Every distinct bundle is created by one item addition, so at most $m + 1$ sets ever exist and each is evaluated by at most n agents: $N \le n(m + 1)$. Rotations permute existing sets and consult nothing new. Under an ε'-accurate answer table the margin also bounds the total number of rotations, since each raises every participant's true value by more than $\tau - 2\varepsilon' = \varepsilon'$, capping them at n/ε'. The bound carries no n^2 factor: the n^2 envy comparisons are made on $m + 1$ distinct sets, and estimates are shared across them. Appx. M.2 gives the argument in full.

4.3 TWO AGENTS WITHOUT A FIXED ALLOCATION: THE BARRIER DISAPPEARS

Thm. 3.1 is proved in EEAG, which leaves open whether the barrier belongs to the universal quantifier or to the fixed allocation. For two agents it is the latter, and sharply so: with nothing fixed, the additive relaxation makes EFX polynomial for arbitrary monotone valuations, though exact EFX takes exponentially many queries there (Plaut & Roughgarden, 2020b).

The algorithm is cut-and-choose. Agent a_1 splits the goods into two pieces Π_1, Π_2 that are ε-EFX from their own perspective, by a local search moving one item at a time from the richer piece to the poorer; a_2 then takes the piece they prefer. All the difficulty is in stopping the search after polynomially many moves when it is driven by estimates, and two observations do it.

Retirement. Within one pass, an item found not to violate need never be examined again. If $\hat{v}_1(\Pi \setminus \{g\}) \le \hat{v}_1(\Pi^{\mathrm{c}}) + \tau$ at some point in the pass, then $f(\Pi \setminus \{g\}) \le f(\Pi^{\mathrm{c}}) + \varepsilon$; and since Π only shrinks and Π^{c} only grows for the rest of the pass, monotonicity preserves that inequality. The estimate that would contradict it is never requested, which is what denies an adversary its leverage.

Free certification of the poorer piece. If a pass moves an item at all, the receiving piece is genuinely the poorer one, so every single-item removal from it is dominated outright and needs no consultation.

Each pass therefore costs $O(m)$ consultations, and passes are few: a move that leaves the two pieces comparable raises $\min(f(\Pi_1), f(\Pi_2))$ by $\varepsilon/4$, and that potential is bounded by 1.

Theorem 4.3 (EFX for Two Agents, Standard Setting)**.** *For $\varepsilon, \delta \in (0, 1)$ and $m \ge 2$, in the standard setting with two agents and arbitrary monotone valuations, the epoch cutter of Alg. 1, simulated as in Thm. 4.1 with $\mathcal{F} =$ EFX, outputs an ε-EFX allocation with probability at least $1 - \delta$ using $O\big(m\varepsilon^{-3}\log(m/(\varepsilon\delta))\big)$ value observations.*

The two counts behind this are $N = O(m/\varepsilon)$ consulted bundles and, at the accuracy $\varepsilon_0 = \varepsilon/8$ the proof uses, $O(\varepsilon^{-2}\log(N/\delta))$ observations apiece; Appx. D gives the algorithm and the proof. Against this, Thm. 4.2 prices ε-EF1 for two agents at $O(m\varepsilon^{-2}\log(m/\delta))$, so in the standard setting the two notions sit within a factor $1/\varepsilon$, both near-linear in m, whereas Thm. 3.1 puts ε-EFX at $\Omega(m^2)$ for $\varepsilon \le 1/128$ once the initial allocation is fixed. Fixing the initial allocation thus raises the known complexity from near-linear to quadratic, and on that family the exponent is exact.

Proposition 4.4 (Releasing the Initial Allocation)**.** *Take the two agents and the valuations of $\mathcal{I}_m(\theta)$ in the standard setting on the $m+2$ goods $\{g^*, g^\circ\} \cup R$. For every θ, the fixed allocation $B_{a_1} = \{g^*, r_1\}$ and $B_{a_2} = \{g^\circ\} \cup (R \setminus \{r_1\})$ is exactly envy-free and consults no observations.*

The whole $\Omega(m^2)$ is created by fixing $\sigma(a_2) = \{g^*, g^\circ\}$. Being unable to split those two items is what forces an algorithm to reconstruct S_θ; once it may split them, there is nothing to reconstruct.

Fig. 1 shows that retirement is what buys the linear count. On the instance of Prop. D.1, a search that re-examines every remaining item after each move consults $\Theta(m^2)$ bundles, fitted exponent 1.93 with bootstrap interval $[1.91, 1.96]$, while the epoch rule consults $\Theta(m)$, exponent 0.95 with interval $[0.92, 0.97]$; the intervals are disjoint and each excludes the other's exponent. On random instances with independent noise we find no clear separation between the two variants. We report that too, because Prop. D.1 is a worst-case statement and the estimate pattern it needs is not what independent noise produces: retirement guards against adversarial error rather than speeding up the average case.

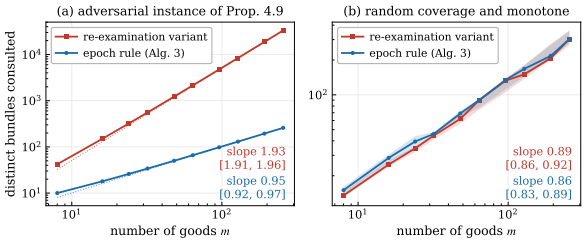

Figure 1: Distinct bundles consulted by the cutter's local search, $\varepsilon = 0.1$. **(a)** On the instance of Prop. D.1, re-examination costs $\Theta(m^2)$ while the epoch rule costs $\Theta(m)$; dotted lines are $0.506m^2$ and $1.01m$. **(b)** On random coverage and monotone instances with independent noise, 40 seeds per point, medians with bootstrap 95% intervals: no clear separation between the two. Annotated slopes are log-log fits with bootstrap intervals.

4.4 The Universal Notions in EEAG

Above the barrier, ε-EF and hence ε-EFX can still be priced, and the price is set by the budget rather than the valuation class. If some extension within budget is exactly envy-free, then estimating every achievable bundle to accuracy $\varepsilon/4$ and searching exhaustively returns an ε-EF extension, at $N = n^2 \sum_{j \le k} \binom{m}{j}$ consulted values (Prop. E.1). Monotonicity carries the guarantee down to ε-EFX and ε-EF1 (Cor. E.3), and for classes determined by their singleton values the count drops to $O(nq)$ with $q = |P| + m$ (Prop. E.4). At $k = 1$ the cost is $O(n^2 m \varepsilon^{-2} \log(nm/\delta))$, which is $o(m^2)$, so the barrier of Thm. 3.1 does not survive a budget of one item (Cor. E.2); where between $k = 1$ and $k = m/2$ it appears is open. And the certificate used here contains no removed bundles at all: under exact envy-free feasibility the cheapest route to the universal notion runs through exact envy-freeness.

Whether the exponential dependence on the budget is necessary is open, but the reason the linear consultation count underlying Thm. 4.2 is unavailable here is clear. Envy-cycle elimination keeps its bundle count at $m + 1$ by *permuting bundles among agents*, and a fixed initial allocation forbids that: $B_a = \sigma(a) \cup \rho(a)$ is anchored to a by σ. Rotation is not an optional step: without it the procedure already fails on exact values (Appx. M.1), so the transfer principle has nothing linear to transfer. That is also why the barrier lives in EEAG: a fixed σ is what lets an agent hold two items an algorithm cannot take apart, and Prop. F.1 shows that holding two items is when the quantifier acquires force.

5 The Sample Complexity Landscape

Table 1 collects what the preceding sections establish, organized by the quantifier over the removed item. The existential notion is learnable from a number of observations near-linear in m wherever an oracle algorithm with linear consultation number exists, and the valuation class enters only through that consultation number. The universal notions face a barrier quadratic in m throughout the small-tolerance regime of Thm. 3.1, and for ε-EFX the exponent is exact on the family that exhibits it. Every upper bound in the table is proved here. The one lower bound we do not prove is inherited: the noiseless value-query model is the special case of our observation model in which the observation is deterministic, so the $\Omega(n \log m)$ bound for binary valuations carries over, for deterministic algorithms.

For unit-demand valuations the two universal notions coincide as soon as the envied bundle holds two items, so the removed item carries no information and any barrier there would have to be purely informational (Prop. F.1). A fixed initial allocation is what creates such a pair. Appx. G discusses these unresolved bounds and related questions; the first asks whether ε-EF1 in the standard setting really costs $\Omega(nm/\varepsilon^2)$, which would make the upper bound of Thm. 4.2 tight.

Table 1: Sample complexity landscape. Upper bounds suppress logarithmic factors; the cited theorem or proposition gives the exact dependence. Here $\binom{m}{\le k} = \sum_{j \le k} \binom{m}{j}$. The lower bound for ε-EFX and ε-EF holds for every $\varepsilon \le 1/128$, for $n = 2$, and already within coverage valuations. The inherited EF1 lower bound applies to deterministic algorithms with $\varepsilon < 1/2$ and in the regime $m \ge n^\alpha$ for a constant $\alpha > 1$. Entries marked open are unresolved; related questions are discussed in Appx. G. The entry marked $[*]$ is inherited from the noiseless query model (Oh et al., 2021, Thm. 5.5), as explained in Appx. G; the dagger records that it applies only to deterministic algorithms.

Notion	Setting and class	Upper		Lower
ε-EF1	standard, monotone	$O(nm/\varepsilon^2)$	[Thm. 4.2]	$\Omega(n \log m)^\dagger$ [$*$]
ε-EF1	EEAG, unit-demand	$O(nq/\varepsilon^2)$	[Prop. E.4]	open
ε-EF1	EEAG, additive	$O(nq^3/\varepsilon^2)$	[Prop. E.4]	open
ε-EF1	EEAG, monotone, budget k	$O\big(n^2(1+\lvert P\rvert+k)\binom{m}{\le k}/\varepsilon^2\big)$	[Cor. E.3]	open
ε-EF1	family of Thm. 3.1	0	[Lem. 3.2]	
ε-EFX	standard, monotone, $n = 2$	$O(m/\varepsilon^3)$	[Thm. 4.3]	open
ε-EFX	released family of Thm. 3.1	0	[Prop. 4.4]	
ε-EFX, ε-EF	EEAG, monotone, budget k, EF-feasible	$O\big(n^2\binom{m}{\le k}/\varepsilon^2\big)$	[Prop. E.1]	
ε-EFX, ε-EF	EEAG, monotone, $k = 1$, EF-feasible	$O(n^2 m/\varepsilon^2)$	[Prop. E.1]	
ε-EFX, ε-EF	EEAG, coverage, $n = 2$, $k = m/2$, EF-feasible	$2^{O(m)}/\varepsilon^2$	[Prop. E.1]	$\Omega(m^2)$ [Thm. 3.1]
ε-EFX	family of Thm. 3.1	$O(m^2)$	[Prop. 3.4]	$\Omega(m^2)$ [Thm. 3.1]

6 Experiments

The theory makes claims about counts of observations, and those counts can be measured on the very instances the proofs are about. Every quantity below is measured over independent instances drawn afresh for each seed. Counts are skewed, so we summarize them by medians. Success rates carry Wilson intervals; scaling exponents are least-squares fits on log-log data, with resampling intervals where reported. Comparisons between two notions are made on the *same* instance with the *same* observations, so the test is the exact McNemar test with Holm correction. Observations in Figures 2 and 3 are Bernoulli, the model of Thm. 3.1. Appx. K gives the numerical summaries and the main experimental protocol; the released scripts contain the complete construction and sampling details.

6.1 The Separation, and What Causes It

Fig. 2 measures the one claim the paper is named for. We fix the family $\mathcal{I}_m(\theta)$ of Thm. 3.1 at $m = 32$ and $\varepsilon = 1/128$, give an algorithm a budget of T observations, and record which of ε-EF1, ε-EFX, and ε-EF its output satisfies under the *true* valuations. One algorithm produces one allocation, checked three ways, so no notion gets its own method and no difference can come from the algorithm; the algorithm is told T but not which notion will judge it.

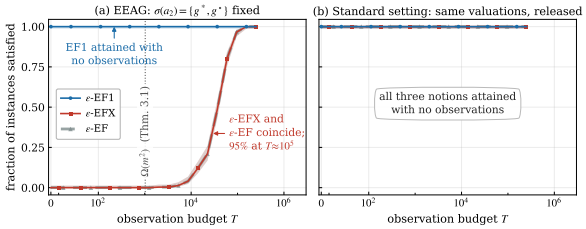

Figure 2: Fraction of instances on which an allocation produced from T noisy observations satisfies each notion, under the true valuations. Family $\mathcal{I}_m(\theta)$ with $m = 32$, $\varepsilon = 1/128$, 300 instances, Bernoulli observations, the estimator of Prop. 3.4; bands are Wilson 95% intervals. **(a)** With $\sigma(a_2) = \{g^*, g^\circ\}$ fixed, EF1 costs nothing and the universal notions cost about 10^5 observations. **(b)** The same valuations with nothing fixed: all three notions hold everywhere at $T = 0$. Markers are staggered so coincident curves remain distinguishable.

In the fixed setting ε-EF1 is satisfied on every instance at $T = 0$ and stays there, while ε-EFX reaches half the instances only at $T \approx 5.9 \times 10^4$ and 95% at $T \approx 9.6 \times 10^4$, putting the measured constant in front of m^2 at 58 and 93. The algorithm is the bit-sharing estimator of Prop. 3.4, and the measured constant sits between the $1/32$ of the lower bound and the 700 of the proof. Paired exact McNemar tests reject equality of the two notions at every budget, with Holm-adjusted p below 10^{-40} at the first four and $p = 0.0015$ at the largest, where ε-EFX has nearly caught up. Releasing the initial allocation, which changes nothing about the valuations, moves every one of those tests to non-significance: all three notions hold on all 300 instances at every budget

including $T = 0$, exactly as Prop. 4.4 predicts. The measured gap is thus attributable to the freeze and not to the valuations or to the accuracy demanded of them. Two features of the figure are themselves predictions: the ε-EFX and ε-EF curves are indistinguishable because on this family the two notions differ by at most $\eta = 2^{-10}$: ε-EFX tolerates $h \le \frac{8m}{3}(\eta + \varepsilon)$ wrong coordinates and ε-EF tolerates $h \le \frac{8m}{3}\varepsilon$, and at $m = 32$, $\varepsilon = 1/128$ both round to the same integer. And the transition sits three and a half orders of magnitude above the $\Omega(m^2)$ line, well inside the gap between the constants $1/32$ and 700; the exponent, which is what the theorem concerns, is measured in Fig. 3.

6.2 Where the Barrier Lives in the (m, ε) Plane

Thm. 3.1 and Prop. L.5 bracket the barrier from two sides: it grows in m, and it disappears once ε exceeds $\frac{3}{32} - \eta \approx 0.0928$. Fig. 3 measures the whole plane rather than either slice, recording for each (m, ε) the empirical observation threshold at which the sampled success rate first reaches 95%, located by bisection to within a factor 1.45.

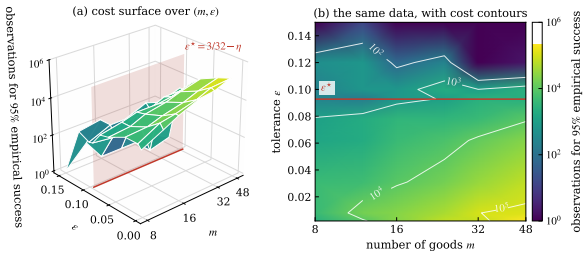

Figure 3: Empirical observation threshold at which ε-EFX first reaches a 95% success rate on the sampled instances of $\mathcal{I}_m(\theta)$, over the (m, ε) plane; 60 instances per estimate, thresholds located by bisection to within a factor 1.45, the bit-sharing estimator of Prop. 3.4 with Bernoulli observations; Appx. K.1 reports a 200-instance near-threshold replication on the $\varepsilon = 0.002$ slice. **(a)** The surface, with the plane $\varepsilon = \varepsilon^\star = \frac{3}{32} - \eta$ of Prop. L.5 shown in red. **(b)** The same values as a heat map with contours. Cost rises sharply in m and only gently in $1/\varepsilon$ until $\varepsilon^\star$, beyond which it collapses to nothing at all.

Along m the cost climbs steeply. On the $\varepsilon = 0.002$ slice, which lies inside the range of Thm. 3.1, we reran near-threshold budgets for $m \in \{8, 12, 16, 24, 32, 48\}$ with 200 matched-trial instances per budget. Taking the first tested budget whose empirical success rate reaches 95% gives a log-log slope of 2.35, with a 95% resampling interval of $[2.10, 2.51]$; requiring instead that the Wilson 95% lower bound on success itself exceed 0.95 gives 2.24. These are finite-grid empirical exponents, and the asymptotic $\Theta(m^2)$ comes from Thm. 3.1 and Prop. 3.4; Appx. K.1 gives the per-budget counts. Along ε the surface is comparatively flat: relaxing ε from 0.002 to 0.031 buys only a factor of about 2.5, the quantitative content of the claim that the barrier is in m and not in $1/\varepsilon$. Past $\varepsilon^\star$ the surface drops away. At $\varepsilon = 0.150$ a single observation suffices at every m except $m = 12$, where 13 are needed; the concentration behind Prop. L.5 is $1 - e^{-\Omega(m)}$, and at the smallest m a random half-set can still miss the required overlap by chance. The collapse begins just above the red line at $\varepsilon^\star$, which is drawn from theory and not fitted to the data.

What else was measured. Appx. J discusses the ablation of Fig. 1 and reports further experiments.

7 Conclusion

How many noisy observations does it take to certify a fair allocation? The answer turns on one syntactic feature: whether the verifier may discount some item of the envied bundle or every item. For the existential notion, a transfer principle reduces learning to the oracle problem at $O(\varepsilon'^{-2} \log(N/\delta))$ per consulted value at accuracy ε', and standard-setting margin envy-cycle elimination gives $\widetilde{O}(nm/\varepsilon^2)$ observations. For the universal notions, no algorithm escapes $\Omega(m^2)$ for any fixed $\varepsilon \le 1/128$, even with two agents and an exactly envy-free extension available; for ε-EFX, $O(m^2)$ suffices on that family, so its exponent is exact. Estimation error transfers to the two notions identically. What differs is resolution: EFX must track valuations at the scale of single items, and EF1 need not.

The quantifier does not act alone. Release the initial allocation and the barrier vanishes, on the hard family outright and in the standard setting down to $\widetilde{O}(m/\varepsilon^3)$ for two agents with arbitrary monotone valuations, although exact EFX there takes exponentially many queries. A fixed allocation is what puts two items in an agent's hands that an algorithm cannot separate, and gives the universal quantifier something to bite on. Appx. G discusses the open bounds and related questions.

REPRODUCIBILITY STATEMENT

Complete proofs of all theoretical results appear in the main text and the appendices, with the lower-bound details in Appx. L and the learning-algorithm details in Appx. M; the experiments of §6 are reported in Appx. J and documented in Appx. K, with the main protocol stated there and the complete construction and sampling details in the released scripts, and the scripts that regenerate every figure and the data underlying every experimental table, together with the verification code and the raw measurement files, are available at `https://anonymous.4open.science/r/slack`. The lower-bound construction of §3 is fully explicit with all constants stated. Five ingredients were computationally cross-checked. Four were verified exhaustively for $m \in \{2, 4, 6, 8\}$ in exact rational arithmetic: well-formedness (value range, monotonicity, submodularity), the feasibility and EF1 claims of Lem. 3.2, the recovery inequality of Lem. 3.3, and the mean-shift bound of Lem. L.2. The constant arithmetic in the Assouad step was checked separately, symbolically and in exact rational arithmetic. The counterexample of Appx. M.1 was executed for $m = 1, \ldots, 8$. As an additional stress test, Alg. 1 was run against an adversary that, at every consultation, branches on the estimate taking either endpoint of its accuracy interval. This explores endpoint perturbations rather than every accurate answer table, so it supplements the proof of Thm. 4.3 without replacing it; over $770{,}576$ complete adversarial paths on random coverage, random monotone, and the anchor instances with $m \leq 6$, every output was two-sided ε-EFX under the true valuation and within the stated consultation and epoch bounds. The verification scripts are in the same repository.

AI USE STATEMENT

During the preparation of this paper, the authors used generative AI tools for copyediting, grammar checking, and improvements to clarity and readability. They were not used to develop the theoretical framework, to prove any of the theorems or propositions, to propose or refine hypotheses, to design the experiments or to provide feedback on the methodology, to implement the method, to generate or clean any dataset, or to interpret results. All content was developed and finalized by the authors, who take full responsibility for the accuracy, originality, and integrity of this paper.

ETHICS STATEMENT

This paper is theoretical and we collected no data ourselves. The two preference datasets used in §6 were gathered and released publicly by others and carry no identifying information: one records rankings of courses by students, the other rankings of web pages by search engines. We use them unmodified apart from the fixed rank-to-value transformations described in Appx. J and specified in the released scripts, and we release no data. The subject matter is fairness in allocation, and the results bear on how much evidence a fairness guarantee requires before it means anything; we see no application of these techniques that creates risk beyond that of fair division itself. The authors have read and adhere to the ICLR Code of Ethics.

REFERENCES

Hannaneh Akrami and Nidhi Rathi. Epistemic EFX Allocations Exist for Monotone Valuations. In *Proc. AAAI'25*, pp. 13520–13528. AAAI Press, 2025. URL `https://doi.org/10.1609/aaai.v39i13.33476`.

Hannaneh Akrami, Alexander Mayorov, Kurt Mehlhorn, Shreyas Srinivas, and Christoph Weidenbach. A Counterexample to EFX; n $\geq$ 3 Agents, m $\geq$ n + 5 Items, Monotone Valuations; via SAT-Solving. *CoRR*, abs/2604.18216, 2026. URL `https://doi.org/10.48550/arXiv.2604.18216`.

Georgios Amanatidis, Evangelos Markakis, and Apostolos Ntokos. Multiple birds with one stone: Beating 1/2 for EFX and GMMS via envy cycle elimination. *Theor. Comput. Sci.*, 841:94–109, 2020. URL `https://doi.org/10.1016/j.tcs.2020.07.006`.

Georgios Amanatidis, Georgios Birmpas, Aris Filos-Ratsikas, and Alexandros A. Voudouris. Fair Division of Indivisible Goods: A Survey. In *Proc. IJCAI'22*, pp. 5385–5393. ijcai.org, 2022. URL `https://doi.org/10.24963/ijcai.2022/756`.

Georgios Amanatidis, Haris Aziz, Georgios Birmpas, Aris Filos-Ratsikas, Bo Li, Hervé Moulin, Alexandros A. Voudouris, and Xiaowei Wu. Fair division of indivisible goods: Recent progress and open questions. *Artif. Intell.*, 322:103965, 2023. URL https://doi.org/10.1016/j.artint.2023.103965.

Christian Arnsperger. Envy-freeness and distributive justice. *Journal of Economic Surveys*, 8(2): 155–186, 1994. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1467-6419.1994.tb00098.x.

Maria-Florina Balcan and Nicholas J. A. Harvey. Learning submodular functions. In *Proc. STOC'11*, pp. 793–802. ACM, 2011. URL https://doi.org/10.1145/1993636.1993741.

Maria-Florina Balcan, Tuomas Sandholm, and Ellen Vitercik. Sample complexity of automated mechanism design. In *Advances in Neural Information Processing Systems 29 (NeurIPS'16)*, pp. 2083–2091, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/c667d53acd899a97a85de0c201ba99be-Abstract.html.

Maria-Florina Balcan, Travis Dick, Ritesh Noothigattu, and Ariel D. Procaccia. Envy-Free Classification. In *Advances in Neural Information Processing Systems 32 (NeurIPS'19)*, pp. 1238–1248, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/e94550c93cd70fe748e6982b3439ad3b-Abstract.html.

Eric Balkanski, Aviad Rubinstein, and Yaron Singer. The Limitations of Optimization from Samples. *J. ACM*, 69(3):21:1–21:33, 2022. URL https://doi.org/10.1145/3511018.

Siddharth Barman and Sanath Kumar Krishnamurthy. Approximation Algorithms for Maximin Fair Division. *ACM Trans. Economics and Comput.*, 8(1):5:1–5:28, 2020. URL https://doi.org/10.1145/3381525.

Gerdus Benadè, Daniel Halpern, and Alexandros Psomas. Dynamic Fair Division with Partial Information. In *Advances in Neural Information Processing Systems 35 (NeurIPS'22)*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/17bb0edcc02bd1f74e771e23b2aa1501-Abstract-Conference.html.

Matthias Bentert, Robert Bredereck, Eva Michelle Deltl, Pallavi Jain, and Leon Kellerhals. How to Resolve Envy by Adding Goods. In *Proc. IJCAI'25*, pp. 3753–3761. ijcai.org, 2025. URL https://doi.org/10.24963/ijcai.2025/417.

Niclas Boehmer, Robert Bredereck, Klaus Heeger, Dusan Knop, and Junjie Luo. Multivariate algorithmics for eliminating envy by donating goods. *Auton. Agents Multi Agent Syst.*, 38(2):43, 2024. URL https://doi.org/10.1007/s10458-024-09674-5.

Sylvain Bouveret and Jérôme Lang. Efficiency and Envy-freeness in Fair Division of Indivisible Goods: Logical Representation and Complexity. *J. Artif. Intell. Res.*, 32:525–564, 2008. URL https://doi.org/10.1613/jair.2467.

Sylvain Bouveret and Jérôme Lang. A General Elicitation-Free Protocol for Allocating Indivisible Goods. In *Proc. IJCAI'11*, pp. 73–78. IJCAI/AAAI, 2011. URL https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-024.

Sylvain Bouveret, Yann Chevaleyre, and Nicolas Maudet. Fair Allocation of Indivisible Goods. In *Handbook of Computational Social Choice*, pp. 284–310. Cambridge University Press, 2016. URL https://doi.org/10.1017/CBO9781107446984.013.

Steven J. Brams and Alan D. Taylor. *Fair division - from cake-cutting to dispute resolution*. Cambridge University Press, 1996.

Johannes Brustle, Jack Dippel, Vishnu V. Narayan, Mashbat Suzuki, and Adrian Vetta. One Dollar Each Eliminates Envy. In *Proc. EC'20*, pp. 23–39. ACM, 2020. URL https://doi.org/10.1145/3391403.3399447.

Xiaolin Bu, Zihao Li, Shengxin Liu, Jiaxin Song, and Biaoshuai Tao. Logarithmic comparison-based query complexity for fair division of indivisible goods. In *Proc. WINE'24*, volume 15534 of *LNCS*, pp. 348–365. Springer, 2024. URL https://doi.org/10.1007/978-3-032-08560-3_20.

Eric Budish. The combinatorial assignment problem: approximate competitive equilibrium from equal incomes. In *Proc. BQGT'10*, pp. 74:1. ACM, 2010. URL https://doi.org/10.1145/1807406.1807480.

Ioannis Caragiannis, Nick Gravin, and Xin Huang. Envy-Freeness Up to Any Item with High Nash Welfare: The Virtue of Donating Items. In *Proc. EC'19*, pp. 527–545. ACM, 2019a. URL https://doi.org/10.1145/3328526.3329574.

Ioannis Caragiannis, David Kurokawa, Hervé Moulin, Ariel D. Procaccia, Nisarg Shah, and Junxing Wang. The Unreasonable Fairness of Maximum Nash Welfare. *ACM Trans. Economics and Comput.*, 7(3):12:1–12:32, 2019b. URL https://doi.org/10.1145/3355902.

Bhaskar Ray Chaudhury, Telikepalli Kavitha, Kurt Mehlhorn, and Alkmini Sgouritsa. A Little Charity Guarantees Almost Envy-Freeness. *SIAM J. Comput.*, 50(4):1336–1358, 2021. URL https://doi.org/10.1137/20M1359134.

Bhaskar Ray Chaudhury, Jugal Garg, and Kurt Mehlhorn. EFX Exists for Three Agents. *J. ACM*, 71 (1):4:1–4:27, 2024. URL https://doi.org/10.1145/3616009.

Trung Dang, Daniel Halpern, Anuran Makur, Alexandros Psomas, Japneet Singh, and Paritosh Verma. Fair Division Under Inaccurate Preferences. *CoRR*, abs/2602.24169, 2026. URL https://doi.org/10.48550/arXiv.2602.24169.

Britta Dorn, Ronald de Haan, and Ildikó Schlotter. Obtaining a Proportional Allocation by Deleting Items. *Algorithmica*, 83(5):1559–1603, 2021. URL https://doi.org/10.1007/s00453-020-00794-4.

Lester E. Dubins and Edwin H. Spanier. How to Cut A Cake Fairly. *The American Mathematical Monthly*, 68(1):1–17, 1961. URL https://www.jstor.org/stable/2311357.

Piotr Faliszewski. AGH Course Selection (PrefLib dataset 00009). https://preflib.github.io/PrefLib-Jekyll/dataset/00009, 2013. File 00009-00000001.soc, 2003 registration: 146 complete strict rankings over 9 courses.

Uriel Feige. Low communication protocols for fair allocation of indivisible goods. In *Proc. EC'25*, pp. 358–382. ACM, 2025. URL https://doi.org/10.1145/3736252.3742553.

Vitaly Feldman, Pravesh Kothari, and Jan Vondrák. Representation, Approximation and Learning of Submodular Functions Using Low-rank Decision Trees. In *Proc. COLT'13*, volume 30 of *JMLR Workshop and Conference Proceedings*, pp. 711–740. JMLR.org, 2013. URL http://proceedings.mlr.press/v30/Feldman13.html.

Satoru Fujishige. *Submodular Functions and Optimization*. Annals of Discrete Mathematics. North Holland, 1991.

Mohammad Ghodsi, Mohammad Taghi Hajiaghayi, Masoud Seddighin, Saeed Seddighin, and Hadi Yami. Fair Allocation of Indivisible Goods: Improvements and Generalizations. In *Proc. EC'18*, pp. 539–556. ACM, 2018. URL https://doi.org/10.1145/3219166.3219238.

Mohammad Ghodsi, Mohammad Taghi Hajiaghayi, Masoud Seddighin, Saeed Seddighin, and Hadi Yami. Fair allocation of indivisible goods: Beyond additive valuations. *Artif. Intell.*, 303:103633, 2022. URL https://doi.org/10.1016/j.artint.2021.103633.

Michel X. Goemans, Nicholas J. A. Harvey, Satoru Iwata, and Vahab S. Mirrokni. Approximating submodular functions everywhere. In *Proc. SODA'09*, pp. 535–544. SIAM, 2009. URL https://doi.org/10.1137/1.9781611973068.59.

Paul W. Goldberg, Kasper Høgh, and Alexandros Hollender. The frontier of intractability for EFX with two agents. *Theor. Comput. Sci.*, 1052:115367, 2025. URL https://doi.org/10.1016/j.tcs.2025.115367.

Daniel Halpern and Nisarg Shah. Fair Division with Subsidy. In *Proc. SAGT'19*, volume 11801 of *LNCS*, pp. 374–389. Springer, 2019. URL https://doi.org/10.1007/978-3-030-30473-7_25.

Jason D. Hartline. Bayesian Mechanism Design. *Found. Trends Theor. Comput. Sci.*, 8(3):143–263, 2013. URL https://doi.org/10.1561/0400000045.

Vishwa Prakash HV, Ayumi Igarashi, and Rohit Vaish. Fair and Efficient Completion of Indivisible Goods. In *Proc. AAAI'25*, pp. 14045–14053. AAAI Press, 2025. URL https://doi.org/10.1609/aaai.v39i13.33537.

David Kurokawa, Ariel D. Procaccia, and Junxing Wang. Fair Enough: Guaranteeing Approximate Maximin Shares. *J. ACM*, 65(2):8:1–8:27, 2018. URL https://doi.org/10.1145/3140756.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

Benny Lehmann, Daniel Lehmann, and Noam Nisan. Combinatorial auctions with decreasing marginal utilities. *Games Econ. Behav.*, 55(2):270–296, 2006. URL https://doi.org/10.1016/j.geb.2005.02.006.

Zihan Li, Pasin Manurangsi, Jonathan Scarlett, and Warut Suksompong. Complexity of round-robin allocation with potentially noisy queries. *Inf. Comput.*, 306:105332, 2025. URL https://doi.org/10.1016/j.ic.2025.105332.

Zihan Li, Yan Hao Ling, Jonathan Scarlett, and Warut Suksompong. Envy-Free Allocation of Indivisible Goods via Noisy Queries. In *Proc. ICML'26*, volume 306 of *Proceedings of Machine Learning Research*. PMLR, 2026. URL https://doi.org/10.48550/arXiv.2602.06361.

Richard J. Lipton, Evangelos Markakis, Elchanan Mossel, and Amin Saberi. On approximately fair allocations of indivisible goods. In *Proc. EC'04*, pp. 125–131. ACM, 2004. URL https://doi.org/10.1145/988772.988792.

Simon Mackenzie and Mashbat Suzuki. Counterexamples to EFX for Submodular and Subadditive Valuations. *CoRR*, abs/2605.06451, 2026. URL https://doi.org/10.48550/arXiv.2605.06451.

Nicholas Mattei and Toby Walsh. PrefLib: A Library for Preferences http://www.preflib.org. In *Proc. ADT'13*, volume 8176 of *LNCS*, pp. 259–270. Springer, 2013. URL https://doi.org/10.1007/978-3-642-41575-3_20.

Luisa Montanari, Ulrike Schmidt-Kraepelin, Warut Suksompong, and Nicholas Teh. Weighted envy-freeness for submodular valuations. *Soc. Choice Welf.*, 65(4):983–1016, 2025. URL https://doi.org/10.1007/s00355-025-01599-6.

Hervé Moulin. *Fair Division and Collective Welfare*. MIT Press, 2003.

Hoon Oh, Ariel D. Procaccia, and Warut Suksompong. Fairly Allocating Many Goods with Few Queries. *SIAM J. Discret. Math.*, 35(2):788–813, 2021. URL https://doi.org/10.1137/20M1313349.

Dominik Peters, Ariel D. Procaccia, and David Zhu. Robust Rent Division. In *Advances in Neural Information Processing Systems 35 (NeurIPS'22)*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/5a1a10c2c2c9b9af1514687bc24b8f3d-Abstract-Conference.html.

Benjamin Plaut and Tim Roughgarden. Communication Complexity of Discrete Fair Division. *SIAM J. Comput.*, 49(1):206–243, 2020a. URL https://doi.org/10.1137/19M1244305.

Benjamin Plaut and Tim Roughgarden. Almost Envy-Freeness with General Valuations. *SIAM J. Discret. Math.*, 34(2):1039–1068, 2020b. URL https://doi.org/10.1137/19M1243977X.

Ariel D. Procaccia, Ben Schiffer, and Shirley Zhang. Honor Among Bandits: No-Regret Learning for Online Fair Division. In *Advances in Neural Information Processing Systems 37 (NeurIPS'24)*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/17bae482066287ce54814514cf12f248-Abstract-Conference.html.

Hugo Steinhaus. The problem of fair division. *Econometrica*, 16(1):101–104, 1948.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer series in statistics. Springer, 2009. URL https://doi.org/10.1007/b13794.

Hal R Varian. Equity, Envy, and Efficiency. *Journal of Economic Theory*, 9(1):63–91, 1974.

Tatsuhito Yamagata and Hanna Sumita. Witness-Certified Fair Division with Comparison Queries. *CoRR*, abs/2608.16109, 2026. URL https://doi.org/10.48550/arXiv.2608.16109.

APPENDIX

Contents

A Valuation Classes and the Hierarchy . . . **16**
definitions of additive, monotone submodular and unit-demand; strictness of the inclusions

B Transfer of Estimation Error . . . **16**
the EF1 robustness lemma, its EFX counterpart, and path independence

C Simulation and Margin Envy-Cycle Elimination . . . **17**
Algorithms 1–3, the consultation number, and the robustness definition

D Two-Agent Cut-and-Choose . . . **17**
the proof of Thm. 4.3 and tightness of re-examination

E Upper Bounds in EEAG . . . **20**
exhaustive search above the barrier; budget and singleton-determined cases

F The Landscape: Supporting Results and Discussion . . . **22**
why unit-demand collapses EFX onto EF

G Open Problems in Detail . . . **22**
unresolved bounds from Table 1 and related questions

H Limitations . . . **25**
what the bounds do not cover

I Extended Related Work . . . **25**
subsidy repairs, other relaxations, learning with samples, misreported values, foundations

J Further Experiments . . . **26**
More than two agents
What the epoch rule buys
Which assumption is actually load-bearing
Non-constructed preference and ranking data

K Experimental Details . . . **29**
Finite-sample uncertainty for the phase diagram
Success rates and paired tests
Consultation counts
Robustness grid
Hyperparameters
Computational verification
Compute

L Omitted Proofs from §3 . . . **32**
Proofs of Lemmas 3.2 and 3.3
From set recovery to coordinate recovery
Per-observation divergence
Transcript divergence under adaptivity
Assouad's lemma and the conclusion
Proof of Prop. 3.4

M Analysis of the Learning Algorithms . . . **35**
Why a gap-driven greedy rule fails
Proof of Thm. 4.2

A Valuation Classes and the Hierarchy

This appendix records the valuation classes referred to in §2 and the inclusions between them.

Definition A.1 (Additive Valuations)**.** A valuation v is *additive* if $v(S) = \sum_{g \in S} v(\{g\})$ for all S.

Definition A.2 (Monotone Submodular Valuations)**.** A valuation $v : 2^{P \cup R} \to \mathbb{R}_{\geq 0}$ is *monotone* if $S \subseteq T$ implies $v(S) \leq v(T)$. It is *submodular* if for all $S \subseteq T$ and $g \notin T$:

$$v(S \cup \{g\}) - v(S) \geq v(T \cup \{g\}) - v(T). \tag{4}$$

We denote the *marginal gain* of g given S by $\Delta_v(g \mid S) = v(S \cup \{g\}) - v(S)$.

Definition A.3 (Unit-Demand Valuations)**.** A valuation is *unit-demand* if $v(S) = \max_{g \in S} v(\{g\})$ for all non-empty S, with $v(\emptyset) = 0$.

Lemma A.4 (Valuation Hierarchy)**.** *Additive, coverage, and unit-demand valuations are all monotone submodular. Moreover, all inclusions are strict: Unit-Demand $\subsetneq$ Coverage $\subsetneq$ Submodular, Additive $\subsetneq$ Submodular.*

Proof. Coverage valuations are monotone and submodular because $S \mapsto \bigcup_{r \in S} C_r$ is monotone and the weight of a union is submodular in the collection; additive valuations are the special case of disjoint C_r. For unit-demand, order the items so that $w_1 \geq \cdots \geq w_m$, take $U = \{u_1, \ldots, u_m\}$ with $C_i = \{u_i, u_{i+1}, \ldots, u_m\}$ and $w(u_i) = w_i - w_{i+1}$, where $w_{m+1} = 0$. The weights are non-negative by the ordering, and for nonempty S with least index $i^* = \min S$ we have $\bigcup_{i \in S} C_i = C_{i^*}$, whose weight telescopes to $w_{i^*} = \max_{i \in S} w_i$. So unit-demand valuations are coverage. Strictness: a coverage valuation with two disjoint singleton sets of equal weight is additive but not unit-demand; the rank function of a uniform matroid of rank 2 on three items is submodular but, being symmetric with $v(\{g\}) = 1$ and $v(M) = 2$, is neither additive nor unit-demand. Nor is it coverage: realizing sets would need $w(C_i) = 1$ and $w(C_i \cup C_j) = 2$, hence $w(C_i \cap C_j) = 0$, giving $w(C_1 \cup C_2 \cup C_3) = 3 \neq 2$. □

Envy gap. For proofs it is convenient to name the quantity compared in (1). Under extension ρ the *envy gap* from a to a' is $\gamma_\rho(a, a') = v_a(B_{a'}) - v_a(B_a)$, so a envies a' exactly when $\gamma_\rho(a, a') > 0$, and ρ is ε-EF when $\gamma_\rho(a, a') \leq \varepsilon$ for every ordered pair.

B Transfer of Estimation Error

Remark B.1 (The same transfer holds for EFX). Nothing in the proof uses the existential quantifier. If $|\hat{v}_a(S) - v_a(S)| \leq \varepsilon'$ on every bundle of the form B_a and $B_{a'} \setminus \{g\}$, and ρ is EFX under $\{\hat{v}_a\}$, then for every $g \in B_{a'}$ the chain $v_a(B_a) \geq \hat{v}_a(B_a) - \varepsilon' \geq \hat{v}_a(B_{a'} \setminus \{g\}) - \varepsilon' \geq v_a(B_{a'} \setminus \{g\}) - 2\varepsilon'$ holds, so ρ is $2\varepsilon'$-EFX under $\{v_a\}$. The universal quantifier passes through unchanged. What separates EF1 from EFX in Thm. 3.1 is therefore not the transfer step but which item the notion allows the verifier to discount, and consequently how fine a structure in the valuations must be resolved before any extension can be certified.

Lemma B.2 (EF1 Robustness)**.** *Let $\{v_a\}_{a \in A}$ be the true valuations and $\{\hat{v}_a\}_{a \in A}$ be estimates satisfying $|\hat{v}_a(S) - v_a(S)| \leq \varepsilon'$ for every agent a and every bundle S of the form B_a or $B_{a'} \setminus \{g\}$ under ρ. If ρ is s-EF1 under $\{\hat{v}_a\}$ for some $s \geq 0$, then ρ is $(s + 2\varepsilon')$-EF1 under $\{v_a\}$. In particular, EF1 under the estimates implies $2\varepsilon'$-EF1 under the truth.*

Proof. Fix agents a, a' with $B_{a'} \neq \emptyset$. Since ρ is s-EF1 under the estimated valuations, there exists a witness $g \in B_{a'}$ such that $\hat{v}_a(B_a) \geq \hat{v}_a(B_{a'} \setminus \{g\}) - s$.

By the estimation guarantee, we have $v_a(B_a) \geq \hat{v}_a(B_a) - \varepsilon'$ and $\hat{v}_a(B_{a'} \setminus \{g\}) \geq v_a(B_{a'} \setminus \{g\}) - \varepsilon'$. Combining these bounds:

$$\begin{aligned} v_a(B_a) &\geq \hat{v}_a(B_a) - \varepsilon' \geq \hat{v}_a(B_{a'} \setminus \{g\}) - s - \varepsilon' \\ &\geq v_a(B_{a'} \setminus \{g\}) - s - 2\varepsilon'. \end{aligned}$$

Thus g witnesses $(s + 2\varepsilon')$-EF1 for the pair (a, a') under the true valuations. □

Remark B.3 (Path Independence). The guarantee depends only on the *final allocation* and the values consulted in certifying it, not on the execution path of the oracle algorithm being simulated. Even if estimation errors lead the simulated algorithm to different choices than it would make with exact values, the output remains $2\varepsilon'$-EF1 under the true valuations whenever it is EF1 under the estimated ones.

C Simulation and Margin Envy-Cycle Elimination

Definition C.1 (Consultation Number)**.** Let $\mathcal{B}$ be an algorithm for EEAG that accesses valuations only through requests for values $v_a(S)$. The *consultation number* $N_{\mathcal{B}}(n, m, |P|, k, \varepsilon')$ is the maximum, over instances with n agents, m pool goods, $|P|$ initial items and budget k, and over all ε'-accurate answer tables, of the number of distinct pairs (a, S) that $\mathcal{B}$ requests.

Definition C.2 (Robustness)**.** Fix a notion $\mathcal{F} \in \{\mathrm{EF1}, \mathrm{EFX}\}$. An oracle algorithm $\mathcal{B}$ for a class $\mathcal{C}$ is (ε', s)-robust for $\mathcal{F}$ if, on every instance in $\mathcal{C}$, whenever every answer $\mathcal{B}$ receives lies within ε' of the true value of the pair requested, $\mathcal{B}$ outputs an extension within budget that is s-$\mathcal{F}$ under the true valuations.

Proof of Thm. 4.1. If $N = 0$ then $\mathcal{B}$ makes no request under any ε'-accurate answer table, in particular the true one, and its output is s-$\mathcal{F}$ by robustness; no observations are needed. So assume $N \geq 1$.

Run the simulation of Alg. 2, drawing T_0 fresh observations for each distinct pair requested and answering with their mean, aborting if an $(N+1)$st distinct pair is requested. Let $R_1, R_2, \ldots$ be the distinct pairs in the order requested, and for $j \geq 1$ let E_j be the event that the jth answer is within ε' of the true value of R_j, taken to hold vacuously if fewer than j pairs are requested. Put $E = \bigcap_{j \leq N} E_j$.

Let $\mathcal{H}_{j-1}$ be the full history before the jth distinct request: the pairs requested so far and the numerical answers returned. Because $\mathcal{B}$ is deterministic, R_j is a measurable function of $\mathcal{H}_{j-1}$, so conditioning on $\mathcal{H}_{j-1} = h$ fixes R_j before its T_0 observations are drawn. Those observations are fresh, independent and $[0,1]$-valued, so Hoeffding gives $\Pr[E_j^{\mathrm{c}} \mid \mathcal{H}_{j-1} = h] \leq 2\exp(-2T_0\varepsilon'^2) \leq \delta/N$ for every such h. Averaging over h with the tower property, $\Pr[E_j^{\mathrm{c}} \mid E_1, \ldots, E_{j-1}] \leq \delta/N$, and a union bound over $j \leq N$ gives $\Pr[E] \geq 1 - \delta$.

The argument is not circular. Suppose an $(N+1)$st distinct request were issued on E. Keep the answers already returned on the first N distinct pairs and extend this partial table by assigning the exact value $v_a(S)$ to every other pair, giving an answer table that is ε'-accurate everywhere. Because $\mathcal{B}$ is deterministic, running it against this table reproduces the same prefix and so issues the same $(N+1)$st distinct request, contradicting the definition of N. Hence the abort never fires on E. Off E it may, and the output is then arbitrary but still feasible; that case is inside the δ we have already paid.

On E, every answer $\mathcal{B}$ receives is within ε' of the truth, so robustness for $\mathcal{F}$ makes its output an s-$\mathcal{F}$ extension within budget. The simulation draws T_0 observations for each of at most N distinct pairs, so the total is $NT_0 = O(N\varepsilon'^{-2}\log(N/\delta))$. □

D Two-Agent Cut-and-Choose

Proof of Prop. 4.4.

Proof. Write $I = \mathbf{1}[r_1 \in S_\theta]$ and $\Delta = \frac{3}{8m}$. Both bundles meet R, so each collects the common term $\frac{3}{16}$, and $|(R \setminus \{r_1\}) \cap S_\theta| = \frac{m}{2} - I$, whence

$$v_1^\theta(B_{a_1}) = \tfrac{9}{16} - \eta + \Delta I, \qquad v_1^\theta(B_{a_2}) = \tfrac{3}{8} + \eta - \Delta I.$$

Their difference is $\frac{3}{16} - 2\eta + 2\Delta I \geq \frac{3}{16} - 2\eta = \frac{95}{512} > 0$, so a_1 does not envy a_2. Agent a_2 holds g° and not g^*, so $v_2(B_{a_2}) = \frac{1}{2} = v_2(B_{a_1})$, and it does not envy either. □

Algorithm 1 Two-agent cut-and-choose from estimates

Require: Value access $\hat{v}_1, \hat{v}_2$; accuracy parameter $\varepsilon_0 = \varepsilon/8$
$\Pi_1 \leftarrow R, \Pi_2 \leftarrow \emptyset, \tau \leftarrow 6\varepsilon_0$
repeat {each iteration is an *epoch*; stop after $\lceil 4/\varepsilon \rceil + 1$ epochs}
  for $(\Pi, \Pi^{\mathrm{c}}) \in \big((\Pi_1, \Pi_2), (\Pi_2, \Pi_1)\big)$ **do**
    for each g in Π, taken in a fixed order and skipped if it has already left Π **do**
      if $\hat{v}_1(\Pi \setminus \{g\}) > \hat{v}_1(\Pi^{\mathrm{c}}) + \tau$ **then**
        move g from Π to Π^{c}
        if $\hat{v}_1(\Pi^{\mathrm{c}} \cup \{g\}) \geq \hat{v}_1(\Pi \setminus \{g\}) - 2\varepsilon_0$ **(evaluated before the move) then**
          end this epoch and begin a new one {strict move; retirements are cleared}
        end if
      end if{otherwise g is *retired* for this epoch and not consulted again}
    end for
    if this pass moved at least one item **then**
      stop {this pass made only weak moves, so Π^{c} is now the poorer piece}
    end if
  end for
until an epoch ends with no move, or a pass ends with only weak moves
a_2 takes Π_1 if $\hat{v}_2(\Pi_1) \geq \hat{v}_2(\Pi_2)$ and Π_2 otherwise; a_1 takes the remaining piece

Algorithm 2 Simulate-with-Estimates$(\mathcal{B}, \varepsilon', \delta)$

Require: Oracle algorithm $\mathcal{B}$ with consultation number N; observation access to $\{v_a\}$; required accuracy ε' and failure probability δ, both in $(0, 1)$
Ensure: Extension ρ
if $N = 0$ **then**
  Run $\mathcal{B}$, which makes no request
  return its output
end if
$T_0 \leftarrow \lceil \ln(2N/\delta)/(2\varepsilon'^2) \rceil$
Run $\mathcal{B}$; if it requests an $(N+1)$st distinct pair, abort and return an arbitrary extension feasible for the instance class. Whenever $\mathcal{B}$ requests $v_a(S)$ for a pair (a, S) not requested before, draw T_0 observations of (a, S), store their mean as $\hat{v}_a(S)$, and return it; on a repeated request return the stored value.
return the extension output by $\mathcal{B}$

Theorem (Thm. 4.3 restated). For $\varepsilon, \delta \in (0, 1)$ and $m \geq 2$, in the standard setting with two agents and arbitrary monotone valuations, the epoch cutter of Alg. 1, simulated as in Thm. 4.1 with $\mathcal{F} = \mathrm{EFX}$, outputs an ε-EFX allocation with probability at least $1 - \delta$ using

$$O\Big(\frac{m}{\varepsilon^3} \log \frac{m}{\varepsilon\delta}\Big)$$

value observations.

Proof. Set $\varepsilon_0 = \varepsilon/8$, $\tau = 6\varepsilon_0$, and write $f = v_1$. We argue deterministically: fix any answer table under which every consulted estimate lies within ε_0 of the truth, and show that Alg. 1 terminates on it with an ε-EFX partition after at most $N = O(m/\varepsilon)$ distinct consultations. The theorem then follows from Thm. 4.1 with $\mathcal{F} = \mathrm{EFX}$, $\varepsilon' = \varepsilon_0$ and $s = \varepsilon$, which supplies the sampling, the failure probability and the count $NT_0 = O(m\varepsilon^{-3} \log(m/(\varepsilon\delta)))$. A *move* transfers g from Π to Π^{c} because $\hat{v}_1(\Pi \setminus \{g\}) > \hat{v}_1(\Pi^{\mathrm{c}}) + \tau$; it is *strict* if $\hat{v}_1(\Pi^{\mathrm{c}} \cup \{g\}) \geq \hat{v}_1(\Pi \setminus \{g\}) - 2\varepsilon_0$ and *weak* otherwise.

At every move, Π^{c} is truly the poorer piece. We have $f(\Pi) \geq f(\Pi \setminus \{g\}) > \hat{v}_1(\Pi \setminus \{g\}) - \varepsilon_0 > \hat{v}_1(\Pi^{\mathrm{c}}) + \tau - \varepsilon_0 \geq f(\Pi^{\mathrm{c}}) + 4\varepsilon_0$.

Within an epoch, moves go one way. After a weak move, $f(\Pi^{\mathrm{c}} \cup \{g\}) < \hat{v}_1(\Pi^{\mathrm{c}} \cup \{g\}) + \varepsilon_0 < \hat{v}_1(\Pi \setminus \{g\}) - \varepsilon_0 \leq f(\Pi \setminus \{g\})$, so the shrunken Π remains truly richer. Consequently every move in an epoch is out of the piece that was richer at the epoch's first move, that piece only loses items,

Algorithm 3 Margin Envy-Cycle Elimination (oracle algorithm $\mathcal{B}_{\mathrm{ECE}}$)

Require: Value access $\hat{v}$; items $r_1, \dots, r_m$; margin τ; write $\hat{\gamma}(a, a') = \hat{v}_a(B_{a'}) - \hat{v}_a(B_a)$ for the estimated envy of a toward a'

1: $B_a \leftarrow \emptyset$ for all a
2: **for** $t = 1, \dots, m$ **do**
3: Build the graph G on A with an edge $a \to a'$ whenever $\hat{\gamma}(a, a') > \tau$
4: **while** G contains a directed cycle; let $a^{(1)} \to \cdots \to a^{(\ell)} \to a^{(1)}$ be the lexicographically least one **do**
5: Rotate: $B_{a^{(j)}} \leftarrow B_{a^{(j+1)}}$ for all j (indices modulo ℓ); rebuild G
6: **end while**
7: Choose the lowest-indexed agent a with no incoming edge in G; $B_a \leftarrow B_a \cup \{r_t\}$
8: **end for**
9: **return** ρ with $\rho(a) = B_a$

and the other only gains them. In particular the potential $\Phi = \min(f(\Pi_1), f(\Pi_2))$, which equals the true value of the poorer piece, does not decrease at a weak move.

Retirement is sound. Suppose g is retired during a pass on Π when the pieces are Π', Π'^{c}: then $\hat{v}_1(\Pi' \setminus \{g\}) \le \hat{v}_1(\Pi'^{\mathrm{c}}) + \tau$, so $f(\Pi' \setminus \{g\}) \le f(\Pi'^{\mathrm{c}}) + \tau + 2\varepsilon_0 = f(\Pi'^{\mathrm{c}}) + \varepsilon$. Later in the same epoch the pieces are $\Pi \subseteq \Pi'$ and $\Pi^{\mathrm{c}} \supseteq \Pi'^{\mathrm{c}}$, and monotonicity gives $f(\Pi \setminus \{g\}) \le f(\Pi' \setminus \{g\}) \le f(\Pi'^{\mathrm{c}}) + \varepsilon \le f(\Pi^{\mathrm{c}}) + \varepsilon$. So a retired item satisfies the ε-EFX condition for the rest of the epoch without being consulted again.

Every item of the final Π was consulted. A pass iterates over the items of Π as they stood when the pass began, in a fixed order, skipping only items that have since been moved; it does not recompute the order after a move. So every item that remains in Π at the end of the pass was consulted while in Π, and was retired. (An implementation that re-sorts Π after each move can skip an item, and the skipped item is neither retired nor moved; the guarantee below then fails. We found this in testing.)

The poorer piece is certified without consultation. If a pass on Π makes at least one move and no strict move, then at its end $f(\Pi^{\mathrm{c}}) < f(\Pi)$ by the second paragraph, so for every $g' \in \Pi^{\mathrm{c}}$, $f(\Pi^{\mathrm{c}} \setminus \{g'\}) \le f(\Pi^{\mathrm{c}}) < f(\Pi) \le f(\Pi) + \varepsilon$. Together with the two preceding paragraphs, both pieces are ε-EFX toward each other under v_1 when the algorithm stops after such a pass. If instead an epoch ends with no move at all, every item of each piece was consulted against the final pieces and found not to violate, and the same bound applies.

Strict moves raise the potential. At a strict move, $f(\Pi^{\mathrm{c}} \cup \{g\}) \ge \hat{v}_1(\Pi^{\mathrm{c}} \cup \{g\}) - \varepsilon_0 \ge \hat{v}_1(\Pi \setminus \{g\}) - 3\varepsilon_0 > \hat{v}_1(\Pi^{\mathrm{c}}) + \tau - 3\varepsilon_0 \ge f(\Pi^{\mathrm{c}}) + 2\varepsilon_0$, and $f(\Pi \setminus \{g\}) > f(\Pi^{\mathrm{c}}) + 4\varepsilon_0$; so Φ rises by at least $2\varepsilon_0 = \varepsilon/4$. Since $0 \le \Phi \le 1$ and Φ never decreases, there are at most $4/\varepsilon$ strict moves, hence at most $4/\varepsilon + 1$ epochs, so the epoch cap of Alg. 1 never binds under an ε_0-accurate answer table; under any other table the cap makes the algorithm halt regardless.

Consultation count. In an epoch each item is consulted at most once as $\Pi \setminus \{g\}$, the value $\hat{v}_1(\Pi^{\mathrm{c}})$ is consulted once per distinct Π^{c}, of which there are at most one more than the number of moves, and $\hat{v}_1(\Pi^{\mathrm{c}} \cup \{g\})$ once per move. With at most m moves per epoch this is at most $2(m + (m+1) + m) = 6m + 2 \le 7m$ consultations per epoch for $m \ge 2$, so $N \le 7m(4/\varepsilon + 1) + 2 = O(m/\varepsilon)$. Every step above uses only that each consulted estimate is within ε_0 of the truth, so the count holds under every ε_0-accurate answer table, which is what Thm. 4.1 requires of a consultation number. In this count, the final 2 is the comparison made by the chooser.

The chooser. If a_2 takes Π_1 then $\hat{v}_2(\Pi_1) \ge \hat{v}_2(\Pi_2)$, so $v_2(\Pi_1) \ge v_2(\Pi_2) - 2\varepsilon_0 \ge v_2(\Pi_2 \setminus \{g\}) - \varepsilon/4$ for every g; symmetrically for Π_2. She is $(\varepsilon/4)$-EFX, and a_1 is ε-EFX toward whichever piece a_2 took, by the fourth paragraph. □

Proposition D.1 (Re-examination costs $\Theta(m^2)$)**.** *Fix $\varepsilon \in (0, 1)$ and set $\varepsilon_0 = \varepsilon/8$ and $\tau = 6\varepsilon_0$ as in Alg. 1. Consider the variant of Alg. 1 that, after every move, restarts its scan over all items of the current Π. For every even $m \ge 4$ there is a monotone valuation for the cutter a_1 (the chooser's valuation may be any monotone valuation) and a choice of estimates, each within ε_0 of the truth, on which this variant consults $\Omega(m^2)$ distinct bundles, although the initial partition is already ε-EFX under the cutter's valuation.*

Proof. Fix two *anchor* items and let the remaining $m-2$ be *movers*. Set $f(S) = (\tau+\varepsilon_0)\cdot \mathbf{1}[S \text{ contains both anchors}]$; f is monotone with range $[0, 7\varepsilon_0] \subseteq [0,1]$. From $\Pi_1 = R$, $\Pi_2 = \emptyset$, removing a mover leaves $f(\Pi_1 \setminus \{g\}) = \tau+\varepsilon_0$ while $f(\Pi_2) = 0$, so every mover has true bracket $\tau+\varepsilon_0$, inside the band $(\tau, \tau+2\varepsilon_0]$ on which estimates within ε_0 may fall on either side of the threshold; removing an anchor drops the value to 0, so anchors are never candidates. Movers are worthless to both pieces, so moving one changes no bracket. Choose estimates so that every bundle consisting of movers alone reads exactly 0, its true value, and so that, each time the variant scans the movers of the current Π_1, every mover but the last examined reads τ and the last reads $\tau+\varepsilon_0$; both readings lie within ε_0 of the true value $\tau+\varepsilon_0$, and against $\hat{v}(\Pi_2) = 0$ only the last exceeds the threshold τ. Each scan then moves exactly one mover after consulting all the others, the bundles consulted are new because Π_1 has changed, and the process continues until Π_1 holds only the anchors, which costs $\sum_{j=2}^{m-1} j = \Omega(m^2)$ consultations in all. Meanwhile $f(\Pi_1 \setminus \{g\}) - f(\Pi_2) = 7\varepsilon_0 < \varepsilon$ throughout, so $(R, \emptyset)$ was ε-EFX from the start. Alg. 1 itself consults each mover once, retires or moves it, and stops after $O(m)$ consultations on the same instance. □

E Upper Bounds in EEAG

The next bound assumes EF-feasibility, which is stronger than the standing EF1-feasibility assumption of §2: it asks for an exactly envy-free extension within budget, not merely an EF1 one.

Proposition E.1 (Upper Bound for ε-EF and ε-EFX)**.** *Let $\varepsilon, \delta \in (0,1)$ and consider EEAG with any fixed σ, budget k, and arbitrary monotone valuations, and suppose the instance is* EF-feasible*: some extension within budget is exactly envy-free. Put*

$$N \;=\; n^2 \sum_{j=0}^{k} \binom{m}{j},$$

which reduces to $n\sum_{j\le k}\binom{m}{j}$ in the standard setting, where all initial bundles are empty. Then ε-EF, and therefore ε-EFX, is achievable within budget with probability at least $1-\delta$ using $N\cdot\lceil 8\varepsilon^{-2}\ln(2N/\delta)\rceil$ value observations. In particular the cost is $O\big(n^2 m\,\varepsilon^{-2}\log(nm/\delta)\big)$ when $k=1$, and $O\big(n^2 m^k \varepsilon^{-2}(k\log m + \log(n/\delta))\big)$ for any constant k.

Proof. Every bundle that can arise under an extension within budget has the form $\sigma(a')\cup S'$ for some agent a' and some $S'\subseteq R$ with $|S'|\le k$, so there are at most $n\sum_{j\le k}\binom{m}{j}$ achievable bundles, and the ε-EF certificate may require any agent's value for any of them; the number of distinct pairs (agent, achievable bundle) is therefore at most N. When every $\sigma(a')$ is empty the achievable bundles do not depend on a' and the count drops by a factor of n. Estimate each to accuracy $\varepsilon/4$, drawing $\lceil 8\varepsilon^{-2}\ln(2N/\delta)\rceil$ observations per pair; by Hoeffding and a union bound over N pairs, all estimates are accurate simultaneously with probability at least $1-\delta$.

On that event, let ρ^* be an exactly envy-free extension within budget, which exists by EF-feasibility. For all a, a' we have $\hat{v}_a(B^*_a) \ge v_a(B^*_a) - \varepsilon/4 \ge v_a(B^*_{a'}) - \varepsilon/4 \ge \hat{v}_a(B^*_{a'}) - \varepsilon/2$, so ρ^* is $(\varepsilon/2)$-EF under $\hat{v}$ and the search is nonempty. Exhaustive search over extensions within budget returns some ρ that is $(\varepsilon/2)$-EF under $\hat{v}$, and reversing the same chain shows ρ is ε-EF under v. Monotonicity gives $v_a(B_{a'}\setminus\{g\}) \le v_a(B_{a'})$ for every g, so ε-EF implies ε-EFX. For the two special cases use $\sum_{j\le k}\binom{m}{j} \le (em/k)^k$, and $N = n^2(m+1)$ when $k=1$. □

Corollary E.2 (The barrier fails at a unit budget)**.** *The conclusion of Thm. 3.1 cannot hold at $k=1$: on any EF-feasible instance with a budget of one item and any $\varepsilon, \delta \in (0,1)$, ε-EFX is achievable from $O(n^2 m\varepsilon^{-2}\log(nm/\delta))$ observations, which is $o(m^2)$ for fixed n, ε and δ.*

Proof. Take $k=1$ in Prop. E.1: then $N = n^2\big(\binom{m}{0}+\binom{m}{1}\big) = n^2(m+1)$ and the observation count is $n^2(m+1)\lceil 8\varepsilon^{-2}\ln(2n^2(m+1)/\delta)\rceil = O(n^2 m\varepsilon^{-2}\log(nm/\delta))$, which is $o(m^2)$ for fixed n, ε, δ. □

Corollary E.3 (EF1 in EEAG for Arbitrary Monotone Valuations)**.** *Let $\varepsilon, \delta \in (0,1)$ and consider EEAG with any fixed σ, budget k, and arbitrary monotone valuations.*

1. *Under EF-feasibility, ε-EF1 is achievable within budget with probability $1-\delta$ using $N\lceil 8\varepsilon^{-2}\ln(2N/\delta)\rceil$ observations with $N = n^2\sum_{j\le k}\binom{m}{j}$.*
2. *Under the standing EF1-feasibility assumption of §2 alone, the same holds with $N = n^2\,(1+|P|+k)\sum_{j\le k}\binom{m}{j}$.*

Both are $O\big(n^2m(|P|+2)\varepsilon^{-2}\log(n^2m(|P|+2)/\delta)\big)$ for $k=1$ and polynomial in m for any constant k.

Proof. (1) On the event of Prop. E.1 the returned extension is ε-EF, and monotonicity gives $v_a(B_{a'}\setminus\{g\}) \le v_a(B_{a'}) \le v_a(B_a)+\varepsilon$ for every g, so ε-EF1 holds with any witness.

(2) An ε-EF1 certificate consults $v_a(B_a)$ and $v_a(B_{a'}\setminus\{g\})$; the bundles $B_{a'}\setminus\{g\}$ range over the sets $(\sigma(a')\cup S')\setminus\{g\}$ with $|S'|\le k$ and $g\in\sigma(a')\cup S'$, at most $(|P|+k)$ removals per achievable bundle, giving the stated N. Estimate every pair to accuracy $\varepsilon/4$. Let ρ^* be the EF1 extension guaranteed by the standing EF1-feasibility assumption of §2, with witness g for the pair (a,a'); then $\hat v_a(B_a^*) \ge v_a(B_a^*)-\varepsilon/4 \ge v_a(B_{a'}^*\setminus\{g\})-\varepsilon/4 \ge \hat v_a(B_{a'}^*\setminus\{g\})-\varepsilon/2$, so ρ^* is $(\varepsilon/2)$-EF1 under $\hat v$ and exhaustive search returns some ρ that is; Lem. B.2 with $s=\varepsilon/2$ and $\varepsilon'=\varepsilon/4$ makes ρ ε-EF1 under v. □

Proposition E.4 (Singleton-Determined Classes)**.** *Let $\varepsilon,\delta\in(0,1)$ and consider EEAG with any fixed σ, any budget k, and the standing EF1-feasibility assumption of §2. Let $q=|P|+m$.*

1. *For unit-demand valuations, ε-EF1 is achievable with probability $1-\delta$ using $O\big(\frac{nq}{\varepsilon^2}\log\frac{nq}{\delta}\big)$ observations; and if it admits an s-EFX extension within budget for some $0\le s<\varepsilon$, then ε-EFX is achievable using $O\big(\frac{nq}{(\varepsilon-s)^2}\log\frac{nq}{\delta}\big)$ observations.*
2. *For additive valuations, ε-EF1 is achievable with probability $1-\delta$ using $O\big(\frac{nq^3}{\varepsilon^2}\log\frac{nq}{\delta}\big)$ observations.*

The algorithms estimate all nq singleton values and then search over extensions within budget; that search enumerates $O\big(n^k\binom{m}{\le k}\big)$ candidates, each checked in time polynomial in n and q, so it is polynomial-time for fixed k and superpolynomial once k grows with m.

Proof. Estimate each $v_a(\{g\})$, $g\in P\cup R$, to accuracy ε_0 with $\lceil(2/\varepsilon_0^2)\ln(2nq/\delta)\rceil$ observations; a union bound gives uniform accuracy with probability $1-\delta$. For unit-demand, $\hat v_a(\emptyset)=0$ and $\hat v_a(S) := \max_{g\in S}\hat v_a(\{g\})$ for nonempty S is within ε_0 of v_a everywhere; take $\varepsilon_0=\varepsilon/4$. For additive, $\hat v_a(S) := \sum_{g\in S}\hat v_a(\{g\})$ is within $|S|\varepsilon_0\le q\varepsilon_0$; take $\varepsilon_0=\varepsilon/(4q)$. In either case $|\hat v_a(S)-v_a(S)|\le\varepsilon/4$ for all (a,S). By feasibility there is an extension ρ^* within budget that is EF1 under v, hence $\varepsilon/2$-EF1 under $\hat v$; exhaustive search finds some extension ρ within budget that is $\varepsilon/2$-EF1 under $\hat v$, and Lem. B.2 (applied with slack $\varepsilon/2$ and error $\varepsilon/4$) gives that ρ is ε-EF1 under v. For the EFX clause, estimate the singletons to accuracy $e=(\varepsilon-s)/4$ instead. An s-EFX extension ρ^* under v is $(s+2e)$-EFX under $\hat v$, so exhaustive search returns some ρ that is $(s+2e)$-EFX under $\hat v$; for every ordered pair and every g in the envied bundle, $v_a(B_a)\ge\hat v_a(B_a)-e\ge\hat v_a(B_{a'}\setminus\{g\})-s-3e\ge v_a(B_{a'}\setminus\{g\})-s-4e = v_a(B_{a'}\setminus\{g\})-\varepsilon$. The observation counts follow from the choices of ε_0. □

Proposition E.5 (The margin exponent is tight)**.** *Let $\varepsilon=\frac14$ and $0<\mu<\frac18$, and put $s=\varepsilon-\mu$. There are two unit-demand EEAG instances with $n=2$, $|P|=2$, $m=2$, budget $k=1$ and Bernoulli value observations, each admitting an s-EFX extension, on at least one of which any algorithm that outputs an ε-EFX extension with probability at least $2/3$ on both needs $\Omega(\mu^{-2})$ observations.*

Proof. Give a_2 both fixed items, $P=\{x,y\}$, and let $R=\{r_1,r_2\}$. Set $v_1(\{x\})=\frac12$ and $v_1(\{y\})=0$, and $v_2(\{x\})=v_2(\{y\})=1$, $v_2(\{r\})=0$ for $r\in R$, so a_2 never envies; each queried pair (a,S) returns an independent $\mathrm{Bernoulli}(v_a(S))$ observation. The empty extension is exactly EF1, since dropping x leaves a_1 the value 0 of $\{y\}$, so the instance meets the standing feasibility assumption. In H_+ put $v_1(\{r_1\})=\frac14+\mu$ and $v_1(\{r_2\})=\frac14-\mu$; in H_- swap the two.

Both are s-EFX-feasible. In H_+ give r_1 to a_1. Deleting y leaves $\{x\}$, worth $\frac12$ to a_1, so s-EFX asks $v_1(B_{a_1})\ge\frac12-s=\frac14+\mu$, which holds with equality; deleting x leaves $\{y\}$, worth 0, which is slack. H_- is symmetric.

Only the high item works. Deleting y makes the target ask $v_1(B_{a_1}) \geq \frac{1}{2} - \varepsilon = \frac{1}{4}$, and the low item gives $\frac{1}{4} - \mu < \frac{1}{4}$. So an ε-EFX output must hand a_1 the high pool item, which differs between H_+ and H_-.

Each observation carries $O(\mu^2)$ information. Every query on a_2 has a value independent of θ. For a_1, a bundle containing x has value $\frac{1}{2}$ under both hypotheses, one containing both pool items but not x has value $\frac{1}{4} + \mu$ under both, and one containing neither pool item has value 0 or $\frac{1}{2}$ under both; only a bundle containing exactly one pool item and not x separates them, with Bernoulli means $\frac{1}{4} \pm \mu$. Since $\frac{1}{4} - \mu \geq \frac{1}{8}$, the Bernoulli divergence is at most $(2\mu)^2/\left(\frac{1}{8} \cdot \frac{5}{8}\right) = \frac{256}{5}\mu^2$. By the divergence decomposition for adaptive sampling (Lattimore & Szepesvári, 2020, Lem. 15.1), T observations give $D(\mathbb{P}_{H_+} \| \mathbb{P}_{H_-}) \leq \frac{256}{5} T\mu^2$. Success probability $2/3$ on both forces total variation at least $\frac{1}{3}$, so Pinsker's inequality gives $\frac{256}{5} T\mu^2 \geq \frac{2}{9}$ and $T = \Omega(\mu^{-2})$. □

F The Landscape: Supporting Results and Discussion

Proposition F.1 (Unit-demand collapses EFX onto EF)**.** *Let v_a be unit-demand and let ρ be an extension under which $|B_{a'}| \geq 2$. Then ρ satisfies the ε-EFX condition for the ordered pair (a, a') if and only if it satisfies the ε-EF condition for that pair.*

Proof. Write $g_1 \in \arg\max_{g \in B_{a'}} v_a(\{g\})$, so $v_a(B_{a'}) = v_a(\{g_1\})$. Since $|B_{a'}| \geq 2$, there is some $g \in B_{a'}$ with $g \neq g_1$, and for that g we have $v_a(B_{a'} \setminus \{g\}) = v_a(\{g_1\}) = v_a(B_{a'})$. So the ε-EFX condition, which quantifies over all g, includes the requirement $v_a(B_a) \geq v_a(B_{a'}) - \varepsilon$, which is ε-EF. The converse is immediate from monotonicity. □

G Open Problems in Detail

We discuss selected unresolved bounds from Table 1 together with several related questions. Three matter most: a lower bound for ε-EF1; an oracle algorithm with polynomial consultation number for EF1 in EEAG, or a proof that none exists; and whether coverage instances harder than $\{\mathcal{I}_m(\theta)\}$ push the barrier above $\Theta(m^2)$. Each is discussed in turn below, and Appx. H sets out what the results do not cover.

A lower bound for ε-EF1. A tight bound would have to avoid the trap that constant-value structure sets. If a family is to force $\Omega(nm/\varepsilon^2)$ observations, no fixed extension may be ε-EF1 on all of it, and the budget must be tight against the worst-case requirement; an information-theoretic step must also account for observations that are informative about several agents at once.

One lower bound is inherited rather than proved here. The noiseless value-query model is the special case of Def. 2 in which the observation is deterministic, so any query lower bound is an observation lower bound. Oh et al. (2021, Thm. 5.5) show that for binary valuations with $m \geq n^\alpha$ for a constant $\alpha > 1$, any *deterministic* algorithm computing an EF1 allocation uses $\Omega(n \log m)$ value queries, by an adversary argument that halves a candidate set at each answer. Their valuations are integer-valued with bundles worth at most 2, so rescaling by $\frac{1}{2}$ puts them in our class and leaves every EF1 violation worth at least $\frac{1}{2}$; for $\varepsilon < 1/2$ the notions ε-EF1 and EF1 therefore coincide on them, and the bound carries over. Two caveats: it is restricted to deterministic algorithms, whereas Thm. 3.1 holds against randomized adaptive ones, and at fixed ε and δ it is a factor of about m from Thm. 4.2 while saying nothing about ε.

Open Problem G.1 (Lower bounds for EF1). Is the sample complexity of ε-EF1 in the standard setting with monotone valuations $\Omega(nm/\varepsilon^2)$, matching Thm. 4.2 up to logarithmic factors? We are not confident that it is, and the reason is instructive. With exact values and two agents, EF1 takes only $O(\log m)$ queries, by arranging the goods on a line and binary searching for a cut point that its owner is content to give away (Oh et al., 2021, Thm. 3.1), and $\Omega(\log m)$ for deterministic algorithms (Oh et al., 2021, Prop. 3.2). Envy-cycle elimination, which Thm. 4.2 simulates, is wasteful by comparison for small n. Whether that binary search survives noise is exactly the sort of question the noisy-search literature addresses, and if it does, the truth for constant n may be polylogarithmic in m rather than linear. A second reason for doubt is that the object being sought is small. Since EF1 is attainable by a contiguous allocation for every monotone profile, the contiguous allocations

form a hitting set of size $\binom{m}{n-1}n!$, so an EF1 allocation can always be named in $O(n \log m)$ bits (Feige, 2025). An $\Omega(nm)$ sample bound would therefore have to arise entirely from the cost of identifying which member of a small family works, not from the size of the answer. It is worth adding that the corresponding gap is open in a setting with no noise at all: for non-identical additive valuations the best communication upper bound for EF1 is the $O(m \log m)$ of round-robin against a description complexity of $O(n \log m)$, and whether $\mathrm{poly}(n, \log m)$ suffices is open already for four agents (Yamagata & Sumita, 2026). Our gap is the statistical shadow of that one. Three requirements must hold simultaneously, and they pull against each other. The budget, here the constraint that all m items be allocated, must be tight against the worst-case parameter setting; EF1-feasibility must hold under that same setting; and the per-observation divergence between neighboring parameters must be bounded for observations on arbitrary bundles, as in Lem. L.3, not merely for singletons. Loosening any one of them admits a fixed extension that is ε-EF1 across the family, collapsing the bound.

Open Problem G.2 (EF1 in EEAG at a large budget). Cor. E.3 prices ε-EF1 in EEAG at $\widetilde{O}(n^2(1 + |P| + k)\binom{m}{\leq k}/\varepsilon^2)$ under the EF1-feasibility assumed throughout, which is polynomial for constant k and superpolynomial when k grows with m. Is that dependence necessary, or is there an oracle algorithm that reaches an EF1 extension within a large budget while consulting polynomially many bundle values? The obstacle is algorithmic rather than statistical: rotation is unavailable once σ is fixed, the rotation-free greedy is incorrect (Appx. M.1), and we know of no substitute. Computational hardness does not settle the question, since sample complexity permits exponential computation: for the related problem of completing a partial allocation, where every remaining good must be assigned, deciding EF1-completability is NP-hard even for binary additive valuations (HV et al., 2025), and a budgeted extension need not allocate the whole pool, so that hardness does not transfer directly either.

The two notions we separate are the endpoints of a scale. Yamagata & Sumita (2026), working in a comparison-query model where an EF1 certificate must be produced explicitly because it cannot otherwise be verified, define k-witness EF1, in which bundles carry a fixed order and the last k goods of every bundle, when present, must each serve as a witness for every agent. We use the pair-specific weakening that asks, for each ordered pair (a, a'), that the envy toward a' be removable by each of $\min\{t, |B_{a'}|\}$ distinct goods, writing t for the count since k is our budget; $t = 1$ is then EF1, and taking t at least the largest bundle size gives EFX. On our hard family Thm. 3.1 already prices $t = 2$ at $\Omega(m^2)$: two witnesses include one other than g^*, and dropping it leaves g^* in a_2's bundle, so $v_1^\theta(\rho(a_1)) \geq \frac{3}{8} - \eta - \varepsilon$, which is the inequality the proof of Lem. 3.3 runs on. In the standard setting Thm. 4.3 prices even $t = m$, the EFX endpoint there, at $\widetilde{O}(m/\varepsilon^3)$ for two agents. The question is what happens between these settings.

Open Problem G.3 (The price of extra witnesses). Fix EEAG with a nonempty initial allocation, restricted to instances where a t-witness EF1 extension exists. What is the sample complexity of ε-t-witness EF1 as a function of t, and for which initial allocations does the quadratic barrier first appear?

Open Problem G.4 (The cost of ε-EFX at a large budget, beyond the family). On the family $\{\mathcal{I}_m(\theta)\}$, for every fixed $\varepsilon \in (0, 1/128]$, the sample complexity of ε-EFX at constant success probability is $\Theta(m^2)$ (Thm. 3.1 and Prop. 3.4). For two agents with EF-feasible coverage valuations and budget $k = m/2$, the general bounds are $\Omega(m^2)$ from that family and $2^{O(m)}$ from Prop. E.1. Where in that range the worst case over coverage valuations lies is open.

The family settles its own exponent, so what remains is whether harder coverage instances exist. The two endpoints fail for different reasons.

The exponential in Prop. E.1 comes from enumerating the $\binom{m}{\leq k}$ bundles an extension might create and pricing each. That is a cost of the proof, not a cost that has been shown necessary. Phrased in the language of Thm. 4.1, the question is whether there is an oracle algorithm that finds an EFX extension within a large budget while consulting only polynomially many bundle values. Such an algorithm may run in exponential time, so the PLS-completeness of exact EFX for two agents with submodular valuations (Goldberg et al., 2025) does not rule it out; what it does rule out, unless PLS = P, is finding one by local search in polynomial time. We know of no obstruction to a polynomial consultation number, and we know of no such algorithm.

On the lower side, our construction hides one bit per pair in a per-item perturbation of size $\Theta(1/m)$, the shape Li et al. (2026) call the naive hard instance in their setting, and Prop. 3.4 confirms it cannot give more than m^2: a single observation is a linear measurement of every hidden bit, and $O(m^2)$

such measurements recover a constant fraction of them. Lem. 3.3 asks for no more, because v_1^θ accumulates the hidden bits monotonically, so each correct coordinate buys margin and a constant tolerance can be met while $\Theta(m)$ coordinates are wrong. In the instance of Li et al. (2026) the envy is a signed sum, fluctuations must be canceled rather than accumulated, and all but $O(\sqrt{m})$ coordinates must be right; their remedy is a second population of items that both agents value far above or far below the rest. A coverage instance in which a constant additive tolerance still forces recovery to within $o(m)$ coordinates, or in which the hidden structure is not linearly measurable, would raise the general lower bound above m^2.

A second reading points the same way. The exponential query bounds for submodular valuations hide an arbitrary function in the middle layer of the hypercube, perturbing a nearly constant value by $\Theta(1/m)$ and thereby encoding $2^{\Theta(m)}$ degrees of freedom (Plaut & Roughgarden, 2020b; Feldman et al., 2013). We perturb at the same scale but encode $m/2$ bits, aggregated linearly, and Prop. 3.4 shows that linear aggregation is exactly what makes them cheap to read. The distance from m^2 to $2^{O(m)}$ is a question of how much is hidden, and how nonlinearly.

Open Problem G.5 (Other settings for the barrier). For two agents in the standard setting, Thm. 4.3 bounds ε-EFX by $\widetilde{O}(m/\varepsilon^3)$ and Thm. 4.2 bounds ε-EF1 by $\widetilde{O}(m/\varepsilon^2)$, both near-linear in m. Is the $1/\varepsilon$ between them real, and is there any lower bound beyond the $\Omega(\log m)$ that holds for deterministic algorithms? Does a super-linear barrier for ε-EFX hold in the standard setting for $n \geq 3$, restricted to instances that admit an EFX allocation since some do not (Akrami et al., 2026)? For unit-demand valuations Prop. E.4 settles the whole positive-margin regime, at $O(nq(\varepsilon - s)^{-2} \log(nq/\delta))$ when an s-EFX extension exists; some margin is necessary: by Prop. E.5 the cost grows as $(\varepsilon - s)^{-2}$, so it is unbounded over instances whose slack $\varepsilon - s$ approaches zero, with $n, m, \varepsilon, \delta$ fixed. Prop. E.5 shows the margin exponent is tight up to logarithmic factors; what remains is whether the linear dependence on q is necessary.

Any positive answer in the standard setting must avoid instances in which the agents rank the goods alike, since on those envy-cycle elimination already returns an EFX allocation (Barman & Krishnamurthy, 2020, §3) and EFX is computable in polynomial time (Plaut & Roughgarden, 2020b, §6); the two agents in $\mathcal{I}_m(\theta)$ rank the goods differently, which is not incidental.

The unit-demand half is not vacuous either, and for a reason that makes it the extreme case of the mechanism in §3.2: there the slack disappears entirely.

So for unit-demand valuations the two universal notions coincide once the envied bundle has two items, and a barrier for one is a barrier for the other there. Any such barrier would be purely informational, and it could not come from the combinatorial source that drives the exponential bounds for submodular valuations. With exact values and two agents, EFX is computable in polynomial time for every class up to and including gross substitutes, which contains unit-demand, and becomes PLS-complete only at submodular (Goldberg et al., 2025); and m queries reconstruct an additive valuation outright, so no query lower bound beyond m is possible there in the noiseless model (Plaut & Roughgarden, 2020b, Fn. 10); the same holds for unit-demand, whose values are determined by the m singletons. Under noise, reconstruction is never exact, but ε-accuracy on the nq singletons transfers to every bundle, which is why Prop. E.4 applies. Note that the question is specific to EEAG. In the standard setting, serial dictatorship gives an EFX allocation for unit-demand valuations using about $\log m$ bits per agent (Feige, 2025), and it does so because every bundle it creates before the last is a singleton, on which Prop. F.1 has no force. A fixed initial allocation removes that escape: once an agent's bundle starts with at least two items, certifying ε-EFX against it is the same as certifying ε-EF.

Why the quantifier decides. The results separate two questions that the computational view of fair division treats together. Under oracle access, EF1 is always attainable in the standard setting and computable in polynomial time for two agents, while deciding exact envy-freeness already encodes PARTITION for two agents with identical additive valuations; the gap between the two notions is a familiar one. Under noisy observations, Thm. 3.1 places ε-EF and ε-EFX at $\Omega(m^2)$ observations for every fixed $\varepsilon \in (0, 1/128]$ on instances where ε-EF1 costs nothing, and Thm. 4.2 bounds ε-EF1 by $\widetilde{O}(nm/\varepsilon^2)$ in the standard setting. Sample complexity sees two classes. The distinction is not the $1/\varepsilon^2$ rate, which every estimation problem pays; it is whether the notion requires the allocation to track the valuations at the scale of individual items. EF1 does not, because it may discount the largest item in the envied bundle. EFX and EF do, and pay for it in m.

It is worth saying plainly why none of this contradicts the impossibility of optimization from samples. Balkanski et al. (2022) show that coverage functions, which are maximizable under a cardinality constraint within $1-1/e$ given a value oracle and are $(1-\varepsilon)$-PMAC-learnable over every distribution, nonetheless cannot be maximized under that constraint from polynomially many samples, drawn from any distribution, to within a factor $2^{-\Omega(\sqrt{\log n})}$; this holds even when the covered sets are known and only the weights are hidden. What separates that result from ours is not the access model. Their construction partitions the ground set into good, bad, and masking elements so that every set is either small enough that good and bad contribute identically, or large enough that the masking elements swamp the difference; since the answers carry no information about which part is good, adaptively chosen queries inherit that independence, and choosing where to look does not obviously help. What separates the two is the objective. Optimization from samples asks for a set whose value is within a constant factor of the best, so an algorithm must identify which elements are good. EF1 asks for no such thing: it is a feasibility problem with a one-item slack, and the slack is what makes the identity of individual goods irrelevant. EFX removes that slack, and by Thm. 3.1 the fine structure returns.

Beyond goods and beyond envy. The framework carries to chores, where Lem. B.2 is unchanged because it depends only on the additive structure of estimation error, and to PROP1 and PROPX, whose quantifiers mirror those of EF1 and EFX. In each case existence has to be settled first, and we have not carried either extension out.

H Limitations

The upper bound that is near-linear in m holds only in the standard setting. In EEAG with a fixed nonempty initial allocation our general bound is exponential in the budget, and polynomial only for constant budgets or for singleton-determined valuation classes. All of these procedures use exhaustive search: polynomial-time for fixed k, superpolynomial once k grows. For singleton-determined classes the observation count is polynomial because bundle values are reconstructed from $O(nq)$ singleton estimates, but the search itself may still be exponential-time. The lower bound of Thm. 3.1 is proved for two agents within coverage valuations and is tight on that family; whether harder coverage instances raise the exponent is open; for unit-demand valuations the bound is near-linear in $q = |P| + m$ once an s-EFX extension exists for a fixed $s < \varepsilon$ (Prop. E.4). For ε-EF1 the available lower bound is the $\Omega(n \log m)$ inherited from the noiseless query model, which needs $\varepsilon < 1/2$ and $m \geq n^{\alpha}$ for a constant $\alpha > 1$, and binds only deterministic algorithms; closing the gap to Thm. 4.2 is open. The observation model assumes bounded, unbiased, independent noise on each request. Boundedness enters the proofs only through Hoeffding's inequality, and §6 finds the algorithms still successful on 97% or more of instances under unbounded Gaussian and heavy-tailed noise, while a biased bounded model breaks them; a formal treatment of the unbounded case would replace Hoeffding by a concentration bound for the tail at hand, which for t_3 means working from a bounded variance rather than from a sub-Gaussian or sub-exponential hypothesis, and we have not carried it out; the measurements are in Appx. J. Our real-preference experiments use two public datasets with 9 and 240 items respectively, and convert rankings to cardinal valuations by a fixed rule; preferences elicited as bundle values directly, rather than induced from rankings, would be a more faithful test.

I Extended Related Work

This appendix records the context compressed out of §1.

Restoring exact fairness by augmentation. Adding goods is one of several ways to repair an unfair allocation. Envy can also be removed by *subsidy*, paying agents to accept their bundles (Halpern & Shah, 2019), where one dollar per agent suffices for additive valuations with values in $[0, 1]$ (Brustle et al., 2020); by *donation*, setting aside a charity pile (Caragiannis et al., 2019a; Chaudhury et al., 2021); or by *deletion* of items (Dorn et al., 2021). Boehmer et al. (2024) give the parameterized complexity of the dual framework in which envy is repaired by removing resources rather than adding them.

Other relaxations and valuation classes. Beside EF1 and EFX, the maximin share guarantee (Kurokawa et al., 2018; Barman & Krishnamurthy, 2020) is the other standard relaxation; surveys

covering all three are Amanatidis et al. (2023) and Amanatidis et al. (2022). Submodular valuations capture diminishing marginal returns in combinatorial auctions (Lehmann et al., 2006; Fujishige, 1991) and are studied extensively in fair division (Ghodsi et al., 2022; 2018; Montanari et al., 2025). Under ordinal access, $O(\log m)$ comparison queries suffice for EF1 with a constant number of agents and identical additive valuations (Bu et al., 2024).

Learning and mechanism design with samples. Two literatures are adjacent but answer a different question. In sample-based mechanism design the samples are draws of the whole valuation profile, and the issue is whether empirical performance predicts performance on the underlying distribution (Balcan et al., 2016; Hartline, 2013). The same contrast appears in envy-free classification, where a classifier envy-free on a sample of individuals must remain so on the population, and bounded Natarajan dimension is what keeps the sample small (Balcan et al., 2019). Our samples are repeated observations of a single fixed profile, and the question is certification rather than generalization.

Robustness to misreported values. Peters et al. (2022) study rent division when reported valuations may be wrong, seeking divisions that remain envy-free within as large a radius of the reports as possible, their *lexislack* solution, and show an almost optimal division can be identified from polynomially many samples of the reports. Dang et al. (2026) perturb each reported valuation once, by stochastic or adversarial additive noise, and bound the envy that is then unavoidable; round-robin's worst-case envy under bounded noise is one of their bounds. Both fix the perturbation and ask what fairness survives it. We instead let an algorithm buy accuracy, and ask how much it must buy.

Foundations. The problem of dividing a resource fairly dates to Steinhaus (1948) and Dubins & Spanier (1961); book-length treatments are Brams & Taylor (1996) and Moulin (2003). For the place of envy-freeness in distributive justice see Arnsperger (1994). That envy-free allocations of indivisible goods may fail to exist is standard (Bouveret & Lang, 2008).

J Further Experiments

This appendix reports the experiments mentioned at the end of §6: the notions compared across agent counts, what the epoch rule buys on the instances of Fig. 1, Alg. 1 run outside its hypotheses, and non-constructed preference and ranking data. The consultation counts behind Fig. 1, the robustness grid, and the hyperparameter ablation are tabulated in Appx. K.

J.1 More than two agents

Thm. 4.3 is for two agents, and Open Problem G.5 asks what happens beyond. Table 2 reports the standard setting with $n \in \{2, 3, 4\}$ agents, $m = 6$ goods, and additive valuations drawn from a Dirichlet prior, so that EFX is guaranteed to exist for $n = 3$ (Chaudhury et al., 2024). One generic budget-T algorithm estimates every bundle value, searches exhaustively for the allocation with the largest minimum estimated envy slack, and its single output is checked three ways.

The gap between the existential and the universal notion tends to widen with n on this grid. With no observations, an index-round-robin allocation is ε-EF1 on 86%, 76%, and 61% of instances for $n = 2, 3, 4$, but ε-EFX on 50%, 17%, and 3%; at $T = 300$ the pairs are $100/99$, $99/82$, and $97/68$. ε-EF1 reaches every instance by $T = 10^3$ at every n, while ε-EFX needs $T = 3 \times 10^3$ at $n = 4$ and ε-EF never exceeds 91% there, because with four agents and six goods an envy-free allocation often does not exist. Nothing here is a proof, but the direction is the one the quantifier predicts: each additional agent adds ordered pairs on which the universal condition must hold for every item, and the one-item slack absorbs more of them than the data does.

J.2 What the epoch rule buys

Thm. 4.3 rests on a single discipline, that an item found not to violate is not consulted again within an epoch, and Prop. D.1 says a search without it can be forced to spend quadratically many consultations. Fig. 1 measures both halves.

On the adversarial instance the separation is exactly as predicted. The re-examination variant consults $\Theta(m^2)$ bundles, with fitted log-log slope 1.93 and bootstrap interval $[1.91, 1.96]$, while Alg. 1

Table 2: Standard setting, additive Dirichlet valuations, $m = 6$, $\varepsilon = 0.1$; fraction of instances on which one generic algorithm's output satisfies each notion, 150 instances per cell (100 for $n = 4$).

	T	0	300	10^3	3×10^3	10^4	3×10^4
$n = 2$	ε-EF1	.860	1.00	1.00	1.00	1.00	1.00
	ε-EFX	.500	.987	1.00	1.00	1.00	1.00
	ε-EF	.420	.967	.993	1.00	.993	.993
$n = 3$	ε-EF1	.760	.993	1.00	1.00	1.00	1.00
	ε-EFX	.173	.820	.960	.993	1.00	1.00
	ε-EF	.080	.620	.927	.987	.973	.980
$n = 4$	ε-EF1	.610	.970	1.00	1.00	1.00	1.00
	ε-EFX	.030	.680	.830	.990	.970	.980
	ε-EF	.010	.370	.630	.850	.810	.910

consults $1.01m$ bundles at $m = 256$, slope 0.95 with interval $[0.92, 0.97]$. The two intervals are disjoint and each excludes the other's exponent, so the gap is not an artifact of the fitting range.

On random instances we find no clear separation between the two variants, both linear, with slopes 0.89 and 0.86 and overlapping intervals. We report this because it is the honest shape of the result: Prop. D.1 is a worst-case statement, and the estimate pattern it needs, in which every band item but one reads just below the threshold on every pass, is not what independent noise produces. The epoch rule is insurance rather than an average-case improvement, which is the right reading of a guarantee that must hold against adversarial estimation error.

J.3 Which assumption is actually load-bearing

Def. 2 asks for observations that are bounded in $[0, 1]$ and unbiased for the bundle's value, and Thm. 4.3 uses boundedness through Hoeffding's inequality. Fig. 4 runs Alg. 1 on six valuation classes under five noise models, two of which satisfy the definition and three of which do not, with the same per-bundle sample budget everywhere.

The algorithm remains highly successful under both violations. Unbounded Gaussian noise gives success rates between 0.99 and 1.00, and t_3 noise, whose tails are heavy enough that no sub-Gaussian bound applies, gives between 0.97 and 1.00. What breaks the guarantee is a noise model that satisfies boundedness: truncating Gaussian noise to $[0, 1]$ drops the success rate as low as 0.80. The reason is visible in panel (b). Truncation is the only one of the five that is biased, and at a bundle worth 0 its bias is 0.0997, which is ε itself at the accuracy demanded here, so the estimator is wrong by the whole tolerance before any sampling error is considered.

In these experiments the only substantial degradation occurs under the biased model, consistent with bias rather than unbounded support or tail weight being the relevant failure mode here; unbiasedness alone, however, does not give a distribution-free finite-sample guarantee. Boundedness enters the proof only to supply a concentration inequality, and another concentration hypothesis would serve equally well; repetition alone does not remove a fixed bias from the sample mean. We state Def. 2 with bounded observations because it is the simplest hypothesis under which the constants are explicit, not because the result is fragile outside it.

J.4 Non-constructed preference and ranking data

The instances above are synthetic, and the hard family in particular is built to make the separation as large as possible. To see what the same quantities look like outside constructed instances, we use two datasets from PrefLib (Mattei & Walsh, 2013). The first is the AGH course-selection survey, in which 146 students each gave a complete strict ranking of 9 courses (Faliszewski, 2013); course allocation is the setting for which EF1 was originally proposed (Budish, 2010). The second is the cleanweb collection, in which four web search engines each rank the same 240 items, which we use as four rankers of a large item set. Rankings are converted to normalized Borda weights and then to additive, budget-additive, and unit-demand valuations. Fig. 5 reports both.

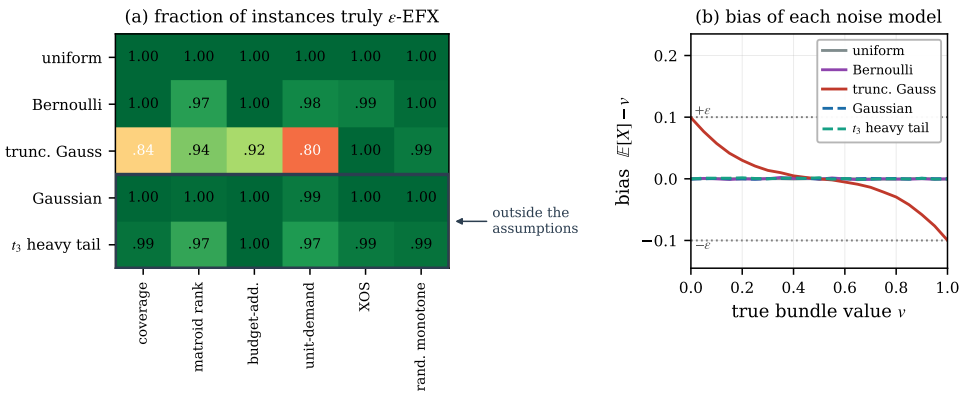

Figure 4: Robustness of Thm. 4.3 outside its hypotheses. $m = 24$, $\varepsilon = 0.1$, 300 instances per cell, identical sample budget per bundle. **(a)** Fraction of instances whose output is ε-EFX under the true valuations; the boxed rows violate the boundedness of Def. 2 and remain highly successful in these experiments, while truncated Gaussian noise, which respects boundedness, is the one model that degrades. **(b)** The explanation: truncation is the only biased model, and its bias reaches ε at the ends of the value range.

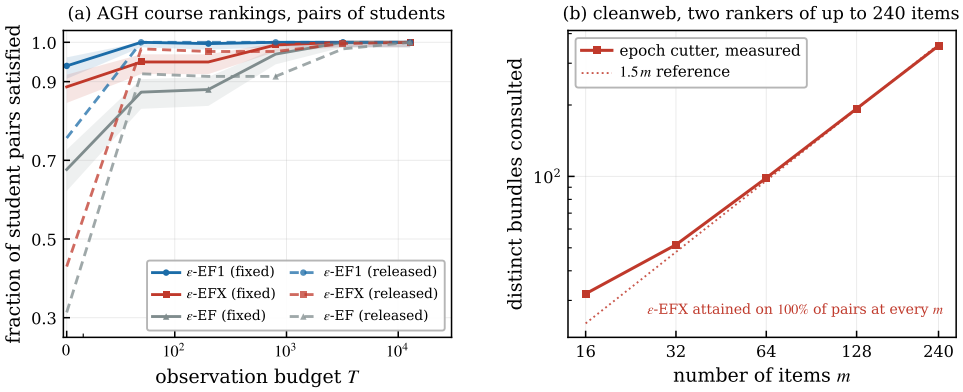

Figure 5: Non-constructed preference and ranking data, $\varepsilon = 0.1$. **(a)** AGH course rankings (Faliszewski, 2013), 300 random student pairs per budget, Borda-additive valuations, Wilson bands on the fixed curves. The order of the three notions is the one the theory predicts, and the gap is widest at $T = 0$. **(b)** cleanweb rankings, six ranker pairs at each of five item counts: the two-agent cutter attains ε-EFX everywhere, consulting a linear number of bundles.

On pairs of students, the ordering ε-EF1 $\geq$ ε-EFX $\geq$ ε-EF holds at every budget in both the fixed and the released setting, and every notion is attained on essentially all pairs by $T \approx 3 \times 10^3$: with nine courses and Borda weights there is not much to learn, and all three notions are cheap in absolute terms. Two things nonetheless carry over from the theory. The gap between the notions is largest when information is scarcest: with no observations at all and the pool split in half, ε-EF1 holds on 76% of released pairs, ε-EFX on 43%, and ε-EF on 31%, so on real course preferences one item of slack is worth 33 points of success before any data is seen. And freezing each student's top course, which is the natural way a partial allocation arises here, moves every curve up rather than down, because a fixed top course is a large bundle only in the sense of Prop. F.1 when it has company, and here it does not. The instances that make the barrier bite are not what course preferences look like, and we say so.

On the algorithmic side the theory transfers exactly. Margin envy-cycle elimination reaches ε-EF1 on 100% of sampled groups of $n \in \{2, \ldots, 6\}$ students under all three valuation inductions, consulting $9n$ distinct bundles, inside the $n(m + 1)$ of Thm. 4.2. On cleanweb the epoch cutter reaches ε-EFX on 100% of ranker pairs at every m from 16 to 240, consulting $2.0m$ bundles at $m = 16$ and $1.5m$ by $m = 240$, inside the $O(m/\varepsilon)$ of Thm. 4.3 with a constant far below the one in the proof.

K Experimental Details

This appendix gives the numbers behind §6. Every quantity was produced by the scripts in the supplementary material, which regenerate every figure and Tables 4, 5 and 6 from scratch; Tables 2 and 7 are transcribed from the scripts' JSON output.

K.1 Finite-sample uncertainty for the phase diagram

The surface in Fig. 3 is a visualization of the (m, ε) plane rather than a confidence bound on a population success threshold: each bisection evaluation draws 60 instances, and at that size an empirical rate of $57/60$ has a Wilson 95% interval of $[0.86, 0.98]$. To quantify threshold uncertainty on the slice the main text uses for scaling, we reran near-threshold budgets at $\varepsilon = 0.002$ for $m \in \{8, 12, 16, 24, 32, 48\}$ with 200 instances per tested budget, drawing θ coordinatewise uniform, sampling a fresh uniform $m/2$-subset for each observation, recovering $\hat{\theta}$ by the A_i/B_i empirical-mean rule of Prop. 3.4, and scoring success as ε-EFX under the true valuations. The 200 instances are shared across the budgets tested at a given m.

m	T/m^2	T	success	Wilson 95%
8	45	2,880	0.930	[0.886, 0.958]
8	50	3,200	0.960	[0.923, 0.980]
8	80	5,120	0.995	[0.972, 0.999]
12	55	7,920	0.940	[0.898, 0.965]
12	60	8,640	0.960	[0.923, 0.980]
12	85	12,240	0.985	[0.957, 0.995]
16	70	17,920	0.940	[0.898, 0.965]
16	80	20,480	0.970	[0.936, 0.986]
16	90	23,040	0.990	[0.964, 0.997]
24	70	40,320	0.945	[0.904, 0.969]
24	80	46,080	0.955	[0.917, 0.976]
24	100	57,600	0.990	[0.964, 0.997]
32	70	71,680	0.915	[0.868, 0.946]
32	80	81,920	0.955	[0.917, 0.976]
32	110	112,640	0.995	[0.972, 0.999]
48	90	207,360	0.945	[0.904, 0.969]
48	100	230,400	0.960	[0.923, 0.980]
48	120	276,480	0.990	[0.964, 0.997]

Table 3: Near-threshold replication at $\varepsilon = 0.002$: 200 instances per row, success scored as ε-EFX under the true valuations.

Two threshold definitions give two finite-grid exponents. Taking the first tested budget whose empirical rate reaches 95% gives thresholds T/m^2 of $50, 60, 80, 80, 80, 100$ and a log-log slope of 2.35; resampling the shared instances and reselecting the threshold within each replicate gives a 95% interval of $[2.10, 2.51]$, from 19,994 of 20,000 replicates in which a threshold falls inside the tested range. Requiring instead that the Wilson 95% lower bound itself exceed 0.95 gives T/m^2 of $80, 85, 90, 100, 110, 120$ and a slope of 2.24. Both are finite-grid quantities, sensitive to the grid and to the threshold definition; the asymptotic exponent is supplied by Thm. 3.1 and Prop. 3.4, not by these fits. Success on shared instances is not monotone in T: a larger budget can flip a coordinate estimate, and 11 of the 2,400 adjacent budget pairs reverse.

K.2 Success rates and paired tests for Fig. 2

Table 4 lists the measured success rates at five budgets in each setting, together with the discordant counts that enter the paired test. Here b counts instances on which ε-EF1 holds and ε-EFX does not, and c counts the reverse; the exact McNemar test is the two-sided binomial test of b against $b + c$ at rate $\frac{1}{2}$, and Holm's correction is applied across all ten comparisons. That $c = 0$ at every budget is itself informative: we never observed an instance on which the universal notion held and the existential one failed, which is what monotonicity predicts. The observations here are Bernoulli, as in Thm. 3.1.

Table 4: Per-budget success rates and paired tests for Fig. 2. $m = 32$, $\varepsilon = 1/128$, 300 instances per cell; budgets are the measured grid points nearest the nominal values. Here b counts instances on which ε-EF1 holds and ε-EFX does not, and c the reverse. That $c = 0$ at every budget matches monotonicity, and the Holm-adjusted p-values reject equality at all five.

setting	T	ε-EF1	ε-EFX	ε-EF	b	c	raw p	Holm p
fixed	0	1.000	0.000	0.000	300	0	9.8×10^{-91}	9.8×10^{-90}
fixed	769	1.000	0.000	0.000	300	0	9.8×10^{-91}	9.8×10^{-90}
fixed	8577	1.000	0.027	0.027	292	0	2.5×10^{-88}	2.0×10^{-87}
fixed	36467	1.000	0.513	0.513	146	0	2.2×10^{-44}	1.6×10^{-43}
fixed	95709	1.000	0.957	0.957	13	0	2.4×10^{-4}	0.00146
released	0	1.000	1.000	1.000	0	0	1	1
released	769	1.000	1.000	1.000	0	0	1	1
released	8577	1.000	1.000	1.000	0	0	1	1
released	36467	1.000	1.000	1.000	0	0	1	1
released	95709	1.000	1.000	1.000	0	0	1	1

K.3 Consultation counts for Fig. 1

Table 5 gives the counts underlying both panels. The adversarial instance is deterministic once the estimate pattern is fixed, so no interval is reported for it; the random columns give medians over 40 seeds, with a bias-corrected bootstrap interval from 2000 resamples for the epoch rule.

Table 5: Distinct bundles consulted by the cutter's local search, $\varepsilon = 0.1$. Left: the instance of Prop. D.1, which is deterministic once the estimates are fixed, so no interval is reported. Right: random coverage and monotone instances, 40 seeds per point, reported as medians, with a bias-corrected bootstrap 95% interval for the epoch rule. Re-examination is quadratic in m on the left and shows no clear separation from the epoch rule on the right.

	adversarial instance		random instances (40 seeds)		
m	epoch	re-exam.	epoch median	re-exam. median	epoch 95% CI
8	10	42	15	14	[14, 16]
16	18	150	28	25	[26, 30]
24	26	322	40	34	[35, 44]
32	34	558	46	44	[46, 48]
48	50	1222	69	62	[64, 73]
64	66	2142	90	90	[77, 91]
96	98	4750	134	132	[111, 136]
128	130	8382	168	150	[144, 180]
192	194	18718	217	207	[205, 269]
256	258	33150	304	304	[279, 357]

K.4 Robustness grid for Fig. 4

Table 6 repeats the grid of Fig. 4 with Wilson 95% intervals attached to every cell. Rows marked † fall outside Def. 2, through bias or unbounded support.

Table 6: Fraction of instances whose output is ε-EFX under the true valuations, with Wilson 95% intervals. $m = 24$, $\varepsilon = 0.1$, 300 instances per cell. † marks noise models outside Def. 2. Truncated Gaussian is bounded but biased; Gaussian and t_3 are unbiased but unbounded.

noise model	coverage	matroid rank	budget-add.	unit-demand	XOS	rand. mon.
uniform	1.000 [0.99, 1.00]	1.000 [0.99, 1.00]	1.000 [0.99, 1.00]	1.000 [0.99, 1.00]	1.000 [0.99, 1.00]	1.000 [0.99, 1.00]
Bernoulli	1.000 [0.99, 1.00]	0.970 [0.94, 0.98]	1.000 [0.99, 1.00]	0.983 [0.96, 0.99]	0.987 [0.97, 0.99]	0.997 [0.98, 1.00]
trunc. Gauss†	0.843 [0.80, 0.88]	0.940 [0.91, 0.96]	0.923 [0.89, 0.95]	0.800 [0.75, 0.84]	1.000 [0.99, 1.00]	0.990 [0.97, 1.00]
Gaussian†	1.000 [0.99, 1.00]	0.997 [0.98, 1.00]	1.000 [0.99, 1.00]	0.990 [0.97, 1.00]	1.000 [0.99, 1.00]	1.000 [0.99, 1.00]
t_3 heavy tail†	0.993 [0.98, 1.00]	0.967 [0.94, 0.98]	1.000 [0.99, 1.00]	0.973 [0.95, 0.99]	0.990 [0.97, 1.00]	0.993 [0.98, 1.00]

K.5 Hyperparameters of Alg. 1, and δ

The proof of Thm. 4.3 fixes the accuracy at $\varepsilon_0 = \varepsilon/8$ and the margin at $\tau = 6\varepsilon_0$; the exit condition needs $\tau + 2\varepsilon_0 \le \varepsilon$ and the strict-move condition needs $\tau > 4\varepsilon_0$. The left block of Table 7 varies τ at fixed $\varepsilon_0 = \varepsilon/8$; the right block follows the coupled path $\tau = 6\varepsilon_0$, simultaneously coarsening the estimates and widening the margin. Both use random coverage and monotone instances with $m = 20$, $\varepsilon = 0.1$, 150 per row and bounded uniform noise; all tested instances succeed at the proof setting; along the tested margin values and along the coupled path, success degrades once the settings are made sufficiently coarse.

Table 7: Ablation of the margin τ (with $\varepsilon_0 = \varepsilon/8$) and of the accuracy ε_0 (with $\tau = 6\varepsilon_0$). The proof's values are in bold. All tested instances succeed at the proof setting; success degrades once the settings are made sufficiently coarse.

τ/ε	ε-EFX rate	consulted	$\varepsilon_0/\varepsilon$	τ/ε	ε-EFX rate	samples per bundle
0.25	1.000	39	1/32	0.19	1.000	2000
0.50	1.000	34	1/16	0.38	1.000	2000
0.75	**1.000**	**33.5**	**1/8**	**0.75**	**1.000**	**576**
1.00	0.987	31	1/4	1.50	0.633	144
1.25	0.700	30	1/2	3.00	0.280	36

The constants in the proof are the largest the argument allows, not the largest that work. A margin of $\tau = 0.75\varepsilon$ is the largest that satisfies $\tau + 2\varepsilon_0 \le \varepsilon$, and the success rate begins to fall at $\tau = \varepsilon$ and collapses at 1.25ε; every tested instance also succeeded at the smaller margins we evaluated, including settings outside the sufficient range the proof uses, at the cost of a few more consultations. Likewise $\varepsilon_0 = \varepsilon/8$ is the coarsest accuracy for which $\tau = 6\varepsilon_0$ still satisfies $\tau + 2\varepsilon_0 \le \varepsilon$, and along that path, moving from $\varepsilon_0 = \varepsilon/8$ to $\varepsilon/4$ reduces the observed success rate from 1.00 to 0.63. The dependence on δ enters only through $\ln(2N/\delta)$ in the per-bundle sample count: at $N = 100$ and $\varepsilon = 0.1$, lowering δ from 10^{-1} to 10^{-4} raises the count from 24,323 to 46,428, a factor 1.9 for three orders of magnitude in δ.

K.6 Computational verification

The epoch cutter of Alg. 1 was checked against an adversary that branches, at every consultation, on the estimate taking either endpoint of its accuracy interval, so every endpoint-choice sequence along every execution path is explored. Over 770,576 complete adversarial paths on random coverage, random monotone, and anchor instances with $m \le 6$, every leaf was a two-sided ε-EFX partition under the true valuation, within the consultation and epoch bounds of Thm. 4.3. A checker that recomputes the scan order after each move, rather than iterating over the order fixed at the start of the pass, reports failures; this is why Thm. 4.3 states the snapshot requirement explicitly.

The construction of Thm. 3.1 was verified separately in exact rational arithmetic: for every $m \in \{2, 4, 6, 8\}$ and every θ, we confirmed the value range $[0, \frac{3}{4}]$, monotonicity, and submodularity of v_1^θ on all subsets, the coverage representation of Lem. 3.2(i), the equality $v_1^\theta(S_\theta) = v_1^\theta(\{g^*, g^\circ\}) = \frac{3}{8}$ of part (ii), and the zero-observation EF1 extension of part (iii). At the theorem-boundary value $\varepsilon = 1/128$, the recovery inequality of Lem. 3.3 was checked by enumerating every extension within budget for every θ: across 10, 132, 1,864 and 27,152 pairs (θ, ρ) at $m = 2, 4, 6, 8$, every ε-EFX extension had Hamming distance zero to S_θ, inside βm; since a smaller ε only strengthens the EFX condition, the same holds for every $\varepsilon \le 1/128$ at these sizes. The mean-shift bound of Lem. L.2 was checked over every θ, every coordinate and every subset of $P \cup R$, for 32, 512, 6,144 and 65,536 exact comparisons at $m = 2, 4, 6, 8$; the maximum shift equals $3/(8m)$ at every size, so the bound is attained. The constants of the Assouad step were checked symbolically in exact rational arithmetic: $T \le m^2/32$ gives total variation at most $1/8$ and an Assouad lower bound of $7m/32$ on expected Hamming error, while the success decomposition with $\beta \le 3/64$ allows at most $41m/192$, and $7/32 - 41/192 = 1/192 > 0$; equivalently, avoiding the contradiction would need $\eta + \varepsilon \ge 5/512$, whereas the theorem's parameters give $\eta + \varepsilon \le 9/1024$. The greedy counterexample of Appx. M.1 was checked for $m = 1, \ldots, 8$: the initial raw envy gaps in the unscaled representation are $(1.2, 5, 0)$, so the rule selects a_2 first, and an exhaustive search over continuations confirms that no EF1 extension remains reachable, although one exists before that step.

K.7 Compute

All experiments run on a single CPU core. The figure suite takes about 40 minutes to regenerate, dominated by the bisection sweep of Fig. 3, which evaluates roughly 1.4×10^5 instances. No GPU is used and no experiment requires more than 1 GB of memory.

L Omitted Proofs from §3

This appendix completes the proof of Thm. 3.1. Throughout, $\mathcal{I}_m(\theta)$, S_θ, η, and β are as in §3. For $\theta \in \{0,1\}^{m/2}$ and $i \in [m/2]$ we write $\theta^{\oplus i}$ for θ with coordinate i flipped, and d_H for Hamming distance on $\{0,1\}^{m/2}$ as well as on subsets of R.

L.1 Proofs of Lemmas 3.2 and 3.3

Proof of Lem. 3.2. (i) Since $|S \cap S_\theta| \le |S_\theta| = m/2$ for every S, the minimum in (3) equals $2|S \cap S_\theta|/m$, so $v_1^\theta(S) = \frac{3}{16}\mathbf{1}[S \cap R \neq \emptyset] + \frac{3}{8m}|S \cap S_\theta| + (\frac{3}{8} - \eta)\mathbf{1}[g^* \in S] + \eta\,\mathbf{1}[g^\circ \in S]$. This is the coverage function of the universe consisting of one element u_0 of weight $\frac{3}{16}$ covered by every $r \in R$, one element u_r of weight $\frac{3}{8m}$ covered only by r for each $r \in S_\theta$, one element u_* of weight $\frac{3}{8} - \eta$ covered only by g^*, and one element $u_\circ$ of weight η covered only by g°. Its maximum is $\frac{3}{16} + \frac{3}{16} + \frac{3}{8} = \frac{3}{4}$. The valuation v_2 is additive, hence coverage, with maximum 1.

(ii) $|S_\theta| = m/2 = k$. Agent a_1 has $v_1^\theta(S_\theta) = \frac{3}{16} + \frac{3}{8m} \cdot \frac{m}{2} = \frac{3}{8} = v_1^\theta(\{g^*, g^\circ\})$ and does not envy. Agent a_2 has $v_2(\{g^*, g^\circ\}) = 1 > 0 = v_2(S_\theta)$ and does not envy.

(iii) Under $\rho(a_1) = \{r_1\}$, dropping g^* from a_2's bundle leaves $\{g^\circ\}$, worth η to a_1, while $v_1^\theta(\{r_1\}) \ge \frac{3}{16} > \eta$; so a_1 is EF1. Agent a_2 values pool goods at 0 and never envies. Under the empty extension, $v_1^\theta(\emptyset) = 0 \ge \eta - \eta$, so it is η-EF1. □

Proof of Lem. 3.3. The ε-EFX condition for a_1 against a_2, with g° dropped, reads $v_1^\theta(S) \ge v_1^\theta(\{g^*\}) - \varepsilon = \frac{3}{8} - \eta - \varepsilon$. The hypothesis $\beta \le \frac{3}{64}$ forces $\eta + \varepsilon \le \frac{9}{1024}$, so the right side is positive; if $S = \emptyset$ the left side is 0, a contradiction. Hence $S \neq \emptyset$ and $v_1^\theta(S) = \frac{3}{16} + \frac{3}{8m}|S \cap S_\theta|$. Rearranging, $|S \cap S_\theta| \ge \frac{m}{2} - \frac{8m}{3}(\eta + \varepsilon)$, that is, $|S_\theta \setminus S| \le \frac{8m}{3}(\eta + \varepsilon)$. Since $|S| \le k = |S_\theta|$, we have $|S \setminus S_\theta| \le |S_\theta \setminus S|$, and therefore $d_H(S, S_\theta) = |S_\theta \setminus S| + |S \setminus S_\theta| \le \frac{16m}{3}(\eta + \varepsilon) = \beta m$. For ε-EF the condition is $v_1^\theta(S) \ge \frac{3}{8} - \varepsilon$, and the same steps give $d_H(S, S_\theta) \le \frac{16m}{3}\varepsilon$. □

L.2 From Set Recovery to Coordinate Recovery

Lemma L.1 (Coordinate Recovery). *For $S \subseteq R$ define $\hat\theta(S) \in \{0,1\}^{m/2}$ by $\hat\theta_i = 1$ if $r_{2i} \in S$ and $r_{2i-1} \notin S$, and $\hat\theta_i = 0$ otherwise. Then $d_H(\hat\theta(S), \theta) \le d_H(S, S_\theta)$ for every θ.*

Proof. Fix a pair i and let r^{in} be the member of $\{r_{2i-1}, r_{2i}\}$ that lies in S_θ and r^{out} the other. If S contains r^{in} and not r^{out}, then $\hat\theta_i = \theta_i$ and the pair contributes 0 to both distances. If S contains r^{out} and not r^{in}, then $\hat\theta_i \neq \theta_i$, and the pair contributes 1 to $d_H(\hat\theta, \theta)$ and 2 to $d_H(S, S_\theta)$. If S contains both or neither, then $\hat\theta_i = 0$, which is wrong only when $\theta_i = 1$, contributing at most 1 to $d_H(\hat\theta, \theta)$; and the pair contributes exactly 1 to $d_H(S, S_\theta)$. Summing over pairs gives the claim. □

L.3 Per-Observation Divergence

Lemma L.2 (Mean Shift). *For every θ, every $i \in [m/2]$, and every $S \subseteq P \cup R$,*

$$\left| v_1^\theta(S) - v_1^{\theta^{\oplus i}}(S) \right| \le \tfrac{3}{8m},$$

and whenever the two values differ, both lie in $[\frac{3}{16}, \frac{3}{4}]$. The valuation v_2 does not depend on θ.

Proof. Flipping θ_i exchanges which of r_{2i-1}, r_{2i} belongs to S_θ, so $|S \cap S_\theta|$ changes by at most 1, and the second summand of (3) changes by at most $\frac{3}{16} \cdot \frac{2}{m}$. No other summand depends on θ. If the

two values differ, then $S \cap R \neq \emptyset$, so the first summand contributes $\frac{3}{16}$ and the value is at least $\frac{3}{16}$; it is at most $\frac{3}{4}$ by Lem. 3.2(i). □

Lemma L.3 (Kullback–Leibler Bound). *For $\mu, \mu' \in [\frac{3}{16}, \frac{3}{4}]$ with $|\mu - \mu'| \leq \frac{3}{8m}$,*

$$\mathrm{KL}\big(\mathrm{Bernoulli}(\mu) \,\big\|\, \mathrm{Bernoulli}(\mu')\big) \;\leq\; \frac{(\mu-\mu')^2}{\mu'(1-\mu')} \;\leq\; \frac{9/(64m^2)}{39/256} \;=\; \frac{12}{13\,m^2} \;<\; \frac{1}{m^2}.$$

Proof. The first inequality is the standard bound $\mathrm{KL}(p\|q) \leq \chi^2(p\|q)$. On the interval $[\frac{3}{16}, \frac{3}{4}]$ the product $\mu'(1-\mu')$ is minimized at the endpoint $\mu' = \frac{3}{16}$, where it equals $\frac{3}{16} \cdot \frac{13}{16} = \frac{39}{256}$. □

L.4 Transcript Divergence under Adaptivity

Let $\mathcal{A}$ be a randomized algorithm that draws T value observations, choosing the t-th pair (a_t, S_t) as a function of its internal randomness and the history $H_t = \big((a_1, S_1, X_1), \ldots, (a_{t-1}, S_{t-1}, X_{t-1})\big)$, and outputs an extension ρ. Let $\mathbb{P}_\theta$ denote the joint law of the internal randomness, the full transcript H_{T+1}, and the output, when the instance is $\mathcal{I}_m(\theta)$.

The next lemma is the standard divergence decomposition for adaptive sampling, in the form familiar from the bandit literature (Lattimore & Szepesvári, 2020, Lem. 15.1); we give the proof since the objects here are (agent, bundle) pairs rather than arms.

Lemma L.4 (Transcript Divergence). *For every θ and every $i \in [m/2]$, $\mathrm{KL}(\mathbb{P}_\theta \,\|\, \mathbb{P}_{\theta^{\oplus i}}) \leq T/m^2$.*

Proof. The internal randomness and the choice of (a_t, S_t) given H_t have the same conditional law under both instances, and the output is a function of the transcript and the randomness. By the chain rule for Kullback–Leibler divergence,

$$\mathrm{KL}(\mathbb{P}_\theta \| \mathbb{P}_{\theta^{\oplus i}}) \;=\; \sum_{t=1}^{T} \mathbb{E}_{\mathbb{P}_\theta}\Big[\, \mathrm{KL}\big(\mathbb{P}_\theta(X_t \mid H_t) \,\big\|\, \mathbb{P}_{\theta^{\oplus i}}(X_t \mid H_t)\big)\Big].$$

Given H_t, the observation X_t is $\mathrm{Bernoulli}(v_{a_t}(S_t))$ under the respective instance. If $a_t = a_2$, or if $S_t \cap R = \emptyset$, the two means coincide and the term vanishes. Otherwise Lem. L.2 places both means in $[\frac{3}{16}, \frac{3}{4}]$ at distance at most $\frac{3}{8m}$, and Lem. L.3 bounds the term by $1/m^2$. □

L.5 Assouad's Lemma and the Conclusion

We use the total-variation form of Assouad's lemma (Tsybakov, 2009, Thm. 2.12(ii)): for any family $\{\mathbb{P}_\theta\}_{\theta \in \{0,1\}^d}$ and any estimator $\hat{\theta}$ measurable with respect to the data,

$$\max_{\theta \in \{0,1\}^d} \mathbb{E}_\theta\big[d_H(\hat{\theta}, \theta)\big] \;\geq\; \frac{d}{2}\Big(1 - \max_{\theta,\, i} \mathrm{TV}\big(\mathbb{P}_\theta, \mathbb{P}_{\theta^{\oplus i}}\big)\Big). \tag{5}$$

Proof of Thm. 3.1. Let $\varepsilon \leq 2^{-7}$ and suppose, for contradiction, that $\mathcal{A}$ draws $T \leq m^2/32$ observations and, for every θ, outputs with probability at least $2/3$ an extension ρ within budget that is ε-EFX under v^θ. Define the estimator $\hat{\theta} = \hat{\theta}(\rho(a_1))$ of Lem. L.1; it is determined by $\mathcal{A}$'s output and hence by the data.

Upper bound on the error. On the success event, Lemmas 3.3 and L.1 give $d_H(\hat{\theta}, \theta) \leq d_H(\rho(a_1), S_\theta) \leq \beta m$. On the failure event, $d_H(\hat{\theta}, \theta) \leq m/2$ trivially. Since the failure probability is at most $1/3$, for every θ,

$$\mathbb{E}_\theta\big[d_H(\hat{\theta}, \theta)\big] \;\leq\; \beta m + \tfrac{1}{3} \cdot \tfrac{m}{2} \;=\; \Big(\beta + \tfrac{1}{6}\Big)m.$$

Lower bound from Assouad. By Lem. L.4 and Pinsker's inequality, $\mathrm{TV}(\mathbb{P}_\theta, \mathbb{P}_{\theta^{\oplus i}}) \leq \sqrt{\mathrm{KL}/2} \leq \sqrt{T/(2m^2)} \leq \sqrt{1/64} = \frac{1}{8}$ for every θ and i. Applying (5) with $d = m/2$,

$$\max_{\theta} \mathbb{E}_\theta\big[d_H(\hat{\theta}, \theta)\big] \;\geq\; \frac{m}{4} \cdot \frac{7}{8} \;=\; \frac{7m}{32}.$$

Contradiction. Combining the two bounds, $\beta + \frac{1}{6} \geq \frac{7}{32}$, that is, $\beta \geq \frac{5}{96}$, that is, $\eta + \varepsilon \geq \frac{5}{96} \cdot \frac{3}{16} = \frac{5}{512}$. But $\eta + \varepsilon \leq 2^{-10} + 2^{-7} = \frac{9}{1024} < \frac{10}{1024} = \frac{5}{512}$. Hence no such $\mathcal{A}$ exists, and every algorithm meeting the guarantee draws more than $m^2/32$ observations.

For ε-EF, Lem. 3.3 gives the smaller value $\beta = \frac{16}{3}\varepsilon$, and the same inequalities hold a fortiori. □

The threshold on ε deserves comment, because the barrier cannot hold for all ε and it is worth knowing where it stops.

Proposition L.5 (Where the barrier is ruled out)**.** *Fix $\varepsilon > \frac{3}{32} - \eta$. On every instance $\mathcal{I}_m(\theta)$, an extension giving a_1 a uniformly random set of $m/2$ pool items is ε-EFX with probability $1 - e^{-\Omega(m)}$, and consults no observations.*

Proof. Write $\kappa = |S \cap S_\theta|$ for the random set S. Then κ is hypergeometric with mean $m/4$, so $\kappa \geq m/4 - m^{2/3}$ with probability $1 - e^{-\Omega(m^{1/3})}$, and in fact $\kappa \geq m/4 - \gamma m$ with probability $1 - e^{-\Omega(m)}$ for any fixed $\gamma > 0$. As in the proof of Lem. 3.3, the binding requirement is $v_1^\theta(S) \geq \frac{3}{8} - \eta - \varepsilon$, and $v_1^\theta(S) = \frac{3}{16} + \frac{3}{16} \cdot \frac{2\kappa}{m} \to \frac{9}{32}$. Since $\frac{9}{32} \geq \frac{3}{8} - \eta - \varepsilon$ exactly when $\varepsilon \geq \frac{3}{32} - \eta$, a fixed surplus above that threshold absorbs the fluctuation. □

So the barrier lives in a window. Thm. 3.1 establishes it for $\varepsilon \leq 1/128$ and Prop. L.5 in Appx. L rules it out for $\varepsilon > 3/32 - \eta \approx 0.0928$. The proof of Thm. 3.1 is not tied to the stated constants: for any $\lambda \in (0, 1)$ and any required success probability $1 - \delta$, the same argument yields $T > 2\lambda^2 m^2$ provided

$$\eta + \varepsilon \;<\; \tfrac{3}{16}\Big(\tfrac{1-\lambda}{4} - \tfrac{\delta}{2}\Big), \tag{6}$$

the stated theorem being $\delta = 1/3$, $\lambda = 1/8$. Demanding success probability $9/10$ rather than $2/3$ raises the admissible ε to below 0.030 at the same $T > m^2/32$, and $19/20$ raises it to below 0.035. What remains is roughly a factor of three, and it traces to a single step: Lem. L.1 charges a pair on which the extension errs the same amount whether the extension took the wrong item or took both, whereas the set distance in Lem. 3.3 charges those cases 2 and 1. A version of the argument that keeps the two accounts separate would close most of the window.

Remark L.6 (On the Constants). The threshold $\varepsilon \leq 1/128$ is what the coefficients $\frac{3}{16}$ and $\frac{3}{8}$ in (3) permit under a $2/3$ success requirement; (6) gives the general form, and rescaling the valuations moves the threshold and the constant $1/32$ together while leaving the exponent of m unchanged. The bound holds for the worst case over θ, and also on average over uniformly random θ, since Assouad's lemma lower-bounds the Bayes risk under the uniform prior. That it applies to randomized algorithms is a consequence of the same fact rather than extra work: randomization cannot help against a fixed prior. This is worth noting because the query lower bounds in this literature are usually adversary arguments, which bind only deterministic algorithms (Oh et al., 2021, Thm. 5.5), (Akrami & Rathi, 2025, Thm. 4.7), and extending them to randomized algorithms takes a separate argument that may cost a square root (Plaut & Roughgarden, 2020b, Thm. 3.4). The natural strategy of estimating each singleton separately costs $\Theta(m^3)$; the shared-observation decoder of Prop. 3.4 brings this down to $\Theta(m^2)$.

L.6 Proof of Prop. 3.4

Proof. Write $d = m/2$, $\Delta = \frac{3}{8m}$, and $T = \lceil 700\, m^2 \rceil$. The algorithm draws T independent uniformly random d-subsets $S_1, \ldots, S_T$ of R, observes $X_t \sim \mathrm{Bernoulli}(v_1^\theta(S_t))$ for each, and for every pair $i \in [d]$ forms the index sets $A_i = \{t : r_{2i} \in S_t,\ r_{2i-1} \notin S_t\}$ and $B_i = \{t : r_{2i-1} \in S_t,\ r_{2i} \notin S_t\}$. It sets $\hat{\theta}_i = 1$ if both A_i and B_i are nonempty and $\bar{X}_{A_i} > \bar{X}_{B_i}$, where $\bar{X}_A$ denotes the mean of X_t over $t \in A$, and $\hat{\theta}_i = 0$ otherwise; it outputs $\rho(a_1) = \hat{S} := \{r_{2i-1} : \hat{\theta}_i = 0\} \cup \{r_{2i} : \hat{\theta}_i = 1\}$, which has size $d = k$.

Conditional mean gap. Since $|S \cap S_\theta| \leq d$ for every S, the minimum in (3) never binds, and for nonempty $S \subseteq R$, $v_1^\theta(S) = \frac{3}{16} + \Delta\,|S \cap S_\theta|$. For $m = 2$ there is no pair $j \neq i$ and the claim below is immediate; assume $m \geq 4$. Conditional on $t \in A_i$, the remaining $d - 1$ elements of S_t form a uniform $(d-1)$-subset of the other $m - 2$ goods, so each pair $j \neq i$ places its S_θ-member in S_t

with probability $\frac{d-1}{m-2} = \frac{1}{2}$; the same holds conditional on $t \in B_i$. Pair i contributes θ_i under A_i and $1-\theta_i$ under B_i. Hence $\mathbb{E}[X_t \mid t \in A_i] - \mathbb{E}[X_t \mid t \in B_i] = \Delta\,(2\theta_i - 1)$.

Per-coordinate error. Fix i and let G_i be the event $\{|A_i| \geq T/5,\ |B_i| \geq T/5\}$. Each of $|A_i|, |B_i|$ is $\mathrm{Bin}(T,p)$ with $p = \binom{m-2}{d-1}/\binom{m}{d} = \frac{m}{4(m-1)} \geq \frac{1}{4}$, so a Chernoff bound gives $\Pr[G_i^{\mathrm{c}}] \leq 2e^{-T/200}$; when either group is empty, G_i^{c} holds and $\hat{\theta}_i$ is set to 0. Conditional on the index sets A_i and B_i, the observations X_t for $t \in A_i$ are independent, $[0,1]$-valued, each with mean $\mathbb{E}[X_t \mid t \in A_i]$, and likewise for B_i. If both empirical means lie strictly within $\Delta/2$ of their expectations the sign of $\bar{X}_{A_i} - \bar{X}_{B_i}$ is correct, so on G_i Hoeffding's inequality gives $\Pr[\hat{\theta}_i \neq \theta_i \mid A_i, B_i] \leq 4\exp(-T\Delta^2/10)$. Unconditionally, $\Pr[\hat{\theta}_i \neq \theta_i] \leq 2e^{-T/200} + 4e^{-T\Delta^2/10} = 2e^{-3.5m^2} + 4e^{-630/64} < 2.14 \times 10^{-4}$ for every even $m \geq 2$.

From coordinate errors to EFX. If h coordinates are wrong then $|\hat{S} \cap S_\theta| = d - h$ and $v_1^\theta(\hat{S}) = \frac{3}{8} - \Delta h$. Dropping g° from a_2's bundle leaves $\{g^*\}$, worth $\frac{3}{8} - \eta$ to a_1, so $\hat{S}$ is ε-EFX for a_1 whenever $\Delta h \leq \eta + \varepsilon$, that is, $h \leq \frac{8m}{3}(\eta + \varepsilon)$; dropping g^* instead leaves $\{g^\circ\}$, worth $\eta < \frac{3}{16} \leq v_1^\theta(\hat{S})$. Agent a_2 values only g^* and g°, both of which it keeps, and never envies.

Success probability. By linearity of expectation, $\mathbb{E}[h] \leq \frac{m}{2} \cdot 2.14 \times 10^{-4}$. The worst case is $\varepsilon \to 0$, which tolerates $h \leq \frac{8m}{3}\eta = m/384$; by Markov's inequality $\Pr[h > m/384] \leq 192 \cdot 2.14 \times 10^{-4} < 0.042$. The algorithm therefore outputs an ε-EFX extension with probability at least $0.958 > 3/4$. □

M Analysis of the Learning Algorithms

M.1 Why a Gap-Driven Greedy Rule Fails

A natural alternative to Alg. 3 is a greedy rule that repeatedly gives one pool item to the agent with the largest raw envy gap $\max_{a'}[v_a(B_{a'}) - v_a(B_a)]$ and stops when that agent's marginal gains vanish. The rule is incorrect even with exact valuations, because the agent with the largest gap need not be in any EF1-violating pair, and feeding it exhausts the pool. The following instance is EF1-feasible and the rule fails on it.

Three agents; $m \geq 1$ pool goods; no binding budget ($k = m$). $\sigma(a_1) = \emptyset$, $\sigma(a_2) = \{g_1, g_2\}$, $\sigma(a_3) = \{g_3\}$. Valuations:

$$v_1(S) = \min(1, |S \cap R|) + 0.6\,|S \cap \{g_1, g_2\}|, \qquad v_2(S) = 0.1\,|S \cap R| + 5 \cdot \mathbf{1}[g_3 \in S],$$
$$v_3(S) = \mathbf{1}[g_3 \in S].$$

All three are monotone submodular; dividing each by $C = 5 + 0.1m$ brings the values into $[0,1]$, and every inequality below is homogeneous, so we keep the unscaled form for readability. The extension $\rho(a_1) = \{r_1\}$ is EF1: $v_1(\{r_1\}) = 1 \geq 0.6 = v_1(\{g_1, g_2\} \setminus \{g_1\})$; a_2 envies both a_1 and a_3, but each envied bundle is a single item, so deleting it removes the envy; a_3 envies no one. Under the gap rule, however, a_2 has envy gap 5 toward a_3 (though the pair (a_2, a_3) satisfies EF1 with witness g_3), while a_1's gap toward a_2 is 1.2. So a_2 is selected first and receives a pool item, and from that point the pair (a_1, a_2) is beyond repair. Whatever the rule does next, a_1's value for its own bundle is at most $\min(1, |B_{a_1} \cap R|) \leq 1$, since g_1 and g_2 are fixed with a_2. Meanwhile $B_{a_2} \supseteq \{g_1, g_2, r\}$, and deleting any one item leaves a_1 at least 1.2: deleting r leaves $\{g_1, g_2\}$, worth 1.2, and deleting g_1 or g_2 leaves the other fixed item with a pool item, worth 1.6. Hence $v_1(B_{a_1}) \leq 1 < 1.2 \leq v_1(B_{a_2} \setminus \{g\})$ for every $g \in B_{a_2}$, and no continuation reaches EF1.

Restricting the rule to agents in violating pairs does not repair it: an agent whose own violation requires many items (because its marginal gains are small) can accumulate a bundle that another, already saturated, agent envies beyond repair. This is consistent with, though not implied by, the NP-hardness of deciding EF1-completability (HV et al., 2025), which concerns the variant in which every remaining good must be allocated. Whether EEAG with a fixed nonempty σ admits an oracle algorithm with polynomial consultation number for general monotone submodular valuations is open; §4.4 settles the singleton-determined classes.

M.2 Proof of Thm. 4.2

Throughout, $\varepsilon' = \varepsilon/8$ and $\tau = 3\varepsilon'$. We fix an arbitrary ε'-accurate answer table and argue deterministically, showing that margin envy-cycle elimination is $(\varepsilon', 8\varepsilon')$-robust for EF1 with consultation number at most $n(m+1)$; Thm. 4.1 then supplies the sampling, the failure probability and the observation count.

Consultation number. That envy-cycle elimination can be run with $O(nm)$ value queries for any number of agents with arbitrary monotone valuations is due to Oh et al. (2021, Thm. 5.1); we record the argument here only to confirm that the margin variant does not increase the count. A bundle, viewed as a set, is created only in line 7 of Alg. 3, once per item, so at most $m+1$ distinct sets (including $\emptyset$) ever arise. Rotations reassign existing sets and create none. Building G requires $\hat{v}_a(S)$ for the n current sets S and all n agents a, and every such pair has the form (agent, existing set). Hence $N_{\mathcal{B}_{\mathrm{ECE}}} \le n(m+1)$, and the observation count in Thm. 4.1 is $O\big(\frac{nm}{\varepsilon^2}\log\frac{nm}{\delta}\big)$.

Termination. Fix any answer table $\hat{v}$. During one execution of the inner loop the current bundles, viewed as sets, are fixed, and rotations only permute them among the agents, so there are at most $n!$ allocation states. Define $\widehat{\Phi} = \sum_a \hat{v}_a(B_a)$. Every cycle rotation gives each participant a bundle it values more than τ above its current one, and so strictly increases $\widehat{\Phi}$. No allocation state can therefore repeat, and every inner loop terminates and exits with an acyclic graph, on every answer table. This is the unconditional termination Thm. 4.1 assumes. Under an ε'-accurate answer table the true welfare rises too: define $\Phi = \sum_a v_a(B_a)$ under the true valuations; $0 \le \Phi \le n$. Line 7 never decreases Φ, by monotonicity. A rotation along a cycle gives each participant $a^{(j)}$ the set $B_{a^{(j+1)}}$ with $\hat{v}_{a^{(j)}}(B_{a^{(j+1)}}) > \hat{v}_{a^{(j)}}(B_{a^{(j)}}) + 3\varepsilon'$; when the estimates are ε'-accurate, the true values satisfy $v_{a^{(j)}}(B_{a^{(j+1)}}) > v_{a^{(j)}}(B_{a^{(j)}}) + \varepsilon'$. Non-participants are unaffected. So each rotation increases Φ by more than ε', and there are fewer than n/ε' rotations in the whole run. Every inner iteration rotates or exits, so the algorithm terminates in finite time.

A source exists. When the inner loop exits, G is acyclic, so some agent a has in-degree zero: for every a', $\hat{\gamma}(a', a) \le \tau$, i.e. $\hat{v}_{a'}(B_{a'}) \ge \hat{v}_{a'}(B_a) - \tau$.

Value monotonicity for each agent. From any time onward, an agent's true value for its own bundle never decreases: additions increase it by monotonicity of v, and rotations increase it by more than ε' under the fixed ε'-accurate table, as shown above. (Note that this uses true monotonicity; the estimates $\hat{v}$ need not be monotone.) In the noiseless setting this property is due to Lipton et al. (2004); see Amanatidis et al. (2020, Thm. 2.7(b)) for the statement we mirror. What the margin buys is that it survives when the rotation rule is driven by estimates rather than true values.

Correctness. Let ρ be the output and fix agents $a' \ne a$ with $B_a \ne \emptyset$. Let r be the item most recently added to the set B_a (as a set; it may have changed hands since), say at time t, when it was added to the set $C = B_a \setminus \{r\}$ then held by a source $\tilde{a}$. At that moment the source condition applied to a' (if $a' = \tilde{a}$ the inequality below is trivial) gives

$$\hat{v}_{a'}(B_{a'}^{(t)}) \;\ge\; \hat{v}_{a'}(C) - \tau,$$

where both values were consulted in building G at time t. Accuracy transfers this to $v_{a'}(B_{a'}^{(t)}) \ge v_{a'}(C) - \tau - 2\varepsilon' = v_{a'}(C) - 5\varepsilon'$. By value monotonicity from time t to the end, $v_{a'}(B_{a'}) \ge v_{a'}(B_{a'}^{(t)})$. Hence

$$v_{a'}(B_{a'}) \;\ge\; v_{a'}(B_a \setminus \{r\}) - 5\varepsilon' \;\ge\; v_{a'}(B_a \setminus \{r\}) - \varepsilon,$$

so r witnesses ε-EF1 for the pair (a', a). Since C was consulted at time t and never changes as a set, the witness value is one of the at most $n(m+1)$ stored estimates.

This shows $\mathcal{B}_{\mathrm{ECE}}$ with margin $3\varepsilon'$ is $(\varepsilon', 8\varepsilon')$-robust for EF1 in the sense of Def. C.2, with consultation number at most $n(m+1)$. Thm. 4.1 with $\varepsilon' = \varepsilon/8$ then gives the stated observation count. □