

Black-Box Lifting and Robustness Theorems for Multi-Agent Contracts*

Paul Dütting[†] Tomer Ezra[‡] Michal Feldman[§] Thomas Kesselheim[¶]

Abstract

Multi-agent contract design has largely evaluated contracts through the lens of pure Nash equilibria (PNE). This focus, however, is *not* without loss: In general, the principal can strictly gain by recommending a complex, possibly correlated, distribution over actions, while preserving incentive compatibility. In this work, we extend the analysis of multi-agent contracts beyond pure Nash equilibria to encompass more general equilibrium notions, including mixed Nash equilibria as well as (coarse-)correlated equilibria (CCE). The latter, in particular, captures the limiting outcome of agents engaged in learning dynamics.

Our main result shows that for submodular and, more generally, XOS rewards, such complex recommendations yield at most a constant-factor gain: there exists a contract and a PNE whose utility is within a constant factor of the best CCE achievable by any contract. This provides a black-box *lifting*: results established against the best PNE automatically apply with respect to the best CCE, with only a constant factor loss. For submodular rewards, we further show how to transform a contract and a PNE of that contract into a new contract such that any of its CCEs gives a constant approximation to the PNE. This yields black-box *robustness*: up to constant factors, guarantees established for a specific contract and PNE automatically extend to the modified contract and any of its CCEs. We thus expand prior guarantees for multi-agent contracts and lower the barrier to new ones. As an important corollary, we obtain poly-time algorithms for submodular rewards that achieve constant approximations in any CCE, against the best CCE under the best contract. Such worst-case guarantees are provably unattainable for XOS rewards. Finally, we bound the gap between different equilibrium notions for subadditive, supermodular, and general rewards.

1 Introduction

A classic contract setting features a *principal* who must incentivize a self-interested *agent* to take costly actions that generate value for the principal. Examples include a firm designing incentive schemes for employees or a platform rewarding contributors. The central difficulty is that actions

*This project has been partially funded by the European Research Council (ERC) under the European Union’s Horizon Europe Program (FACT, grant agreement No. 101170373), by an Amazon Research Award, by the NSF-BSF (grant number 2020788), by the Israel Science Foundation Breakthrough Program (grant No. 2600/24), and by a grant from TAU Center for AI and Data Science (TAD). T. Ezra is supported by the Harvard University Center of Mathematical Sciences and Applications.

[†]Google Research, Zürich, Switzerland. Email: duetting@google.com

[‡]Harvard University, Cambridge, USA. Email: tomer@cmsa.fas.harvard.edu

[§]Tel Aviv University, Tel Aviv, Israel. Email: mfeldman@tauex.tau.ac.il

[¶]University of Bonn, Bonn, Germany. Email: thomas.kesselheim@uni-bonn.de

are typically *hidden*: the principal cannot directly monitor actions and must instead align incentives through payments that depend only on observable outcomes.

In recent years, many such relationships have migrated to computational platforms. These environments are large-scale and complex, calling for an algorithmic treatment. *Algorithmic contract design* has therefore emerged at the intersection of economics and computation, developing models and schemes for designing incentives in these settings (we refer to Dütting, Feldman, and Talgam-Cohen, 2024 for a comprehensive survey).

A natural focal point in this literature is *combinatorial contracts*: the principal may interact with teams of agents, a single agent may choose among combinations of actions, and the overall reward depends on the joint selection through a combinatorial reward function that can exhibit both substitutability and complementarity (e.g., Babaioff, Feldman, and Nisan, 2006; Castiglioni, Marchesi, and Gatti, 2023; Deo-Campo Vuong, Dughmi, Patel, and Prasad, 2024; Dütting, Ezra, Feldman, and Kesselheim, 2021, 2023, 2025; Dütting, Feldman, and Gal-Tzur, 2024; Dütting, Feldman, Gal-Tzur, and Rubinstein, 2026; Ezra, Feldman, and Schlesinger, 2024; Feldman, 2025; Hann-Caruthers and Goel, 2024). The principal’s objective is to select a contract that maximizes her expected utility given the agents’ equilibrium behavior.

A key limitation of much of this literature is its (near) exclusive focus on *pure Nash equilibria* (PNE). In multi-agent settings this restriction is *not* without loss of generality. As observed by Babaioff, Feldman, and Nisan (2010), the gap can arise already for submodular rewards (see Example A.1). In particular, a principal can sometimes do strictly better by inducing agents to play a *distribution* over actions from which no agent wishes to deviate. This raises a natural question: *how much additional utility can the principal obtain by moving beyond PNE to richer equilibrium concepts that allow randomization or correlation?* This question is especially salient because natural learning dynamics typically converge to these broader, and often correlated, equilibrium notions.

1.1 Model and Research Problem

We consider the multi-agent combinatorial contracts model of Dütting, Ezra, Feldman, and Kesselheim, 2025, which generalizes the models of Dütting, Ezra, Feldman, and Kesselheim, 2021 and Dütting, Ezra, Feldman, and Kesselheim, 2023. In the multi-agent combinatorial contracts model, a single principal interacts with n agents. Each agent $i \in [n]$ can take any subset of actions S_i from an available action set A_i . We let $A = \bigcup_i A_i$, and use $m = |A|$ to denote the total number of actions. We refer to the special case where $|A_i| = 1$ for all $i \in [n]$, i.e., each agent can either take action or not, as the binary-actions case. The agents’ choice of actions $S \subseteq A$, with $S_i = S \cap A_i$ the set of actions chosen by agent $i \in [n]$, determines a reward $f(S)$ through a reward function $f : 2^A \rightarrow \mathbb{R}_{\geq 0}$. We generally assume that the reward function is normalized (so that $f(\emptyset) = 0$) and monotone (so that $S \subseteq T$ implies $f(S) \leq f(T)$). Access to f may be given through a *value oracle* or a *demand oracle* (see Section 2). The agents each have a cost c_j for each action $j \in A$, and the cost for a set of actions $S_i \subseteq A_i$ is $c(S_i) = \sum_{j \in S_i} c_j$. To incentivize the agents, the principal designs a linear contract $\vec{\alpha} = (\alpha_1, \dots, \alpha_n) \in [0, 1]^n$. The interpretation is that the principal pays each agent i an α_i fraction of the reward $f(S)$ that results from the agents’ actions S . The principal’s utility is $(1 - \sum_i \alpha_i)f(S)$. The agents, in turn, have a utility of $\alpha_i f(S) - c(S_i)$. Since the agents’ utilities depend on each other, we are interested in equilibria among agents.

While prior work has focused on pure Nash equilibria (PNE), here we are interested in exploring the more general equilibrium concepts of mixed Nash equilibria (MNE), correlated equilibria (CE), and coarse-correlated equilibria (CCE). In these more general equilibrium notions the principal

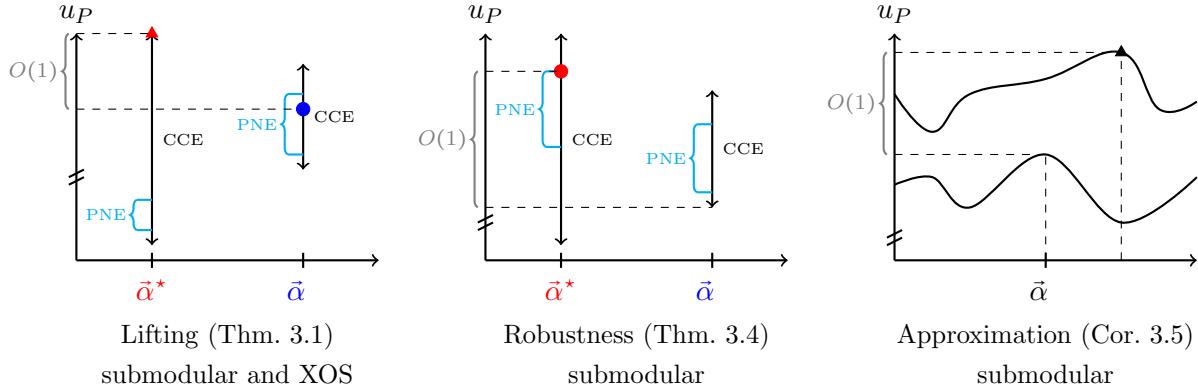


Figure 1: Visualization of our results for submodular and XOS rewards. Lifting (left panel): We are given as input a contract $\vec{\alpha}^*$ and a CCE (triangle) and we construct a contract $\vec{\alpha}$ and a PNE (circle). Robustness (middle panel): We are given a contract $\vec{\alpha}^*$ and a PNE (circle) and we construct a contract $\vec{\alpha}$ with a guarantee for any CCE under that contract. In both panels, the ranges of principal utilities under PNE and CCE are drawn as continuous intervals for illustration. Approximation (right panel): The lower curve shows the worst-performing CCE for each contract; the upper curve shows the corresponding best-performing CCE. We show that it is possible to compute a contract $\vec{\alpha}$ such that the worst CCE under that contract is within a constant-factor of the best CCE under any contract (black triangle).

induces the agents to play a distribution over actions, while ensuring that no agent has an incentive to deviate. The randomization can be either independent (as in MNE), or correlated (as in CE and CCE), where CE and CCE differ in how they capture the property that no agent wants to deviate (for formal definitions see Section 2). It is well known that these equilibrium concepts are successive generalizations, i.e.,

$$\text{PNE} \subseteq \text{MNE} \subseteq \text{CE} \subseteq \text{CCE}.$$

Our goal in this work is to bound the gap in the principal’s utility between different equilibria under different equilibrium notions, potentially achieved by different contracts, and provide algorithms for computing contracts with strong robustness guarantees. We consider both rewards from the hierarchy of complement-free set functions (submodular \subseteq XOS \subseteq subadditive) as well as supermodular and general monotone rewards.

1.2 Our Contribution

We discuss our main results, for submodular and XOS reward functions, in Section 1.2.1. In Section 1.2.2, we cover additional results that map the broader landscape of gaps between different equilibrium concepts under prominent classes of reward functions. We provide an illustration and an overview of our results in Figure 1 and Table 1.

1.2.1 Main Results: Submodular and XOS Rewards

For submodular rewards, Babaioff, Feldman, and Nisan, 2010 demonstrated that the principal’s utility under the best mixed Nash equilibrium can be strictly greater than under the best pure

Nash equilibrium. They conjectured that, for submodular rewards, this (multiplicative) gap is bounded by a constant. Despite substantial progress on combinatorial contract design since then, this conjecture has remained unresolved.

Our main result not only settles this conjecture but in fact establishes a much stronger result. We show that, in a model with arbitrary (non-binary) actions and for XOS rewards (a strict generalization of submodular rewards), for each contract $\vec{\alpha}^*$ and any coarse-correlated equilibrium (CCE) of that contract, there is a (typically different) contract $\vec{\alpha}$ and a pure Nash equilibrium (PNE) under $\vec{\alpha}$ that achieves a constant fraction of the principal's utility under the CCE. See Figure 1 (left) for an illustration. Thus, the conjectured constant gap is strengthened along three dimensions: extending from binary to arbitrary actions, from MNE to CCE, and from submodular to XOS rewards.

Black-Box Lifting Theorem (Theorem 3.1). For the multi-agent combinatorial contracts model with submodular or XOS rewards, given any contract $\vec{\alpha}^*$ and any CCE of $\vec{\alpha}^*$, there exists a contract $\vec{\alpha}$ and a PNE S of $\vec{\alpha}$ such that the principal's utility under S is a constant fraction of the principal's utility under the CCE of $\vec{\alpha}^*$. Moreover, given $\vec{\alpha}^*$ and the CCE, one can find $\vec{\alpha}$ and S in polynomial time (in n, m and the support size of the CCE) with value and demand oracle access to f .

This result provides a black-box lifting: any guarantee established with respect to the best-PNE benchmark automatically holds also against the stronger benchmark of the best CCE, with only a constant factor loss, thereby strengthening existing bounds for multi-agent contracts and simplifying the path to new ones. Indeed, to obtain guarantees against the best CCE, one no longer needs to grapple with the intricate notion of CCE (with its randomization and correlation), but can instead reason about the simpler and more transparent concept of pure Nash equilibrium. Readers familiar with the smoothness framework for bounding the price of anarchy (Roughgarden, 2015) may find this connection conceptually reminiscent; see Section 1.4 for further discussion.

As an immediate corollary, all approximation results established in Dütting, Ezra, Feldman, and Kesselheim, 2023, 2025, for instance, hold against the stronger benchmark of the best CCE under any contract. One such result is that, for submodular rewards, there is a poly-time algorithm that finds a contract under which any PNE attains a constant-factor approximation to the best PNE under any contract (Dütting, Ezra, Feldman, and Kesselheim, 2025). Our lifting theorem immediately extends this guarantee against the best CCE under any contract.

We further strengthen this result by establishing the following (black-box) robustness result for submodular rewards, which shows how to transform any contract and pure Nash equilibrium of that contract into a different contract that comes with a guarantee for any CCE under the modified contract. See Figure 1 (middle) for an illustration.

Black-Box Robustness Theorem (Theorem 3.4). For the multi-agent combinatorial contracts model with submodular rewards, given any contract $\vec{\alpha}^*$ and pure Nash equilibrium of $\vec{\alpha}^*$, there is an algorithm that runs in polynomial time (in n and m) using value queries to f , that computes a contract $\vec{\alpha}$ such that any CCE under $\vec{\alpha}$ achieves an $O(1)$ -approximation to the principal's utility under the PNE of $\vec{\alpha}^*$.

Combining this with the results in (Dütting, Ezra, Feldman, and Kesselheim, 2023, 2025) and the Black-Box Lifting Theorem, shows that for the multi-agent combinatorial contracts model with submodular rewards there is a poly-time algorithm (with value and demand oracles) that computes a contract such that any CCE under that contract provides a $O(1)$ -approximation to the best CCE under any contract (see Corollary 3.5). In the special case of binary actions, such a contract can

Reward function		Gap between ... and PNE		
		MNE	CE	CCE
Submodular/XOS	Binary/arbitrary actions	$\Theta(1)$ Lower: Ex. A.1, Upper: Thm. 3.1		
Subadditive	Binary/arbitrary actions	$\Theta(\text{poly}(n))$ Lower: Prop. 4.1, Upper: Prop. 4.5		
Supermodular	Binary actions	No gap Thm. 5.1		
	Arbitrary actions	No gap Thm. 5.2	Unbounded	Prop. 5.4
General	Binary/arbitrary actions	Unbounded Prop. 6.1		

Table 1: Gaps between MNE/CE/CCE and PNE, for different reward functions and binary actions vs. arbitrary actions. In merged cells, lower bounds are proved for simpler action and equilibrium notions and extend to more general ones; upper bounds are proved for more general cases and apply to the simpler ones. (E.g., for submodular/XOS rewards, the lower bound is proved for binary actions and MNE, while the upper bound is proved for arbitrary actions and CCE.)

be computed with value queries only. We provide an illustration of the guarantees achieved by the corresponding algorithms in Figure 1 (right).

Similar to the Black-Box Lifting Theorem, the Black-Box Robustness Theorem is conceptually related to the price of anarchy framework. It shows that, for submodular rewards, one only needs to show the existence of a good contract and pure Nash equilibrium, and can then rely on the extension result to transform the original contract into a contract under which all CCE (and hence learning outcomes) are near-optimal. Notably, for XOS rewards, such a worst-case approximation guarantee is unattainable, even for the weaker target of ensuring that any PNE under the given contract is within a constant factor of the best PNE under any contract (as already observed in Dütting, Ezra, Feldman, and Kesselheim, 2025).

1.2.2 Beyond XOS Rewards: Mapping the Landscape

We next examine the gap between different equilibrium concepts beyond XOS rewards. We start with subadditive rewards, then we move to supermodular rewards, and finally we consider general (monotone) rewards.

Subadditive Rewards. For subadditive rewards, we construct a carefully designed instance that exhibits a polynomial lower bound of order $\Omega(\sqrt{n})$ on the gap between the principal’s utility under the best mixed Nash equilibrium (MNE) and the best pure Nash equilibrium (PNE), where n denotes the number of agents. Remarkably, this gap already arises in the binary-actions case (so n is also the number of actions here). We also show that this gap is at most $O(n)$, even between

CCE and PNE and the case of general actions.¹

Proposition (Subadditive Rewards) (Proposition 4.1 and Proposition 4.5). For the multi-agent combinatorial contracts model with subadditive rewards, the gap between the principal's utility from the best MNE and the best PNE is $\Omega(\text{poly}(n))$, even in the binary-actions case. On the other hand, the gap between the principal's utility under the best CCE and the best PNE is at most $O(n)$, even for an arbitrary number of actions.

Supermodular Rewards. For supermodular rewards, we show that in the case of *binary actions*, there is no gap between the principal's utility from the best coarse-correlated equilibrium and the best pure Nash equilibrium. This strengthens a result of Babaioff, Feldman, and Nisan, 2010, who established this no-gap result only with respect to mixed NE.

Theorem (Supermodular Rewards, Binary Actions) (Theorem 5.1). For the multi-agent model with binary actions and supermodular rewards, there is no gap between the principal's utility from the best CCE and the best PNE.

In contrast, for arbitrary actions there is no gap only when comparing the best *correlated* equilibrium (CE) to the best PNE; for the more general notion of *coarse-correlated* equilibria (CCE), the gap to PNE may be unbounded.

Theorem (Supermodular Rewards, General Actions) (Theorem 5.2 and Proposition 5.4). For the multi-agent model with arbitrary actions and supermodular rewards, there is no gap between the principal's utility under the best CE and the best PNE; while the gap between CCE and PNE is unbounded.

General Rewards. Finally, we turn to general monotone rewards, in particular rewards that are neither subadditive nor supermodular. We show that in this regime, even the gap between the best MNE and the best PNE can be unbounded. Moreover, such unbounded gaps arise already with a constant number of agents and binary actions.

Proposition (General Rewards) (Proposition 6.1). In the multi-agent combinatorial contract model with general rewards (neither subadditive, nor supermodular), the gap between the principal's utility under the best MNE and the best PNE is unbounded, even with only four agents and binary actions.

1.3 Challenges and Techniques

Black-Box Lifting and the Scaling-for-Existence Lemma. Our key tool for establishing an upper bound on the gap between coarse-correlated equilibria and pure Nash equilibria, for submodular and XOS rewards, is a simple yet powerful *Scaling-for-Existence Lemma* (Lemma 3.3). This lemma starts from a contract $\bar{\alpha}^*$ and a distribution over action sets that satisfies a mild dropout-stability condition (Definition 2.6). Dropout-stability requires that each agent weakly prefers to follow the actions in the given distribution over unilaterally deviating to taking no action. This condition is naturally satisfied by all equilibrium concepts we study.

¹A weaker upper bound of m follows from a recent result by Dütting, Ezra, Feldman, and Kesselheim, 2025, showing that the gap between social welfare and the best PNE is at most m .

The Scaling-for-Existence Lemma shows that, after appropriately scaling the contract to $\tilde{\alpha}$, there exists a PNE that guarantees high reward relative to the original distribution. In particular, if there is a subset of agents that achieves high expected reward under the original distribution and whose total share $\sum_i \alpha_i^*$ is bounded away from 1, then the lemma can be used to show that the induced pure Nash equilibrium recovers a constant fraction of the principal’s utility under the original distribution. To obtain the constant-factor gap, we show that in any dropout-stable distribution, either (i) there is a “significant” agent and a good pure Nash equilibrium that incentivizes only this agent, or (ii) there exists a subset of agents to which we can apply the Scaling-for-Existence Lemma to obtain a good pure Nash equilibrium.

Our proof of the Scaling-for-Existence Lemma leverages that, for each contract, the induced game among the agents is a potential game (e.g., Deo-Campo Vuong, Dughmi, Patel, and Prasad, 2024; Dütting, Ezra, Feldman, and Kesselheim, 2025). We show that for the original contract $\tilde{\alpha}^*$ and any dropout-stable distribution under this contract, the XOS structure implies that the expected potential of any set S in the support of the dropout-stable distribution is non-negative. We then scale the contract to $\tilde{\alpha}$ and select S as a global maximizer of the potential function for $\tilde{\alpha}$. This ensures that $(S, \tilde{\alpha})$ forms a pure Nash equilibrium. Finally, the fact that $\tilde{\alpha}$ is obtained from $\tilde{\alpha}^*$ through scaling, together with the non-negativity of the expected potential under $\tilde{\alpha}^*$, implies that this pure Nash equilibrium achieves high reward relative to the original distribution.

Notably, for subadditive rewards, our proof for the Scaling-for-Existence Lemma breaks: it is no longer guaranteed that the expected potential of any set S in the support of a dropout stable distribution for contract $\tilde{\alpha}^*$ is non-negative. This failure not only undermines the proof technique, but points to a deeper structural issue. Namely, such a result is provably impossible for subadditive rewards: the gap between PNE and MNE (and thus PNE and CCE) is superconstant.

Black-Box Robustness and the Scaling-for-Robustness Lemma. The Scaling-for-Existence Lemma derives a good PNE from a dropout-stable distribution over actions (such as CE or CCE). Our Black-Box Robustness Theorem builds on a lemma, the *Scaling-for-Robustness Lemma* (Lemma 3.6), that achieves an orthogonal goal, namely showing that *every* CCE is good relative to a reference equilibrium (or rather relaxation thereof).

Our lemma generalizes the Doubling Lemma of Dütting, Ezra, Feldman, and Kesselheim, 2025. The Doubling Lemma showed that, for submodular rewards, for every contract $\tilde{\alpha}^*$ and PNE of $\tilde{\alpha}^*$, there exists a scaled contract $\tilde{\alpha}$ such that any PNE of $\tilde{\alpha}$ provides a constant-approximation to the original PNE. We present a new argument that achieves a parallel result for any CCE (rather than PNE) under the scaled contract, losing only another constant factor. This enables a comparison of any CCE under a given contract to the best PNE under any contract.

There are two main differences between the two scaling lemmas: First, they give different types of guarantees: the Scaling-for-Existence Lemma ensures that there *exists* a good equilibrium, whereas the Scaling-for-Robustness Lemma ensures that *every* equilibrium is good. Second, their domains differ: the existence lemma holds for XOS rewards, while the robustness lemma holds only for submodular rewards. A lemma unifying both results provably can’t hold. Indeed, for XOS rewards, there are instances for which the gap between the best and the worst equilibrium is large (Dütting, Ezra, Feldman, and Kesselheim, 2025).

Transitions for Supermodular Rewards. Our results for supermodular rewards exhibit two perhaps surprising transitions, namely (i) between binary and general actions and (ii) between

CE and CCE. Compared to the binary-actions case, arbitrary actions introduce two additional challenges. First, agents may take actions not present in the original distribution. We address this by showing that there exists a PNE whose action profile contains the union of the sets in the support of the correlated equilibrium, rather than matching it exactly, as in the binary case. Second, a distribution that is only a CCE (and not a CE) may place positive probability on actions that reduce an agent’s utility under the contract; such actions are not best responses and therefore cannot appear in any PNE of that contract. Consequently, our proof extends to correlated equilibria but not to coarse-correlated equilibria. Indeed, our lower-bound construction for CCE shows that the gap can be unbounded.

1.4 Further Related Work

Inefficiency of Equilibria and Price of Anarchy. Bounding the inefficiency of equilibria is one of the staples of algorithmic game theory e.g., Koutsoupias and Papadimitriou, 1999; Roughgarden and Tardos, 2002; Syrgkanis and Tardos, 2013, and among the main conceptual contributions of theoretical computer science to economics.

In this regard, our work bears some conceptual resemblance to the smoothness framework for bounding the price of anarchy (Roughgarden, 2015), in that to bound a ratio under a general equilibrium concept (CCE), it suffices to bound it under a simpler one (PNE), thereby simplifying the analysis. Similarly, as in the smoothness framework, it suffices to consider “simple” deviations: deviation to one’s strategy in an optimal profile in the smoothness framework, and deviation to taking no action in our case (as captured by the dropout-stability notion).

Two important differences to the Price of Anarchy literature are that the Price of Anarchy literature is typically interested in the social welfare (or social cost) achieved by equilibria, and focuses on worst-case equilibria for a fixed game. In contrast, here we are interested in the principal’s utility and we compare equilibria across different induced games.

Related Studies on Combinatorial Contracts. Our work builds on the growing body of research on combinatorial contracts (Feldman, 2025), and in particular contributes to the rapidly expanding literature on multi-agent contracts. The work that is most related to our work is Babaioff, Feldman, and Nisan, 2010, who—working in the combinatorial agency model of Babaioff, Feldman, and Nisan, 2006; Babaioff, Feldman, Nisan, and Winter, 2012—quantify the ratio between the principal’s utility under mixed and pure strategies, for reward functions exhibiting increasing and decreasing marginals (corresponding to supermodular and submodular functions, respectively). We significantly extend this work from binary actions to arbitrary actions, and from mixed Nash equilibria to more general equilibrium concepts such as correlated and coarse-correlated equilibria—and on the way we resolve the main open question from this prior work, concerning the gap between PNE and MNE for submodular rewards.

The very recent work of Dütting, Ezra, Feldman, and Kesselheim, 2025 is another important reference. Although their work focuses on pure Nash equilibria, just like most of the existing literature on multi-agent contracts, they do show a related result: namely, that for the multi-agent combinatorial actions model that we study in this work, and for submodular rewards, there exists a contract α such that the worst-equilibrium under that contract, yields a constant-factor approximation to the best equilibrium under the best contract. We strengthen this result in two ways: First, we show that in fact any CCE of the contract returned by the algorithm obtains a

constant fraction to the best PNE under any contract. Second, by our lifting result, we extend this guarantee to the best-CCE benchmark.

Alon, Castiglioni, Chen, Ezra, Li, and Talgam-Cohen, 2025 extend the multi-agent (binary action) contract design framework to settings with many projects, where the principal needs to partition the agents among the projects, and within each project, the principal incentivizes the agents through a contract.

Cacciamani, Bernasconi, Castiglioni, and Gatti, 2024 explore a related multi-agent and (non-combinatorial) multi-action model, and highlight the value of randomized contracts. In their model, the principal also uses a randomized contract, which is different from our approach which assumes that the principal posts a single deterministic contract. They then propose and study an equilibrium notion, which can be interpreted as a correlated equilibrium when both the principal and the agents randomize. They show that this type of randomized correlated equilibria are more powerful than pure Nash equilibria, demonstrating for instance that the principal’s utility can be unboundedly higher when the setting has a supermodular reward structure. An important difference between our work and their work is that we consider deterministic contracts.

Finally, slightly more removed from our question and work, Dasaratha, Golub, and Shah, 2025 study a multi-agent multi-action model in which the agents choose from a continuum of effort levels, and the agent’s choices determine both their costs and the principal’s reward, through cost and reward functions. The conceptual similarity lies in their treatment of “fractional” actions, akin to our model where an agent selects different action sets with varying probabilities. However, unlike our work, they do not consider mixed or correlated equilibria.

2 Preliminaries

Multi-Agent Combinatorial Contracts. Every agent $i \in [n]$ has a (finite) set of actions A_i . The set of actions A_i and $A_{i'}$ of any two agents $i \neq i'$ are disjoint. We denote the set of all actions by $A = \bigcup_{i \in [n]} A_i = [m]$. An important special case is when for all $i \in [n]$, it holds that $|A_i| = 1$. In this case, each agent can either take action or not. We refer to this as the binary-actions case.

Every agent i can take any subset of actions $S_i \subseteq A_i$. We use $S = \bigcup_{i \in [n]} S_i$ to denote the set of actions chosen by the agents, and for each agent $i \in [n]$ we let $S_{-i} = \bigcup_{i' \neq i} S_{i'}$. (Similarly, for any $S \subseteq A$ we let $S_i = S \cap A_i$.) There is a reward function $f : 2^A \rightarrow \mathbb{R}_{\geq 0}$. We generally assume that the reward function is monotone (non-decreasing) and normalized so that $f(\emptyset) = 0$. In addition, each agent has a cost c_j for each action $j \in A_i$. The cost of a set of actions $S_i \subseteq A_i$ is $c(S_i) = \sum_{j \in S_i} c_j$. For singletons $\{j\} \subseteq A$, we sometimes use the shorthands $f(j) = f(\{j\})$ and $c(j) = c(\{j\})$.

A (linear) contract $\vec{\alpha} = (\alpha_1, \dots, \alpha_n)$ defines a share $\alpha_i \in [0, 1]$ of the reward $f(S)$ that is to be paid to agent i . The utility of agent i under contract $\vec{\alpha}$ when the set of actions chosen by the agents is S is

$$u_i(S, \vec{\alpha}) = \alpha_i f(S) - c(S_i).$$

The utility of the principal in that case is

$$u_P(S, \vec{\alpha}) = \left(1 - \sum_{i \in [n]} \alpha_i\right) \cdot f(S).$$

Note that, for any set S , the utilities of the agents and the utility of the principal sum up to $f(S) - c(S)$. We refer to this quantity as the *welfare* of the set of actions S .

We assume that both the principal and the agents are expected utility maximizers; their expected utilities for a distribution \mathcal{D} over sets $S \subseteq A$ is $u_i(\mathcal{D}, \vec{\alpha}) = \mathbf{E}_{S \sim \mathcal{D}}[u_i(S, \vec{\alpha})]$ and $u_P(\mathcal{D}, \vec{\alpha}) = \mathbf{E}_{S \sim \mathcal{D}}[u_P(S, \vec{\alpha})]$. We adopt the perspective of the principal and seek a contract that maximizes the principal's expected utility.

Equilibrium Concepts. In multi-agent combinatorial contracts, the principal defines a contract $\vec{\alpha}$ and the agents choose their actions in reply to this contract. We are thus looking at a single-leader multiple-followers Stackelberg game. Specifically, for any fixed contract $\vec{\alpha}$ we study a simultaneous move game among the agents. We consider different equilibrium concepts.

A pure Nash equilibrium is a set of actions S such that no agent wants to deviate from their choice of action S_i to some other action T_i .

Definition 2.1 (Pure Nash Equilibrium). Given reward function f , a *pure Nash equilibrium* (PNE) of contract $\vec{\alpha}$ is a set of actions $S \subseteq A$ such that for every agent i and every $T_i \subseteq A_i$ it holds that

$$\alpha_i \cdot f(S_{-i} \cup S_i) - c(S_i) \geq \alpha_i \cdot f(S_{-i} \cup T_i) - c(T_i).$$

A mixed Nash equilibrium, in turn, is a product distribution $\mathcal{D} = \mathcal{D}_1 \times \dots \times \mathcal{D}_n$ over sets of actions where \mathcal{D}_i is a distribution over sets of action among A_i , where no agent i strictly wants to deviate from their distribution \mathcal{D}_i to a pure strategy $T_i \subseteq A_i$.

Definition 2.2 (Mixed Nash Equilibrium). Given reward function f , a *mixed Nash equilibrium* (MNE) of contract $\vec{\alpha}$ is a product distribution $\mathcal{D} = \prod_{i \in [n]} \mathcal{D}_i$ over sets $S \subseteq A$ such that for every agent i and every $T_i \subseteq A_i$ it holds that

$$\mathbf{E}_{S \sim \mathcal{D}}[\alpha_i \cdot f(S_{-i} \cup S_i) - c(S_i)] \geq \mathbf{E}_{S \sim \mathcal{D}}[\alpha_i \cdot f(S_{-i} \cup T_i) - c(T_i)].$$

Correlated equilibria are defined as a joint (possibly) correlated distribution over sets of actions. We can interpret S_i as the action recommended to agent i . Then the equilibrium requirement is that each agent i should prefer to follow their recommended action S_i , rather than switching to some other action $\pi_i(S_i)$ whenever they are told S_i .

Definition 2.3 (Correlated Equilibrium). Given reward function f , a *correlated equilibrium* (CE) of contract $\vec{\alpha}$ is a distribution \mathcal{D} over sets $S \subseteq A$ such that for every agent i and every mapping $\pi_i : 2^{A_i} \rightarrow 2^{A_i}$ it holds that

$$\mathbf{E}_{S \sim \mathcal{D}}[\alpha_i \cdot f(S_{-i} \cup S_i) - c(S_i)] \geq \mathbf{E}_{S \sim \mathcal{D}}[\alpha_i \cdot f(S_{-i} \cup \pi_i(S_i)) - c(\pi_i(S_i))].$$

The notion of a coarse-correlated equilibrium weakens the equilibrium requirement of a correlated equilibrium. Rather than being able to map each recommended action S_i to some other action $\pi_i(S_i)$, agent i only considers a single alternative action T_i .

Definition 2.4 (Coarse-Correlated Equilibrium). Given reward function f , a *coarse-correlated equilibrium* (CCE) of contract $\vec{\alpha}$ is a distribution \mathcal{D} over sets $S \subseteq A$ such that for every agent i and every $T_i \subseteq A_i$ it holds that

$$\mathbf{E}_{S \sim \mathcal{D}}[\alpha_i \cdot f(S_{-i} \cup S_i) - c(S_i)] \geq \mathbf{E}_{S \sim \mathcal{D}}[\alpha_i \cdot f(S_{-i} \cup T_i) - c(T_i)].$$

Note that the inequality in the definition of CCE is identical to that in the definition of MNE. The difference is that in an MNE, the distribution \mathcal{D} must be a product distribution.

It is not difficult to see that the equilibrium notions are successive relaxations, i.e.,

$$\text{PNE} \subseteq \text{MNE} \subseteq \text{CE} \subseteq \text{CCE}.$$

While in general games, a pure Nash equilibrium may not exist, games induced by multi-agent combinatorial action contracts admit a weighted potential function (Deo-Campo Vuong, Dughmi, Patel, and Prasad, 2024; Dütting, Ezra, Feldman, and Kesselheim, 2025). Recall that any (finite) game that admits such a potential function has the finite-improvement property, and hence at least one pure Nash equilibrium. Another useful property of such games is that any local maximum of the potential function corresponds to a pure Nash equilibrium of the underlying game. (For a detailed discussion of potential games see Monderer and Shapley, 1996.)

Recall that a weighted potential function is a function $\phi : 2^A \rightarrow \{-\infty\} \cup \mathbb{R}$ such that $u_i(S_i, S_{-i}, \vec{\alpha}) > u_i(S'_i, S_{-i}, \vec{\alpha})$ implies $\phi(S_i, S_{-i}) > \phi(S'_i, S_{-i})$. For any set S and contract $\vec{\alpha}$, let

$$\Phi(S, \vec{\alpha}) = f(S) - \sum_{i \in [n]} \frac{c(S_i)}{\alpha_i}, \quad (1)$$

where, if $\alpha_i = 0$ we define $c(S_i)/\alpha_i = \infty$ when $c(S_i) > 0$, and we let $c(S_i)/\alpha_i = 0$ when $c(S_i) = 0$.

Proposition 2.5 (Deo-Campo Vuong, Dughmi, Patel, and Prasad, 2024; Dütting, Ezra, Feldman, and Kesselheim, 2025). *For every contract $\vec{\alpha}$, the function $\Phi(\cdot, \vec{\alpha})$ is a weighted potential function in the game induced by the contract $\vec{\alpha}$.*

Dropout Stability. A crucial ingredient in our analysis is the following property, which captures stability with respect to a unilateral deviation to taking no action.

Definition 2.6 (Dropout Stability). Given reward function f , a distribution \mathcal{D} over sets of actions S is *dropout-stable* with respect to $\vec{\alpha}$ if for every agent i it holds that

$$\mathbf{E}_{S \sim \mathcal{D}}[\alpha_i \cdot f(S) - c(S_i)] \geq \mathbf{E}_{S \sim \mathcal{D}}[\alpha_i \cdot f(S_{-i})].$$

The following observation is immediate.

Observation 2.7. *Any coarse-correlated equilibrium of $\vec{\alpha}$ is also dropout-stable with respect to $\vec{\alpha}$.*

In what follows, it will be useful to have the notion of marginal contribution of an individual action or set of actions to a given set of actions S . For an action $j \in A$, we write $f(j | S) = f(S \cup j) - f(S)$ for the marginal contribution of action j to S . Similarly, for a set of actions $T \subseteq A$, we write $f(T | S) = f(S \cup T) - f(S)$ for the marginal contribution of the set of actions T to S .

Classes of Reward Functions. Our main interest will be in the following classes of (non-negative) reward functions $f : 2^A \rightarrow \mathbb{R}_{\geq 0}$. A set function is:

- *additive* if there exist values $f_1, \dots, f_m \in \mathbb{R}_{\geq 0}$ such that $f(S) = \sum_{j \in S} f_j$.
- *gross-substitutes* if for any two vectors $p \leq q \in \mathbb{R}_+^m$ and any $S \subseteq A$ such that $S \in \arg \max_{S' \subseteq A} (f(S') - \sum_{j \in S'} p_j)$ there is a $T \subseteq A$ with $\{j \in S \mid q_j \leq p_j\} \subseteq T$ such that $T \in \arg \max_{T' \subseteq A} (f(T') - \sum_{j \in T'} q_j)$.

- *submodular* if for every $S, S' \subseteq A$ with $S \subseteq S'$ and any $j \in A$ it holds that $f(j | S) \geq f(j | S')$.
- *XOS* if there exists a collection of additive functions $\{a_\ell : 2^A \rightarrow \mathbb{R}_{\geq 0}\}_{\ell=1,\dots,k}$ such that for each $S \subseteq A$ it holds that $f(S) = \max_{\ell=1,\dots,k} a_\ell(S)$.
- *subadditive* if for every $S, S' \subseteq A$, it holds that $f(S) + f(S') \geq f(S \cup S')$.
- *supermodular* if for every $S, S' \subseteq A$ with $S \subseteq S'$ and any $j \in A$ it holds that $f(j | S) \leq f(j | S')$.

All functions in this list except for supermodular belong to the hierarchy of complement-free set functions. It is well known that additive \subseteq gross substitutes \subseteq submodular \subseteq XOS \subseteq subadditive, and that all containment relations are strict (B. Lehmann, D. Lehmann, and Nisan, 2006).

We consider two standard primitives for accessing combinatorial set functions. A *value oracle* is given a set S and returns $f(S)$. A *demand oracle* is given a set of non-negative prices $p_1, \dots, p_m \in \mathbb{R}_{\geq 0}$ and returns a set S that maximizes $f(S) - \sum_{j \in S} p_j$.

3 Submodular and XOS Rewards

In this section, we show that, when the rewards are submodular or more generally XOS, there is a constant gap between the principal's utility under the best coarse-correlated equilibrium and the best pure Nash equilibrium. This result is tight in several ways. First, as we demonstrate in Example A.1, there is at least a constant gap between mixed Nash equilibria and pure Nash equilibria, even with gross-substitutes rewards and binary actions. Moreover, as we show in Proposition 4.1, for subadditive rewards there is at least a polynomial gap, even between mixed and pure Nash equilibria and the special case of binary actions.

Theorem 3.1 (Black-Box Lifting Theorem). *Suppose f is XOS. Let $\vec{\alpha}^*$ be any contract and let \mathcal{D}^* be any coarse-correlated equilibrium of $\vec{\alpha}^*$. Then there exists a contract $\vec{\alpha}$ and a pure Nash equilibrium S of $\vec{\alpha}$ such that $(1 - \sum_i \alpha_i) \cdot f(S) \geq \Omega(1) \cdot (1 - \sum_i \alpha_i^*) \cdot \mathbf{E}_{S^* \sim \mathcal{D}^*}[f(S^*)]$. Moreover, given $\vec{\alpha}^*$ and \mathcal{D}^* , one can find such $\vec{\alpha}$ and S in polynomial time (in n, m and in the support size of \mathcal{D}^*) using value and demand oracles to f .*

To prove Theorem 3.1 we proceed in two steps. In Section 3.1, we present our key new lemma driving this result, the Scaling-for-Existence Lemma. Afterwards, in Section 3.2, we show how to use this lemma to establish the theorem. Finally, in Section 3.3 we show how to leverage the theorem to obtain a poly-time algorithm for submodular rewards, that computes a contract such that any CCE under that contract provides a constant approximation to the best CCE under the best contract.

Remark 3.2. We note that the proof of Theorem 3.1 does not use the additivity of the cost functions, and thus Theorem 3.1 holds even when each agent has an arbitrary normalized (not necessarily monotone) non-negative combinatorial cost function $c_i : 2^{A_i} \rightarrow \mathbb{R}_{\geq 0}$.

3.1 The Scaling-for-Existence Lemma

Our key tool for establishing Theorem 3.1 is the following Scaling-for-Existence Lemma. The starting point of this lemma is a coarse-correlated equilibrium \mathcal{D} for some contract $\vec{\alpha}$ — or, more precisely, any distribution \mathcal{D} over sets that satisfies the weaker dropout-stability condition for contract $\vec{\alpha}$. The lemma then establishes the existence of a pure Nash equilibrium at an appropriately scaled contract $\vec{\alpha}'$, that achieves high reward relative to the original distribution \mathcal{D} . Our proof relies

Algorithm 1 Scaling-for-Existence for XOS Rewards

Input: Costs $c_1, \dots, c_m \in \mathbb{R}_{\geq 0}$, demand oracle access to a XOS function $f : 2^A \rightarrow \mathbb{R}_{\geq 0}$, a subset of agents $N' \subseteq N$, a parameter $\gamma > 1$, a contract $\vec{\alpha}$ and a corresponding CCE \mathcal{D} .

Output: A contract $\vec{\alpha}'$ and a PNE S' of $\vec{\alpha}'$ with $f(S') \geq (1 - \frac{1}{\gamma}) \cdot \mathbf{E}_{S \sim \mathcal{D}}[f(\bigcup_{i \in N'} S_i)]$.

- 1: Let $\alpha'_i = \gamma \cdot \alpha_i \cdot \mathbb{1}[i \in N']$
 - 2: Let \vec{p} be the price vector where $p(j) = \frac{c_j}{\alpha'_i}$ for each $j \in A_i$ \triangleright Here, $\frac{0}{0} = 0$ and $\frac{c}{0} = \infty$ for $c > 0$
 - 3: Let $S' \in \arg \max_{S \subseteq A} (f(S) - \sum_{j \in S} p(j))$
 - 4: **return** $\vec{\alpha}', S'$
-

on the property that the expected potential value of any dropout-stable distribution must be non-negative for XOS reward functions (even when restricted to a subset of agents). This implies that if a contract is scaled, then there is a PNE with high potential, which bounds from below the reward.

Lemma 3.3 (Scaling-for-Existence Lemma). *Suppose f is XOS. Let \mathcal{D} be a dropout-stable distribution with respect to $\vec{\alpha}$. For any set of agents $N' \subseteq N$ and $\gamma > 1$, let $\vec{\alpha}'$ be defined by $\alpha'_i = \gamma \cdot \alpha_i$ for $i \in N'$ and $\alpha'_i = 0$ otherwise. Then, there exists a pure Nash equilibrium S' with respect to $\vec{\alpha}'$, satisfying*

$$f(S') \geq (1 - \frac{1}{\gamma}) \cdot \mathbf{E}_{S \sim \mathcal{D}}[f(\bigcup_{i \in N'} S_i)].$$

Moreover, $\vec{\alpha}'$ and S' can be found in polynomial time (in n, m and in the support size of \mathcal{D}^*) with demand oracle access to f (see Algorithm 1).

Proof. Consider any N' . We use notation $S_{N'} = \bigcup_{i \in N'} S_i$. Fix a set S in the support of \mathcal{D} . Because f is XOS, there is an additive function a such that $f(S) = a(S)$ and $f(S'') \geq a(S'')$ for all S'' . Therefore, we have

$$\sum_{i \in N'} f(S_i | S_{-i}) = \sum_{i \in N'} (f(S) - f(S_{-i})) \leq \sum_{i \in N'} (a(S) - a(S_{-i})) = \sum_{i \in N'} a(S_i) = \sum_{i \in N'} a(S_{N'}) \leq f(S_{N'}). \quad (2)$$

Now for any set S define $\Phi(S, \vec{\alpha}) = f(S) - \sum_i \frac{c(S_i)}{\alpha_i}$ where $\frac{0}{0}$ is interpreted as 0, and $\frac{c}{0}$ for a positive c is interpreted as ∞ . We observe that for any agent $i \in N$ with $\alpha_i = 0$, by dropout-stability it holds that $\mathbf{E}_{S \sim \mathcal{D}}[c(S_i)] = 0$ and therefore $\Phi(S, \vec{\alpha})$ is always finite for S in the support of a distribution \mathcal{D} that is dropout-stable with respect to $\vec{\alpha}$.

Then we have

$$\Phi(S_{N'}, \vec{\alpha}) = f(S_{N'}) - \sum_{i \in N'} \frac{c(S_i)}{\alpha_i} \geq \sum_{i \in N'} f(S_i | S_{-i}) - \sum_{i \in N'} \frac{c(S_i)}{\alpha_i} = \sum_{i \in N'} \left(f(S_i | S_{-i}) - \frac{c(S_i)}{\alpha_i} \right),$$

where the inequality follows by Inequality (2).

Observe that dropout-stability is equivalent to $\mathbf{E}_{S \sim \mathcal{D}}[\alpha_i \cdot f(S_i | S_{-i}) - c(S_i)] \geq 0$ for all i . Then, by linearity of expectation and dropout-stability, it follows that

$$\mathbf{E}_{S \sim \mathcal{D}}[\Phi(S_{N'}, \vec{\alpha})] \geq 0. \quad (3)$$

Let S' be a set of actions maximizing $\Phi(S', \vec{\alpha}')$ (when fixing $\vec{\alpha}'$ as defined in the statement of the lemma). Then we have

$$\Phi(S', \vec{\alpha}') \geq \mathbf{E}_{S \sim \mathcal{D}}[\Phi(S_{N'}, \gamma \vec{\alpha})] = \mathbf{E}_{S \sim \mathcal{D}} \left[f(S_{N'}) - \sum_{i \in N'} \frac{c(S_i)}{\gamma \vec{\alpha}_i} \right]$$

$$= \left(1 - \frac{1}{\gamma}\right) \mathbf{E}_{S \sim \mathcal{D}}[f(S_{N'})] + \frac{1}{\gamma} \mathbf{E}_{S \sim \mathcal{D}}[\Phi(S_{N'}, \vec{\alpha})] \geq \left(1 - \frac{1}{\gamma}\right) \mathbf{E}_{S \sim \mathcal{D}}[f(S_{N'})],$$

where the first inequality follows by the maximality of S' , and the last inequality follows by Eq. (3).

As S' is a global maximum of $\Phi(\cdot, \vec{\alpha}')$, it is also a local maximum. Since $\Phi(\cdot, \vec{\alpha}')$ is a potential function for the game induced by the contract $\vec{\alpha}'$ (see Proposition 2.5), this means that S' is a pure Nash equilibrium with respect to contract $\vec{\alpha}'$. Finally, observe that one can find S' using a single demand query with prices $\frac{c_j}{\alpha'_j}$ for each action $j \in A_i$ (where $\frac{0}{0}$ is interpreted as 0, and $\frac{c}{0}$ for $c > 0$ is interpreted as ∞). The “moreover” part of the lemma thus follows by noting that Algorithm 1 sets $\vec{\alpha}'$ as stated in the lemma, and chooses S' as a demand set at these prices. \square

3.2 Proof of Theorem 3.1

We are now ready to prove Theorem 3.1. The high-level idea is to distinguish cases based on the correlated equilibrium \mathcal{D}^* at contract α^* , and whether there is a “significant” agent, namely, an agent such that $\alpha_i^* > 3/4$ and $(1 - \alpha_i^*) \cdot \mathbf{E}[f(S_i^*)] \geq 4 \cdot \mathbf{E}[f(S_{-i}^*)]$, or not. If there is a significant agent, we show that we can get a good pure Nash equilibrium by incentivizing that agent alone.

If there is no significant agent, then either (i) there is an agent with $\alpha_i^* > 3/4$ but $(1 - \alpha_i^*) \cdot \mathbf{E}[f(S_i^*)] \leq 4 \cdot \mathbf{E}[f(S_{-i}^*)]$, or (ii) $\alpha_i^* < 3/4$ for all agents i . In case (i), we show that dropping agent i and applying the Scaling-for-Existence Lemma to the remaining agents yields a good pure Nash equilibrium. (Note that there can be at most one agent with $\alpha_i^* > 3/4$ and that $\sum_{i' \neq i} \alpha_{i'}^* \leq 1/4$.) In case (ii), we argue that the agents can be partitioned into two groups B_1, B_2 such that $\sum_{i' \in B_\ell} \alpha_{i'}^* \leq 3/4$ for $\ell \in \{1, 2\}$, and applying the Scaling-for-Existence Lemma to the better of the groups gives a good pure Nash equilibrium.

Proof of Theorem 3.1. Consider any contract $\vec{\alpha}^*$ and any coarse-correlated equilibrium \mathcal{D}^* with respect to $\vec{\alpha}^*$, given as input to Algorithm 2. We analyze the guarantee provided by the contract computed by this algorithm. In the remainder of the proof, all expectations are over S^* that is distributed according to \mathcal{D}^* . We consider three cases:

Case A: There exists an agent i with $\alpha_i^* > 3/4$ and $(1 - \alpha_i^*) \cdot \mathbf{E}[f(S_i^*)] \geq 4 \cdot \mathbf{E}[f(S_{-i}^*)]$. In this case, since $\alpha_i^* > 3/4$, we have $\mathbf{E}[f(S_i^*)] \geq 16 \cdot \mathbf{E}[f(S_{-i}^*)]$. By subadditivity of f , this implies that $\mathbf{E}[f(S_i^*)] \geq 16 \cdot \mathbf{E}[(f(S^*) - f(S_i^*))]$, or equivalently, $\mathbf{E}[f(S_i^*)] \geq \frac{16}{17} \cdot \mathbf{E}[f(S^*)]$.

Consider contract $\vec{\alpha}$ with $\alpha_i = \frac{1+\alpha_i^*}{2}$ and $\alpha_j = 0$ for $j \neq i$. Note that $\alpha_i > \alpha_i^*$. Also note that for agents $j \neq i$ doing nothing is a best response no matter what the other agents do. Let $S = (S_i, \emptyset)$ be any pure Nash equilibrium of $\vec{\alpha}$ (such that $S'_j = \emptyset$ for all $j \neq i$). We next show that $f(S_i) \geq \frac{1}{2} \cdot \mathbf{E}[f(S_i^*)]$.

First observe that since $S = (S_i, \emptyset)$ is a pure Nash equilibrium of $\vec{\alpha}$, we have

$$\alpha_i f(S_i) - c(S_i) \geq \mathbf{E}[\alpha_i f(S_i^*) - c(S_i^*)]. \quad (4)$$

On the other hand, since \mathcal{D} is a coarse-correlated equilibrium of $\vec{\alpha}^*$, it must hold that

$$\mathbf{E}[\alpha_i^* f(S_i^* | S_{-i}^*) - c(S_i^*)] \geq \mathbf{E}[\alpha_i^* f(S_i | S_{-i}^*) - c(S_i)].$$

By subadditivity of f , we have $f(S_i^* | S_{-i}^*) = f(S_i^* \cup S_{-i}^*) - f(S_{-i}^*) \leq f(S_i^*)$. By monotonicity of f , we have $f(S_i | S_{-i}^*) = f(S_i \cup S_{-i}^*) - f(S_{-i}^*) \geq f(S_i) - f(S_{-i}^*)$. We thus obtain

$$\mathbf{E}[\alpha_i^* f(S_i^*) - c(S_i^*)] \geq \mathbf{E}[\alpha_i^* f(S_i) - \alpha_i^* f(S_{-i}^*) - c(S_i)]. \quad (5)$$

Algorithm 2 Black-Box Lifting for XOS Rewards

Input: Costs $c_1, \dots, c_m \in \mathbb{R}_{\geq 0}$, value and demand oracle access to a XOS function $f : 2^A \rightarrow \mathbb{R}_{\geq 0}$, a contract $\vec{\alpha}^*$, and a corresponding CCE \mathcal{D}^* .

Output: A contract $\vec{\alpha}$, and a PNE S with $(1 - \sum_i \alpha_i) f(S) \geq \Omega(1)(1 - \sum_i \alpha_i^*) E_{S^* \sim \mathcal{D}^*}[f(S^*)]$.

```

1: if  $\max_j \alpha_j^* > 3/4$  then
2:   Let  $i = \arg \max_j \alpha_j^*$                                 ▷ There must be a unique maximum
3:   if  $(1 - \alpha_i^*) \cdot \mathbf{E}_{S^* \sim \mathcal{D}^*}[f(S_i^*)] \geq 4 \cdot \mathbf{E}_{S^* \sim \mathcal{D}^*}[f(S_{-i}^*)]$  then
4:     Set  $\vec{\alpha}$  such that  $\alpha_i = \frac{1+\alpha_i^*}{2}$  and  $\alpha_j = 0$  for  $j \neq i$ 
5:     Let  $S \in \arg \max_{S' \subseteq A_i} (f(S') - \frac{c(S')}{\alpha_i})$ 
6:     return  $\vec{\alpha}, S$ 
7:   else
8:     Apply Algorithm 1 with  $\mathcal{D}^*$ ,  $\vec{\alpha}^*$ ,  $N' = [n] \setminus \{i\}$  and  $\gamma = 2$  to obtain  $\vec{\alpha}$  and  $S$ 
9:     return  $\vec{\alpha}, S$ 
10:  end if
11: else
12:  Partition  $N$  into two bundles  $B_1, B_2$ , where for each  $\ell \in \{1, 2\}$ ,  $\sum_{i \in B_\ell} \alpha_i^* \leq 3/4$ 
13:  Let  $\ell \in \arg \max_{\ell' \in \{1, 2\}} \mathbf{E}_{S^* \sim \mathcal{D}^*}[f(\bigcup_{i \in B_{\ell'}} S_i^*)]$ 
14:  Apply Algorithm 1 with  $\mathcal{D}^*$ ,  $\vec{\alpha}^*$ ,  $N' = B_\ell$  and  $\gamma = \frac{7}{6}$  to obtain  $\vec{\alpha}$  and  $S$ 
15:  return  $\vec{\alpha}, S$ 
16: end if

```

By summing up (4) and (5), using linearity of expectation, we get

$$(\alpha_i - \alpha_i^*) f(S_i) \geq (\alpha_i - \alpha_i^*) \cdot \mathbf{E}[f(S_i^*)] - \alpha_i^* \cdot \mathbf{E}[f(S_{-i}^*)]. \quad (6)$$

Thus, using that $\alpha_i > \alpha_i^*$,

$$\begin{aligned} f(S_i) &\geq \mathbf{E}[f(S_i^*)] - \frac{\alpha_i^*}{\alpha_i - \alpha_i^*} \mathbf{E}[f(S_{-i}^*)] = \mathbf{E}[f(S_i^*)] - \frac{2\alpha_i^*}{1 - \alpha_i^*} \mathbf{E}[f(S_{-i}^*)] \\ &\geq \mathbf{E}[f(S_i^*)] - \frac{2\alpha_i^*}{1 - \alpha_i^*} \cdot \frac{1 - \alpha_i^*}{4} \cdot \mathbf{E}[f(S_i^*)] \geq \frac{1}{2} \mathbf{E}[f(S_i^*)], \end{aligned} \quad (7)$$

where the first inequality is by rearranging Inequality (6), the equality is by the definition of α_i , the second inequality is by the assumption of the case, and the last inequality is since $\alpha_i^* \leq 1$.

Overall, the principal's utility under contract $\vec{\alpha}$ and equilibrium S is

$$\begin{aligned} \left(1 - \sum_j \alpha_j\right) \cdot f(S) &= (1 - \alpha_i) \cdot f(S_i) = \frac{1}{2}(1 - \alpha_i^*) \cdot f(S_i) \geq \frac{1}{4}(1 - \alpha_i^*) \cdot \mathbf{E}[f(S_i^*)] \\ &\geq \frac{1}{4} \left(1 - \sum_j \alpha_j^*\right) \cdot \mathbf{E}[f(S_i^*)] \geq \frac{4}{17} \left(1 - \sum_j \alpha_j^*\right) \cdot \mathbf{E}[f(S^*)], \end{aligned}$$

where the second equality follows by the definition of α_i . This concludes the argument for this case.

Case B: There exists an agent i with $\alpha_i^* > 3/4$ and $(1 - \alpha_i^*) \cdot \mathbf{E}[f(S_i^*)] \leq 4 \cdot \mathbf{E}[f(S_{-i}^*)]$. In this case, let $\vec{\alpha}$ be the contract where $\alpha_i = 0$ and $\alpha_j = 2\alpha_j^*$ for $j \neq i$. By applying Lemma 3.3 on \mathcal{D}^* , $\vec{\alpha}^*$,

$N' = [n] \setminus \{i\}$ and $\gamma = 2$ we get that there exists a pure Nash equilibrium S with respect to contract $\vec{\alpha}$ such that

$$f(S) \geq \frac{1}{2} \mathbf{E} \left[f \left(\bigcup_{j \in N'} S_j^* \right) \right]. \quad (8)$$

We can bound the principal's utility under $\vec{\alpha}^*$ and \mathcal{D}^* by

$$\begin{aligned} \left(1 - \sum_j \alpha_j^*\right) \mathbf{E} [f(S^*)] &\leq \left(1 - \sum_j \alpha_j^*\right) \mathbf{E} [f(S_i^*)] + \left(1 - \sum_j \alpha_j^*\right) \mathbf{E} [f(S_{-i}^*)] \\ &\leq (1 - \alpha_i^*) \mathbf{E} [f(S_i^*)] + \mathbf{E} [f(S_{-i}^*)] \leq 5 \cdot \mathbf{E} [f(S_{-i}^*)], \end{aligned} \quad (9)$$

where the first inequality is by subadditivity, and the last inequality is by the assumption of the case.

One the other hand, under contract $\vec{\alpha}$ (for which $\sum_j \alpha_j = \sum_{j \neq i} 2\alpha_j^* \leq 2(1 - \alpha_i^*) \leq \frac{1}{2}$), and equilibrium S , the principal's utility is

$$(1 - \sum_j \alpha_j) f(S) \stackrel{(8)}{\geq} (1 - \sum_{j \neq i} 2\alpha_j^*) \cdot \frac{1}{2} \mathbf{E} \left[f \left(\bigcup_{j \in N'} S_j^* \right) \right] \geq \frac{1}{4} \mathbf{E} [f(S_{-i}^*)] \stackrel{(9)}{\geq} \frac{1}{20} \left(1 - \sum_j \alpha_j^*\right) \mathbf{E} [f(S^*)],$$

which concludes the proof of the case.

Case C: $\alpha_i^* \leq 3/4$ for every i . We claim that the agents can be partitioned into two bundles B_1, B_2 , where for each $\ell \in \{1, 2\}$, $\sum_{i \in B_\ell} \alpha_i^* \leq 3/4$. To see this, consider the following process. We start by creating a separate bundle for each agent. Note that this way, by the assumption of the case, each bundle has sum of α_i^* at most $3/4$. Then, as long as there are two bundles with sum of α_i^* less than $3/4$ we merge them. This process is well-defined and terminates with two bundles with the desired property since as long as there are more than two bundles, since $\sum_i \alpha_i^* \leq 1$, there must be two bundles with sum of α_i^* at most $2/3$, so we can merge two of them.

Now, assume without loss of generality that

$$\mathbf{E} [f(\bigcup_{i \in B_1} S_i^*)] \geq \mathbf{E} [f(\bigcup_{i \in B_2} S_i^*)]. \quad (10)$$

Let $\vec{\alpha}$ be the contract where $\alpha_i = 0$ for $i \in B_2$ and $\alpha_i = \frac{7}{6}\alpha_i^*$ for $i \in B_1$. By applying Lemma 3.3 on \mathcal{D}^* , $\vec{\alpha}^*$, $N' = B_1$ and $\gamma = 7/6$ we get that there exists an equilibrium S with respect to contract $\vec{\alpha}$ such that

$$f(S) \geq \frac{1}{7} \mathbf{E} \left[f \left(\bigcup_{i \in B_1} S_i^* \right) \right]. \quad (11)$$

The principal's utility from $\vec{\alpha}$ and S is

$$\begin{aligned} (1 - \sum_i \alpha_i) f(S) &\geq (1 - \frac{7}{6} \sum_{i \in B_1} \alpha_i^*) \cdot \frac{1}{7} \cdot \mathbf{E} \left[f \left(\bigcup_{i \in B_1} S_i^* \right) \right] \geq (1 - \frac{7}{6} \cdot \frac{3}{4}) \cdot \frac{1}{7} \cdot \mathbf{E} \left[f \left(\bigcup_{i \in B_1} S_i^* \right) \right] \\ &\geq \frac{1}{56} \frac{\mathbf{E} [f(\bigcup_{i \in B_1} S_i^*)] + \mathbf{E} [f(\bigcup_{i \in B_2} S_i^*)]}{2} \geq \frac{\mathbf{E} [f(S^*)]}{112}, \end{aligned}$$

where the first inequality is by the definition of $\vec{\alpha}$ and by Inequality (11), the second inequality is since $\sum_{i \in B_1} \alpha_i^* \leq 3/4$, the third inequality is since by Inequality (10), and the last inequality is by subadditivity. This concludes the proof of the theorem. \square

3.3 Robustness and Tractability

We next derive our robust approximation results. To this end, we show that for submodular rewards it is possible to turn any contract $\vec{\alpha}^*$ and pure Nash equilibrium S^* of that contract into a contract $\vec{\alpha}$ such that any CCE under $\vec{\alpha}$ achieves a constant approximation to the principal's utility under S^* .

Theorem 3.4 (Black-Box Robustness Theorem). *Let f be a submodular reward function. There exists an algorithm (Algorithm 3 in Appendix B) that runs in polynomial time (in n and m) using only value queries to f , that, given a contract $\vec{\alpha}^*$ and a corresponding PNE S^* , outputs a contract $\vec{\alpha}$ such that for every CCE \mathcal{D} of $\vec{\alpha}$,*

$$\left(1 - \sum_i \alpha_i\right) \cdot \mathbf{E}_{S \sim \mathcal{D}}[f(S)] \geq \Omega(1) \cdot \left(1 - \sum_i \alpha_i^*\right) \cdot f(S^*).$$

Using (Dütting, Ezra, Feldman, and Kesselheim, 2023) and (Dütting, Ezra, Feldman, and Kesselheim, 2025) and combining with Theorem 3.1 we obtain the following corollary:

Corollary 3.5 (Efficient Robust Approximation Algorithms). *For submodular rewards there is an algorithm that runs in polynomial time using value and demand queries and finds a contract $\vec{\alpha}$ such that any CCE of $\vec{\alpha}$ obtains an $O(1)$ -approximation to the principal's optimal utility under the best CCE under any contract. For binary actions the same guarantee can be achieved with value queries only.*

The key tool for proving Theorem 3.4 is a Scaling-for-Robustness Lemma—a strengthened version of the Doubling Lemma from Dütting, Ezra, Feldman, and Kesselheim, 2025.² This lemma plays a role orthogonal to the Scaling-for-Existence Lemma: rather than showing the existence of a contract and a PNE which is good relative to a reference CCE, it is used to derive a contract under which *every* CCE is good with respect to a reference PNE. Similar to the Scaling-for-Existence Lemma, rather than working directly with the reference equilibrium, it starts from a dropout stable distribution over sets of actions.

Lemma 3.6 (Scaling-for-Robustness Lemma). *Suppose f is submodular. Let $\epsilon > 0$ and let $\vec{\epsilon} = (\epsilon, \dots, \epsilon) \in \mathbb{R}_+^n$. Let $\gamma > 1$. Let \mathcal{D} be a dropout-stable distribution with respect to $\vec{\alpha}$. Then any coarse correlated equilibrium (CCE) \mathcal{D}^\dagger with respect to $\gamma\vec{\alpha} + \vec{\epsilon}$ satisfies $\mathbf{E}_{S^\dagger \sim \mathcal{D}^\dagger}[f(S^\dagger)] \geq \frac{1}{2}(1 - \frac{1}{\gamma})\mathbf{E}_{S \sim \mathcal{D}}[f(S)]$.*

Proof. Let \mathcal{D} be a dropout-stable distribution with respect to $\vec{\alpha}$, and let \mathcal{D}^\dagger be a CCE with respect to $\gamma\vec{\alpha} + \vec{\epsilon}$. As \mathcal{D}^\dagger is a CCE with respect to $\gamma\vec{\alpha} + \vec{\epsilon}$, agent i weakly prefers taking action set S_i^\dagger drawn from \mathcal{D}^\dagger over S_i drawn independently from \mathcal{D} . That is,

$$\mathbf{E}_{S^\dagger \sim \mathcal{D}^\dagger}[(\gamma\alpha_i + \epsilon)f(S^\dagger) - c(S_i^\dagger)] \geq (\gamma\alpha_i + \epsilon)\mathbf{E}_{S^\dagger \sim \mathcal{D}^\dagger, S \sim \mathcal{D}}[f(S^\dagger_{-i} \cup S_i)] - c(S_i).$$

Rearranging the terms and taking the sum over all agents i , we obtain

$$\mathbf{E}_{S^\dagger \sim \mathcal{D}^\dagger, S \sim \mathcal{D}} \left[\sum_{i=1}^n f(S_i^\dagger | S_{-i}^\dagger) - \sum_{i=1}^n f(S_i | S_{-i}^\dagger) + \sum_{i=1}^n \frac{c(S_i) - c(S_i^\dagger)}{\gamma\alpha_i + \epsilon} \right] \geq 0. \quad (12)$$

²While the Doubling Lemma by Dütting, Ezra, Feldman, and Kesselheim, 2025 applies to any PNE of the scaled contract, our Scaling-for-Robustness Lemma applies even with respect to any CCE of the scaled contract.

By submodularity of f , for any S^\dagger it holds that

$$\sum_{i=1}^n f(S_i^\dagger | S_{-i}^\dagger) \leq f(S^\dagger), \quad (13)$$

while for any S, S^\dagger it holds that

$$\sum_{i=1}^n f(S_i | S_{-i}^\dagger) \geq f(S \setminus S_i^\dagger | S^\dagger) \quad (14)$$

Also, as \mathcal{D} is dropout-stable, we furthermore have for every agent i that $\mathbf{E}_{S \sim \mathcal{D}}[\alpha_i f(S) - c(S_i)] \geq \mathbf{E}_{S \sim \mathcal{D}}[\alpha_i f(S_{-i})]$, implying that

$$\frac{\mathbf{E}_{S \sim \mathcal{D}}[c(S_i)]}{\alpha_i} \leq \mathbf{E}_{S \sim \mathcal{D}}[f(S_i | S_{-i})]. \quad (15)$$

Therefore, it holds that

$$\begin{aligned} \mathbf{E}_{S^\dagger \sim \mathcal{D}^\dagger, S \sim \mathcal{D}} \left[\sum_{i=1}^n \frac{c(S_i) - c(S_i^\dagger)}{\gamma \alpha_i + \epsilon} \right] &\leq \sum_{i=1}^n \frac{\mathbf{E}_{S \sim \mathcal{D}}[c(S_i)]}{\gamma \alpha_i + \epsilon} \\ &\leq \frac{1}{\gamma} \sum_{i=1}^n \frac{\mathbf{E}_{S \sim \mathcal{D}}[c(S_i)]}{\alpha_i} \leq \frac{1}{\gamma} \sum_{i=1}^n \mathbf{E}_{S \sim \mathcal{D}}[f(S_i | S_{-i})] \leq \frac{\mathbf{E}_{S \sim \mathcal{D}}[f(S)]}{\gamma}, \end{aligned} \quad (16)$$

where the next-to-last inequality follows by Equation (15), and the final one holds by submodularity of f .

Combining Inequalities (13), (14), we get that for any S, S^\dagger ,

$$\sum_{i=1}^n f(S_i^\dagger | S_{-i}^\dagger) - \sum_{i=1}^n f(S_i | S_{-i}^\dagger) \leq f(S^\dagger) - f(S \setminus S_i^\dagger | S^\dagger) = 2f(S^\dagger) - f(S \cup S^\dagger) \leq 2f(S^\dagger) - f(S).$$

Taking expectation over the last inequality and combining it with Equation (12) and (16) gives

$$\begin{aligned} 0 &\leq \mathbf{E}_{S^\dagger \sim \mathcal{D}^\dagger, S \sim \mathcal{D}} \left[\sum_{i=1}^n f(S_i^\dagger | S_{-i}^\dagger) - \sum_{i=1}^n f(S_i | S_{-i}^\dagger) + \sum_{i=1}^n \frac{c(S_i) - c(S_i^\dagger)}{\gamma \alpha_i + \epsilon} \right] \\ &\leq \mathbf{E}_{S^\dagger \sim \mathcal{D}^\dagger, S \sim \mathcal{D}} \left[2f(S^\dagger) - \left(1 - \frac{1}{\gamma}\right)f(S) \right], \end{aligned}$$

yielding $\mathbf{E}_{S^\dagger \sim \mathcal{D}^\dagger, S \sim \mathcal{D}} \left[2f(S^\dagger) - \left(1 - \frac{1}{\gamma}\right)f(S) \right] \geq 0$, and thus $\mathbf{E}_{S^\dagger \sim \mathcal{D}^\dagger}[f(S^\dagger)] \geq \frac{1}{2} \left(1 - \frac{1}{\gamma}\right) \mathbf{E}_{S \sim \mathcal{D}}[f(S)]$, as desired. \square

The proof of Theorem 3.4 concludes by applying Lemma 3.6 to analyze the contract returned by Algorithm 3 (see Appendix B for details).

4 Subadditive Rewards

In this section, we show that for subadditive rewards, there is a $\Theta(\text{poly}(n))$ gap between the principal's utility under the best coarse-correlated equilibrium and the best pure Nash equilibrium. We first show that this gap arises already between mixed and pure Nash equilibria, and even with binary actions.

Proposition 4.1 (Subadditive Rewards). *There exists a binary-action instance with a subadditive reward function in which the gap between the principal's utility under the best MNE and under the best PNE is $\Omega(\sqrt{n})$.*

Proof. Consider a setting with $2n + 2$ agents denoted by $A = \{x, y\} \cup [2n]$. We define the reward function $f : 2^A \rightarrow \mathbb{R}_{\geq 0}$ in Table 2. The costs of agents x and y are 0, and the costs of all remaining agents are $c_i = \frac{2}{3n}$.

$ S \cap \{x, y\} $	0	$0 < i < 2n - 1$	$2n - 1$		$2n$
			$[n] \subseteq S$	$[n] \notin S$	
0	0	$2 + \frac{i}{\sqrt{n}}$	$2 + \frac{2n-1}{\sqrt{n}}$	$3 + \frac{2n-1}{\sqrt{n}}$	$4 + \frac{2n-1}{\sqrt{n}}$
1	4	$4 + \frac{i}{\sqrt{n}}$	$5 + \frac{2n-1}{\sqrt{n}}$	$4 + \frac{2n-1}{\sqrt{n}}$	$6 + \frac{2n-1}{\sqrt{n}}$
2	5	$5 + \frac{i}{\sqrt{n}}$	$5 + \frac{2n-1}{\sqrt{n}}$	$6 + \frac{2n-1}{\sqrt{n}}$	$7 + \frac{2n-1}{\sqrt{n}}$

Table 2: Definition of the monotone subadditive function $f : 2^{[2n] \cup \{x, y\}} \rightarrow \mathbb{R}_{\geq 0}$. The value of a set S is determined by $|S \cap \{x, y\}|$ and $|S \cap [2n]|$, and in the case $|S \cap [2n]| = 2n - 1$, by whether the missing element belongs to $[n]$.

Claim 4.2. *The function f is monotone and subadditive.*

Proof. We first prove that the function f is monotone. Afterwards, we show that it is subadditive.

Monotonicity. To see that f is monotone, we can observe that all of the marginals (presented in Tables 3 and 4) are non-negative.

$ S \cap \{x, y\} $	0	$0 < i < 2n - 1$	$2n - 1$		$2n$
			$[n] \subseteq S$	$[n] \notin S$	
0	4	2	3	1	2
1	1	1	0	2	1

Table 3: The marginal $f(j | S)$ for $j \in \{x, y\}$.

Subadditivity. To see that f is subadditive, observe that for every $T \neq \emptyset$, it holds that

$$f(T) \geq 2 + 2 \cdot \mathbb{1}[|\{x, y\} \cap T| \geq 1] + \mathbb{1}[|\{x, y\} \cap T| \geq 2] + \frac{|T \setminus \{x, y\}|}{\sqrt{n}}, \quad (17)$$

on the other hand

$$f(T) \leq 4 + 2 \cdot \mathbb{1}[|\{x, y\} \cap T| \geq 1] + \mathbb{1}[|\{x, y\} \cap T| \geq 2] + \frac{|T \setminus \{x, y\}|}{\sqrt{n}}. \quad (18)$$

S ∩ {x, y}	S ∩ [2n]	0	0 < i < 2n - 2	2n - 2		2n - 1	
				[n] ∖ {j} ⊆ S	[n] ∖ {j} ∉ S	j ∈ [n]	j ∉ [n]
0	0	$2 + \frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n}}$	$1 + \frac{1}{\sqrt{n}}$	1	2
1	1	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n}}$	$1 + \frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n}}$	2	1
2	2	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n}}$	$\frac{1}{\sqrt{n}}$	$1 + \frac{1}{\sqrt{n}}$	1	2

Table 4: The marginal $f(j | S)$ for $j \in [2n]$.

Thus,

$$\begin{aligned}
f(S) + f(T) &\geq 2 + 2 + 2 \cdot \mathbb{1}[|\{x, y\} \cap S| \geq 1] + 2 \cdot \mathbb{1}[|\{x, y\} \cap T| \geq 1] \\
&+ \mathbb{1}[|\{x, y\} \cap S| \geq 2] + \mathbb{1}[|\{x, y\} \cap T| \geq 2] + \frac{|S \setminus \{x, y\}|}{\sqrt{n}} + \frac{|T \setminus \{x, y\}|}{\sqrt{n}} \\
&\geq 4 + 2 \cdot \mathbb{1}[|\{x, y\} \cap (S \cup T)| \geq 1] + \mathbb{1}[|\{x, y\} \cap (S \cup T)| \geq 2] + \frac{|S \cup T \setminus \{x, y\}|}{\sqrt{n}} \\
&\geq f(S \cup T),
\end{aligned}$$

where the first inequality is by Inequality (17), the second inequality is by subadditivity of the indicator function, and since $\frac{|S \cap [2n]|}{\sqrt{n}}$ is an additive function, and the last inequality is by Inequality (18). This concludes the proof of the claim. \square

Claim 4.3. *No PNE achieves a principal utility of more than 6.5.*

Proof. Let S be some PNE. If $|S \cap [2n]| \leq 1$ then the utility of the principal is bounded by the reward which is bounded by $5 + \frac{1}{\sqrt{n}}$. If $1 < |S \cap [2n]| < 2n - 1$ then the marginal of all agents in $S \cap [2n]$ are $\frac{1}{\sqrt{n}}$, which means that the principal's utility is bounded by $f(S)(1 - |S \cap [2n]| \cdot \frac{c_i}{\sqrt{n}})$, which is negative for $|S \cap [2n]| > \frac{3\sqrt{n}}{2}$, thus, the utility of principal is bounded by $5 + \frac{3\sqrt{n}/2}{\sqrt{n}} = 6.5$. For S with $|S \cap [2n]| = 2n - 1$, the marginals of all the $2n - 1$ agents in $|S \cap [2n]|$ is bounded by $1 + \frac{1}{\sqrt{n}}$, thus the fraction that remains with the principal is $1 - (2n - 1) \frac{2/3n}{1 + \frac{1}{\sqrt{n}}} < 0$, which means that the principal obtains a negative utility from incentivizing this set. If $|S \cap [2n]| = 2n$, then half of the agents (in $[2n]$) have a marginal of 1, and the remaining half have marginals of 2. Thus, the fraction of the reward that remains with the principal is $1 - n \frac{2/3n}{1} - n \frac{2/3n}{2} = 0$, which means that the principal cannot obtain positive utility from incentivizing this set. \square

Claim 4.4. *There exists a MNE that achieves a principal utility of $\Omega(\sqrt{n})$.*

Proof. Consider the MNE where all agents in $[2n]$ take action with a probability of 1, and agents x, y take action (each) with a probability of $\frac{1}{2}$. Since the costs of x, y are zero, we could use $\alpha_x = \alpha_y = 0$. For the remaining agents, the expected marginal contribution of them to the reward is 1.5 since with probability $\frac{1}{2}$ their marginal contribution is 1, and with probability half their marginal contribution is 2. Thus, it is sufficient to use contract $\frac{c_i}{1.5} = \frac{4}{9n}$ to incentivize them. Overall, the

utility of the principal from this MNE is $E[f(S)](1 - 2n \cdot \frac{4}{9n}) = (\frac{1}{4} \cdot 4 + \frac{1}{2} \cdot 6 + \frac{1}{4} \cdot 7 + \frac{2n-1}{\sqrt{n}}) \cdot \frac{1}{9} = \Omega(\sqrt{n})$, which concludes the proof. \square

The proof of the proposition follows from Claims 4.2, 4.3, and 4.4. \square

We next show that the gap between the principal's utility under the best CCE and the best PNE is at most polynomial in the number of agents.

Proposition 4.5. *The gap between the principal's utility under the best CCE and under the best PNE is at most $O(n)$.*

The proof of Proposition 4.5 is deferred to Appendix C.

Remark 4.6. We note that Dütting, Ezra, Feldman, and Kesselheim (2025) show that the principal's utility under the best PNE is at least a $\frac{1}{m}$ -fraction of the optimal welfare (which trivially bounds the principal's utility under the best CCE). This implies an upper bound of m on the gap between the utility under the best CCE and the best PNE for subadditive reward functions. Proposition 4.5 establishes a stronger upper bound that does not depend on the number of actions, but only on the number of agents.

5 Supermodular Rewards

In this section, we establish our results for supermodular rewards. We first show that with binary actions, there is no gap between the principal's utility in the best coarse-correlated equilibrium and the best pure Nash equilibrium.

Theorem 5.1 (Supermodular Rewards, Binary Actions). *There exists no gap between the principal's utility under the best CCE and the best PNE in binary-action settings.*

Proof. Consider a contract $\tilde{\alpha}$, and a CCE \mathcal{D} . Let S' be the union of the sets of agents that are in the support of \mathcal{D} . Let $\tilde{\alpha}'$ be the contract for which $\alpha'_i = 0$ for $i \notin S'$, and $\alpha'_i = \alpha_i$ for $i \in S'$. We prove that S' is a pure Nash equilibrium with respect to $\tilde{\alpha}'$.

In the remainder of the proof, all expectations (and probabilities) are over S , which is distributed according to \mathcal{D} . As \mathcal{D} is a coarse-correlated equilibrium, no agent can improve their utility by unilaterally not working anymore. That is, for all $i \in S'$, we have

$$\mathbf{E}[\alpha_i \cdot f(S) - c_i \mathbb{1}[i \in S]] \geq \mathbf{E}[\alpha_i \cdot f(S \setminus \{i\})],$$

or equivalently

$$\mathbf{E}[\alpha_i \cdot (f(S) - f(S \setminus \{i\})) - c_i \mathbb{1}[i \in S]] \geq 0.$$

For every set S in the support of \mathcal{D} and every agent $i \in A$ we have

$$f(i | S \setminus \{i\}) = (f(S) - f(S \setminus \{i\})) \cdot \mathbb{1}[i \in S] \leq f(i | S' \setminus \{i\}) \cdot \mathbb{1}[i \in S],$$

where the last inequality is since f is supermodular, and $S \subseteq S'$. In combination, this means that

$$\Pr[i \in S](\alpha_i \cdot f(i | S' \setminus \{i\}) - c_i) = \mathbf{E}[\alpha_i \cdot f(i | S' \setminus \{i\}) \cdot \mathbb{1}[i \in S] - c_i \cdot \mathbb{1}[i \in S]] \geq 0.$$

So, if $i \in S'$, then $\alpha_i \cdot f(i | S' \setminus \{i\}) - c_i \geq 0$, meaning that $\alpha_i f(S') - c_i \geq \alpha_i f(S' \setminus \{i\})$, and therefore $\alpha'_i f(S') - c_i \geq \alpha'_i f(S' \setminus \{i\})$, which means that agent i does not want to deviate. For $i \notin S'$, it holds that $\alpha'_i \cdot f(i | S') = 0 \leq c_i$, which means that agent i does not want to deviate. Overall, we deduce that S' is a pure Nash equilibrium with respect to contract $\tilde{\alpha}'$. \square

We next consider the general (non-binary) setting for which we show that there is no gap between CE and PNE.

Theorem 5.2 (Supermodular Rewards, General Actions). *There exists no gap between the utility of the principal under the best CE and the best PNE.*

Proof. Consider a contract $\bar{\alpha}$ and a correlated equilibrium \mathcal{D} with respect to this contract. Let $T = \cup_{S \in \text{sup}(\mathcal{D})} S$. Before proving the theorem we are first going to prove the following key lemma.

Lemma 5.3. *For every agent i , assuming that every agent $j \neq i$ selects a set of actions S_j such that $T_j \subseteq S_j$ (where we denote by $S_{-i} = \cup_{j \neq i} S_j$), then there exists a set S_i for which $T_i \subseteq S_i$ that is agent i 's best response. I.e.,*

$$S_i \in \arg \max_{X_i \subseteq A_i} \alpha_i f(X_i \cup S_{-i}) - c(X_i).$$

Proof. Let S_i^1, \dots, S_i^k be the support of the sets taken by agent i according to \mathcal{D} (ordered arbitrarily). Consider an arbitrary set R_i in $\arg \max_{X_i \subseteq A_i} \alpha_i f(X_i \cup S_{-i}) - c(X_i)$. For $j = 0, \dots, k$, let $R_i^j = R_i \cup \cup_{\ell=1}^j S_i^\ell$ (where $R_i^0 = R_i$). We denote by $\mathcal{D}_{-i}(S_i^j)$ the distribution of the set of actions suggested to agents $N \setminus \{i\}$, conditioned on agent i suggested the action set S_i^j . By that \mathcal{D} is a correlated equilibrium, we have that

$$\alpha_i \cdot \mathbf{E}_{X_{-i} \sim \mathcal{D}_{-i}(S_i^j)} [f(S_i^j \cup X_{-i})] - c(S_i^j) \geq \alpha_i \cdot \mathbf{E}_{X_{-i} \sim \mathcal{D}_{-i}(S_i^j)} [f((S_i^j \cap R_i^{j-1}) \cup X_{-i})] - c(S_i^j \cap R_i^{j-1}),$$

or by rearranging, and since $S_i^j \setminus R_i^{j-1} = R_i^j \setminus R_i^{j-1}$ we get that

$$\alpha_i \cdot \mathbf{E}_{X_{-i} \sim \mathcal{D}_{-i}(S_i^j)} [f(R_i^j \setminus R_i^{j-1} \mid (S_i^j \cap R_i^{j-1}) \cup X_{-i})] \geq c(R_i^j \setminus R_i^{j-1}).$$

Now, since (1) $X_{-i} \subseteq S_{-i}$ for each realization in the support of $\mathcal{D}_{-i}(S_i^j)$, (2) $S_i^j \cap R_i^{j-1} \subseteq R_i^{j-1}$, (3) $R_i^j \setminus R_i^{j-1}$ is disjoint from $R_i^{j-1} \cup S_{-i}$, and (4) f is supermodular, we deduce that

$$\alpha_i \cdot \mathbf{E}_{X_{-i} \sim \mathcal{D}_{-i}(S_i^j)} [f(R_i^j \setminus R_i^{j-1} \mid R_i^{j-1} \cup S_{-i})] \geq c(R_i^j \setminus R_i^{j-1}).$$

Since the last expression is deterministic (as it does not depend on the realization of X_{-i}), we conclude that

$$\alpha_i \cdot f(R_i^j \setminus R_i^{j-1} \mid R_i^{j-1} \cup S_{-i}) \geq c(R_i^j \setminus R_i^{j-1}).$$

By summing over $j \in [k]$ we obtain that

$$\alpha_i \cdot (f(R_i^k \cup S_{-i}) - f(R_i^0 \cup S_{-i})) \geq c(R_i^k \setminus R_i^0),$$

which by rearrangement we conclude that

$$\alpha_i \cdot f(R_i^k \cup S_{-i}) - c(R_i^k) \geq \alpha_i \cdot f(R_i^0 \cup S_{-i}) - c(R_i^0).$$

Since $R_i^0 = R_i$, and $R_i \in \arg \max_{X_i \subseteq A_i} \alpha_i f(X_i \cup S_{-i}) - c(X_i)$ this implies that

$$R_i^k \in \arg \max_{X_i \subseteq A_i} \alpha_i f(X_i \cup S_{-i}) - c(X_i).$$

This concludes the proof of the lemma since $T_i = \cup_j S_i^j \subseteq R_i^k$. \square

We are now ready to prove the theorem. For this, we observe that if we start with T , and as long as some agent i can strictly improve her utility, she improves her utility to a set containing T_i (which is without loss because of Lemma 5.3), this process maintains the property that the action profile always contains T . Moreover, the process must terminate after a finite number of steps, as at each step, the value of the potential function (see Equation (1)) strictly increases, and this is a finite function. \square

We next prove that Theorem 5.2 is tight by presenting an instance with a supermodular reward function, for which there is an unbounded gap in the principal's utility that can be obtained from a CCE compared to a PNE.

Proposition 5.4 (Supermodular Rewards, General Actions). *There exists an instance in which the gap between the principal's utilities under the best CCE and under the best PNE is unbounded.*

Proof. Consider an instance with two agents where $A_1 = \{1, 2\}$, and $A_2 = \{3\}$. Let $f : 2^A \rightarrow \mathbb{R}_{\geq 0}$ be as defined as in Table 5. The function f is monotone and supermodular. The costs of the actions are $c(1) = 0.75$, $c(2) = 8.5$ and $c(3) = 0.25$.

S	\emptyset	$\{1\}$	$\{2\}$	$\{3\}$	$\{1, 2\}$	$\{1, 3\}$	$\{2, 3\}$	$\{1, 2, 3\}$
$f(S)$	0	0	1	0	5.5	1	1	10

Table 5: A supermodular reward function f with unbounded gap between the best CCE and the best PNE.

Note that the only set of actions with a positive welfare is $\{1, 2, 3\}$. Thus, this is the only candidate for a PNE with a positive utility for the principal. To incentivize $\{1, 2, 3\}$, the principal must use a contract $\vec{\alpha}$ such that $\alpha_1 \cdot f(\{1, 2\} \mid \{3\}) - c(\{1, 2\}) \geq \alpha_1 \cdot f(\{1\} \mid \{3\}) - c(1)$ and $\alpha_2 f(\{3 \mid \{1, 2\}\}) \geq c(3)$, thus $\alpha_1 \geq 8.5/9 = 17/18$, and $\alpha_2 \geq 0.25/4.5 = 1/18$, thus the principal cannot obtain a positive utility.

Consider the contract $\vec{\alpha}_3 = (0.925, 1/18)$. It holds that the distribution \mathcal{D} where $\{1, 2, 3\}$ is suggested with probability 0.8, and \emptyset is suggested with probability 0.2 is a CCE. Indeed this is a CCE since the utility of agent 2 when following the suggestion is $0.8 \cdot (\alpha_2 \cdot f(\{1, 2, 3\}) - c(3)) = 11/45$, while if agent 2 does nothing, his utility is $0.8(\alpha_2 \cdot f(\{1, 2\})) = 11/45$, and if he always takes action 3, his utility is $0.8\alpha_2 f(\{1, 2, 3\}) + 0.2\alpha_2 f(3) - c(3) = 7/36 < 11/45$. For agent 1, following the suggestion of the principal leads to a utility of $0.8 \cdot (0.925 \cdot 10 - 9.25) = 0$. Now deviating to \emptyset leads to a utility of 0, deviating to $\{1\}$ leads to a utility of $0.8 \cdot 0.925 \cdot 1 - 0.75 = -0.01$, deviating to $\{2\}$ leads to a utility of $\alpha_1 \cdot 1 - c(2) = -7.575$, and deviating to $\{1, 2\}$ leads to a utility of $0.8 \cdot 0.925 \cdot 10 + 0.2 \cdot 0.925 \cdot 5.5 - 9.25 = -0.8325$. Thus, it is a CCE.

The utility of the principal under this CCE is $(1 - 0.925 - 1/18) \cdot 0.8 \cdot 10 = 7/45 > 0$, which concludes the proof. \square

6 General Rewards

In this section, we show a lower bound on the gap between the principal's utility from the best MNE and the best PNE for general reward functions (that are neither subadditive, nor supermodular). We show that this gap is unbounded, even for a constant number of agents and binary actions.

Proposition 6.1 (General Rewards). *There exists a binary-action instance in which the gap between the principal's utility under the best MNE and under the best PNE is unbounded.*

Proof. Consider a binary-action instance with four agents, i.e., $A = \{1, 2, 3, 4\}$. The reward function is defined by the monotone closure of the following values (i.e., the maximum over all defined subsets). Let

$$f(\{1, 2\}) = 2, f(\{1, 3\}) = f(\{2, 4\}) = 1, f(\{1, 2, 3\}) = f(\{1, 2, 4\}) = \phi + 1,$$

where $\phi = \frac{1+\sqrt{5}}{2} \approx 1.618$ (i.e., ϕ is the golden ratio). Note that f is not subadditive (as $f(1) + f(2) < f(\{1, 2\})$), nor supermodular (as $f(\{1, 2, 3\}) + f(\{1, 2, 4\}) \geq f(\{1, 2\}) + f(\{1, 2, 3, 4\})$).

The costs are

$$c(1) = c(2) = 1, \text{ and } c(3) = c(4) = 0.$$

We next calculate the utility of the principal under a PNE corresponding to all non-redundant sets of agents (sets that do not contain an agent with a 0 marginal, and non-zero cost). For this, we utilize the characterization of Babaioff, Feldman, and Nisan, 2006 for the principal's utility function as a function of the PNE for binary instances:

$$g(S) = f(S)(1 - \sum_{i \in S} \frac{c(i)}{f(i | S \setminus \{i\})}).$$

It holds that:

$$\begin{aligned} g(\{1, 2\}) &= f(\{1, 2\})(1 - \frac{c(1)}{f(1 | \{2\})} - \frac{c(2)}{f(2 | \{1\})}) = 0 \\ g(\{1, 3\}) &= f(\{1, 3\})(1 - \frac{c(1)}{f(1 | \{3\})}) = 0 \\ g(\{1, 3, 4\}) &= f(\{1, 3, 4\})(1 - \frac{c(1)}{f(1 | \{3, 4\})}) = 0 \\ g(\{2, 4\}) &= f(\{2, 4\})(1 - \frac{c(2)}{f(2 | \{4\})}) = 0 \\ g(\{2, 3, 4\}) &= f(\{2, 3, 4\})(1 - \frac{c(2)}{f(2 | \{3, 4\})}) = 0 \\ g(\{1, 2, 3\}) &= f(\{1, 2, 3\})(1 - \frac{c(1)}{f(1 | \{2, 3\})} - \frac{c(2)}{f(2 | \{1, 3\})}) = f(\{1, 2, 3\})(1 - \frac{1}{\phi+1} - \frac{1}{\phi}) = 0 \\ g(\{1, 2, 4\}) &= f(\{1, 2, 4\})(1 - \frac{c(1)}{f(1 | \{2, 4\})} - \frac{c(2)}{f(2 | \{1, 4\})}) = f(\{1, 2, 4\})(1 - \frac{1}{\phi} - \frac{1}{\phi+1}) = 0 \\ g(\{1, 2, 3, 4\}) &= f(\{1, 2, 3, 4\})(1 - \frac{c(1)}{f(1 | \{2, 3, 4\})} - \frac{c(2)}{f(2 | \{1, 3, 4\})}) = f(\{1, 2, 3, 4\})(1 - \frac{1}{\phi} - \frac{1}{\phi}) < 0 \end{aligned}$$

Thus, no PNE obtains a positive utility for the principal.

Consider the contract $\vec{\alpha} = (\frac{4}{5\phi}, \frac{4}{5\phi}, 0, 0)$, and the distribution \mathcal{D} over agents where agents 1,2 always take an action with probability 1, and agents 3 and 4 take an action with probability $1 - \frac{\phi}{2}$

each (independently). The expected marginal of agent 1 (similarly, of agent 2) is: $(1 - \frac{\phi}{2})^2 \cdot \phi + (\frac{\phi}{2})(1 - \frac{\phi}{2}) \cdot \phi + (1 - \frac{\phi}{2})(\frac{\phi}{2})(\phi + 1) + (\frac{\phi}{2})^2 \cdot 2 = \frac{5\phi}{4}$. Distribution \mathcal{D} is a MNE since agent 1 (similarly, agent 2) is paid $\frac{c(1)}{E_{S \sim \mathcal{D}}[f(1|S \setminus \{1\})]} = \frac{4}{5\phi}$. The principal's utility under this MNE is

$$(1 - 2 \cdot \frac{4}{5\phi}) \left(\frac{\phi + 1}{4} \cdot 2 + (1 - \frac{\phi + 1}{4}) \cdot (\phi + 1) \right) > 0,$$

which concludes the proof. \square

References

- Alon, Tal, Matteo Castiglioni, Junjie Chen, Tomer Ezra, Yingkai Li, and Inbal Talgam-Cohen (2025). “Multi-Project Contracts”. In: *Proceedings of the 26th ACM Conference on Economics and Computation*, pp. 580–598.
- Babaioff, Moshe, Michal Feldman, and Noam Nisan (2006). “Combinatorial agency”. In: *Proc. of EC 2006*, pp. 18–28.
- (2010). “Mixed Strategies in Combinatorial Agency”. In: *J. Artif. Intell. Res.* 38, pp. 339–369.
- Babaioff, Moshe, Michal Feldman, Noam Nisan, and Eyal Winter (2012). “Combinatorial agency”. In: *J. Econ. Theory* 3.147, pp. 999–1034.
- Cacciamani, Federico, Martino Bernasconi, Matteo Castiglioni, and Nicola Gatti (2024). “Multi-agent contract design beyond binary actions”. In: *Proc. of EC 2024*.
- Castiglioni, Matteo, Alberto Marchesi, and Nicola Gatti (2023). “Multi-agent contract design: How to commission multiple agents with individual outcomes”. In: *Proc. of EC 2023*, pp. 412–448.
- Dasaratha, Krishna, Benjamin Golub, and Anant Shah (2025). “Incentive Design with Spillovers”. In: *Proc. of EC 2025*.
- Deo-Campo Vuong, Ramiro, Shaddin Dughmi, Neel Patel, and Aditya Prasad (2024). “On Supermodular Contracts and Dense Subgraphs”. In: *Proc. of SODA 2024*, pp. 109–132.
- Dütting, Paul, Tomer Ezra, Michal Feldman, and Thomas Kesselheim (2021). “Combinatorial Contracts”. In: *Proc. of FOCS 2021*, pp. 815–826.
- (2023). “Multi-agent Contracts”. In: *Proc. of STOC 2023*, pp. 1311–1324.
- (2025). “Multi-Agent Combinatorial Contracts”. In: *Proc. of SODA 2025*, pp. 1857–1891.
- Dütting, Paul, Michal Feldman, and Yoav Gal-Tzur (2024). “Combinatorial Contracts Beyond Gross Substitutes”. In: *Proc. of SODA 2024*, pp. 92–108.
- Dütting, Paul, Michal Feldman, Yoav Gal-Tzur, and Aviad Rubinstein (2026). “When Contracts Get Complex: Information-Theoretic Barriers”. In: *Proc. of SODA 2026*. Forthcoming.
- Dütting, Paul, Michal Feldman, and Inbal Talgam-Cohen (2024). “Algorithmic Contract Theory: A Survey”. In: *Found. Trends Theor. Comput. Sci.* 16.3-4, pp. 211–412.
- Ezra, Tomer, Michal Feldman, and Maya Schlesinger (2024). “On the (In)approximability of Combinatorial Contracts”. In: *Proc. of ITCS 2024*. Vol. 287, 44:1–44:22.
- Feldman, Michal (2025). “Combinatorial Contract Design: Recent Progress and Emerging Frontiers”. In: *arXiv preprint arXiv:2510.15065*.
- Hann-Caruthers, Wade and Sumit Goel (2024). “Optimality of Weighted Contracts for Multi-agent Contract Design with a Budget”. In: *Proc. of EC 2024*, pp. 1295–1295.
- Koutsoupias, Elias and Christos Papadimitriou (1999). “Worst-case equilibria”. In: *Proc. of STACS 1999*, pp. 404–413.

- Lehmann, Benny, Daniel Lehmann, and Noam Nisan (2006). “Combinatorial auctions with decreasing marginal utilities”. In: *Games Econ. Behav.* 55.2, pp. 270–296.
- Monderer, Dov and Lloyd S. Shapley (1996). “Potential Games”. In: *Games Econ. Behav.* 14, pp. 124–143.
- Roughgarden, Tim (2015). “Intrinsic Robustness of the Price of Anarchy”. In: *J. ACM* 62.5, 32:1–32:42.
- Roughgarden, Tim and Éva Tardos (2002). “How bad is selfish routing?” In: *J. ACM* 49.2, pp. 236–259.
- Syrgkanis, Vasilis and Éva Tardos (2013). “Composable and efficient mechanisms”. In: *Proc. of STOC 2013*, pp. 211–220.

A Separation Example

The following example (adopted from Babaioff, Feldman, and Nisan, 2010, Example 3.1) with two identical agents each having a single action, shows that with submodular rewards the principal can strictly benefit from inducing a mixed Nash equilibrium rather than a pure Nash equilibrium. We remark that the reward function in this example is also gross-substitutes and not just submodular as every submodular function over two elements is also gross-substitutes.

Example A.1 (Separation Example). Consider an instance with two identical agents $A = \{1, 2\}$ with costs $c(1) = c(2) = c = 1$. The submodular reward function is such that $f(\emptyset) = 0$, $f(1) = f(2) = 180$, and $f(\{1, 2\}) = 200$.

Recall the definition of the function $g : 2^A \rightarrow \mathbb{R}$, introduced in the proof of Proposition 6.1. Under pure Nash equilibria, the best contracts for the possible action profiles \emptyset , $\{1\}$, $\{2\}$, and $\{1, 2\}$ yield the principal a utility of

$$\begin{aligned} g(\emptyset) &= 0, \\ g(1) &= f(1) \cdot \left(1 - \frac{c}{f(1)}\right) = 179, \\ g(2) &= f(2) \cdot \left(1 - \frac{c}{f(2)}\right) = 179, \quad \text{and} \\ g(\{1, 2\}) &= f(\{1, 2\}) \cdot \left(1 - \frac{c}{f(1 \mid \{2\})} - \frac{c}{f(2 \mid \{1\})}\right) = 180. \end{aligned}$$

Thus, the highest utility the principal can obtain in a pure Nash equilibrium is 180, by inducing both agents to exert effort.

Consider the contract $\vec{\alpha} = (\frac{1}{36}, \frac{1}{36})$. We show that the product distribution \mathcal{D} in which each agent takes an action with probability 0.9 is a MNE. Assuming that one agent takes an action with probability 0.9, then the utility of the other action as a function of the probability it takes an action is denoted by

$$u_a(p) = \frac{1}{36} (p \cdot (0.9 \cdot f(\{1, 2\}) + 0.1 \cdot f(1)) + (1-p) \cdot (0.9 \cdot f(1) + 0.1 \cdot f(\emptyset))) - p \cdot c = 4.5,$$

which means that the utility of the agent is independent of the probability of the agent taking action, which means that \mathcal{D} is indeed an MNE.

Algorithm 3 Black-Box Robustness for Submodular Rewards

Input: Costs $c_1, \dots, c_m \in \mathbb{R}_{\geq 0}$, value oracle access to a submodular function $f : 2^A \rightarrow \mathbb{R}_{\geq 0}$, a contract $\vec{\alpha}^*$, and a corresponding PNE S^* .

Output: A contract $\vec{\alpha}$ guaranteeing a principal's utility of at least $\Omega(1) \cdot (1 - \sum_i \alpha_i^*) f(S^*)$ for any CCE of $\vec{\alpha}$.

```

1: Let  $A_0 = \{j \in A \mid c_j = 0\}$  be the zero cost actions
2: if  $f(A_0) \geq \frac{2}{17}(1 - \sum_i \alpha_i^*) f(S^*)$  then
3:   return  $\vec{\alpha} = \vec{\epsilon}$  for  $\epsilon = \frac{1}{2n}$ 
4: else if  $\max_i \alpha_i^* > 3/4$  then
5:   Let  $i^* = \arg \max_i \alpha_i^*$  ▷ There must be a unique maximum
6:   if  $(1 - \alpha_{i^*}^*) \cdot f(S_{-i}^*) \geq 4 \cdot f(S_{-i}^*)$  then
7:     return  $\vec{\alpha}$  where  $\alpha_{i^*} = \frac{1 + \alpha_{i^*}^*}{2}$  and  $\alpha_i = 0$  for  $i \neq i^*$ 
8:   else
9:     return  $\vec{\alpha}$  where  $\alpha_{i^*} = 0$  and  $\alpha_i = 2\alpha_i^*$  for  $i \neq i^*$ 
10:  end if
11: else
12:   Partition  $N$  into two bundles  $B_1, B_2$ , where for each  $\ell \in \{1, 2\}$ ,  $\sum_{i \in B_\ell} \alpha_i^* \leq 3/4$ 
13:   Let  $\ell \in \arg \max_{\ell' \in \{1, 2\}} f(\bigcup_{i \in B_{\ell'}} S_i^*)$ 
14:   return  $\vec{\alpha}$  where  $\alpha_i = \frac{7}{6}\alpha_i^*$  for  $i \in B_\ell$  and  $\alpha_i = 0$  for  $i \notin B_\ell$ 
15: end if
```

The expected principal's utility under contract $\vec{\alpha}$ and distribution \mathcal{D} is

$$\mathbf{E}_{S \sim \mathcal{D}}[u_p(\vec{\alpha}, S)] = (1 - \frac{2}{36}) \left(\frac{1}{100} \cdot f(\emptyset) + \frac{9}{100} \cdot f(1) + \frac{9}{100} \cdot f(2) + \frac{81}{100} \cdot f(\{1, 2\}) \right) = \frac{17}{18} \cdot 194.4 = 183.6,$$

which is strictly higher than the optimal principal utility in a pure Nash equilibrium. This shows that the principal can strictly gain from inducing a mixed Nash equilibrium, rather than a pure Nash equilibrium.

B Proof of Theorem 3.4

The high-level idea is to distinguish cases based on the equilibrium S^* at contract $\vec{\alpha}^*$, and whether it is sufficient to incentivize only zero cost actions, or whether there is a “significant” agent, namely, an agent such that $\alpha_i^* > 3/4$ and $(1 - \alpha_i^*) \cdot f(S_i^*) \geq 4 \cdot f(S_{-i}^*)$, or not. If there is a significant agent, we show that we can get a good approximation under any CCE by aiming to incentivize that agent alone. If there is no significant agent, then either (i) there is an agent with $\alpha_i^* > 3/4$ but $(1 - \alpha_i^*) \cdot f(S_i^*) \leq 4 \cdot f(S_{-i}^*)$, or (ii) $\alpha_i^* < 3/4$ for all agents i . In case (i), we show that dropping agent i and applying the Scaling-for-Robustness Lemma to the remaining agents yields a good guarantee under any CCE. (Note that there can be at most one agent with $\alpha_i^* > 3/4$ and that $\sum_{i' \neq i} \alpha_{i'}^* \leq 1/4$.) In case (ii), we argue that the agents can be partitioned into two groups B_1, B_2 such that $\sum_{i' \in B_\ell} \alpha_{i'}^* \leq 3/4$ for $\ell \in \{1, 2\}$, and applying the Scaling-for-Robustness Lemma to the better of the groups guarantees a good principal's utility under any CCE.

Proof of Theorem 3.4. Consider any input $\vec{\alpha}^*$ and PNE S^* with respect to $\vec{\alpha}^*$, given as input to

Algorithm 3. We analyze the guarantee provided by the contract computed by this algorithm by distinguishing the following cases:

Case A: It holds that $f(A_0) \geq \frac{2}{17}(1 - \sum_i \alpha_i^*)f(S^*)$. In this case, since A_0 is a PNE with respect to the zero-contract $\vec{0}$ (a contract that pays nothing to all agents), by applying Lemma 3.6 with $\epsilon = \frac{1}{2n}$ and $\gamma = 2$ we obtain that under contract $\vec{\alpha} = \gamma\vec{0} + \vec{\epsilon} = \vec{\epsilon}$ any CCE \mathcal{D} with respect to $\vec{\alpha}$ satisfies $\mathbf{E}_{S \sim \mathcal{D}}[f(S)] \geq \frac{1}{2}(1 - \frac{1}{\gamma})f(A_0) = \frac{1}{4}f(A_0)$. Thus, the principal's utility under this case is at least

$$(1 - \sum_i \alpha_i)\mathbf{E}_{S \sim \mathcal{D}}[f(S)] \geq (1 - \frac{1}{2})\frac{1}{4}f(A_0) \geq \frac{1}{68}(1 - \sum_i \alpha_i^*)f(S^*),$$

which concludes this case.

Case B: There exists an agent i with $\alpha_i^* > 3/4$ and $(1 - \alpha_i^*) \cdot f(S_i^*) \geq 4 \cdot f(S_{-i}^*)$ (and we are not in Case A). In this case, since $\alpha_i^* > 3/4$, we have $f(S_i^*) \geq 16 \cdot f(S_{-i}^*)$. By subadditivity of f , this implies that $f(S_i^*) \geq 16 \cdot (f(S^*) - f(S_i^*))$, or equivalently,

$$f(S_i^*) \geq \frac{16}{17} \cdot f(S^*). \quad (19)$$

Consider contract $\vec{\alpha}$ with $\alpha_i = \frac{1+\alpha_i^*}{2}$ and $\alpha_j = 0$ for $j \neq i$. Note that $\alpha_i > \alpha_i^*$. Let \mathcal{D} be any CCE of $\vec{\alpha}$. We next show that $\mathbf{E}_{S \sim \mathcal{D}}[f(S)] \geq \frac{1}{4} \cdot f(S_i^*)$.

We have the following:

$$\begin{aligned} \mathbf{E}_{S \sim \mathcal{D}}[\alpha_i f(S_i) - c(S_i)] &\geq \mathbf{E}_{S \sim \mathcal{D}}[\alpha_i(f(S_i \cup S_{-i}) - f(S_{-i})) - c(S_i)] \\ &\geq \mathbf{E}_{S \sim \mathcal{D}}[\alpha_i(f(S_i^* \cup S_{-i}) - f(S_{-i}))] - c(S_i^*) \\ &\geq \alpha_i f(S_i^*) - \alpha_i \mathbf{E}_{S \sim \mathcal{D}}[f(S_{-i})] - c(S_i^*) \\ &\geq \alpha_i f(S_i^*) - \alpha_i f(A_0) - c(S_i^*), \end{aligned} \quad (20)$$

where the first inequality is by subadditivity, the second inequality is since \mathcal{D} is a CCE of $\vec{\alpha}$, the third inequality is by monotonicity of f , and the last inequality is since no agent but i will take non-zero cost actions in a CCE since $\alpha_j = 0$ for $j \neq i$. On the other hand, since S^* is a PNE of $\vec{\alpha}^*$, it must hold that

$$\alpha_i^* f(S_i^* \cup S_{-i}^*) - c(S_i^*) \geq \mathbf{E}_{S \sim \mathcal{D}}[\alpha_i^* f(S_i \cup S_{-i}^*) - c(S_i)].$$

By subadditivity of f , we have $f(S_i^* \cup S_{-i}^*) - f(S_{-i}^*) \leq f(S_i^*)$. Moreover, by monotonicity of f , we have $f(S_i \cup S_{-i}^*) \geq f(S_i)$. We thus obtain

$$\alpha_i^* f(S_i^*) - c(S_i^*) \geq \mathbf{E}_{S \sim \mathcal{D}}[\alpha_i^* f(S_i) - c(S_i)] - \alpha_i^* f(S_{-i}^*). \quad (21)$$

By summing up (20) and (21), we obtain

$$\begin{aligned} &\mathbf{E}_{S \sim \mathcal{D}}[\alpha_i f(S_i) - c(S_i)] + \alpha_i^* f(S_i^*) - c(S_i^*) \\ &\geq \mathbf{E}_{S \sim \mathcal{D}}[\alpha_i^* f(S_i) - c(S_i)] - \alpha_i^* f(S_{-i}^*) + \alpha_i f(S_i^*) - \alpha_i f(A_0) - c(S_i^*). \end{aligned}$$

Using linearity of expectation and that $\alpha_i \leq 1$ and $\alpha_i^* \leq 1$, we get

$$(\alpha_i - \alpha_i^*)\mathbf{E}_{S \sim \mathcal{D}}[f(S_i)] \geq (\alpha_i - \alpha_i^*) \cdot f(S_i^*) - \alpha_i f(A_0) - \alpha_i^* f(S_{-i}^*) \geq (\alpha_i - \alpha_i^*) \cdot f(S_i^*) - f(A_0) - f(S_{-i}^*). \quad (22)$$

Thus, using that $\alpha_i > \alpha_i^*$,

$$\begin{aligned}
\mathbf{E}_{S \sim \mathcal{D}}[f(S_i)] &\geq f(S_i^*) - \frac{f(S_{-i}^*) + f(A_0)}{\alpha_i - \alpha_i^*} = f(S_i^*) - \frac{2(f(S_{-i}^*) + f(A_0))}{1 - \alpha_i^*} \\
&\geq f(S_i^*) - \frac{2}{1 - \alpha_i^*} \cdot \left(\frac{1 - \alpha_i^*}{4} \cdot f(S_i^*) + \frac{2}{17}(1 - \sum_j \alpha_j^*) f(S^*) \right) \\
&\geq f(S_i^*) - \frac{2}{1 - \alpha_i^*} \cdot \left(\frac{1 - \alpha_i^*}{4} \cdot f(S_i^*) + \frac{2}{17}(1 - \alpha_i^*) \frac{17}{16} f(S_i^*) \right) = \frac{1}{4} f(S_i^*). \quad (23)
\end{aligned}$$

where the first inequality is by rearranging Inequality (22), the first equality is by the definition of α_i , the second inequality is by the assumption of the case, and the last inequality is by Inequality (19).

Overall, the principal's utility under contract $\vec{\alpha}$ and CCE \mathcal{D} is

$$\begin{aligned}
\left(1 - \sum_j \alpha_j^*\right) \cdot \mathbf{E}_{S \sim \mathcal{D}}[f(S)] &\geq (1 - \alpha_i) \cdot \mathbf{E}_{S \sim \mathcal{D}}[f(S_i)] = \frac{1}{2}(1 - \alpha_i^*) \cdot \mathbf{E}_{S \sim \mathcal{D}}[f(S_i)] \geq \frac{1}{8}(1 - \alpha_i^*) \cdot f(S_i^*) \\
&\geq \frac{1}{8} \left(1 - \sum_j \alpha_j^*\right) \cdot f(S_i^*) \geq \frac{2}{17} \left(1 - \sum_j \alpha_j^*\right) \cdot f(S^*),
\end{aligned}$$

where the first equality is by definition of $\vec{\alpha}$ and monotonicity, the equality follows by the definition of α_i , the second inequality holds by Inequality (23), and the final inequality is by Inequality (19). This concludes the argument for this case.

Case C: There exists an agent i with $\alpha_i > 3/4$ and $(1 - \alpha_i^*) \cdot f(S_i^*) \leq 4 \cdot f(S_{-i}^*)$ (and we are not in Case A). In this case, let $\vec{\alpha}$ be the contract where $\alpha_i = 0$ and $\alpha_j = 2\alpha_j^*$ for $j \neq i$. By applying Lemma 3.6 on S^* , $\vec{\alpha}^*$, $N' = [n] \setminus \{i\}$ and $\gamma = 2$ we get that any CCE \mathcal{D} with respect to contract $\vec{\alpha}$ satisfies

$$\mathbf{E}_{S \sim \mathcal{D}}[f(S)] \geq \frac{1}{4} f(\bigcup_{j \in N'} S_j^*). \quad (24)$$

We can bound the principal's utility under $\vec{\alpha}^*$, and S^* by

$$\begin{aligned}
\left(1 - \sum_j \alpha_j^*\right) f(S^*) &\leq \left(1 - \sum_j \alpha_j^*\right) f(S_i^*) + \left(1 - \sum_j \alpha_j^*\right) f(S_{-i}^*) \\
&\leq (1 - \alpha_i^*) f(S_i^*) + f(S_{-i}^*) \leq 5 \cdot f(S_{-i}^*),
\end{aligned} \quad (25)$$

where the first inequality is by subadditivity, the last inequality is by the assumption of the case.

One the other hand, under contract $\vec{\alpha}$ (for which $\sum_j \alpha_j = \sum_{j \neq i} 2\alpha_j^* \leq 2(1 - \alpha_i^*) \leq \frac{1}{2}$), and CCE \mathcal{D} , the principal's utility is

$$(1 - \sum_j \alpha_j) \mathbf{E}_{S \sim \mathcal{D}}[f(S)] \stackrel{(24)}{\geq} (1 - \sum_{j \neq i} 2\alpha_j^*) \cdot \frac{1}{4} f(\bigcup_{j \in N'} S_j^*) \geq \frac{1}{8} f(S_{-i}^*) \stackrel{(25)}{\geq} \frac{1}{40} \left(1 - \sum_j \alpha_j^*\right) f(S^*),$$

which concludes the proof of the case.

Case D: $\alpha_i^* \leq 3/4$ for every i (and we are not in case A). Note that we can partition the agents into two bundles B_1, B_2 , where for each $\ell \in \{1, 2\}$, $\sum_{i \in B_\ell} \alpha_i^* \leq 3/4$ (by the same argument as in the proof of Theorem 3.1). Now, assume without loss of generality that

$$f(\bigcup_{i \in B_1} S_i^*) \geq f(\bigcup_{i \in B_2} S_i^*). \quad (26)$$

Let $\vec{\alpha}$ be the contract where $\alpha_i = 0$ for $i \in B_2$ and $\alpha_i = \frac{7}{6}\alpha_i^*$ for $i \in B_1$. By applying Lemma 3.6 on S^* , $\vec{\alpha}^*$, $N' = B_1$ and $\gamma = 7/6$ we get that there any CCE \mathcal{D} with respect to contract $\vec{\alpha}$ satisfies

$$\mathbf{E}_{S \sim \mathcal{D}}[f(S)] \geq \frac{1}{14} f\left(\bigcup_{i \in B_1} S_i^*\right) \quad (27)$$

The principal's utility from $\vec{\alpha}$ and \mathcal{D} is

$$\begin{aligned} (1 - \sum_i \alpha_i) \mathbf{E}_{S \sim \mathcal{D}}[f(S)] &\geq (1 - \frac{7}{6} \sum_{i \in B_1} \alpha_i^*) \cdot \frac{1}{14} \cdot f\left(\bigcup_{i \in B_1} S_i^*\right) \geq (1 - \frac{7}{6} \cdot \frac{3}{4}) \cdot \frac{1}{14} \cdot f\left(\bigcup_{i \in B_1} S_i^*\right) \\ &\geq \frac{1}{112} \frac{f\left(\bigcup_{i \in B_1} S_i^*\right) + f\left(\bigcup_{i \in B_2} S_i^*\right)}{2} \geq \frac{f(S^*)}{224}, \end{aligned}$$

where the first inequality is by the definition of $\vec{\alpha}$ and by Inequality (27), the second inequality is since $\sum_{i \in B_1} \alpha_i^* \leq 3/4$, the third inequality is by Inequality (26), and the last inequality is by subadditivity. This concludes the proof of the theorem. \square

C Proof of Proposition 4.5

We first prove a weaker variant of the Scaling-for-Existence Lemma for subadditive functions.

Lemma C.1 (Scaling-for-Existence Lemma for Subadditive). *Suppose f is subadditive. Let \mathcal{D} be a dropout-stable distribution with respect to $\vec{\alpha}$. For any set of $i \in N$ and $\gamma > 1$, let $\vec{\alpha}'$ be defined by $\alpha'_i = \gamma \cdot \alpha_i$, and $\alpha'_{i'} = 0$ for $i' \neq i$. Then, there exists a pure Nash equilibrium S' with respect to $\vec{\alpha}'$, satisfying*

$$f(S') \geq (1 - \frac{1}{\gamma}) \cdot \mathbf{E}_{S \sim \mathcal{D}}[f(S_i)].$$

Proof. Fix a set S in the support of \mathcal{D} . Now for any set $S'_i \subseteq A_i$ define $\Phi(S'_i, \vec{\alpha}) = f(S'_i) - \frac{c(S'_i)}{\alpha_i}$ where $\frac{0}{0}$ is interpreted as 0, and $\frac{c}{0}$ for a positive c is interpreted as ∞ . Then we have

$$\mathbf{E}_{S \sim \mathcal{D}}[\Phi(S_i, \vec{\alpha})] = \mathbf{E}_{S \sim \mathcal{D}}[f(S_i) - \frac{c(S_i)}{\alpha_i}] \geq \mathbf{E}_{S \sim \mathcal{D}}[f(S_i | S_{-i}) - \frac{c(S_i)}{\alpha_i}] \geq 0, \quad (28)$$

where the first inequality is by subadditivity, and the last inequality is by dropout-stability.

Let $S'_i \subseteq A_i$ be a set of actions maximizing $\Phi(S'_i, \vec{\alpha}')$ (when fixing $\vec{\alpha}'$ as defined in the statement of the lemma). Then we have

$$\begin{aligned} \Phi(S'_i, \vec{\alpha}') &\geq \mathbf{E}_{S \sim \mathcal{D}}[\Phi(S_i, \vec{\alpha}')] = \mathbf{E}_{S \sim \mathcal{D}}\left[f(S_i) - \frac{c(S_i)}{\gamma \alpha_i}\right] \\ &= \left(1 - \frac{1}{\gamma}\right) \mathbf{E}_{S \sim \mathcal{D}}[f(S_i)] + \frac{1}{\gamma} \mathbf{E}_{S \sim \mathcal{D}}[\Phi(S_i, \vec{\alpha})] \geq \left(1 - \frac{1}{\gamma}\right) \mathbf{E}_{S \sim \mathcal{D}}[f(S_i)], \end{aligned}$$

where the first inequality follows by the maximality of S'_i , and the last inequality follows by Eq. (28).

As S'_i is a global maximum of $\Phi(\cdot, \vec{\alpha}')$, it is also a local maximum. Since $\Phi(\cdot, \vec{\alpha}')$ is a potential function for the game induced by the contract $\vec{\alpha}'$ (see Proposition 2.5), this means that S'_i is a pure Nash equilibrium with respect to contract $\vec{\alpha}'$. \square

We are now ready to prove Proposition 4.5. Consider any contract $\vec{\alpha}^*$ and any coarse-correlated equilibrium \mathcal{D} with respect to $\vec{\alpha}^*$. In the remainder of the proof, all expectations are over S^* that is distributed according to \mathcal{D} . We consider three cases:

Case A: There exists an agent i with $\alpha_i^* > 3/4$ and $(1 - \alpha_i^*) \cdot \mathbf{E}[f(S_i^*)] \geq 4 \cdot \mathbf{E}[f(S_{-i}^*)]$. The proof of this case is identical to the proof of Case A of the proof of Theorem 3.1.

Case B: There exists an agent i with $\alpha_i^* > 3/4$ and $(1 - \alpha_i^*) \cdot \mathbf{E}[f(S_i^*)] \leq 4 \cdot \mathbf{E}[f(S_{-i}^*)]$. In this case, let $i^* = \arg \max_{i' \neq i} \mathbf{E}[f(S_{i'})]$. Let $\vec{\alpha}'$ be the contract where $\alpha'_{i'} = 0$ for $i' \neq i^*$, and $\alpha'_{i^*} = 2\alpha_{i^*}$. By applying Lemma C.1 on \mathcal{D} , $\vec{\alpha}^*$, i^* and $\gamma = 2$ we get that there exists an equilibrium S' with respect to contract $\vec{\alpha}'$ such that

$$f(S') \geq \frac{1}{2} \mathbf{E}[f(S_{i^*}^*)]. \quad (29)$$

We can bound the expected principal's utility under $\vec{\alpha}^*$, and \mathcal{D} by

$$\begin{aligned} \left(1 - \sum_j \alpha_j^*\right) \mathbf{E}[f(S^*)] &\leq \left(1 - \sum_j \alpha_j^*\right) \mathbf{E}[f(S_i^*)] + \left(1 - \sum_j \alpha_j^*\right) \mathbf{E}[f(S_{-i}^*)] \\ &\leq (1 - \alpha_i^*) \mathbf{E}[f(S_i^*)] + \mathbf{E}[f(S_{-i}^*)] \leq 5 \cdot \mathbf{E}[f(S_{-i}^*)], \end{aligned} \quad (30)$$

where the first inequality is by subadditivity, the last inequality is by the assumption of the case.

One the other hand, under contract $\vec{\alpha}'$ (for which $\sum_j \alpha'_j = 2\alpha_{i^*}^* \leq 2(1 - \alpha_i^*) \leq \frac{1}{2}$), and equilibrium S' , the principal's utility is

$$(1 - \sum_j \alpha'_j) f(S') \stackrel{(29)}{\geq} (1 - 2\alpha_{i^*}^*) \cdot \frac{1}{2} \mathbf{E}[f(S_{i^*}^*)] \geq \frac{1}{4n} \mathbf{E}[f(S_{-i}^*)] \stackrel{(30)}{\geq} \frac{1}{20n} \left(1 - \sum_j \alpha_j^*\right) \mathbf{E}[f(S^*)],$$

where the second inequality is by subadditivity and the definition of i^* . This concludes the proof of the case.

Case C: $\alpha_i^* \leq 3/4$ for every i . In this case, let $i^* = \arg \max_i \mathbf{E}[f(S_i)]$. Let $\vec{\alpha}'$ be the contract where $\alpha'_i = 0$ for $i \neq i^*$, and $\alpha'_{i^*} = \frac{7}{6}\alpha_{i^*}$. By applying Lemma C.1 on \mathcal{D} , $\vec{\alpha}^*$, i^* and $\gamma = \frac{7}{6}$ we get that there exists an equilibrium S' with respect to contract $\vec{\alpha}'$ such that

$$f(S') \geq \frac{1}{7} \mathbf{E}[f(S_{i^*}^*)]. \quad (31)$$

Under contract $\vec{\alpha}'$ (for which $\sum_j \alpha'_j = \frac{7}{6}\alpha_{i^*}^* \leq \frac{7}{6} \cdot \frac{3}{4} = \frac{7}{8}$), and equilibrium S' , the principal's utility is

$$(1 - \sum_j \alpha'_j) f(S') \stackrel{(31)}{\geq} (1 - \frac{7}{8}) \cdot \frac{1}{7} \mathbf{E}[f(S_{i^*}^*)] = \frac{1}{56} \mathbf{E}[f(S_i^*)] \geq \frac{1}{56n} \mathbf{E}[f(S^*)],$$

where the last inequality is by the definition of i^* and by subadditivity. This concludes the proof of the proposition. \square