Unique Implementation in Team Production: Addressing Coordinated Strategies and Failure of Common Knowledge

Karen Wu*

October 14, 2024

Abstract

A principal sets bonuses for agents to ensure that the induced game will have a unique outcome where all agents work rather than shirk. I explore three notions of outcomes: (1) Nash equilibrium (2) correlated equilibrium (3) rationalizability. It is weakly more expensive for the principal to uniquely implement working under (1) than under (2) which in turn, is more expensive than implementing under (3). I show that when the production technology is anonymous, these weak inequalities in cost hold at equality. When the assumption of anonymity is weakened to a condition I call *aligned marginal contributions*, I show that there can be a strict gap between uniquely implementing under Nash compared to correlated equilibrium, but there remains no gap between implementing under correlated equilibrium compared to rationalizability. Finally, I provide a sufficient condition under which there is a strict gap between achieving unique correlated equilibrium and unique rationalizable strategies.

1 Introduction

Consider a setting where agents choose whether to exert effort on a group project. A principal sets bonuses that will pay out to agents when the project is successful in order to guarantee that all agents work. How low the bonuses can be set while achieving this goal may depend on exactly how the team members interact. Have the agents worked together on projects before? Can agents communicate or otherwise coordinate their actions? Different answers to these questions mean that the principal should use different

*University of Chicago

solution concepts to understand how agents will play the game.

This paper explores three solution concepts in this setting: (1) Nash equilibrium, (2) correlated equilibrium, or (3) rationalizability. Nash and correlated equilibrium both posit that agents are aware of each other's strategies and, taking them as given, choose their own strategy optimally. Nash equilibrium assumes agents take their actions independently while correlated equilibrium allows for rich ways of coordinating agents' actions including communication. Both solution concepts are appropriate for settings where team members have worked on many projects together and are habituated to each others' strategies. Because Nash equilibrium assumes independent mixing, it is more appropriate for settings where agents cannot coordinate their actions, while correlated equilibrium captures environments where team members can freely talk to each other and possibly organize their shirking through communication without taking an explicit stand on the exact protocol through which agents communicate. Rationalizability, on the other hand, does not restrict agents to play best responses to each other's strategies. Agents can play any action that survives an individual, iterative reasoning process that is constrained by common knowledge of the game's payoffs and rationality. Team members that have never worked together and have little awareness of how each other will act likely are probably best described by rationalizalibility rather than solution concepts that presume common knowledge of others' strategies.

In this paper, I study three versions of the principal's problem corresponding to these three solution concepts and identify settings where there are no differences in the values of the problems and settings where there are strict differences. I interpret the existence (or non-existence) of differences as reflecting the increased (or equivalent) difficulty of resolving strategic uncertainty when allowing for correlated actions or the failure of common knowledge of others' strategies.

So far, the literature on unique implementation in the team effort setting has focused on the case where the underlying base game between workers is supermodular. Given an order on agents' actions that the analyst supplies, supermodularity is essentially an assumption on the shape of an agent's difference in payoffs switching to one action over another as the actions of other agents vary. It ensures that the lowest (and highest) rationalizable actions of each agent are mutual best responses[1]. As a result, when the base game is supermodular, bonuses that achieve unique implementation in Nash equilibrium also achieve unique implementation in rationalizable strategies and unique implementation in correlated equilibrium. This paper departs from previous lit-

[1]See Milgrom and Roberts (1990)

erature by setting aside supermodularity and exploring orthogonal assumptions on the base game. Without supermodularity, studying the different solution concepts requires grappling with their mathematical differences. Nash equilibrium and rationalizability correspond to fixed points of particular mappings while correlated equilibria are solutions to linear programs. Generally, comparing the solution concepts seems hard, but this paper is able to make progress in certain important cases.

I first study the three solution concepts when the production technology is *anonymous* so that the probability of project success depends only on the number of agents working rather than their identities. This setting has been regarded as an important special case in previous literature (see Winter (2004) and Halac, Lipnowski, and Rappoport (2021)) but to the best of my knowledge, has not been studied without also assuming supermodularity. The first main result Theorem 1 states that if production technology is anonymous, there is no difference in the costs of implementing in unique Nash, unique correlated equilibrium, or unique rationalizable strategies. The proof proceeds by demonstrating that any bonus profile where working is the unique Nash equilibrium also establishes working as the unique rationalizable strategy. I establish this with an algorithm that takes in any bonus profile and finds a pure strategy Nash equilibrium in the resulting game. The algorithm's outcome gives a bonus that implements working as a unique Nash equilibrium compactly establishes various lower bounds on individuals' bonuses. It turns out that these lower bounds on bonuses are already so high that any bonus profile that implements working in a unique Nash equilibrium must also virtually implement working as the uniquely rationalizable action profile.

Next, to further explore the equality between unique correlated equilibrium and unique rationalizable strategies, I weaken anonymity to a condition I call *aligned marginal contributions*. A production function satisfies aligned marginal contributions if, given some agent is marginally more productive when the set of workers J_1 are working than when J_2 are working, then all agents not in J_1 or J_2 are more marginally productive when J_1 works than when J_2 works. Theorem 2 shows that if the production technology satisfies aligned marginal contributions, then there is no gap between the cost of implementing uniquely in correlated equilibria versus rationalizable strategies; in contrast, there can now be a gap between the cost of implementing in unique Nash equilibrium and implementing in unique correlated equilibrium and rationalizable strategies.

To establish this, I produce a duality based characterization of when a finite game has a unique pure strategy correlated equilibrium (Proposition 2). This characterization leads to a simple necessary condition on bonuses that implement in unique correlated

equilibrium (Lemma 3). Specifically, it can be shown that for any set J of agents that are working, if J is not the set of all agents, there must be an agent $i \notin J$ who would strictly prefer to switch from shirking to working. This condition is enough to demonstrate that all agents must find working uniquely rationalizable when the production technology satisfies aligned marginal contributions.

Intuitively, a production technology satisfies aligned marginal contributions when agents are complementary in a similar way. The last main result of the paper, Proposition 5, explores when agents differ in their productivities in a way that ensures the existence of a strict gap between the cost of implementing in a unique correlated equilibrium compared to implementing in unique rationalizable strategies. To make working an agent's unique rationalizable action, the principal must pay the agent a bonus high enough to motivate working given their worst case conjecture about others' actions. This could require paying an agent a lot if the agent contributes very little to the project in their worst case conjecture. In contrast, in a correlated equilibrium, the principal can use agents' common belief about each others' strategies to instead demonstrate that an agent's worst case conjecture will not occur. For example, if Alice is very unproductive when just Bob is working but not when Bob and Colin are working, then the principal, rather than paying Alice enough to motivate working when just Bob is working, can instead pay Colin enough to work if Bob is working and rely on Alice's awareness that Colin is paid enough to be working if Bob is working to rule out the unfavorable scenario. Implementing in correlated equilibrium, the principal no longer has to pay Alice as much because Alice does not entertain the belief that just Bob is working.

To formalize this idea and build Proposition 5, I provide a characterization of the principal's problem designing to achieve a unique correlated equilibrium that establishes a connection to the principal's problem designing to achieve unique rationalizable strategies. Specifically, the unique rationalizable strategies problem is equivalent to the unique correlated equilibrium problem with an additional constraint. This characterization of the unique correlated equilibrium problem also demonstrates the key additional reasoning agents have available in a correlated equilibrium. While an agent can calculate her rationalizable strategies on her own, achieving a unique correlated equilibrium involves a "group" reasoning process where conjectures about equilibrium play that might result in a member of the group shirking with positive probability are rejected if at least one member of the group would not comply with her recommendation to shirk.

1.1 Related Literature

This paper contributes to the literature on unique implementation by exploring different solution concepts. Early notable papers in this literature are Segal (1999), Segal (2003), and Winter (2004). These papers focus on identifying the minimum cost required for a principal to implement a desired outcome uniquely in Nash equilibrium. Agents' payoff differences are assumed to be increasing or decreasing as more agents take the desired action, so Milgrom and Roberts (1990) implies that unique implementation in Nash equilibrium is equivalent to unique implementation in rationalizable strategies. The main result in these papers is an extreme "divide-and-conquer" logic where agents are endogenously paid different strategic rents in order to build up assurance that a particular action is the unique rationalizable one; if payoff differences are increasing in others' actions, hierarchies of bonuses are extreme with one agent per tier. Sakovics and Steiner (2012) and Halac, Kremer, and Winter (2020) are recent papers that similarly explore unique implementation but in other applications.

Recent literature on unique implementation and the related concern of adversarial equilibrium selection has continued to focus on implementation in rationalizable strategies and added elements of information design. Moriya and Yamashita (2020) allows the principal to design information about an exogenous binary state that affects productivity given commonly known bonuses. Halac, Lipnowski, and Rappoport (2021) allows the principal in the Winter (2004) game to offer bonuses privately; while agents know their own bonuses and know the production technology, they have uncertainty about the bonuses other agents face. In follow up work, Halac, Lipnowski, and Rappoport (2022) relax assumptions about the supermodularity of the base game which possibly drives a wedge between Nash and rationalizability and focus on unique implementation in rationalizable strategies. Inostroza and Pavan (2023) consider design of a public signal in a global games setting where the policy maker evaluates a game according to an adversarially chosen profile of rationalizable actions. Morris, Oyama, and Takahashi (2024) study the problem of an information designer who seeks to implement an outcome as the "smallest" equilibrium in a binary action supermodular base game, where small gets its meaning from the order on actions that defines supermodularity.

This paper is the first (to the best of my knowledge) to study unique implementation in correlated equilibrium. Previous work like Neyman (1997) and Ui (2007) demonstrate that particular classes of games have a unique correlated equilibrium but do not consider how to design to achieve this. Neyman (1997) shows this for games with convex strategy sets and smooth strictly concave potentials. Ui (2007) generalizes Neyman's result to

	project success	no project success
work	$b_i - c_i$	$-c_i$
no work	b_i	0

Table 1: Payoffs for agents

games where fixing other agents' actions, each player's payoff function is smooth and concave in their own action. This paper provides a duality-based characterization of finite games with a unique correlated equilibrium (Lemma 2), applying no assumptions on the shape of payoffs, and then considers how to design the game to satisfy this characterization. Another notable paper studying unique correlated equilibria is Viossat (2008) which shows that the set of finite games with a unique correlated equilibrium is open, and like this paper, uses Myerson (1997) and dual reduction to make progress.

2 Model

There are N agents with typical agent i. I denotes the set of all agents, and J denotes a generic subset of I.

All agents have two actions available to them, working and shirking, denoted w and ϕ respectively. The action profile influences whether the project is successful. Specifically, this relationship is captured by a mapping from who is working to the probability of project success, $P : J \mapsto [0, 1]$. The principal aims to have all her employees work but has limited ability to monitor their actions. She does not observe the true profile of actions that agents take but can observe whether the project is successful or not. As a result, the principal incentivizes agents to work rather than shirk by setting bonuses that pay out only when project success is achieved. These bonuses are given by the vector $b = (b_1, ..., b_N)$. Bonuses are indexed by i so the boss can promise different bonuses to different workers. Each b_i is constrained to be weakly greater than 0.

Given the bonuses the principal sets, agents play a simultaneous, complete information game. Payoffs of this game are given in Table 1. If an agent works, he pays a utility cost of $c_i > 0$, regardless of whether the project is successful and what other agents do.

Say a bonus profile *implements working in Nash equilibrium (NE)* if the bonus profile induces a game with a unique Nash equilibrium where all agents work[2]. A bonus profile b is said to *virtually implement working in NE* if for all $\varepsilon > 0$, $b+\varepsilon := (b_1+\varepsilon, ..., b_N+\varepsilon)$ implements working in NE. I refer to the infimal expected cost of bonuses the principal

[2]I omit the word "uniquely" just to have a shorter phrase.

must pay out in order to implement working in NE as UNE. I analogously define *(virtually) implements working in correlated equilibrium (CE)* and *(virtually) implements working in rationalizable strategies (RS)* and the values of analogous problems UCE and URS, respectively.

This paper compares the values of UNE, UCE and URS. Note that if a profile of bonuses implements working in rationalizable strategies, then these bonuses necessarily implement working in correlated equilibrium and Nash equilibrium. Similarly, if a profile of bonuses implements working in correlated equilibrium, then it must implement working in Nash equilibrium. So, in general, $UNE \leq UCE \leq URS$. I investigate when these inequalities hold at equality and when these inequalities are strict.

Throughout, I require the following to hold:

Assumption 1. *$P(\cdot)$ is **strictly increasing** so that $P(J) > P(J')$ for any $J' \subsetneq J$, $J \subseteq N$.*

This assumption ensures that there exists some finite bonus profile that uniquely implements working in NE/CE/RS.[3]

3 Motivating Examples

This section illustrates how the production technology can result in differences in the cost of implementing in different solution concepts and previews the main results.

Example 1. *Anonymous Technology. $UNE = UCE = URS$.*
Consider a firm with three workers. Each worker has a private cost of working $\frac{1}{2}$. The three workers are interchangeable, so each worker has the same effect on productivity. The probability of project success as a function of who is working is given by

$$\begin{array}{ccccc} & P(1) = \frac{1}{2} & & P(1,2) = \frac{2}{3} & \\ P(\emptyset) = 0 & P(2) = \frac{1}{2} & & P(1,3) = \frac{2}{3} & \qquad P(1,2,3) = 1 \\ & P(3) = \frac{1}{2} & & P(2,3) = \frac{2}{3} & \end{array}$$

The first person to work has a large marginal contribution to success and can get a lot done on their own. A second person working increases the number of hands working on

[3]Specifically, consider the bonus profile where every agent finds working strictly dominant. This bonus profile necessarily uniquely implements in NE, CE and RS and provides an upper bound on the cost of the main problem for each solution concept.

the project, but when the two working team members disagree, they are deadlocked on decisions. This affects their productivity, so the second team member working increases the probability of success only by $\frac{1}{6}$. The third person working adds to the number of hands working and makes ties in decision-making impossible, so the marginal contribution to success of a third person working is larger at $\frac{1}{3}$. This production technology is not beyond imagination but already it is beyond the scope of the existing results about unique implementation since the game that the workers play is not supermodular. It is easy to see that each agent's marginal contribution to the probability of project success is non-monotone in the number of other agents working; it is more laborious but still straightforward to show that no order on agents' actions will make this game supermodular. I have omitted this since it is not interesting.

Let us consider how to implement working as a Nash equilibrium now. Since the marginal contribution to success of a third agent working is 1/3, every agent must be promised a bonus of at least 3/2 in order to motivate them to work when the other two agents are working. If the principal pays all agents this minimum amount, i.e. $b = (3/2, 3/2, 3/2)$, all agents working is a NE, but there are also three NE where exactly one agent works.

One way to eliminate these undesirable equilibria is to go all the way to making working uniquely rationalizable for the agents. To achieve this, the principal must pay agents so that at least one finds working strictly dominant, a second agent finds working strictly dominant given at least one other person is working, and the third agent finds working strictly dominant given at least two other people are working. Given $P(\cdot)$ and $c_i = \frac{1}{2}$ for all agents, this means that the principal must give two agents bonuses strictly greater than 3 and the third agent is given a bonus strictly greater than 3/2. It is clear that the cheapest bonus profile that virtually implements in RS is such that two agents are paid 3 and the third agent is paid 3/2.

In theory, this bonus profile might be unnecessarily expensive if the principal thinks agents play a Nash equilibrium rather than just rationalizable strategies, but it turns out that this bonus profile is also the cheapest that virtually uniquely implements in Nash equilibrium. Consider again $(3/2, 3/2, 3/2)$. Let us see what minimal adjustments the principal has to make in order to get rid of the problematic Nash equilibria where only one agent works. To undermine these NE, the principal can either ensure that working is not a best response to everyone else shirking or that shirking is a best response to exactly one person working. Since each agent's marginal contribution to success is 1/2 when everyone else is shirking and each bonus must be at least 3/2, agents' bonuses are

already high enough that working is a best response to everyone else shirking. Instead, the principal must defeat the undesired Nash equilibria by raising bonuses so that for any strategy profile where exactly one agent is working, there is a non-working agent who is paid too much to find shirking is optimal. As a result, at least two agents must be paid at least 3 so that however we try to pick 1 agent to work and 2 agents to shirk, one of the designated shirkers has a bonus that is too high for them to find shirking optimal. These minimal changes result in a bonus profile where 2 agents have a bonus of 3 and 1 agent has a bonus of 3/2, the same bonus profile that most cheaply virtually implements in RS, so the principal does not have to adjust bonuses further to virtually implement in NE. $\diamond$

Theorem 1 builds on and generalizes the arguments used in Example 1. When $P(\cdot)$ is anonymous, fixing a bonus profile, we can traverse through undesired candidate Nash equilibria in such a way that an action profile fails to be a Nash equilibrium only if a designated shirker would prefer working to shirking rather than a designated worker preferring shirking to working. This establishes lower bounds on individuals' bonuses; it turns out that these lower bounds are already so high that they virtually implement in RS.

Next, I consider a slight modification of the technology in the previous example:

Example 2. *Non-anonymous Technology.* $UNE < UCE < URS$.
All agents have the same cost of working as in Example 1, but agent 2 *has distinct skills from* 1 *and* 3. *The probability of project success as a function of who is working is given by*

$$P(\emptyset) = 0 \qquad \begin{matrix} P(1) = \frac{1}{2} \\ P(2) = \frac{1}{6} \\ P(3) = \frac{1}{2} \end{matrix} \qquad \begin{matrix} P(1,2) = \frac{3}{4} \\ P(1,3) = \frac{2}{3} \\ P(2,3) = \frac{3}{4} \end{matrix} \qquad P(1,2,3) = 1$$

While before all agents were in some sense equal (and so prone to decision-making deadlocks without an odd number involved in the project), agent 2 is now a manager: unproductive on her own but very productive with an underling. In fact, an underling working with the manager avoids the decision deadlock encountered by two underlings working together, so $P(1,2) = P(2,3) > P(1,3)$.

Consider the bonus profile $(2+\varepsilon, 2+\varepsilon, 2+\varepsilon)$ for small $\varepsilon > 0$. There is a unique Nash equilibrium where all agents work at this bonus profile, so UNE is at most 6.

To see this, let σ_i be the probability agent i plays w. Suppose there exists a Nash equilibrium where 1 strictly prefers playing ϕ over w. Then, $\sigma_1 = 0$. Consider the two player game between 2 and 3 where they take as given that 1 never works. 3's marginal contribution to success is at least $\frac{1}{4}$ regardless of whether 2 is working. Given the bonus profile, 3 finds working uniquely rationalizable since 1 is not working. Given that 3 is working and 1 is not working, 2 strictly prefers to work. Since $\sigma_2 = \sigma_3 = 1$, 1 actually strictly prefers to work over not working, so there is no Nash equilibrium where 1 strictly prefers not working over working.

Now, suppose there exists a Nash equilibrium where 1 is indifferent between working and not working. 1's indifference condition is equivalent to the following relationship between other players' strategies:

$$\frac{1}{12}\sigma_2 - \frac{1}{3}\sigma_3 + \frac{1}{2} = \frac{1}{4+2\varepsilon}$$

I can bound σ_3 from below:

$$\sigma_3 \geq \frac{3}{4} + \frac{3}{12}\sigma_2 \geq \frac{3}{4}$$

Now, let's consider 2's payoff difference between working and not working:

$$\frac{1}{12}b_2\sigma_1 + \frac{1}{12}b_2\sigma_3 + \frac{1}{6}b_2 - c_2$$

If σ_1 and σ_3 are such that

$$\frac{1}{12}(\sigma_1 + \sigma_3) + \frac{1}{6} \geq \frac{1}{4}$$

then 2 strictly prefers to work over not working. The above inequality is equivalent to $\sigma_1 + \sigma_3 \geq 1$. So, if $\sigma_3 = 1$, 2 strictly prefers to work, and given that both 2 and 3 are working, 1 will strictly prefer to work. If 1 is indifferent between working and not working, it must be that $\sigma_3 < 1$.

Observe that 1 and 3 are symmetric so that their indexes can be swapped in the expression for 1's payoff difference between working and not working. Identical arguments will establish that for 3 to be indifferent between working and not working, $\sigma_1 \geq \frac{3}{4}$. So $\sigma_1 + \sigma_3 \geq \frac{9}{4}$, and 2 must strictly prefer to work over not working. Given that 2 is working, 1's marginal contribution to success is at least $\frac{1}{4}$, so 1 strictly prefers to work. I conclude that there is no Nash equilibrium where 1 is indifferent between working and

not working.

Finally, suppose there exists a Nash equilibrium where 1 strictly prefers w over ϕ. Then, $\sigma_1 = 1$. 2's marginal contribution to success is guaranteed to be at least $\frac{1}{4}$, so $\sigma_2 = 1$. Finally, since 1 and 2 are working with probability 1, 3 must be working with probability 1. Working is indeed a best response for 1 given others' strategies, so this is the only Nash equilibrium in the game. UNE is at most 6.

In contrast, URS is 7. The cheapest bonus profile that virtually implements in rationalizable strategies is $(2, 3, 2)$ which corresponds to 2 eliminating not work as rationalizable, then 1 eliminating, and finally 3 eliminating. UCE lies strictly between UNE and URS at approximately 6.1. $\diamond$

I will revisit this example in Section 6 of this paper when I present Proposition 5 which provides a sufficient condition for a game to have a gap between implementing in unique correlated equilibrium and unique rationalizable strategies.

4 Anonymous Production

In this section, I restrict attention to games where $P(\cdot)$ satisfies the following definition:

Definition 1. *$P(\cdot)$ is **anonymous** if $P(J) = P(J')$ for all $J, J' \subseteq N$ where $|J| = |J'|$.*

$P(\cdot)$ is anonymous if the probability of achieving project success does not directly depend on the identities of the agents working. As a result, I write $P(k)$ to denote the probability of success induced by any set of k agents working. Anonymity is a particular form of symmetry restriction on the productivities of agents; note that no restriction is put on agents' private costs of working.

Theorem 1. *If $P(\cdot)$ is anonymous, $UNE = UCE = URS$.*

I will show the result by demonstrating that for any profile of bonuses b that implements in unique Nash equilibrium, there is a profile of bonuses $\hat{b}$ where $b_i \geq \hat{b}_i$ for all i, and $\hat{b}$ virtually implements in unique rationalizable strategies. This shows that $URS \leq UNE$. Since $UNE \leq UCE \leq URS$, the inequalities must hold at equality.

Most of the work in the proof is constructing the appropriate $\hat{b}$. To do so, I present an algorithm that takes in any bonus profile, starts with the conjecture that no agents work, and then progressively switches agents from not working to working, terminating

when it reaches a Nash equilibrium. When fed a bonus profile b that uniquely implements working in Nash equilibrium, the algorithm ends with the unique equilibrium. Of course, this is not surprising, but this outcome implies useful lower bounds on agents' bonuses that will produce $\hat{b}$.

To get a sense of how the algorithm's outcome establishes the bounds, consider a strategy profile where agents in J work and agents in $I - J$ do not work. This strategy profile could fail to be a Nash equilibrium for two reasons: either some agent in J does not find it optimal to work ($0 > P(|J|) - P(|J|-1)b_i - c_i$) or some agent in $I - J$ would rather work over not working ($0 < P(|J|+1) - P(|J|)b_i - c_i$). The algorithm rules out the first of these alternatives by switching agents from not working to working in a particular order which guarantees that an agent once designated a worker will always continue to prefer working over shirking at all action profiles the algorithm arrives at.

Now, let's see the algorithm. It takes in any profile of bonuses, b, not necessarily one that implements working in Nash equilibrium, and iteratively builds a set of agents $\hat{J}$ so that there is a pure strategy Nash equilibrium where $\hat{J}$ works and those left out do not:

1. Initialize $k = 1$ and set $J^0 = \emptyset$.

2. Define

$$J^k = \begin{cases} J^{k-1} & \text{if } \forall i \notin J^{k-1}, \frac{c_i}{b_i} > P(k) - P(k-1) \\ J^{k-1} \cup i & \text{otherwise, given } i \in \text{argmin}_{j \notin J^{k-1}} \frac{c_j}{b_j} \end{cases}$$

3. If $J^k = J^{k-1}$, set $J = J^k$ and exit. Otherwise, increment k and return to step 2.

Note that agents are added to $\hat{J}$ starting from the smallest c_i/b_i and progressing to larger c_i/b_i[4]

Agents in $\hat{J}$ working while agents not in $\hat{J}$ shirking is a pure strategy Nash equilibrium.[5] The stopping condition of the algorithm ensures that all agents not in $\hat{J}$ weakly prefer shirking to working. To see that all agents added to $\hat{J}$ prefer working to shirking, let

[4] $\frac{c_i}{b_i}$ is an intuitive measure of how motivated agents are to work over shirk as it is the minimum contribution to success such that i weakly prefers w to ϕ. Agents with a lower c_i/b_i can be interpreted as being more highly motivated to work since they will work over shirk even when their marginal contribution to success is relatively low.

[5] This algorithm bears some resemblance to a standard algorithm that constructs a pure strategy Nash equilibrium in a generalized ordinal potential game. See Monderer and Shapley (1996). The team production game is a weighted potential game.

$|\hat{J}| = k$, and index agents in the order that they are added to $\hat{J}$, if they are added. Agents that are not added to $\hat{J}$ are indexed something strictly greater than k. Consider k, the last agent added to $\hat{J}$. Because k is an agent with the minimum c_i/b_i among agents in $I - J^{k-1}$, it must be that $\frac{c_k}{b_k} \leq P(k) - P(k-1) \iff \big(P(k) - P(k-1)\big)b_k - c_k \geq 0$. Any agent $i < k$ was added to $\hat{J}$ before k and so $\frac{c_i}{b_i} \leq \frac{c_k}{b_k}$. This implies that

$$\big(P(k) - P(k-1)\big)b_i - c_i \geq 0 \quad \forall i < k$$

The anonymity of the production function and the special order in which agents are added are used in this last observation. As the algorithm adds agents to the conjectured worker set, agents that have already been added see their marginal contribution to project success change. In principle, this might jeopardize their preference for working over shirking, but since the production function is anonymous, the marginal contribution to success of the last agent added is the same as that of agents added earlier. Because the algorithm adds agents starting from lowest c_i/b_i to larger c_i/b_i, adding the last agent to the worker set implies agents added earlier continue to prefer work to shirk.

I have shown the following lemma:

Lemma 1. *If b induces a unique NE where all agents work, then $\hat{J} = I$.*

Now, suppose b uniquely implements in Nash equilibrium. Given lemma 1, I can index all agents in the order they are eventually added to $\hat{J}$. An important implication of lemma 1 is that

$$\frac{c_i}{b_i} \leq P(k) - P(k-1) \qquad \forall i \leq k, \quad \forall 1 \leq k \leq N$$

In words, the lowest $k+1$ c_i/b_i are at most $P(k+1) - P(k)$ for all $k = 0, ..., N-1$. Depending on the shape of P, not all these upper bounds are necessarily relevant; some might imply others hold. The following process identifies the most relevant upper bounds:

1. Start with $n = 1$. Define $k_n := \text{argmin}_{1 \leq k \leq N} P(k) - P(k-1)$.
2. Increment n by 1.
3. If $k_{n-1} \neq N$, then define $k_n = \text{argmin}_{k_{n-1} < k \leq N} P(k) - P(k-1)$.
4. If $k_{n-1} = N$, then exit.
5. Return to step 3.

To understand what this process does, consider k_1. k_1 is the size of the worker set where the marginal contribution to success of those who are working is smallest, so for $i \leq k_1$, $c_i/b_i \leq P(k_1) - P(k_1 - 1)$ implies $c_i/b_i \leq P(k) - P(k-1)$ for all $k \leq k_1$. The argument that k_n is the binding upper bound for $i \leq k_n$ is similar. Denote the maximum n reached in the iterative construction of k_n as $\bar{n}$. When the marginal contribution to success is increasing as more agents work, then $k_n = n$ with $\bar{n} = N$. When the marginal contribution to success is decreasing as more agents work, then $\bar{n} = 1$ with $k_1 = N$. I can summarize the relevant upper bounds as:

$$\bar{s}(i) = \begin{cases} P(k_1+1) - P(k_1) & \text{if } i \leq k_1 + 1 \\ P(k_2+1) - P(k_2) & \text{if } k_1 + 1 < i \leq k_2 + 1 \\ ... & \\ P(k_n+1) - P(k_n) & \text{if } k_{n-1} + 1 < i \leq k_n + 1 \\ ... & \\ P(k_{\bar{n}}+1) - P(k_{\bar{n}}) & \text{if } k_{\bar{n}-1} + 1 < i \leq k_{\bar{n}} + 1 \end{cases}$$

Given the upper bounds on c_i/b_i, I can identify our desired lower bound profile of bonuses, $\hat{b}$:

$$\begin{aligned} & \frac{c_i}{b_i} \leq \bar{s}(i) \\ \iff \hat{b}_i := & \frac{c_i}{\bar{s}(i)} \leq b_i \end{aligned}$$

The final step to proving Theorem 1 is observing that $\hat{b}$ virtually implements in URS. To show this, consider a bonus profile where each agent i's bonus is a $\varepsilon > 0$ above $\hat{b}_i$. Denote this bonus profile by $\hat{b}^\varepsilon$ where $\hat{b}_i^\varepsilon = \hat{b}_i + \varepsilon$ for $\varepsilon > 0$. Using the definition of k_1, I can observe that for all $i \leq k_1$,

$$\big(P(k) - P(k-1)\big)\hat{b}_i^\varepsilon - c_i \geq \big(P(k_1+1) - P(k_1)\big)\hat{b}_i^\varepsilon - c_i > 0 \quad \forall 1 \leq k \leq N$$

so agents $1, ..., k_1$ find working strictly dominant. Now, suppose that agents $1, ..., k_n$ are working. Observe that for all $k_n < i \leq k_{n+1}$,

$$\big(P(k) - P(k-1)\big)\hat{b}_i^\varepsilon - c_i \geq \big(P(k_n) - P(k_n - 1)\big)\hat{b}_i^\varepsilon - c_i > 0 \quad \forall k_n \leq k \leq N$$

so agents $k_n + 1$, ..., k_{n+1} find working strictly dominant given that $1, ..., k_n$ are working. This completes the proof. □

The lower bound profile of bonuses $\hat{b}$ has a particular structure:

Definition 2. *A bonus profile b is **tiered according to strategic bottlenecks** (TASB) if b has k_1 agents such that $\frac{c_i}{b_i} = P(k_1) - P(k_1 - 1)$, $k_2 - k_1$ agents such that $\frac{c_i}{b_i} = P(k_2) - P(k_2 - 1)$, ..., and $k_{\bar{n}} - k_{\bar{n}-1}$ agents such that $\frac{c_i}{b_i} = P(k_{\bar{n}}) - P(k_{\bar{n}} - 1)$.*

Recall the final step of proving Theorem 1. That $\hat{b}$ is TASB is enough to make the argument $\hat{b}$ virtually implements in URS. The next corollary summarizes this:

Corollary 1. *Suppose b is tiered according to strategic bottlenecks. Then, b virtually implements in URS.*

So, the value of the UNE = UCE = URS problem in the anonymous $P(\cdot)$ case is the value of the cheapest TASB b. This b can be easily identified:

Proposition 1. *When $P(\cdot)$ is anonymous, the value of the UNE = UCE = URS problem is given by $\hat{b}^*$ where*

$$\hat{b}^* = \begin{cases} \frac{c_i}{P(k_1+1)-P(k_1)} & \text{if} \quad 1 \leq i \leq k_1 \\ \dots & \\ \frac{c_i}{P(k_n+1)-P(k_n)} & \text{if} \quad k_{n-1} < i \leq k_n \\ \dots & \\ \frac{c_i}{P(k_{\bar{n}}+1)-P(k_{\bar{n}})} & \text{if} \quad k_{\bar{n}-1} < i \leq k_{\bar{n}} \end{cases}$$

and agents are indexed so that $c_1 \leq c_2 ... \leq c_N$.

Finding the minimum cost TASB b is essentially a two-sided matching problem. The strategic bottlenecks define "strategic roles": k_1 agents must be paid enough to motivate working when $k_1 - 1$ other agents are working, an additional $k_2 - k_1$ agents must be paid enough to motivate working when $k_2 - 1$ other agents are working, and so on. The role of being motivated to work when $k_n - 1$ other agents are working can be thought of as having an associated type $1/(P(k_1) - P(k_1 - 1))$. Each agent i can be associated with his cost of working c_i. When agent i is matched to the strategic role of working when $k_n - 1$ other agents work, the principal will pay them a bonus of $c_i/(P(k_n) - P(k_n - 1))$. The mapping from agent and role types to required bonus is supermodular, so as shown in Becker (1973), to minimize the total cost of bonuses, lower cost agents should be matched to strategic roles associated with smaller marginal contributions to success.

5 Aligned Marginal Contributions

5.1 Preliminaries

In this section, I explore what drives the equivalence between UCE and URS in the case of anonymous production. I focus on the following restriction on $P(\cdot)$:

Definition 3. *$P(\cdot)$ satisfies **aligned marginal contributions** if there exists a ranking $\succeq$ over worker sets $J \subsetneq N$ where $J \succeq J' \iff$ for all $i \notin J, J'$, $P(J \cup i) - P(J) \geq P(J' \cup i) - P(J')$.*

Remark 1. *If $P(\cdot)$ is anonymous, $P(\cdot)$ satisfies aligned marginal contributions.*

Given $P(\cdot)$ is anonymous, for any worker set $J \subsetneq N$, every $i \notin J$ has the same marginal contribution to success:

$$P(J \cup i) - P(J) = \Delta_J \quad \forall i \notin J \tag{Sym}$$

In fact, $P(J \cup i) - P(J)$ does not directly depend on J but $|J|$ since $P(\cdot)$ is anonymous, but the above observation is enough to demonstrate that $P(\cdot)$ satisfies aligned marginal contributions. The ranking on $J \subsetneq N$ required by the definition can be produced from the natural order on Δ_J so that $J \succeq J'$ if and only if $\Delta_J \geq \Delta_{J'}$. □

Remark 2. *If $N = 2$, then $P(\cdot)$ satisfies aligned marginal contributions.*

Note that a two player game has 3 worker sets: (1) no one working (2) exactly agent 1 working (3) exactly agent 2 working. Intuitively, producing the ranking required for aligned marginal contributions is easy because there does not exist two different worker sets J and J' where both 1 and 2 are not in J or J'. 1 and 2 cannot "disagree" on which worker sets they are relatively productive with.

To show the claim, let's partition the set of all $P(\cdot)$ functions into 4 cases. Table 2 verifies that it is possible to produce the desired ranking over worker sets for each case.

Note that this remark also demonstrates that $P(\cdot)$ satisfying aligned marginal contributions does not imply that $P(\cdot)$ is anonymous or even symmetric in the sense of satisfying Sym for every $J \subsetneq N$. □

Aligned marginal contributions is a sufficient condition under which UCE and URS coincide:

Theorem 2. *If $P(\cdot)$ satisfies aligned marginal contributions, $UCE = URS$.*

Case	Verifying Ranking
$P(1,2) - P(2) \geq P(1) - P(\emptyset)$ $P(1,2) - P(1) \geq P(2) - P(\emptyset)$	$2 \succeq 1 \succeq \emptyset$
$P(1,2) - P(2) \geq P(1) - P(\emptyset)$ $P(1,2) - P(1) < P(2) - P(\emptyset)$	$2 \succeq \emptyset \succeq 1$
$P(1,2) - P(2) < P(1) - P(\emptyset)$ $P(1,2) - P(1) \geq P(2) - P(\emptyset)$	$1 \succeq \emptyset \succeq 2$
$P(1,2) - P(2) < P(1) - P(\emptyset)$ $P(1,2) - P(1) \geq P(2) - P(\emptyset)$	$1 \succeq \emptyset \succeq 2$
$P(1,2) - P(2) < P(1) - P(\emptyset)$ $P(1,2) - P(1) < P(2) - P(\emptyset)$	$\emptyset \succeq 2 \succeq 1$

Table 2: Every two player game satisfies aligned marginal contributions.

5.2 On UCE in Finite Games

In this subsection, I study when a finite game has a unique correlated equilibrium that puts probability one on a particular profile of actions. I then apply this to the team production game.

Consider a finite game with N agents and typical agent i. S_i is the finite set of actions available to each i with typical element $s_i \in S_i$. $u_i(s_i, s_{-i})$ is the utility i gets from this action profile. $\mu \in \Delta S$ is a distribution over action profiles.

Fix a profile of actions $\bar{s} \in S$. Distribution μ such that $\mu(\bar{s}) = 1$ is a correlated equilibrium of the game if the following constraints are satisfied:

$$\sum_{s} \mu(s) = 1$$
$$\sum_{s_{-i}} \mu(s) \Big(u_i(s_i, s_{-i}) - u_i(\tilde{s}_i, s_{-i}) \Big) \geq 0 \quad \forall s_i, \tilde{s}_i \in S_i, \forall i$$

Furthermore, μ is the *unique* correlated equilibrium if there is no correlated equilibrium that puts positive probability on any action profile $s \neq \bar{s}$. In other words, the following linear system is infeasible:

$$\sum_{s_{-i}} \mu(s) \Big(u_i(s_i, s_{-i}) - u_i(\tilde{s}_i, s_{-i}) \Big) \geq 0 \quad \forall s_i, \tilde{s}_i \in S_i, \forall i$$
$$\mu(s) \geq 0 \quad \forall s \in S$$
$$\sum_{s \neq \bar{s}} \mu(s) > 0$$

Using Motzkin's Transposition Theorem (see Tucker (1957) Corollary 2A), the infeasibility of the above system is equivalent to the feasibility of the following linear system:

$$\sum_{i,\tilde{s}_i} \alpha_i(\tilde{s}_i|s_i)\Big(u_i(s_i, s_{-i}) - u_i(\tilde{s}_i, s_{-i})\Big) + \gamma(s) + \beta = 0 \quad \forall s \neq \bar{s}$$

$$\sum_{i,\tilde{s}_i} \alpha_i(\tilde{s}_i|\bar{s}_i)\Big(u_i(\bar{s}_i, \bar{s}_{-i}) - u_i(\tilde{s}_i, \bar{s}_{-i})\Big) + \gamma(\bar{s}) = 0$$

$$\alpha_i(\tilde{s}_i|s_i) \geq 0 \quad \forall i, \forall \tilde{s}_i, s_i \in S_i$$

$$\gamma(s) \geq 0 \quad \forall s \in S$$

$$\beta > 0$$

Rearranging this linear system, normalizing α_i, and dispensing with γ, we can arrive at:

Lemma 2. *$\bar{\mu} \in \Delta S$ such that $\bar{\mu}(\bar{s}) = 1$ is the unique CE if and only if $\exists\ \alpha \geq 0$ such that:*

$$1 = \sum_{\tilde{s}_i} \alpha_i(\tilde{s}_i|s_i) \quad \forall i, \forall s_i$$

$$\sum_i \sum_{\tilde{s}_i} \alpha_i(\tilde{s}_i|s_i)\big(u_i(\tilde{s}_i, s_{-i}) - u_i(s)\big) > 0 \quad \forall s \neq \bar{s}$$

$$\sum_i \sum_{\tilde{s}_i} \alpha_i(\tilde{s}_i|\bar{s}_i)\big(u_i(\tilde{s}_i, \bar{s}_{-i}) - u_i(\bar{s})\big) = 0$$

Using other results in the literature, it is possible to learn even more about the α that satisfies the linear system in lemma 2. Myerson (1997) tells us we can build an auxiliary game from the original game, fixing the multipliers on obedience constraints associated with a given correlated equilibrium. To construct this game, note that since $\sum_{\tilde{s}_i} \alpha_i(\tilde{s}_i|s_i) = 1$, the multipliers on i's obedience constraints α_i define a Markov chain where S_i is the state space. $\alpha_i(\tilde{s}_i|s_i)$ can be interpreted as the probability the process transitions to $\tilde{s}_i$ from state s_i. In the auxiliary game termed the "dual reduction," each player i has finite pure strategies that correspond to certain stationary distributions of α_i.[6] Mixed strategies in the dual reduction are convex combinations of stationary distributions of α_i and so also stationary with respect to α_i. Payoffs are defined from the original game payoffs in the natural way. Since the dual reduction is a finite game, it has a correlated equilibrium. Theorem 1 of Myerson (1997) establishes that a correlated equilibrium of the dual reduction is also a correlated equilibrium of the original

[6]It is clear that the dual reduction is constructed using the dual variables associated with the obedience constraints of a correlated equilibrium. It is a "reduction" because players have weakly fewer pure strategies in the dual reduction than in the original game.

game.

When a game has a unique pure strategy correlated equilibrium $\bar{\mu}$ with multipliers $\alpha \geq 0$ on the obedience constraints, the associated dual reduction also has a unique correlated equilibrium. Then, for each i, playing $\bar{s}_i$ with probability 1 must be a stationary distribution for α_i, so $\alpha_i(\bar{s}_i|\bar{s}_i) = 1$. I record the sum of these observations in the following proposition:

Proposition 2. *$\bar{\mu} \in \Delta S$ such that $\bar{\mu}(\bar{s}) = 1$ is the unique CE if and only if $\exists\ \alpha \geq 0$ such that:*

$$
\begin{aligned}
1 &= \sum_{\tilde{s}_i} \alpha_i(\tilde{s}_i|s_i) \quad \forall i, \forall s_i \\
\sum_i \sum_{\tilde{s}_i} &\alpha_i(\tilde{s}_i|s_i)\big(u_i(\tilde{s}_i, s_{-i}) - u_i(s)\big) > 0 \quad \forall s \neq \bar{s} \\
&\alpha_i(\bar{s}_i|\bar{s}_i) = 1 \quad \forall i
\end{aligned}
$$

5.2.1 UCE in the Team Production Game

Now, I apply Proposition 2 to the team production game, taking $\bar{s}$ to be the action profile where all agents work:

$$
\begin{aligned}
&\inf_{(b_i)_i, \alpha \geq 0} \sum_i P(N) b_i && \text{(UCE)} \\
&\text{s.t. } 1 = \alpha_i(w|\phi) + \alpha_i(\phi|\phi) \quad \forall i && (1) \\
&\qquad \sum_{i \notin J} \alpha_i(w|\phi)\big(P(J \cup i) b_i - c_i - P(J) b_i\big) > 0 \quad \forall J \subsetneq I && (2)
\end{aligned}
$$

Because specifying $\alpha_i(w|\phi)$ pins down $\alpha_i(\phi|\phi)$, I only consider how to choose $\alpha_i(w|\phi)$ and write $\alpha_i(w|\phi)$ as α_i from here on.

5.3 Proof of Theorem 2

To show Theorem 2, I will show that if $P(\cdot)$ satisfies aligned marginal contributions, then any bonus profile that uniquely implements in correlated equilibrium also uniquely implements in rationalizable strategies. Since $UCE \leq URS$, this is enough to show that $UCE = URS$.

I start with an implication of UCE:

Lemma 3. *Suppose b is UCE. For all $J \subsetneq I$, there exists some $i \notin J$ for whom*

$$P(J \cup i)b_i - c_i - P(J)b_i > 0$$

If such an $i \notin J$ did not exist, it would not be possible to produce $\alpha \geq 0$ and satisfy the constraints 2.

Let b implement in unique correlated equilibrium. Aligned marginal contributions implies that there is a complete ranking over all worker sets $J \subsetneq I$. So, there is a bottom ranked worker set J^1 such that for all $J \neq J^1$, $J \succeq J^1$. By lemma 3, there must be an agent $i^1 \notin J^1$ such that $P(J^1 \cup i^1)b_{i^1} - c_{i^1} - P(J^1)b_{i^1} > 0$. By the definition of aligned marginal contributions, for all other worker sets $J \neq J^1$ s.t. $i^1 \notin J$, i^1 has a greater marginal contribution to success when J is working than when J^1 are working, i.e. $P(J \cup i^1) - P(J) \geq P(J^1 \cup i^1) - P(J^1)$. Since $P(J^1 \cup i^1)b_{i^1} - c_{i^1} - P(J^1)b_{i^1} > 0$, working is i^1's uniquely rationalizable action.

Now, suppose agents $i^1, ..., i^k$ have eliminated not working as rationalizable. Let's restrict our attention to the ranking over worker sets that contain $i^1, ..., i^k$. Again, there is a J^{k+1} that is bottom ranked among these sets. By lemma 3, there must be an agent $i^{k+1} \notin J^{k+1}$ such that $P(J^{k+1} \cup i^{k+1})b_{i^{k+1}} - c_{i^{k+1}} - P(J^{k+1})b_{i^{k+1}} > 0$. Since agent i^{k+1}'s smallest marginal contribution to success given $i^1, ..., i^k$ are working is at worker set J^{k+1}, i^{k+1} eliminates not working as rationalizable given that $i^1, ..., i^k$ are working.

Repeatedly applying this argument produces a relabeling of all agents as $i^1, ..., i^N$ where the superscript denotes the order in which they eliminated not working as rationalizable. In other words, b implements in URS as well. □

While aligned marginal contributions implies $UCE = URS$, the next example demonstrates that it does not imply $UNE = UCE$.

Example 3. ***Satisfies aligned marginal contributions but*** $UNE < UCE = URS$

Consider a three person game where $c_1 = 2$, $c_2 = c_3 = 1$ and the probability of project

success is given by:

$$
\begin{array}{llll}
 & P(1) = \frac{1}{4} & P(1,2) = \frac{7}{16} & \\
P(\emptyset) = 0 & P(2) = \frac{3}{8} & P(1,3) = \frac{7}{16} & P(1,2,3) = 1 \\
 & P(3) = \frac{2}{8} & P(2,3) = \frac{5}{8} &
\end{array}
$$

The following ranking verifies that the game satisfies aligned marginal contributions:

$$\{2,3\} \succeq \{1,2\} \succeq \{1,3\} \succeq \emptyset \succeq \{3\} \succeq \{2\} \succeq \{1\}$$

Theorem 2 tells us that $UCE = URS$. To calculate $UCE = URS$, it is easier to focus on URS. The cheapest profile of bonuses that virtually implements in rationalizability is $(\frac{16}{3}, \frac{8}{3}, \frac{16}{3})$ and corresponds to 3 eliminating shirking as rationalizable, then 2 eliminating, and then 1 eliminating. The value of $UCE = URS$ is thus $13\frac{1}{3}$. On the other hand, UNE is weakly less than 12. To show this, I consider the bonus profile $(\frac{16}{3} + \varepsilon, \frac{8}{3} + \varepsilon, 4 + \varepsilon)$ where $\varepsilon > 0$. I will show that for small enough ε, there is a unique Nash equilibrium where all agents work with probability 1.

To see this, first note that whether an agent prefers to work or shirk depends only on their marginal contribution to success. Given these values for the bonuses, let us calculate the minimum marginal contribution to success at which each agent will weakly prefer working to shirking. For 1, the minimal marginal contribution is $\frac{6}{16+\varepsilon}$. For 2, it is $\frac{3}{8+\varepsilon}$, and for 3, it is $\frac{1}{4+\varepsilon}$.

Let σ_i denote the probability that agent i plays work in a Nash equilibrium. Observe that if $\sigma_1 = 0$, then 2's marginal contribution to success is at least $\frac{3}{8}$, so σ_2 must be 1. Similarly, 3's marginal contribution to success is at least $\frac{1}{4}$, so σ_3 must be 1. If 2 and 3 are working with probability 1, then 1's marginal contribution to success is $\frac{3}{8}$, and he must also be working with probability 1. Contradiction. There is no Nash equilibrium where 1 plays work with 0 probability.

Since 1 must be playing work with positive probability in any Nash equilibrium, this implies that 2 and 3 must be playing work with relatively high probability. Specifically, it must be that $\sigma_2\sigma_3 > \frac{1}{2}$. If not, 1's marginal contribution to project success can be

bounded from above:

$$\begin{aligned}
&\underbrace{(1-\sigma_2)(1-\sigma_3)\frac{1}{4}+(1-\sigma_2)\sigma_3\frac{3}{16}+\sigma_2(1-\sigma_2)\frac{1}{16}}_{\leq(1-\sigma_2\sigma_3)\frac{1}{4}}+\sigma_2\sigma_3\frac{3}{8}\\
&\quad\leq(1-\sigma_2\sigma_3)\frac{1}{4}+\sigma_2\sigma_3\frac{3}{8}\\
&\quad\leq\frac{5}{16}<\frac{6}{16+3\varepsilon}
\end{aligned}$$

The last strict inequality holds for small enough ε. Given this little lemma, it is clear that both 2 and 3 must be playing work with positive probability in any Nash equilibrium.

Now, suppose 2 mixes between working and not working in a Nash equilibrium. 2's indifference between the two actions implies that his marginal contribution to success is

$$\begin{aligned}
(1-\sigma_1)(1-\sigma_3)\frac{6}{16}+\sigma_1(1-\sigma_3)\frac{3}{16}+(1-\sigma_1)\sigma_3\frac{6}{16}+\sigma_1\sigma_2\frac{9}{16}\\
=\frac{3}{8}-\frac{3}{16}\sigma_1+\frac{6}{16}\sigma_1\sigma_3=\frac{3}{8+3\varepsilon}
\end{aligned}$$

I can then bound the probability that 3 plays work from above:

$$\sigma_3\leq\frac{1}{2}$$

This implies that $\sigma_2\sigma_3\leq\frac{1}{2}$ if 2 mixes between working and not working, so it cannot be that 2 is mixing. Similarly, if 3 is mixing between working and not working, his indifference implies:

$$\begin{aligned}
(1-\sigma_2)(1-\sigma_1)\frac{4}{16}+\sigma_1(1-\sigma_2)\frac{3}{16}+\sigma_2(1-\sigma_1)\frac{4}{16}+\sigma_1\sigma_2\frac{9}{16}\\
=\frac{4}{16}-\frac{1}{16}\sigma_1+\frac{6}{16}\sigma_1\sigma_2=\frac{1}{4+\varepsilon}
\end{aligned}$$

and it is again possible to bound the probability that 2 plays work:

$$\sigma_2\leq\frac{1}{6}$$

Again, this implies 1 strictly prefers to shirk over work, so it cannot be that 3 mixes between working and not working either. I conclude that $\sigma_2=\sigma_3=1$. Given 2 and 3 are playing work with probability 1, 1's best response is to work with probability 1. This is the unique Nash equilibrium when ε is sufficiently small, so $UNE\leq 12$ and there

is a strict gap between the cost of implementing in unique Nash compared to the cost of implementing in unique correlated equilibrium and unique rationalizable strategies. $\diamond$

Aligned marginal contributions is not necessary for UCE to equal URS but makes stating the result and proving it easy by asserting the existence of a complete ranking over worker sets. With a complete ranking, given that agents $i^1, ..., i^k$ find working uniquely rationalizable, there necessarily exists a worker set where all agents left out of the set have minimal marginal contributions to productivity. The next example demonstrates that it is also possible to satisfy this requirement without satisfying aligned marginal contributions:

Example 4. ***Sufficient Alignment Implies*** $UCE = URS$ ***but*** $UNE < UCE$
Consider a three person game where $c_1 = c_2 = c_3 = \frac{1}{2}$ *where the probability of project success is given by the following:*

$$\begin{array}{ccccc} & & P(1) = \frac{1}{3} & P(1,2) = \frac{2}{3} & \\ P(\emptyset) = 0 & P(2) = \frac{7}{24} & P(1,3) = \frac{1}{2} & P(1,2,3) = 1 \\ & & P(3) = \frac{7}{24} & P(2,3) = \frac{1}{2} & \end{array}$$

Equip agents with bonuses $(\frac{3}{2} + \varepsilon, \frac{3}{2} + \varepsilon, \frac{3}{2} + \varepsilon)$ where $\varepsilon > 0$. When ε is very small, agents are paid just enough for each to strictly prefer working over not working when their marginal contribution to success is $\frac{1}{3}$. This game has a unique Nash equilibrium where all agents work with probability 1.

To establish this, consider whether agent 3 prefers working to not working in a Nash equilibrium of the game. Suppose 3 strictly prefers to shirk over working in equilibrium. Since 3 is not working, 1 strictly prefers working to shirking since his marginal contribution to success is guaranteed to be at least $P(1) - P(\emptyset) = 1/3$. Given that 1 is working with probability 1, 2 strictly prefers to work over not working since his marginal contribution to success is then 1/3. Since 1 and 2 are both working, 3 strictly prefers to work since his marginal contribution to success is then 1/3, so there is no Nash equilibrium where 3 strictly prefers shirking to working.

Consider what happens when 3 weakly prefers working to shirking in a Nash equilibrium. Then, it cannot be that 1 is shirking with probability 1 since this guarantees that 3's marginal contribution to success is at most $\frac{7}{24} < \frac{1}{3}$, and 3 would strictly prefer to shirk

over working. Let's see that 1 cannot be mixing between working and not working in equilibrium. If 1 is mixing, 1 is indifferent between working and not working. Using this indifference condition, it is possible to create an upper bound on σ_2:

$$\sigma_2 \leq \frac{3\sigma_3}{1+6\sigma_3} \leq \frac{3}{7}$$

where the second inequality follows by observing that the RHS of the first inequality is strictly increasing in σ_3 and evaluating the RHS at $\sigma_3 = 1$. We can now bound 3's payoff difference from switching to work from shirk:

$$(1-\sigma_1)\underbrace{\left[(1-\sigma_2)\frac{7}{24}b_3 + \sigma_2\frac{5}{24}b_3\right]}_{\leq \frac{7}{24}b_3} + \sigma_1\underbrace{\left[(1-\sigma_2)\frac{1}{6}b_3 + \sigma_2\frac{8}{24}b_3\right]}_{\leq \frac{5}{21}b_3} - c_3 \leq \frac{7}{24}b_3 - c_3 < 0$$

So, in a Nash equilibrium, it cannot be that 3 prefers to work over shirking while 1 is mixing between working and shirking.

Finally, suppose that 3 prefers to work over shirking in a Nash equilibrium and 1 is playing work with probability 1. Then, 2's marginal contribution to success is at least 1/3, so 2 is also working with probability 1. 3 is also working with probability 1. This is the only Nash equilibrium.

The above argument establishes that $UNE \leq 4.5$. URS is approximately 5.23, given by the cost of $(\frac{12}{9}, \frac{12}{5}, \frac{3}{2})$. This bonus profile is associated with first 2 eliminating not working as rationalizable, followed by 1 eliminating, and then 3 eliminating. In this game, UCE actually coincides with URS. To see this, notice that both 1 and 2's marginal contributions to project success are smallest when only 3 is working. In other words, both 1 and 2 are least marginally productive when just 3 is working. By Lemma 3, if b implements in UCE, then 1 or 2 must be motivated to work when only 3 is working. In other words, 1 or 2 must find working strictly dominant.

Let's say 1 finds working dominant. 2 and 3 are now playing a two-player game since 1 must be working. By Remark 2, $P(\cdot)$, restricted to worker sets that include 1, satisfies aligned marginal contributions. So, if b_1 makes 1 find working strictly dominant, (b_2, b_3) induces 2 and 3 to find working uniquely rationalizable. A similar argument can be made if b_2 makes working strictly dominant for 2. As a result, $UCE = URS$ in this game. $\diamond$

The example above generalizes the logic in the proof of Theorem 2. $P(\cdot)$ fails aligned

marginal contributions: agent 3 is less productive when 1 is working than when no one is working but agent 2 is more productive when 1 is working than when no one is working. Additionally, agent 3 is less productive when 2 is working than when no one is working, but agent 1 is more productive when 2 is working than when no one is working. Once we demonstrate that either 1 or 2 must find working strictly dominant, that 3 is relatively productive at $\emptyset$ while 1 and 2 both are relatively unproductive at $\emptyset$ does not matter. No one working cannot occur in equilibrium anyways.

6 Sufficient Condition for Gap between UCE and URS

Given Theorem 2, a strict gap between the UCE and URS problems implies agents differ in which worker sets they are relatively more productive with. This section further explores what drives a strict gap by providing a different characterization of the UCE problem which can be used to more easily compare the UCE and URS problems. The main result of the section, Proposition 5, uses this characterization to give a sufficient condition on the base game for the existence of a strict gap between the UCE and URS problems.

Let $\mathcal{P}$ denote a partition of the agents $\{I_1, ..., I_K\}$. The cells of the partition are ordered; this is reflected in their indexing. Fix a vector $(\alpha_i) \gg 0$. I define a cell level bonus design problem:

$$V_{\mathcal{P},k}(\alpha) := \min_{(b_i)_{i\in I_k}} \quad \sum_{i\in K_k} b_i \tag{3}$$

$$\text{s.t.} \quad \sum_{i\in I_k - J} \alpha_i \Big((P(J\cup i) - P(J))b_i - c_i \Big) \geq 0 \qquad \forall J \subsetneq N, \cup_{l=1}^{k-1} I_l \subseteq J \tag{4}$$

This problem optimizes the choice of bonuses for just the agents in the kth cell, fixing $(\alpha_i)_{i\in I_k}$ as the multipliers for agents in I_k. The cell level problem has constraints that differ from those in 2 in two ways. First, constraints correspond only to sets J that contain all agents in cells that precede I_k. It is as if agents in cells that precede I_k are presumed to be working from the perspective of agents in cell I_k. Second, each J constraint only contains incentive terms from agents in the kth cell. The incentives of agents in cells that succeed I_k are irrelevant for the kth cell problem.

Now, suppose a bonus profile b uniquely implements working in rationalizable strategies.

b implies an order in which agents eliminate not working as rationalizable. Denote it by $i^1, ..., i^N$ where i^1 eliminates first. This order on agents naturally maps to an ordered partition $\mathcal{P}$ of agents where each cell is a singleton and $I_k = \{i^k\}$. Taking as given that $i^1, ..., i^{k-1}$ work, i^k finds working strictly dominant, so for any $\alpha_{i^k} > 0$, 4 holds strictly and b_{i^k} is feasible for the I_k cell problem for any $\alpha \gg 0$. Given this observation, it is clear that

$$\min_{\mathcal{P}\in\mathcal{O},\alpha=(1,...,1)} \sum_{k=1} V_{\mathcal{P},k}(\alpha) \leq \sum_{i\in I} b_i$$

where $\mathcal{O}$ is the set of ordered partitions such that each cell of the partition is a singleton. From here, it is straightforward to show that URS is given by the optimal choice of ordered singleton partition:

Proposition 3.

$$URS = \min_{\mathcal{P}\in\mathcal{O},\alpha=(1,...,1)} \sum_{k=1}^{N} V_{\mathcal{P},k}(\alpha) \tag{URS-P}$$

where $\mathcal{O}$ is the set of ordered partitions where each cell of the partition is a singleton.

Like URS, UCE is also given by the sum of cell problems induced by a choice of ordered partition and $\alpha \gg 0$. Unlike URS, UCE does not restrict attention to ordered partitions with singleton cells:

Proposition 4.

$$UCE = \min_{\mathcal{P},\alpha\gg 0} \sum_{k} V_{\mathcal{P},k}(\alpha) \tag{UCE-P}$$

To understand the non-singleton cells, fix a bonus profile b that virtually implements in a unique correlated equilibrium. The proof of Proposition 4 shows that b can be mapped to an ordered partition and $\alpha \gg 0$ such that 4 hold strictly for each cell problem. Consider a single cell of the partition, I_k. Using Motzkin's transposition theorem, the existence of $\alpha \gg 0$ that satisfy the inequalities 4 is equivalent to the non-existence of a distribution β over worker sets $J \subsetneq I$ s.t. $\cup_{l=1}^{k-1} I_l \subset J$ and the following inequalities hold:

$$\sum_{J \text{ s.t. } i\notin J, \cup_{l=1}^{k-1} I_l \subset J \subsetneq N} \beta(J)\Big((P(J\cup i) - P(J))b_i - c_i\Big) \leq 0 \quad \forall i \in I_k$$

with strictness for at least one $i \in I_k$. In other words, for every distribution β which has support only on action profiles where agents in $\cup_{l=1}^{k-1} I_l$ are working, either there

exists $i \in I_k$ such that

$$\sum_{J \text{ s.t. } i \notin J, \cup_{l=1}^{k-1} I_l \subset J \subsetneq N} \beta(J)\Big((P(J \cup i) - P(J))b_i - c_i\Big) > 0 \tag{D}$$

or for all $i \in I_k$, the inequality holds at equality. Suppose β involves agent $i \in I_k$ shirking with positive probability. Then, if all agents' bonuses were slightly increased, at the perturbed bonuses, D must hold for agent i.

Consider the special case when I_k is a singleton cell consisting of agent i. When individuals' bonuses are perturbed upwards, for every β, the condition D holds for i, so given that all agents in $I_1, ..., I_{k-1}$ are working, there is no belief over whether agents in $I - \cup_{l=1}^{k-1} I_l$ are working such that i's best response is shirking. In other words, shirking is not rationalizable for i.[7] Now, consider the case when I_k is not a singleton. When bonuses are perturbed upwards, for every distribution β where at least one worker in I_k is shirking with positive probability, D says that there is at least one agent $i \in I_k$ that, when recommended to shirk according to β, would strictly prefer to *disobey* their recommendation and work instead. So, only all agents in I_k working with probability 1 can be part of the outcome of a correlated equilibrium.

As gestured at in the introduction, the difference between rationalizability and correlated equilibrium is roughly the common knowledge of equilibrium play. Proposition 4 characterizes this gap more. The restriction that agents share a common belief about others' strategies "binds" in a particular way: agents internalize that others in their cells must share the same beliefs about the distribution of play and understand that agents in preceding cells are part of their own groups that establish they must be working, but they can hold any beliefs about how agents in succeeding cells play.

The proofs of Propositions 3 and 4 rely on similar observations, so I discuss the high level ideas of Proposition 4 only. First, I show that for any bonus profile b that virtually uniquely implements in CE, I can construct an ordered partition $\mathcal{P}$ and $\alpha \gg 0$ such that $(b_i)_{i \in I_k}$ is feasible for $V_{\mathcal{P},k}(\alpha)$ for each cell I_k in the partition. I do this by considering a sequence of bonuses $\{b^n\}$ that converges to b from above so that $b_i^n \geq b_i$ for each i and for all n. Each b^n uniquely implements in correlated equilibrium and so has associated α^n multipliers such that (b^n, α^n) satisfy 2 and 1.

It is possible that many coordinates of the α^n multipliers converge to 0 as n converges

[7] See Pearce (1984).

to ∞, but these coordinates could be doing so at different speeds. The different rates of convergence produce the ordered cells of the partition with agents in higher indexed cells corresponding to coordinates of the multipliers that converge to 0 quicker. As b^n approaches b, agents in cells with higher indexes contribute less and less to the aggregate incentive constraints that contain agents in cells with low indexes until their gain from deviating disappears from the constraints. This implies the satisfaction of the cell problem constraints 4 and establishes that:

$$\inf_{\mathcal{P},\alpha\gg 0} \sum_k V_{\mathcal{P},k}(\alpha) \leq UCE$$

Next, I show that given an arbitrary partition $\mathcal{P}$ and $\alpha \gg 0$, the solutions to the induced cell problems produce a bonus profile b that virtually uniquely implements in CE. To show this, I consider a perturbation of b where every agent's bonus is increased by a positive ε. It is possible to construct $\hat{\alpha}$ so that the perturbed b and $\hat{\alpha}$ satisfy 1 and 2. The idea is simple: since increasing each agent's bonus by ε allows 4 to be satisfied strictly, I can go from the cell problem constraints to the aggregated IC constraints 2 by keeping $\hat{\alpha}_i$ for i in higher indexed cells sufficiently small relative to $\hat{\alpha}_j$ for j in lower indexed cells. This shows that

$$\inf_{\mathcal{P},\alpha\gg 0} \sum_k V_{\mathcal{P},k}(\alpha) = UCE$$

Finally, it is possible to show that UCE is actually achieved by a bonus profile that virtually uniquely implements in correlated equilibrium, so I can get Proposition 4. The full details of the proofs of Proposition 3 and 4 are in the appendix.

Using Propositions 3 and 4, UCE and URS do not coincide if there is a partition $\notin \mathcal{O}$ and $\alpha \gg 0$ cheaper than any partition in $\mathcal{O}$. Let's consider when this can happen. To ensure that working is the unique rationalizable action, the principal must pay agents enough so that they prefer working to shirking even at their worst case conjecture about who else is working. The principal can be clever and build this assurance by "dividing-and-conquering." Rather than paying all agents a lot to ensure they find working optimal, she can pay a select few a lot. The common knowledge of rationality and payoffs will allow agents that are not paid as much to reason that those who are must be working. This shrinks the space of conjectures about others' actions that these workers can entertain, so even though these workers are not paid as much, they will still find working optimal.

In contrast, when seeking to implement in a correlated equilibrium, the principal can

utilize agents' common belief about equilibrium play to avoid paying out bonuses tailored to individuals' worst case conjectures because these conjectures might involve recommendations that agents would not find obedient. The following result expands on this idea and presents a sufficient condition for there to be a gap between UCE and URS:

Proposition 5. *Let b^R be the bonus profile that virtually implements in unique RS and has cost URS. Let b^R have associated partition $\mathcal{P} \in \mathcal{O}$ and let agents be indexed according to the order on cells in the partition so that $k \in I_k$. Let J^k denote the worker set J s.t. $1, ..., k-1 \in J$ and $(P(J^k \cup k) - P(J^k))b_k^R - c_k = 0$.*

Given generic P, if $P(J^k \cup k+1) - P(J^k) > P(J^{k+1} \cup k+1) - P(J^{k+1})$ for some k, then $UCE < URS$.

Proposition 5 fixes the cost-minimizing order in which agents eliminate shirking as rationalizable. The sufficient condition requires $k+1$ to be more productive when J^k is working than when J^{k+1} is, so k and $k+1$ disagree regarding which workers sets they are more productive with. k has a relatively low marginal contribution to success when J^k is working, but $k+1$ is relatively productive at J^k. The principal implementing in rationalizable strategies must convince k to work even if he hypothesizes J^k works, but the principal implementing in CE can deal with k hypothesizing J^k works by utilizing k and $k+1$'s shared belief about play to demonstrate that J^k cannot occur in equilibrium because $k+1$ is paid enough to ensure he would join in on working.

The proof proceeds by using the results of Propositions 3 and 4. Given that b^R is associated with the partition $\mathcal{P}$, I consider an alternative partition $\hat{\mathcal{P}}$ where all cells are the same as in $\mathcal{P}$ except the kth and the $k+1$-th cell have been merged. Then, I fix the bonus profile b^R. It is possible to produce new limit multipliers $\hat{\alpha}$ so that under $\hat{\alpha}_k$ and $\hat{\alpha}_{k+1}$, the old bonuses (b_k^R, b_{k+1}^R) satisfy the 4 for $\hat{I}_k$ strictly. Agent k's bonus is set to motivate k to work if just J^k are working, but by including $k+1$ in the same cell, k's bonus can be lowered since $k+1$'s bonus already contributes significantly to the aggregated incentive terms 4.

I conclude this section by revisiting a motivating example to apply Proposition 5.

Example 2 Revisited.
I can demonstrate that the gap between UCE and URS exists by applying Proposition

5. Bonus profile b^R corresponds to partition

$$\mathcal{P} = \{I_1 = \{2\}, I_2 = \{1\}, I_3 = \{3\}\}$$

The key worker set that determines 2's bonus is $J^1 = \emptyset$. Given that 2 must be working, the key worker set that determines 1's bonus is $J^2 = \{2, 3\}$, and the analogous set for 3's bonus is $J^3 = \{1, 2\}$. Note that $P(1) - P(\emptyset) > P(1, 2, 3) - P(2, 3)$ so the condition in the proposition is satisfied. In fact, a partition that puts all agents in one cell with $\alpha = (0.15, 1, 0.15)$ will verify bonus profile $(2, 2.1, 2)$ as virtually implementing in UCE.

7 Conclusion

This paper studies unique implementation in the team production game while relaxing the supermodularity assumption used in the literature before. Without supermodularity, there are important distinctions between different solution concepts. The main results in this paper explore the assumptions on the production technology that produce gaps or no gaps between the costs of uniquely implementing in Nash equilibrium, in correlated equilibrium, and in rationalizable strategies.

References

Becker, Gary S. "A theory of marriage: Part I". *Journal of Political economy* 81, no. 4 (1973): 813–846.

Halac, Marina, Ilan Kremer, and Eyal Winter. "Raising capital from heterogeneous investors". *American Economic Review* 110, no. 3 (2020): 889–921.

Halac, Marina, Elliot Lipnowski, and Daniel Rappoport. "Addressing strategic uncertainty with incentives and information". In *AEA Papers and Proceedings*, 112:431–437. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2022.

— . "Rank uncertainty in organizations". *American Economic Review* 111, no. 3 (2021): 757–786.

Inostroza, Nicolas, and Alessandro Pavan. "Adversarial coordination and public information design". *Available at SSRN 4531654* (2023).

Milgrom, Paul, and John Roberts. "Rationalizability, learning, and equilibrium in games with strategic complementarities". *Econometrica: Journal of the Econometric Society* (1990): 1255–1277.

Moriya, Fumitoshi, and Takuro Yamashita. "Asymmetric-information allocation to avoid coordination failure". *Journal of Economics & Management Strategy* 29, no. 1 (2020): 173–186.

Morris, Stephen, Daisuke Oyama, and Satoru Takahashi. "Implementation via Information Design in Binary-Action Supermodular Games". *Econometrica* 92, no. 3 (2024): 775–813.

Myerson, Roger B. "Dual reduction and elementary games". *Games and Economic Behavior* 21, **numbers** 1-2 (1997): 183–202.

Neyman, Abraham. "Correlated equilibrium and potential games". *International Journal of Game Theory* 26, no. 2 (1997): 223–227.

Pearce, David G. "Rationalizable strategic behavior and the problem of perfection". *Econometrica: Journal of the Econometric Society* (1984): 1029–1050.

Sakovics, Jozsef, and Jakub Steiner. "Who matters in coordination problems?" *American Economic Review* 102, no. 7 (2012): 3439–3461.

Segal, Ilya. "Contracting with externalities". *The Quarterly Journal of Economics* 114, no. 2 (1999): 337–388.

— . "Coordination and discrimination in contracting with externalities: Divide and conquer?" *Journal of Economic Theory* 113, no. 2 (2003): 147–181.

Tucker, Albert W. "1. Dual Systems of Homogeneous Linear Relations". *Linear Inequalities and Related Systems.(AM-38)* (1957): 1–18.

Ui, Takashi. "Correlated equilibrium and concave games". *International Journal of Game Theory* 37 (2008): 1–13.

Viossat, Yannick. "Is having a unique equilibrium robust?" *Journal of Mathematical Economics* 44, no. 11 (2008): 1152–1160.

Winter, Eyal. "Incentives and discrimination". *American Economic Review* 94, no. 3 (2004): 764–773.

A Omitted Proofs for Section 5

A.1 Proof of Proposition 4

To show Proposition 4, we first show that any b that virtually implements in unique correlated equilibrium has an associated ordered partition and $\alpha \gg 0$ such that $(b_i)_{i \in I_k}$ is feasible for each I_k cell problem so

$$\inf_{\mathcal{P}, \alpha \gg 0} \sum_k V_{\mathcal{P},k}(\alpha) \leq UCE$$

Then, we show that given a partition $\mathcal{P}$ and $\alpha \gg 0$, every bonus b such that $(b_i)_{i \in I_k}$ solves the I_k cell problem virtually implements in unique correlated equilibrium. From this, we can conclude that there cannot be a strict gap between the LHS and the RHS of the above inequality since this would imply the existence of a bonus profile arbitrarily close in cost to the LHS but strictly below the RHS. We now know that

$$\inf_{\mathcal{P}, \alpha \gg 0} \sum_k V_{\mathcal{P},k}(\alpha) = UCE$$

The last step of the proof is to demonstrate that there is a bonus profile that virtually implements in unique correlated equilibrium (and so is weakly above the LHS in cost) that attains the cost of UCE. The infimum on the LHS is actually attained by some partition and $\alpha \gg 0$.

Lemma 4. *Given that b virtually implements in UCE, $\exists$ a partition of agents $\mathcal{P} = \{I_1, ..., I_K\}$ with ordered cells and $(\alpha_i) \gg 0$ where $(b_i)_{i \in I_k}$ satisfies*

$$\sum_{i \in I_k - J} \alpha_i \Big((P(J \cup i) - P(J)) b_i - c_i \Big) \geq 0 \quad \forall J \subsetneq N, \cup_{l=1}^{k-1} I_l \subseteq J \tag{5}$$

Proof. Fix a sequence of $\{\varepsilon_n\}_n$ where $\varepsilon_n > 0$ for all n and $\varepsilon_n \to 0$. Since b virtually implements in UCE, $b + \varepsilon_n$ implements in UCE for each n. Thus, each $b + \varepsilon_n$ can be associated with some $\alpha^n \geq 0$ multipliers so that $(\alpha^n, b + \varepsilon_n)$ satisfy 2. Recall that $\{\alpha^n\}_n$ are such that $\alpha_i^n \in [0, 1]$ for each i. Note also that we can without loss normalize $\max_i \alpha_i^n$ to be 1; the strict 2 constraints guarantee that fixing n, not all α_i^n can be 0. The $\{\alpha^n\}_n$ form a bounded sequence and so they have a convergent subsequence $\{\alpha^{n_p}\}_p$ with limit point α^1. $\max_i \alpha_i$ is a continuous function of α, so $\max_i \alpha_i^1 = 1$. We have that α^1 must contain some coordinates that are strictly greater than 0. Let I_1 be the

set of all i such that $\alpha_i^1 > 0$. For $J \subsetneq I$ where $I_1 \cap \big(I - J\big) \neq \emptyset$, we have that

$$\sum_{i \in I_1 - J} \alpha_i^1 \Bigg((P(J \cup i) - P(J))b_i - c_i \Bigg) \geq 0$$

For J where $I_1 \subset J$, when we take the limit of 2 as $\alpha^{n_p} \to \alpha^1$, we see that the LHS of 2 converges to 0.

Now, suppose we have defined $I_1, ..., I_{k-1}$ and $\alpha_i \gg 0$ for $i \in \cup_{l=1}^{k-1} I_{k-1}$. We can create a truncation of each α^n, $\hat{\alpha}^n = (\alpha_i^n)_{i \notin \cup_l^{k-1} I_l}$. Consider $J \subsetneq I$ s.t. $\cup_{l=1}^{k-1} I_l \subset J$. From 2, we know the following holds for all such J

$$\sum_{i \notin J} \alpha_i \Bigg((P(J \cup i) - P(J))b_i - c_i \Bigg) > 0$$

We do the same trick as before, normalizing $\max_i \hat{\alpha}_i^n$ to be 1, which does not affect satisfaction of the above J inequalities. Then, we can once again study a convergent subsequence of $\{\hat{\alpha}^n\}$, denoting its limit point α^k. α^k is guaranteed to have a non-zero coordinate since $\max_i \hat{\alpha}_i^n = 1$ for all n, so $\max_i \alpha_i^k = 1$. For $J \subsetneq I$ s.t. $\cup_{l=1}^{k-1} I_l \subsetneq J$,

$$\sum_{i \in I_k - J} \alpha_i^k \Bigg((P(J \cup i) - P(J))b_i - c_i \Bigg) \geq 0$$

By iterating this process, we produce all cells of the partition and inequalities in the statement of the lemma. □

Lemma 5. *Fix a partition $\mathcal{P}$ with ordered cells $\{I_1, ..., I_K\}$ and $\alpha \gg 0$. Let b denote the bonus profile where $(b_i)_{i \in I_k}$ solves the I_k cell problem. b virtually implements in unique correlated equilibrium.*

Proof. To see this, consider an upward perturbation of b, $\hat{b}$, where for all agents, $\hat{b}_i = b_i + \varepsilon$ for some $\varepsilon > 0$. The constraints of each cell level subproblem hold strictly at $\hat{b}$ given the choice of α. To verify that $\hat{b}$ implements in UCE, we need to produce some $\hat{\alpha} \geq 0$ multipliers so that $\hat{\alpha}$ and $\hat{b}$ satisfy 2. To construct $\hat{\alpha}$, we start by setting

$$\hat{\alpha}_i = \alpha_i \quad \forall i \in I_1$$

At $\hat{b}$, 4 hold strictly for I_1. Next, we proceed recursively. Suppose we have set $\hat{\alpha}_i$ for

all $i \in \cup_{l=1}^{k-1} I_l$ so that modified J constraints hold strictly:

$$\sum_{i \in \cup_{l=1}^{k-1} I_l - J} \hat{\alpha}_i \Big((P(J \cup i) - P(J))\hat{b}_i - c_i \Big) > 0 \quad \forall J \subsetneq I \text{ s.t. } \cup_{l=1}^{k-1} I_l \not\subset J$$

Now, it is possible to choose $r_k \in (0, 1]$ such that $\forall J \subsetneq I$ s.t. $\cup_{l=1}^{k-1} I_l \not\subset J$,

$$\sum_{i \in \cup_{l=1}^{k-1} I_l - J} \hat{\alpha}_i \Big((P(J \cup i) - P(J))\hat{b}_i - c_i \Big) + r_k \sum_{i \in I_k - J} \alpha_i \Big((P(J \cup i) - P(J))\hat{b}_i - c_i \Big) > 0$$

since there are a finite number of $J \subsetneq I$ s.t. $\cup_{l=1}^{k-1} I_l \not\subset J$. Additionally, for $J \subsetneq I$ s.t. $\cup_{l=1}^{k-1} I_l \not\subset J$, we have that

$$r_k \sum_{i \in I_k - J} \alpha_i \Big((P(J \cup i) - P(J))\hat{b}_i - c_i \Big) > 0$$

since $\hat{b}_i > b_i$ for all $i \in I_k$ and $r_k > 0$. Set $\hat{\alpha}_i = r_k \alpha_i$ for $i \in I_k$. This iterative construction of $\hat{\alpha}$ enables us to satisfy all 2 constraints given $\hat{b}$. □

Finally, observe that there exists a sequence of bonus profiles $\{b^n\}_n$ such that b^n implements in UCE for all n and $\sum_i b_i^n$ converges to UCE. Without loss, this sequence $\{b^n\}$ is bounded. It has a convergent subsequence, denoted $\{b^{n_p}\}_p$. Let $\bar{b}$ be the limit point of this subsequence. $\{\sum_i b_i^{n_p}\}_p$ is a subsequence of the original cost sequence $\{\sum_i b_i^n\}$, so it must converge to UCE as well. $\sum_i b_i$ is a continuous function of b, so $\sum_i \bar{b}_i = UCE$. Because $\bar{b}$ is the limit of bonuses that implement in UCE, $\bar{b}$ virtually implements in UCE. To see why this is, consider any $\varepsilon > 0$ and define a set of bonuses weakly below $\bar{b} + \varepsilon$

$$B^\varepsilon := \{b | b_i \leq \bar{b}_i + \varepsilon \forall i\}$$

If there exists a $b \in B^\varepsilon$ such that we can produce $\alpha \geq 0$ so that 1 and 2 are satisfied, then we can conclude that $\bar{b} + \varepsilon$ also implements in UCE since $\bar{b} + \varepsilon$ differs from b by a weakly positive vector $\delta \geq 0$. So, using the same α that verify b as implementing in UCE, we can conclude that $b + \varepsilon$ also implements in UCE since every payoff difference term is weakly greater at $\bar{b} + \varepsilon$ compared to b. Since $\{b_i^{n_p}\}_p$ converges to $\bar{b}$ and $\bar{b} \in B^\varepsilon$, we are guaranteed the existence of such a b.

Now, $\bar{b}$ virtually implements in UCE and has cost exactly UCE. Using lemma 4, we

know that

$$\inf_{\mathcal{P},\alpha\gg 0} \sum_k V_{\mathcal{P},k}(\alpha) \leq \sum_i \bar{b}_i$$

So the inequality holds at equality and the infimum can be replaced by min. □

A.2 Proof of Proposition 3

The same observations that drive the proof of Proposition 4 help us here. First, any b that virtually implements in unique rationalizable strategies can be mapped to a partition $\mathcal{P} \in \mathcal{O}$:

$$\sum_{k=1}^{N} V_{\mathcal{P},k}(\alpha) \leq \sum_i b_i$$

where α can be taken without loss to be $(1, ..., 1)$ since each cell of $\mathcal{P}$ is a singleton. This gives us that

$$\min_{\mathcal{P}\in\mathcal{O},\alpha=1} \sum_{k=1}^{N} V_{\mathcal{P},k}(\alpha) \leq URS$$

Next, any $\mathcal{P} \in \mathcal{O}$ induces cell problems and associated bonuses b that virtually implement in unique rationalizable strategies, so the above inequality is an equality.
Now, let's show the intermediate claims to complete the proof. First, suppose b virtually implements in unique rationalizable strategies. We can take a sequence of $\{\varepsilon_n\}_n$ such that $\varepsilon_n > 0$ and $\varepsilon_n \to 0$. Consider the sequence of bonuses $\{b+\varepsilon_n\}_n$ where $\varepsilon_n > 0$ for all n and $\varepsilon_n \to 0$. Each bonus in this sequence implements in unique rationalizable strategies and so is associated with an order over agents in which they eliminate not working as rationalizable. This order can equivalently be represented as an ordered partition $\mathcal{P}_n \in \mathcal{O}$. It is without loss to set $\alpha_i^n = 1$ for all i and for all n in the sequence since the value of $V_{\mathcal{P},k}(\alpha)$ does not depend on α if $I_k \in \mathcal{P}$ is a singleton. There are a finite number of ordered partitions possible, so along the sequence, some partition, call it $\mathcal{P}$, must occur infinitely often. We can then choose a subsequence of $\{b + \varepsilon_n\}_n$, selecting only the elements of the sequence that have associated partition $\mathcal{P}$. This subsequence converges to b, so we have that 5 holds for b, $\alpha = 1$, and partition $\mathcal{P} \in \mathcal{O}$.

Next, fix $\mathcal{P} \in \mathcal{O}$. Index agents so that $k \in I_k$ for each $k = 1, ..., N$. Let bonus profile b be such that b_k solves $V_{\mathcal{P},k}(1)$ for each k. Consider a small perturbation of b, $b + \varepsilon$ for $\varepsilon > 0$. Given bonus $b_1 + \varepsilon$, all the 4 inequalities for the I_1 cell are satisfied strictly,

so 1 finds working a strictly dominant action. Suppose $1, ..., k-1$ find working strictly dominant. Given bonus $b_k + \varepsilon$, the k cell inequalities are satisfied strictly, so k finds working strictly dominant. We conclude that given bonuses $b + \varepsilon$, each agent has a unique rationalizable action, working. b virtually implements in unique rationalizable strategies. □

A.3 Proof of Proposition 5

Proof. To show the statement, we consider b^R and demonstrate that for an ordered partition that is *not* all singleton cells and some $\alpha \gg 0$, $\sum_k V_{\mathcal{P},k}(\alpha) < \sum_i b_i^R$.

Proposition 3 tells us that b^R has associated partition $\mathcal{P} \in \mathcal{O}$. Index agents according to the order of their cell, so 1 is the agent in cell I_1. We consider a minimal change to the partition $\mathcal{P}$ and study $\tilde{\mathcal{P}} := \{\tilde{I}_l\}_{l=1}^{N-1}$ where

$$\tilde{I}_l = \begin{cases} I_l & \text{if } l < k \\ I_k \cup I_{k+1} & \text{if } l = k \\ I_{l+1} & \text{if } l > k \end{cases}$$

$\tilde{\mathcal{P}}$ merges the k and $k+1$th cells of the original partition $\mathcal{P}$. We will consider a particular alternate α to pair $\tilde{\mathcal{P}}$ with as well. We restrict (without loss) $\tilde{\alpha}_i = 1$ for all $i \neq k+1$. We will demonstrate that there is $\tilde{\alpha}_{k+1} > 0$ such that

$$\sum_{l}^{N-1} V_{\tilde{\mathcal{P}},l}(\tilde{\alpha}) < \sum_{l}^{N} V_{\mathcal{P},l} \tag{6}$$

Note that given our restrictions on the cells of $\tilde{P}$,

$$V_{\tilde{\mathcal{P}},l}(\tilde{\alpha}) = V_{\mathcal{P},l}(\alpha) \quad \forall l < k$$
$$V_{\tilde{\mathcal{P}},l}(\tilde{\alpha}) = V_{\mathcal{P},l+1}(\alpha) \quad \forall l > k$$

so the only remaining question is how $V_{\tilde{\mathcal{P}},k}(\tilde{\alpha})$ compares to $V_{\mathcal{P},k}(\alpha) + V_{\mathcal{P},k+1}(\alpha)$.

To get that 6 holds, we'll show that we can produce $\tilde{\alpha}_{k+1}$ so that at $b_k = b_k^R$ and $b_{k+1} = b_{k+1}^R$, all constraints in the $V_{\tilde{\mathcal{P}},k}$ problem hold at > 0 for constraints where b_k appears. b_k can be lowered strictly and the resulting bonus would still be feasible at $\tilde{\alpha}_k$ and $\tilde{\alpha}_{k+1}$. 6 would hold then.

Let's think about the relationship between $V_{\mathcal{P},k}(\alpha)$, $V_{\mathcal{P},k+1}(\alpha)$, and $V_{\tilde{\mathcal{P}},k}(\tilde{\alpha})$ at $b_k = b_k^R$

and $b_{k+1} = b^R_{k+1}$. The original $V_{\mathcal{P},k}(\alpha)$ problem has these constraints:

$$(P(J \cup k) - P(J))b^R_k - c_k \geq 0 \quad \forall J \subsetneq N, \{1, ..., k-1\} \subseteq J, k \notin J$$

where I have imposed that $\alpha_k = 1$. As long as P is generic, only the J^k constraint was binding; all other J constraints held > 0. Similarly, the original $V_{\mathcal{P},k+1}(\alpha)$ problem has constraints

$$(P(J \cup k+1) - P(J))b^R_{k+1} - c_{k+1} \geq 0 \quad \forall J \subsetneq N, \{1, ..., k\} \subseteq J, k+1 \notin J$$

and the genericity of P implies that only the J^{k+1} is binding. Other constraints are slack.

Let's compare to the new problem $V_{\tilde{\mathcal{P}},k}$. The $\{k, k+1\}$ problem has constraints for J s.t.

$$\tilde{\alpha}_k \Big((P(J \cup k) - P(J))b^R_k - c_k \Big) \geq 0 \quad \forall J \subsetneq N, \{1, ..., k-1, k+1\} \subseteq J$$

$$\tilde{\alpha}_k \Big((P(J \cup k) - P(J))b^R_k - c_k \Big) + \tilde{\alpha}_{k+1} \Big((P(J \cup k+1) - P(J))b^R_{k+1} - c_{k+1} \Big) \geq 0 \quad \forall J \subsetneq N, \{1, ..., k-1$$

$$\tilde{\alpha}_{k+1} \Big((P(J \cup k+1) - P(J))b^R_{k+1} - c_{k+1} \Big) \geq 0 \quad \forall J \subsetneq N, \{1, ..., k\} \subseteq J$$

The first set of constraints look exactly like their counterparts in the $V_{\mathcal{P},k}(\alpha)$. Similarly, the third set of constraints look like their counterparts in the $V_{\mathcal{P},k+1}(\alpha)$ problem. The second set of constraints resemble constraints in the $V_{\mathcal{P},k}(\alpha)$ but now have additional terms. Recall that by hypothesis J^k is in the second set of constraints.

At b^R, every J constraint in the first set of constraints holds > 0. Every J constraint $\neq J^k$ in the second set holds strictly > 0 in the original k problem, so it is possible to choose $\tilde{\alpha}_{k+1}$ to small but strictly positive to ensure that every constraint $\neq J^k$ in the second set still holds strictly > 0. For J^k, the hypothesis of the proposition guarantees that $k+1$'s term in the new J^k constraint is strictly positive, so with small but strictly positive $\tilde{\alpha}_{k+1}$, we have that the new J^k constraint holds strictly > 0. This is as we desired. We conclude that 6 holds. □