

## ROBUST INCENTIVES FOR TEAMS

TIANJIAO DAI  
MetaJUUSO TOIKKA  
The Wharton School, University of Pennsylvania

We show that demanding team incentives to be robust to nonquantifiable uncertainty about the game played by the agents leads to contracts that align the agents' interests. Such contracts have a natural interpretation as team-based compensation. Under budget balance they reduce to linear contracts, thus identifying profit-sharing, or equity, as an optimal contract absent a sink or a source of funds. A linear contract also gives the best profit guarantee to an outside residual claimant. These contracts still suffer from the free-rider problem, but a positive guarantee obtains if and only if the technology known to the contract designer is sufficiently productive.

KEYWORDS: Moral hazard, robustness, teams, incentives, interest alignment, optimal contracts, linear contracts, maxmin.

## 1. INTRODUCTION

THE STANDARD CONTRACT-THEORETIC approach to motivating teams, pioneered by [Holmström \(1982\)](#), emphasizes informational aspects of the problem. It holds that any signal informative of an agent's action be used to determine his compensation. A recognized shortcoming of this approach is that it leads to contracts that are sophisticated and highly context-dependent. Moreover, because the focus is on individual performance, team-based pay that aligns the agents' compensation emerges only under very specific assumptions about technology. This contrasts with incentive schemes observed in practice, which tend to be simpler and often include team-based compensation even if information about individual performance may be available. For instance, partnerships commonly operate under profit-sharing agreements, firms use team incentives to motivate employees, and academic economists share credit equally for coauthored papers.

In this paper, we investigate foundations for such simple incentive schemes by considering contracts that are robust to nonquantifiable uncertainty about the game played by the agents. Our model is based on the classic team production problem, where the agents take costly unobservable actions, which jointly determine a stochastic contractible output. Our main specification assumes all parties to be risk-neutral and that the agents are protected by limited liability, but allows for a general production technology.

The game is common knowledge among the agents, perhaps by virtue of their expertise, or because it is simply evident now that they have been called to act. However, inspired by [Carroll's \(2015\)](#) work on the foundations of linear contracts in principal-agent problems, we assume that the principal designing the contract only knows some of the actions available to each agent, and hence she only knows some of the action profiles in the game. Realizing that the game may be bigger than she thinks, but not having a prior on the set of

---

Tianjiao Dai: [daitj207@gmail.com](mailto:daitj207@gmail.com)

Juuso Toikka: [toikka@wharton.upenn.edu](mailto:toikka@wharton.upenn.edu)

We thank Gabriel Carroll, Tommaso Denti, Bob Gibbons, Bengt Holmström, Ayca Kaya, Stephen Morris, Joel Sobel, Alex Wolitzky, and anonymous referees for comments. The usual disclaimer applies.

possible games, the principal evaluates contracts based on their guaranteed performance across all games consistent with her knowledge.

Our first result shows that guaranteeing good performance either in terms of the expected surplus for a budget balanced team, or in terms of the principal's profit if she is the residual claimant, requires the contract to align the agents' interests. That is, each agent's compensation should covary positively and linearly with the compensation of all other agents. Such a contract is affine in some one-dimensional aggregate of the output (but not necessarily in the output itself), so it can be naturally interpreted as prescribing team-based pay. Thus, team-based compensation emerges even though richer measures of performance may be available under the action profiles known to the principal.

The necessity of interest alignment derives from the fact that when a contract induces disagreement about the ranking of (stochastic) outputs among the agents, then—should the game provide an opportunity for it—an agent may seek individual gain at a social cost. We show that for essentially any contract that fails to align the agents' interests, there is a game where such selfish actions create a “race to the bottom,” with the unique equilibrium output distribution concentrated at the worst possible output. In a sense, the construction generalizes the well-known problematic incentive properties of rank-order tournaments (e.g., Lazear (1989)), showing that for robustness, they are the overriding concern. While the result is reminiscent of Carroll's (2015) linearity result, the two are logically independent. On one hand, the definition of interest alignment only involves the payments to the agents, so every contract trivially aligns the agents' interests in the single-agent case. On the other hand, a contract that aligns the agents' interests need not be linear in the value of output.

For contracts that are budget balanced among the agents, interest alignment is equivalent to paying each agent a fixed share of the output's monetary value. We show that some such linear contract achieves the best surplus guarantee subject to budget balance, thus singling out profit-sharing, or equity, as an optimal arrangement. Along the way, we characterize the optimal guarantee, which is made tractable by the fact that any contract that aligns interests induces a potential game among the agents.

We also show that a linear contract achieves the best guarantee for the principal's profit. By our first result, we can focus on contracts that align the agents' interests. The candidate optimal contracts can then be represented as consisting of a function specifying the agents' total compensation for each output, and of shares determining how it is divided among the agents. We show that, holding the shares fixed, total compensation should be linear in the output's value, and so the contract should be linear overall. Heuristically, a linear contract aligns interests across all parties, including the principal.

Whether the optimal guarantees for surplus and profit are positive depends on the severity of the free-rider problem. Unlike in the case of one agent, it is not enough that some known action profile generates a positive surplus. Instead, the condition that characterizes known production technologies for which the optimal guarantees are nontrivial requires a social planner to be able to generate positive surplus in a model where the agents' costs are appropriately inflated to account for the robustness concern. Thus, even absent setup costs, only sufficiently profitable teams are worth forming.

While our results are the strongest with risk-neutrality, Section 6 shows that nontrivial performance guarantees require team-based compensation also when agents are risk-averse. The agents' interests must then be aligned in the utility space, which translates to monetary payments that covary positively across agents in the sense that if one agent's pay increases, so does the pay of all other agents. Thus, the basic logic holds irrespective of risk attitudes. Even collinearity of payments can be recovered for a subset of CRRA preferences.

The question of foundations for linear contracts has received a great deal of attention in the one-agent case, starting with Holmström and Milgrom (1987). See Carroll (2015) for a review of this literature. As we focus on the contracts' guaranteed performance, our work belongs to the literature studying worst-case optimal contracts in various settings; see, for example, Hurwicz and Shapiro (1978), Chung and Ely (2007), Chassang (2013), Frankel (2014), Garrett (2014), Yamashita (2015), Carroll (2017), Carroll and Segal (2019), and Marku and Ocampo Diaz (2019).<sup>1</sup> Similar robustness concerns motivate the work on robust mechanism design following Bergemann and Morris (2005), and the analysis of approximately optimal contracts in locally misspecified models by Madarász and Prat (2017).

Theoretical explanations for the use of profit-sharing and for the prevalence of partnerships as an organizational form have been put forth by Garicano and Santos (2004) and Levin and Tadelis (2005), among others. Che and Yoo (2001) show that team-based pay can be a part of the optimal mix of formal and relational incentives in a repeated partnership problem where the agents can monitor each others' actions.

Finally, the need to align the agents' interests resonates with some themes in the extensive management literature on teams. For example, Hackman (2002) posits that a key enabling condition for work-team effectiveness is the existence of a compelling direction that should specify ends but not means. Interpreting the "means" as the agents' actions and the "ends" as the contractible output, a contract that aligns the agents' interests provides just that.<sup>2</sup>

## 2. MODEL

We consider a principal motivating a team consisting of one or more agents, indexed by  $i = 1, \dots, I$ . The team's observable output  $y$  is an element of a finite set  $Y$  held fixed throughout the analysis. Its intrinsic value is denoted  $v(y)$ . For example,  $v(y)$  may be the expected market value of the team's production conditional on the signal  $y$ , or it may reflect how the principal aggregates different dimensions of performance. We denote by  $y_0$  the least desirable output and set its value to zero:  $v(y_0) = \min v(Y) = 0$ . (Output  $y_0$  can be chosen arbitrarily among the minimizers if there are many.) To avoid trivialities, we assume that  $\max v(Y) > 0$ .

A (*production*) *technology* for the team is a tuple  $(A, c, F)$ , where  $A := \times_{i=1}^I A_i$  is the finite set of action profiles,  $c : A \rightarrow \mathbb{R}_+^I$  is the profile of cost functions, and  $F : A \rightarrow \Delta(Y)$  is the family of output distributions. We restrict attention to technologies where each agent's cost depends only on his own action, that is,  $c_i(a) = c_i(a_i)$ . Any technology describes a version of the classic moral-hazard-in-teams problem: every agent takes an unobservable action  $a_i \in A_i$  at a private cost  $c_i(a_i) \geq 0$ , and the resulting action profile  $a = (a_1, \dots, a_I)$  determines the output distribution  $F(a) \in \Delta(Y)$ .

The principal can motivate the agents with monetary rewards contingent on the realized output. We assume that the agents are protected by limited liability, meaning that payments to them have to be nonnegative. An incentive scheme, or a *contract*, is thus a function  $w : Y \rightarrow \mathbb{R}_+^I$  that specifies a payment profile  $w(y) = (w_1(y), \dots, w_I(y))$  for

<sup>1</sup>The literature has continued to grow since our paper was first circulated. Most closely related are Carroll and Walton (2021) who give an alternative proof for going from our aligned-interest contracts to a linear principal-optimal contract, and Kambhampati (2022) who studies robust performance evaluation of agents who are known to be operating identical unknown technologies, but who cannot affect each others' output.

<sup>2</sup>This is true quite literally: the parameter  $d$  in our Lemma 3.1(iii) is the direction of the ray in  $\mathbb{R}_+^I$  along which all payment profiles lie.

each output  $y$ . Agent  $i$ 's net payoff is then  $w_i(y) - c_i(a_i)$ , with the principal receiving  $v(y) - \sum_i w_i(y)$ . All parties are assumed risk neutral, but we discuss risk-averse agents in Section 6.

We say that the contract  $w$  is *budget balanced* if the value of output is shared by the agents, that is, if  $\sum_i w_i(y) = v(y)$  for all  $y$ .

The principal designs the contract either to maximize surplus subject to budget balance, or to maximize her profits. However, she does so without full knowledge of the game played by the agents. Specifically, we assume that the true technology is common knowledge among the agents, but the principal only knows of some technology  $(A^0, c^0, F^0)$ , referred to as the *known technology*. The principal believes that the true technology may be any technology  $(A, c, F)$  such that  $A \supseteq A^0$  and  $(c, F)|_{A^0} = (c^0, F^0)$ . That is, the true technology contains the action profiles known to the principal, and the true costs and output distributions associated with these profiles conform with the principal's knowledge. To simplify notation, we suppress the cost functions and output distributions, writing  $A^0$  and  $A$  for the known and the true technology, respectively.

We assume that the known technology  $A^0$  contains a zero-cost action for each agent. This simplifies some of the arguments without affecting our results qualitatively.<sup>3</sup> As we assume nothing about the associated output distributions, the loss in scope is minimal.

A contract  $w$  and the (true) technology  $A$  induce a normal form game  $\Gamma(w, A)$ , where agent  $i$ 's expected payoff is  $u_i(a; w, A) := \mathbb{E}_{F(a)}[w_i(y)] - c_i(a_i)$ . We write  $\mathcal{E}(w, A)$  for its set of mixed strategy Nash equilibria. An equilibrium exists because  $A$  was assumed finite. In case there are many, we adopt the usual partial-implementation assumption from contract theory and focus on the equilibrium that is best for the principal's objective.<sup>4</sup> Thus, the expected surplus induced by the contract  $w$  given technology  $A$  is

$$S(w, A) := \max_{\sigma \in \mathcal{E}(w, A)} \left( \mathbb{E}_{F(\sigma)}[v(y)] - \sum_a \sigma(a) \sum_i c_i(a_i) \right),$$

where  $F(\sigma)$  is the outcome distribution induced by  $F$  and the strategy profile  $\sigma$ . Similarly, the principal's expected profit from the contract  $w$  given technology  $A$  is

$$V(w, A) := \max_{\sigma \in \mathcal{E}(w, A)} \mathbb{E}_{F(\sigma)} \left[ v(y) - \sum_i w_i(y) \right].$$

Faced with the uncertainty about the game played by the agents, the principal ranks contracts according to their guaranteed expected performance over all possible (finite) technologies. For the surplus and profits, these guarantees are, respectively,

$$S(w) := \inf_{A \supseteq A^0} S(w, A) \quad \text{and} \quad V(w) := \inf_{A \supseteq A^0} V(w, A).$$

<sup>3</sup>See the working paper version (Dai and Toikka (2018)), where we did not make this assumption.

<sup>4</sup>This minimizes the departure from the standard model and ensures the existence of an optimal contract. However, essentially the same results hold under the alternative assumption that the agents play the worst equilibrium for the principal among equilibria that are not strictly Pareto dominated for the agents, but in this case optimal contracts may only exist in the sense of a limit. We omit the details in the interest of space. In contrast, simply selecting the worst equilibrium for the principal does not work as then all contracts only have a trivial guarantee. This is because we can simply add a profile  $\bar{a}$  of zero-cost actions such that (1)  $\bar{a}$  leads to output  $y_0$  for sure, and (2) so does any profile where only one agent has deviated from  $\bar{a}$ . Then  $\bar{a}$  is an equilibrium with output  $y_0$  no matter the contract or what other actions are available.

We say that a contract is *team-optimal* if it maximizes  $S(w)$  over all budget-balanced contracts. A contract is *principal-optimal* if it maximizes  $V(w)$  over all contracts. Note that  $S(w) \geq 0$ , since each agent can ensure a zero payoff by playing a zero-cost action in  $A_i^0$  given any technology  $A \supseteq A^0$ . On the other hand, the *zero contract*  $w \equiv 0$  yields a nonnegative expected profit from any technology, and hence  $V(0) \geq 0$ .

Some remarks regarding the formulation are in order. It is worth noting that we have deliberately assumed that a contract can only condition on the outcome  $y$ . This assumption captures the essence of our robustness exercise: we are interested in the performance of a fixed contract in varying circumstances. We thus explicitly rule out the possibility of tailoring the contract to the technology by asking the agents to report it to the principal.<sup>5</sup>

The most immediate interpretation is that the principal is designing a contract for a single team, not fully aware of the game the agents are playing. This could reflect the agents' superior knowledge of the situation, or be due to the principal having to design the contract before the details, or the team's members, are known. The principal can, however, envision and evaluate all possible outputs that may arise as a result of the team's activities, that is, she knows the set  $Y$  and the mapping  $v : Y \rightarrow \mathbb{R}$ . That  $Y$  is held fixed is not restrictive as our main results do not require output distributions to have full support.

An alternative interpretation is that the contract is to be used in a number of different situations, perhaps by different teams, and we want it to guarantee good performance in all of them. For the profit guarantee, an example might be a large firm utilizing multiple self-managed teams. The realized situation (captured by  $A$ ) may be apparent to the team, but too costly to communicate or verify. Hence, the firm resorts to designing a contract based only on the aspects common to all situations (captured by  $A^0$ ).

For team-optimal contracts, the principal corresponds to a "social planner" designing a robustly optimal budget balanced contract. While we are agnostic about the interpretation, perhaps the most natural one is to view this as a normative exercise. The multiteam interpretation could then have us looking, for instance, for a standardized contract for different kinds of partnerships. An example of such a contract is a profit-sharing agreement common in professional services.

### 3. NECESSITY OF INTEREST ALIGNMENT

We start the analysis by showing that for a contract to have a meaningful surplus or profit guarantee, it is essentially necessary for it to align the agents' interests in the following sense.

**DEFINITION 3.1:** A contract  $w$  *aligns the agents' interests* if for every pair of agents  $i$  and  $j$ , and every pair of output distributions  $F$  and  $G$  on  $Y$ ,  $\mathbb{E}_F[w_i(y)] > \mathbb{E}_G[w_i(y)]$  implies  $\mathbb{E}_F[w_j(y)] \geq \mathbb{E}_G[w_j(y)]$ .

That is, if some agent strictly prefers output distribution  $F$  to output distribution  $G$  under contract  $w$  (gross of costs), then so does any other agent (at least weakly).

---

<sup>5</sup>If asking the agents to report the technology were allowed, then with two or more agents it would be possible to partially implement the Bayesian profit-maximizing contract for the true technology by using a mechanism that chooses the Bayesian optimal contract for the reported technology whenever the agents' reports agree, and which "punishes" the agents with the zero contract if any reports disagree. With three or more agents, the Bayesian surplus-maximizing contract could be implemented similarly. But as is typical in the implementation literature, the two-agent case is more difficult because then it is not obvious to tell who deviated when reports disagree, and budget balance prevents punishing both agents simultaneously. As this issue is orthogonal to our analysis, we do not pursue it further.

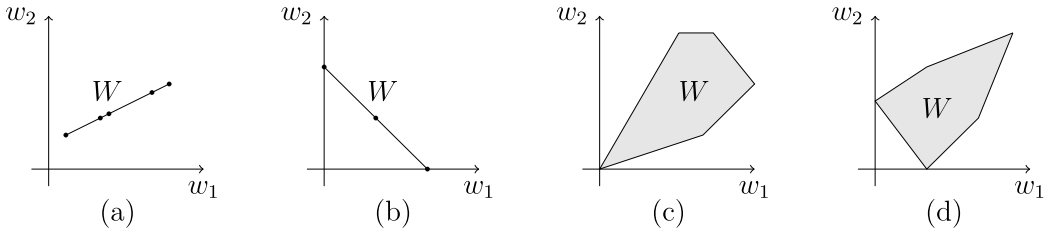


FIGURE 1.—The set  $W = \text{co}(w(Y))$  for  $I = 2$ . (a) A contract that aligns the agents’ interests. (b)–(d) Contracts that fail to align the agents’ interests.

A contract that does not satisfy Definition 3.1 is said to *fail to align the agents’ interests*. Any constant contract, such as the zero contract, aligns the agents’ interests. Note also that all contracts satisfy Definition 3.1 vacuously if the team consists of just one agent. We will use the following geometric characterization of interest alignment.

LEMMA 3.1: *The following four properties of a contract  $w$  are equivalent:*

- (i) *The contract  $w$  aligns the agents’ interests.*
- (ii) *For all pairs of agents  $i$  and  $j$ , the set  $w_{i,j}(Y) := \{(w_i(y), w_j(y)) : y \in Y\}$  is contained in a ray in  $\mathbb{R}_+^2$ , that is,  $w_{i,j}(Y) \subset \{\underline{w}_{i,j} + d_{i,j}t : t \in \mathbb{R}_+\}$  for some  $\underline{w}_{i,j}, d_{i,j} \in \mathbb{R}_+^2$ .*
- (iii) *All payment profiles are contained in a ray in  $\mathbb{R}_+^I$ , that is,  $w(Y) \subset \{\underline{w} + dt : t \in \mathbb{R}_+\}$  for some  $\underline{w}, d \in \mathbb{R}_+^I$ .*
- (iv) *There exist outputs  $\bar{y}$  and  $\underline{y}$  with  $w(\bar{y}) \geq w(\underline{y})$  such that, for every output  $y \in Y$ , we have  $w(y) = (1 - \lambda)w(\underline{y}) + \lambda w(\bar{y})$  for some  $\lambda \in [0, 1]$ .*

We provide a more detailed proof in the Appendix, but Lemma 3.1 follows essentially just by observing that the set of possible expected payment profiles under contract  $w$  is

$$W := \text{co}(w(Y)) = \{x \in \mathbb{R}_+^I : x = \mathbb{E}_F[w(y)] \text{ for some } F \in \Delta(Y)\}.$$

Therefore,  $w$  aligns the agents’ interests in the sense of Definition 3.1 if and only if for all  $x, x' \in W$ ,  $x_i > x'_i$  implies  $x_j \geq x'_j$  for all agents  $i$  and  $j$ , or equivalently, if  $W$  (and hence  $w(Y)$ ) is contained in a ray in  $\mathbb{R}_+^I$ . A contract that aligns the agents’ interests thus prescribes team-based compensation in a strong sense: the agents’ payments covary positively and linearly. Moreover, the parameter  $\lambda = \lambda(y)$  in Lemma 3.1(iv) can naturally be interpreted as measuring the team’s performance on a scale from zero to one.

Figure 1 depicts the set  $W$  for some contracts that do or do not align the agents’ interests in the case of two agents. Panel (b) corresponds to a tournament where the agents’ interests are diametrically opposed—the antithesis of interest alignment.

It is worth noting that Definition 3.1 only concerns the agents, and so it is silent on how the payments relate to the value of output  $v(y)$ . For example, while any contract where each  $w_i$  is linear in  $v(y)$  aligns the agents’ interests, so does a contract that pays a bonus  $b_i$  to all agents conditional on some output  $\hat{y}$  and that otherwise pays nothing. Thus, Definition 3.1 does not imply linearity in output value in general, but the next lemma shows that it does do so under budget balance.

LEMMA 3.2: *A contract  $w$  is budget balanced and aligns the agents’ interests if and only if there are shares  $(\alpha_1, \dots, \alpha_I) \in [0, 1]^I$  such that  $\sum_i \alpha_i = 1$  and  $w_i(y) = \alpha_i v(y)$  for all  $i$  and  $y$ .*

PROOF: Clearly, a contract of this form is budget balanced and aligns the agents' interests. For the converse, note that by budget balance we can take  $\bar{y} \in \arg \max_y v(y)$  and  $\underline{y} = y_0$  in Lemma 3.1(iv). Fixing  $y$ , we sum over  $i$  and use budget balance again to get  $v(y) = \sum_i w_i(y) = (1 - \lambda) \sum_i w_i(y_0) + \lambda \sum_i w_i(\bar{y}) = (1 - \lambda)v(y_0) + \lambda v(\bar{y}) = \lambda v(\bar{y})$ . Hence,  $\lambda = v(y)/v(\bar{y})$ . Noting that  $w_i(y_0) = 0$  by limited liability and budget balance, we thus have  $w_i(y) = (w_i(\bar{y})/v(\bar{y}))v(y)$ , so taking  $\alpha_i = w_i(\bar{y})/v(\bar{y})$  yields the result. Q.E.D.

Our first main result shows that any contract that fails to align the agents' interests can be easily improved upon regardless of whether we are interested in profits or surplus.

**THEOREM 3.1:** *If a contract  $w$  fails to align the agents' interests, then  $V(w) \leq V(0)$ . If, in addition,  $w$  is budget balanced, then  $S(w) \leq S(w')$  for every contract  $w'$  that is budget balanced and aligns the agents' interests.*

That is, the profit guarantee of a contract that fails to align the agents' interests is no better than that of the zero contract. And if the contract is also budget balanced, then its guaranteed expected surplus is weakly worse than the guarantee obtained by arbitrarily distributing shares to the agents. These results imply, inter alia, that we can restrict attention to contracts that align the agents' interests when searching for optimal ones.

Before turning to the proof, we note that Theorem 3.1 can be strengthened under a mild restriction on the known technology  $A^0$ . We say that an action profile  $a \in A^0$  satisfies *full support* if either  $F(a)$  has full support on  $Y$ , or  $F(a)$  equals  $\delta_{y_0}$ , the Dirac measure at the worst output  $y_0$ . We say that the action profile  $a$  satisfies *costly production* if  $\mathbb{E}_{F(a)}[v(y)] > 0$  implies  $c_i(a_i) > 0$  for some agent  $i$ . The next result shows that under either assumption, interest alignment is a necessary condition for a contract to improve on the trivial guarantee.

**THEOREM 3.2:** *Suppose that each action profile in the known technology satisfies full support or costly production. If a contract  $w$  fails to align the agents' interests, then  $V(w) \leq 0$  and  $S(w) \leq 0$ .*

We note for future reference that if  $w$  is budget balanced, then the necessity of interest alignment obtains under an even milder condition, requiring only that any action profile that yields the maximum output value with certainty not be costless to all agents.

**LEMMA 3.3:** *Suppose that for all  $a \in A^0$ ,  $\text{supp } F(a) \subseteq \arg \max_{y \in Y} v(y)$  implies  $c_i(a) > 0$  for some agent  $i$ . If a budget balanced contract  $w$  fails to align the agents' interests, then  $S(w) \leq 0$ .*

Of course, the above results are silent on whether contracts that do align the agents' interests actually improve on the trivial guarantees. We address this question in Sections 4 and 5, which consider, respectively, team-optimal and principal-optimal contracts.

The key idea behind Theorem 3.1 is that when a contract fails to align the agents' interests, it is possible to use the agents' disagreement over output distributions to add actions to the game so as to eliminate any equilibria in known actions and to drive the equilibrium output to  $y_0$ . The next example illustrates this in essentially the simplest non-trivial case: a rank-order tournament with two agents.

	$A_2^0$	$a'_2$
$A_1^0$	$\dots$	$-c_1(a_1^0), b$
$a'_1$	$b, -c_2(a_2^0)$	$b/2, b/2$

FIGURE 2.—The game  $\Gamma(w, A)$  for Example 3.1. To see that  $a'$  is the unique equilibrium, fix a mixed strategy equilibrium  $\sigma$ . If the support of  $\sigma$  is contained in  $A^0$ , then some agent  $i$ 's expected payoff is at most  $b/2$ , whereas deviating to  $a'_i$  yields  $b$  for sure. Hence,  $a'_i$  must be in the support of  $\sigma_i$  for some agent  $i$ . But then  $a'_{-i}$  is the unique best response for agent  $-i$ , and thus  $\sigma_{-i}(a'_{-i}) = 1$ . This in turn implies  $\sigma_i(a'_i) = 1$ . Therefore,  $\sigma$  is the pure-strategy profile  $a'$ .

EXAMPLE 3.1: Let  $I = 2$ . An output is a pair  $(y_1, y_2) \in Y \subset \mathbb{R}_+^2$ , with  $Y = \{0, 1, \dots, \bar{y}\}^2$  for some integer  $\bar{y} > 0$ . Let  $v(y) = y_1 + y_2$  so that the worst output is  $y_0 = (0, 0)$ . The known technology  $A^0$  can be arbitrary. Say, one could assume that any  $a_i^0 \in A_i^0$  only affects the distribution of  $y_i$ .

A tournament is a contract  $w$  that gives a prize  $b$  to the agent who produces the most, and splits the prize equally in case of a tie. That is,  $w_i(y) = b > 0$  if  $y_i > y_{-i}$ ,  $w_i(y) = b/2$  if  $y_i = y_{-i}$ , and  $w_i(y) = 0$  if  $y_i < y_{-i}$ . See Figure 1(b). As is well known, this motivates agent  $i$  to not only increase  $y_i$ , but also to reduce  $y_{-i}$  via refusing help, stealing, or sabotage (e.g., Lazear (1989)). Formally, consider a technology  $A \supset A^0$  such that  $A_i := A_i^0 \cup \{a'_i\}$  with  $c_i(a'_i) = 0$  for  $i = 1, 2$ . Let  $F(a'_1, a_2^0) = \delta_{(1,0)}$  and  $F(a_1^0, a'_2) = \delta_{(0,1)}$  for all  $a_i^0 \in A_i^0$ ,  $i = 1, 2$ . Then playing the new action  $a'_i$  wins the tournament for sure against any  $a_{-i}^0 \in A_{-i}^0$ . Finally, let  $F(a') = \delta_{(0,0)}$  so that nothing is produced if both agents play the new action.

Figure 2 shows that  $a'$  is the unique equilibrium of the game  $\Gamma(w, A)$  induced by the technology  $A$ . Thus,  $V(w, A) = v(0, 0) - b = -b$  and  $V(w) \leq V(w, A) = -b < 0 \leq V(0)$ , that is, the zero contract has a better guarantee than the tournament, consistent with Theorem 3.1.

### 3.1. Proof of Theorems 3.1 and 3.2

The main part of the proof is showing that for essentially any contract that fails to align the agents' interests, one can find a game whose unique equilibrium output distribution assigns arbitrarily high probability to the worst output  $y_0$ . This is Lemma 3.5 below, which generalizes Example 3.1 by using a more elaborate construction than the one used for the tournament. However, there is a “nuisance case” that has to be treated separately, and we do so first in Lemma 3.4. The theorems are then shown to follow from these two lemmas.

Given a contract  $w$ , let

$$Y^* := \bigcap_{i=1}^I \arg \max_{y \in Y} w_i(y).$$

By definition, any  $y \in Y^*$  simultaneously maximizes the payment to every agent. The set  $Y^*$  may be nonempty even if  $w$  fails to align the agents' interests (see Figure 1(d), where  $Y^*$  consists of the outputs that map to the tip of the pointy end of the set  $W$ ). The nuisance case arises if the agents can ensure that the output is in  $Y^*$  at zero cost to all agents as there is then in general no way to drive the output to  $y_0$ . The next lemma shows that such contracts are nevertheless uninteresting as they perform worse than the zero contract or any budget balanced contract that aligns interests, and that under the assumptions of Theorem 3.2 they have at most a zero profit guarantee and cannot satisfy budget balance.



LEMMA 3.4: *Let  $w$  be a contract different from the zero contract. Suppose there exists  $a^* \in A^0$  such that  $\text{supp } F(a^*) \subseteq Y^*$  and  $c(a^*) = 0$ . Then the following properties hold:*

- (i)  $V(w) < V(0)$ .
- (ii) *If  $w$  fails to align the agents' interests and if each action profile in the known technology has full support or costly production, then  $V(w) < 0$ .*
- (iii) *If  $w$  is budget balanced, then  $S(w) \leq S(w')$  for every contract  $w'$  that is budget balanced and aligns the agents' interests.*
- (iv) *If each action profile in the known technology has full support or costly production, then  $w$  is not budget balanced.*

We relegate the proof to the Appendix, but the idea is straightforward: As the action profile  $a^*$  in Lemma 3.4 gives the agents their maximum payments under contract  $w$  at zero cost, it is an equilibrium given any technology  $A \supseteq A^0$ . This may potentially yield a positive profit or surplus guarantee. However, as  $c(a^*) = 0$ , the agents would be happy to play  $a^*$  also under the zero contract, which of course would give even more profit. This observation underlies part (i) of Lemma 3.4. The assumptions in part (ii) strengthen the conclusion from  $V(w) < V(0)$  to  $V(w) < 0$  as they imply that the value of output given  $a^*$  is zero. Part (iii) follows by observing that under budget balance, the outputs in  $Y^*$  must maximize the value of output. Thus,  $a^*$  yields  $\max v(Y)$  at zero cost, and hence it is a surplus-maximizing equilibrium under any contract that pays a fixed share to each agent, no matter what other actions are available. Finally, part (iv) follows as generating  $\max v(Y)$  with certainty at zero cost is inconsistent with full support or costly production.

With Lemma 3.4 out of the way, the proof comes down to the following result.

LEMMA 3.5: *Let  $w$  be a contract that fails to align the agents' interests. Suppose that for all  $a \in A^0$ ,  $\text{supp } F(a) \subseteq Y^*$  implies  $c(a) \neq 0$ . Then there exists a sequence of technologies  $A^n \supseteq A^0$ , each with a unique equilibrium output distribution  $F^n \in \Delta(Y)$ , such that  $F^n \rightarrow \delta_{y_0}$ .*

Before turning to the proof, let us verify that Theorems 3.1 and 3.2 indeed follow from the preceding two results. Observe first that under the assumptions of Lemma 3.5, we have  $V(w) \leq V(w, A^n) \leq \mathbb{E}_{F^n}[v(y)] \rightarrow 0$  and  $S(w) \leq S(w, A^n) \leq \mathbb{E}_{F^n}[v(y)] \rightarrow 0$ . Therefore, Theorem 3.1 follows by noting that any contract that fails to align the agents' interests is covered by either parts (i) and (iii) of Lemma 3.4, or by Lemma 3.5. As for Theorem 3.2, the assumption about  $A^0$  matters only in the case covered by Lemma 3.4, where part (ii) then gives  $V(w) < 0$ , and part (iv) shows that that case is then impossible under budget balance. Hence, we have  $V(w) \leq 0$  and  $S(w) \leq 0$ .

It remains to establish Lemma 3.5. Throughout the proof, we fix a contract  $w$  that fails to align the agents' interests, and assume that, for all  $a \in A^0$ ,  $\text{supp } F(a) \subseteq Y^*$  implies  $c(a) \neq 0$ . Observe that  $w$  fails Lemma 3.1(ii) for some pair of agents. Relabeling if necessary, we assume without loss of generality that this is agents 1 and 2.

We proceed in two steps. First, we construct a preliminary technology that eliminates equilibria in known actions. We then amend it to drive the equilibrium output to  $y_0$ .<sup>6</sup>

---

<sup>6</sup>The proof has a “divide and conquer” flavor, making it similar in spirit to [Abreu and Matsushima \(1992\)](#). However, as we construct the worst-case game separately for each contract, the designer of the game knows the agents' preferences here. Hence, viewed as an implementation problem, ours is nontrivial only because of the existence of the known actions, which have no counterpart in the implementation literature. Thus, even setting aside the different solution concepts, Abreu and Matsushima's construction cannot be applied in our setting, nor can ours be applied in theirs.

*Preliminary Technology  $A^*$ .* Define the technology  $A^* \supset A^0$  as follows. For each agent  $i$ , let  $A_i^* := A_i^0 \cup \{a_i^*\}$  and  $c_i(a_i^*) = 0$ . Note that  $a_i^*$  is a least-cost action for agent  $i$ . It will be helpful to think of  $a_i^*$  as an action that allows agent  $i$  to “veto” any outcome that could arise under the known technology.

We define output distributions for action profiles in  $A^* \setminus A^0$  by specifying first the corresponding expected payment profiles with the help of the following lemma.

LEMMA 3.6: *There exist (not necessarily all distinct) points  $z^1, \dots, z^I$  in  $W$  such that*

$$\sum_i z_i^i > \sum_i \mathbb{E}_{F(a)}[w_i(y) - c_i(a_i)] \quad \text{for all } a \in A^0, \tag{3.1}$$

$$z_j^j > z_j^i \quad \text{for all } j \in \{1, 2\}, i \in \{1, \dots, I\} \text{ with } i \neq j, \text{ and} \tag{3.2}$$

$$z_j^j \geq z_j^i \quad \text{for all } i, j \in \{3, \dots, I\}. \tag{3.3}$$

That is, each agent  $j$  prefers  $z^j$  to any other point  $z^i$  (strictly so if  $j \in \{1, 2\}$ ), but it is infeasible for all agents to simultaneously get their preferred payoff  $z_j^j$  under the known technology.

To sketch the proof of Lemma 3.6, consider the case of two agents so that condition (3.3) holds vacuously. Let  $z^1$  and  $z^2$  be agent 1’s and agent 2’s favorite points in  $W$ . On one hand, if  $z^1$  and  $z^2$  are distinct as in Figures 1(b) and (c), then (3.1) and (3.2) are clearly satisfied. On the other hand, if  $z^1 = z^2$ , then  $Y^* \neq \emptyset$  and  $W$  must have a nonempty interior so that it resembles Figure 1(d). We can then choose  $z^1$  and  $z^2$  in the interior of  $W$  to satisfy (3.2). Moreover, since any  $a$  with  $\text{supp } F(a) \subseteq Y^*$  has  $c_i(a_i) > 0$  for some  $i$  by assumption, choosing  $z^1$  and  $z^2$  close enough to the pointy end will also satisfy (3.1). The proof in the Appendix provides the details and also covers the case of more than two agents.

Now fix points  $z^1, \dots, z^I$  in  $W$  satisfying (3.1)–(3.3). For each  $a \in A^* \setminus A^0$ , let

$$x(a) := \begin{cases} z^j & \text{if } (a_1, a_2) \neq (a_1^*, a_2^*) \text{ and } j = \min\{i : a_i = a_i^*\}, \\ x^* := \frac{1}{2}z^1 + \frac{1}{2}z^2 & \text{if } (a_1, a_2) = (a_1^*, a_2^*), \end{cases} \tag{3.4}$$

and let the corresponding output distribution be any distribution  $F(a) \in \Delta(Y)$  with  $\mathbb{E}_{F(a)}[w(y)] = x(a)$ . This completes the description of the technology  $A^*$ .

Continuing with the interpretation of  $a_i^*$  as a veto action, (3.4) says that any agent  $j$  can veto the play and force the payment profile  $z^j$  if all other agents play known actions. If multiple agents veto, then the tie is broken in favor of the agent with the lowest index, except when both agents 1 and 2 veto, in which case  $x^*$ , the average of  $z^1$  and  $z^2$ , is chosen.

The next lemma lists some key properties that any  $A \supseteq A^*$  inherits from  $A^*$ .

LEMMA 3.7: *Every technology  $A \supseteq A^*$  satisfies the following properties:*

- (i) *If  $\sigma$  is a mixed strategy profile with  $\text{supp } \sigma \subseteq A^0$ , then there exists an agent  $i$  for whom  $u_i(a_i^*, \sigma_{-i}; w, A) > u_i(\sigma; w, A)$ .*
- (ii)  *$u_i(a_i^*, a_{-i}; w, A) \geq u_i(a_i^0, a_{-i}; w, A)$  for all  $i$ , all  $a_i^0 \in A_i^0$ , and all  $a_{-i} \in A_{-i}^* \setminus A_{-i}^0$ .*
- (iii) *The inequality in part (ii) is strict for  $i = 1, 2$ .*

Part (i) rules out equilibria in known actions. It follows because, by (3.4), any agent  $i$  who unilaterally deviates from  $\sigma$  to  $a_i^*$  earns  $z_i^i$  (as  $c_i(a_i^*) = 0$ ), whereas the sum of the

agents' expected payments under  $\sigma$  is less than  $\sum_i z_i^i$  by (3.1). Part (ii) shows that  $a_i^*$  weakly dominates any known action  $a_i^0$  if at least one other agent plays the veto action, and part (iii) shows that this dominance is strict for agents 1 and 2. Both claims follow by inspection of (3.4) given our choice of  $z^1, \dots, z^l$ ; see the proof in the Appendix for the details.

The ‘‘tie-breaking’’ built into (3.4) in case multiple agents play  $a_i^*$  favors agents 1 and 2. This can be shown to imply that in any equilibrium of  $\Gamma(w, A^*)$ , agents 1 and 2 play  $(a_1^*, a_2^*)$ . More generally, for all suitably chosen  $A \supseteq A^*$ , at least one of them forgoes all actions in  $A^0$ :

LEMMA 3.8: *Suppose  $A \supseteq A^*$  is a technology where  $A_i = A_i^*$  for all  $i > 2$  and where*

$$u_i(a_i^*, a_{-i}; w, A) \geq u_i(a_i^0, a_{-i}; w, A) \quad \text{for all } i, \text{ all } a_i^0 \in A_i^0, \text{ and all } a_{-i} \in A_{-i} \setminus A_{-i}^*. \quad (3.5)$$

Then  $\sigma_1(A_1^0)\sigma_2(A_2^0) = 0$  for every  $\sigma \in \mathcal{E}(w, A)$ .

The proof in the Appendix uses Lemma 3.7(i) and the weak dominance conditions in Lemma 3.7(ii) and (3.5) to first show that if all agents assign positive probability to known actions, then  $a_i^*$  strictly dominates all  $a_i^0 \in A_i^0$  for some agent  $i$ , which contradicts agent  $i$  assigning positive probability to  $A_i^0$ . Thus, some set of agents must play known actions with probability zero. We then use Lemma 3.7(iii) to show that this set contains agent 1 or 2.

We are now in a position to prove Lemma 3.5. There are two cases to consider depending on how the profile  $x^*$  defined in (3.4) is located relative to  $w(y_0)$ .

*Case 1:  $w_j(y_0) \geq x_j^*$  for some  $j \in \{1, 2\}$ .* Without loss of generality, assume  $w_1(y_0) \geq x_1^*$ . To ensure strict incentives, fix  $\varepsilon \in (0, 1)$  and take  $F_\varepsilon \in \Delta(Y)$  such that  $F_\varepsilon(y_0) > 1 - \varepsilon$  and  $\mathbb{E}_{F_\varepsilon}[w_1(y)] > x_1^*$ . This is feasible for any  $\varepsilon > 0$  as  $x_1^* = \frac{1}{2}z_1^1 + \frac{1}{2}z_1^2 < z_1^1 \leq \max w_1(Y)$  by (3.2).

Define the technology  $\bar{A} \supset A^* \supset A^0$  by letting  $\bar{A}_1 := A_1^* \cup \{\bar{a}_1\}$  with  $c_1(\bar{a}_1) = 0$ , and letting  $\bar{A}_i := A_i^*$  for  $i \neq 1$ . Let  $F(a) := F_\varepsilon$  for all  $a \in \bar{A}$  such that  $a_1 = \bar{a}_1$ . That is, agent 1 is the only one who has an additional action  $\bar{a}_1$  beyond the actions in the preliminary technology  $A^*$ . By playing  $\bar{a}_1$ , agent 1 can unilaterally force the distribution  $F_\varepsilon$ , and he will do so in every equilibrium:

LEMMA 3.9: *If  $\sigma$  is an equilibrium of  $\Gamma(w, \bar{A})$ , then  $\sigma_1(\bar{a}_1) = 1$ .*

PROOF: We verify first that technology  $\bar{A}$  satisfies (3.5). It suffices to consider  $i \neq 1$  as  $\bar{A}_{-1} \setminus A_{-1}^* = \emptyset$ . Note that if  $i \neq 1$ , then every  $a_{-i} \in \bar{A}_{-i} \setminus A_{-i}^*$  has  $a_1 = \bar{a}_1$ . Thus, for any such  $a_{-i}$ , we have  $u_i(a_i^*, a_{-i}; w, \bar{A}) - u_i(a_i^0, a_{-i}; w, \bar{A}) = c_i(a_i^0) \geq 0$  as agent  $i$  cannot affect the output if agent 1 plays  $\bar{a}_1$ . Thus, Lemma 3.8 applies, and we have  $\sigma_1(A_1^0)\sigma_2(A_2^0) = 0$  for all  $\sigma \in \mathcal{E}(w, \bar{A})$ .

Suppose first that  $\sigma_2(A_2^0) = 0$  so that agent 2 plays  $a_2^*$  with probability 1. We then have  $u_1(\bar{a}_1, a_2^*, a_{-(1,2)}; w, \bar{A}) = \mathbb{E}_{F_\varepsilon}[w_1(y)] > x_1^* = u_1(a_1^*, a_2^*, a_{-(1,2)}; w, \bar{A}) > u_1(a_1^0, a_2^*, a_{-(1,2)}; w, \bar{A})$  for all  $a_1^0 \in A_1^0$  and all  $a_{-(1,2)} \in \bar{A}_{-(1,2)}$ , where the last inequality is by Lemma 3.7(iii). Therefore,  $\bar{a}_1$  is agent 1's unique best response and  $\sigma_1(\bar{a}_1) = 1$  as desired.

Suppose then that  $\sigma_1(A_1^0) = 0$ , but  $\sigma_1(\bar{a}_1) < 1$ . Then  $\sigma_1(a_1^*) = 1 - \sigma_1(\bar{a}_1) > 0$  and we have  $u_2(a_2^*, \sigma_{-2}; w, \bar{A}) - u_2(a_2^0, \sigma_{-2}; w, \bar{A}) = \sigma_1(\bar{a}_1)c_2(a_2^0) + \sigma_1(a_1^*)(x_2^* - z_2^1 + c_2(a_2^0)) > 0$  for all  $a_2^0 \in A_2^0$ , where the strict inequality follows as  $x_2^* - z_2^1 = \frac{1}{2}(z_2^2 - z_2^1) > 0$  by (3.2). This

implies  $\sigma_2(A_2^0) = 0$ , which was shown above to imply  $\sigma_1(\bar{a}_1) = 1$ , contradicting  $\sigma_1(\bar{a}_1) < 1$ . We conclude that  $\sigma_1(\bar{a}_1) = 1$  as desired. *Q.E.D.*

Lemma 3.9 implies that the unique equilibrium output distribution in  $\Gamma(w, \bar{A})$  is  $F_\varepsilon$ , which converges to  $\delta_{y_0}$  as  $\varepsilon \rightarrow 0$ . This establishes the claim in Lemma 3.5 for Case 1.

*Case 2:*  $w_j(y_0) < x_j^*$  for all  $j \in \{1, 2\}$ . In this case,  $W_{1,2} := \text{co}(w_{1,2}(Y))$  has a nonempty interior relative to  $\mathbb{R}^2$ , denoted  $\text{int}(W_{1,2})$ . (Otherwise, as  $w$  fails Lemma 3.1(ii) for agents 1 and 2 by assumption,  $W_{1,2}$  must be a strictly decreasing line segment. But then  $w_1(y_0) < x_1^*$  implies  $w_2(y_0) > x_2^*$ .) Given any  $x \in \mathbb{R}_+^2$ , write  $x_{1,2} := (x_1, x_2) \in \mathbb{R}_+^2$ .

We construct a technology  $\hat{A} \supset A^* \supset A^0$  as follows. Let  $\hat{A}_i := A_i^*$  for each agent  $i > 2$ . Let  $\hat{A}_1 := A_1^* \cup \{a_1^1, a_1^3, \dots, a_1^{K-1}\}$ , where  $K \geq 2$  is an even number to be specified below, and let  $\hat{A}_2 := A_2^* \cup \{a_2^2, a_2^4, \dots, a_2^K\}$ . Let  $c_i(a_i^k) = 0$  for  $i \in \{1, 2\}$  and  $1 \leq k \leq K$ . That is, only agents 1 and 2 have additional actions beyond those in the preliminary technology  $A^*$ , and hence most of the arguments that follow will only involve the two of them.

We will next specify expected payment profiles in  $W$ , which will be used to define the output distributions associated with action profiles involving new actions. Figures 3 and 4 illustrate schematically the projections of these profiles to agents 1 and 2, and how they are assigned to their actions; it may be helpful to refer to them along the way.

By perturbing it if necessary, we can assume that  $x_{1,2}^* \in \text{int}(W_{1,2})$ .<sup>7</sup> We approximate  $w(y_0) \in W$  by fixing  $\varepsilon > 0$  and taking  $F_\varepsilon \in \Delta(Y)$  to be a distribution with full support on  $Y$  such that  $F_\varepsilon(y_0) > 1 - \varepsilon$ . Then  $w_{1,2}^\varepsilon(y_0) := \mathbb{E}_{F_\varepsilon}[w_{1,2}(y)] \in \text{int}(W_{1,2})$ , and for  $\varepsilon$  small

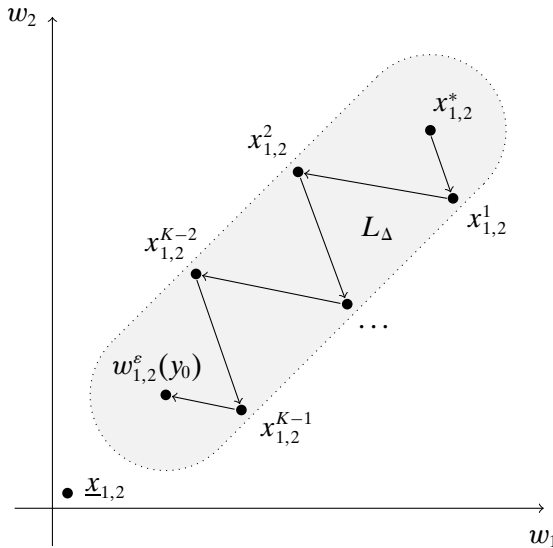


FIGURE 3.—Construction of the sequence  $(x_{1,2}^0 = x_{1,2}^*, x_{1,2}^1, \dots, x_{1,2}^{K-1}, x_{1,2}^K = w_{1,2}^\varepsilon(y_0))$  in the proof of Lemma 3.5. (Here,  $K = 6$ .) The shaded area is  $L_\Delta$ .

<sup>7</sup>Conditions (3.1) and (3.2) involve finitely many strict inequalities. Thus, when  $\text{int}(W_{1,2})$  is nonempty, we can perturb  $z^1$  and  $z^2$  so that  $z_{1,2}^1$  and  $z_{1,2}^2$  lie in  $\text{int}(W_{1,2})$ . Then  $x_{1,2}^* = \frac{1}{2}z_{1,2}^1 + \frac{1}{2}z_{1,2}^2 \in \text{int}(W_{1,2})$  by convexity of  $W$ . Moreover, for small enough perturbations, we continue to have  $w_j(y_0) < x_j^*$  for all  $j \in \{1, 2\}$ .

	$A_2^0$	$a_2^*$	$a_2^2$	$a_2^4$	$\dots$	$a_2^{K-2}$	$a_2^K$
$A_1^0$	$\dots$	$z^2$	$\eta^0$	$\eta^0$	$\dots$	$\eta^0$	$\eta^0$
$a_1^*$	$z^1$	$\mathbf{x}^0$	$\eta^1$	$\eta^1$	$\dots$	$\eta^1$	$\eta^1$
$a_1^1$	$\eta^0$	$\mathbf{x}^1$	$\mathbf{x}^2$	$\eta^2$	$\dots$	$\eta^2$	$\eta^2$
$a_1^3$	$\eta^0$	$\eta^1$	$\mathbf{x}^3$	$\mathbf{x}^4$	$\dots$	$\eta^3$	$\eta^3$
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\ddots$	$\dots$	$\dots$
$a_1^{K-3}$	$\eta^0$	$\eta^1$	$\eta^2$	$\eta^3$	$\dots$	$\mathbf{x}^{K-2}$	$\eta^{K/2}$
$a_1^{K-1}$	$\eta^0$	$\eta^1$	$\eta^2$	$\eta^3$	$\dots$	$\mathbf{x}^{K-1}$	$\mathbf{x}^K$

FIGURE 4.—Assignment of expected payment profiles to action profiles  $a \in A$  such that  $a_1 \notin A_1^0$  or  $a_2 \notin A_2^0$ . Note that for such  $a$ , the expected payment profile depends only on  $a_1$  and  $a_2$ . Furthermore, we have  $c_i(a_i^*) = c_i(a_i^k) = 0$  for  $i \in \{1, 2\}$  and  $1 \leq k \leq K$ . Thus, this matrix also directly shows  $u_1(a_1, a_2, a_{-\{1,2\}}; w, A)$  for  $a_1 \notin A_1^0$  and  $u_2(a_1, a_2, a_{-\{1,2\}}; w, A)$  for  $a_2 \notin A_2^0$ .

enough, we have  $x_{1,2}^* > w_{1,2}^\varepsilon(y_0)$ , which we assume to be the case henceforth. Moreover, we can then fix  $\underline{x} \in W$  such that  $\underline{x}_{1,2} \in \text{int}(W_{1,2})$  and  $\underline{x}_{1,2} < w_{1,2}^\varepsilon(y_0)$ .

Let  $L$  be the line segment connecting  $x_{1,2}^*$  and  $w_{1,2}^\varepsilon(y_0)$  in  $\mathbb{R}_+^2$ . Given  $\Delta > 0$ , let  $L_\Delta$  be the  $\Delta$ -neighborhood of  $L$ , that is,  $L_\Delta := \{x_{1,2} \in \mathbb{R}_+^2 : \|x_{1,2} - z_{1,2}\| < \Delta \text{ for some } z_{1,2} \in L\}$ . Because  $x_{1,2}^* > w_{1,2}^\varepsilon(y_0) > \underline{x}_{1,2}$  and all three points lie in the convex open set  $\text{int}(W_{1,2})$ , we can take  $\Delta > 0$  small enough so that  $L_\Delta \subseteq \text{int}(W_{1,2})$  and  $\underline{x}_{1,2} < z_{1,2}$  for all  $z_{1,2} \in L_\Delta$ . We then choose the even number  $K \geq 2$  and points  $x_{1,2}^1, \dots, x_{1,2}^{K-1}$  in  $L_\Delta$  such that the sequence  $(x_{1,2}^0 = x_{1,2}^*, x_{1,2}^1, \dots, x_{1,2}^{K-1}, x_{1,2}^K = w_{1,2}^\varepsilon(y_0))$  satisfies the following conditions:

1. Among any two consecutive points, agent 1 prefers the odd one: for all  $k$  odd,

$$x_1^k > x_1^{k+1} \quad \text{and} \quad x_1^k > x_1^{k-1}.$$

2. Agent 2 has the opposite preference: for all  $k$  odd,

$$x_2^k < x_2^{k+1} \quad \text{and} \quad x_2^k < x_2^{k-1}.$$

See Figure 3 for an illustration. We map each point  $x_{1,2}^k, 0 \leq k \leq K$ , to a point  $x^k$  in  $W$  by letting  $x^0 = x^*$  and  $x^K = w^\varepsilon(y_0) := \mathbb{E}_{F_\varepsilon}[w(y)]$ , and by taking  $x^k$  for  $k \notin \{0, K\}$  to be any point in  $W$  whose image in  $W_{1,2}$  is  $x_{1,2}^k$ .

We will need another sequence  $(\eta^0, \dots, \eta^{K/2})$  in  $W$  with  $\eta_{1,2}^0 < \dots < \eta_{1,2}^{K/2} < \underline{x}_{1,2}$ . Such a sequence can be found, because  $\underline{x}_{1,2} \in \text{int}(W_{1,2})$ . Note that we have  $\eta_{1,2}^l < \underline{x}_{1,2} < x_{1,2}^k$  for all  $k$  and  $l$ , because each  $x_{1,2}^k$  was chosen from  $L_\Delta$ , which dominates  $\underline{x}_{1,2}$ .

To complete the description of technology  $\hat{A}$ , we assign the above expected payment profiles to action profiles in  $\hat{A} \setminus A^*$  according to Figure 4. Finally, we assume that the profile  $x^K = w^\varepsilon(y_0)$  is generated by  $F_\varepsilon$ , that is,  $F(a_1^{K-1}, a_2^K, a_{-\{1,2\}}) = F_\varepsilon$  for all  $a_{-\{1,2\}} \in \hat{A}_{-\{1,2\}}$ . For any other expected payment profile, the corresponding output distribution can be taken to be any  $F \in \Delta(Y)$  that generates it. We will write  $u_i(a) := u_i(a; w, \hat{A})$  as  $(w, \hat{A})$  are held fixed in what follows.

Let us verify that the technology  $\hat{A}$  so defined satisfies the assumptions of Lemma 3.8.

LEMMA 3.10: *Technology  $\hat{A}$  satisfies (3.5) and the inequality in it is strict for  $i = 1, 2$ .*

PROOF: For agent 1, we have  $u_1(a_1^*, a_{-1}) - u_1(a_1^0, a_{-1}) = \eta_1^1 - \eta_1^0 + c_1(a_1^0) > 0$  for all  $a_{-1} \in \hat{A}_{-1} \setminus A_{-1}^*$ , since any such  $a_{-1}$  has  $a_2 \in \{a_2^2, \dots, a_2^K\}$ , in which case we can read agent 1's payoff from the top two rows in Figure 4 up to the cost of action  $a_1^0$ , and since  $\eta^1 > \eta^0$  by construction.

Agent 2 is treated analogously. The only difference arises if  $a_{-2} \in \hat{A}_{-2} \setminus A_{-2}^*$  contains  $a_1^1$ . Then  $u_2(a_2^*, a_{-2}) - u_2(a_2^0, a_{-2}) = x_2^1 - \eta_2^0 + c_2(a_2^0) > 0$ , since  $x_2^1 > \eta_2^0$  by construction.

If  $i > 2$ , then  $u_i(a_i^*, a_{-i}) - u_i(a_i^0, a_{-i}) = c_i(a_i^0) \geq 0$  for all  $a_i^0 \in A_i^0$  and  $a_{-i} \in \hat{A}_{-i} \setminus A_{-i}^*$ , since any such  $a_{-i}$  has  $a_1 \in \hat{A}_1 \setminus A_1^*$  or  $a_2 \in \hat{A}_2 \setminus A_2^*$ , implying that agent  $i$  cannot affect  $y$ .  
*Q.E.D.*

LEMMA 3.11: *If  $\sigma$  is an equilibrium of  $\Gamma(w, \hat{A})$ , then  $\sigma_1(a_1^{K-1}) = 1$  and  $\sigma_2(a_2^K) = 1$ .*

PROOF: Let  $\sigma \in \mathcal{E}(w, \hat{A})$ . We show first that  $\sigma_1(A_1^0) = \sigma_2(A_2^0) = 0$ . By Lemma 3.8, we have  $\sigma_1(A_1^0)\sigma_2(A_2^0) = 0$ . Suppose  $\sigma_2(A_2^0) = 0$  so that  $\text{supp } \sigma_{-1} \subseteq \hat{A}_{-1} \setminus A_{-1}^0$ . We claim that then  $u_1(a_1^*, \sigma_{-1}) > u_1(a_1^0, \sigma_{-1})$  for all  $a_1^0 \in A_1^0$ . To see this, note that the inequality  $u_1(a_1^*, a_{-1}) - u_1(a_1^0, a_{-1}) > 0$  holds for all  $a_{-1} \in A_{-1}^* \setminus A_{-1}^0$  by Lemma 3.7(iii) and for all  $a_{-1} \in \hat{A}_{-1} \setminus A_{-1}^*$  by Lemma 3.10. Thus, it holds for all  $a_{-1} \in \text{supp } \sigma_{-1}$ . This implies  $\sigma_1(A_1^0) = 0$ . The same argument shows that  $\sigma_1(A_1^0) = 0$  implies  $\sigma_2(A_2^0) = 0$ . Hence,  $\sigma_1(A_1^0) = \sigma_2(A_2^0) = 0$  as desired.

Because  $\sigma_1(A_1^0) = \sigma_2(A_2^0) = 0$ , we can eliminate row  $A_1^0$  and column  $A_2^0$  in Figure 4. The remaining matrix is solvable by iterated elimination of strictly dominated strategies:  $a_1^1$  dominates  $a_1^*$  since  $x_1^1 > x_1^0$ ,  $x_1^2 > \eta_1^1$ , and  $\eta_1^2 > \eta_1^1$  by construction. And once  $a_1^*$  is eliminated,  $a_2^2$  dominates  $a_2^*$  since  $x_2^1 < x_2^2$ ,  $\eta_2^1 < x_2^3$ , and  $\eta_2^1 < \eta_2^2$ .

To complete the argument, suppose we have already eliminated all rows and columns except for rows  $a_1^k, \dots, a_1^{K-1}$  and columns  $a_2^{k+1}, \dots, a_2^K$  for some  $1 \leq k \leq K - 3$  odd. Then  $a_1^{k+2}$  dominates  $a_1^k$  because  $x_1^{k+2} > x_1^{k+1}$  for  $k$  odd, and because  $x_1^{k+3} > \eta_1^{(k+3)/2}$  and  $\eta_1^{(k+5)/2} > \eta_1^{(k+3)/2}$  (if  $k < K - 3$ ). Thus, we can eliminate row  $a_1^k$ . But then  $a_2^{k+3}$  dominates  $a_2^{k+1}$  because  $x_2^{k+3} > x_2^{k+2}$  for  $k$  odd, and because  $x_2^{k+4} > \eta_2^{(k+3)/2}$  and  $\eta_2^{(k+5)/2} > \eta_2^{(k+3)/2}$  (if  $k < K - 3$ ). Thus, we can eliminate column  $a_2^{k+1}$ . Therefore, only row  $a_1^{K-1}$  and column  $a_2^K$  survive, implying that  $\sigma_1(a_1^{K-1}) = 1$  and  $\sigma_2(a_2^K) = 1$  as desired.  
*Q.E.D.*

By Lemma 3.11, the unique equilibrium output distribution in  $\Gamma(w, \hat{A})$  is  $F_\varepsilon$ , which converges to  $\delta_{y_0}$  as  $\varepsilon \rightarrow 0$ . This establishes Lemma 3.5 for Case 2, completing the proof.

#### 4. TEAM-OPTIMAL CONTRACTS

We now turn to team-optimal contracts, which maximize the surplus guarantee  $S(w)$  subject to budget balance. The following result collects our main findings. For the statement, say that a contract  $w$  is *linear (in value of output)* if  $w_i = \alpha_i v$  for some  $\alpha_i \in [0, 1]$  for each agent  $i$ . Under budget balance, this is equivalent to  $w$  aligning the agents' interests by Lemma 3.2.

THEOREM 4.1:

- (i) *There exists a linear team-optimal contract.*

(ii) *A team-optimal contract  $w$  guarantees positive expected surplus (i.e.,  $S(w) > 0$ ) if and only if the known technology  $A^0$  satisfies*

$$\max_{a \in A^0} \left( \mathbb{E}_{F(a)}[v(y)] - \sum_i c_i(a_i) - 2 \sum_{i,j:i \neq j} \sqrt{c_i(a_i)c_j(a_j)} \right) > 0. \tag{4.1}$$

(iii) *A linear team-optimal contract can be found by taking any profile  $a \in A^0$  that attains the maximum in (4.1) and defining each agent’s share by*

$$\alpha_i = \frac{\sqrt{c_i(a_i)}}{\sum_{j=1}^I \sqrt{c_j(a_j)}} \quad (\text{with } 0/0 = 1/I \text{ by convention}). \tag{4.2}$$

(iv) *Suppose there does not exist  $a \in A^0$  such that  $c(a) = 0$  and  $\mathbb{E}_{F(a)}[v(y)] = \max v(Y)$ . Then every team-optimal contract that guarantees positive expected surplus is linear.*

Part (i) of Theorem 4.1 shows that profit-sharing, or equity, is a robustly optimal contract for a team absent a sink or a source of funds. By part (ii), such a contract has a positive surplus guarantee if and only if the known technology is sufficiently productive as the maximand in (4.1) is the expected surplus minus an extra cost term. Part (iii) provides a formula for the optimal shares. Finally, part (iv) gives a sufficient condition for all team-optimal contracts to be linear, which is weak enough to hold in most cases of interest.

Theorem 3.1 implies that in establishing the above results, it is enough to consider budget balanced contracts that align the agents’ interests, or equivalently, budget balanced linear contracts. We start by deriving a formula for the surplus guarantee  $S(w)$  for such contracts.

Fix a budget balanced linear contract  $w$  with associated shares  $(\alpha_1, \dots, \alpha_I)$  for the agents. Given any technology  $A \supseteq A^0$ , let  $P$  be the mapping  $A \rightarrow \mathbb{R} \cup \{-\infty\}$  defined by

$$P(a) := \mathbb{E}_{F(a)}[v(y)] - \sum_i \frac{c_i(a_i)}{\alpha_i}, \tag{4.3}$$

where  $0/0 = 0$  and  $x/0 = \infty$  for  $x > 0$  by convention. Let  $U^0(w) := \max_{a \in A^0} P(a)$ . Note that  $U^0(w)$  depends only on  $A^0$ , and it is well-defined and nonnegative even if  $\alpha_i = 0$  for some agent  $i$  because each  $A_i^0$  contains a zero-cost action by assumption.

LEMMA 4.1: *If  $w$  is a budget balanced linear contract, then  $S(w) = U^0(w)$ .*

Before proving the result, let us heuristically interpret the guarantee  $U^0(w)$ . Suppose momentarily that  $A^0$  consists of one action profile  $a^0$  with  $c(a^0) > 0$ . Consider trying to reduce surplus relative to  $a^0$  by giving agent 1 a new zero-cost action  $a'_1$ . We can get agent 1 to deviate to  $a'_1$  as long as  $\mathbb{E}_{F(a'_1, a_{-1}^0)}[\alpha_1 v(y)] > \mathbb{E}_{F(a^0)}[\alpha_1 v(y)] - c_1(a_1^0)$ , or

$$\mathbb{E}_{F(a'_1, a_{-1}^0)}[v(y)] > \mathbb{E}_{F(a^0)}[v(y)] - \frac{c_1(a_1^0)}{\alpha_1}.$$

That is, we can lower the expected value of output by at most  $c_1(a_1^0)/\alpha_1$ , which is more than agent 1’s cost saving of  $c_1(a_1^0)$  if  $\alpha_1 < 1$ . This is just the usual free-rider problem: agent 1’s deviation imposes a negative externality on the other agents. Having added the action  $a'_1$ , we can then also give agent 2 a zero-cost action  $a'_2$  to reduce the expected value of output further by  $c_2(a_2^0)/\alpha_2$ . Continuing this way, we obtain a zero-cost profile  $a'$  with expected surplus  $\mathbb{E}_{F(a')} [v(y)] \approx \mathbb{E}_{F(a^0)} [v(y)] - \sum_i c_i(a_i^0)/\alpha_i$ , which equals  $U^0(w)$  when  $A^0$  is a singleton. In other words, the guarantee is obtained by exhausting opportunities for free-riding.

Lemma 4.1 follows from the next two lemmas. The first one shows that there always exists an equilibrium with surplus weakly greater than  $U^0(w)$ , which implies  $S(w) \geq U^0(w)$ .

LEMMA 4.2: *Suppose  $w$  is a budget balanced linear contract,  $A$  is a technology that contains  $A^0$ , and  $a^* \in \arg \max_{a \in A} P(a)$ . Then  $a^* \in \mathcal{E}(w, A)$  and  $\mathbb{E}_{F(a^*)} [v(y)] - \sum_i c_i(a_i^*) \geq U^0(w)$ .*

PROOF: The key observation is that a budget balanced linear contract  $w$  induces a potential game between the agents.<sup>8</sup> To see this, fix a technology  $A \supseteq A^0$ . Then each agent’s payoff is  $u_i(a) = \mathbb{E}_{F(a)} [\alpha_i v(y)] - c_i(a_i)$ , and for any agent  $i$  with a positive share  $\alpha_i > 0$ , we have

$$u_i(a_i, a_{-i}) - u_i(a'_i, a_{-i}) = \alpha_i (P(a_i, a_{-i}) - P(a'_i, a_{-i})) \quad \text{for all } a_i, a'_i \in A_i, \text{ and all } a_{-i} \in A_{-i}.$$

Hence, if all agents’ shares are positive, then  $P$  is a weighted potential for the game  $\Gamma(w, A)$ , which implies  $\arg \max_{a \in A} P(a) \subseteq \mathcal{E}(w, A)$ . Moreover, this inclusion holds even if some agents’ shares are zero. This is because any  $a^* \in \arg \max_{a \in A} P(a)$  has  $c_i(a_i^*) = 0$  if  $\alpha_i = 0$ . Thus,  $a_i^*$  maximizes  $u_i(a_i, a_{-i}^*) = -c_i(a_i)$  for such agents, whereas  $a_j^*$  remains a best response for all agents with  $\alpha_j > 0$  as their payoff function can still be derived from the function  $P$  as above.<sup>9</sup> To complete the proof, we note that the inequality follows, since we have

$$\mathbb{E}_{F(a^*)} [v(y)] - \sum_i c_i(a_i^*) \geq \mathbb{E}_{F(a^*)} [v(y)] - \sum_i \frac{c_i(a_i^*)}{\alpha_i} = \max_{a \in A} P(a) \geq \max_{a \in A^0} P(a) = U^0(w),$$

where the second inequality is by  $A \supseteq A^0$  and the last equality is by definition of  $U^0(w)$ . *Q.E.D.*

We then show that the surplus guarantee is no higher than  $U^0(w)$ .

LEMMA 4.3: *Suppose  $w$  is a budget balanced linear contract, and  $G$  is a distribution on  $Y$  such that  $\mathbb{E}_G [v(y)] > U^0(w)$ . Then there exists a technology  $A \supseteq A^0$  such that  $F(\sigma) = G$  and  $\sum_i u_i(\sigma; w, A) = \mathbb{E}_G [v(y)]$  for all  $\sigma \in \mathcal{E}(w, A)$ .*

Note that letting  $\mathbb{E}_G [v(y)] \rightarrow U^0(w)$  implies  $S(w) \leq U^0(w)$ , establishing Lemma 4.1.<sup>10</sup>

<sup>8</sup>All concepts and results related to potential games used in the analysis can be found in [Monderer and Shapley \(1996\)](#). Any contract that aligns the agents’ interests can be shown to induce a type of potential game between the agents, but we will not need the general form of this result.

<sup>9</sup>If some agents’ shares are zero, then  $P$  is not a potential for  $\Gamma(w, A)$ , but it can be shown that  $\Gamma(w, A)$  is still a generalized ordinal potential game. As this will not be used in the analysis, we omit the details.

<sup>10</sup>Lemma 4.3 holds vacuously if  $U^0(w) = \max v(Y)$ . But in that case we have  $S(w) \leq U^0(w)$  a fortiori.



We prove Lemma 4.3 in the Appendix. To outline the idea, suppose all agents' shares are positive for simplicity. Consider a technology  $A \supseteq A^0$  where each agent has a new action  $a'_i$  with  $c_i(a'_i) = 0$ , and  $F(a') = G$ . Completing the description of  $A$  so that  $a' \in \mathcal{E}(w, A)$  is easy. This is because  $\mathbb{E}_{F(a')} [v(y)] = \mathbb{E}_G [v(y)] > U^0(w) = \max_{a \in A^0} P(a)$ , implying that  $a'$  maximizes  $P$  on  $A^0 \cup \{a'\}$ . Thus, if we set  $F(a) = \delta_{y_0}$  for all  $a \notin A^0 \cup \{a'\}$ , then  $\{a'\} = \arg \max_{a \in A} P(a)$  and  $a'$  is an equilibrium of  $\Gamma(w, A)$  by Lemma 4.2. However, the complication is that the game  $\Gamma(w, A)$  so defined could have other equilibria. The proof deals with this by carefully choosing  $F(a)$  for  $a \notin A^0 \cup \{a'\}$  to make  $a'$  the only equilibrium.

We are now ready to prove Theorem 4.1. Part (i) follows by Theorem 3.1 and a continuity and compactness argument, which we relegate to the Appendix. As for parts (ii) and (iii), Lemma 4.1 implies that linear team-optimal contracts and the optimal surplus guarantee correspond to the solutions and the value of the following maximization problem:

$$\max_{\alpha \in [0,1]^I : \sum_i \alpha_i = 1} U^0(w) = \max_{a \in A^0} \max_{\alpha \in [0,1]^I : \sum_i \alpha_i = 1} \left( \mathbb{E}_{F(a)} [v(y)] - \sum_i \frac{c_i(a_i)}{\alpha_i} \right).$$

By standard arguments, the inner maximum is achieved for any  $a \in A^0$  by (4.2). Substituting these shares back into the objective yields (4.1). Part (iv) is immediate from Lemma 3.3.

### 5. PRINCIPAL-OPTIMAL CONTRACTS

We then consider principal-optimal contracts, which maximize the principal's guaranteed expected profit  $V(w)$ . By Theorem 3.1, we can restrict attention to contracts that align the agents' interests as any contract that fails to do so is no better than the zero contract. However, unlike in the case of budget balanced contracts studied in the previous section, this restriction by itself does not imply any relationship between the value of output and the agents' compensation. Nevertheless, we show that a linear contract is still optimal.

- THEOREM 5.1:** (i) *There exists a linear principal-optimal contract.*  
 (ii) *A principal-optimal contract  $w$  guarantees a positive expected profit (i.e.,  $V(w) > 0$ ) if and only if the known technology  $A^0$  satisfies (4.1).*  
 (iii) *If every action profile in the known technology  $A^0$  satisfies full support, then every principal-optimal contract that guarantees a positive expected profit is linear.*

Theorem 5.1 parallels our findings for team-optimal contracts. Part (i) shows the optimality of linear contracts, and part (iii) provides a condition for their uniqueness. By part (ii), the same condition that characterizes whether a team-optimal contract can guarantee positive surplus is necessary and sufficient for the optimal profit guarantee to be positive. Note that the optimal contract depends only on  $v(y)$  even if the signal  $y$  is richer in the sense that  $v$  is not invertible. When  $I = 1$ , these results replicate Carroll's (2015) result for principal-agent problems (save for  $Y$  here being finite as opposed to compact). In that case, (4.1) simplifies to the requirement that some known action yield positive expected surplus.

The rest of this section is devoted to deriving the results in Theorem 5.1. Along the way, we obtain a formula for the profit guarantee of a linear contract, which can be used to find an optimal one. We comment on this at the end of the section.

As a first step, we give a convenient representation for the candidate optimal contracts. By Theorem 3.1, we can focus on contracts that align the agents’ interest. Moreover, we can take the lowest payment to each agent to be zero, because otherwise we could weakly increase the principal’s profit by subtracting a constant from the agent’s compensation as this does not affect incentives.<sup>11</sup> By Lemma 3.1(iv), there then exists some output  $\underline{y} \in Y$  that gives the zero payment simultaneously to all agents, that is,  $w(\underline{y}) = 0$ . A contract with this property is said to be *anchored at the origin*.

Any contract  $w$  that aligns the agents’ interests and is anchored at the origin has the following representation. Let  $\bar{w}(y) := \sum_i w_i(y)$  denote the agents’ total compensation at the output  $y$ . Then there are shares  $\alpha = (\alpha_1, \dots, \alpha_I) \in [0, 1]^I$ ,  $\sum_i \alpha_i = 1$ , such that for all  $i$ ,

$$w_i(y) = \alpha_i \bar{w}(y) \quad \text{for all } y \in Y. \tag{5.1}$$

Conversely, any pair  $(\bar{w}, \alpha)$ , where the total compensation  $\bar{w} : Y \rightarrow \mathbb{R}_+$  has  $\min \bar{w}(Y) = 0$ , defines via (5.1) a contract that aligns the agents’ interests and is anchored at the origin. To see this, consider an auxiliary model where the value of each output  $y$  is  $\bar{w}(y)$  instead of  $v(y)$ ; the analog of output  $y_0$  exists since  $w$  is anchored at the origin. The contract  $w$  is budget balanced and aligns the agents’ interests in this model, so (5.1) follows by Lemma 3.2.<sup>12</sup> The converse is obvious. We will think of any linear contract as a special case of (5.1) where  $\bar{w}(y) = \beta v(y)$  for some  $\beta \geq 0$  so that  $w_i(y) = \alpha_i \beta v(y)$ .

Using (5.1), we can view even a nonlinear contract  $w$  as a budget balanced linear contract in the auxiliary model, allowing us to recycle results from Section 4. To avoid confusion, let

$$\bar{P}(a) := \mathbb{E}_{F(a)}[\bar{w}(y)] - \sum_i \frac{c_i(a_i)}{\alpha_i},$$

where we have replaced  $v(y)$  with  $\bar{w}(y)$  in the definition of  $P(a)$  in (4.3), and where  $\alpha_i$  is agent  $i$ ’s share in (5.1). We then define  $\bar{U}^0(w) := \max_{a \in A^0} \bar{P}(a)$ .

To shorten the statement of results, we say that a contract  $w$  is *eligible* if it (i) aligns the agents’ interests, (ii) is anchored at the origin, and (iii) satisfies  $V(w) > 0$  and  $V(w) \geq V(0)$ . This adapts Carroll’s (2015) notion of an eligible contract to the multiagent setting, parts (i) and (ii) being the novel requirements. The set of eligible contracts may be empty as the best profit guarantee may be zero in violation of (iii). However, if  $V(w) > 0$  for some anchored contract  $w$ , then this contract is eligible unless  $V(w) < V(0)$ , in which case the zero contract is eligible. In particular, any principal-optimal contract with a positive guarantee is eligible.

The following characterization is similar to the single-agent case.

LEMMA 5.1: *Let  $w$  be an eligible contract, different from the zero contract. Then*

$$V(w) = \min_{G \in \Delta(Y)} \mathbb{E}_G[v(y) - \bar{w}(y)] \quad \text{subject to} \quad \mathbb{E}_G[\bar{w}(y)] \geq \bar{U}^0(w). \tag{5.2}$$

Moreover, if  $G$  achieves the minimum, then  $\mathbb{E}_G[\bar{w}(y)] = \bar{U}^0(w)$ .

<sup>11</sup>Given a contract  $w$ , define the contract  $w'$  by  $w'(y) := w(y) - (\min w_1(Y), \dots, \min w_I(Y))$  ( $y \in Y$ ). Then  $w' \leq w$  and  $\mathcal{E}(w, A) = \mathcal{E}(w', A)$  for all  $A \supseteq A^0$ , and hence  $V(w') \geq V(w)$ .

<sup>12</sup>If  $w$  is the zero contract, then the auxiliary model does not satisfy  $\max v(Y) = \max \bar{w}(Y) > 0$ . However, the zero contract clearly satisfies (5.1).

The proof of Lemma 5.1 builds directly on our results for team-optimal contracts. To see that  $V(w)$  is not less than the minimum, interpret  $w$  as a budget balanced linear contract in a model where  $v(y) = \bar{w}(y)$ . Then Lemma 4.2 implies that any technology  $A \supseteq A^0$  has an equilibrium  $a^*$  such that  $\mathbb{E}_{F(a^*)}[\bar{w}(y)] \geq \mathbb{E}_{F(a^*)}[\bar{w}(y)] - \sum_i c_i(a_i^*) \geq \bar{U}^0(w)$ . Thus,  $V(w)$  is at least the minimum profit under distributions satisfying the constraint in (5.2).

In the other direction, take any  $G \in \Delta(Y)$  such that  $\mathbb{E}_G[\bar{w}(y)] > \bar{U}^0(w)$ . Interpret  $w$  again as a budget balanced linear contract with  $v(y) = \bar{w}(y)$ . Then Lemma 4.3 gives us a technology  $A \supseteq A^0$  where  $G$  is the unique equilibrium output distribution. The proof in the Appendix uses this fact to bound  $V(w)$  from above by the minimum in (5.2).

Lemma 5.1 yields a formula for the profit guarantee for any eligible linear contract  $w$ , which is valid also for the zero contract.

LEMMA 5.2: *Let  $w$  be an eligible linear contract, with  $w_i(y) = \alpha_i \beta v(y)$  for each  $i$ . Then*

$$V(w) = (1 - \beta) \max_{a \in A^0} \left( \mathbb{E}_{F(a)}[v(y)] - \sum_i \frac{c_i(a_i)}{\beta \alpha_i} \right), \tag{5.3}$$

where  $0/0 = 0$  and  $x/0 = \infty$  for  $x > 0$  by convention.

Note that if  $I = 1$ , then  $\alpha_1 = 1$  and (5.3) reduces to the single-agent formula by Chas-sang (2013) and Carroll (2015).

PROOF: If  $w$  is different from the zero contract and  $G$  achieves the minimum in (5.2), then

$$V(w) = (1 - \beta) \mathbb{E}_G[v(y)] = \frac{1 - \beta}{\beta} \mathbb{E}_G[\bar{w}(y)] = \frac{1 - \beta}{\beta} \bar{U}^0(w).$$

The formula in (5.3) now follows by writing out  $\bar{U}^0(w)$ .

If  $w$  is the zero contract, then the agents can only play zero-cost actions. (Recall that each  $A_i^0$  contains such an action by assumption.) Any  $a \in A^0$  with  $c(a) = 0$  is an equilibrium given any technology  $A \supseteq A^0$ , and hence  $V(0) = \max \mathbb{E}_{F(a)}[v(y)]$  over  $a \in A^0$  such that  $c(a) = 0$ . This agrees with the formula in the lemma, given the conventions involving 0. Q.E.D.

LEMMA 5.3: *There exists a linear contract  $w^*$  such that  $V(w^*) \geq V(w)$  for every linear contract  $w$ . Moreover,  $V(w^*) > 0$  if and only if the known technology satisfies (4.1).*

PROOF: If no linear contract is eligible, then  $V(0) = 0$  and the zero contract is optimal within the class of linear contracts. If there exists an eligible linear contract, then the claim follows by continuity of (5.3) in  $\beta$  and  $\alpha_i$ .

If (4.1) holds, then  $S(w) = U^0(w) = \max_{a \in A^0} (\mathbb{E}_{F(a)}[v(y)] - \sum c_i(a_i)/\alpha_i) > 0$  for some  $\alpha \in [0, 1]^I$  with  $\sum \alpha_i = 1$  by Theorem 4.1(ii) and Lemma 4.1. By continuity, (5.3) is positive for this same  $\alpha$  for  $\beta < 1$  close enough to 1. Conversely, if (4.1) is not satisfied, then  $(1 - \beta) \max_{a \in A^0} (\mathbb{E}_{F(a)}[v(y)] - \sum c_i(a_i)/\beta \alpha_i) \leq (1 - \beta) U^0(w) = (1 - \beta) S(w) \leq 0$  for all  $\alpha \in [0, 1]^I$  with  $\sum \alpha_i = 1$  and all  $\beta \leq 1$  by Theorem 4.1(ii) and Lemma 4.1. Therefore, (5.3) is nonpositive for any linear contract with  $\beta \leq 1$ . (If  $\beta > 1$ , then profit is nonpositive a fortiori.) Q.E.D.

With these facts about linear contracts established, it remains to show that the best linear contract is optimal among all contracts. Consider the representation (5.1). We will show that any contract can be (weakly) improved by making the total compensation  $\bar{w}$  linear in the value of output while keeping the agents' shares,  $\alpha$ , fixed. As  $\bar{w}$  is one-dimensional, we draw here on the single-agent case. In particular, the following lemma, whose proof we relegate to the Appendix, identifies a supporting hyperplane to the set of pairs  $(\bar{w}(y), v(y) - \bar{w}(y))$  under contract  $w$ , which will be used to define the improvement contract.

LEMMA 5.4: *Let  $w$  be an eligible contract, different from the zero contract. Then there exist numbers  $\kappa$  and  $\lambda$ , with  $\lambda > 0$ , such that*

$$v(y) - \bar{w}(y) \geq \kappa + \lambda \bar{w}(y) \quad \text{for all } y \in Y, \tag{5.4}$$

$$V(w) = \kappa + \lambda \bar{U}^0(w). \tag{5.5}$$

Given an eligible contract  $w \neq 0$  and numbers  $\kappa, \lambda$  satisfying (5.4) and (5.5), define the affine contract  $w'$  by

$$\bar{w}'(y) := \frac{1}{1 + \lambda} v(y) - \frac{\kappa}{1 + \lambda} \quad \text{and} \quad w'_i(y) := \alpha_i \bar{w}'(y), \tag{5.6}$$

where  $\alpha_i$  is agent  $i$ 's share in the representation (5.1) of the original contract  $w$ . Then  $\bar{w}'(y) \geq \bar{w}(y) \geq 0$  for all  $y \in Y$  by (5.4), and thus  $w'_i \geq 0$  for all  $i$  as required by our definition of a contract. Note that  $\bar{w}'(y_0) = -\kappa/(1 + \lambda) \geq 0$  implies  $\kappa \leq 0$ .

The affine contract  $w'$  can be improved by removing the constant payment, which does not affect the agents' incentives. That is, define the linear contract  $w''$  by setting

$$w''_i(y) := \frac{\alpha_i}{1 + \lambda} v(y) = w'_i(y) + \frac{\alpha_i \kappa}{1 + \lambda} \leq w'_i(y), \tag{5.7}$$

where the inequality holds because, as noted above,  $\kappa \leq 0$ .

LEMMA 5.5: *Suppose  $w$  is an eligible contract, different from the zero contract, that satisfies (5.4) and (5.5), and  $w''$  is the linear contract defined by (5.7). Then  $V(w'') \geq V(w)$ . Moreover, if every  $a \in A^0$  satisfies full support and  $w$  is not linear, then  $V(w'') > V(w)$ .*

PROOF: Fix the contract  $w$ . Let  $w'$  and  $w''$  be the affine and linear contracts defined by (5.6) and (5.7). Observe first that since  $\bar{w}(y) \leq \bar{w}'(y) = \bar{w}''(y) - \kappa/(1 + \lambda)$  for all  $y \in Y$ , we have

$$\bar{U}^0(w) \leq \max_{a \in A^0} \left( \mathbb{E}_{F(a)} \left[ \bar{w}''(y) - \frac{\kappa}{1 + \lambda} \right] - \sum_i \frac{c_i(a_i)}{\alpha_i} \right) = \bar{U}^0(w'') - \frac{\kappa}{1 + \lambda}. \tag{5.8}$$

The contract  $w''$ , being linear, satisfies (5.1). Reinterpreting it as a budget-balanced contract we apply Lemma 4.2 (with the substitutions  $v(y) = \bar{w}''(y)$  and  $U^0(w'') = \bar{U}^0(w'')$ ) to find for any technology  $A \supseteq A^0$  a pure-strategy equilibrium  $a^* \in \mathcal{E}(w'', A)$  such that  $\mathbb{E}_{F(a^*)}[\bar{w}''(y)] \geq \bar{U}^0(w'')$ . But  $\mathcal{E}(w', A) = \mathcal{E}(w'', A)$  as the constants do not affect incentives. Thus,  $a^* \in \mathcal{E}(w', A)$  and

$$\mathbb{E}_{F(a^*)}[\bar{w}'(y)] = \mathbb{E}_{F(a^*)}[\bar{w}''(y)] - \frac{\kappa}{1 + \lambda} \geq \bar{U}^0(w'') - \frac{\kappa}{1 + \lambda} \geq \bar{U}^0(w),$$

where the last inequality is by (5.8). Moreover,  $w'$  satisfies (5.4) by construction, and thus

$$V(w', A) \geq \mathbb{E}_{F(a^*)}[v(y) - \bar{w}'(y)] \geq \kappa + \lambda \mathbb{E}_{F(a^*)}[\bar{w}'(y)] \geq \kappa + \lambda \bar{U}^0(w) = V(w), \quad (5.9)$$

where the last step is by (5.5). Because  $A$  was arbitrary, this implies  $V(w') \geq V(w)$ . Now  $V(w'') = V(w') - \kappa/(1 + \lambda) \geq V(w)$  shows that  $V(w'') \geq V(w)$  as desired.

It remains to show strict inequality for nonlinear contracts under full support. If  $w'$  is not linear (i.e., if  $\kappa < 0$ ), then  $V(w'') > V(w') \geq V(w)$ . So suppose  $w'$  is linear, and every  $a \in A^0$  has full support, that is,  $F(a) \neq \delta_{y_0}$  implies  $\text{supp } F(a) = Y$  for all  $a \in A^0$ . If  $w$  is not linear, then  $\bar{w}(y) \leq \bar{w}'(y) = \bar{w}''(y) - \kappa/(1 + \lambda)$  holds with strict inequality for some  $y \in Y$ . Furthermore, because  $w$  is eligible, we have  $\bar{U}^0(w) > 0$  by Lemma 5.1, and so the maximum in  $\bar{U}^0(w)$  is achieved by some  $a \in A^0$  such that  $F(a)$  has full support. This implies that the inequality in (5.8) is strict. The strict inequality carries through to imply that in (5.9),  $V(w', A)$  is bounded above  $V(w)$  uniformly in  $A \supseteq A^0$ . Therefore,  $V(w'') \geq V(w') > V(w)$ . Q.E.D.

We can now show the claims in Theorem 5.1 using the previous lemmas.

**PROOF OF THEOREM 5.1:** For part (i), let  $w$  be a contract that aligns the agents' interests and is anchored at the origin. If  $w$  is not eligible, then  $V(w) \leq V(0)$ . If  $w$  is eligible, then  $V(w) \leq V(w'')$  for a linear contract  $w''$  by Lemma 5.5. Thus, either way,  $V(w)$  is less than the guarantee of some linear contract, and hence the claim follows by Lemma 5.3.

Part (ii) follows from part (i) and Lemma 5.3.

For part (iii), suppose all  $a \in A^0$  have full support, and  $w$  is a nonlinear principal-optimal contract and thus anchored at the origin. If  $V(w) > 0$ , then  $w$  is eligible, and  $V(w) < V(w'')$  for a linear contract  $w''$  by Lemma 5.5, contradicting optimality of  $w$ . Q.E.D.

When (4.1) holds, a linear principal-optimal contract can be found by maximizing (5.3) with respect to  $\beta$  and  $\alpha$ . (Otherwise, the zero contract is optimal.) It is easier to first find the optimal shares for each  $a \in A^0$ , and then maximize with respect to  $a$ . So fix  $a \in A^0$ . Given any  $\beta > 0$ , maximizing (5.3) with respect to  $\alpha$  amounts to finding team-optimal shares in a model where the value of output is  $\beta v(y)$ . These shares  $\alpha(a)$  are given by (4.2), and for fixed  $a$  they are independent of  $\beta$ . Similarly, maximizing (5.3) with respect to  $\beta$  given any  $\alpha$  gives the principal-optimal share in a single-agent model where each action  $a \in A^0$  costs  $C(a, \alpha) := \sum_i c_i(a_i)/\alpha_i$ . Carroll (2015) shows that this is given by  $\beta(a, \alpha) = \sqrt{C(a, \alpha)/\mathbb{E}_{F(a)}[v(y)]}$ . Substituting these shares back into (5.3) gives

$$\left( \sqrt{\mathbb{E}_{F(a)}[v(y)]} - \sqrt{C(a, \alpha(a))} \right)^2 = \left( \sqrt{\mathbb{E}_{F(a)}[v(y)]} - \sum_i \sqrt{c_i(a_i)} \right)^2.$$

Maximizing the above expression over the set of known action profiles yields a maximizer  $a^* \in A^0$ , which can be substituted back into the formulas for the shares to obtain the optimal contract  $\alpha^* = \alpha(a^*)$  and  $\beta^* = \beta(a^*, \alpha^*)$ .

### 6. RISK-AVERSE AGENTS

The necessity of interest alignment can be generalized to risk-averse agents. To this end, suppose that each agent's payoff under a contract  $w : Y \rightarrow \mathbb{R}_+^I$  is of the form

$\hat{u}_i(w_i(y)) - c_i(a_i)$  for an increasing, concave function  $\hat{u}_i : \mathbb{R}_+ \rightarrow \mathbb{R}$  that is continuous at 0 with  $\hat{u}_i(0) = 0$ . The model is otherwise as in Section 2. In particular, the principal is still a risk-neutral utilitarian, and so the expected surplus given a contract  $w$  and a technology  $A$  now takes the form  $S(w, A) = \max_{\sigma \in \mathcal{E}(w, A)} \sum_i u_i(\sigma; w, A)$ , with  $u_i(a; w, A) = \mathbb{E}_{F(a)}[\hat{u}_i(w_i(y)) - c_i(a_i)]$ . The surplus guarantee is defined the same way as before:  $S(w) = \inf_{A \supseteq A^0} S(w, A)$ .

The following definition generalizes the condition in Lemma 3.1(iii) to this setting.

**DEFINITION 6.1:** A contract  $w$  aligns the agents' interests in utilities if the set of all payment-utility profiles  $\hat{u}(w(Y)) := \{(\hat{u}_1(w_1(y)), \dots, \hat{u}_I(w_I(y))) : y \in Y\}$  is contained in a ray in  $\mathbb{R}_+^I$ , that is, if  $\hat{u}(w(Y)) \subset \{\underline{u} + dt : t \in \mathbb{R}_+\}$  for some  $\underline{u}, d \in \mathbb{R}_+^I$ .

Any contract satisfying Definition 6.1 prescribes team-based compensation in the sense that all payment profiles  $(w_1(y), \dots, w_I(y))$ ,  $y \in Y$ , lie on a one-dimensional path in  $\mathbb{R}_+^I$  and  $w_i(y) > w_i(y')$  implies  $w_j(y) \geq w_j(y')$  for all  $i, j, y, y'$ . Thus, the agents' compensation covaries positively, but not necessarily linearly.

**THEOREM 6.1:** If a contract  $w$  fails to align the agents' interests in utilities, then  $V(w) \leq V(0)$ . If, in addition,  $w$  is budget balanced, then there exists a budget-balanced contract  $w'$  that aligns the agents' interests in utilities such that  $S(w) \leq S(w')$ .

Theorem 6.1 generalizes Theorem 3.1 to show the necessity of interest alignment for risk-averse agents. (A minor difference is that now, in the budget-balanced case, we only have the existence of some interest-aligned contract  $w'$  that improves on  $w$ .) The proof is essentially the same after we change variables by defining  $\tilde{w}_i(y) := \hat{u}_i(w_i(y))$  so that the principal's payoff is  $v(y) - \sum_i \hat{u}_i^{-1}(\tilde{w}_i(y))$ . Then  $w$  aligns the agents' interests in utilities if and only if  $\tilde{w}$  satisfies Definition 3.1. Lemma 3.5 applies to  $\tilde{w}$  verbatim as its proof makes no reference to the cost of the contract to the principal. The adjustment required to Lemma 3.4 is handled in the Appendix.

With interests aligned in the utility space, the properties of payments depend on the utility functions. For example, collinearity can be recovered for particular CRRRA preferences.

**LEMMA 6.1:** Suppose the agents' preferences over money are represented by symmetric power utility functions (i.e.,  $\hat{u}_i(x) = x^\rho$  for some  $\rho \in (0, 1]$  independent of  $i$ ). Let  $w$  be a contract that is anchored at the origin. Then  $w$  aligns the agents' interests in utilities if and only if it aligns the agents' interests in the sense of Definition 3.1.

The proof is immediate from the properties of the power utility, and hence omitted.

Any budget balanced contract is necessarily anchored at the origin, and so is any principal-optimal contract. Thus, with symmetric power utility functions, optimal contracts have the agents' compensation covarying positively and linearly as in the risk-neutral case.

### 7. CONCLUDING REMARKS

We have shown that demanding team incentives to be robust to nonquantifiable uncertainty about the game played by the agents leads to contracts that align the agents' interests. Under budget balance such contracts are equivalent to linear contracts, implying

that a linear scheme has the best surplus guarantee for a budget balanced team. A linear contract was also shown to have the best profit guarantee to an outside residual claimant. These optimal linear contracts still suffer from the free-rider problem, but if the known technology is sufficiently productive, a positive guarantee remains.

Aligning the agents’ interests not only limits a contract’s downside, it may also increase its upside as it motivates the agents to take advantage of unexpected opportunities to help and to allocate tasks efficiently. This upside is lost on our worst-case analysis. (See Itoh (1991) and Garicano and Santos (2004) for incentives to help and to refer clients in Bayesian models.) Note that if the opportunities are not unexpected (i.e., if they are part of the known technology), then they do affect our analysis: withholding help or not referring a task to a better agent are examples of the kind of negative actions that drive the worst case.

Finally, considering the worst case over all games consistent with the known technology is a strong assumption, which facilitates tractable analysis and yields sharp predictions about optimal contracts. Robustness is, however, only one of many considerations affecting contract design, and contracts observed in practice reflect it to varying degrees. A natural way to try to incorporate this into the analysis would be to restrict the set of games deemed possible, with smaller sets resulting in less limitations on contract form. Identifying subsets of games for which the analysis remains tractable is a nontrivial problem left for future work.

APPENDIX

A.1. Proofs for Section 3

PROOF OF LEMMA 3.1: Suppose in negation of (i) that there exist agents  $i, j$  and distributions  $F, G \in \Delta(Y)$  such that  $\mathbb{E}_F[w_i(y)] > \mathbb{E}_G[w_i(y)]$  and  $\mathbb{E}_F[w_j(y)] < \mathbb{E}_G[w_j(y)]$ . Then  $\text{co}(w_{i,j}(Y)) = \{\mathbb{E}_F[(w_i(y), w_j(y))] : F \in \Delta(Y)\}$  is not contained in a ray in  $\mathbb{R}_+^2$ , and hence neither is  $w_{i,j}(Y)$  in negation of (ii). Thus, (ii) implies (i). Conversely, if (ii) fails, then  $\text{co}(w_{i,j}(Y))$  is not contained in a ray in  $\mathbb{R}_+^2$  and we can find  $F, G \in \Delta(Y)$  such that  $\mathbb{E}_F[w_i(y)] > \mathbb{E}_G[w_i(y)]$  and  $\mathbb{E}_F[w_j(y)] < \mathbb{E}_G[w_j(y)]$  in negation of (i). Thus, (i) implies (ii).

Suppose  $w$  satisfies (iii). Then  $w(Y)$  is contained in a ray in  $\mathbb{R}_+^I$ , the projection of which to agents  $i$  and  $j$ ’s payments is a ray in  $\mathbb{R}_+^2$ , which contains  $w_{i,j}(Y)$ . Thus, (iii) implies (ii). Conversely, suppose  $w$  satisfies (ii). If  $w$  is constant, then it clearly satisfies (iii), so suppose  $w_i$  is not constant for some agent  $i$ . By (ii), we can write  $w_j = \alpha_j w_i + \beta_j$  for some  $\alpha_j \geq 0$  and  $\beta_j \in \mathbb{R}$  for all  $j$ . Then  $w(y) = (\alpha_1 w_i(y) + \beta_1, \dots, w_i(y), \dots, \alpha_I w_i(y) + \beta_I)$ , so every  $w(y)$  clearly lies on the same ray in  $\mathbb{R}_+^I$ , which is (iii).

The equivalence of (iii) and (iv) is immediate as  $Y$  is finite.

*Q.E.D.*

PROOF OF LEMMA 3.3: Suppose  $w$  is a budget balanced contract that fails to align the agents’ interests. By budget balance, we have  $Y^* \subseteq \arg \max_{y \in Y} v(y)$ . The condition in Lemma 3.3 then implies that for all  $a \in A^0$ ,  $\text{supp } F(a) \subseteq Y^* \subseteq \arg \max_{y \in Y} v(y)$  implies  $c(a) \neq 0$ . Thus, Lemma 3.5 applies, which gives  $S(w) \leq 0$ .

*Q.E.D.*

PROOF OF LEMMA 3.4: Part (i): Fix  $a^* \in A^0$  as in the lemma. Consider a technology  $A' \supseteq A^0$  such that  $A'_i = A^0_i \cup \{a'_i\}$  with  $c_i(a'_i) = 0$  and  $F(a'_i, a_{-i}) = F(a^*)$  for all  $i$  and all  $a_{-i} \in A'_{-i}$ . Then each agent can ensure his highest feasible payoff  $\max w_i(Y)$  by playing

$a'_i$ . This implies that any equilibrium  $\sigma \in \mathcal{E}(w, A')$  can assign positive probability only to  $a$  such that  $c(a) = 0$  and  $\text{supp } F(a) \subseteq Y^*$ . Hence,

$$\begin{aligned} V(w, A') &\leq \max_{a \in A': c(a)=0 \text{ and } \text{supp } F(a) \subseteq Y^*} \mathbb{E}_{F(a)} \left[ v(y) - \sum_i w_i(y) \right] \\ &= \max_{a \in A^0: c(a)=0 \text{ and } \text{supp } F(a) \subseteq Y^*} \mathbb{E}_{F(a)} \left[ v(y) - \sum_i w_i(y) \right] \\ &< \max_{a \in A^0: c(a)=0 \text{ and } \text{supp } F(a) \subseteq Y^*} \mathbb{E}_{F(a)} [v(y)] \\ &\leq V(0). \end{aligned} \tag{A.1}$$

Above, the second line follows from the first one, since the set of distributions associated with zero-cost profiles is the same in  $A'$  and  $A^0$ ; the strict inequality follows, since  $w_i(y) > 0$  for all  $y \in Y^*$  for some agent  $i$  because  $w$  is different from the zero contract; the last inequality follows since every  $a \in A^0$  with  $c(a) = 0$  is an equilibrium under the zero contract given any  $A \supseteq A^0$ . Thus,  $V(w) \leq V(w, A') < V(0)$ .

Part (ii): Consider the third line in (A.1). If  $a \in A^0$  has costly production, then  $c(a) = 0$  implies  $\mathbb{E}_{F(a)}[v(y)] = 0$ . If it has full support, then  $\text{supp } F(a) \subseteq Y^*$  implies  $F(a) = \delta_{y_0}$ , since  $Y^* \subsetneq Y$  whenever  $w$  fails to align the agents' interests. Thus,  $\mathbb{E}_{F(a)}[v(y)] = 0$  for all feasible  $a$ , and so the maximum is zero. Thus,  $V(w) \leq V(w, A') < 0$ .

Part (iii): Suppose that  $w$  is budget balanced so that  $\sum_i w_i(y) = v(y)$  for all  $y$ . Then  $Y^* \subseteq \arg \max_{y \in Y} \sum_i w_i(y) = \arg \max_{y \in Y} v(y)$ . Thus,  $\text{supp } F(a^*) \subseteq \arg \max_{y \in Y} v(y)$ . We claim that  $a^* \in \mathcal{E}(w', A)$  for any budget-balanced contract  $w'$  that aligns the agents' interests and any technology  $A \supseteq A^0$ . Indeed,  $w'_i$  is of the form  $w'_i(y) = \alpha_i v(y)$  for some  $\alpha_i \geq 0$  by Lemma 3.2. So  $a^*$  gives all agents their highest feasible payoff under  $w'$  as it maximizes  $v(y)$  at zero cost. Hence,  $a^*$  is an equilibrium. Because  $A$  was arbitrary, this implies that  $S(w') \geq \mathbb{E}_{F(a^*)}[v(y)] = \max v(Y) \geq S(w)$ .

Part (iv): Suppose  $w$  is budget balanced so that  $Y^* \subseteq \arg \max_{y \in Y} v(y) \subsetneq Y$ . If  $a^*$  has costly production, then  $\text{supp } F(a^*) \subseteq Y^* = \arg \max_{y \in Y} v(y)$  implies  $c(a^*) \neq 0$ , which contradicts  $c(a^*) = 0$ . If  $a^*$  has full support, then  $\text{supp } F(a^*) \subseteq Y^* \subsetneq Y$  implies  $F(a^*) = \delta_{y_0}$ , which contradicts  $\text{supp } F(a^*) \subseteq Y^* \subseteq \arg \max_{y \in Y} v(y)$ . Q.E.D.

PROOF OF LEMMA 3.6: Consider condition (3.1). Take any  $z^1, \dots, z^l$  in  $W$  such that  $z^i_i = \max w_i(Y)$  for all  $i$ . Let  $a \in A^0$ . If  $\text{supp } F(a) \not\subseteq Y^*$ , then  $\sum_i z^i_i = \sum_i \max w_i(Y) > \sum_i \mathbb{E}_{F(a)}[w_i(y)]$ , and hence the inequality in (3.1) holds at  $a$  as costs are nonnegative. If  $\text{supp } F(a) \subseteq Y^*$  so that  $\sum_i z^i_i = \sum_i \max w_i(Y) = \sum_i \mathbb{E}_{F(a)}[w_i(y)]$ , then the assumption in Lemma 3.5 implies that we have  $\sum_i c_i(a_i) > 0$ , and thus the inequality in (3.1) again holds at  $a$ . Therefore, (3.1) is satisfied. Moreover, by finiteness of  $A^0$ , we can fix  $\eta > 0$  such that (3.1) is satisfied by any collection of points  $z^1, \dots, z^l$  such that  $z^i_i \geq \max w_i(Y) - \eta$  for all  $i$ .

We will choose  $z^i \in W$  for all  $i > 2$  as follows. Fix  $y^i \in \arg \max_{y \in Y} w_i(y)$  and write  $H$  for the uniform distribution on  $Y$ . Let  $G_i := (1 - \theta)\delta_{y^i} + \theta H$ , and let  $z^i := \mathbb{E}_{G_i}[w(y)]$ , where  $\theta > 0$  is common for all  $i > 2$  and small enough so that  $z^i_i \geq \max w_i(Y) - \eta$ . This choice satisfies (3.3) as each  $z^i$  is perturbed by the same amount of uniform noise.

It remains to fix  $z^1$  and  $z^2$ . Suppose first that  $\arg \max_{y \in Y} w_1(y) \cap \arg \max_{y \in Y} w_2(y) = \emptyset$ . Let  $z^j = w(y)$  for some  $y \in \arg \max_{y \in Y} w_j(y)$  for  $j = 1, 2$ . Then  $z^j_j = \max w_j(Y) > z^j_i$  for



all  $i \neq j$ , where the inequality for  $i > 2$  follows because such  $z^i$  is generated by  $G_i$  with full support and  $w_j$  is not constant. Thus, (3.2) holds. Moreover, we have  $z_i^i \geq \max w_i(Y) - \eta$  for all  $i$ , which implies (3.1), proving the lemma for this case.

Consider then the remaining case where  $\arg \max_{y \in Y} w_1(y) \cap \arg \max_{y \in Y} w_2(y)$  contains some output  $\bar{y}$ . Then the projection of  $W$  to the payments to agents 1 and 2, or  $W_{1,2}$ , is a convex set with a nonempty interior relative to  $\mathbb{R}_+^2$ . (Otherwise it would be a nondecreasing line segment, which would align the interests of agents 1 and 2.) This implies that we can find points  $z^1$  and  $z^2$  in  $W$ , arbitrarily close to  $w(\bar{y})$ , such that  $z_j^j > z_j^i$  for  $i, j \in \{1, 2\}$ ,  $i \neq j$ .<sup>13</sup> Choosing  $z^1$  and  $z^2$  sufficiently close to  $w(\bar{y})$  ensures that we also have  $z_j^j > z_j^i$  for  $i > 2$ , because  $w_j(\bar{y}) = \max w_j(Y) > z_j^i$  for  $j \in \{1, 2\}$  and  $i \notin \{1, 2\}$  as each  $z^i$  is generated by a full support distribution. This establishes (3.2). Moreover, for  $z^1$  and  $z^2$  close enough to  $w(\bar{y})$  we also have  $z_i^i \geq \max w_i(Y) - \eta$  for all  $i$ , which in turn implies (3.1). *Q.E.D.*

**PROOF OF LEMMA 3.7:** Part (i): Fix a mixed strategy profile  $\sigma$  with  $\text{supp } \sigma \subseteq A^0$ . By (3.4), any agent  $i$  who deviates to  $a_i^*$  gets  $u_i(a_i^*, \sigma_{-i}; w, A) = z_i^i - c_i(a_i^*) = z_i^i$ . Summing over  $i$  gives

$$\sum_i u_i(a_i^*, \sigma_{-i}; w, A) = \sum_i z_i^i > \sum_{a \in A^0} \sigma(a) \sum_i \mathbb{E}_{F(a)}[w_i(y) - c_i(a_i)] = \sum_i u_i(\sigma; w, A),$$

where the inequality is by (3.1). Therefore,  $u_i(a_i^*, \sigma_{-i}; w, A) > u_i(\sigma; w, A)$  for some agent  $i$ .

Part (ii): Fix agent  $i$ ,  $a_i^0 \in A^0$ , and  $a_{-i} \in A_{-i}^* \setminus A_{-i}^0$ . Let  $j$  be the lowest index among agents  $-i$  for which  $a_j \notin A_j^0$ . If  $i > j$ , then (3.4) implies that  $i$  can affect the expected payments only if  $i = 2$ , and hence

$$u_i(a_i^*, a_{-i}; w, A) - u_i(a_i^0, a_{-i}; w, A) = \mathbf{1}_{\{i=2\}}(x_2^* - z_2^1) + c_i(a_i^0) \geq 0, \tag{A.2}$$

where  $x_2^* - z_2^1 = \frac{1}{2}(z_2^2 - z_2^1) > 0$  by (3.2). On the other hand, if  $i < j \neq 2$ , then (3.4) implies

$$u_i(a_i^*, a_{-i}; w, A) - u_i(a_i^0, a_{-i}; w, A) = z_i^i - z_i^j + c_i(a_i^0) \geq 0, \tag{A.3}$$

where  $z_i^i - z_i^j \geq 0$  by (3.2) or (3.3). Finally, if  $i < j = 2$ , then  $i = 1$  and (3.4) implies  $u_1(a_1^*, a_{-1}; w, A) - u_1(a_1^0, a_{-1}; w, A) = x_1^* - z_1^2 + c_1(a_1^0) > 0$ , where  $x_1^* - z_1^2 > 0$  by (3.2).

Part (iii): Inspecting the proof of part (ii), the inequality in (A.2) is strict if  $i = 2$ . Similarly, the inequality in (A.3) is strict for  $i \in \{1, 2\}$  by (3.2). *Q.E.D.*

**PROOF OF LEMMA 3.8:** Let  $\sigma \in \mathcal{E}(w, A)$  where  $A \supseteq A^*$  satisfies (3.5). We will write  $u_i(a) := u_i(a; w, A)$  to simplify notation. We show first that  $\sigma_i(A_i^0) = 0$  for some agent  $i$ . Suppose not. Let  $\hat{\sigma}_i(a_i) := \sigma_i(a_i) / \sigma_i(A_i^0)$  for all  $a_i \in A_i^0$ , all  $i$ . Then  $\hat{\sigma}$  is a mixed strategy profile with  $\text{supp } \hat{\sigma} \subseteq A^0$ . By Lemma 3.7(i), there exists an agent  $i$  and an action  $a_i^0 \in \text{supp } \hat{\sigma}_i \subseteq \text{supp } \sigma_i$  such that  $u_i(a_i^*, \hat{\sigma}_{-i}) > u_i(a_i^0, \hat{\sigma}_{-i})$ . Moreover, Lemma 3.7 and (3.5)

<sup>13</sup>Choose 2-vectors  $(z_1^1, z_2^1)$  and  $(z_1^2, z_2^2)$  in  $\text{int}(W_{1,2}) \subset \mathbb{R}_+^2$  with the desired properties, and then choose any  $I$ -vectors  $z^1$  and  $z^2$  in  $W \subset \mathbb{R}_+^I$  whose projections in  $W_{1,2}$  are  $(z_1^1, z_2^1)$  and  $(z_1^2, z_2^2)$ , respectively.

imply  $u_i(a_i^*, a_{-i}) \geq u_i(a_i^0, a_{-i})$  for all  $a_{-i} \in A_{-i} \setminus A_{-i}^0$ . We thus have

$$\begin{aligned} u_i(a_i^*, \sigma_{-i}) &= \prod_{j \neq i} \sigma_j(A_j^0) \sum_{a_{-i} \in A_{-i}^0} \hat{\sigma}_{-i}(a_{-i}) u(a_i^*, a_{-i}) + \sum_{a_{-i} \notin A_{-i}^0} \sigma_{-i}(a_{-i}) u_i(a_i^*, a_{-i}) \\ &> \prod_{j \neq i} \sigma_j(A_j^0) \sum_{a_{-i} \in A_{-i}^0} \hat{\sigma}_{-i}(a_{-i}) u(a_i^0, a_{-i}) + \sum_{a_{-i} \notin A_{-i}^0} \sigma_{-i}(a_{-i}) u_i(a_i^0, a_{-i}) \\ &= u_i(a_i^0, \sigma_{-i}), \end{aligned}$$

which contradicts  $a_i^0 \in \text{supp } \sigma_i$ .

If  $I = 2$ , then we are done. So suppose  $I > 2$  and assume towards contradiction that  $\sigma_1(A_1^0)\sigma_2(A_2^0) > 0$ . Consider agent 1. As argued above, we have  $\sigma_i(A_i^0) = 0$  for some  $i > 2$ , and thus  $\text{supp } \sigma_{-1} \subseteq A_{-1} \setminus A_{-1}^0$ . Therefore,  $u_1(a_1^*, a_{-1}; w, A) - u_1(a_1^0, a_{-1}; w, A) \geq 0$  for all  $a_1^0 \in A_1^0$  and all  $a_{-1} \in \text{supp } \sigma_{-1}$  by (3.5) and Lemma 3.7(ii). Furthermore, if  $a_2 \in A_2^0$ , then the inequality is strict as then  $u_1(a_1^*, a_{-1}; w, A) - u_1(a_1^0, a_{-1}; w, A) = z_1^1 - z_1^j + c_1(a_1^0) > 0$ , where  $j = \min\{i > 2 : a_i = a_i^*\}$  and  $z_1^1 > z_1^j$  by (3.2). Because  $\sigma_2(A_2^0) > 0$ , this implies that  $a_1^*$  is a strictly better response than any  $a_1^0 \in A_1^0$ , contradicting  $\sigma_1(A_1^0) > 0$ . *Q.E.D.*

### A.2. Proofs for Section 4

**PROOF OF THEOREM 4.1:** Parts (ii)–(iv) are proven in the main text. For part (i), identify the space of budget balanced linear contracts with the set  $B := \{\alpha \in [0, 1]^I : \sum_i \alpha_i = 1\}$ , and write  $S(\alpha)$  for the surplus guarantee of contract  $\alpha$ . By Theorem 3.1 and Lemma 3.2, it suffices to show that  $S(\alpha)$  is an upper semicontinuous function of  $\alpha$  on the compact set  $B$ . Fix a sequence  $(\alpha^n)$  in  $B$  converging to some  $\alpha \in B$ . (Since  $B$  is finite dimensional, any norm will do.) We need to show that  $S(\alpha) \geq \limsup_n S(\alpha^n)$ . By moving to a subsequence if necessary, we can assume that  $S(\alpha^n)$  converges to  $\limsup_n S(\alpha^n)$ . Fix any technology  $A \supseteq A^0$  and denote by  $\sigma^n$  the equilibrium of  $\Gamma(\alpha^n, A)$  that achieves  $S(\alpha^n, A)$ . Extracting a further subsequence if necessary, we can assume that the sequence  $(\sigma^n)$  converges to some  $\sigma \in \Delta(A)$ . Since the agents’ payoffs are continuous in  $\alpha$ , the profile  $\sigma$  is an equilibrium of  $\Gamma(\alpha, A)$  by the upper hemicontinuity of the Nash equilibrium correspondence. We thus have

$$\begin{aligned} S(\alpha, A) &\geq \mathbb{E}_{F(\sigma)}[v(y)] - \sum_a \sigma(a) \sum_i c_i(a) \\ &= \lim_n \left( \mathbb{E}_{F(\sigma^n)}[v(y)] - \sum_a \sigma^n(a) \sum_i c_i(a) \right) = \lim_n S(\alpha^n, A) \geq \lim_n S(\alpha^n). \end{aligned}$$

Since  $A \supseteq A^0$  was arbitrary, this implies  $S(\alpha) \geq \lim_n S(\alpha^n)$  as desired. *Q.E.D.*

**PROOF OF LEMMA 4.3:** Fix a budget balanced linear contract  $w$  and a distribution  $G \in \Delta(Y)$  such that  $\mathbb{E}_G[v(y)] > U^0(w)$ . Let  $O := \{i : \alpha_i > 0\}$  and  $N := \{i : \alpha_i = 0\}$  so that  $O$  is the set of “owners” and  $N$  the set of “nonowners.” By budget balance,  $0 \leq |N| < I$ .

Construct a technology  $A \supseteq A^0$  as follows. Let  $A_i := A_i^0 \cup \{a_i^*\}$  with  $c_i(a_i^*) = 0$  for all  $i \in O$ , and let  $A_i := A_i^0$  for all  $i \in N$ . Let  $F(a) = G$  for all  $a \in A$  such that  $a_O = a'_O$ . We will complete the construction of  $A$  so that in every equilibrium  $\sigma$ , the owners play  $a'_O$  and the nonowners play (possibly mixed) zero-cost actions. Then  $F(\sigma) = G$  as desired.

We define the outcome distributions for the remaining action profiles as follows. Let  $0 < \varepsilon_{|N|+1} < \varepsilon_{|N|+2} < \dots < \varepsilon_I := \mathbb{E}_G[v(y)] - U^0(w)$ , to be used to provide strict incentives. We assume that each  $\varepsilon_k$  is small enough to satisfy the finitely many restrictions imposed on it by (A.5) below. Fix any  $a \in A \setminus A^0$ . Let  $J := \{i \in O : a_i = a'_i\} \cup N$ . Then  $|J| > |N|$  because at least one owner  $i$  plays  $a'_i$  in all the profiles in  $A \setminus A^0$ . Let

$$E(a) := \max_{\hat{a}_J \in A_J^0} \left( \mathbb{E}_{F(\hat{a}_J, a_{-J})}[v(y)] - \sum_{j \in J} \frac{c_j(\hat{a}_j)}{\alpha_j} \right). \tag{A.4}$$

Note that  $E(a)$  depends only on the profile  $a_{-J}$  (i.e., on the actions of the agents in  $O \setminus J$ ), and it is nonnegative and well-defined even if  $N \neq \emptyset$  as the maximum selects zero-cost actions for all  $j \in N$ . We define the output distribution for  $a$  to be some  $F(a) \in \Delta(Y)$  such that

$$\mathbb{E}_{F(a)}[v(y)] = \begin{cases} E(a) + \varepsilon_{|J|} & \text{if } E(a) < \max v(Y), \\ E(a) & \text{if } E(a) = \max v(Y), \end{cases} \tag{A.5}$$

where the second case is a nuisance that has to be accounted for as  $v(y)$  is bounded. (Note that in the first case, feasibility requires  $\varepsilon_{|J|} \leq \max v(Y) - E(a)$ ; letting  $a$  range over  $A \setminus A^0$  then defines at most finitely many inequalities involving  $\varepsilon_{|J|}$  as we commented above.)

We verify that this construction is consistent with us already having defined  $F(a) = G$  for all  $a \in A$  with  $a_O = a'_O$ . Suppose  $|J| = I$  so that the profile  $a$  above has  $a_O = a'_O$ . Then (A.4) gives  $E(a) = U^0(w) < \mathbb{E}_G[v(y)] \leq \max v(Y)$ . Thus, by (A.5) the distribution  $F(a)$  satisfies  $\mathbb{E}_{F(a)}[v(y)] = E(a) + \varepsilon_I = U^0(w) + \varepsilon_I = \mathbb{E}_G[v(y)]$  by definition of  $\varepsilon_I$ .

LEMMA A.1: *Suppose  $a \in A$  is such that  $c_j(a_j) = 0$  for all  $j \in N$ . Then  $u_i(a'_i, a_{-i}) \geq u_i(a)$  for all  $i \in O$ . Moreover, if  $a_O \neq a'_O$ , then the inequality is strict for some  $i \in O$ .*

PROOF: Consider agent  $i \in O$ . Suppose first that  $a \in A^0$  and  $c_j(a_j) = 0$  for all  $j \in N$ . Then

$$\begin{aligned} u_i(a) &= \alpha_i \left( \mathbb{E}_{F(a)}[v(y)] - \frac{c_i(a_i)}{\alpha_i} - \sum_{j \in N} \frac{c_j(a_j)}{\alpha_j} \right) \\ &\leq \alpha_i \min \{ E(a'_i, a_{-i}) + \varepsilon_{|N|+1}, \max v(Y) \} \\ &= \alpha_i \mathbb{E}_{F(a'_i, a_{-i})}[v(y)] = u_i(a'_i, a_{-i}), \end{aligned} \tag{A.6}$$

where the first equality follows as  $c_j(a_j) = 0$  for all  $j \in N$ , the inequality uses (A.4) with  $J = N \cup \{i\}$ , and the penultimate equality is by (A.5). To see that the inequality is strict for some  $i \in O$ , note that if equality holds in (A.6), then the minimum must select the second case as  $\varepsilon_{|N|+1} > 0$ , which implies that  $\mathbb{E}_{F(a)}[v(y)] = \max v(Y)$  and  $c_i(a_i) = 0$ . Therefore, if this holds for all  $i \in O$ , then the profile  $a \in A^0$  yields  $\max v(Y)$  at zero cost, implying that  $U^0(w) = \max v(Y)$ , which contradicts  $\mathbb{E}_G[v(y)] > U^0(w)$ .

Assume then that  $a \in A \setminus A^0$ , with  $c_j(a_j) = 0$  for all  $j \in N$ . If  $a_i = a'_i$ , then the claim is true trivially, so suppose there is an agent  $i \in O$  with  $a_i \in A_i^0$ . Then  $i \notin J$  and (A.5) gives

$$u_i(a) = \alpha_i \left( \mathbb{E}_{F(a)}[v(y)] - \frac{c_i(a_i)}{\alpha_i} \right) = \alpha_i \left( \min \{ E(a) + \varepsilon_{|J|}, \max v(Y) \} - \frac{c_i(a_i)}{\alpha_i} \right). \tag{A.7}$$

Suppose first that the minimum in (A.7) is attained only by the first case. Then (A.4) implies that the far right-hand side of (A.7) equals

$$\begin{aligned}
 & \alpha_i \left( \max_{\hat{a}_J \in A_J^0} \left( \mathbb{E}_{F(\hat{a}_J, a_{-J})} [v(y)] - \sum_{j \in J} \frac{c_j(\hat{a}_j)}{\alpha_j} \right) + \varepsilon_{|J|} - \frac{c_i(a_i)}{\alpha_i} \right) \\
 & < \alpha_i \min \left\{ \max_{\hat{a}_{J \cup \{i\}} \in A_{J \cup \{i\}}^0} \left( \mathbb{E}_{F(\hat{a}_{J \cup \{i\}}, a_{-J \cup \{i\}})} [v(y)] - \sum_{j \in J \cup \{i\}} \frac{c_j(\hat{a}_j)}{\alpha_j} \right) + \varepsilon_{|J \cup \{i\}|}, \max v(Y) \right\} \\
 & = \alpha_i \min \{ E(a'_i, a_{-i}) + \varepsilon_{|J \cup \{i\}|}, \max v(Y) \} \\
 & = \alpha_i \mathbb{E}_{F(a'_i, a_{-i})} [v(y)] \\
 & = u_i(a'_i, a_{-i}), \tag{A.8}
 \end{aligned}$$

where the inequality is strict because  $\varepsilon_{|J|} < \varepsilon_{|J \cup \{i\}|}$ , the equality that follows it is by (A.4), and the penultimate equality is by (A.5). Thus,  $u_i(a) < u_i(a'_i, a_{-i})$ .

If instead the minimum in (A.7) is attained by the second case, then  $E(a) = \max v(Y)$  by (A.5), and (A.4) implies that the far right-hand side of (A.7) equals

$$\begin{aligned}
 \alpha_i \left( \max v(Y) - \frac{c_i(a_i)}{\alpha_i} \right) & = \alpha_i \left( \max_{\hat{a}_J \in A_J^0} \left( \mathbb{E}_{F(\hat{a}_J, a_{-J})} [v(y)] - \sum_{j \in J} \frac{c_j(\hat{a}_j)}{\alpha_j} \right) - \frac{c_i(a_i)}{\alpha_i} \right) \\
 & \leq \alpha_i \max_{\hat{a}_{J \cup \{i\}} \in A_{J \cup \{i\}}^0} \left( \mathbb{E}_{F(\hat{a}_{J \cup \{i\}}, a_{-J \cup \{i\}})} [v(y)] - \sum_{j \in J \cup \{i\}} \frac{c_j(\hat{a}_j)}{\alpha_j} \right) \\
 & \leq \alpha_i \min \{ E(a'_i, a_{-i}) + \varepsilon_{|J \cup \{i\}|}, \max v(Y) \} \\
 & = \alpha_i \mathbb{E}_{F(a'_i, a_{-i})} [v(y)] \\
 & = u_i(a'_i, a_{-i}), \tag{A.9}
 \end{aligned}$$

where the second inequality uses (A.4), and the penultimate equality is by (A.5). Therefore,  $u_i(a) \leq u_i(a'_i, a_{-i})$ .

It remains to show that the inequality is strict for some  $i \in O$  if  $a_O \neq a'_O$ . Because (A.8) contains a strict inequality, agent  $i \in O$  with  $a_i \neq a'_i$  can be indifferent between  $a_i$  and  $a'_i$  only if (A.9) holds as a chain of equalities. Suppose this is the case. Because  $\varepsilon_{|J \cup \{i\}|} > 0$ , then the minimum on the third line in (A.9) must select the second case, which together with the left-hand side of (A.9) implies that  $c_i(a_i) = 0$ . Therefore, if  $a_O \neq a'_O$  and every agent  $i \in O$  is indifferent between  $a_i$  and  $a'_i$ , then  $c_i(a_i) = 0$  for all  $i \notin J$ . Furthermore, the case covered by (A.9) only arises if  $E(a) = \max v(Y)$ , and thus (A.4) implies that we have

$$\begin{aligned}
 \max v(Y) & = \max_{\hat{a}_J \in A_J^0} \left( \mathbb{E}_{F(\hat{a}_J, a_{-J})} [v(y)] - \sum_{j \in J} \frac{c_j(\hat{a}_j)}{\alpha_j} \right) \\
 & = \max_{\hat{a}_J \in A_J^0} \left( \mathbb{E}_{F(\hat{a}_J, a_{-J})} [v(y)] - \sum_{j \in J} \frac{c_j(\hat{a}_j)}{\alpha_j} - \sum_{i \notin J} \frac{c_i(a_i)}{\alpha_i} \right) \leq U^0(w) < \mathbb{E}_G[v(y)],
 \end{aligned}$$

where the second equality follows because  $c_i(a_i) = 0$  for all  $i \notin J$ , and the weak inequality is by definition of  $U^0(w)$ . Thus,  $\mathbb{E}_G[v(y)] > \max v(Y)$ , a contradiction. Q.E.D.

To complete the proof of Lemma 4.3, note that all agents in  $N$  can clearly only play zero-cost actions with positive probability in any equilibrium because their shares are zero. Consider any mixed strategy profile  $\sigma$  where this is true. Suppose  $\sigma_j(a'_j) < 1$  for some agent  $j \in O$ . Then some profile  $a \in A$  with  $a_O \neq a'_O$  is realized with positive probability under  $\sigma$ . By Lemma A.1, there exists an agent  $i \in O$  (who could be agent  $j$ ) for whom  $a'_i$  is a strictly better response to  $a_{-i}$  than  $a_i$ . Moreover, Lemma A.1 also shows that  $a'_i$  is a weakly better response than any  $a_i^0 \in A_i^0$  against all  $\tilde{a}_{-i} \in A_{-i}$  such that  $c_k(\tilde{a}_k) = 0$  for all  $k \in N$ . Therefore, playing  $a'_i$  with probability 1 is a profitable deviation from  $\sigma_i$  for agent  $i$ , and thus  $\sigma$  is not an equilibrium. We conclude that all  $\sigma \in \mathcal{E}(w, A)$  have  $\sigma_i(a'_i) = 1$  for all  $i \in O$ . Q.E.D.

### A.3. Proofs for Section 5

PROOF OF LEMMA 5.1: That  $V(w)$  is not less than the minimum is shown in the main text.

To prove the converse, note that the feasible set in (5.2) is compact, so the minimum is achieved at some  $G^*$ . Let  $\pi := \mathbb{E}_{G^*}[v(y) - \bar{w}(y)]$ . We show below that  $\bar{U}^0(w) < \max \bar{w}(Y)$ . Thus, we can approximate  $G^*$  with a sequence  $(G^n)$  such that  $\mathbb{E}_{G^n}[\bar{w}(y)] > \bar{U}^0(w)$  and  $\mathbb{E}_{G^n}[v(y) - \bar{w}(y)] \rightarrow \pi$  as the objective is continuous in  $G$ .<sup>14</sup> Interpreting  $w$  as a budget balanced linear contract in a model where  $v(y) = \bar{w}(y)$ , Lemma 4.3 implies that for each  $G^n$  there exists a technology  $A^n \supseteq A^0$  for which  $G^n$  is the unique equilibrium output distribution. Hence,  $V(w) \leq V(w, A^n) = \mathbb{E}_{G^n}[v(y) - \bar{w}(y)] \rightarrow \pi$  as desired.

To show that  $\bar{U}^0(w) < \max \bar{w}(Y)$ , suppose to the contrary that  $\bar{U}^0(w) = \max \bar{w}(Y)$ . The definition of  $\bar{U}^0(w)$  then implies that there exists a profile  $a \in A^0$  such that  $c(a) = 0$  and  $\text{supp } F(a) \subseteq \arg \max_{y \in Y} \bar{w}(y) = Y^*$ , where the equality holds because  $w$  aligns the agents' interests. Thus,  $V(w) < V(0)$  by Lemma 3.4(i), contradicting the eligibility of  $w$ .

It remains to show that any minimizer satisfies the constraint with equality. Let  $G^*$  be a minimizer. Because  $w$  is eligible, we have  $V(w) = \mathbb{E}_{G^*}[v(y) - \bar{w}(y)] > 0$ . Observe that if  $\mathbb{E}_{G^*}[\bar{w}(y)] > \bar{U}^0(w)$ , then the mixture  $G := (1 - \varepsilon)G^* + \varepsilon\delta_{y_0}$  is feasible for any  $\varepsilon > 0$  small enough. But  $v(y_0) - \bar{w}(y_0) = -\bar{w}(y_0) \leq 0$ , which in turns implies that we have  $\mathbb{E}_G[v(y) - \bar{w}(y)] \leq (1 - \varepsilon)V(w) < V(w)$ , which contradicts  $G^*$  being a minimizer. We conclude that  $\mathbb{E}_{G^*}[\bar{w}(y)] = \bar{U}^0(w)$ . Q.E.D.

PROOF OF LEMMA 5.4: We adapt the proof of Lemma 3 in Carroll (2015) to the present setting. Let  $B := \text{co}(\{(\bar{w}(y), v(y) - \bar{w}(y)) : y \in Y\})$  and  $C := \{(u, v) : u > \bar{U}^0(w) \text{ and } v < V(w)\}$ . Then  $B$  and  $C$  are disjoint subsets of  $\mathbb{R}^2$  by Lemma 5.1. Thus, by the separating hyperplane theorem, there exist numbers  $\kappa, \lambda$ , and  $\mu$ , with  $(\lambda, \mu) \neq (0, 0)$ , such that

$$\kappa + \lambda u - \mu v \leq 0 \quad \text{for all } (u, v) \in B, \tag{A.10}$$

$$\kappa + \lambda u - \mu v \geq 0 \quad \text{for all } (u, v) \in C. \tag{A.11}$$

Furthermore, letting  $G^*$  be some distribution that achieves the minimum in (5.2), the point  $(\mathbb{E}_{G^*}[\bar{w}(y)], \mathbb{E}_{G^*}[v(y) - \bar{w}(y)])$  lies in the closure of both  $B$  and  $C$ , and we thus have

$$\kappa + \lambda \mathbb{E}_{G^*}[\bar{w}(y)] - \mu \mathbb{E}_{G^*}[v(y) - \bar{w}(y)] = 0. \tag{A.12}$$

<sup>14</sup>For example, take  $G^n$  to be the mixture  $(1 - \frac{1}{n})G^* + \frac{1}{n}\delta_{\bar{y}}$  for some  $\bar{y} \in \arg \max_{y \in Y} \bar{w}(y)$ .

We will show that  $\lambda > 0$  and  $\mu > 0$ . Inequality (A.11) implies that  $\lambda \geq 0$  and  $\mu \geq 0$ , so it suffices to show that these inequalities are strict.

Suppose toward contradiction that  $\mu = 0$ . Then  $\lambda > 0$ , and inequalities (A.10) and (A.11) imply  $\max_{y \in Y} \bar{w}(y) \leq -\kappa/\lambda \leq \bar{U}^0(w)$ . Thus,  $\bar{U}^0(w) = \max \bar{w}(Y)$ . But in the proof of Lemma 5.1 above it is shown that this contradicts the eligibility of  $w$ . On the other hand, if  $\lambda = 0$ , then  $\mu > 0$ , and (A.10) and (A.11) imply  $\min_{y \in Y} (v(y) - \bar{w}(y)) \geq \kappa/\mu \geq V(w)$ . But  $\min_{y \in Y} (v(y) - \bar{w}(y)) \leq (v(y_0) - \bar{w}(y_0)) \leq 0$ , so  $V(w) \leq 0$ , contradicting eligibility of  $w$ .

Now rescale  $(\kappa, \lambda, \mu)$  so that  $\mu = 1$ . Then (A.10) and (A.12) imply (5.4) and (5.5). Q.E.D.

#### A.4. Proof for Section 6

**PROOF OF THEOREM 6.1:** Fix a contract  $w : Y \rightarrow \mathbb{R}_+^I$  that fails to align the agents' interests in utilities. Then the contract  $\tilde{w} := (\hat{u}_1 \circ w_1, \dots, \hat{u}_I \circ w_I) : Y \rightarrow \mathbb{R}_+^I$  fails to align the agents' interests in the equivalent model with risk-neutral agents where the principal's payoff is given by  $v(y) - \sum_i \hat{u}_i^{-1}(\tilde{w}_i(y))$ . Furthermore, because each function  $\hat{u}_i$  is increasing, we have  $Y^* := \cap_i \arg \max_{y \in Y} \tilde{w}_i(y) = \cap_i \arg \max_{y \in Y} w_i(y)$ .

Suppose first that for all  $a \in A^0$ ,  $\text{supp } F(a) \subseteq Y^*$  implies  $c(a) \neq 0$ . Applying Lemma 3.5 to  $\tilde{w}$  gives a sequence of technologies  $A^n \supseteq A^0$  with unique equilibrium output distributions  $F^n \rightarrow \delta_{y_0}$ . Thus,  $V(\tilde{w}) \leq \mathbb{E}_{F^n}[v(y) - \sum_i \hat{u}_i^{-1}(\tilde{w}_i(y))] \leq \mathbb{E}_{F^n}[v(y)] \rightarrow 0$ . Moreover, if  $w$  is budget balanced, then limited liability implies  $w_i(y_0) = 0$  for all  $i$ . We thus have  $S(\tilde{w}) \leq \mathbb{E}_{F^n}[\sum_i \tilde{w}_i(y)] = \mathbb{E}_{F^n}[\sum_i \hat{u}_i(w_i(y))] \rightarrow \sum_i \hat{u}_i(w_i(y_0)) = 0$ . That is, this case is covered by Lemma 3.5 just like in the risk-neutral case.

For the rest of the proof, suppose there exists  $a^* \in A^0$  such that  $\text{supp } F(a^*) \subseteq Y^*$  and  $c(a^*) = 0$ . Then Lemma 3.4(i) applies to  $\tilde{w}$  verbatim because, by inspection, the cost of payments to the principal only appears in the proof in display (A.1), which clearly still holds when the principal's payoff is  $v(y) - \sum_i \hat{u}_i^{-1}(\tilde{w}_i(y))$ . Thus,  $V(\tilde{w}) \leq V(0)$ .

It remains to handle the case where  $w$  is budget balanced. Denote by  $(\bar{u}_1, \dots, \bar{u}_I)$  the maximum payment-utility levels under contract  $w$ , that is,  $\bar{u}_i := \hat{u}_i(w_i(y))$  for all  $y \in Y^*$ . Define a budget balanced contract  $w'$  as follows. Let  $y \in Y$ . Take  $\lambda \in [0, 1]$  such that

$$\sum_i \hat{u}_i^{-1}(\lambda \bar{u}_i) = v(y) \in [0, \max v(Y)].$$

To see that this is feasible, note that the left-hand side equals 0 when  $\lambda = 0$  as each  $u_i$  is an increasing function with  $u_i(0) = 0$ . On the other hand, when  $\lambda = 1$  the left-hand side equals  $\sum_i \hat{u}_i^{-1}(\bar{u}_i) = \sum_i \max w_i(Y) = \max v(Y)$  by definition of  $\bar{u}_i$  and budget balance, which implies that  $Y^* = \arg \max_y v(y)$  whenever the former is nonempty. Thus, equality holds for some  $\lambda \in [0, 1]$  by the intermediate value theorem. Now define  $w'(y)$  by letting  $w'_i(y) = \hat{u}_i^{-1}(\lambda \bar{u}_i)$  for all  $i$ . The contract  $w'$  so defined is budget balanced by construction. It also aligns the agents' interests in utilities as for each  $y \in Y$ , we have  $\hat{u}_i(w'_i(y)) = \lambda \bar{u}_i$  for all  $i$  for some constant  $\lambda \in [0, 1]$  that is independent of  $i$ .

To finish the proof, we observe that  $a^* \in \mathcal{E}(w', A)$  for all  $A \supseteq A^0$ , because under  $a^*$  the output is sure to be in  $Y^*$  and  $c(a^*) = 0$ , giving each agent  $i$  the maximal feasible payoff of  $\bar{u}_i$ . Thus,  $S(w') = \sum_i \bar{u}_i \geq S(w)$ . Q.E.D.

#### REFERENCES

ABREU, DILIP, AND HITOSHI MATSUSHIMA (1992): "Virtual Implementation in Iteratively Undominated Strategies: Complete Information," *Econometrica*, 60 (5), 993–1008. [1591]

- BERGEMANN, DIRK, AND STEPHEN MORRIS (2005): “Robust Mechanism Design,” *Econometrica*, 73 (6), 1771–1813. [1585]
- CARROLL, GABRIEL (2015): “Robustness and Linear Contracts,” *American Economic Review*, 105 (2), 536–563. [1583-1585,1599-1601,1603,1611]
- (2017): “Robustness and Separation in Multi-Dimensional Screening,” *Econometrica*, 85 (2), 453–488. [1585]
- CARROLL, GABRIEL, AND ILYA SEGAL (2019): “Robustly Optimal Auctions With Unknown Resale Opportunities,” *Review of Economic Studies*, 86 (4), 1527–1555. [1585]
- CARROLL, GABRIEL, AND DANIEL WALTON (2021): “A General Framework for Robust Contracting Models,” Discussion paper, University of Toronto. [1585]
- CHASSANG, SYLVAIN (2013): “Calibrated Incentive Contracts,” *Econometrica*, 81 (5), 1935–1971. [1585,1601]
- CHE, YEON-KOO, AND SEUNG-WEON YOO (2001): “Optimal Incentives for Teams,” *American Economic Review*, 91 (3), 525–541. [1585]
- CHUNG, KIM-SHAU, AND JEFFREY ELY (2007): “Foundations of Dominant-Strategy Mechanisms,” *Review of Economic Studies*, 74 (2), 447–476. [1585]
- DAI, TIANJIAO, AND JUUSO TOIKKA (2018): “Robust Incentives for Teams,” Discussion paper, Massachusetts Institute of Technology. [1586]
- FRANKEL, ALEXANDER (2014): “Aligned Delegation,” *American Economic Review*, 104 (1), 66–83. [1585]
- GARICANO, LUIS, AND TANO SANTOS (2004): “Referrals,” *American Economic Review*, 94 (3), 499–525. [1585, 1605]
- GARRETT, DANIEL (2014): “Robustness of Simple Menus of Contracts in Cost-Based Procurement,” *Games and Economic Behavior*, 87, 631–641. [1585]
- HACKMAN, RICHARD (2002): *Leading Teams: Setting the Stage for Great Performances*. Boston, MA: Harvard Business School Press. [1585]
- HOLMSTRÖM, BENGT (1982): “Moral Hazard in Teams,” *The Bell Journal of Economics*, 13 (2), 324–340. [1583]
- HOLMSTRÖM, BENGT, AND PAUL MILGROM (1987): “Aggregation and Linearity,” *Econometrica*, 55 (2), 303–328. [1585]
- HURWICZ, LEONID, AND LEONARD SHAPIRO (1978): “Incentive Structures Maximizing Residual Gain Under Incomplete Information,” *The Bell Journal of Economics*, 9 (1), 180–191. [1585]
- ITOH, HIDESHI (1991): “Incentives to Help in Multi-Agent Situations,” *Econometrica*, 59 (3), 611–636. [1605]
- KAMBHAMPATI, ASHWIN (2022): “Robust Performance Evaluation,” Discussion paper, United States Naval Academy. [1585]
- LAZEAR, EDWARD (1989): “Pay Equality and Industrial Politics,” *Journal of Political Economy*, 97 (3), 561–580. [1584,1590]
- LEVIN, JONATHAN, AND STEVEN TADELIS (2005): “Profit Sharing and the Role of Professional Partnerships,” *Quarterly Journal of Economics*, 120 (1), 131–171. [1585]
- MADARÁSZ, KRISTÓF, AND ANDREA PRAT (2017): “Sellers With Misspecified Models,” *Review of Economic Studies*, 84 (2), 790–815. [1585]
- MARKU, KELER, AND SERGIO OCAMPO DIAZ (2019): “Robust Contracts in Common Agency,” Discussion paper, University of Minnesota. [1585]
- MONDERER, DOV, AND LLOYD SHAPLEY (1996): “Potential Games,” *Games and Economic Behavior*, 14, 124–143. [1598]
- YAMASHITA, TAKURO (2015): “Implementation in Weakly Undominated Strategies: Optimality of Second-Price Auction and Posted-Price Mechanism,” *Review of Economic Studies*, 82 (3), 1223–1246. [1585]

---

Co-editor Dirk Bergemann handled this manuscript.

*Manuscript received 20 April, 2018; final version accepted 9 January, 2022; available online 25 January, 2022.*