Robust Incentives for Teams*

Tianjiao Dai† Juuso Toikka‡

April 20, 2018

Abstract

This paper considers team incentive schemes that are robust to nonquantifiable uncertainty about the game played by the agents. A principal designs a contract for a team of agents, each taking an unobservable action that jointly determine a stochastic contractible outcome. The game is common knowledge among the agents, but the principal only knows some of the available action profiles. Realizing that the game may be bigger than he thinks, the principal evaluates contracts based on their guaranteed performance across all games consistent with his knowledge. All parties are risk neutral and the agents are protected by limited liability.

A contract is said to align the agents' interests if each agent's compensation covaries positively and linearly with the other agents' compensation. It is shown that contracts that fail to do so are dominated by those that do, both in terms of the surplus guarantee under budget balance, and in terms of the principal's profit guarantee when he is the residual claimant. It thus suffices to base compensation on a one-dimensional aggregate even if richer outcome measures are available. The best guarantee for either objective is achieved by a contract linear in the monetary value of the outcome. This provides a foundation for practices such as team-based pay and profit-sharing in partnership.

1 Introduction

Much of economic activity is performed by teams, broadly defined to encompass groups of agents such as partnerships, committees, research groups or start-ups, and work teams in manufacturing and services. The classical contract-theoretic approach to motivating such

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

†MIT, Department of Economics, daitj207@mit.edu

‡MIT, Department of Economics, toikka@mit.edu

teams, pioneered by Holmström (1982), emphasizes informational aspects of the problem. It holds that any signal informative of an agent's action should optimally be used to determine his compensation. This leads to contracts that are sophisticated and highly context-dependent. Moreover, there is no reason for compensation to be team-based, except in the extreme case where only the team's aggregate performance is observable. These predictions are at odds with incentive schemes typically observed in practice, which tend to be simpler and often include team-based pay even if information about individual performance is available. For instance, partnerships commonly operate under a simple profit-sharing agreement, firms use team incentives to motivate employees, and academic economists share credit equally for co-authored 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 Holmström's (1982) team production problem, where each agent takes an unobservable action at a private cost, and the profile of actions stochastically determines a contractible outcome that may convey information about both aggregate as well as individual performance. We assume that all parties are risk-neutral and the agents are protected by limited liability, but impose no particular structure on the 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 he only knows some of the action profiles in the game. Realizing that the game may be bigger than he thinks, but not having a prior on the set of possible games, the principal evaluates contracts based on their guaranteed performance across all games consistent with his 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 own profit if he is the residual claimant, requires that a contract align the agents' interests. In particular, each agent's compensation should covary positively and linearly with the compensation of all other agents. Such a contract has a natural representation in terms of a one-dimensional aggregate of the outcome, the value of which determines everyone's compensation, so we can reasonably interpret the contract as providing team-based compensation. Contracts of this form dominate all other contracts. Thus, team-based compensation is optimal even though richer measures of performance may be available.

The necessity of interest alignment derives from the fact that when a contract induces disagreement about the ranking of outcomes among the agents, then—should the game provide the opportunity for it—an agent may seek personal gain at the others' expense. We show that it is possible to find games where such selfish actions create a "race to the bottom," with the unique equilibrium leading to the worst possible outcome. We illustrate the basic intuition in the context of a rank-order tournament after having introduced the model. While the result is reminiscent of Carroll's (2015) linearity result for principal-agent problems, the two are logically distinct: the definition of interest alignment only involves the payments to the agents, so every contract trivially aligns the agent's interests in the single-agent case.

If a contract that aligns the agents' interests is budget balanced among the agents, then it is in fact a linear contract where each agent is paid a fixed share of the monetary value generated by the contractible outcome. We show that some such linear contract achieves the best possible surplus guarantee within the class of budget-balanced contracts. This provides a possible foundation for profit-sharing agreements in partnerships.

We also show that a linear contract achieves the best possible guarantee for the principal's profit in the case where the principal is the residual claimant for the team's output. By our first result, the search for the principal-optimal contract can be restricted to contracts that align the agents' interests. Moreover, the candidate optimal contracts can be represented as consisting of a function specifying the agents' total compensation for each outcome, and of shares that determine how it is divided amont the agents. By holding the shares fixed, we show that the total compensation should be a linear function of the monetary value generated by the outcome, and so the contract should be linear overall. Heuristically, a contract that aligns the agents' interests ensures that no agent can seek personal gain at the expense of the other agents. Requiring that this not happen at the expense of the principal, either, implies that the agents' compensation must covary linearly with the principal's payoff as well, leading to a linear contract.

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 generate a positive surplus. The condition that characterizes known production technologies for which the optimal guarantees are non-trivial comprises of a virtual surplus calculation: a social planner should be able to generate positive surplus in a model where the agents' costs are appropriately inflated to account for the robustness concern. So the theory here predicts that, even absent setup costs, only sufficiently profitable teams are worth forming.

While the results are the strongest under risk-neutrality, we show in Section 6 that non-trivial performance guarantees require team-based compensation also when the 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 behind interest alignment and team-based pay holds irrespective of the agents' risk attitudes. Even linearity can be recovered if the agents' preferences over monetary lotteries are given by a class of symmetric CRRA preference

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 (2017), and Marku and Ocampo Diaz (2017). 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).

Other theoretical explanations have been put forth for the use of profit-sharing, and for the prevalence of partnerships as an organizational form in the professional services industry—see, for example, Garicano and Santos (2004) or Levin and Tadelis (2005). These papers either exogenously restrict the contract space, or solve for the optimal contract in particular parametric models. Che and Yoo (2001) show that team-based compensation can be a part of the optimal mix of formal and relational incentives in a repeated partnership problem where the agents can observe each others' actions. Our work provides a complementary perspective, showing that team-based compensation arises as a robustly optimal contract in a static setting where the agents cannot monitor each other.

Finally, there is an extensive management literature on teams. The result that contracts should align the agents' interests connects with some of the themes in this literature. For example, Hackman (2002) posits that one of the key enabling conditions 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 outcome, a contract that aligns the agents' interests provides just that.[1]

2 Model

We consider the problem of a principal incentivizing a team of agents, indexed $i = 1, \ldots, I$. Each agent takes an unobservable action a_i from a finite set A_i at a private cost $c_i(a_i) \geq 0$.

[1] This is true quite literally: the parameter d in our Definition 1 is the direction of the ray in $\mathbb{R}^I_+$ along which all payment profiles lie.

The cost can be interpreted as monetary, or as simply describing the agent's preferences over the available actions. The resulting action profile $a = (a_1, \dots, a_I) \in A := \times_{i=1}^{I} A_i$ determines stochastically the team's observable output y, an element of a finite set Y of possible outcomes. The distribution of y given a is denoted $F(a) \in \Delta(Y)$. We refer to the tuple (A, c, F), where $c = (c_1, \dots, c_I) : A_1 \times \cdots \times A_I \to \mathbb{R}_+^I$ is the profile of cost functions and $F : A \to \Delta(Y)$ is the family of output distributions, as the *technology*.

The outcome y provides a measure of the team's performance, possibly along multiple dimensions, and serves as a signal about the agents' actions. Its intrinsic monetary 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 outcome and normalize its value to zero, i.e., $v(y_0) = \min v(Y) = 0$. (y_0 can be chosen arbitrarily among the minimizers if the worst outcome is not unique.) To avoid trivialities, we assume $\max v(Y) > 0$.

The agents do not have preferences over the outcomes per se, but the principal can guide them by designing an incentive scheme that rewards the agents based on the team's output. We assume that the agents are protected by limited liability, meaning that payments to them have to be non-negative. An incentive scheme, or a *contract*, is thus a function $w : Y \to \mathbb{R}_+^I$ that specifies a payment profile $w(y) = (w_1(y), \dots, w_I(y))$ for every possible outcome $y \in Y$. The net payoff of agent i 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. (We discuss risk-averse agents in Section 6.)

Given a contract w, the convex hull of all payment profiles is denoted $W := \text{co}(w(Y))$. We say that the contract w is *budget balanced* if the value of output is shared by the agents, i.e., if $\sum_i w_i(y) = v(y)$ for all y.

The principal designs the contract either to maximize total surplus subject to budget balance, or to maximize his profits. However, he does so without full knowledge of the game played by the agents. Specifically, inspired by Carroll (2015), we assume that the technology (A, c, F) is common knowledge among the agents, but the principal only knows some finite set $A^0 = \times_{i=1}^{I} A_i^0$ of action profiles with an associated profile of cost functions $c^0 : A^0 \to \mathbb{R}_+^n$ and outcome distributions $F^0 : A^0 \to \Delta(Y)$, collectively referred to as the *known technology*. The principal believes that the true technology may be any (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. (Note that the set of possible outcomes Y is held fixed; it is known by all parties.) To simplify notation, we suppress the cost functions and outcome distributions when this causes no confusion, writing A^0 and A for the known and the true technology, respectively.

Together a contract w and the (true) technology A induce a normal form game $\Gamma(w, A)$ between the agents. We let $\mathcal{E}(w, A)$ denote 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.[2] Thus, the expected total surplus induced by the contract w given technology A is

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

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

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

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 total surplus and profits, these guarantees are, respectively,

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

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 the guaranteed expected surplus satisfies $S(w) \geq -\sum_i \underline{c}_i^0$, where $\underline{c}_i^0 := \min c_i(A_i^0)$, since each agent can ensure a payoff of $-\underline{c}_i^0$ by playing the least-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.

The most immediate interpretation is that the principal is designing the contract for a single team, not fully aware of the game the agents are playing. For example, this uncertainty

[2] This minimizes the departure from the standard model and ensures the existence of an optimal contract. Essentially the same results obtain 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.

may reflect the agents' superior knowledge of the situation. Or it could be due to the principal having to design the contract before the details of the team's operating environment are known, or even who the team's members will be. Importantly, however, the principal can envision and evaluate all possible outcomes that may arise as a result of the teams activities, i.e., he knows the set Y and the mapping $v : Y \to \mathbb{R}$. The fact that Y is held fixed is not restrictive as our main findings do not require the output distributions in the known technology to have full support. Thus Y can contain outcomes that are impossible under the known technology. (We selectively invoke a full-support assumption to state stronger versions of some of our results.)

An alternative interpretation of the model is that the principal is designing a contract to be used in a number of different situations, perhaps by many different teams, and wants the contract to guarantee good expected performance in all of them. In case of the profit guarantee, an example might be a large firm utilizing multiple self-managing teams. Each team may face different situations and operating environments. The realized situation may be apparent to all team members (captured by the agents knowing A), but it may be too difficult and costly to communicate or verify all of this information about the circumstances to a third party for the contract to depend on it. And so the firm resorts to designing a contract based only on the aspects common to all situations (captured by A^0).

In case of the surplus guarantee for a budget-balanced team, we may interpret the principal as trying to design a standard contract to be employed across different kinds of partnerships. An example of such a contract is a profit-sharing agreement, used commonly in professional services. Here the known technology A^0 then captures the common aspects of the partnerships to which the standard contract or organizational form is to be applied.

3 Necessity of Interest Alignment

We start the analysis by showing that contracts that fail to align the agents' interests can be easily improved upon. We also show that, under mild additional assumptions about the known technology, interest alignment becomes necessary for a contract to have a nontrivial surplus or profit guarantee.

In order to motivate our definition of interest alignment and to develop intuition for the result, it is useful to first consider a contract under which the agents are in direct competition—a rank-order tournament. To this end, suppose for a moment that there are just two agents and that an outcome is a pair $y = (y_1, y_2)$ listing their outputs. Let Y be any finite grid on $\mathbb{R}^2_+$ containing the origin and at least one outcome with $y_1 > y_2$ and another with $y_2 > y_1$. The value to the principal is the sum $v(y) = y_1 + y_2$, and thus $y_0 = (0, 0)$.

	A_2^0	a_2'
A_1^0	$\cdots$	$-c_1(a_1), b$
a_1'	$b, -c_2(a_2)$	$b/2, b/2$

Figure 1. The game $\Gamma(w, A)$ for the tournament example. Since b is the highest feasible payoff, a_i' is a weakly dominant strategy. To see that a' is the unique equilibrium, fix a mixed strategy equilibrium σ. If the support of σ 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 σ_i for some agent i. But then a_{-i}' is the unique best response for agent $-i$, and thus σ_{-i} must play it with certainty. This in turn implies that σ_i must play a_i' with certainty. Therefore, σ is simply the pure-strategy profile a'.

A rank-order tournament is a contract specifying three payment levels: $w_i(y) = b > 0$ if $y_i > y_{-i}$, $w_i(y) = b/2$ if $y_i = y_{-i}$, and $w_i(y) = 0$ if $y_i < y_{-i}$. That is, the agent with the highest output gets a bonus b, which is shared equally in case of a tie.

Because of the form of the contract, agent 1 has an incentive to pursue actions that increase the likelihood that his output is (weakly) greater than agent 2's output, or $y_1 \geq y_2$. If the only way to do this is by increasing y_1, then the tournament does incentivize the agent to exert effort towards increasing total output.[3] But as pointed out by Lazear (1989), it is also in agent 1's interest to sabotage agent 2 to lower y_2, for example, by refusing to help. He may also try to claim credit for some of agent 2's output—or even outright steal it—to shift some of y_2 to y_1. To the extent that such actions distract from agent 1's productive efforts, they lead to lower total output. Indeed, if both agents can engage in such activities, even the best equilibrium for the principal may yield no output.

More formally, given any known technology A^0, consider a technology A where each agent has an additional zero-cost action a_i' so that $A_i = A_i^0 \cup \{a_i'\}$ for $i = 1, 2$. The action a_i' results deterministically in some outcome y^i that has agent i winning the tournament if agent $-i$ plays any action in A_{-i}^0 (i.e., $y_i^i > y_{-i}^i$). Think of a_i' as an activity that benefits agent i at the other agent's expense as discussed above. However, suppose that if both agents engage in this activity, then nothing is produced: the profile $a' = (a_1', a_2')$ leads to the outcome $y_0 = (0, 0)$ with certainty. It is easy to verify that a_i' is then a weakly dominant strategy for each agent in the game $\Gamma(w, A)$, and a' is the unique equilibrium—see Figure 1.

The principal's profit given technology A is $V(w, A) = v(y_0) - b = -b$, and thus the tournament's profit guarantee is negative: $V(w) \leq V(w, A) < 0$. The principal would be better off offering the zero contract. In fact, as the unique equilibrium of $\Gamma(w, A)$ yields y_0

[3] Holmström (1982) showed that for some specific choices of technology, a tournament is the optimal contract for a principal who knows the game played by the agents and designs the contract to maximize his expected payoff. The (sub-)optimality of the tournament in this sense plays no role in the example.

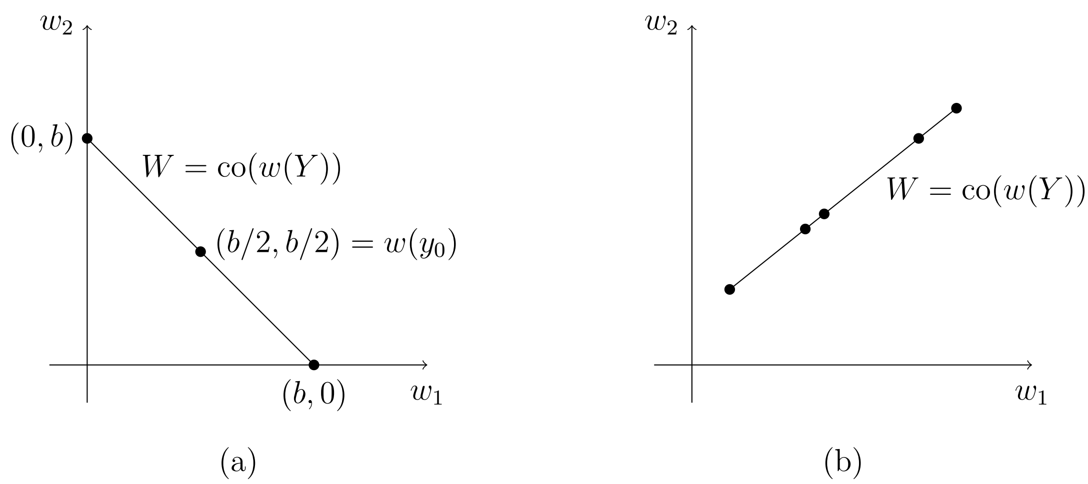


Figure 2. (a) The rank-order tournament. (b) A contract that aligns the agents' interests.

with certainty, the tournament's profit guarantee would be nonpositive even if rewarding the agents was costless to the principal (i.e., if his payoff was just $v(y)$).

To motivate our definition of interest alignment, it is useful to represent the above argument graphically. In Figure 2.a, the line segment between $(b, 0)$ and $(0, b)$ is the convex hull of payment profiles $W = \text{co}(\{(0, b), (b/2, b/2), (b, 0)\})$. The new action a_i' allows agent i to force at zero cost his most preferred point in W (i.e., $(0, b)$ or $(b, 0)$) if agent $-i$ plays any action in A_{-i}^0. As the expected payment profile $\mathbb{E}_{F(\sigma)}[w(y)]$ under any mixed strategy profile σ lies somewhere in W, at least one agent thus has a profitable deviation if the other agent's strategy puts positive probability only on known actions. This rules out equilibria with support in A^0. Finally, a' yields rewards $w(y_0) = (b/2, b/2)$ at the midpoint of the line segment; this point is better for each agent than the other agent's most preferred point, so a' is the unique equilibrium.

A rank-order tournament is special in that the agents' interests are in direct conflict: for any outcome distributions $F, G \in \Delta(Y)$, whenever $\mathbb{E}_F[w_1(y)] > \mathbb{E}_G[w_1(y)]$ so that agent 1 prefers F to G, we have $\mathbb{E}_G[w_2(y)] > E_F[w_2(y)]$ so that agent 2 prefers G to F. Graphically, this is equivalent to W being a downward-sloping line segment as in Figure 2.a so that the agents have opposite preferences over the points in W. However, it turns out that the same perverse incentives that undermine the tournament can arise in contracts that induce far less disagreement about the desirability of different outcome distributions. To completely rule out such disagreement, W must consist of a (weakly) increasing line segment as in Figure 2.b. This suggests the following definition.

Definition 1. A contract w *aligns the agents' interests* if all payment profiles lie on the same ray in $\mathbb{R}_+^I$, i.e., if $w(Y) \subset \{\underline{w} + dt : t \in \mathbb{R}_+\}$ for some $\underline{w}, d \in \mathbb{R}_+^I$.

A contract that does not satisfy the definition is said to *fail to align the agents' interests*. The tournament in Figure 2.a is an example. Note that any contract for which the interior of W is non-empty as in Figures 3 and 4 below fails to align the agents' interests.

Definition 1 is equivalent to the requirement that for all agents i and j and all outcome distributions $F, G \in \Delta(Y)$, $\mathbb{E}_F[w_i(y)] > \mathbb{E}_G[w_i(y)]$ implies $\mathbb{E}_F[w_j(y)] \geq \mathbb{E}_G[w_j(y)]$. That is, no two agents disagree on the ranking of any pair of outcome distributions, albeit one of the agent's preference may be strict and the other's weak (if the latter is globally indifferent). The equivalence follows by noting that each point in the convex hull of payment profiles W is the expected profile $\mathbb{E}_F[w(y)]$ for some $F \in \Delta(Y)$. Hence, the agents do not disagree on the ranking of distributions precisely when W, and thus $w(Y)$, lies along a ray in $\mathbb{R}^I_+$.

A second equivalent condition is the existence of outcomes $\overline{y}$ and $\underline{y}$ with $w(\overline{y}) \geq w(\underline{y})$ such that, for all $y \in Y$, we have $w(y) = (1-\lambda)w(\underline{y}) + \lambda w(\overline{y})$ for some $\lambda \in [0,1]$. The parameter $\lambda = \lambda(y)$ has a natural interpretation as a measure of the team's performance on a scale from zero to one. A contract that aligns the agents' interests thus prescribes team-based compensation in a strong sense: agents' payments covary positively and linearly.

Note that any constant contract satisfies Definition 1. For example, the zero contract aligns the agents' interests. It is also worth noting that the definition only concerns the agents, and so in general it is silent on how the payments relate to the value of the outcome. However, if the contract is also budget balanced, then interest alignment is equivalent to the agents dividing the value $v(y)$ amongst themselves according to some fixed shares.

Lemma 1. *A contract w is budget balanced and aligns the agents' interests if and only if there exists $\alpha = (\alpha_1, \ldots, \alpha_n) \in [0,1]^I$ 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 $\underline{y} = y_0$ and $\overline{y} \in \arg\max_y v(y)$ in the second equivalent condition above. Fixing y, we sum over i and use budget balance again to get $v(y) = \sum_i w_i(y) = (1-\lambda)\sum_i w_i(y_0) + \lambda \sum_i w_i(\overline{y}) = (1-\lambda)v(y_0) + \lambda v(\overline{y}) = \lambda v(\overline{y})$. Hence, $\lambda = v(y)/v(\overline{y})$. Noting that $w_i(y_0) = 0$ by limited liability and budget balance, we thus have $w_i(y) = (w_i(\overline{y})/v(\overline{y}))v(y)$, so taking $\alpha_i = w_i(\overline{y})/v(\overline{y})$ yields the result. $\square$

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 total surplus.

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

We prove Theorem 1 by finding for any contract that fails to align the agents' interests a (sequence of) game(s) with poor performance. The construction is more involved, but the basic idea is the same as in the tournament example: misalignment erodes a contract's guaranteed performance because, if given the opportunity, agents will seek personal gain at the expense of others, and this can lead to all equilibria being bad for the principal.

Before turning to the proof, we note that Theorem 1 can be strengthened under additional assumptions about the known technology to show that interest alignment is necessary in order to obtain any nontrivial performance guarantees.

We need the following definitions. An action profile $a \in A^0$ satisfies *full support* if $F(a) \neq \delta_{y_0}$ (where δ_{y_0} is the Dirac measure at y_0) implies that $F(a)$ has full support on Y. It satisfies *costly production* if $\mathbb{E}_{F(a)}[v(y)] > 0$ implies $c_i(a_i) > 0$ for some agent i. The following corollary shows that if each known action profile satisfies either of these, the worst case for any contract that fails to align the agents' interests is that no value is created.

Corollary 1. *Suppose every action profile in the known technology A^0 satisfies full support or costly production. If a contract w fails to align the agents' interests, then there exists a sequence of technologies $A^n \supseteq A^0$ such that*

$$\sup_{\sigma \in \mathcal{E}(w, A^n)} F(\{y \in Y : v(y) > 0\} | \sigma) \to 0.$$

The value of the team's equilibrium output converges to zero as $n \to \infty$, so the principal's profit is nonpositive in the limit (and it would be so even if the principal didn't have to pay the agents' compensation out of pocket). Moreover, the construction in the proof uses actions whose costs are no lower than the costs of the known actions, implying that the equilibrium total surplus converges to its theoretical lower bound. This establishes the following corollary.

Corollary 2. *Under the assumptions of Corollary 1, if a contract w fails to align the agents' interests, then $V(w) \leq 0$ and $S(w) \leq -\sum_i \underline{c}_i^0$.*

Corollary 2 shows that under relatively weak additional assumptions, contracts that fail to align the agents' interests are not only dominated in the sense of Theorem 1; they fail to improve on the trivial bounds both for profits and total surplus. In fact, if the contract is also budget balanced, an even weaker assumption will do.

Corollary 3. *Suppose the known technology A^0 does not contain an action profile a such that $c(a) = 0$ and $\operatorname{supp} F(a) \subseteq \arg\max_y v(y)$. If a budget balanced contract w fails to align the agents' interests, then $S(w) \leq -\sum_i \underline{c}_i^0$.*

That is, unless it is costless to produce the most valuable outcome(s) with certainty under the known technology, any budget balanced contract that fails to align the agents' interests has only the trivial surplus guarantee.

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

3.1 Proof of Theorem 1

We present here the proof of Theorem 1, relegating those of the corollaries to the Appendix. Throughout this section, fix a contract w that fails to align the agents' interests. Let

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

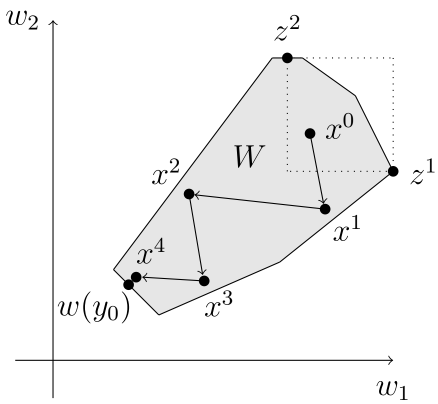
By definition, any $y \in Y^*$ simultaneously maximizes the payment to every agent. Graphically, this means that $w(y) \geq x$ for every $x \in W$. Note that Y^* may well be empty.

There are three cases to consider, corresponding to the following three lemmas. The first case is that the set Y^* is empty. This case is similar to the tournament example in that then no single point in W is the best point for all agents.

Lemma 2. *If $Y^* = \emptyset$, then there exists a sequence of technologies $A^n \supseteq A^0$, with unique equilibrium distributions $F^n \in \Delta(Y)$ and $\min c_i(A_i^n) = \underline{c}_i^0$ for all i, such that $F^n \to \delta_{y_0}$.*

The proof of this and that of the next lemma make use of the fact that the agents' payoffs depend on the outcome distribution $F(a) \in \Delta(Y)$ only through the expected payment profile $\mathbb{E}_{F(a)}[w(y)] \in W$. Conversely, any $x \in W$ is the expected payment profile for some $F \in \Delta(Y)$. Therefore, when constructing a technology A, as far as the agents' incentives are concerned, it suffices to specify the expected payment profiles $x(a) \in W$, $a \in A$.

Proof Sketch. For simplicity, we sketch the proof for the case of two agents, assuming further that the lowest known cost is zero for each agent (i.e., $\underline{c}_i^0 = 0$, $i = 1, 2$).



	A_2^0	a_2^1	a_2^2	a_2^3
A_1^0	$\cdots$	$\mathbf{z^2}$	u^0	u^0
a_1^1	$\mathbf{z^1}$	$\mathbf{x^0}$	u^1	u^1
a_1^2	u^0	$\mathbf{x^1}$	$\mathbf{x^2}$	u^2
a_1^3	u^0	u^1	$\mathbf{x^3}$	$\mathbf{x^4}$

Figure 3. A contract that fails to align the agents' interests with $Y^* = \emptyset$. The gray area is W, the convex hull of payment profiles. The payoff matrix represents the game $\Gamma(w, A)$ in the proof sketch for Lemma 2.

When $Y^* = \emptyset$ and $I = 2$, the set W is either a downward-sloping line segment as in Figure 2.a, or it has a non-empty interior as in Figure 3. The first case essentially reduces to the tournament example, so we focus on Figure 3.

Consider a technology A where $A_i = A_i^0 \cup \{a_i^1, a_i^2, a_i^3\}$ with $c_i(a_i^k) = 0$ for all i and k (so that each new action is a least-cost action) and where any $a \in A$ involving new actions is assigned an expected payment profile as specified in the right panel of Figure 3.

Note that z^i is (one of) agent i's most preferred point(s) in W. Such points z^1 and z^2 are necessarily distinct when $\arg\max_y w_1(y) \cap \arg\max_y w_2(y) = \emptyset$. Taking z^i to be the expected payment profile if agent i plays a_i^1 and the other agent plays any $a_{-i} \in A_{-i}^0$ eliminates equilibria in known actions (i.e., with support in A^0): in any such equilibrium some agent i would necessarily get less than z^i, and hence he could profitably deviate to a_i^1.

If $w(y_0)$ was in the dotted rectangle in Figure 3, we could then set $x(a_1^1, a_2^1) = w(y_0)$ with no need for actions a_i^2 and a_i^3. Then (a_1^1, a_2^1) would be the unique equilibrium of the game $\Gamma(w, (A_1^0 \cup \{a_1^1\}) \times (A_2^0 \cup \{a_2^1\}))$, similarly to the tournament example. However, when $w(y_0)$ lies outside the rectangle, as depicted here, this no longer works as at least one agent i prefers the other agent's most preferred point z^{-i} to $w(y_0)$. (They both do in Figure 3.)

Instead, we choose a sequence $(x^0, \ldots, x^4)$ in W as in Figure 3. That is, each agent i prefers x^0 to z^{-i}, and given any two adjacent elements of the sequence $(x^0, \ldots, x^4)$, agent 1 strictly prefers the odd one and agent 2 the even one. The last element, x^4, is chosen in the interior of W so that we can populate the remaining cells in the payoff matrix in Figure 3 with points $u^0 < u^1 < u^2$ such that $u^l < x^k$ for all l, k. (u^i are not shown in the left panel; they can be chosen in the gap between x^4 and $w(y_0)$ if x_2^3 is close enough to x_2^4.)

When the row(s) and column(s) corresponding to A_1^0 and A_2^0 have been eliminated—they are not necessarily strictly dominated, but no agent will play them with positive probability in any equilibrium—the remaining matrix is by construction dominance solvable, with (a_1^3, a_2^3) the unique outcome. Letting $x^4 \to w(y_0)$ thus yields a sequence of technologies whose unique equilibrium expected payment profiles converge to $w(y_0)$, and thus the distributions generating them can be taken to converge to δ_{y_0} as desired.

Note that the number of steps in the path $(x^0, \ldots, x^4)$ and the number of actions in the technology A is dictated by the shape of W, with a narrower set requiring more steps. But any points x and $x' < x$ in the interior of W can be connected by some such finite path. $\square$

The proof in the Appendix generalizes the above argument to I agents. When $I > 2$, there is an additional complication arising from the fact that even though $Y^* = \emptyset$, the intersection $\arg\max_{y \in Y} w_i(y) \cap \arg\max_{y \in Y} w_j(y)$ may still be nonempty for every pair of agents i and j. Because of this, we endow every agent with one new action to eliminate equilibria in known actions. The ensuing "race to the bottom" that drives the equilibrium outcome (close) to y_0 can then be constructed using just two players as above.

If $Y^* \neq \emptyset$, the projection of W to some pair of agents' payments is of the form depicted in Figure 4. (The interior relative to $\mathbb{R}_+^2$ is non-empty, since otherwise w would align the agents' interests.) There no longer exist distinct most preferred points for the agents that got the construction in the proof of Lemma 2 started. However, if no known zero-cost action profile maps deterministically to the point z in Figure 4, then we can still eliminate equilibria involving profiles in A^0 and drive the outcome to y_0 with essentially the same construction.

Lemma 3. *Let $Y^* \neq \emptyset$. Suppose that, for all $a \in A^0$, $\operatorname{supp} F(a) \subseteq Y^*$ implies $c(a) \neq 0$. Then there exists a sequence of technologies $A^n \supseteq A^0$, with unique equilibrium distributions $F^n \in \Delta(Y)$, such that $\min c_i(A_i^n) \to \underline{c}_i^0$ for all i and $F^n \to \delta_{y_0}$.*

Proof Sketch. We again assume for simplicity that $I = 2$ and $\underline{c}_i^0 = 0$, $i = 1, 2$.

Consider a technology A that assigns one new zero-cost action a_i^1 to each agent so that $A_i = A_i^0 \cup \{a_i^1\}$. Choose x^0, z^1, z^2 as in Figure 4, i.e., x^0 is in the interior of W and $z_i^i > x_i^0 > z_i^{-i}$. Let the expected payment profile be z^i if only agent i plays the new action a_i^1; let it be x^0 if both agents play the new action.

The profile a^1 is an equilibrium of the game $\Gamma(w, A)$, because a_i^1 is the unique best-response to a_{-i}^1 by construction. In fact, for x^0 close enough to z, it is the only equilibrium. To see this, choose x^0 close enough to z such that

$$x_1^0 + x_2^0 > \mathbb{E}_{F(a)}[w_1(y)] - c_1(a_1) + \mathbb{E}_{F(a)}[w_2(y)] - c_2(a_2) \quad \forall a \in A^0.$$

This is possible, because by assumption every $a \in A^0$ with $\mathbb{E}_{F(a)}[w(y)] = z$ has some agent

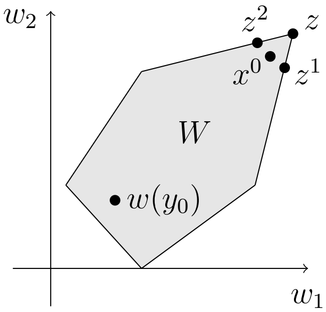


Figure 4. A contract that fails to align the agents' interests with $Y^* \neq \emptyset$.

playing a costly action, and A^0 is finite. The inequality implies that for all $a \in A^0$, we have $z_i^i > x_i^0 > \mathbb{E}_{F(a)}[w_i(y)] - c_i(a_i)$ for some agent i, who thus can profitably deviate to a_i^1. This rules out other pure-strategy equilibria. With some more work one can establish a^1 as the unique mixed-strategy equilibrium as well.

Having escaped the point z, we can now add actions $\{a_i^2, \dots, a_i^K\}$ to the technology A and use the construction in the proof of Lemma 2 to drive the equilibrium outcome to y_0. See the Appendix for the details. □

Finally, there remains the possibility that some known action profile $a^* \in A^0$ ensures that the outcome is in Y^* at no cost to the agents. Then a^* is an equilibrium for any technology $A \supseteq A^0$, and hence the contract w can potentially give a nontrivial profit or surplus guarantee. But a^* is also an equilibrium under the zero contract as well as under any budget balanced contract that aligns the agents' interests; such contracts can be shown to improve upon w. More precisely, we have the following lemma.

Lemma 4. *Suppose there exists $a \in A^0$ such that* $\operatorname{supp} F(a) \subseteq Y^*$ *and $c(a) = 0$. Then $V(w) < V(0)$. Moreover, if w is budget balanced, then $S(w) \leq S(w')$ for every contract w' that is budget balanced and aligns the agents' interests.*

We note for future reference that this lemma holds also for any contract w that aligns the agents' interests, different from the zero contract.

Proof. Let $a^* \in A^0$ satisfy the assumption in the lemma. Consider a technology A where $A_i = A_i^0 \cup \{a_i'\}$ and $c_i(a_i') = 0$ for all i. Let $F(a_i', a_{-i}) = F(a^*)$ for 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

$\operatorname{supp} F(a) \subseteq Y^*$. Hence,

$$V(w, A) \leq \max_{a \in A\,:\, c(a)=0 \text{ and } \operatorname{supp} F(a) \subseteq Y^*} \mathbb{E}_{F(a)}\Big[v(y) - \sum_i w_i(y)\Big]$$
$$= \max_{a \in A^0\,:\, c(a)=0 \text{ and } \operatorname{supp} F(a) \subseteq Y^*} \mathbb{E}_{F(a)}\Big[v(y) - \sum_i w_i(y)\Big] < \max_{a \in A^0\,:\, c(a)=0} \mathbb{E}_{F(a)}[v(y)] \leq V(0).$$

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 $y \in Y^*$ 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, $V(w) \leq V(w, A) < V(0)$, establishing the first part of the lemma.

For the second part, suppose that w is budget balanced so that $\sum_i w_i(y) = v(y)$ for all y. Then $Y^* \subseteq \arg\max_{y \in Y} \sum_i w_i(y) = \arg\max_{y \in Y} v(y)$. By assumption, there thus exists $a^* \in A^0$ such that $c(a^*) = 0$ and $\operatorname{supp} F(a^*) \subseteq \arg\max_{y \in Y} v(y)$. 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 1. So a^* gives each agent his highest feasible payoff under w' as it maximizes $v(y)$ at zero cost. Hence, a^* is an equilibrium and $S(w') \geq \mathbb{E}_{F(a^*)}[v(y)] = \max v(Y) \geq S(w)$. □

Theorem 1 follows from Lemmas 2–4 by noting that each of Lemmas 2 and 3 implies that the profit guarantee satisfies $V(w) \leq V(w, A^n) \leq \mathbb{E}_{F^n}[v(y)] \to 0$, and the surplus guarantee satisfies $S(w) \leq S(w, A^n) \leq \mathbb{E}_{F^n}[v(y)] - \sum \min c_i(A_i^n) \to -\sum \underline{c}_i^0$.

To prove the stronger results in Corollaries 1–3, it suffices to show that the additional assumptions allow us to strengthen Lemma 4. Heuristically, they ensure that there is no known zero-cost action profile $a \in A^0$ such that $\operatorname{supp} F(a) \subseteq Y^*$, or, if one exists, that the outcomes in Y^* have value zero. See the Appendix for details.

4 Team-Optimal Contracts

We now consider team-optimal contracts, i.e., contracts that give the best possible expected surplus guarantee $S(w)$ subject to budget balance. By the preceding analysis, it is without loss of generality to restrict attention to contracts that align the agents' interests. We show here that an optimal contract of this form exists, and give a necessary and sufficient condition for it to yield a non-trivial surplus guarantee. We also derive a necessary and sufficient condition for the guarantee to be positive so that the team is worth forming, and discuss how to find the optimal shares.

The following result collects our main findings on team-optimal contracts. For the statement, say that a contract is *linear* if each agent is paid some fixed share $\alpha_i \in [0,1]$ of the value $v(y)$ for every $y \in Y$. Under budget balance this is equivalent to the contract aligning the agents' interests by Lemma 1.

Theorem 2. *(i) There exists a linear team-optimal contract.*

(ii) A team-optimal contract w guarantees non-trivial expected surplus (i.e., $S(w) > -\sum_i \underline{c}_i^0$) if and only if the known technology A^0 satisfies

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

(iii) If no known action profile $a \in A^0$ satisfies both $c(a) = 0$ and $\mathbb{E}_{F(a)}[v(y)] = \max v(Y)$, then every team-optimal contract that guarantees non-trivial expected surplus is linear.

Part (i) of Theorem 2 implies that profit-sharing is a robustly optimal arrangement for a partnership absent a principal who could act as a sink or a source of funds. Part (ii) gives a necessary and sufficient condition on the known technology A^0 for the optimal surplus guarantee to be non-trivial. The first two terms on the left-hand side of (4.1) correspond to the expected surplus. The presence of the third term means that a non-trivial surplus guarantee requires the known technology to be sufficiently productive; simply being able to generate surplus in excess of the agents' minimum costs is not enough. As we explain below, this is a manifestation of the familiar free-rider problem. Finally, part (iii) gives a sufficient condition for all team-optimal contracts to be linear when (4.1) holds. The condition is weak enough that we expect it to hold in most cases of interest.

Part (iii) of Theorem 2 follows immediately from Corollary 3. As for part (i), since budget-balanced linear contracts outperform other budget-balanced contracts by Theorem 1, it suffices to show the existence of an optimal contract within the class of linear contracts. For this we identify the space of such contracts with the compact set $\{\alpha \in [0,1]^I : \sum_i \alpha_i = 1\}$ by Lemma 1. We then use the upper hemi-continuity of the Nash equilibrium correspondence and our equilibrium selection to show that the surplus guarantee $S(w)$ is an upper semi-continuous function of α. As such it achieves a maximum at some α^*; this α^* is a team-optimal contract. See Lemma A.4 for details.

To prove part (ii) of Theorem 2, and to describe the properties of linear team-optimal contracts further, we characterize the guaranteed expected surplus $S(w)$ for any budget-balanced linear contract. Fix any such contract w. Define $U^0(w) \in \mathbb{R} \cup \{-\infty\}$ by setting

$$U^0(w) := \sup_{a \in A^0} \left(\mathbb{E}_{F(a)}[v(y)] - \sum_i \frac{c_i(a_i)}{\alpha_i} \right), \tag{4.2}$$

where $0/0 = 0$ and $x/0 = \infty$ for $x > 0$ by convention. We shall see below that $U^0(w)$ can be interpreted as a type of virtual surplus. It turns out to be the key object in the determination of the guarantee $S(w)$.

We first deal quickly with the possibility that $U^0(w) = -\infty$. This is the case if and only if $\alpha_i = 0$ for some agent i with positive lowest known cost $\underline{c}_i^0 > 0$. Any such contract w only gives the trivial guarantee $S(w) = -\sum_i \underline{c}_i^0$. To see this, let i be an agent with $\underline{c}_i^0 > 0$ for whom $\alpha_i = 0$. Consider a technology where this agent i has a new action that costs slightly less than $\underline{c}_i^0$, but leads to the outcome y_0 with certainty. Agent i must play this new action in all equilibria as he has no stake in the outcome, and the other agents must play their least-cost actions as they cannot affect the outcome. Thus all equilibria yield the lowest feasible surplus with certainty.[4] This implies that we can at least weakly improve upon any contract w such that $U^0(w) = -\infty$ with a contract w' such that $U^0(w') > -\infty$: any contract w' where each agent gets a strictly positive share will do. We can thus focus on the case $U^0(w) > -\infty$ in what follows.

For this case, we have the following characterization.

Lemma 5. *Let w be a budget-balanced contract that aligns the agents' interests. Suppose $U^0(w) > -\infty$. Then*

$$S(w) = \min_{E \in [0, \max v(Y)],\, b \in \mathbb{R}_+^I} E - \sum_i b_i \quad \textit{subject to}$$

$$E - \sum_i \frac{b_i}{\alpha_i} \geq U^0(w), \tag{4.3}$$

$$\alpha_i E - b_i \geq -\underline{c}_i^0 \quad \forall i = 1, \ldots, I. \tag{4.4}$$

If (E, b) achieves the minimum, then it satisfies (4.3) with equality. Furthermore, there exists a minimizer (E, b) with $E = \max\{U^0(w), 0\}$, $Eb_i = 0$, and $b_i \leq \underline{c}_i^0$ for all i.

The interpretation of the above minimization problem is that we are trying to construct a worst-case technology $A \supseteq A^0$ where the best equilibrium $a \in A$ results in the expected value of output $E = \mathbb{E}_{F(a)}[v(y)]$ and cost $b_i = c_i(a_i)$ for each agent i.

[4] More formally, consider $A \supseteq A^0$ where $A_i = A_i^0 \cup \{a_i'\}$ for some agent i with $\alpha_i = 0$ and $\underline{c}_i^0 > 0$, and $A_j = A_j^0$ for $j \neq i$. Let $c_i(a_i') = \eta \underline{c}_i^0$ for some $\eta \in (0,1)$, and let $F(a) = \delta_{y_0}$ whenever $a_i = a_i'$. Because $\alpha_i = 0$, agent i plays a_i' in all equilibria of $\Gamma(w, A)$. Since the other agents cannot affect the outcome, they must play their least-cost actions in A^0. Thus, $S(w) \leq S(w, A) = -\eta \underline{c}_i^0 - \sum_{j \neq i} \underline{c}_j^0 \to -\sum_j \underline{c}_j^0$ as $\eta \to 1$.

The proof of Lemma 5 requires some work, the details of which we relegate to the Appendix. We outline here the main ideas. The key observation is that any budget-balanced linear contract w induces a potential game between the agents.[5] To see this, fix any technology $A \supseteq A^0$. Suppose for simplicity that each agent's share α_i is positive under w. Define the function $P : A \to \mathbb{R}$ by

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

Denote agent i's payoff by $u_i(a) := \mathbb{E}_{F(a)}[\alpha_i v(y)] - c(a_i)$. Then, for every a_i, a_i', and a_{-i},

$$u_i(a_i, a_{-i}) - u_i(a_i', a_{-i}) = \alpha_i(P(a_i, a_{-i}) - P(a_i', a_{-i})).$$

That is, the function P is a weighted potential for the game $\Gamma(w, A)$. This implies that any $a^* \in \arg\max_{a \in A} P(a)$ is a pure strategy Nash equilibrium.

That $S(w)$ is not less than the minimum in Lemma 5 is now easy to establish. Any equilibrium a^* of $\Gamma(w, A)$ that maximizes the potential P satisfies

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

where the inequality follows because $A \subseteq A^0$, and the last equality is by definition of $U^0(w)$. Thus, there exists an equilibrium where the expected value of output and costs satisfy (4.3). Moreover, we have $\alpha_i \mathbb{E}_{F(a^*)}[v(y)] - c_i(a_i^*) \geq -\underline{c}_i^0$, as otherwise agent i would deviate to the least-cost action in A_i^0, and thus this equilibrium also satisfies (4.4). So the highest equilibrium surplus is no less than the minimum surplus subject to (4.3) and (4.4).

We generalize the above argument in the Appendix to the case where $\alpha_i = 0$ for some agent(s). Then P is no longer a potential, but the game can still be shown to be a generalized ordinal potential game. We state the general version of the result here for future reference.

Lemma 6. *Let w be a budget-balanced contract that aligns the agents' interests. For every technology $A \supseteq A^0$, there exists a pure-strategy Nash equilibrium $a^* \in \mathcal{E}(w, A)$ such that*

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

and $\alpha_i \mathbb{E}_{F(a^*)}[v(y)] - c_i(a_i^*) \geq -\underline{c}_i^0$ for all i. ($-\infty \geq -\infty$ is allowed in the first inequality.)

[5] All concepts and results related to potential games used in the analysis can be found in Monderer and Shapley (1996). It can be shown that any contract that aligns the agents' interests induces a potential game between the agents, but we do not need the general form of this result.

To prove the other direction of Lemma 5—that $S(w)$ is not greater than the minimum—we construct a sequence of technologies $A^n \supseteq A^0$ such that $S(w, A^n)$ converges to the minimum. The idea is again the easiest to illustrate when every agent's share is positive, so suppose this is the case. Let (E, b) achieve the minimum in Lemma 5. Consider a technology A that assigns one new action a_i' to each agent. Let $\mathbb{E}_{F(a')}[v(y)] = E$ and $c_i(a_i') = b_i$.

We note first that it is easy to complete the description of A such that the profile of new actions a' is an equilibrium by considering the potential P. To see this, note that (4.3) says that $P(a') \geq U^0(w) = \max_{a \in A^0} P(a)$. Thus, if we set $F(a) = \delta_{y_0}$ for every $a \notin A^0 \cup \{a'\}$, then $P(a) = -\sum c_i(a_i)/\alpha_i \leq -\sum b_i/\alpha_i \leq P(a')$ for all such a, since we can assume that the minimizer satisfies $b_i \leq \underline{c}_i^0$ by the last part of Lemma 5. Then a' maximizes the potential on A and hence it is an equilibrium with the desired surplus $E - \sum b_i$.

However, in general the equilibrium that maximizes the potential need not be the one with the highest surplus. To rule out equilibria with a higher surplus, we modify the construction to ensure that a' is the unique equilibrium. Roughly, we slightly increase the payoff to the profile a' and then use the gap between $P(a')$ and $U^0(w)$ so created to carefully construct the distributions $F(a)$ for actions $a \notin A^0 \cup \{a'\}$ to eliminate all other equilibria. The details can be found in the Appendix. Our discussion of the free-rider problem below contains a heuristic derivation of the worst-case surplus under particular assumptions, which essentially provides a non-technical outline of the proof for that case.

We are now in a position to complete the proof of Theorem 2.

Proof of Theorem 2.(ii). Let w be a linear team-optimal contract. Without loss of generality, we may assume $U^0(w) > -\infty$ so that Lemma 5 applies. Let (E, b) be a minimizer. Suppose that w only has the trivial guarantee $S(w) = -\sum_i \underline{c}_i^0 \leq 0$. Then (4.3) implies

$$0 \geq S(w) = E - \sum_i b_i \geq E - \sum_i \frac{b_i}{\alpha_i} \geq U^0(w).$$

By Lemma 5 we can then take $E = 0$ and $b_i \leq \underline{c}_i^0$, and hence we must have $b_i = \underline{c}_i^0$ for all i. Since (4.3) binds, this in turn implies that $-\sum_i \underline{c}_i^0/\alpha_i = U^0(w)$, or equivalently, that

$$\max_{a \in A^0} \left(\mathbb{E}_{F(a)}[v(y)] - \sum_i \frac{c_i(a_i) - \underline{c}_i^0}{\alpha_i} \right) = 0. \tag{4.5}$$

Conversely, if (4.5) holds, then $(E, b) = (0, (\underline{c}_1^0, \ldots, \underline{c}_I^0))$ is feasible in the minimization problem, and thus $S(w) \leq -\sum_i \underline{c}_i^0$. We conclude that (4.5) is a necessary and sufficient condition for w to yield the trivial guarantee $S(w) = -\sum_i \underline{c}_i^0$.

The left-hand side of (4.5) is non-negative, since the sum can be made equal to zero by choosing a profile of least-cost actions. Hence, w has a non-trivial surplus guarantee if and only if the maximum is positive. Maximizing with respect to α we obtain a necessary and sufficient condition for this to be true for some (and hence also for the optimal) α:

$$\max_{a\in A^0} \max_{\alpha\in[0,1]^I:\sum_i \alpha_i=1} \left(\mathbb{E}_{F(a)}[v(y)] - \sum_i \frac{c_i(a_i) - \underline{c}_i^0}{\alpha_i}\right) > 0. \tag{4.6}$$

For any $a \in A^0$, the inner maximum in (4.6) is achieved by setting

$$\alpha_i(a) = \frac{\sqrt{c_i(a_i) - \underline{c}_i^0}}{\sum_j \sqrt{c_j(a_j) - \underline{c}_j^0}},$$

with $0/0 = 1/I$ by convention. Substituting these shares back into (4.6) and rearranging then yields the expression in (4.1) and establishes part (ii) of Theorem 2. □

4.1 Positive guarantee and optimal shares

Part (ii) of Theorem 2 characterizes all known technologies for which a team-optimal contract gives a non-trivial surplus guarantee. However, if the lowest known cost $\underline{c}_i^0$ is positive for some of the agents, this guarantee may still be negative. One possible interpretation of such a positive lowest cost is that the agent has to incur a fixed cost. With this in mind, it is natural to ask when does a team-optimal contract achieve a positive surplus guarantee $S(w) > 0$ so that the fixed cost is worth incurring and the team is worth forming.

The properties of the minimizers in Lemma 5 immediately imply the following result.

Lemma 7. *Let w be a budget-balanced contract that aligns the agents' interests. We have $S(w) > 0$ if and only if $U^0(w) > 0$. Furthermore, if $U^0(w) \geq 0$, then $S(w) = U^0(w)$.*

Therefore, $S(w) > 0$ for some contract w (and thus also for the optimal contract) if and only if $U^0(w)$ can be made strictly positive by maximizing it with respect to α, or

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

Analogously to (4.6), the inner maximum is achieved for any $a \in A^0$ by setting

$$\alpha_i(a) = \frac{\sqrt{c_i(a_i)}}{\sum_j \sqrt{c_j(a_j)}}, \tag{4.7}$$

with $0/0 = 1/I$ by convention. Substituting these shares back into the objective establishes the following result.

Theorem 3. *A team-optimal contract guarantees positive expected surplus if and only if*

$$\max_{a\in A^0}\Big(\mathbb{E}_{F(a)}[v(y)]-\sum_i c_i(a_i)-2\sum_{i,j:i\neq j}\sqrt{c_i(a_i)c_j(a_j)}\Big)>0. \tag{4.8}$$

If (4.8) *holds and* w *is a team-optimal contract, then* $S(w)$ *equals the left-hand side of* (4.8)*. Moreover, a linear team-optimal contract can then be found by substituting any maximizer from* (4.8) *into* (4.7)*.*

Analogously to (4.1), the left-hand side of condition (4.8) includes an extra term in addition to the expected surplus, implying that the maximized expected surplus from the known technology has to be sufficiently positive for the guarantee to be positive. Of course, (4.1) and (4.8) coincide if the lowest known cost is zero for each agent.

Theorem 3 shows that when the optimal surplus guarantee is positive, a linear team-optimal contract can be found by first solving the maximization problem in (4.8) to obtain a maximizer a^*, and then backing out the shares $\alpha_i^*=\alpha_i(a^*)$ using (4.7). This provides a way to solve for an optimal contract in parametric examples. The simplest such example is found by taking the known technology to be a singleton, in which case the optimal shares are just (4.7) evaluated at the only known action profile.

Even without any additional structure, it is immediate from (4.7) that an agent's share should be larger, the costlier is his action in the profile a^*. However, in general the optimal shares reflect more than just the agents' costs, because a^* depends also on how the actions affect the distribution of outcomes.

For completeness, we note that if (4.1) holds, but (4.8) is not satisfied, then a linear team-optimal contract can still be found by maximizing $S(w)$ with respect to α using the characterization in Lemma 5. But this is a tedious exercise at best. Finally, if (4.1) is not satisfied, any contract just gives the trivial guarantee $-\sum \underline{c}_i^0$, so any shares are optimal.

4.2 On the free-rider problem

In order to see the economic intuition behind the surplus guarantee, it is useful to consider the case where the lowest known cost is zero for each agent (i.e., $\underline{c}_i^0=0$, $i=1,\dots,I$) so that $U^0(w)$ is nonnegative. The surplus guarantee from any linear budget-balanced contract w is then simply $S(w)=U^0(w)$ by Lemma 7.

The surplus guarantee $U^0(w)$ can be interpreted as a type of *virtual surplus*: it accounts for the incentive and robustness concerns by inflating the agents' costs with their shares. Its roots are in the free-rider problem. Namely, if there is only one agent, then the virtual surplus equals the true surplus as the agent receives the full value of output. But with two or more agents, it is impossible to promise the full value to every agent—captured by some of the shares α_i being less than one—resulting in the virtual surplus being lower than the true surplus. As is evident from the derivation of Theorem 3 above, the third term in (4.8) captures what is left of this effect after the shares have been optimized. It is also responsible for the extra term in (4.1).

The following thought experiment explains the functional form of $U^0(w)$. Suppose that the known technology consists of just one action profile: $A^0 = \{a^0\}$. Consider trying to lower the expected surplus relative to a^0 by giving agent 1 a new lower-cost action a_1'. To keep agent 1 indifferent and thus willing to play the new action, we can lower the expected value of output by $(c_1(a_1^0) - c_1(a_1'))/\alpha_1$. If $\alpha_1 < 1$, this reduces the total surplus as the reduction in the expected value of output is larger than agent 1's cost saving. That is, the deviation by agent 1 imposes a negative externality on the other agents; this is precisely the free-rider problem. Letting $c_1(a_1') = 0$ we obtain the maximal reduction of $c_1(a_1^0)/\alpha_1$ in the expected value of output. We can then give agent 2 a zero-cost action a_2' to obtain a further reduction of $c_2(a_2^0)/\alpha_2$, and so on. Continuing this process, we obtain a zero-cost action profile a' whose expected value of output is given by $\mathbb{E}_{F(a')}[v(y)] = \mathbb{E}_{F(a^0)}[v(y)] - \sum c_i(a_i^0)/\alpha_i$. As the costs are zero, this is also the expected surplus from the profile a'.[6] By inspection of (4.2), we have thus arrived at the formula for $U^0(w)$ for the case of one known action profile.

Heuristically, when there are multiple known action profiles, the maximum in (4.2) identifies the one for which the above process yields the highest remaining surplus.

Two further remarks are in order: First, if some agents have fixed costs so that $\underline{c}_i^0 > 0$, the above logic still applies. Indeed, as long as $U^0(w) \geq 0$, the worst-case can still be obtained with zero-cost actions—that is, Lemma 5 has a minimizer with $b = 0$—and hence we have $S(w) = U^0(w)$ exactly as above. But because of the fixed costs, we may now have $U^0(w) < 0$. In this case, we can reduce the expected value of output all the way to zero in the above process without having to reduce all agents' costs to zero. The worst case will then involve some agents having actions with positive costs (corresponding to having a minimizer with $b_i > 0$ for some i in Lemma 5). The determination of these costs is why the solution to the minimization problem in Lemma 5 is more complicated when $U^0(w) < 0$. While there does not appear to be a simple closed-form expression for $S(w)$ in this case, the solution has the following regularity: The cost of the new action is set to zero for the agents with the smallest shares as this provides the largest reduction in the expected value of output for a given reduction in costs; the agents with the largest shares will have their costs reduced only

[6] This is essentially what the worst-case technology used in the proof of Lemma 5 reduces to in this case, modulo the fact that there we require incentives to be strict to ensure uniqueness of the equilibrium outcome.

to the minimum cost $\underline{c}_i^0$, which may be positive. We omit the details.

Second, it is worth emphasizing that even when $U^0(w) \geq 0$, the fact that the profit guarantee from a linear contract is given by the virtual surplus $U^0(w)$ depends crucially on the true technology being unknown. Namely, suppose instead that the principal believes the true technology to be equal to the known technology A^0. Any maximizer $a^* \in A^0$ in (4.2) is an equilibrium of $\Gamma(w, A^0)$ as it maximizes the potential P on A^0. (It can be verified that any such a^* is an equilibrium even if $\alpha_i = 0$ for some i even though P is not a potential in this case.) So the principal can expect at least the surplus $\mathbb{E}_{F(a^*)}[v(y)] - \sum_i c_i(a_i^*)$. Consequently, a linear team-optimal contract is in general not an optimal linear contract for a model where the principal knows the technology.

5 Principal-Optimal Contracts

We then turn to principal-optimal contracts that maximize the principal's guaranteed expected profit $V(w)$. By Theorem 1, we can restrict attention to contracts that align the agents' interests, because any contract that fails to do so is dominated by 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 particular relationship between the value of the outcome, $v(y)$, and the agents' compensation. Thus, even though a linear contract still turns out to be optimal, showing this requires more work than in the case of team-optimal contracts, where optimality of linear contracts was implied by Theorem 1.

As a first step, we derive a convenient representation of the candidate optimal contracts. Note that if w is a principal-optimal contract, then the lowest payment to each agent must be zero. Otherwise we could strictly increase the principal's profit with the contract w' defined by $w'(y) := w(y) - (\min w_1(Y), \ldots, \min w_I(Y))$ for all $y \in Y$, because subtracting the constants does not affect the agents' incentives (i.e., $\mathcal{E}(w, A) = \mathcal{E}(w', A)$ for all A), but it reduces the principal's wage bill. Moreover, because w aligns the agents' interests, there exists an outcome $\underline{y} \in Y$ that yields the zero payment simultaneously to all agents so that $w(\underline{y}) = 0$. 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 under w given outcome y. Then there exist shares $\alpha = (\alpha_1, \ldots, \alpha_I) \in [0,1]^I$, with $\sum_i \alpha_i = 1$, such that for every agent i,

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

That is, each agent is paid some fixed share of the total compensation for any outcome.

Conversely, any contract that can be written in this form and where $\bar{w}(y) = 0$ for some y aligns the agents' interests and is anchored at the origin. This result is almost immediate from the definitions; it can be proven the same way as Lemma 1.[7]

With the above representation in hand, we can think of the problem of finding a principal-optimal contract in two stages. First, given the agents' shares α, determine how the total compensation $\bar{w}(y)$ should depend on the outcome y. Second, optimize over the agents' shares. As the total compensation is one dimensional, the first stage has a connection to the single-agent case, which we exploit to show that, here too, it is optimal to tie total compensation linearly to the value of the outcome, i.e., to set $\bar{w}(y) = \beta v(y)$ for some $\beta \in [0, 1]$. The payment to each agent is then $w_i(y) = \alpha_i \beta v(y)$, implying that a linear contract is optimal. (Recall that a contract is said to be linear if each agent is paid a fixed share of the value $v(y)$ for each y; the shares may or may not sum to one across agents.) The second stage then consists of finding the optimal linear contract by optimizing jointly with respect to β and α.

More precisely, we have the following result:

Theorem 4. *(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 (4.8).

(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 4 parallels the findings for team-optimal contracts in Theorem 2. Part (i) shows the optimality of linear contracts, and part (iii) provides a sufficient condition for all principal-optimal contracts to be linear. Heuristically, a linear contract aligns interests not just among the agents, but among all the parties. Thus it prevents the agents from seeking personal gain at the other agents' or the principal's expense. In the special case of one agent, Theorem 4 strengthens Carroll's (2015) result about the optimality of a linear contract by showing that a principal-optimal contract should tie the agent's compensation linearly to the value of output $v(y)$ to the principal, even though the signal y may be richer.

Part (ii) of Theorem 4 gives a necessary and sufficient condition for the optimal profit guarantee to be positive. This condition is the same one that characterizes whether a team-optimal contract can guarantee positive expected surplus. To see why this is the case, note that the principal can ensure a positive profit by choosing β close enough to 1 and taking

[7] Replace v with $\bar{w}$ everywhere in the proof of Lemma 1, and observe that $\min \bar{w}(Y) = 0$ because w is anchored at the origin.

α to coincide with the shares in the team-optimal contract whenever the latter generates a positive expected surplus. In contrast, if the condition fails, then the free-rider problem makes the surplus guarantee negative for every linear contract with $\beta = 1$; decreasing the agents' shares by lowering β can only make matters worse. Note that with one agent, (4.8) reduces to the requirement that some known action generates a positive expected surplus, which is precisely the non-triviality assumption needed in that case.

The rest of this section is devoted to the proof of Theorem 4. Along the way, we obtain a formula for the profit guarantee for a linear contract, which can be used to find the agents' shares in a linear principal-optimal contract. We remark on this after the proof.

We will make use of the fact that—as far as the agents are concerned—any contract w that aligns the agents' interests and is anchored at the origin can be interpreted as a budget-balanced contract in an auxiliary model where the value of each outcome y is given by $\bar{w}(y)$. We can thus recycle results, most notably the characterization behind Lemma 5, from the analysis of team-optimal contracts. (In case of the zero contract $w \equiv 0$, the auxiliary model does not satisfy the non-triviality assumption $\max v(Y) = \max \bar{w}(Y) > 0$, so we treat it separately.) This and the above mentioned connection to the single-agent case allow us to proceed relative fast here.

The virtual surplus $U^0(w)$ continues to play a key role in the analysis. To avoid confusion, we write

$$\bar{U}^0(w) := \sup_{a \in A^0} \left(\mathbb{E}_{F(a)}[\bar{w}(y)] - \sum_i \frac{c_i(a_i)}{\alpha_i} \right) \in \mathbb{R} \cup \{-\infty\},$$

where we have simply replaced $v(y)$ with the total compensation $\bar{w}(y)$ in the definition of $U^0(w)$ in (4.2), and where $0/0 = 0$ and $x/0 = \infty$ for $x > 0$ by convention.

To shorten the statements of some of the lemmas that follow, 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 definition adapts Carroll's (2015) notion of an eligible contract to the multi-agent setting, parts (i) and (ii) being the novel requirements. For example, any linear contract satisfies (i) and (ii). But the set of eligible contracts may nevertheless be empty, since the best profit guarantee may be zero in violation of (iii). However, if $V(w) > 0$ for some 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 analogous to the single-agent case.

Lemma 8. *Let w be an eligible contract, different from the zero contract. Then*

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

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

The proof of Lemma 8 relies on the characterization behind Lemma 5. To see that $V(w)$ is not less than the minimum, interpret w as a budget-balanced contract in a model where $v(y) = \bar{w}(y)$. Then Lemma 6 implies that every technology $A \supseteq A^0$ has an equilibrium a^* such that $\mathbb{E}_{F(a^*)}[\bar{w}(y)] \geq \mathbb{E}_{F(a^*)}[\bar{w}(y)] - \sum c_i(a_i^*)/\alpha_i \geq \bar{U}^0(w)$. Thus the principal's profit is at least the minimum profit under distributions satisfying the constraint in (5.2).

We prove the other direction in the Appendix. The key observation is that $(E, b) \in \mathbb{R}_+^{I+1}$ defined by $E = \mathbb{E}_G[\bar{w}(y)]$ and $b = 0$ is a feasible point in the minimization problem in Lemma 5. Thus, the constructive direction in the proof of Lemma 5 gives us a technology $A \supseteq A^0$ where the unique equilibrium distribution of outcomes is approximately G.

An application of Lemma 8 yields a formula for the profit guarantee for any eligible linear contract w, where $\bar{w}(y) = \beta v(y)$ for some $\beta \in (0, 1]$ and $w_i(y) = \beta\alpha_i v(y)$. It turns out that the formula is also valid for the zero contract whenever it is eligible.[8]

Lemma 9. *Let w be an eligible linear contract with $\beta \in [0, 1]$ and $\alpha \in [0, 1]^I$, $\sum \alpha_i = 1$. 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 Lemma 9 applies also when $I = 1$, in which case $\alpha_1 = 1$ and (5.3) reduces to the formula derived for the single-agent case by Chassang (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 claim now follows by writing out $\bar{U}^0(w)$ and noting that the supremum in $\bar{U}^0(w)$ is achieved, since $\bar{U}^0(w) = \mathbb{E}_G[\bar{w}(y)] \geq 0 > -\infty$.

If the zero contract is eligible, then there exists a profile $a \in A^0$ such that $c(a) = 0$ and $\mathbb{E}_{F(a)}[v(y)] > 0$. (We can't have $\underline{c}_i^0 > 0$ for any i, because $V(0)$ could then be driven to zero by giving agent i a cheaper action—cf. footnote 4.) 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. □

[8] The zero contract has a continuum of parameterizations by α and β, since any α_i with $\sum_i \alpha_i = 1$ will do if $\beta = 0$. This multiplicity creates no problem in the analysis, but of course it could be avoided by denoting the agents' shares by $\gamma_i := \beta\alpha_i$, in which case $\beta = \sum \gamma_i \leq 1$ and the zero contract corresponds to $\gamma = 0$.

From (5.3) we can deduce the existence of a best linear contract and see that (4.8) is necessary and sufficient for the profit that it guarantees to be positive.

Lemma 10. *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 (4.8).*

Proof. If no linear contract is eligible, then the zero contract has $V(0) = 0$ and thus it 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 β and α_i.

The derivation leading to Theorem 3 in Section 4 shows that if (4.8) holds, then we have $U^0(w) = \max_{a \in A^0}(\mathbb{E}_{F(a)}[v(y)] - \sum c_i(a_i)/\alpha_i) > 0$ for some $\alpha_i \in [0,1]^I$ with $\sum \alpha_i = 1$. Thus, (5.3) is positive for β close enough to 1. Conversely, if (4.8) does not hold, then we have $(1-\beta)\max_{a \in A^0}(\mathbb{E}_{F(a)}[v(y)] - \sum c_i(a_i)/\beta\alpha_i) \leq (1-\beta)U^0(w) \leq 0$ for every β and α, showing that (5.3) is nonpositive. Hence, no linear contract is eligible. □

With these facts regarding linear contracts established, to prove Theorem 4 it suffices to show that any eligible contract can be improved upon by a linear contract, strictly so if every known action profile satisfies full support. To this end, fix an eligible contract and consider the representation (5.1). We will show that the contract can be (weakly) improved upon by making the total compensation $\bar{w}$ a linear function of the output value $v(y)$, while keeping each agent's share α_i of the total compensation fixed. As the total compensation is one-dimensional, this allows us to draw on the proof of the single-agent case.

We need the following key lemma from the single-agent case. It identifies a particular supporting hyperplane to the set of pairs $(\bar{w}(y), v(y) - \bar{w}(y))$ under contract w that will be used to define the improvement contract. The proof in the appendix adapts the argument from the single-agent case to the present setting.

Lemma 11. *Let w be an eligible contract, different from the zero contract. Then there exist numbers κ and λ, with $\lambda > 0$, such that*

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

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

Given an eligible contract w and numbers κ, λ satisfying (5.4) and (5.5), define the 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 α_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' so defined will be shown to improve on the contract w. Moreover, w' can be further improved upon simply by removing the constant payment, which does not affect the agents' incentives, and which is nonnegative by limited liability. Because w' is affine, it is not anchored at the origin and hence it neither has the representation (5.1), nor is it eligible. So for technical reasons, it is convenient to show both improvements at once.

To this end, define the linear contract w'' by setting

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

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

Lemma 12. *Let w be an eligible contract, different from the zero contract, that satisfies (5.4) and (5.5), and let w'' be the linear contract defined by (5.7). Then $V(w'') \geq V(w)$. Moreover, if every known action profile satisfies full support and w is not linear, then $V(w'') > V(w)$.*

Proof. We 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 6 (with the substitutions $v(y) = \bar{w}''(y)$ and $U^0(w'') = \bar{U}^0(w'')$) to find for any $A \supseteq A^0$ a pure-strategy equilibrium $a^* \in \mathcal{E}(w'', A)$ with $\mathbb{E}_{F(a^*)}[\bar{w}''(y)] \geq \bar{U}^0(w'')$. 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

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

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 the strict inequality for non-linear contracts under the full support assumption. Observe first that if w' is not linear (i.e., if $\kappa < 0$), then $V(w'') > V(w') \geq V(w)$. So suppose w' is linear. Let every action profile in A^0 satisfy full support, i.e., $F(a) \neq \delta_{y_0}$ implies $\operatorname{supp} F(a) = Y$ 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 8, 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, $V(w'') \geq V(w') > V(w)$. □

We can now summarize how the claims in Theorem 4 follow from the previous lemmas.

Proof of Theorem 4. For part (i), we may restrict attention to contracts that align the agents' interests and are anchored at the origin. Let w be any such contract. If w is not eligible, then it is dominated by the zero contract, which is linear. If w is eligible, then Lemmas 11 and 12 imply that there exists a linear contract that does at least as well as w. Thus, either way, w is weakly dominated by a linear contract, and so the existence of a linear principal-optimal contract follows from Lemma 10, which shows the existence of a best linear contract.

Part (ii) follows from part (i) and Lemma 10.

For part (iii), suppose that every action profile in A^0 satisfies full support. Let w be a nonlinear principal-optimal contract with $V(w) > 0$. Then w is eligible and it can be strictly improved upon with a linear contract by Lemmas 11 and 12, a contradiction. □

A linear principal-optimal contract can be found as follows. If (4.8) is not satisfied, then the zero contract is optimal. Otherwise, maximize (5.3) with respect to β and α.

In the latter case, similarly to team-optimal contracts, we may first find the optimal shares for each $a \in A^0$, and then maximize with respect to a. So fix $a \in A^0$. Note that given any β, maximizing (5.3) with respect to α gives us the team-optimal shares in an auxiliary model where the value of each outcome is $\beta v(y)$. These are given by $\alpha(a)$ defined in (4.7). Similarly, maximizing (5.3) with respect to β given any α gives us the principal-optimal share in an auxiliary single-agent model where each action $a \in A^0$ costs the agent $C(a, \alpha) := \sum_i c_i(a_i)/\alpha_i$. 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 the objective gives

$$\left(\sqrt{E_{F(a)}[v(y)]} - \sqrt{C(a, \alpha(a))}\right)^2 = \left(\sqrt{E_{F(a)}[v(y)]} - \sqrt{\sum_{i,j} \sqrt{c_i(a_i)c_j(a_j)}}\right)^2.$$

Maximizing the above expression (or the square root thereof to eliminate the square) yields a maximizer a^*, which can be substituted back into the formulas for the shares to obtain the optimal contract $\alpha^* = \alpha(a^*)$ and $\beta^* = \beta(a^*, \alpha^*)$.

The simplest non-trivial examples can be created by assuming that there is only one known action profile so that $A^0 = \{a^0\}$. In that case there is no maximization over actions, so the principal-optimal linear contract is given by $\alpha^* = \alpha(a^0)$ and $\beta^* = \beta(a^0, \alpha^*)$.

6 Risk-Averse Agents

The above analysis assumes the agents to be risk-neutral. We explain here how the logic of interest alignment extends to risk-averse agents.

Each agent's payoff under a contract $w : Y \to \mathbb{R}_+^I$ is now of the form $u_i(w_i(y)) - c_i(a_i)$ for some increasing, concave function $u_i : \mathbb{R} \to \mathbb{R}$. For simplicity, we assume $u_i(0) = 0$. The model is otherwise exactly as specified in Section 2; in particular, the principal is risk-neutral.

We can apply our previous arguments in the utility space to show that a specific form of interest alignment remains necessary for non-trivial performance guarantees.

Definition 2. A contract w *aligns the agents' interests in utilities* if all utility-payment profiles $u(w(y)) := (u_1(w_1(y)), \dots, u_I(w_I(y)))$, $y \in Y$, lie on the same ray in $\mathbb{R}_+^I$, i.e., if $u(w(Y)) \subset \{\underline{u} + dt : t \in \mathbb{R}_+\}$ for some $\underline{u}, d \in \mathbb{R}_+^I$.

Note that if a contract w aligns the agents' interests in utilities, then w prescribes team-based compensation in the sense that the 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'. So the agents' compensation covaries positively, but not necessarily linearly. Such team-based compensation is necessary:

Theorem 5. *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 $S(w) \leq S(w')$.*

This result parallels Theorem 1, except that now in the budget-balanced case, we are only guaranteed the existence of some budget-balanced contract that improves on w.

The proof of Theorem 5 is essentially the same as that of Theorem 1. Change variables by defining $\tilde{w}_i(y) := u_i(w_i(y))$ so that the principal's payoff becomes $v(y) - \sum_i u_i^{-1}(\tilde{w}_i(y))$. Then w aligns the agents' interests in utilities if and only if $\tilde{w}$ aligns the agents' interests in the sense of Definition 1. Lemmas 2 and 3 then apply verbatim as their proofs do not rely on the form of the principal's cost function. A minor adjustment is required to Lemma 4; it is handled in the Appendix.

Theorem 5 shows that even with risk-averse agents, the concern for robust performance gives rise to team-based compensation. But with interests aligned in the utility space, the exact form of the payments depends on the agents' utility functions. However, there is one case where interest alignment in utilities is equivalent to interest alignment in payments.

Lemma 13. *Suppose the agents' preferences over money are represented by symmetric power utility functions (i.e., $u_i(x) = x^\rho$ 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 1.*

Any budget-balanced contract is necessarily anchored at the origin, and so is any candidate for a principal-optimal contract. Thus, with symmetric power utility functions (a particular case of CRRA preferences), optimal contracts have the agents' compensation co-varying positively and linearly just like in the risk-neutral case. And if the contract is also budget balanced, it is simply a linear contract. Thus profit sharing is a team-optimal contract when agents have symmetric power utility functions.

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. Such contracts have a natural interpretation as being team-based. Under budget balance they reduce to linear contracts, showing that profit-sharing, or equity, is a team-optimal contract. And the contract with the best profit guarantee for the principal is similarly linear.

These incentive schemes have two additional robustness properties, which play no role in our analysis, but likely contribute to their popularity. First, interest alignment limits the scope for collusion among subsets of agents as all agents' payoffs are already similar to each other, save for the costs. Moreover, in case of a linear principal-optimal contract, the agents' compensation varies linearly with the value of the outcome, so any collusive scheme that increases the agents' total compensation also increases the principal's payoff.

Second, interest alignment not only limits a contract's downside, it also has the potential to increase the upside as it gives the agents an incentive to take advantage of unexpected opportunities to help each other and to allocate tasks efficiently. This upside potential is lost on our worst-case analysis, and it is unclear how to capture it short of moving to a fully Bayesian framework. (See Itoh (1991) and Garicano and Santos (2004) for the analysis of incentives to help and to allocate tasks in Bayesian models.) Note, however, that if the said 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-equipped agent are examples of the kind of negative actions that lead to the worst case.

The driving force behind our results is that only by completely eliminating conflict in the agents' preferences over outcomes can the contract guarantee good performance in all games. Taking the worst case over all games consistent with the known technology is arguably a strong assumptions. On one hand, it leads to a tractable analysis and yields sharp predictions about optimal contracts. On the other hand, the prediction that all teams be governed via linear schemes is obviously empirically false.

With this in mind, one may view the contribution of this paper as identifying robustness as a force pushing towards contracts that align the agents' interests. The analysis here shows that if this concern is strong enough, only linear schemes survive. Robustness is, however, only one of many considerations affecting contract design. Consequently, the contracts we observe in practice reflect it to varying degrees. A natural way to try to incorporate this into the worst-case analysis would be to restrict the set of games deemed possible, with smaller sets then resulting in less limitations on contract form. Identifying subsets of games for which the analysis remains tractable is a nontrivial problem which we leave for future work.

The Appendix

A.1 Proofs for Section 3

Proof of Lemma 2. We assume throughout the proof that the set Y^* is empty.

We use the following notation. Let $\bar{w}_i := \max_{y \in Y} w_i(y)$ and $Y_i^* := \arg\max_{y \in Y} w_i(y)$. The projection of W to the payments of agents i and j is denoted $W_{i,j} \subset \mathbb{R}_+^2$. The interior of $W_{i,j}$ relative to $\mathbb{R}^2$ is $\text{int}(W_{i,j})$. For any $x \in W$, we write $x_{i,j}$ for the image of x in $W_{i,j}$.

To simplify the exposition, we assume that w_i is not constant for any agent i. We comment at the end of the proof how the argument needs to be adjusted to accomodate such agents. (These agents can be essentially ignored when constructing the worst-case technology, so the issue is mostly notational.)

A Preliminary Technology. We construct a technology with a unique equilibrium expected payment profile in W. It forms the basis of all other technologies used in the proof.

Define A^1 by letting $A_i^1 = A_i^0 \cup \{a_i^1\}$ for all i. Let $c_i(a_i^1) = \underline{c}_i^0$ for all i so that the new action is a least-cost action for each agent.

For each i, fix $z^i \in W$ such that $z_i^i = \bar{w}_i$. Let $F^i \in \Delta(Y)$ be such that $\mathbb{E}_{F^i}[w(y)] = z^i$.

Given an action profile a in A^1 such that at least one agent plays the new action a_i^1, let $n = n(a) := |\{i : a_i = a_i^1\}|$. We then define the outcome distribution by setting

$$F(a) = \Big(1 - \frac{I - n + 1}{I}\xi\Big)\frac{1}{n}\sum_{i\,:\,a_i = a_i^1} F^i + \frac{I - n + 1}{I}\xi H,$$

where $\xi \in (0, 1)$, and H is the uniform distribution on Y. The corresponding profile of expected payments to the agents is

$$x(a) = \Big(1 - \frac{I-n+1}{I}\xi\Big)\frac{1}{n}\sum_{i\,:\,a_i=a_i^1} z^i + \frac{I-n+1}{I}\xi\mathbb{E}_H[w(y)]. \tag{A.1}$$

We record the following observations for future reference:

1. Since $w_i(y)$ is not constant in y, a_i^1 is the unique best-response to any profile a_{-i} where some agent $j \neq i$ plays a_j^1. This is because playing a_i^1 shifts the convex combination in (A.1) in the direction of z^i and reduces the weight on the full-support distribution H; the latter effect gives uniqueness even if we have $z_i^i = z_i^j$ for all j such that $a_j = a_j^1$.

2. The distribution $F(a)$ has full support on Y. Hence, $x_i(a) < \bar{w}_i$ for every agent for whom w_i is not constant. Moreover, $x_{i,j}(a)$ is in the interior of the projection $W_{i,j}$ (relative to $\mathbb{R}^2$) for all agents i and j for which the interior is nonempty.

Lemma A.1. *The profile a^1 is the unique equilibrium of $\Gamma(w, A^1)$ for all $\xi > 0$ small enough.*

Proof. Observation 1. implies that a^1 is an equilibrium, and that it is the only equilibrium where at least one agent i plays the new action a_i^1 with probability 1. Thus it only remains to show that this is the case in every equilibrium.

Let $E := \cap_i E_i$, where $E_i := \{a \in A^1 : a_i \in A_i^0\}$. Then E is the event that every agent plays some known action. Suppose towards contradiction that there exists an equilibrium σ such that $\sigma(E) = \prod_i \sigma_i(A_i^0) > 0$. Because Y^* is empty, there exists some agent j such that, conditional on E, $F(\sigma)$ assigns probability at most $(I-1)/I$ to $\arg\max_{y\in Y} w_j(y)$. Thus, agent j's payoff, given E, is at most $\frac{I-1}{I}\bar{w}_j + \frac{1}{I}\max\{w_j(y) : y \notin Y_j^*\}$. In particular, some $\hat{a}_j \in A_j^0$ with $\sigma_j(\hat{a}_j) > 0$ yields agent j a payoff no greater than this conditional on $\cap_{i\neq j} E_i$.

Note that agent j's payoff from a_j^1 is $(1-\xi)z_j^j + \xi\mathbb{E}_H[w_j(y)] = (1-\xi)\bar{w}_j + \xi\mathbb{E}_H[w_j(y)]$ conditional on $\cap_{i\neq j} E_i$. Moreover, a_j^1 gives a strictly higher payoff than $\hat{a}_j$ if some agent $i \neq j$ plays $a_i \notin A_i^0$ by observation 1. Thus, for $\xi > 0$ small enough, a_j^1 yields a strictly higher (unconditional) expected payoff than $\hat{a}_j$, contradicting $\sigma_j(\hat{a}_j) > 0$. The cutoff for ξ depends on j, but not on σ, so it can be chosen uniformly as there are finitely many agents. □

In what follows, we assume that ξ is small enough for the result to apply.

The Worst-case, Case 1. There are two cases to consider. We first deal with the easier case where $w_i(y_0) \geq x_i(a^1)$ for some agent i, where $x(a^1)$ is the expected payment profile defined by equation (A.1). For concreteness, suppose the inequality holds for agent 1. We will ensure strict incentives by using the following perturbation. Let $F_\varepsilon \in \Delta(Y)$ be such that $F_\varepsilon(y_0) > 1-\varepsilon$ and $\mathbb{E}_{F_\varepsilon}[w_1(y)] > x_1(a^1)$. Such a distribution can be found because $x_1(a^1) < \bar{w}_1$ by observation 2. F_ε will be our equilibrium distribution, so letting $\varepsilon \to 0$ will then give the desired sequence of technologies as $F_\varepsilon \to \delta_{y_0}$.

Define A by setting $A_1 = A_1^1 \cup \{a_1^2\}$, with $c_1(a_1^2) = \underline{c}_1^0$, and $A_i = A_i^1$ for $i > 1$, where A^1 is the technology constructed above. Let $F(a) = F_\varepsilon$ for all $a \in A$ such that $a_1 = a_1^2$.

We claim that F_ε is the unique equilibrium outcome distribution for the game $\Gamma(w, A)$. Indeed, the profile $(a_1^2, a_2^1, \ldots, a_I^1)$ is an equilibrium, since $\mathbb{E}_{F_\varepsilon}[w_1(y)] > x_1(a^1) > x_1(a_1, a_{-i}^1)$ for all $a_1 \in A_1^0$, and a_i^1 is optimal for agents $i > 1$ as their actions do not affect the outcome when $a_1 = a_1^2$. Moreover, agent 1 must play a_1^2 in any equilibrium where $\sigma_i(A_i^0) = 0$ for some agent i. To see this, suppose to the contrary that some such equilibrium has $\sigma_1(a_1^2) < 1$. Since $\sigma_i(A_i^0) = 0$ for some i, observation 1. then implies that a_j^1 strictly dominates all $a_j \in A_j^0$ for every agent j. Thus, $\sigma_j(A_j^0) = 0$ for all j, and hence $a_j = a_j^1$ for all $j > 1$. But then a_1^2 is agent 1's unique best-response, contradicting $\sigma_1(a_1^2) < 1$.

It remains to show that we have $\sigma_i(A_i^0) = 0$ for at least one agent i in every equilibrium of $\Gamma(w, A)$. This follows by the same argument as Lemma A.1. Define the events E_i analogously and suppose there exists an equilibrium σ with $\sigma(E) > 0$. The second paragraph in the proof of Lemma A.1 applies verbatim. The only difference is in the third paragraph. Now conditional on some agent $i \neq j$ playing $a_i \notin A_i^0$, a_j^1 may be only weakly better than $\hat{a}_j$: if $j \neq 1$ and $a_1 = a_1^2$, then agent j's action doesn't affect the outcome, but a_j^1 is still optimal as it is a least-cost action. This is enough to get the contradiction, because a_j^1 is strictly better than $\hat{a}_j$ conditional on $\cap_{i \neq j} E_i$. This completes the proof for the first case.

The Worst-case, Case 2. The more challenging case obtains if $w_i(y_0) < x_i(a^1)$ for all i. Then some projection $W_{i,j}$ of W has a nonempty interior relative to $\mathbb{R}^2$. To see this, note that if $\operatorname{int} W_{i,j}$ is empty for all pairs i, j, then each $W_{i,j}$ is a (possibly degenerate) line segment. But Y^* is empty, so some line segment $W_{i,j}$ must be strictly decreasing, implying that $w_k(y_0) \geq x_k(a^1)$ for $k = i$ or $k = j$. Relabeling if necessary, we assume that $\operatorname{int} W_{1,2} \neq \emptyset$.

Consider a technology A where $A_i = A_i^1 \cup \{a_i^2, \ldots, a_i^K\}$ for $i = 1, 2$, with K a number to be specified, and where $A_i = A_i^1$ for $i > 2$. We let $c_i(a_i^k) = \underline{c}_i^0$ for all i and k so that any action $a_i \notin A_i^0$ is a least-cost action for agent i.

We first define a collection of points used to define expected payments to action profiles containing actions in $A_i \setminus A_i^1$, $i = 1, 2$. Fix $\varepsilon > 0$ and $F_\varepsilon \in \Delta(Y)$ such that $F_\varepsilon(y_0) > 1-\varepsilon$ and $\mathbb{E}_{F_\varepsilon}[w_{1,2}(y)] \in \operatorname{int}(W_{1,2})$. F_ε will be our equilibrium outcome distribution. (We can simply take $F_\varepsilon = \delta_{y_0}$, if $w_{1,2}(y_0) \in \operatorname{int}(W_{1,2})$.) Fix $\underline{x} \in W$ such that $\mathbb{E}_{F_\varepsilon}[w_{1,2}(y)] > \underline{x}_{1,2}$.

Let x^0 be some point in $X := \{x(a) : a \in A^1,\ a \notin A^0\}$ (with $x(a)$ defined by (A.1)) that maximizes agent 1's payoff on X. As both $x_{1,2}^0$ and $\mathbb{E}_{F_\varepsilon}[w_{1,2}(y)]$ are in the interior of the

	A_2^0	a_2^1	a_2^2	a_2^3	$\cdots$	a_2^{K-1}	a_2^K
A_1^0	$\cdots$	$\cdot$	u^0	u^0	$\cdots$	u^0	u^0
a_1^1	$\cdot$	$\cdot$	u^1	u^1	$\cdots$	u^1	u^1
a_1^2	u^0	$\mathbf{x^1}$	$\mathbf{x^2}$	u^2	$\cdots$	u^2	u^2
a_1^3	u^0	u^1	$\mathbf{x^3}$	$\mathbf{x^4}$	$\cdots$	u^3	u^3
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\ddots$	$\cdots$	$\cdots$
a_1^{K-1}	u^0	u^1	u^2	u^3	$\cdots$	$\mathbf{x^{2K-4}}$	u^{K-1}
a_1^K	u^0	u^1	u^2	u^3	$\cdots$	$\mathbf{x^{2K-3}}$	$\mathbf{x^{2(K-1)}}$

Figure 5. Expected payments to agents 1 and 2 (for any fixed $a_{-\{1,2\}} \in A_{-\{1,2\}}$) in the game $\Gamma(w, A)$ in the proofs of Lemmas 2 and 3.

convex set $W_{1,2}$ (the former by observation 2.), we can choose from W points $x^1, \ldots, x^{2(K-1)}$, with $x^{2(K-1)} = \mathbb{E}_{F_\varepsilon}[w(y)]$, such that the sequence $(x^0, \ldots, x^{2(K-1)})$ satisfies the following conditions (where any cases involving $k < 0$ or $k > 2(K-1)$ can be ignored):

1. Among two consecutive points, agent 1 prefers the odd one: for all $k = 0, \ldots, K-1$,

$$x_1^{2k-1} > x_1^{2k} \quad \text{and} \quad x_1^{2k+1} > x_1^{2k}.$$

2. Agent 2 has the opposite preference: for all $k = 0, \ldots, K-1$,

$$x_2^{2k-1} < x_2^{2k} \quad \text{and} \quad x_2^{2k+1} < x_2^{2k}.$$

3. Both prefer each point in the sequence to $\underline{x}$: for all $k = 0, \ldots, 2(K-1)$, $x_{1,2}^k > \underline{x}_{1,2}$.

(See Figure 3, where this sequence is $(x^0, x^1, x^2, x^3, x^4)$, so $K = 3$.) Note that the sequence can be constructed by first choosing a desired sequence in $W_{1,2}$, and then defining the remaining coordinates arbitrarily subject to feasibility.

With K fixed by the above sequence, we choose another sequence $(u^0, \ldots, u^{K-1})$ in W such that $u_{1,2}^{K-1} > \cdots > u_{1,2}^0$ and $x_{1,2}^k > u_{1,2}^l$ for any k and l. Say, as $x_{1,2}^k > \underline{x}_{1,2}$ for each k, we can use convex combinations of $x^{2(K-1)}$ and $\underline{x}$ with large enough weights on $\underline{x}$. (In Figure 3 we can take $\underline{x} = w(y_0)$, so these points would lie between x^4 and $w(y_0)$, close to $w(y_0)$.)

To complete the description of A, we assume that any action profile involving actions $\{a_i^2, \ldots, a_i^K\}$, $i = 1, 2$, leads to an expected payment profile as specified in Figure 5. Furthermore, we assume that the profile $x^{2(K-1)} = \mathbb{E}_{F_\varepsilon}[w(y)]$ is generated by the distribution F_ε, i.e., $F(a_1^K, a_2^K, a_{-\{1,2\}}) = F_\varepsilon$ for all $a_{-\{1,2\}} \in A_{-\{1,2\}}$. For all other expected payment profiles, any distribution $F \in \Delta(Y)$ that generates them will do.

Uniqueness of the Equilibrium Outcome Distribution in Case 2. We claim that every equilibrium of $\Gamma(w, A)$ has agents 1 and 2 playing actions a_1^K and a_2^K. Note that this leads to the outcome distribution F_ε that puts at least probability $1-\varepsilon$ on y_0. Therefore, letting $\varepsilon \to 0$ yields a sequence of technologies with the properties listed in the lemma.

The claim follows from the following two lemmas:

Lemma A.2. *Let $\sigma \in \mathcal{E}(w, A)$. If $\sigma_i(A_i^0) = 0$ for some i, then $\sigma_1(a_1^K) = 1$ and $\sigma_2(a_2^K) = 1$.*

Proof. Let $\sigma \in \mathcal{E}(w, A)$. Suppose first that $\sigma_2(A_2^0) = 0$. (The case $\sigma_1(A_1^0) = 0$ is handled analogously.) We can then eliminate column A_2^0 in Figure 5. Then a_1^1 strictly dominates any $a_1 \in A_1^0$ for agent 1. To see this, note that $u_1^1 > u_1^0$ by construction. And if agent 2 plays a_2^1, then a_1^1 is the unique best-response among actions in $A_1^1 = A_1^0 \cup \{a_1^1\}$ by observation 1. Therefore, $\sigma_1(A_1^0) = 0$, and we can eliminate row A_1^0 in Figure 5. But now that row A_1^0 and column A_2^0 are both eliminated, the remaining matrix is by construction solvable by iterated elimination of strictly dominated strategies: a_1^2 dominates a_1^1 for agent 1 since his payoff in the cell (a_1^1, a_2^1) in Figure 5 is at most x_1^0, and we have $x_1^0 < x_1^1$, $u_1^1 < x_1^2$, and $u_1^1 < u_1^2$. Similarly, once a_1^1 is eliminated, a_2^2 dominates a_2^1 for agent 2 since $x_2^1 < x_2^2$, $u_2^1 < x_2^3$, and $u_2^1 < u_2^2$. Continuing iteratively, we see that only the cell (a_1^K, a_2^K) remains. (Note that we need not consider costs here as each new action is equally costly.)

Suppose then that $\sigma_i(A_i^0) = 0$ for some $i > 2$. Then the payoffs in the four cells in the top-left corner of the matrix in Figure 5 are given by equation (A.1). This implies that a_2^1 strictly dominates all $a_2 \in A_2^0$ for agent 2 by observation 1., and the fact that $u_2^1 > u_2^0$ by construction. Therefore, $\sigma_2(A_2^0) = 0$, implying that we are back in the first case. □

Lemma A.3. *In every equilibrium σ of $\Gamma(w, A)$, we have $\sigma_i(A_i^0) = 0$ for some agent i.*

The proof is essentially the same as that of Lemma A.1.

Proof. Define the events $E := \cap_i E_i$, $E_i := \{a \in A : a_i \in A^0\}$, and suppose $\sigma(E) > 0$ for some $\sigma \in \mathcal{E}(w, A)$. The second paragraph in the proof of Lemma A.1 applies verbatim. The only difference is in the third paragraph:

If $j > 2$, then conditional on some agent $i \neq j$ playing $a_i \notin A_i^0$, a_j^1 may now be only weakly better than $\hat{a}_j$. This is because, if $a_i = a_i^k$ for some $i \in \{1, 2\}$ and $k \geq 2$, then agent j's action doesn't affect the outcome. However, a_j^1 is still a best-response as it is a least-cost action. This is enough to get the contradiction, because a_j^1 does strictly better than $\hat{a}_j$ conditional on $\cap_{i \neq j} E_i$.

If $j \in \{1, 2\}$, then a_j^1 does strictly better than $\hat{a}_j$ whenever some agent $i \neq j$ plays $a_i \notin A_i^0$ by observation 1., and the fact that $u_j^1 > u_j^0$ by construction. A contradiction. □

As promised, we comment here on how to accommodate agents for whom w_i is constant. Let J be the set of such agents. We have $|J| \leq I-2$, since w fails to align the agents' interests. We can then apply the above construction to agents $\{1, \ldots, I\} \setminus J$, setting $A_j = A_j^0$ for all $j \in J$ and letting the new actions of agents not in J dictate the outcome in all of the technologies considered. The above analysis still applies, with the obvious modification that any "all i"-statement now means "all $i \notin J$" where relevant.[9] In every equilibrium, any agent $j \in J$ will then play some least-cost action(s) in A_j^0, without affecting the outcome. For example, in Lemma A.1, the result is now that every equilibrium has each agent $i \notin J$ playing a_i^1. All other results hold as stated. □

Proof of Lemma 3. The proof has many elements in common with that of Lemma 2, but here we also need to manipulate the agents' costs, leading to some important differences.

Since $Y^* \neq \emptyset$, there exists $z \in W$ such that $z \geq x$ for all $x \in W$. (Just take $z = w(y)$ for any $y \in Y^*$.) This implies that the projection $W_{i,j}$ of W is two-dimensional for some i and j as otherwise w would align the agents' interests. Without loss, take this pair to consist of agents 1 and 2. Note that the interior of $W_{1,2}$ relative to $\mathbb{R}_+^2$, or $\text{int}(W_{1,2})$, is nonempty. Recall that $x_{1,2}$ denotes the image of $x \in W$ in $W_{1,2}$.

We will construct a technology A where $A_i = A_i^0 \cup \{a_i^1, \ldots, a_i^K\}$ for $i = 1, 2$, with K to be specified, and where $A_i = A_i^0 \cup \{a_i^1\}$ for $i > 2$.

Fix $\eta \in (0, 1)$. Let $c_i(a_i^k) = \eta \underline{c}_i^0$ for each agent i and $k = 1, \ldots, K$. Then the new actions are strictly cheaper than any known action with a positive cost: if $c_i(a_i) > 0$ for some $a_i \in A_i^0$, then at least one inequality in $c_i(a_i) \geq \underline{c}_i^0 \geq \eta \underline{c}_i^0$ is strict.

In order to define the expected payments, fix $\varepsilon > 0$ and let $F_\varepsilon \in \Delta(Y)$ be a distribution such that $F_\varepsilon(y_0) > 1-\varepsilon$ and $\mathbb{E}_{F_\varepsilon}[w_{1,2}(y)] \in \text{int}(W_{1,2})$. (If $w_{1,2}(y_0) \in \text{int}(W_{1,2})$, we can simply take $F_\varepsilon = \delta_{y_0}$.) F_ε will be our equilibrium outcome distribution. Because $\mathbb{E}_{F_\varepsilon}[w_{1,2}(y)] < z_{1,2}$, we can then find a point $x^0 \in W$ such that (i) $z_{1,2} > x_{1,2}^0 > \mathbb{E}_{F_\varepsilon}[w_{1,2}(y)]$, (ii) $x_{1,2}^0 \in \text{int}(W_{1,2})$, and (iii) x^0 is close enough to z so that

$$\sum_i (x_i^0 - \eta \underline{c}_i^0) > \sum_i \mathbb{E}_{F(a)}[w_i(y) - c_i(a_i)] \quad \text{for all } a \in A^0. \tag{A.2}$$

To see why (A.2) can be satisfied, fix $a \in A^0$. If $\text{supp}\, F(a) \subseteq Y^*$ so that $\mathbb{E}_{F(a)}[w_i(y)] = z_i$ for all i, then by the assumption in the lemma, $c_j(a_j) > 0$ for some j. Then $\eta \underline{c}_j^0 < c_j(a_j)$ and thus the inequality in (A.2) holds for any x^0 such that $\sum_i x_i^0$ is close enough to $\sum_i z_i$. If $\text{supp}\, F(a) \not\subseteq Y^*$, then $\sum_i z_i > \sum_i \mathbb{E}_{F(a)}[w_i(y)]$, and again the inequality holds for any x^0

[9] In particular, Case 1 is now defined as having $w_i(y_0) \geq x_i(a_{-\{J\}}^1)$ for some agent $i \notin J$, whereas in Case 2, we have $w_i(y_0) < x_i(a_{-\{J\}}^1)$ for all $i \notin J$. (The notation $a_{-\{J\}}^1$ reflects the fact that $x(a)$ is still defined by (A.1) to only depend on actions of agents not in J.)

such that $\sum_i x_i^0$ is close enough to $\sum_i z_i$. As A^0 is finite, some $x^0 < z$ thus satisfies (A.2).

We let the profile of expected payments to be $x(a) = x^0$ for any $a \in A$ such that $a_i \in A_i^0$ for $i = 1, 2$ and $a_j = a_j^1$ for at least one agent $j > 2$.

We then define expected payments for action profiles involving new actions of agent 1 and 2. By the above conditions (i) and (ii), we can pick z^1 and z^2 in W such that $z_i^i > \mathbb{E}_{F_\varepsilon}[w_i(y)]$, $z_i^i > x_i^0 > z_i^j$, $j \neq i$. (See Figure 4.) Then we complete the construction of the sequence $(x^0, \ldots, x^{2(K-1)})$ and choose the sequence $(u^0, \ldots, u^{K-1})$ exactly as in the proof of Lemma 2. Finally, given any $a_{-\{1,2\}} \in A_{-\{1,2\}}$, we assign the expected payment profiles to agent 1 and 2's actions as in Figure 5, where the top-left corner is now given by the matrix

	A_2^0	a_2^1
A_1^0	$\cdots$	z^2
a_1^1	z^1	x^0

that reflects elements specific to the current construction. Any distributions generating these payoffs will do, except for $x^{2(K-1)}$, which is generated by the distribution $F_\varepsilon \in \Delta(Y)$.

We claim that every equilibrium of the game $\Gamma(w, A)$ so constructed has agents 1 and 2 playing a_1^K and a_2^K, and thus the unique equilibrium distribution is F_ε, which assigns at least probability $1 - \varepsilon$ to y_0. As $\min c_i(A_i) = \eta \underline{c}_i^0$, letting $\varepsilon \to 0$ and $\eta \to 1$ simultaneously (say, put $1 - \eta = \varepsilon \to 0$) yields a sequence of technologies with the desired properties.

To prove the claim, we show first that agents 1 and 2 must play a_1^K and a_2^K with probability 1 in every equilibrium σ where $\sigma_j(A_j^0) = 0$ for some agent j. Indeed, suppose this holds for some $j > 2$. Then the payoff in the cell (A_1^0, A_2^0) in Figure 5 is x^0. But then a_i^1 dominates all actions in A_i^0 for agents $i = 1, 2$, and iterated elimination leads to the profile (a_1^K, a_2^K) as desired. If instead $j \in \{1, 2\}$, say, $j = 1$, then the top row in Figure 5 is eliminated, and so a_2^1 dominates all actions in A_2^0 for agent 2, and iterated elimination again leads to (a_1^K, a_2^K). The case $j = 2$ is handled similarly.

It remains to show that every $\sigma \in \mathcal{E}(w, A)$ has $\sigma_j(A_j^0)$ for some j. Let $E := \cap_i E_i$, with $E_i := \{a \in A : a \in A^0\}$. Suppose towards contradiction that $\sigma(E) > 0$. Then (A.2) implies that conditional on E, the expected payoff of some agent j is strictly less than $x_j^0 - \eta \underline{c}_j^0$. Thus, some $\hat{a}_j \in A_j^0$ with $\sigma_j(\hat{a}_j) > 0$ yields agent j a payoff strictly less than $x_j^0 - \eta \underline{c}_j^0$ conditional on $\cap_{i \neq j} E_i$. Hence, a_j^1 gives a strictly higher payoff than $\hat{a}_j$ given $\cap_{i \neq j} E_i$. Moreover, if any agent $i \neq j$ plays $a_i \notin A_i^0$, then a_j^1 still gives at least as high a payoff as $\hat{a}_j$: for $j > 2$ this is because agent j's action then doesn't affect the outcome and a_j^1 is a least-cost action; for $j = 1, 2$ this is because a_j^1 then dominates any $a_j \in A_j^0$ by construction. Therefore, a_j^1 yields a higher (unconditional) expected payoff than a_j, contradicting $\sigma_j(\hat{a}_j) > 0$. $\square$

Proof of Corollary 1. By Lemmas 2 and 3, it suffices to show that if there exists $a^* \in A^0$ such that $\operatorname{supp} F(a^*) \subseteq Y^*$ and $c(a^*) = 0$, then there exists a technology $A \supseteq A^0$ such that for all $\sigma \in \mathcal{E}(w, A)$, we have $F(\{y \in Y : v(y) = 0\}|\sigma) = 1$. So fix any such a^*.

Note that if a^* satisfies full support, then $F(a^*) = \delta_{y_0}$ (since $Y^* \neq Y$ if w fails to align the agents' interests), and hence $y_0 \in Y^*$. If a^* satisfies costly production, then $\mathbb{E}_{F(a^*)}[v(y)] = 0$, and again $\{y \in Y : v(y) = 0\} \cap Y^* \neq \emptyset$. Thus, either way, $v(y^*) = 0$ for some $y^* \in Y^*$.

Let A be the technology constructed in the first part of the proof of Lemma 4 with $v(y^*) = 0$. As noted there, if $\sigma \in \mathcal{E}(w, A)$, then the agents only play zero cost actions under σ and $\operatorname{supp} F(\sigma) \subseteq Y^*$, since otherwise some agent i could profitably deviate to a_i'. Therefore, $\sigma(a) > 0$ for $a \in A^0$ only if (i) $c(a) = 0$ and (ii) $\operatorname{supp} F(a) \subseteq Y^*$. But we saw above that $F(\{y \in Y : v(y) = 0\}|a) = 1$ for all such a. On the other hand, if $a \notin A^0$, then $F(a) = \delta_{y^*}$, where $v(y^*) = 0$. Thus A has the desired property. □

Proof of Corollary 3. If w is budget balanced so that $\sum_i w_i(y) = v(y)$ for all $y \in Y$, then $Y^* \subseteq \arg\max_{y \in Y} \sum_i w_i(y) = \arg\max_{y \in Y} v(y)$. The assumption in Corollary 3 then implies that the case covered by Lemma 4 cannot arise, so the claim follows by Lemmas 2 and 3. □

A.2 Proofs for Section 4

We first restate and prove part (i) of Theorem 2:

Lemma A.4. *There exists a linear team-optimal contract.*

Proof. By Lemma 1, we can identify the space of linear budget-balanced contracts with the compact set $B := \{\alpha \in [0,1]^I : \sum_i \alpha_i = 1\}$. We will denote such a contract simply by α. As noted in the discussion following Theorem 2 in the main text, it suffices to show that the guaranteed expected surplus $S(\alpha)$ is an upper semi-continuous function of α on B.[10] Fix a sequence (α^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 σ^n the equilibrium of $\Gamma(\alpha, A)$ that achieves $S(\alpha^n, A)$. Extracting a further subsequence if necessary, we can assume that the sequence (σ^n) converges to some $\sigma \in \Delta(A)$. Since the agents' payoffs are continuous in α, the profile σ is an equilibrium of

[10] The argument that follows parallels Carroll's (2015) proof of existence of an optimal linear contract with general cost lower bounds in the single-agent case.

$\Gamma(\alpha, A)$ by the upper hemi-continuity of the Nash equilibrium correspondence. We thus have

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

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

We then prove Lemma 6.

Proof of Lemma 6. Fix any w and A as in the statement of the Lemma. Define positive "auxiliary shares" $\tilde{\alpha}_i \in (0, 1]$, $i = 1, \ldots, n$, by setting $\tilde{\alpha}_i = \alpha_i$ if $\alpha_i > 0$, and otherwise letting $\tilde{\alpha}_i > 0$ be any number small enough such that

$$\tilde{\alpha}_i < \frac{\min\{|c_i(a_i) - c_i(a_i')| : a_i, a_i' \in A_i,\ c_i(a_i) \neq c_i(a_i')\}}{\max v(Y)}.$$

(Note that $\sum_i \tilde{\alpha}_i > 1$ if $\alpha_i = 0$ for some i.) Define the function $\tilde{P} : A \to \mathbb{R}$ by

$$\tilde{P}(a) := \mathbb{E}_{F(a)}[v(y)] - \sum_i \frac{c_i(a_i)}{\tilde{\alpha}_i}. \tag{A.3}$$

It is straightforward to verify that then, for every agent i and every a_i, a_i', and a_{-i},

$$u_i(a_i, a_{-i}) - u_i(a_i', a_{-i}) > 0 \quad \text{implies} \quad \tilde{P}(a_i, a_{-i}) - \tilde{P}(a_i', a_{-i}) > 0.$$

Indeed, if $\alpha_i > 0$, then $u_i(a_i, a_{-i}) - u_i(a_i', a_{-i}) = \alpha_i(\tilde{P}(a_i, a_{-i}) - \tilde{P}(a_i', a_{-i}))$. If $\alpha_i = 0$, then $u_i(a_i, a_{-i}) - u_i(a_i', a_{-i}) > 0$ implies $c_i(a_i) < c_i(a_i')$, and so the choice of $\tilde{\alpha}_i$ implies

$$\tilde{P}(a_i, a_{-i}) - \tilde{P}(a_i', a_{-i}) = \mathbb{E}_{F(a_i, a_{-i})}[v(y)] - \mathbb{E}_{F(a_i', a_{-i})}[v(y)] - \frac{c_i(a_i) - c_i(a_i')}{\tilde{\alpha}_i} > 0.$$

Thus, $\tilde{P}$ is a generalized ordinal potential for $\Gamma(w, A)$, and hence $\arg\max_{a \in A} \tilde{P}(a) \subseteq \mathcal{E}(w, A)$. In particular, there exists a pure-strategy equilibrium.

It remains to establish the inequalities. Fix an equilibrium $a^* \in \arg\max_{a \in A} \tilde{P}(a)$. Then $\alpha_i \mathbb{E}_{F(a^*)}[v(y)] - c_i(a_i^*) \geq -\underline{c}_i^0$, as otherwise agent i could deviate to a least-cost action in A_i^0. Consider then the first inequality. If $U^0(w) = -\infty$, it holds vacuously. So let $U^0(w) > -\infty$. As noted after the definition of $U^0(w)$ in (4.2), then $\alpha_j = 0$ implies $\underline{c}_j^0 = 0$. Moreover, we have $c_j(a_j^*) = 0$ for any such agent j, since otherwise he could deviate to a zero-cost action in A_j^0. But then $c_i(a_i^*)/\tilde{\alpha}_i = c_i(a_i^*)/\alpha_i$ for all i (since $\tilde{\alpha}_i = \alpha_i > 0$ or $c_i(a_i^*) = 0$), and we have

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

where the first inequality follows since $A \supseteq A^0$, and the second follows since $\alpha_i \leq \tilde{\alpha}_i$. □

As preparation for the proof of Lemma 5, the following lemma establishes some properties of the solutions to the minimization problem.

Lemma A.5. *If (E, b) achieves the minimum in Lemma 5, then*

(i) $\alpha_i = 0$ *implies* $b_i = 0$ $(\forall i = 1, \ldots, I)$.

(ii) (4.3) *holds with equality.*

Furthermore, there exists a minimizer (E, b) satisfying the following additional properties:

(iii) $Eb_i = 0$ $(\forall i = 1, \ldots, I)$.

(iv) $E = \max\{U^0(w), 0\}$.

(v) $b_i \leq \underline{c}_i^0$ $(\forall i = 1, \ldots, I)$.

(The proof shows that properties (iii)–(v) are in fact necessary unless $\alpha_i = 1$ for some i.)

Proof. A minimizer exists since we are minimizing a continuous function over a compact set.

We first deal with the case $U^0(w) = \max v(Y)$. It is straightforward to verify that then $(E, b) = (U^0(w), 0)$ is the only feasible point, and that it satisfies properties (i)–(v).

From now on, let $\max v(Y) > U^0(w) > -\infty$. Note that if $\alpha_i = 0$, then $\underline{c}_i^0 = 0$ (since $U^0(w) > -\infty$), and so (4.4) implies that only $b_i = 0$ is feasible. This shows property (i).

To show (ii), suppose to the contrary that $E - \sum_i b_i/\alpha_i > U^0(w)$ for some minimizer (E, b). Then equality must hold in (4.4) for all i, since otherwise we could increase some b_i. Thus, $b_i = \alpha_i E + \underline{c}_i^0$, which we can substitute for b_i in (4.3). Rearranging then gives

$$E(1 - |\{i : \alpha_i > 0\}|) > U^0(w) + \sum_i \frac{\underline{c}_i^0}{\alpha_i} = \max_{a \in A^0} \Big(\mathbb{E}_{F(a)}[v(y)] - \sum_i \frac{c_i(a_i) - \underline{c}_i^0}{\alpha_i} \Big) \geq 0,$$

where the equality is by definition of $U^0(w)$, and the last inequality follows because $a \in A^0$ with $c_i(a_i) = \underline{c}_i^0$ for all i is feasible. But $|\{i : \alpha_i > 0\}| \geq 1$, and thus $E(1 - \{i : \alpha_i > 0\}|) \leq 0$, contradicting the strict inequality in the other direction above. This establishes (ii).

If there exists some agent i with $\alpha_i = 1$, then we can clearly satisfy properties (iii)-(v) by letting $(E, b) = (U^0(w), 0)$ if $U^0(w) \geq 0$, and letting $(E, b_i, b_{-i}) = (0, -U^0(w), 0)$ otherwise.[11]

For the rest of the proof, we assume that $\alpha_i < 1$ for all i, and we show that any minimizer satisfies properties (iii)–(v). Let $b_j = 0$ for all $j \in J^0 := \{i : \alpha_i = 0\}$. Consider minimization only over E and b_i, $i \in \{1, \dots, I\} \setminus J^0$. By inspection, the feasible set is a convex polyhedron in $\mathbb{R}^{I+1-|J^0|}$ with a nonempty interior, and the objective function is affine. The following Kuhn-Tucker conditions are thus necessary for a minimum:

$$\beta \geq 0,\ \mu_i \geq 0,\ \eta \geq 0,\ \lambda \geq 0,\ \theta_i \geq 0, \tag{A.4}$$

$$1 - \lambda - \sum_i \theta_i \alpha_i - \eta + \beta = 0, \tag{A.5}$$

$$-1 + \frac{\lambda}{\alpha_i} - \mu_i + \theta_i = 0, \tag{A.6}$$

$$\beta(\bar{V} - E) = 0,\ \mu_i b_i = 0,\ \eta E = 0,\ \lambda\Big(E - \sum_i \frac{b_i}{\alpha_i} - U^0(w)\Big) = 0,\ \theta_i(\alpha_i E - b_i + \underline{c}_i^0) = 0, \tag{A.7}$$

where i ranges over all agents with $\alpha_i > 0$, and where $\bar{V} := \max v(Y)$.

We can now show property (iii). Let (E, b) be a minimizer. If $E = 0$, then we are done, so let $E > 0$. We show first that $\theta_i = 0$ for all i. Suppose to the contrary that $\theta_i > 0$ for some i. Note that if $\theta_i > 0$, then $b_i = \alpha_i E + \underline{c}_i^0 > 0$ by (A.7), implying that $\mu_i = 0$. Thus, multiplying both sides of (A.6) by α_i and then summing over all i such that $\theta_i > 0$ gives

$$0 = \sum_{i:\theta_i>0} (-\alpha_i + \lambda) + \sum_i \theta_i \alpha_i = \sum_{i:\theta_i>0} (-\alpha_i + \lambda) + 1 - \lambda + \beta,$$

where the second equality substitutes for $\sum_i \theta_i \alpha_i$ using (A.5), noting that $\eta = 0$ by (A.7). Rearranging the terms yields

$$\sum_{i:\theta_i>0} \alpha_i = 1 + \beta + (|\{i : \theta_i > 0\}| - 1)\lambda \geq 1 + \beta \geq 1,$$

since $\theta_i > 0$ for at least one agent by assumption. But $\sum_{i:\theta_i>0} \alpha_i \geq 1$ implies that we must have $\theta_i > 0$ for all i such that $\alpha_i > 0$, and hence (4.4) holds as equality for every such agent by (A.7). Since (4.4) is an equality also for each agent in J^0, it thus holds as an equality for every agent, leading to a contradiction with (4.3) as in the proof of property (i) above. We

[11] To see that (4.4) and property (v) are satisfied for agent i, note that

$$\alpha_i E - b_i = E - b_i = U^0(w) = \max_{a \in A^0 \,:\, c_j(a_j)=0\ \forall j \neq i} \big(\mathbb{E}_{F(a)}[v(y)] - c_i(a_i)\big) \geq -\underline{c}_i^0.$$

conclude that $\theta_i = 0$ for all i. This implies that $\lambda = 1 + \beta$ by (A.5), so that (A.6) becomes

$$-1 + \frac{1+\beta}{\alpha_i} - \mu_i = 0,$$

which in turn implies $\mu_i > 0$, as $\alpha_i < 1$ by assumption. Hence, $b_i = 0$ by (A.7), as desired.

Property (iv) follows from properties (ii) and (iii). Namely, if $U^0(w) \geq 0$, then only $(E, b) = (U^0(w), 0)$ is consistent with both (ii) and (iii). On the other hand, if $U^0(w) < 0$, then $b_i > 0$ for some i by (ii), which by (iii) implies $E = 0$.

It remains to show property (v). If $U^0(w) \geq 0$, then (ii) and (iv) imply $b_i = 0 \leq \underline{c}_i^0$ for all i. If $U^0(w) < 0$, then $E = 0$ by (iv), and (4.4) implies $b_i \leq \underline{c}_i^0$ for all i. □

The next lemma will be used to show that $S(w)$ is not greater than the minimum in Lemma 5. It will also be used to characterize $V(w)$ in Section 5, which is why we state it in a form that emphasizes the uniqueness of the equilibrium distribution of outcomes.

Lemma A.6. *Let w be a budget-balanced contract that aligns the agents' interests. Suppose $\max v(Y) > U^0(w) > -\infty$. Let $(E, b) \in [0, \max v(Y)] \times \mathbb{R}_+^I$ and $G \in \Delta(Y)$ be such that*

(i) $E > \max\{U^0(w), 0\}$,

(ii) (4.3) *holds with strict inequality,*

(iii) $b_i \leq \underline{c}_i^0$ *for all i (and thus* (4.4) *is satisfied), and*

(iv) $\mathbb{E}_G[v(y)] = E$.

Then there exists a technology $A \supseteq A^0$ such that every $\sigma \in \mathcal{E}(w, A)$ satisfies $F(\sigma) = G$ (and hence $\mathbb{E}_{F(\sigma)}[v(y)] = E$) and $\sum_{a_i} \sigma_i(a_i) c_i(a_i) = b_i$ for all i.

Proof of Lemma A.6. Fix (E, b) and G as in the lemma. Let $J^0 = \{i : \alpha_i = 0\}$. Note that $0 \leq |J^0| < I$, because w is budget balanced. Moreover, we have $\underline{c}_i^0 = 0 = b_i$ for all $i \in J^0$, since $U^0(w) > -\infty$. We will construct a technology $A \supseteq A^0$ where $A_i = A_i^0 \cup \{a_i'\}$ with $c_i(a_i') = b_i$ for all $i \notin J^0$, and $A_i = A_i^0$ for all $i \in J^0$. Equilibria of $\Gamma(w, A)$ will consist of profiles where each agent $i \notin J^0$ plays a_i' and agents in J^0 mix over zero-cost actions.

We define outcome distributions for action profiles involving the new actions as follows. Let $\varepsilon_I := E - \max\{U^0(w), 0\}$ and let $0 \equiv \varepsilon_{|J^0|} < \varepsilon_{|J^0|+1} < \cdots < \varepsilon_I$, to be used to provide strict incentives. We will assume that each ε_k is small enough to satisfy the finitely many restrictions imposed on it by (A.9) below. Fix $a \in A$ and let $J := \{i : a_i = a_i'\} \cup J^0$. Suppose

$|J^0| < |J|$ so that at least one agent plays the new action in a. If equality holds in[12]

$$\max_{a_J \in A_J^0} \left(\mathbb{E}_{F(a_J, a_{-J})}[v(y)] - \sum_{j \in J} \frac{c_j(a_j) - b_j}{\alpha_j} \right) \leq \max v(Y), \tag{A.8}$$

then we take $F(a)$ to be any distribution such that $\mathbb{E}_{F(a)}[v(y)] = \max v(Y)$. Since $b_j \leq \underline{c}_j^0$ for every agent by assumption (iii), the only other possibility is that (A.8) holds with strict inequality instead. In that case we let $F(a)$ to be any distribution such that

$$\mathbb{E}_{F(a)}[v(y)] = \left[\max_{a_J \in A_J^0} \left(\mathbb{E}_{F(a_J, a_{-J})}[v(y)] - \sum_{j \in J} \frac{c_j(a_j) - b_j}{\alpha_j} \right) \right]^+ + \varepsilon_{|J|} < \max v(Y), \tag{A.9}$$

where $[r]^+ := \max\{r, 0\}$ for $r \in \mathbb{R}$. (This defines at most finitely many inequalities involving $\varepsilon_{|J|}$, because A is finite.)

Note that if $|J| = I$ so that every agent $i \notin J^0$ plays a_i' in the profile a, then the left-hand side of (A.8) equals $U^0(w) + \sum_i b_i/\alpha_i < E \leq \max v(Y)$, where the strict inequality is by assumption (ii). Thus $F(a)$ satisfies (A.9) and $\mathbb{E}_{F(a)}[v(y)] = [U^0(w)]^+ + \varepsilon_I = E$. We may thus put $F(a) = G$ for any such a.

The following lemma is the first step towards a characterization of $\mathcal{E}(w, A)$.

Lemma A.7. *Let $a \in A$ be an action profile where every agent in J^0 plays a zero-cost action. Then $u_i(a_i', a_{-i}) \geq u_i(a)$ for every agent $i \notin J^0$.*

Proof. Let $i \notin J^0$ and $\hat{a}_i \in A_i^0$. Fix $a_{-i} \in A_{-i}$ such that $c_j(a_j) = 0$ for all $j \in J^0$. Then

$$u_i(\hat{a}_i, a_{-i}) \leq \max_{a_i \in A_i^0} u_i(a_i, a_{-i}) = \alpha_i \max_{a_i \in A_i^0} \left(\mathbb{E}_{F(a_i, a_{-i})}[v(y)] - \frac{c_i(a_i)}{\alpha_i} \right). \tag{A.10}$$

Let $N := \{j \neq i : a_j = a_j'\}$. Suppose first that the maximum in (A.10) is achieved by a_i such that $F(a_i, a_{-i})$ is defined by (A.9), with $J = N \cup J^0$. We can then write out the far

[12]Recalling that $\underline{c}_i^0 = b_i = 0$ for all $i \in J^0$ shows that the maximum on the left-hand side is finite, since we can always choose zero-cost actions for every agent with $\alpha_i = 0$.

right-hand side of (A.10) as[13]

$$
\begin{aligned}
&\alpha_i \max_{a_i\in A_i^0}\left(\left[\max_{a_J\in A_J^0}\left(\mathbb{E}_{F(a_J,a_i,a_{-J\cup\{i\}})}[v(y)]-\sum_{j\in J}\frac{c_j(a_j)-b_j}{\alpha_j}\right)\right]^+ +\varepsilon_{|J|}-\frac{c_i(a_i)}{\alpha_i}\right)\\
&\le \alpha_i \min\left\{\left[\max_{a_{J\cup\{i\}}\in A^0_{J\cup\{i\}}}\left(\mathbb{E}_{F(a_{J\cup\{i\}},a_{-J\cup\{i\}})}[v(y)]-\sum_{j\in J\cup\{i\}}\frac{c_j(a_j)-b_j}{\alpha_j}\right)\right]^+ +\varepsilon_{|J|}-\frac{b_i}{\alpha_i},\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\left.\max v(Y)-\frac{b_i}{\alpha_i}\right\}\\
&\le \alpha_i\left(\mathbb{E}_{F(a_i',a_{-i})}[v(y)]-\frac{b_i}{\alpha_i}\right)=u_i(a_i',a_{-i}),
\end{aligned}
\tag{A.11}
$$

where the first inequality uses $b_i \le \underline{c}_i^0$ and the second inequality follows by definition of $F(a_i', a_{-i})$ (with $J = N \cup J^0 \cup \{i\}$), since $\varepsilon_{|J|} < \varepsilon_{|J\cup\{i\}|}$. Thus $u_i(\hat{a}_i, a_{-i}) \le u_i(a_i', a_{-i})$.

If instead the maximum in (A.10) is achieved by $\tilde{a}_i$ such that $\mathbb{E}_{F(\tilde{a}_i,a_{-i})}[v(y)] = \max v(Y)$, then (A.8) holds with equality. This implies that there exists some $\tilde{a}_J \in A_J^0$ such that $c_j(a_j) - b_j = 0$ for all $j \in J$ and $\mathbb{E}_{F(\tilde{a}_J,\tilde{a}_i,a_{-J\cup\{i\}})}[v(y)] = \max v(Y)$. The right-hand side of (A.10) now becomes

$$
\begin{aligned}
\alpha_i\left(\max v(Y)-\frac{c_i(\tilde{a}_i)}{\alpha_i}\right) &= \alpha_i\left(\mathbb{E}_{F(\tilde{a}_J,\tilde{a}_i,a_{-J\cup\{i\}})}[v(y)]-\sum_{j\in J\cup\{i\}}\frac{c_j(\tilde{a}_j)-b_j}{\alpha_j}-\frac{b_i}{\alpha_i}\right)\\
&= \alpha_i \max_{a_{J\cup\{i\}}\in A^0_{J\cup\{i\}}}\left(\mathbb{E}_{F(a_{J\cup\{i\}},a_{-J\cup\{i\}})}[v(y)]-\sum_{j\in J\cup\{i\}}\frac{c_j(a_j)-b_j}{\alpha_j}\right)-b_i\\
&\le \alpha_i\mathbb{E}_{F(a_i',a_{-i})}[v(y)]-b_i = u_i(a_i',a_{-i}),
\end{aligned}
\tag{A.12}
$$

where the last line follows by definition of $F(a_i', a_{-i})$ (with $J = N \cup J^0 \cup \{i\}$). This shows that $u_i(\hat{a}_i, a_{-i}) \le u_i(a_i', a_{-i})$ in this case as well. □

Lemma A.7 implies that any $\sigma \in \Delta(A)$ with $\sigma_i(a_i') = 1$ for $i \notin J^0$ and $\sigma_j(a_j)c_j(a_j) = 0$ for $j \in J^0$ is an equilibrium. In any such equilibrium, only profiles $a \in A$ with $|J| = I$ arise with positive probability. Therefore, $F(\sigma) = G$, $\mathbb{E}_{F(\sigma)}[v(y)] = E$ and $\sum_{a_i} \sigma_i(a_i)c_i(a_i') = b_i$. To rule out other equilibria, we need the following result.

Lemma A.8. *Let $a \in A$ be an action profile where every agent in J^0 plays a zero-cost action. If $a_{-J^0} \neq a'_{-J^0}$, then $u_i(a_i', a_{-i}) > u_i(a_i, a_{-i})$ for some agent i.*

Proof. Fix $a \in A$ as in the lemma. Define J as above. Assume towards contradiction that $u_i(a_i', a_{-i}) = u(a_i, a_{-i})$ for all $i \notin J$. Then, for each agent $i \notin J$, (A.11) or (A.12) holds

[13] If $N = \emptyset$, then by convention we ignore the maximization over $a_J \in A_J^0$ and the operator $[\cdot]^+$ on the first line of (A.11).

as a chain of equalities. By definition of the distributions $F(a_i', a_{-i})$, this is possible only if $\mathbb{E}_{F(a_i',a_{-i})}[v(y)] = \mathbb{E}_{F(a)}[v(y)] = \max v(Y)$ and $c_i(a_i) = b_i$ for all $i \notin J$; otherwise the choice of ε_k results in at least one strict inequality in each case. But note that, by definition of $F(a)$, we have $\mathbb{E}_{F(a)}[v(y)] = \max v(Y)$ only if

$$\begin{aligned}\max v(Y) &= \max_{a_J \in A_J^0} \left(\mathbb{E}_{F(a_J,a_{-J})}[v(y)] - \sum_{j \in J} \frac{c_j(a_j) - b_j}{\alpha_j} \right) \\ &= \max_{a_J \in A_J^0} \left(\mathbb{E}_{F(a_J,a_{-J})}[v(y)] - \sum_{i=1}^{I} \frac{c_i(a_i) - b_i}{\alpha_i} \right) \leq U^0(w) + \sum_i \frac{b_i}{\alpha_i} < E,\end{aligned}$$

where the second equality follows because $c_i(a_i) = b_i$ for all $i \notin J$, the weak inequality follows because in $U^0(w)$ the maximization is over all agents' actions, and the strict inequality is because (E, b) satisfies (4.3) with strict inequality by assumption (ii). Thus, $E > \max v(Y)$, contradicting $E \in [0, \max v(Y)]$. □

To complete the proof, observe that the agents in J^0 can clearly only play zero-cost actions in any equilibrium. Consider $\sigma \in \Delta(A)$ where this is true. Suppose $\sigma_i(a_i') < 1$ for some agent $i \notin J^0$. Then some profile $\hat{a}$ satisfying the assumptions of Lemma A.8 arises with positive probability under σ. Let i be an agent who can profitably deviate from $\hat{a}$ to a_i'; the existence of such an agent follows by Lemma A.8. If agent i deviates from σ_i to playing a_i' for sure, then his payoff increases strictly when the other agents play $\hat{a}_{-i}$ (which happens with positive probability under σ_{-i}), and it increases weakly against all other a_{-i} by Lemma A.7. Thus σ is not an equilibrium. □

We are now in a position to prove Lemma 5.

Proof of Lemma 5. By Lemma 6, every technology $A \supseteq A^0$ has a pure strategy equilibrium where the expected value of output E and costs b satisfy (4.3) and (4.4). Hence $S(w)$ is not less than the minimum of $E - \sum b_i$ over $E \in [0, \max v(Y)]$ and $b \in \mathbb{R}_+^I$ such that (4.3) and (4.4) are satisfied.

In the other direction, suppose first that $U^0(w) < \max v(Y)$. By Lemma A.5.(iv) and (v), there exists a minimizer (E, b) with $E = \max\{U^0(w), 0\}$ and $b_i \leq \underline{c}_i^0$ for all i. Let $E < E' \leq \max v(Y)$. Then (E', b) together with any G such that $\mathbb{E}_G[v(y)] = E'$ satisfies the assumptions in Lemma A.6, and hence there exists a technology $A \supseteq A^0$ such that $S(w) \leq S(w, A) = E' - \sum b_i$. Letting $E' \to E$ shows that $S(w)$ is not greater than the minimum.

If on the other hand $U^0(w) = \max v(Y)$, then $(E, b) = (\max v(Y), 0)$ achieves the minimum by Lemma A.5.(iii) and (iv). As $S(w) \leq \max v(Y)$ by feasibility, $S(w)$ is less than the minimum in this case as well. We conclude that $S(w)$ equals the minimum.

The properties of the minimizers follow from Lemma A.5. □

A.3 Proofs for Section 5

Proof of Lemma 8. 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 $E^n := \mathbb{E}_{G^n}[\bar{w}(y)] > \bar{U}^0(w)$ and $\mathbb{E}_{G^n}[v(y) - \bar{w}(y)] \to \pi$ as the objective is continuous in G.[14] Every $(E^n, 0)$ and G^n satisfy the assumptions of Lemma A.6 (with the substitutions $v(y) = \bar{w}(y)$ and $U^0(w) = \bar{U}^0(w)$), and thus there exists a technology $A^n \supseteq A^0$ for which G^n is the unique equilibrium distribution of outcomes. Hence, $V(w) \le V(w, A^n) = \mathbb{E}_{G^n}[v(y) - \bar{w}(y)] \to \pi$ 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 $\operatorname{supp} F(a) \subseteq \arg\max_{y \in Y} \bar{w}(y) = Y^*$, where the equality holds because w aligns the agents' interests. Thus $V(w) < V(0)$ by Lemma 4, 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 $\varepsilon > 0$ small enough. But $v(y_0) - \bar{w}(y_0) = -\bar{w}(y_0) \le 0$, implying that $\mathbb{E}_G[\bar{w}(y) - \bar{w}(y)] \le (1-\varepsilon)V(w) < V(w)$, which contradicts G^* being a minimizer. We conclude that $\mathbb{E}_{G^*}[\bar{w}(y)] = \bar{U}^0(w)$. □

Proof of Lemma 11. We adapt the proof of Lemma 3 in Carroll (2015) to the present setting. Let $B \subset \mathbb{R}^2$ be the convex hull of points $(\bar{w}(y), v(y) - \bar{w}(y))$, $y \in Y$. Let $C \subset \mathbb{R}^2$ be the set of all points (u, v) such that $u > \bar{U}^0(w)$ and $v < V(w)$. The sets B and C are disjoint by Lemma 8. Thus, by the separating hyperplane theorem, there exist numbers κ, λ, and μ, with $(\lambda, \mu) \neq (0, 0)$, such that

$$\kappa + \lambda u - \mu v \le 0 \quad \text{for all } (u, v) \in B, \tag{A.13}$$

and

$$\kappa + \lambda u - \mu v \ge 0 \quad \text{for all } (u, v) \in C. \tag{A.14}$$

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

$$\kappa + \lambda \mathbb{E}_{G^*}[\bar{w}(y)] - \mu \mathbb{E}_{G^*}[v(y) - \bar{w}(y)] = 0. \tag{A.15}$$

[14] 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.14) implies that $\lambda \geq 0$ and $\mu \geq 0$, so it suffices to show that these inequalities are strict.

Suppose towards contradiction that $\mu = 0$. Then $\lambda > 0$, and inequalities (A.13) and (A.14) 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 8 above it is shown that this contradicts the eligibility of w. On the other hand, if $\lambda = 0$, then $\mu > 0$, and (A.13) and (A.14) 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.

To complete the proof, rescale κ, λ, and μ such that $\mu = 1$. Then (A.13) implies (5.4) and (A.15) implies (5.5). □

A.4 Proofs for Section 6

Proof of Theorem 5. As noted in the main text, only Lemma 4 needs to be adjusted. Futhermore, the first part of Lemma 4, which shows that $V(w) < V(0)$, still goes through as before when the principal's payoff is $v(y) - \sum u_i^{-1}(w_i(y))$. (The cost of payments to the principal only appears in the proof in the displayed chain of inequalities establishing $V(w, A) < V(0)$, which clearly continues to hold after the cost is modified.) So it suffices to adjust the claim about the budget balanced case.

To this end, suppose a budget balanced contract w fails to align the agents' interests in utilities. As in Lemma 4, suppose there exists $a^* \in A^0$ such that $\operatorname{supp} F(a^*) \subseteq Y^*$ and $c(a^*) = 0$. Note that, by definition, any $y \in Y^*$ simultaneously maximizes the payments to all agents, and since the functions u_i are increasing, it also simultaneously maximizes their utilities gross of costs. Let $\bar{u}$ be the profile of these maximized utility levels, i.e., $\bar{u} = u(w(y))$ for any $y \in Y^*$. Note that $S(w) \leq \sum_i \bar{u}_i$.

Define the budget balanced improvement contract w' as follows. Let $w'(y) = w(y)$ for all $y \in Y^*$. For any other y, define $w'(y)$ such that

$$\sum_i w'_i(y) = v(y) \quad \text{and} \quad u(w'(y)) = \lambda \bar{u} \ \text{ for some } \lambda \in [0, 1]. \tag{A.16}$$

To see that this is feasible, fix some $y \notin Y^*$. We can rewrite (A.16) as

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

Note that w being budget balanced implies that any $y \in Y^*$ satisfies $v(y) = \max v(Y)$, and hence the left-hand side equals $\max v(Y)$ for $\lambda = 1$ by our choice of $\bar{u}$. On the other hand, it equals 0 when $\lambda = 0$, since $u_i(0) = 0$ by assumption. Thus by the intermediate value theorem, equality holds for some $\lambda \in [0, 1]$.

To finish the proof, we observe that $a^* \in \mathcal{E}(w', A)$ for all $A \supseteq A^0$. Indeed, a^* gives each agent his highest feasible payoff under w' (i.e., $\bar{u}_i$) as the outcome is then guaranteed to be in Y^* and $c(a^*) = 0$. Thus, $S(w') \geq \sum_i \bar{u}_i \geq S(w)$. □

Proof of Lemma 13. The proof is immediate from the properties of the power utility function, but we provide it for completeness. Fix w anchored at the origin. Suppose w aligns the agents' interests in utilities. Then all payment-utility profiles lie on a line segment from the origin to some $u(\bar{w}) \in \mathbb{R}^I_+$, where $\bar{w}$ is the profile of maximum payments to all agents. Fix any y. We claim that $w(y) = \lambda \bar{w}$ for some $\lambda \in [0, 1]$, and thus w aligns the agents' interests in the sense of Definition 1. To see this, note that there exists $\xi \in [0, 1]$ such that $w_i(y)^\rho = \xi \bar{w}_i^\rho$ for all i because w aligns the agents' interests in utilities. Hence, $w_i(y) = \xi^{\frac{1}{\rho}} \bar{w}_i$ as desired. The other direction is proven similarly. □

References

Bergemann, D., and S. Morris (2005): "Robust Mechanism Design," *Econometrica*, 73(6), 1771–1813.

Carroll, G. (2015): "Robustness and Linear Contracts," *American Economic Review*, 105(2), 536–563.

——— (2017): "Robustness and Separation in Multi-Dimensional Screening," *Econometrica*, 85(2), 453–488.

Carroll, G., and I. Segal (2017): "Robustly Optimal Auctions with Unknown Resale Opportunities," Discussion paper, Stanford University.

Chassang, S. (2013): "Calibrated Incentive Contracts," *Econometrica*, 81(5), 1935–1971.

Che, Y.-K., and S.-W. Yoo (2001): "Optimal Incentives for Teams," *American Economic Review*, 91(3), 525–541.

Chung, K.-S., and J. Ely (2007): "Foundations of Dominant-Strategy Mechanisms," *Review of Economic Studies*, 74(2), 447–476.

Frankel, A. (2014): "Aligned Delegation," *American Economic Review*, 104(1), 66–83.

Garicano, L., and T. Santos (2004): "Referrals," *American Economic Review*, 94(3), 499–525.

GARRETT, D. (2014): "Robustness of Simple Menus of Contracts in Cost-based Procurement," *Games and Economic Behavior*, 87, 631–641.

HACKMAN, R. (2002): *Leading Teams: Setting the Stage for Great Performances*. Harvard Business School Press, Boston, MA.

HOLMSTRÖM, B. (1982): "Moral Hazard in Teams," *The Bell Journal of Economics*, 13(2), 324–340.

HOLMSTRÖM, B., AND P. MILGROM (1987): "Aggregation and Linearity," *Econometrica*, 55(2), 303–328.

HURWICZ, L., AND L. SHAPIRO (1978): "Incentive Structures Maximizing Residual Gain Under Incomplete Information," *The Bell Journal of Economics*, 9(1), 180–191.

ITOH, H. (1991): "Incentives to Help in Multi-Agent Situations," *Econometrica*, 59(3), 611–636.

LAZEAR, E. (1989): "Pay Equality and Industrial Politics," *Journal of Political Economy*, 97(3), 561–580.

LEVIN, J., AND S. TADELIS (2005): "Profit Sharing and the Role of Professional Partnerships," *Quarterly Journal of Economics*, 120(1), 131–171.

MADARÁSZ, K., AND A. PRAT (2017): "Sellers with Misspecified Models," *Review of Economic Studies*, 84(2), 790–815.

MARKU, K., AND S. OCAMPO DIAZ (2017): "Robust Contracts in Common Agency," Discussion paper, University of Minnesota.

MONDERER, D., AND L. SHAPLEY (1996): "Potential Games," *Games and Economic Behavior*, 14, 124–143.

YAMASHITA, T. (2015): "Implementation in Weakly Undominated Strategies: Optimality of Second-Price Auction and Posted-Price Mechanism," *Review of Economic Studies*, 82(3), 1223–1246.