



**UNIVERSITY
OF WARWICK**

Economics

Contracts with Interdependent Preferences

Marek Weretka and Debraj Ray

October 2026

No: 1634

**Warwick
Economics
Research Papers**

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

Contracts with Interdependent Preferences

Marek Weretka (r) Debraj Ray[†]

August 2025, revised July 2026

Abstract. A principal contracts with multiple agents, as in Lazear and Rosen (1981) and Green and Stokey (1983). Agents have interdependent preferences, but otherwise the setting is classical. We characterize optimal contracts, and relate the direction of co-movement in rewards — “joint liability” (positive) or “tournaments” (negative) — to the assumed structure of preference interdependence. At the same time, as we range over a symmetric family of positively and negatively interdependent preferences, the corresponding distribution of optimal contracts leans towards commonality in payoffs (e.g., joint liability) rather than differences (e.g., tournaments), in a sense made precise in the paper. This asymmetry is qualified to some degree when the mechanism-design problem is augmented by robustness constraints designed to eliminate multiple equilibria.

1. INTRODUCTION

Imagine yourself as an economist designing contracts for a plantation in a developing country. The workers are fruit pickers or tea pluckers, drawn from nearby villages so that many are acquaintances or family, or have some social connection. Upon observing the prevalent payment structure, which consists of a fixed wage accompanied by a bonus tied to the quantity of individually harvested output, you — a firm believer in economic incentives — may be tempted to further enhance the incentive scheme. We don’t mean a reduction in the base wage and a larger slope on harvested output; these may be restricted either because of limited liability or mandated minimum wage payments (Jayaraman et al. 2016). You might resort to more intricate, *inter-agent incentives*, where the payoff to one worker is somehow related to the performance of her compatriots. Some classical references are Lazear and Rosen (1981), Holmstrom (1982), Green and Stokey (1983), Nalebuff and Stiglitz (1983) and Mookherjee (1984).

[†]Ray: New York University and University of Warwick, debraj.ray@nyu.edu; Weretka: University of Wisconsin-Madison and FAME/GRAPE. We thank Dilip Abreu, Yeon-Koo Che, Parikshit Ghosh, Navin Kartik, Meg Meyer, Dilip Mookherjee, Nick Netzer, Marco Ottaviani, David Pearce, Fernando Vega-Redondo, Lones Smith, Ennio Stacchetti and Eyal Winter for insightful comments. We are grateful to Claude Opus 4.8 for interactions that improved several technical arguments. Ray acknowledges support under National Science Foundation Grant SES-2315720. Weretka acknowledges the support of National Center for Science (grant 2019/33/B/HS4/00151).

Examples of such mechanisms are tournaments and joint liability mechanisms. They are often used; see Wantchekon (1994) and Bandiera, Barankay and Rasul (2005) for instances of the former, and Banerjee, Besley and Guinnane (1994), Besley and Coate (1995), Ghatak (1999) or Ghatak and Guinnane (1999) for examples of the latter. In a tournament, a worker is rewarded if she outperforms her colleague, creating a competitive environment between them. In contrast, joint liability mechanisms link higher worker compensation to better compatriot performance, generating an environment in which coordination or cooperation can be rewarding. Or one could envision less extreme variants, with standard pay-for-performance augmented by winners' prizes or joint bonuses. We are going to study the relationship between interdependent preferences and optimal contracting.

Substantial empirical evidence suggests that agents not only value material rewards but are also affected by the payoffs of their fellow agents. Sobel (2005) and Fehr and Gächter (2000) review the empirical literature on interdependent preferences. The importance of such preferences in the workplace is recognized in Bandiera et al. (2005), List (2009) and Luft (2016). Interdependence is particularly relevant in teams drawn from a particular geographical location such as a village, where agents are more likely to know one another.

It is with this emphasis in mind that we introduce payoff externalities — altruistic or adversarial — into an otherwise standard team agency problem. Our focus is firmly on how competition and cooperation relate to the sign of those externalities, a theme on which we elaborate below. Evolutionary arguments suggest that both altruism and envy could be present depending on the ambient environment, with independence being the exception rather than the rule; see Axelrod and Hamilton (1981), Koçkesen, Ok and Sethi (2000), Trivers (2006), Marsh (2016), and Robson (2017). But quite apart from evolutionary arguments, it is intuitive that in settings where individuals know one another, preferences will naturally intertwine owing to ties of family and kinship, and additionally for reasons of socioeconomic similarity, friendship, shared experiences, and past interactions. These connections could be friendly or antagonistic.

We study a principal who employs a number of risk-averse identical agents, known to one another. Each chooses a binary effort (work, shirk). Our basic assumption is simple: each agent assigns symmetric weight α to the welfare of other agents. This weight can be positive or negative, and it is commonly known to all parties. The principal's objective is to efficiently encourage "high effort" by offering contracts to the agents. A contract for any agent specifies a non-negative monetary transfer to that agent, conditional on a vector of signals, one for each agent. We deliberately constrain the analysis by supposing

that each agent’s signal only depends on her *own* effort. Therefore all cross-agent payoff-dependence must arise solely from payoff externalities and not via the signal correlations that have already been extensively studied in the literature. (It is easy to accommodate such correlations, but at the cost of blunting our desired focus.)

Incentivizing agents with interdependent preferences requires us to accommodate a central idea which can quickly turn quite complex. It is that a contract not only affects an agent’s incentives via her own material payoff, but also via the payoffs she generates for others. An optimal contract must take such payoff spillovers fully into account, leading to interdependent payments that depend on the entire vector of agent performances. In particular, we are interested in the *cross-impact* ξ on an agent’s contractual monetary reward when another agent’s performance changes, relative to the *own-impact* from self-performance. This relative cross-impact is pinned down by a particular sum of likelihood ratios that quantify agent performances, which entirely determines the extent of cooperativeness or competition in the joint contract. In this way, we are able to transparently characterize optimal contracts under nonpaternalistic preference interdependence and a general (though independent) signal structure. Moreover, while our baseline exercise is one of implementing the desired outcome as *an* equilibrium, we also pay attention to robust implementation, or the elimination of other undesirable equilibria.

But we do not wish to remain in the technical realm of simply linking cross- and own-impacts, or emphasizing generality for its own sake. We use our method to say something about the qualitative features of optimal contracts. Our key findings are outlined below:

(1) Optimal contracts span an entire spectrum, ranging from competitive to cooperative. Tournaments, in which agents are rewarded for outperforming their compatriots, exemplify the former. Joint liability, in which agents are rewarded for their own successes *and* those of others, exemplifies the latter. We link such contracts to altruistic and adversarial preferences (Proposition 1). Agent reward strictly increases in own performance, of course, but co-moves positively with partner performance when preference interdependence is altruistic ($\alpha > 0$), and negatively with partner performance when interdependence is antagonistic ($\alpha < 0$). Altruism generates cooperative contracts, antagonism generates competitive contracts. This is an intuitive yet foundational baseline result.

(2) Our next result, Proposition 2, describes the relative cross-impact on any agent’s reward. If there are n agents with symmetric interdependent preferences summarized by the parameter α , that relative cross-impact satisfies a simple formula that is independent

of risk attitudes or the performance vector at which cross-impact is evaluated:

$$\xi = \frac{\alpha}{1 - \alpha(n - 2)}.$$

Indeed, ξ only depends on team size and the extent of preference interdependence.

(3) The observation above tells us that altruistic teams with three or more members display greater relative cross-impact (in absolute terms) compared to adversarial teams with the same magnitude of preference interdependence. The asymmetry is heightened in larger teams. In an example with 20 team members, the relative cross-impact is 14% when agents place a 4% weight on a compatriot payoff. But a weight of *minus* 4% results in a relative cross-impact of about 2.3% in absolute terms. The asymmetry is particularly stark when passing to the two limits with maximally altruistic and maximally adversarial attitudes.¹ In the altruistic limit, every agent’s contract converges to pure joint liability, where payment depends on the vector of performances but is the same for every agent, regardless of how her own performance ranks against the rest. In the corresponding adversarial limit, there is convergence to a winner-take-all tournament, *but only when there are just two team members*. With three or more agents, the limit contract pays an agent for her own performance even when that performance is not ranked the best (Corollary 1). This dilution of adversarial contracts aligns with patterns observed in labor markets; see Che and Yoo (2001) and Luft (2016).²

(4) Turning now to the “robust implementation” of high effort as a *unique* equilibrium, new features emerge. Under conditions described in Proposition 3, and in a two-agent setting, our baseline contracts generate submodularity for adversarial agents, and so induce a unique Nash equilibrium. Implementation of the high-effort equilibrium therefore comes with no strategic ambiguity, and with no need to depart from symmetry. This contrasts with a recent literature that emphasizes the benefits of asymmetric contracts for robust implementation.³

The same is not generally true when agents are altruistic; again, see Proposition 3. Joint liability typically generates supermodularity across efforts, raising the specter of multiple

¹These limits are defined by natural upper bounds on the degree of interdependence that prevent the interdependent utility system from becoming “explosive”. For these bounds and their implications in other contexts, see Pearce (2008), Hori and Kanaya (1989), Bergstrom (1999) and Ray & Vohra (2020).

²For more in-depth discussion of the literature that documents these trends see Che and Yoo (2001). For the literature in the field of managerial studies, see Luft (2016).

³See Segal (1999, 2003), Winter (2004), Genicot and Ray (2006), Bernstein and Winter (2012), Halac, Kremer and Winter (2020), Halac, Lipnowski and Rappoport (2021), Halac, Kremer and Winter (2023), and Camboni and Porcellacchia (2023). On multiple equilibria in the joint liability setting, see Besley and Coate (1995).

equilibria in the post-contract game across agents. For unique implementation, the principal needs to dial back on team rewards in some regions of the performance space, but ramp up group rewards elsewhere. The asymmetry of relative cross-impact under altruism vs antagonism is rendered even more complicated under robust implementation.

(5) Finally, in our baseline setting, it is shown that preference interdependence of either kind invariably advantages the principal (Proposition 4). Principal payoffs under the optimal contract, viewed as a function of the common preference weight placed by an agent on others, must climb as that weight moves away from zero *in either direction*.

We conclude by pointing out some limitations and possible extensions of our exercise. An especially stark assumption concerns the equality, both in magnitude and sign, of the interdependence parameter across all agents. If we drop this restriction, the resulting contract will delicately depend on the full matrix of weights. Differences in the extent of altruism or antagonism will now matter, as also the possibility that there may well be friends and enemies within the same team. So the optimal contract will exhibit elements of competition and cooperation in complex ways. Such a construction relies on details that are unlikely to be available to the principal, who might have some overall idea of where her labor force is from (e.g., the same village), but not much more than that. That raises the interesting question of designing another kind of robust mechanism in this setting, where robustness is now defined relative to heterogeneity in preferences.

Second, there is potential endogeneity in preference interdependence. Both the intensity and sign of preference weights may well depend on the past experiences (and contracts) of team members. But the dependence is subtle. At first glance it seems natural that adversarial interdependence would arise in tournament-like environments; see, e.g., Lanzetta and Englis (1989) and Zillman and Cantor (1977). But it is far from clear — to say the least — that tennis players or students graded on the curve are any more adversarial than anyone else. After all, they *do* understand the rules of the game, and why their compatriots seek to outdo them. Indeed, it is in “cooperative” joint liability settings, in which low payments might be blamed on some shirkers within the team, that antagonism could more easily be provoked (see Ghatak and Guinnane 1999 for some empirical observations in the context of microfinance). It is even possible that an overall drift towards adversarial attitudes might come to dominate the present, *regardless* of the contractual environment of the past (Koçkesen, Ok and Sethi 2000). There are potentially interesting dynamics involved here, but these are beyond the scope of the current paper.

2. RELATED LITERATURE

The early theoretical literature on team agency emphasizes informational externalities among agents. Lazear and Rosen (1981), Holmstrom (1982), Green and Stokey (1983), Nalebuff and Stiglitz (1983) and Mookherjee (1984) offer an explanation for why common informational shocks might cause principals to resort to tournaments. The literature on joint liability (Stiglitz 1990, Ghatak 1999, or Ghatak and Guinnane 1999) shows how peer monitoring or agent information about other agents might cause rewards to co-move positively. Both literatures ignore the payoff externalities that are generated via interdependent preferences. Therefore, our problem differs from the ones studied in these papers; our framework allows us to separately analyze the effects of interdependent preferences in isolation from the informational considerations already studied in the literature.

There is also a literature on asymmetry across contracts with positive and negative co-movement of rewards. Meyer and Mookherjee (1987), Che and Yoo (2001) and DeMarzo and Kaniel (2023) all contrast group rewards to tournament-like structures. Indeed, Meyer and Mookherjee (1987) warn against the use of tournaments when the principal cares for equality ex post. Such ex post concerns precipitate a social preference for positively correlated agent payoffs, thereby reducing the use of competitive contracts but eroding agent incentives in the process. They conclude that “welfare-optimal compensation schemes in general depend separately on an equity and an incentive component that tend to correlate agent compensations in different directions.” Che and Yoo (2001) study a dynamic setting; the incentive effect of positive correlation is different there. Cooperative rewards encourage peer monitoring of effort among team members and can be used effectively as punishments in the event of shirking — with positively correlated rewards, it is relatively easy for a team member to shirk and lower the reward of a compatriot. With negative correlation, he must work harder to generate the same end.⁴ DeMarzo and Kaniel (2023) study a setting in which agents assess their wage against a weighted average of wages paid to fellow agents. Their exercise emphasizes the impact of wage transparency and inter-principal competition on equilibrium wage structure.

Relative to these papers, our focus on interdependent preferences is distinct, as are the specific results we obtain. Our model is static, so the forces basic to the Che-Yoo argument are absent here. Indeed, the asymmetry we uncover relies on teams with at least *three*

⁴As a countervailing force, Che and Yoo (2001) also incorporate an informational externality in the form of a common shock to outputs, which tends to favor competitive contracts. Their study then highlights the optimality of joint liability mechanisms when the discount factor is large relative to the size of the informational externality, and the optimality of a tournament otherwise.

agents, whereas Che and Yoo (2001) consider a two-agent model throughout. And unlike Meyer and Mookherjee (1987) or DeMarzo and Kaniel (2023), our principal or agents have no particular preference for *ex post* equality, which is of course central to their exercise.

Other contributions to contracting with interdependent preferences ask different questions in distinct models. Letina, Liu and Netzer (2020) study multiple agents with a single intermediary — a “reviewer” — positioned between principal and agents. This reviewer observes agent performance and has social preferences towards them. The study explores the effect of reviewer bias; e.g., a leniency bias in which the reviewer hesitates to report instances of agent shirking to avoid subjecting them to contractual punishment. Abernethy et al. (2023) study social norms that embody altruism; these affect the cost of agent effort in a team. The resulting contract depends on both agent output and overall output, and the authors study comparative statics of the optimal contract with respect to changing weights on the cost of effort. Bartling (2011) studies a two-agent problem in a CARA-normal setting with binary effort, in which agents dislike wage inequality. They observe (as we do) that altruistic preferences can tone down the extent of competition.

Itoh (2004) considers a two-state model with a principal and two inequality-averse agents, each of whom compare their wages to those paid to the other. Inequality-aversion captures both altruism (no one wants to forge far ahead) and antagonism (no one wants to be left too far behind). Depending on the relative magnitudes of these two components, contracts can be competitive or cooperative. Dhillon and Herzog-Stein (2009) is a version of this exercise in which only the wage *rank* matters to agents (and not the absolute wage value). Their results indicate that with incomplete information and convex rank-dependent preferences, contracts are interdependent and might be asymmetric. Sarkisian (2017) studies a two-agent binary effort setting with common binary output, resulting in a wage contract contingent on two states. Agents are either altruistic in our sense or *homo moralis*, with utility directly influenced by the other’s actions. They compare preference interdependence to independence in both these specifications, with contractible effort. Vásquez and Weretka (2021) study firms who hire workers with interdependent preferences, but they presume that all workers are paid an unconditional, uniform wage. Performance-based pay plays no role in their exercise.

Our paper is distinct from these contributions. Our specification is directly based on agent payoffs and interdependent effects from those payoffs, rather than on inequality aversion per se or other payoff spillovers arising from agent choices. We demonstrate that an agent’s wage in the optimal contract moves positively with the performance of others when agents are altruistic, and negatively when they are spiteful. Moreover, the

strength of this dependence grows with the degree of preference interdependence. Our baseline characterizations do not rely on binary signals, two-agent settings, or the linear-normal framework (though our analysis of robust implementation is conducted with two agents). Additionally, we identify quantitative asymmetries between altruistic and adversarial preferences — an insight that none of the other contributions capture.

3. MODEL

3.1. Signals and Performance. Each of n agents makes a binary effort choice $e \in \{\ell, h\}$, at cost 0 for $e = \ell$ and cost $c > 0$ for $e = h$. Given $e \in \{\ell, h\}$, an agent produces an observed signal y from some set Y with probability measure μ^e , identical for all agents and independent across them. (Interdependent signals are easy enough to incorporate, but our focus is on interdependent *preferences* instead.) We assume that μ^ℓ is absolutely continuous with respect to μ^h , so that the Radon–Nikodym derivative $[d\mu^\ell/d\mu^h](y)$ is always well-defined.⁵ We need not place any other substantive restriction on Y , though it might help to think of it as a space of single-dimensional outputs that each agent might feasibly produce. It will be convenient to define

$$(1) \quad \lambda(y) \equiv 1 - \frac{d\mu^\ell}{d\mu^h}(y).$$

This function is a likelihood ratio: the larger is λ , the more likely it is that effort was high. We will refer to $\lambda(y)$ as the agent’s *performance* when she generates signal y . Given any effort level, λ is a bona fide random variable, and its realizations are perfectly deducible from y . Denote by γ the derived probability measure for λ when the agent exerts high effort. Note from (1) that λ can vary from a possible low of $-\infty$ to a possible maximum of one, and also that $\int \lambda d\gamma = 0$. Use Λ to denote its support. Below, we make the obvious argument that $\Lambda \equiv \Lambda^n$ is a sufficient domain for determining agent rewards.

3.2. Preferences. We presume that agents have symmetric interdependent preferences:

$$(2) \quad U_i = \underbrace{[u(m) - c\mathbb{1}_i^h]}_{\text{material payoff}} + \underbrace{\alpha \sum_{j \neq i} U_j}_{\text{interdependence}}, \text{ for } i = 1, \dots, n,$$

where $u(m)$ is a vNM felicity on money, $\mathbb{1}_i^h$ is an indicator for high effort, α is a measure of altruism or antipathy across agents, and $\{U_j\}$ are the utilities of the other agents. The felicity function is increasing, unbounded, and twice continuously differentiable, with

⁵As is well known, this derivative reduces to the ratio of probabilities over $y \in Y$ when Y is discrete, and to the ratio of densities for distributions on Y , when those densities exist.

$u(0) = 0, u'(m) > 0, u''(m) < 0, u'(0) = \infty$ and $u'(\infty) = 0$. Observe that such preferences are fully *non-paternalistic*, in the sense in which a Bergson-Samuelson welfare function is non-paternalistic: an agent (and in Bergson-Samuelson, the social planner) derives utility from other agents' *utilities*, and not, say, their "material payoffs" $u(m) - c\mathbb{1}_i^h$.

We can — and will — reduce such preferences to ones that are directly defined on material outcomes. But simply *starting* there lacks a clear foundation. Why should an agent care about another's material payoff if, for other agents, material payoffs are not the only source of utility? This adaptation of the Bergson-Samuelson view to interdependent preferences is in the spirit of Pearce's 1984 dissertation published as Pearce (2008); see also Ray (1987), Hori and Kanaya (1989), and Bergstrom (1999). Our motivation rests on a somewhat philosophical observation: an agent who "genuinely cares" about others must care for their total well-being, which includes both material and social components. Because we can reduce such preferences to utilities defined over material payoffs, much of the analysis does not rely on whether we insist on such a philosophy. But some of it does; specifically, our predicted contractual asymmetry between positive and negative interdependencies in large teams. We will return to a discussion of these matters.

3.3. Reduced-Form Payoff Representation. We follow Pearce (2008), Hori and Kanaya (1989), Bergstrom (1999) and Ray & Vohra (2020) to obtain a "coherent" utility representation of this interactive system. Briefly, coherence asks for a unique utility representation in action space, with no unbounded utility spirals driven by interactive utility alone, *for the same vector of material payoffs*; see Ray & Vohra (2020) for a formal statement. The case studied in this paper is particularly simple. For every i , $U_i = [u(m_i) - c\mathbb{1}_i^h] + \alpha \sum_{j \neq i} U_j$, so that both uniqueness of the utility representation and boundedness of utility are guaranteed by the necessary and sufficient condition $|\alpha(n-1)| < 1$, or loosely, preference-dependence is mild enough.⁶ We impose this restriction in all that follows, and note that

$$(3) \quad (1 + \alpha)U_i = [u(m_i) - c\mathbb{1}_i^h] + \alpha A,$$

where $A \equiv \sum_j U_j$. Adding (3) over all i and dividing by $1 - \alpha(n-1)$, we obtain

$$(4) \quad A = \frac{\sum_j [u(m_j) - c\mathbb{1}_j^h]}{1 - \alpha(n-1)},$$

⁶Ray & Vohra (2020) assume that for a system $U_i = f_i(m_i, \{U_j\}_{j \neq i})$, there is $B(m) < \infty$ such that $|U_i| < \max_{j \neq i} |U_j|$ whenever $\max_{j \neq i} |U_j| > B(m)$. This is equivalent to $|\alpha(n-1)| < 1$ in our simpler setting.

and using (4) in (3) and transposing terms, we see that for every i ,

$$(5) \quad V_i \equiv (1 + \alpha) \frac{1 - \alpha(n-1)}{1 - \alpha(n-2)} U_i = [u(m_i) - c1_i^h] + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq i} [u(m_j) - c1_j^h],$$

where U_i has been payoff-renormalized by $(1 + \alpha)[1 - \alpha(n-1)]/[1 - \alpha(n-2)]$ to obtain V_i . While own material payoffs affect this utility one for one, the combined weight of the material payoff of other individuals is always less than one, given that $|\alpha(n-1)| < 1$.

3.4. Contracts and Equilibrium. A principal hires the agents and offers each agent i a limited liability contract $\hat{m}_i(\mathbf{y})$ with $\hat{m}_i(\mathbf{y}) \geq 0$ for each $\mathbf{y}$. Then agents interact, each choosing effort and seeking to maximize the payoff function in (5). Throughout, we assume that the principal wants to implement high effort for all agents. She chooses a collection of nonnegative functions $\{\hat{m}_i\}$ so as to

$$(6) \quad \text{minimize } \sum_i \mathbb{E}(\hat{m}_i | \mathbf{e} = \mathbf{h}), \text{ subject to the constraint: } \mathbf{e} = \mathbf{h} \text{ is incentive-compatible.}$$

We develop “incentive compatibility” in the next Section. We make three remarks here. First, to facilitate comparison with some of the literature discussed in Section 2, Appendix A contains an example with binary outputs, which may be of independent interest. Second, we haven’t included a participation constraint for the agents. The limited liability constraint already constrains payments on the downside, so that our problem is non-trivial as it stands. Moreover, the extension to participation constraints is straightforward (see Appendix B for details). Third, “implementing high effort” as described in (6) means that *some* equilibrium involves high effort for all. In contrast, “robust implementation” would require *every* second stage equilibrium to exhibit high effort. This is an extension that does affect our results to some degree, and we take it up in Section 6.

4. OPTIMAL CONTRACTS

4.1. The Incentive Constraint. Our notion of incentive compatibility presumes that all efforts and payments are revealed *ex-post* to every agent. So when an agent chooses effort, she understands that all agents will see her effort and payment, and experience their payoffs accordingly. Other variants are possible (see Section 8). But we assume throughout that effort is not verifiable by the *principal*, so no payment can be conditioned on it.

Recall that Y is the common support of outputs or signals. Define $\mathbf{Y} \equiv Y^n$, where n is the number of agents. We set $d\boldsymbol{\mu}^e \equiv \prod_j d\mu_j^{e_j}$ to indicate integration with respect to the joint probability of $\mathbf{y}$ when agents supply effort vector $\mathbf{e}$, and set $d\boldsymbol{\mu}_{-i}^e \equiv \prod_{j \neq i} d\mu_j^e$.

Define the random vector of performances by $\lambda(\mathbf{y}) \equiv \{\lambda(y_i)\}_{i=1}^n$, and let $\Lambda \equiv \Lambda^n$ be the product support of λ . The following observation formalizes the idea that without loss of generality, we can focus on contracts that are measurable with respect to performance vectors as defined by λ .

Observation 1. *If optimal contracts exist, they must be defined on performances alone. That is, for every i , there is $m_i : \Lambda \rightarrow \mathbb{R}_+$ such that the i th component of the optimal contract, $\hat{m}_i$, satisfies*

$$(7) \quad \hat{m}_i(\mathbf{y}) = m_i(\lambda(\mathbf{y})).$$

See Appendix C for a proof. Intuitively, all incentive constraints are measurable with respect to λ , provided $\hat{m}_i(\mathbf{y})$ is. However, conditioning rewards on different signals with the same performance λ only increases the variance of the principal's payout to risk-averse agents, and so is suboptimal. So, keeping (7) in mind, we work below with $\{m_i\}$.

We now argue that every agent receives the same contract in performance space. For $\mathbf{z} \in \mathbb{R}^n$, let $\mathbf{z}^{[i]}$ be any vector with $z_1^{[i]} = z_i$ as its first entry and some permutation of $\mathbf{z}_{-i}$ for the rest. By Observation 1, the i th contract can be viewed as a function of $\lambda(\mathbf{y})^{[i]}$. We now show that there is a unique optimal contract (up to any permutation of peers) with the same functional form for every player; for a proof see Appendix C.

Observation 2. *There is a unique function m on Λ , invariant to permutations of its last $n - 1$ entries, such that $m_i(\lambda(\mathbf{y})) \stackrel{\mu^h\text{-a.s.}}{=} m(\lambda(\mathbf{y})^{[i]})$ for every i under an optimal contract defined on performances λ . In summary, each reward function under the optimal contract is symmetric in the performances of peers, and the same contractual form is offered to all agents.*

Given Observation 2, we proceed by supposing that every agent gets the same contract m , the form of which is symmetric across her peers. Suppose that agent i 's compatriots all choose $e_j = h$. Then by (5), if agent i chooses $e_i = h$, her (renormalized) payoff is

$$(8) \quad V_i^h \equiv \left[\int_{\mathbf{Y}} u(m(\lambda(\mathbf{y})^{[i]})) d\mu^h - c \right] + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq i} \left[\int_{\mathbf{Y}} u(m(\lambda(\mathbf{y})^{[j]})) d\mu^h - c \right].$$

In similar fashion, if our agent chooses $e = \ell$, her expected payoff V_i^ℓ is given by

$$(9) \quad V_i^\ell \equiv \left[\int_{\mathbf{Y}} u(m(\lambda(\mathbf{y})^{[i]})) d\mu_i^\ell d\mu_{-i}^h - c \right] + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq i} \left[\int_{\mathbf{Y}} u(m(\lambda(\mathbf{y})^{[j]})) d\mu_i^\ell d\mu_{-i}^h - c \right].$$

So the incentive constraint is given by $V_i^h \geq V_i^\ell$, or, using (8) and (9),

$$\int_{\mathbf{Y}} u(m(\lambda(\mathbf{y})^{[i]})) \lambda(y_i) d\mu^h + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq i} \left[\int_{\mathbf{Y}} u(m(\lambda(\mathbf{y})^{[j]})) \lambda(y_i) d\mu^h \right] \geq c,$$

where recall that $\lambda(y) = 1 - [d\mu^\ell/d\mu^h](y)$. This permits us to suppress $\mathbf{y}$ altogether and write the incentive constraint entirely in terms of likelihoods instead of outputs:

$$\int_{\Lambda} u(m(\boldsymbol{\lambda}^{[i]}))\lambda_i d\boldsymbol{\gamma} + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq i} \left[\int_{\Lambda} u(m(\boldsymbol{\lambda}^{[j]}))\lambda_i d\boldsymbol{\gamma} \right] \geq c,$$

where recall that $\boldsymbol{\gamma}$ is the probability measure over the performance λ when $e = h$. We now re-index the variables in the inequality above. First set $i = 1$ and note that $\boldsymbol{\lambda}^{[1]} = \boldsymbol{\lambda}$. Next, for each $j \neq 1$, we “rotate” the entries in $\boldsymbol{\lambda}^{[j]}$ so that $[j]$ is replaced by $[1]$, with all other indices permuted accordingly, including the index $i = 1$ under the second integral, which will run the entire gamut of values $\{2, \dots, n\}$ as different values of j are thus replaced. Because the optimal contract is symmetric, that gives us an equivalent representation of the incentive constraint as

$$(10) \quad \int_{\Lambda} u(m(\boldsymbol{\lambda}))\Psi(\boldsymbol{\lambda})d\boldsymbol{\gamma} \geq c,$$

where we’ve defined

$$(11) \quad \Psi(\boldsymbol{\lambda}) \equiv \lambda_1 + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq 1} \lambda_j.$$

The function Ψ will play a central role in the design of optimal contracts. It serves as an “incentive index” that measures the extent to which the provision of an additional util to an agent under the performance realization $\boldsymbol{\lambda}$ encourages her effort (or discourages it if negative), appropriately accounting for the utility spillovers from other agents.

4.2. The Principal’s Objective. The principal’s expected wage bill under m is given by

$$(12) \quad \int_{\Lambda} \sum_{j=1}^n m(\boldsymbol{\lambda}^{[j]})d\boldsymbol{\gamma} = n \int_{\Lambda} m(\boldsymbol{\lambda})d\boldsymbol{\gamma}$$

and she will seek to minimize this payout, subject to meeting the incentive constraint (10). The use of the incentive index Ψ in that constraint enables us to effectively reduce a team agency setting with interdependent preferences to a single-agent problem.

4.3. Independent Preferences. As a warm-up, shut down utility interdependence by setting $\alpha = 0$. Then $\Psi(\boldsymbol{\lambda}) = \lambda_1$, and so (10) reduces to

$$\int_{\Lambda} u(m(\boldsymbol{\lambda}))\lambda_1 d\boldsymbol{\gamma} = \int_{\Lambda} \left[\int_{\Lambda^{n-1}} u(m(\lambda_1, \boldsymbol{\lambda}_{-1}))d\boldsymbol{\gamma}_{-1} \right] \lambda_1 d\boldsymbol{\gamma}_1 \geq c,$$

which shows that conditioning m on λ_{-1} is absurd when u is strictly concave. Calling the solution $m(\lambda_1)$ by a slight abuse of notation, the incentive constraint reduces to

$$(13) \quad \int_{\Lambda} u(m(\lambda_1)) \lambda_1 d\gamma_1 \geq c$$

for agent 1, and by symmetry, (13) applies to any agent i , once we replace 1 by i . Because the principal seeks to minimize $\int_{\Lambda} m(\lambda_1) d\gamma_1$, it must be the case that the “marginal product” of money across any set of states with positive payouts must be constant, or, using the end-point conditions $u'(0) = \infty$ and $u'(\infty) = 0$,

$$(14) \quad \text{If } \lambda_1 \leq 0, \text{ then } m(\lambda_1) = 0; \text{ if } \lambda_1 > 0, \text{ then } m(\lambda_1) > 0 \text{ and } u'(m(\lambda_1))\lambda_1 \text{ is constant in } \lambda_1.$$

Condition (14) implies that $m(\lambda_i)$ is flat at 0 up to $\lambda^* \equiv 0$, and thereafter increases in λ_i .

4.4. Interdependent Preferences. We maximize a suitable Lagrangian subject to the limited liability constraint. For a reward function m , define the Lagrangian functional

$$(15) \quad \mathcal{L}(m) \equiv - \int_{\Lambda} m(\lambda) d\gamma + \nu \left[\int_{\Lambda} u(m(\lambda)) \Psi(\lambda) d\gamma - c \right],$$

where ν is a multiplier for the incentive compatibility constraint, and $\Psi(\lambda)$ is the incentive index defined in (11). The first order conditions are given by

$$-1 + \nu u'(m(\lambda)) \Psi(\lambda) \leq 0, \text{ with equality if } m(\lambda) > 0.$$

But $m(\lambda)$ must be positive for some λ otherwise no incentives can be provided. In particular, it follows that $\nu > 0$, and so the first order condition above reduces to

$$(16) \quad \text{If } \Psi(\lambda) \leq 0, \text{ then } m(\lambda) = 0; \text{ if } \Psi(\lambda) > 0, \text{ then } m(\lambda) > 0 \text{ and } u'(m(\lambda))\Psi(\lambda) \text{ is constant in } \lambda,$$

where we’ve used the end-point conditions $u'(0) = \infty$ and $u'(\infty) = 0$. Equation (16) connects monetary payments directly to the index $\Psi(\lambda)$. Payments $m(\lambda)$ are positive if and only if $\Psi(\lambda) > 0$. Within this zone, assuming λ is continuously distributed, we see from differentiating (16) that the isoquants $m(\lambda)$ and $\Psi(\lambda)$ are aligned via the condition:

$$(17) \quad \frac{\nabla m}{m} = \theta^{-1} \frac{\nabla \Psi}{\Psi},$$

where $\theta \equiv -u''(m)m/u'(m)$ is the coefficient of relative risk aversion, which in general will vary with m , so the formula (17) is not entirely closed-form but still instructive.

Now unpack Ψ as described in (11). If λ_1 alone increases, so does $\Psi(\lambda)$. So an agent’s reward increases in her own performance, provided that her reward was strictly positive to begin with. As for the cross-effects, they depend on whether the agents are altruistic ($\alpha > 0$) or adversarial ($\alpha < 0$). When $\alpha > 0$, an increase in another agent j ’s performance

will raise λ , and by extension — using (11) — it will raise Ψ and therefore the reward to agent 1. This is in the spirit of joint liability. Exactly the opposite happens when $\alpha < 0$, which is in the spirit of a tournament. These observations together formally establish:

Proposition 1. *The reward to an agent is positive whenever $\Psi(\lambda) > 0$, and zero when $\Psi(\lambda) \leq 0$. Moreover, when $\Psi(\lambda) > 0$, an agent's reward moves with performance as summarized in (17). In particular, it strictly increases in own performance, co-moves positively with partner performance when $\alpha > 0$ and co-moves negatively when $\alpha < 0$. The former arrangement is cooperative; the latter is competitive.*

*In either case, as long as the payment to an agent is positive, the marginal reward of an increase in own performance always exceeds the marginal reward (under cooperation) or marginal loss (under competition) from the same increase in compatriot performance.*⁷

Proposition 1 is more subtle than it appears at first glance, because cross-payments serve different functions. Assume that (a) there are just two agents and (b) we are in the altruistic case. One reason for a cross-payment at configuration $\lambda = (\lambda_1, \lambda_2)$ arises when 2 is being paid so much more than 1 that $u'(m_2) < \alpha u'(m_1)$. Then giving 1 some reward from 2's better performance makes 2 happier than if she were to get it all herself. It also incentivizes her better. This is an effect that arises from the curvature of 2's utility function.

But there are dangers. Such an arrangement might incentivize 1 to free-ride, anticipating cross-payments from 2's success. This is all nicely taken care of in our optimal contract. In fact, if player 1's Ψ -index at the configuration $\lambda = (\lambda_1, \lambda_2)$, given by $\lambda_1 + \alpha\lambda_2$, is negative, 1 won't get paid for 2's extra success; see the solution as summarized in (16).

Another scenario is one in which 2 is doing badly and so gets *no* reward from better marginal performance: the baseline performance region at $\lambda = (\lambda_1, \lambda_2)$ is one that the principal is not happy with. Notice that this could happen even though “it is not 2's fault”: $\Psi_2 = \lambda_2 + \alpha\lambda_1$ might be negative because 1's performance is poor. But then it's easy to see that $\Psi_1 = \lambda_1 + \alpha\lambda_2$ must be negative as well, so no payments, direct or cross, will be made for marginal performance.⁸ On the other hand, if $\Psi_2 < 0$ but $\Psi_1 > 0$ (so that it is “genuinely 2's fault”), 1 — but not 2 — will be rewarded for a marginal increase in 2's performance, because the principal wants to deploy resources to incentivize 1 to stay where she currently is. Thus cross-payments serve more than one function. Proposition 1 combines them in a strikingly simple fashion.

⁷Using the coherence condition $|\alpha|(n-1) < 1$, it is easy to see that the coefficient $\alpha/[1 - \alpha(n-2)]$ in the expression $\Psi(\lambda)$ is smaller than one in absolute value, which verifies this claim.

⁸This situation is reminiscent of joint liability, in which 2 is held liable for the failings of 1.

4.5. A Comment on Risk Aversion. Risk attitudes, or the curvature of felicity u , affect the extent to which rewards are concentrated on signals associated with the highest values of the incentive index Ψ . This can be seen to some extent from (17), which translates $\nabla\Psi$ into ∇m . A lower value of relative risk aversion means that money payments grow faster than they did before, with the multiplier ν adjusting so that initial levels of money payment are lower. As we converge to the complete absence of risk aversion, the resulting concentration of payments becomes unrealistically high and focused only on the highest possible value of Ψ . Therefore some degree of risk aversion is needed for reasonable contractual forms. The following intuitive example is useful in this regard.

Assuming constant relative risk aversion felicity $u(m) = m^{1-\theta}$ with $\theta \in (0,1)$,⁹ optimal payment equals $\max\{\Psi/k, 0\}^{1/\theta}$, where the constant k is such that the incentive constraint (10) binds. The payment moves almost linearly with Ψ in the payment zone as θ approaches 1. Payments turn increasingly convex with Ψ as risk aversion declines. For near-risk-neutral agents ($\theta \simeq 0$), rewards are sharply increasing only in the vicinity of outcomes that maximize the incentive index. In particular, even a mildly altruistic but near-risk-neutral agent receives meaningful rewards only when *every* agent achieves peak performance, and an adversarial near-risk-neutral agent is rewarded significantly only when she attains the highest output values and compatriot performance is very poor.¹⁰

5. COOPERATION AND COMPETITION: AN ASYMMETRY

When λ varies continuously, the cross-impact for agent 1 is given by the value of $\partial m / \partial \lambda_j$ for any $j \neq 1$, and the own-impact is given by $\partial m / \partial \lambda_1$. By symmetry and the linearity of Ψ in λ , the former effect is the same across all peers $j \neq 1$ under an optimal contract, and by symmetry again, the fact that we're discussing agent 1 in particular is irrelevant. It follows that the *relative cross-impact* within a contract can be defined as

$$\xi(\alpha, n) \equiv \frac{\partial m / \partial \lambda_j}{\partial m / \partial \lambda_1},$$

where, because the use of agents 1 and j is arbitrary, there are no indices or subscripts for ξ , though we keep the dependence on α and group size n explicit. As we shall see below,

⁹Because felicity has a finite lower bound, we rule out additional curvature — logarithmic or higher.

¹⁰In such situations, optimal contracts display an extreme form of joint liability. *Every* team member must *fully* excel for anyone to receive any compensation at all. When $\alpha < 0$ and Λ is compact, the contract becomes an extreme tournament where compensation requires that an agent outperform *every* other agent by *the largest possible* margin, $\max \Lambda - \min \Lambda$. This could lead to non-existence of an optimal contract for near-risk-neutral agents — certainly for risk-neutral agents — when performances have atomless density, but these are technical matters that we ignore. (When Λ is finite, optimal contracts can be shown to exist.)

ξ is unaffected by risk aversion or the baseline realization of the performance vector λ or, as shown in Appendix B, by the agent's outside options when the model is extended to include participation constraints. Anticipating this, we've removed these variables as explicit arguments of ξ . The next proposition follows immediately from condition (17):

Proposition 2. *Assume λ is continuously distributed. Fix α and any λ for which $\Psi(\lambda) > 0$; i.e., positive payments are made. Then the relative cross-impact under the optimal contract is given by*

$$(18) \quad \xi(\alpha, n) = \frac{\alpha}{1 - \alpha(n-2)},$$

which is independent of u or λ , and only depends on the size and sign of α , and group size n .

With just two agents, (18) reduces to $\xi(\alpha, 2) = \alpha$ for any λ , so that

$$|\xi(\alpha, 2)| = |\xi(-\alpha, 2)|$$

for every λ with positive rewards to both agents. That is, the relative cross-impact of the contract is symmetric across α and $-\alpha$; only its sign changes. Coherence tells us that $|\alpha| < 1$. As α converges to its upper altruistic limit $+1$, the contract converges to pure joint liability, while at the opposite adversarial extreme -1 , it becomes a pure tournament.

But this symmetry across α and $-\alpha$ is broken when there are three or more agents. Suppose that $n = 3$. Then we see from (18) that

$$(19) \quad |\xi(\alpha, 3)| = \left(\frac{1 + \alpha}{1 - \alpha} \right) |\xi(-\alpha, 3)|$$

for every λ and every α satisfying the coherence bound $|\alpha| \in (0, 1/2)$. For instance, when $\alpha = 1/3$, (18) tells us that the relative cross-impact is 50%. That same number drops to 25% for the adversarial case $\alpha = -1/3$, so the ratio of these two relative cross-impacts is 2. Altruism delivers twice the relative cross-impact compared to equal-sized antipathy.

This asymmetry is magnified in still larger teams. With 20 agents and $|\alpha| \approx 0.04$, the ratio of relative cross-impacts across positive and negative interdependence jumps to over 6. (The relative cross-impact is around 14% for altruistic preferences but only a bit over 2% for adversarial preferences.) In summary, *the principal exploits altruistic motives to a greater degree than she does adversarial motives*, the discrepancy expanding with larger teams.

This phenomenon is general, and extends to all environments with three or more agents and arbitrary distributions of output conditional on effort.

Corollary 1 (Proposition 2). Fix $\alpha \in (0, 1/(n-1))$. Then for all λ such that payments are strictly positive,

$$(20) \quad \frac{|\xi(\alpha, n)|}{|\xi(-\alpha, n)|} = 1 + \frac{2\alpha(n-2)}{1-\alpha(n-2)}.$$

This ratio of relative cross-impacts increases both in n and in α (as long as $n \geq 3$). For fixed n , as $\alpha \uparrow 1/(n-1)$, the ratio converges to $2n-3$.¹¹

Corollary 1, which highlights the asymmetric implications of $+\alpha$ and $-\alpha$ for the same magnitude of α , is the only part of the paper where our non-paternalistic formulation plays a crucial role. Our comparison is for some given *absolute* value of α , held fixed across altruism and antagonism. But this benchmarking at the non-paternalistic level nevertheless translates into an asymmetry as we pass to the reduced form. That reduced form defined on material preferences is given by equation (5) in Section 3.3, reproduced here for convenience:

$$V_i = [u(m_i) - c\mathbb{1}_i^h] + \frac{\alpha}{1-\alpha(n-2)} \sum_{j \neq i} [u(m_j) - c\mathbb{1}_j^h].$$

The induced coefficient on material preferences is $\frac{\alpha}{1-\alpha(n-2)}$. Comparing its absolute value for $\alpha > 0$ and $-\alpha < 0$, we see that for all $n > 2$,

$$\underbrace{\frac{\alpha}{1-\alpha(n-2)}}_{\text{absolute interdependence under altruism}} \neq \underbrace{\frac{\alpha}{1+\alpha(n-2)}}_{\text{absolute interdependence under antagonism}},$$

which shows that the symmetry in non-paternalistic interdependence does not translate into symmetry in the reduced form. This is *not* a mere technicality. It captures the difference between an altruistic and an antagonistic environment. Mutual altruism creates an additional effective multiplier on material payoffs, heightening the positive impact of those payoffs. With mutual antagonism, the effect is just the opposite: the induced negative coefficient on material payoffs, while still negative, is dampened.

One reaction to this kind of argument is that preference interdependence should be defined directly on material payoffs, with the symmetry imposed on that direct definition. As already discussed, we find it difficult to see why that should be the case. If A cares for B 's state of happiness (whether positively or negatively) and understands that B 's state

¹¹Because the coherence bound $|\alpha| < 1/(n-1)$ tightens with n , holding α fixed as the team grows moves the environment closer to that bound. The asymmetry is not an artifact of this proximity, however. Holding *total* interdependence $\kappa \equiv \alpha(n-1)$ fixed instead, the ratio in (20) becomes $[1 + \kappa \frac{n-2}{n-1}] / [1 - \kappa \frac{n-2}{n-1}]$, which still exceeds one, increases in n , and converges to $(1 + \kappa)/(1 - \kappa)$ as $n \rightarrow \infty$. The normalization of aggregate preference interdependence bounds the asymmetry, but in no way eliminates it.

depends in turn on the well-being of others, we find it unreasonable to presume that A would ignore these considerations when reacting to B 's payoffs. The entire environment of sympathy or antipathy matters.

The asymmetry in (20) morphs into a qualitative difference at extreme levels of preference interdependence. At the altruistic limit as $\alpha \rightarrow \frac{1}{n-1}$ for any value of n , every agent's contract converges to a *common* contract that depends on the entire vector of performances λ , and rewards every agent in the same way. The limit payment to every agent is zero for any λ such that $\sum_j \lambda_j \leq 0$, and if $\sum_j \lambda_j > 0$, it is given by a common payment

$$(21) \quad m(\lambda) = (u')^{-1} \left(k / \sum_j \lambda_j \right)$$

made to every agent, where the constant k ensures that the incentive constraint (10) binds along any sequence of solutions going to this limit. At the limit solution, agents not only get a symmetric contract modulo permutations of λ , they literally get the *same contract*, entirely conditioned on team performance. This mechanism is optimal, even though each agent is far from being a symmetric utilitarian. They don't need to be: third-party feedback effects sufficiently magnify their altruistic tendencies.

In contrast, under extreme adversarial preferences ($\alpha \approx -\frac{1}{n-1}$), third-party interactions attenuate preference interdependence. For instance, in a three-person team, the limit gradient of payment for agent 1 is aligned with the vector $(1, -\frac{1}{3}, -\frac{1}{3})$, which is not the mirror image of its altruistic counterpart. An agent's payment is now contingent on surpassing a benchmark set at two-thirds of average compatriot performance. So she could receive compensation even if her performance falls below that of the other two agents.

More generally, for any n , as $\alpha \downarrow -\frac{1}{n-1}$, we see from (11) that

$$\Psi(\lambda) \rightarrow \lambda_1 - \frac{1}{2n-3} \sum_{j \neq 1} \lambda_j,$$

for every vector of realized performances λ . In the adversarial limit agent 1 receives a positive reward if and only if

$$(22) \quad \lambda_1 > \frac{1}{2n-3} \sum_{j \neq 1} \lambda_j,$$

and in that case her reward is given by

$$(23) \quad m(\lambda) = (u')^{-1} \left(k' / \left[\lambda_1 - \frac{1}{2n-3} \sum_{j \neq 1} \lambda_j \right] \right),$$

for some k' that makes the incentive constraints (10) bind along the sequence of solutions. While this contract does possess competitive features, it is not a “pure” tournament except in the two-agent case, where we’ve already seen that an agent is never paid if her performance is below that of her compatriot. When $n \geq 3$, an agent can be rewarded for her own performance even when it is not the best. Using (23), we see that agent 1 gets a positive payment if and only if her performance, measured in λ -units, exceeds $(n-1)/(2n-3)$ of the average performance of her compatriots. That is, as n grows large, she is paid if her performance surpasses just half the average performance of the others, driving home the asymmetry between altruistic and adversarial limits.

The literature recognizes that competitive structures are not widely employed, whereas team-based bonuses are more commonly favored. For instance, Lazear (1989) notes that despite the theoretical support for rank-order tournaments, their use is infrequent. Taken together, our discussion of local contract competitiveness sheds some light on this phenomenon. At the same time, we will see that there is a different force in our setting that pushes towards more competitive contracts, especially in smaller teams. However, in the next Section on robust implementation, we qualify some of these claims.

6. ROBUST IMPLEMENTATION AND COMPETITION

An older literature in implementation theory (Mookherjee 1984, Ma 1988, Ma, Moore and Turnbull 1988, Mookherjee and Reichelstein 1992, Bergemann and Morris 2009) observes that under an optimal contract, there could be multiple equilibria among agents. See Segal (1999, 2003), Winter (2004), Genicot and Ray (2006), Halac, Kremer and Winter (2020, 2023), or Halac, Lipnowski and Rappoport (2021) for more recent treatments. A central point in this literature is that coordination on an equilibrium that favors the principal is by no means guaranteed — especially if there is another equilibrium that favors the agents.

Specifically, our incentive constraint (10) requires that an agent weakly prefers working to shirking, *assuming that all the other agents are working*. However, that may not eliminate effort profiles that involve shirking by some agents. Moreover, these might constitute potentially undesirable equilibria from the principal’s perspective. To ensure robust implementation, we augment the principal’s problem by adding constraints that ensure effort from an agent even under the presumption that one or more of her compatriots might be shirking. As is standard in the literature on robust design, we consider only those equilibria satisfying an “indifference refinement” condition, under which workers choose to exert effort when they are indifferent between working and shirking.

Certainly, once an agent's incentives are thus dealt with, other agents need only be incentivized under a smaller set of constraints. This well-known argument could give rise quite naturally to contracts that are asymmetric across identical agents, as already observed in the literature cited above.¹²

6.1. Robustness Constraint. We proceed in parallel to our derivation of (10) but work with just two agents, setting $n = 2$. Assuming agent 2 chooses effort e , if agent 1 chooses $e_1 = h$, her (renormalized) payoff is given by

$$\tilde{V}_1^h = \int_{\mathbf{Y}} u(\hat{m}_1(\mathbf{y})) d\mu_1^h d\mu_2^e - c + \alpha \left[\int_{\mathbf{Y}} u(\hat{m}_2(\mathbf{y})) d\mu_1^h d\mu_2^e - \mathbb{1}_e c \right],$$

where $\mathbb{1}_e$ is the indicator for $e = h$. Likewise, if agent 1 chooses $e_1 = \ell$, her payoff is:

$$\tilde{V}_1^\ell = \int_{\mathbf{Y}} u(\hat{m}_1(\mathbf{y})) d\mu_1^\ell d\mu_2^e + \alpha \left[\int_{\mathbf{Y}} u(\hat{m}_2(\mathbf{y})) d\mu_1^\ell d\mu_2^e - \mathbb{1}_e c \right].$$

Given that e is supplied by 2, agent 1 is incentivized to expend high effort if $\tilde{V}_1^h \geq \tilde{V}_1^\ell$. Just as in Observation 1, there is no loss in writing the contract $\hat{m}_1(\mathbf{y})$ as a function $m_1(\boldsymbol{\lambda})$ of $\boldsymbol{\lambda}$ alone. Therefore, switching to performance space in the two expressions above, and subtracting each side of the latter from the former, we have, using $\tilde{V}_1^h \geq \tilde{V}_1^\ell$:

$$(24) \quad \int_{\Lambda} u(m_1(\boldsymbol{\lambda})) \lambda_1 d\boldsymbol{\gamma} + \alpha \int_{\Lambda} u(m_2(\boldsymbol{\lambda})) \lambda_1 d\boldsymbol{\gamma} \geq c.$$

when agent 2's effort $e = h$, giving us the usual incentive constraint, and

$$(25) \quad \int_{\Lambda} u(m_1(\boldsymbol{\lambda})) \lambda_1 (1 - \lambda_2) d\boldsymbol{\gamma} + \alpha \int_{\Lambda} u(m_2(\boldsymbol{\lambda})) \lambda_1 (1 - \lambda_2) d\boldsymbol{\gamma} \geq c.$$

when 2's effort $e = \ell$, giving us the new robustness constraint.

Following Winter (2004) and other authors, note that the robustness constraint (25) need only be imposed on just *one* of the two players, thereby ensuring that high effort is dominant for her. So we can suppose that (24) and (25) must both hold for, say, agent 1. The contract for agent 2 will then need to satisfy just the incentive constraint:

$$(26) \quad \int_{\Lambda} u(m_2(\boldsymbol{\lambda})) \lambda_2 d\boldsymbol{\gamma} + \alpha \int_{\Lambda} u(m_1(\boldsymbol{\lambda})) \lambda_2 d\boldsymbol{\gamma} \geq c,$$

the low-effort choice by agent 1 having been iteratively removed. Robust optimal contracts minimize $\int_{\Lambda} [m_1(\boldsymbol{\lambda}) + m_2(\boldsymbol{\lambda})] d\boldsymbol{\gamma}$, subject to the pair of incentive constraints (24) and (26) for both agents, and the robustness constraint (25) for agent 1, as just described.

¹²Even with symmetry imposed, the form of the contracts change relative to those in the baseline model.

Following the argument in Observation 2, it can be seen that such an optimal arrangement is unique, only modulo a renaming of the players. That said, the two contracts $m_1(\lambda)$ and $m_2(\lambda)$ cannot be solved for separately, because preferences are interdependent.

6.2. Robustness and the Direction of Interdependence. It is instructive to evaluate (24) and (25) at the baseline symmetric optimum $m(\lambda) \equiv m_1(\lambda_1, \lambda_2) = m_2(\lambda_2, \lambda_1)$. In this case, it is easy to see that the left hand side of the incentive constraint (24) exceeds the left hand side of the robustness constraint (25) by the extra expression

$$(27) \quad S(\alpha) \equiv (1 + \alpha) \int_{\Lambda} u(m(\lambda)) \lambda_1 \lambda_2 d\gamma,$$

which can be interpreted as the extent of (endogenous) effort supermodularity built into the baseline optimum $m(\lambda)$. If this term is positive, then the robustness condition cannot hold at the baseline optimum, because the incentive constraints bind. On the other hand, if $S(\alpha) \leq 0$ under the baseline optimum contract, *the baseline optimum is automatically robust*. We can exploit this observation to obtain sufficient conditions for either possibility.

Proposition 3. *Assume that λ has compact support Λ and that its associated measure γ under high effort admits a continuously differentiable density. Then:*

- (a) *There exists a threshold $\epsilon > 0$ such that whenever $|\alpha| < \epsilon$:*
 - (i) *The baseline optimal contract with adversarial agents automatically satisfies the robustness constraint, and is therefore robustly implementable with no change. In contrast:*
 - (ii) *The baseline optimal contract with altruistic agents is not robustly implementable, and so must be adjusted to meet the robustness constraint for one of the agents.*

Statements (i) and (ii) are also true if at least one of the following conditions is met:

- (b) *The increasing function $f(z) \equiv u \circ u'^{-1}(1/z)$ defined for all $z > 0$ is strictly convex, or*
- (c) *Performance λ is uniformly distributed on some interval when an agent exerts effort.*

Proposition 3 suggests that optimal contracts under *adversarial* preference interdependence automatically have good robustness properties. Intuitively, when an agent's counterpart is expected to put in low effort, an adversarial agent is even more willing to choose high effort under the baseline optimal contract. (Recall that in the two-agent case, the baseline optimal contract is tournament-like and rewards the agent's effort more sharply when her counterpart expends low effort.) In contrast, when agents are altruistic, the baseline optimal contract causes them to reinforce each other's conduct, and the robustness constraint needs to be separately imposed to eliminate bad equilibria in which both

agents expend low effort. Proposition 3 establishes this result for any u satisfying our assumptions when the extent of preference interdependence is small enough.

These results also hold under either of the conditions (b) or (c) in the statement of the Proposition. Part (c) is self-explanatory, simply stating that the performance indicator has uniform density on the exertion of high effort. For part (b), it might be useful to apply (16) and note that $f(\Psi(\lambda)/k)$ can be interpreted as equilibrium felicity in the baseline model, in the region of positive payments. The requirement that f is convex pertains to scenarios in which agents are relatively risk-tolerant. For instance, if $u(m) = m^{1-\theta}$ for $\theta \in (0, 1)$, condition (b) will hold when $\theta < 1/2$.

That said, the intuition is not entirely general. The restriction that $|\alpha|$ is small can bite when neither conditions (b) nor (c) additionally hold. As an example (see Appendix C for details), use the same constant-elasticity specification in the previous paragraph with $\theta = \frac{9}{10}$, with $\alpha = \frac{1}{2}$. Assume just two possible output realizations, namely 0 and 1, with probabilities $\mu^h(\{0\}) = \frac{4}{5}$, and $\mu^\ell(\{0\}) = 1$. As already noted, condition (b) fails, because the convexity of f relies on $\theta < 1/2$. Condition (c) also fails, given the concentrations in the probability mass of λ . Now it can be shown that conclusion (ii) fails in general. For instance, if the cost of effort is around 0.336, the baseline optimal contract for altruistic agents generates effort *sub*-modularity, thereby ensuring robust implementation under the baseline equilibrium. An analogous example that violates conclusion (i) in Proposition 3 can be constructed for adversarial agents, generating *super*-modularity and the consequent non-robustness of the baseline optimal contract, even though $\alpha < 0$.

Such examples notwithstanding, our analysis of interdependent preferences suggests that at least with small adversarial interdependence, the cost-minimizing contract is automatically robust. But adjustments may be required with altruistic agents.

6.3. Optimal Contracts When Robustness Considerations Bind. We now discuss the adjustments that are entailed when robustness considerations bind. Asymmetric contracts will be typically called for, even across agents who are *ex ante* identical. Over and above such considerations, the contractual choice remains unique up to a permutation of the agents. In what follows, we presume that agent 1 is contracted with both incentive and robustness considerations in mind, so that agent 2 can have no doubt about agent 1's action, and must therefore only satisfy the incentive constraint. As before, the solution to

this problem can be obtained with the Lagrangian method. Write the Lagrangian as

$$\begin{aligned}\mathcal{L}(m) \equiv & - \int_{\Lambda} [m_1(\lambda) + m_2(\lambda)] d\gamma + v_1 \left[\int_{\Lambda} u(m_1(\lambda)) \lambda_1 d\gamma + \alpha \int_{\Lambda} u(m_2(\lambda)) \lambda_1 d\gamma - c \right] \\ & + v_2 \left[\int_{\Lambda} u(m_2(\lambda)) \lambda_2 d\gamma + \alpha \int_{\Lambda} u(m_1(\lambda)) \lambda_2 d\gamma - c \right] \\ & + \rho \left[\int_{\Lambda} u(m_1(\lambda)) \lambda_1 (1 - \lambda_2) d\gamma + \alpha \int_{\Lambda} u(m_2(\lambda)) \lambda_1 (1 - \lambda_2) d\gamma - c \right],\end{aligned}$$

where v_1 and v_2 are the multipliers on each incentive constraint, and ρ is the multiplier on the robustness constraint applied just to agent 1, all nonnegative.

Following the same arguments as in Observation 2, the first-order conditions are necessary and sufficient to describe the optimum. We write these separately for agents 1 and 2 as they enter the problem asymmetrically. To ease the writing, define $\zeta_1 \equiv v_1 + \rho$ and $\zeta_2 \equiv v_2$; each of them is strictly positive at an optimum. For agent 1, we have:

$$(28) \quad m_1(\lambda) > 0 \text{ iff } \zeta_1 \lambda_1 + \alpha \zeta_2 \lambda_2 - \rho \lambda_1 \lambda_2 > 0, \text{ with } u'(m_1(\lambda)) [\zeta_1 \lambda_1 + \alpha \zeta_2 \lambda_2 - \rho \lambda_1 \lambda_2] = 1$$

for all such λ , while for agent 2,

$$(29) \quad m_2(\lambda) > 0 \text{ iff } \zeta_2 \lambda_2 + \alpha \zeta_1 \lambda_1 - \alpha \rho \lambda_1 \lambda_2 > 0, \text{ with } u'(m_2(\lambda)) [\zeta_2 \lambda_2 + \alpha \zeta_1 \lambda_1 - \alpha \rho \lambda_1 \lambda_2] = 1$$

for all such λ . Equations (28) and (29) are entirely symmetric when the robustness constraint is not binding, in which case $\rho = 0$ and we are back to the baseline model, where Observation 2 implies $\zeta_1 = \zeta_2$. We therefore presume from now on that the robustness constraint binds, with $\rho > 0$. Then, as already discussed, the contract pair turns asymmetric. While a full description of the resulting changes necessitates a more complete analysis, equations (28) and (29) tell us exactly how these changes are wrought. They flow through the robustness multiplier ρ — which applies differently to agents 1 and 2 — and via the two multipliers ζ_1 and ζ_2 , which now depart in value from each other.

Let us examine the marginal reward to each agent following an increase in own performance. Equations (28) and (29) tell us that in the zone where payments are positive, the responses of m_1 to λ_1 and m_2 to λ_2 are given by

$$(30) \quad -\frac{u''(m_1)}{u'(m_1)^2} \frac{\partial m_1(\lambda)}{\partial \lambda_1} = \zeta_1 - \rho \lambda_2 \text{ and } -\frac{u''(m_2)}{u'(m_2)^2} \frac{\partial m_2(\lambda)}{\partial \lambda_2} = \zeta_2 - \alpha \rho \lambda_1.$$

In contrast to the baseline model, where these marginal own-responses for i were independent of compatriot performance λ_j , they now do indeed depend on λ_j . Suppose that $\alpha > 0$, which is the more relevant case for a binding robustness constraint. Restricting ourselves to the sub-domain where payments to i are positive, marginal own-responses

are highest when agent j 's performance is at its worst, flattening out as that performance improves. This is intuitive: robustness considerations must sharpen the incentives to agent 1 when 2's performance is expected to be low, relative to the baseline problem with just the incentive constraint alone.

But the same arguments hold for 2: agent 1's performance can reduce 2's marginal reward to own performance, even though agent 2 faces no robustness constraint. This might seem counterintuitive. However, agent 2's utility does enter agent 1's utility (discounted by α), tightening agent 1's robustness constraint. The higher the value of λ_1 , the more pronounced this effect becomes. That is, the robustness constraint for agent 1 is not relaxed solely through direct monetary incentives to that agent; it is also relaxed via the social component by adjusting payments to agent 2.

Now consider the cross-rewards to each agent. Using (28) and (29) again, we see that in the zone where payments are positive, the responses of m_1 to λ_2 and m_2 to λ_1 are given by

$$(31) \quad -\frac{u''(m_1)}{u'(m_1)^2} \frac{\partial m_1(\boldsymbol{\lambda})}{\partial \lambda_2} = \alpha \zeta_2 - \rho \lambda_1 \text{ and } -\frac{u''(m_2)}{u'(m_2)^2} \frac{\partial m_2(\boldsymbol{\lambda})}{\partial \lambda_1} = \alpha \zeta_1 - \alpha \rho \lambda_2.$$

Once again, these effects depend on the level of performance (unlike in the baseline model), and in contrast to (30), this time they depend on own performance. Specifically, if own performance is low, the cross-rewards are high. So the cooperative nature of contracts in the baseline model when $\alpha > 0$ is still preserved but only conditional on low performance. All this starkly reflects the need — generated by robustness — to provide incentives to produce a large λ_i when λ_j is small rather than large, so as to generate high effort under the imagined scenario that the team member is providing low effort. In contrast, when performance is high, these cross-effects are attenuated and may even turn negative, injecting a sense of competition into sub-regions of the performance space.

Despite these qualitative differences, the solution to the robustness problem shares a strong similarity with the baseline case. The new robustness considerations effectively adjust the two weights assigned to the contract components, but the overall structure of contracts remains the same for both agents. Instead of the old incentive index Ψ , the extended incentive index under robustness can be written as

$$\Psi_i - \rho_i \times \lambda_1 \times \lambda_2,$$

where the new statistics $\Psi_i = \lambda_i + (\zeta_j / \zeta_i) \alpha \lambda_j$ adjust (relative to Ψ) the weight put on compatriot performance, using the ratio of the Lagrange coefficients. The weight on the

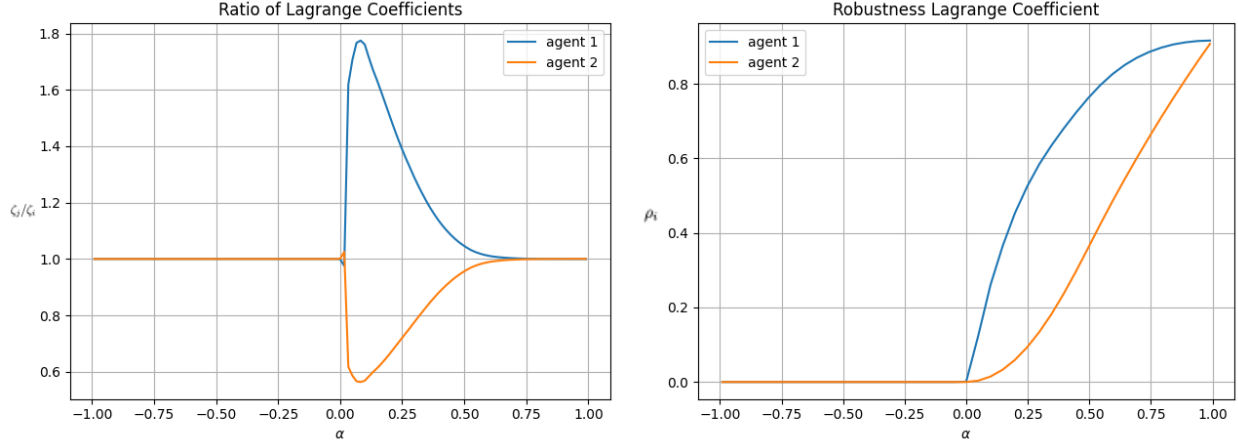


Figure 1. Asymmetry in the optimal contracts: Example: $u(m) = \sqrt{m}$, $\mu_i^h((0, y_i]) = y_i$, and $\mu_i^\ell((0, y_i]) = 2y_i - y_i^2$, both on $Y = [0, 1]$. The robustness constraint is binding (with $\rho > 0$) when $\alpha > 0$, not otherwise. Left panel: weight ζ_j/ζ_i in the Ψ_i statistic, as a function of α . Right panel: weight ρ_i as a function of α .

negative of the super-modularity term $\lambda_1 \times \lambda_2$ — the term needed for robust implementation — is

$$\rho_1 \equiv \frac{\rho}{\zeta_1} \quad \text{for agent 1, and} \quad \rho_2 \equiv \left(\frac{\rho}{\zeta_2}\right) \alpha \quad \text{for agent 2.}$$

Thus, the additional adjustments are fully characterized by the “weight distortions” given by ζ_j/ζ_i , and by ρ_i , each determined endogenously from the Lagrangian coefficients. Incentives are thereby made asymmetric in two ways. First, in the standard part Ψ_i , the agent with the larger (or smaller) value of ζ_i becomes respectively less (or more) sensitive to the other agent’s performance compared to the standard implementation. Second, the agents assign different weights to the super-modularity term.

Figure 1 plots contract coefficients as a function of α in the following example: $u(m) = \sqrt{m}$, $\mu_i^h((0, y_i]) = y_i$, and $\mu_i^\ell((0, y_i]) = 2y_i - y_i^2$, with $Y = [0, 1]$, so that λ is uniformly distributed on $\Lambda = [-1, 1]$. When $\alpha < 0$, the incentive constraints bind and the robustness constraint is slack (i.e., $\rho = 0$). When $\alpha > 0$, the robustness constraint binds, and ρ is positive and monotonically increasing in α , and agent 1, who is assigned to take care of robustness, has weights $\zeta_2 > \zeta_1$ and $\rho_1 > \rho_2$. As already indicated in a more general setting, the baseline optimal contract from Section 4.4 is robust for $\alpha < 0$, but this robustness no longer holds when $\alpha > 0$. In the example no qualification on the size of α is needed for this observation.

That negative cross-effects could arise is shown by the same example. When $\alpha = 1/2$, it is easy to check that agent 1’s cross-sensitivity to agent 2’s performance, given by $\alpha(\zeta_2/\zeta_1) -$

$\rho_1 \lambda_1$, turns negative when $\lambda_1 > 0.8$. In this sense, considerations of robustness restore some degree of competitiveness even when team members are altruistic, so as to eliminate equilibria that are unfavorable from the principal's perspective.

Both pairs of effects in equations (30) and (31) are intended to liberate agents from the robustness constraint. They are designed to shift effort incentives from regions in which performance is high to regions in which performance is low, thereby seeking to eliminate the specter of a low-effort equilibrium. That necessitates some competitive elements.

In Section 5, we argued that under the baseline model, there is a bias towards the use of cooperative or joint-liability contracts. This was formalized by the idea that in the altruistic setting, the absolute magnitudes of (normalized) cross-effects are larger than they are in the mirror-image adversarial setting. In short, the tournament-like features of optimal contracts were muted in the adversarial case, but in the altruistic case, the cooperative features were enhanced. The analysis of this Section suggests that the analysis needs to be qualified somewhat, once the model is extended to handle robust implementation. Now optimal contracts under altruism generate an induced game across agents that has effort complementarities, and possibly multiple equilibria. As we have shown, such multiplicities can be addressed by making contracts more competitive in some performance regions than they would otherwise be, but still more group-based in others. Thus, in the presence of robustness, our previous assertions do need to be qualified to some degree. A fuller analysis of these competing effects must await further research.

7. PREFERENCE INTERDEPENDENCE AND PAYOFFS TO THE PRINCIPAL

We briefly remark on principal payoffs with interdependent preferences.

Proposition 4. *In the baseline model with incentive constraints but no robustness or participation constraints, the expected payout by the principal to incentivize high effort is decreasing in α when $\alpha > 0$ and increasing in α when $\alpha < 0$, reaching a maximum when $\alpha = 0$. In particular, the principal prefers interdependent preferences of either kind — altruistic or adversarial — to independent preferences.*

Figure 2 uses the example in Section 6.2 to depict expected payout by an optimizing principal as a function of α . It is inverted-U shaped, with payouts first rising in α until $\alpha = 0$, and thereafter falling. Not only does the principal unambiguously prefer teams with interdependent preferences, she prefers a greater degree of interdependence no matter on

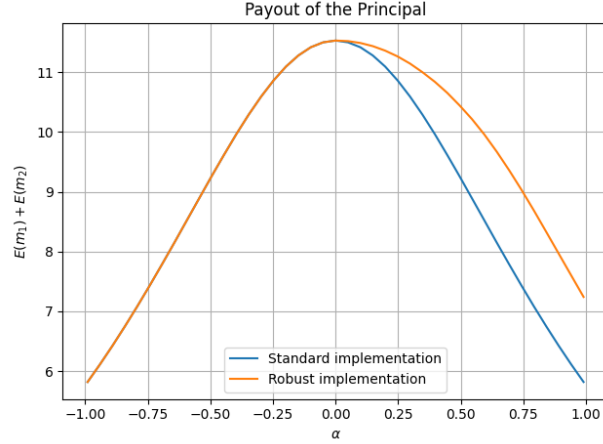


Figure 2. Expected payout by the principal, as a function of α , in the example with linear performance and $\theta = \frac{1}{2}$. The blue curve gives the payout under standard implementation; the orange curve gives the payout under robust implementation, in which high effort is the unique equilibrium.

which side of the independent benchmark those preferences lie. We should note, however, that there is no general presumption of payoff symmetry on either side of $\alpha = 0$.

Under robustness, it is necessary to amend Proposition 4. In what follows, we suppose that at least one of the sufficient conditions (b) or (c) in Proposition 3 holds. Then we know that when agents are adversarial, the robustness constraint never binds, so that robust implementation with adversarial agents would come at no additional cost. The principal's payoff looks exactly as it did before in the region $\alpha \leq 0$, and continues to increase with the absolute value of α . However, when agents are altruistic, the robustness implementation of high effort will generally require the principal to distort the original contracts. Robust implementation is indeed costlier than standard implementation in this setting.

Figure 2 illustrates this additional robustness cost as a function of α .

One lesson from this exercise that extends to all environments covered by Proposition 3 is that a principal seeking to robustly implement effort might be inclined to do so with teams of adversaries rather than friends. Additionally, confronted with altruistic teams, the principal will generally want to inject greater elements of competition into the contractual structure. This inclination runs counter to the asymmetry of contractual structure that we discussed earlier. That result emphasized how the contracts themselves lean towards group liability rather than tournaments. A full assessment of the broader picture must rest on these conflicting considerations.

Agent welfare under preference interdependence is also of interest, although it is problematic to compare or rank such welfare across situations when preferences are changing. However, it is easy to see that altruistic preferences could be detrimental to *material* agent payoffs. An agent with independent preferences and a limited liability constraint at zero consumption cannot be driven below zero payoff — she can always set effort equal to zero. However, with interdependent preferences, agents could experience negative material payoffs even with limited liability in place. Altruistic agents might choose costly effort so as to protect their partners. Adversarial agents might do the same to *reduce* favorable outcomes for others. Thus, preference interdependence can be leveraged to effectively relax the limited liability constraint, with negative implications for agent payoffs.

8. SUMMARY AND FUTURE DIRECTIONS

We’ve examined optimal contracting when agent preferences are interdependent, focusing on how these preferences affect the structure of contracts. In our framework, agents derive utility not only from their own material rewards and costs but also — via altruistic or adversarial preferences — from the payoffs of their fellow team members. We’ve shown that positive interdependence fosters cooperative joint-liability contracts, and negative interdependence leans towards competitive tournament-style contracts. That much is intuitive. But in addition, our analysis reveals a tendency for optimal contracts to favor joint-liability mechanisms over tournaments, in a sense that we’ve attempted to make precise in the paper.

Specifically, we’ve argued that altruistic teams display greater cross-impact of performance on payoffs (relative to own performance) compared to adversarial teams with the same absolute magnitude of preference interdependence. This asymmetry appears when there are at least three team members, and is heightened in larger teams. Moreover, for any team size, every agent’s contract converges to a *common* contract as altruism converges up to a limit value (given by coherence), in which payment is exactly the same for every agent. Individual performance matters for that common payment, but only insofar as it affects aggregate team performance. In the corresponding adversarial limit, there is convergence to a pure tournament with winner taking all, *but only when there are just two team members*. With three or more agents, the limit contract pays an agent for her own performance even when that performance is not ranked the best.

We then ask whether robustness constraints, designed to eliminate multiple equilibria, might impact the principal’s design and valuation of contracts. We argue that altruistic

preferences tend to endogenously generate supermodularity in incentives, which can lead to multiple equilibria. To ensure robust implementation, the principal will need to make adjustments in the baseline optimal contract. We show that the principal achieves this by increasing the competitiveness of the contract in some regions of the performance space, but accentuating group rewards in other regions. Overall, then, while our analysis highlights an asymmetry in the structure of contracts that pushes towards group liability, that asymmetry is made more complex and nuanced in the presence of robustness.

Some of the characterizations in the main text (e.g., Proposition 2 and Corollary 1) assume continuous signal distributions. This restriction is not essential. Appendix A offers more details regarding contract structure, using a two-agent example with discrete (binary) signals.

On the other hand, we do make some strong assumptions. First, we assume that the structure of preference interdependence is common knowledge, and in particular that it is known to the principal. One way to drop this assumption is to design contracts that perform well over a class of preference configurations. This would involve a robustness analysis of a different kind, in the spirit of Hurwicz and Shapiro (1978), Carroll (2015), or Dai and Toikka (2022). Given the asymmetry highlighted in this paper, we would conjecture that such a robust contract would favor group-based payments rather than tournaments.

Our specification presumes that efforts and payments are revealed *ex-post* to every agent. These conditions are not the only ones that can be studied. Two plausible alternatives suggest themselves. The first is that effort choices are not observed *ex post*, but all payments still are. The second is that neither the effort choices nor the payments to other agents are observed. The preferred interpretation that we use in the paper is, therefore, not devoid of qualification.

Finally, our model is static and in particular presumes an exogenous degree of interdependence that remains unaffected by the choices made by agents or by the contracts offered to them. However, it is important to acknowledge that this assumption may be considered restrictive in certain environments. For example, if an agent deviates by choosing low effort, thereby jeopardizing the prospects of other agents, and if other agents come to know of this deviation, a presumed altruistic value of $\alpha > 0$ *ex ante* may not apply *ex post*.¹³ Likewise, it has been observed that in certain competitive environments, initial

¹³Anger and a sense of betrayal then enter into the analysis — at least in an extended dynamic model. For related discussions in a team setting, see Ray and Ueda (1996), Section 4.1., and in the specific context of microfinance, Ghatak and Guinnane (1999), Section 3.3.

strangers who interact over time tend to develop adversarial attitudes, while cooperative environments foster friendship. This indicates that the type of contract offered can significantly influence preference linkages among agents in the long term. However, considering these phenomena involves *endogenous* preference interdependence, which falls outside the scope of this paper. Understanding this aspect is crucial for comprehending optimal design, but we defer it to future research.

REFERENCES

- Abernethy, Margaret A, Jan Bouwens, Christian Hofmann, and Laurence van Lent (2023) "Altruism, Social norms, and Incentive Contract Design," *Review of Accounting Studies*, 28 (2), 570–614.
- Axelrod, Robert and William D Hamilton (1981) "The Evolution of Cooperation," *Science*, 211 (4489), 1390–1396.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2005) "Social Preferences and the Response to Incentives: Evidence from Personnel Data," *The Quarterly Journal of Economics*, 120 (3), 917–962.
- Banerjee, Abhijit V, Timothy Besley, and Timothy W Guinnane (1994) "Thy Neighbor's Keeper: The Design of a Credit Cooperative with Theory and a Test," *The Quarterly Journal of Economics*, 109 (2), 491–515.
- Bartling, Björn (2011) "Relative performance or team evaluation? Optimal contracts for other-regarding agents," *Journal of Economic Behavior & Organization*, 79 (3), 183–193.
- Bergemann, Dirk and Stephen Morris (2009) "Robust Implementation in Direct Mechanisms," *The Review of Economic Studies*, 76 (4), 1175–1204.
- Bergstrom, Theodore C (1999) "Systems of Benevolent Utility Functions," *Journal of Public Economic Theory*, 1 (1), 71–100.
- Bernstein, Shai and Eyal Winter (2012) "Contracting with Heterogeneous Externalities," *American Economic Journal: Microeconomics*, 4 (2), 50–76.
- Besley, Timothy and Stephen Coate (1995) "Group Lending, Repayment Incentives and Social Collateral," *Journal of Development Economics*, 46 (1), 1–18.
- Camboni, Matteo and Michael Porcellacchia (2023) "Monitoring Team Members: Information Waste and Monitoring Team Members: Information Waste and the Self-Promotion Trap," *working paper*.
- Carroll, Gabriel (2015) "Robustness and Linear Contracts," *American Economic Review*, 105 (2), 536–63.
- Che, Yeon-Koo and Seung-Weon Yoo (2001) "Optimal Incentives for Teams," *American Economic Review*, 91 (3), 525–541.

- Dai, Tianjiao and Juuso Toikka (2022) "Robust Incentives for Teams," *Econometrica*, 90 (4), 1583–1613.
- DeMarzo, Peter M and Ron Kaniel (2023) "Contracting in Peer Networks," *The Journal of Finance*, 78 (5), 2725–2778.
- Dhillon, Amrita and Alexander Herzog-Stein (2009) "Games of Status and Discriminatory Contracts," *Games and Economic Behavior*, 65 (1), 105–123.
- Fehr, Ernst and Simon Gächter (2000) "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives*, 14 (3), 159–182.
- Genicot, Garance and Debraj Ray (2006) "Contracts and Externalities: How Things Fall Apart," *Journal of Economic Theory*, 131 (1), 71–100.
- Ghatak, Maitreesh (1999) "Group Lending, Local Information and Peer Selection," *Journal of Development Economics*, 60 (1), 27–50.
- Ghatak, Maitreesh and Timothy W Guinnane (1999) "The Economics of Lending with Joint Liability: Theory and Practice," *Journal of Development Economics*, 60 (1), 195–228.
- Green, Jerry R and Nancy L Stokey (1983) "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, 91 (3), 349–364.
- Halac, Marina, Ilan Kremer, and Eyal Winter (2020) "Raising Capital from Heterogeneous Investors," *American Economic Review*, 110 (3), 889–921.
- (2023) "Monitoring teams," *American Economic Journal: Microeconomics*.
- Halac, Marina, Elliot Lipnowski, and Daniel Rappoport (2021) "Rank Uncertainty in Organizations," *American Economic Review*, 111 (3), 757–86.
- Holmstrom, Bengt (1982) "Moral Hazard in Teams," *The Bell Journal of Economics*, 324–340.
- Hori, Hajime and Sadao Kanaya (1989) "Utility Functionals with Nonpaternalistic Inter-generational Altruism," *Journal of Economic Theory*, 49 (2), 241–265.
- Hurwicz, Leonid and Leonard Shapiro (1978) "Incentive Structures Maximizing Residual Gain Under Incomplete Information," *The Bell Journal of Economics*, Spring, 180–191.
- Itoh, Hideshi (2004) "Moral Hazard and Other-Regarding Preferences," *The Japanese Economic Review*, 55, 18–45.
- Jayaraman, Rajshri, Debraj Ray, and Francis De Véricourt (2016) "Anatomy of a Contract Change," *American Economic Review*, 106 (2), 316–358.
- Koçkesen, Levent, Efe A. Ok, and Rajiv Sethi (2000) "The Strategic Advantage of Negatively Interdependent Preferences," *Journal of Economic Theory*, 92 (2), 274–299, <https://doi.org/10.1006/jeth.1999.2587>.
- Lanzetta, John T and Basil G Englis (1989) "Expectations of Cooperation and Competition and Their Effects on Observers' Vicarious Emotional Responses," *Journal of Personality and Social Psychology*, 56 (4), 543.

- Lazear, Edward P (1989) "Pay Equality and Industrial Politics," *Journal of Political Economy*, 97 (3), 561–580.
- Lazear, Edward P and Sherwin Rosen (1981) "Rank-order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89 (5), 841–864.
- Letina, Igor, Shuo Liu, and Nick Netzer (2020) "Delegating Performance Evaluation," *Theoretical Economics*, 15 (2), 477–509.
- List, John (2009) "Social Preferences: Some Thoughts from the Field," *Annual Review of Economics*, 1, 563–583.
- Luft, Joan (2016) "Cooperation and Competition Among Employees: Experimental Evidence on the Role of Management Control Systems," *Management Accounting Research*, 31, 75–85.
- Ma, Ching-To (1988) "Unique Implementation of Incentive Contracts with Many Agents," *The Review of Economic Studies*, 55 (4), 555–572.
- Ma, Ching-To, John Moore, and Stephen Turnbull (1988) "Stopping Agents from "Cheating"," *Journal of Economic Theory*, 46 (2), 355–372.
- Marsh, Abigail A (2016) "Neural, Cognitive, and Evolutionary Foundations of Human Altruism," *Wiley Interdisciplinary Reviews: Cognitive Science*, 7 (1), 59–71.
- Meyer, Margaret and Dilip Mookherjee (1987) "Incentives, Compensation, and Social Welfare," *The Review of Economic Studies*, 54 (2), 209–226.
- Mookherjee, Dilip (1984) "Optimal Incentive Schemes with Many Agents," *The Review of Economic Studies*, 51 (3), 433–446.
- Mookherjee, Dilip and Stefan Reichelstein (1992) "Dominant Strategy Implementation of Bayesian Incentive Compatible Allocation Rules," *Journal of Economic Theory*, 56 (2), 378–399.
- Nalebuff, Barry J and Joseph E Stiglitz (1983) "Prizes and Incentives: Towards a General Theory of Compensation and Competition," *The Bell Journal of Economics*, 21–43.
- Pearce, David G (2008) "Nonpaternalistic Sympathy and the Inefficiency of Consistent Intertemporal Plans," *Foundations in Microeconomic Theory*, 213.
- Ray, Debraj (1987) "Nonpaternalistic intergenerational altruism," *Journal of Economic Theory*, 41 (1), 112–132, [https://doi.org/10.1016/0022-0531\(87\)90008-1](https://doi.org/10.1016/0022-0531(87)90008-1).
- Ray, Debraj and Kaoru Ueda (1996) "Egalitarianism and Incentives," *Journal of Economic Theory*, 71, 324–348.
- Ray, Debraj & Rajiv Vohra (2020) "Games of Love and Hate," *Journal of Political Economy*, 128 (5), 1789–1825.
- Robson, Arthur J (2017) "Group Selection: A Review Essay on Does Altruism Exist?" *Journal of Economic Literature*, 55 (4), 1570–1582.

- Sarkisian, Roberto (2017) “Team Incentives under Moral and Altruistic Preferences: Which Team to Choose?” *Games*, 8 (3), 37.
- Segal, Ilya (1999) “Contracting with Externalities,” *The Quarterly Journal of Economics*, 114 (2), 337–388.
- (2003) “Coordination and Discrimination in Contracting with Externalities: Divide and Conquer?” *Journal of Economic Theory*, 113 (2), 147–181.
- Sobel, Joel (2005) “Interdependent Preferences and Reciprocity,” *Journal of Economic Literature*, 43 (2), 392–436.
- Stiglitz, Joseph E (1990) “Peer Monitoring and Credit Markets,” *World Bank Economic Review*, 4 (3), 351–366.
- Trivers, Robert (2006) “Reciprocal Altruism: 30 Years Later,” *Cooperation in Primates and Humans: Mechanisms and Evolution*, 67–83.
- Vásquez, Jorge and Marek Weretka (2021) “Co-worker Altruism and Unemployment,” *Games and Economic Behavior*, 130, 224–239.
- Wantchekon, L (1994) “Tournaments and Optimal Contracts for Teams in Rural West Africa,” *mimeo*, Department of Economics, Northwestern University.
- Winter, Eyal (2004) “Incentives and Discrimination,” *American Economic Review*, 94 (3), 764–773.
- Zillman, Dolf and Joanne R Cantor (1977) “Affective Responses to the Emotions of a Protagonist,” *Journal of Experimental Social Psychology*, 13 (2), 155–165.

APPENDIX

A. An Example With Two Agents and Binary Signals. This appendix contains a simple example with binary signals $Y = \{0, 1\}$, where $y = 0$ represents failure, and $y = 1$ success. Let $\mu^e(y)$ denote the probability of y when an agent exerts effort $e \in \{\ell, h\}$. Recall that

$$\lambda(y) = 1 - \frac{\mu^\ell(y)}{\mu^h(y)}.$$

Given that λ is strictly increasing, and has expected value of 0 under μ^h , we have $\lambda(0) < 0$ and $\lambda(1) > 0$. For two agents, i and j , a typical (symmetric) compensation scheme is shown in Table 1.

This binary example provides clearer interpretations for the various components of a contract. We can define its “base wage” component to be $\min_{i,j} \hat{m}(y_i, y_j)$. Then there

	$y_j = 0$	$y_j = 1$
$y_i = 0$	$\hat{m}(0,0)$	$\hat{m}(0,1)$
$y_i = 1$	$\hat{m}(1,0)$	$\hat{m}(1,1)$

Table 1. A typical two-agent contract, showing compensation to row agent i

are various additional bonuses, depending on the outcomes of the agent and her compatriot. For instance, one might define a “performance bonus” as the difference between $\min_j \hat{m}(1, y_j)$ and the base wage, or a “winner’s bonus” as $\hat{m}(1, 0)$ net of both the base wage and the performance bonus. Similarly, a “team bonus” can be thought of as $\hat{m}(1, 1) - \text{base wage} - \text{performance bonus}$. And the “consolation prize” paid to a losing agent in the event $(0, 1)$ is $\hat{m}(0, 1)$ net of the base wage. As in the general model, the principal’s objective is to choose this entire pay and bonus system to minimize the expected wage bill while ensuring that high effort levels constitute a Nash equilibrium.

When preferences are independent ($\alpha = 0$), the incentive compatibility constraint reduces to $\mathbb{E}[\lambda(y_i)u(\hat{m}(y_i, y_j))] \geq c$, where the expectation is taken with respect to the joint probability $d\mu_i^h d\mu_j^h$. Cost-minimizing agent compensation does not depend on compatriot output, as each agent is risk-averse. It is simply the smallest non-negative payment, con-

	$y_j = 0$	$y_j = 1$
$y_i = 0$	0	0
$y_i = 1$	$u^{-1}(\frac{c}{\mu^h(1)\lambda(1)})$	$u^{-1}(\frac{c}{\mu^h(1)\lambda(1)})$

Table 2. Optimal compensation to i with independent preferences.

ditioned on y_i but *not* on y_j , that satisfies the incentive compatibility constraint. There is a bonus for good performance, but it is independent of compatriot output; see Table 2.

Equal Odds. First, we consider what might be called the *equal odds case*, with $\mu^h(1) = \mu^h(0) = \frac{1}{2}$, for which the performance function is easily seen to be symmetric, with $\lambda(0) = -\lambda(1)$, each realization happening with equal probability.¹⁴ This is a binary counterpart of our example in the main text with continuous outputs and a linear performance function. Equal odds is a knife-edge case that we drop below, but it helps the exposition.

¹⁴The expectation being zero implies $-\lambda(0)/\lambda(1) = \mu^h(1)/\mu^h(0)$. Any asymmetry in μ^h directly affects λ .

For any α with $|\alpha| < 1$, both $\lambda(0) + \alpha\lambda(0)$ and $\lambda(0) + \alpha\lambda(1)$ are negative. Therefore payment on failure is zero, which is perhaps unsurprising given the absence of a participation constraint. However, in the event of success, there are now two *distinct* payments to i , conditioned on j 's output. When there is altruism, a higher payment for i 's effort in events of the form $(y_i, 0)$ comes at a cost. By contract symmetry, they also correspond to higher payments (to j) in events of the form $(0, y_j)$, which *dampen* i 's incentives to supply effort in such scenarios (i shares in j 's success via altruism). This leads the principal to tone down the performance bonus — $\hat{m}(1, 0)$ in this situation — though she is happy to retain a positive team bonus: $\hat{m}(1, 1) - \hat{m}(1, 0) > 0$. The opposite occurs when agents are adversarial, resulting in a winner's bonus: $\hat{m}(1, 0) - \hat{m}(1, 1) > 0$.

The advantage of the binary example is that it is easy to compute the full path of performance and team bonuses for any α . For instance, when felicity functions have constant risk aversion $u(m) = m^{1-\theta}$ for $\theta \in (0, 1)$, payments can be derived in closed form:

$$\hat{m}(1, 0) = \left[\frac{1}{1-\alpha} \frac{1}{\left(\frac{1+\alpha}{1-\alpha}\right)^{\frac{1}{\theta}} + 1} \frac{4c}{\lambda(1)} \right]^{\frac{1}{1-\theta}} \quad \text{and} \quad \hat{m}(1, 1) = \left[\frac{1+\alpha}{1-\alpha} \right]^{\frac{1}{\theta}} \left[\frac{1}{1-\alpha} \frac{1}{\left(\frac{1+\alpha}{1-\alpha}\right)^{\frac{1}{\theta}} + 1} \frac{4c}{\lambda(1)} \right]^{\frac{1}{1-\theta}}.$$

Figure 3 shows the optimal payments under equal odds for different degrees of pref-

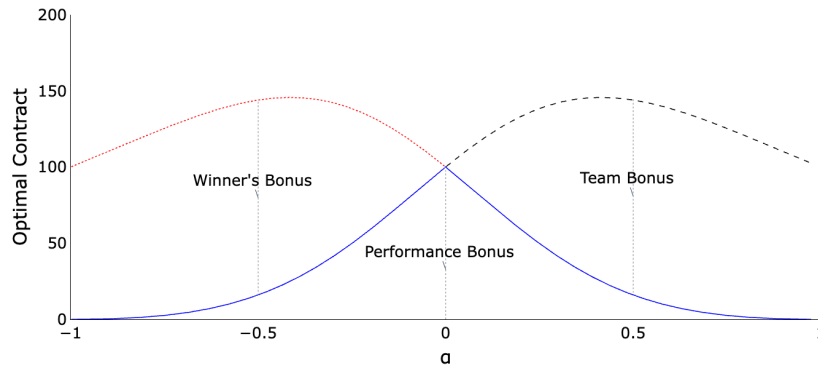


Figure 3. Agent payouts as a function of α , computed for $\theta = 0.5$ in the equal odds case.

erence interdependence, with α on the horizontal axis and various bonuses on the vertical axis, and with θ set to $\frac{1}{2}$. The blue line represents the payment when the agent performs, the vertical distance between the red and blue lines is the winner's bonus, and the distance between the black and blue lines is the team bonus. The optimal mechanism continuously evolves from competitive to cooperative as α varies between -1 and 1 . The mechanism is a pure tournament for $\alpha \approx -1$, a two-bonus contract with a winner's bonus

for $\alpha \in (-1, 0)$, independent contracts for $\alpha = 0$, a two-bonus contract with a team bonus for $\alpha \in (0, 1)$, and finally, as $\alpha \approx 1$, a pure joint liability mechanism.

With strictly concave felicity, optimal contracts satisfy $\hat{m}(0, 0) = \hat{m}(0, 1) = 0$ and $\hat{m}(1, 0) \geq \hat{m}(1, 1)$ when $\alpha \leq 0$. As already seen, the differences are highlighted when $|\alpha| \rightarrow 1$, so that agents become extremely altruistic or adversarial. The limit contracts take pure forms:

	$y_j = 0$	$y_j = 1$		$y_j = 0$	$y_j = 1$
$y_i = 0$	0	0	$y_i = 0$	0	0
$y_i = 1$	0	$u^{-1}\left(\frac{c}{\mu^h(1)\lambda(1)}\right)$	$y_i = 1$	$u^{-1}\left(\frac{c}{\mu^h(1)\lambda(1)}\right)$	0

Table 3. Altruistic and adversarial limits. Left: i 's reward as $\alpha \uparrow 1$. Right: i 's reward as $\alpha \downarrow -1$.

Table 3 collects these limit results, which differ from the independent case in Table 2. The principal uses contracts with two types of bonuses. For agents with altruistic preferences, there is a bonus $\hat{m}(1, 0)$ for individual performance, but the emphasis is on the *team* bonus, $\hat{m}(1, 1) - \hat{m}(1, 0)$, paid when both agents are successful. This team bonus fully dominates as α approaches one, with the individual performance bonus dwindling to zero. In that limit, the mechanism is one of joint liability: it *only* pays out when both agents succeed.

In contrast, for adversarial preferences, an agent's performance bonus is equal to $\hat{m}(1, 1)$, but the emphasis is now on the *winner's* bonus, $\hat{m}(1, 0) - \hat{m}(1, 1)$, a transfer paid if an agent outperforms their opponent. When α is close to negative one, a pure tournament is offered with just a winner's bonus: a player receives compensation *only* if they win.

Dropping the Equal-Odds Benchmark. With two agents, the equal-odds benchmark creates a symmetric mirror around positive and negative α (Figure 3), which could be misleading. When $\mu^h(1) \neq \frac{1}{2}$, as it generically must be, λ is no longer symmetric around success and failure, resulting in qualitative changes to optimal contracts. Consider first the “high-failure” case, in which $\mu^h(0) > \mu^h(1)$. Here, when $\alpha > \frac{\mu^h(1)}{\mu^h(0)}$, contracts also include a consolation prize $\hat{m}(0, 1) > 0$, paid when an agent fails and is outperformed by her peer. This prize and the performance bonus remain bounded away from zero even as α approaches one.

It may seem counterintuitive that, in the absence of a participation constraint, an agent receives positive compensation even when their output indicates shirking. However, in the high-failure case, the absolute value of $\lambda(0)$ is relatively small, which limits the direct disincentive effect of a payment received upon failure. At the same time, with a high

degree of altruism, the consolation prize offered to one agent is internalized by the other agent if the latter wins. Because $\lambda(1)$ is relatively large, such additional sources of utility serve as powerful motivators for effort. This latter effect dominates when α is larger than the threshold mentioned earlier. Similar forces cause the performance bonus to persist even for α arbitrarily close to one.

Conversely, in the “high-success” case with $\mu^h(1) > \mu^h(0)$, payoffs received upon failure are strong *de*-motivators, because $\lambda(0)$ is now large in absolute value. Consequently, the payment $\hat{m}(1, 0)$ received by an agent when her compatriot loses discourages the compatriot’s work. Therefore, for $\alpha > \frac{\mu^h(0)}{\mu^h(1)}$, the principal has incentives to set the payments in this event to zero. As a result, the optimal mechanism is one of pure joint liability, without a winner’s bonus.

Contracts for asymmetric λ can be derived in closed form and are depicted in Figure 4.

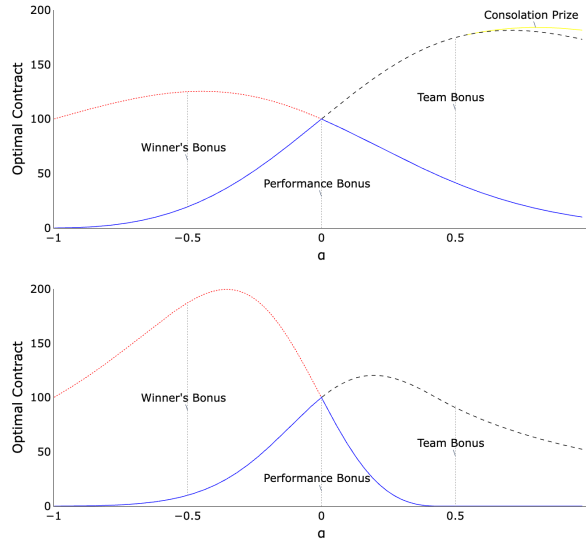


Figure 4. Agent payouts as a function of α , computed for $\theta = 0.5$ and $\mu^h(1) = 0.35$ (left panel) and $\mu^h(1) = 0.7$ (right panel).

In summary, probability distributions play a crucial role in how strongly agents respond to their own and others’ material incentives in the case of success or failure. The effectiveness of the mechanisms varies with α , as optimal contracts utilize the interdependence of preferences in success and failure events differently, leading to asymmetrical outcomes.

Interdependent Preferences with Many Agents. Return to equal odds, but now suppose there are three agents. Then the coherence condition is $|\alpha| < 1/2$, and in this case, we have:

$$(32) \quad V_i = [\mathbb{E}u(\hat{m}_i|\mathbf{e}) - c\mathbf{1}_i^h] + \frac{\alpha}{1-\alpha} \sum_{j \neq i} [\mathbb{E}u(\hat{m}_j|\mathbf{e}) - c\mathbf{1}_j^h].$$

Optimal contracts now potentially depend on the outputs of all three agents, specifying payments for eight possible outcomes, $\hat{m} : \{0,1\}^3 \rightarrow \mathbb{R}_+$. The incentive constraint for agent i is:

$$(33) \quad \mathbb{E} \left(\left[\lambda_i + \frac{\alpha}{1-\alpha} \sum_{j \neq i} \lambda_j \right] u(\hat{m}(y_i, y_j, y_k)) \right) \geq c,$$

where, as before, the expectations operator is taken with respect to the high-effort joint probability.

Now optimal contracts exhibit qualitative differences across positively and negatively interdependent preferences. Consider first the scenario of adversarial agents, where $\alpha < 0$. In such cases, agent i will be rewarded *if and only if* she is successful, i.e., $\hat{m}(0, y_j, y_k) = 0$ and $\hat{m}(1, y_j, y_k) > 0$ for every y_j and y_k . Moreover, that payment is negatively correlated with the performance of her compatriots:

$$\hat{m}(1, 0, 0) > \hat{m}(1, 1, 0) = \hat{m}(1, 0, 1) > \hat{m}(1, 1, 1) > 0.$$

This scheme can be implemented by offering a performance bonus along with two additional payments: one a winner's bonus for outperforming one team member and the other a top-up for outperforming both agents. This scheme *does not converge to a pure tournament* as adversarial preferences become extreme. In particular, the performance bonus $\hat{m}(1, 1, 1)$ converges to a strictly positive value as α approaches the adversarial limit. (Some tournament-like features are retained via competitive bonuses.) We depict the optimal contracts in Figure 5.

Now consider the altruistic case. The optimal contract is “collaborative” in the sense that payments are positively correlated with the success of others. But there are other nuances. In situations with mild interdependence, indicated by $\alpha < \frac{1}{3}$, an agent receives a performance bonus $\hat{m}(1, 0, 0) > 0$ and additional rewards for each successful team member:

$$\hat{m}(1, 1, 1) > \hat{m}(1, 1, 0) = \hat{m}(1, 0, 1) > \hat{m}(1, 0, 0) > 0.$$

When $\alpha > \frac{1}{3}$, that performance bonus disappears (recall it is defined as the minimum payment if the agent performs). That is, $\hat{m}(1, 0, 0) = 0$. In order to receive monetary compensation, an agent must demonstrate joint success with at least one other team member.

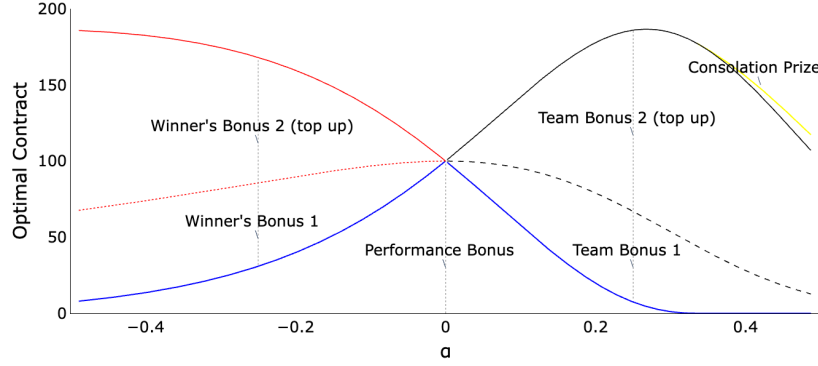


Figure 5. Agent payouts as a function of α , computed for $\theta = 0.5$ and $n = 3$.

Additionally, each agent is guaranteed a consolation prize if they fail but the other two agents perform successfully ($\hat{m}(0, 1, 1) > 0$). These contracts are also shown in Figure 5.

With three agents involved, it becomes evident that optimal contracts are not simple mirror images as the interdependence coefficient goes from negative to positive. This phenomenon illustrates the inherent asymmetry described in the main text. The cooperativeness of the contract for any given $\alpha > 0$ is more pronounced than its corresponding competitiveness at $-\alpha$.

These observations extend to a setting with an arbitrary number of agents. With binary outputs, the optimal contract is determined entirely by the individual performance of an agent and the number of other successful agents. When there is negative or moderately positive interdependence, i.e., $\alpha < \frac{1}{2n-3}$, a contract rewards an agent only if she is herself successful. The monetary compensation is positively (negatively) related to the successes of other agents in the case of positive (negative) interdependence. When there is significant positive interdependence, i.e., $\alpha > \frac{1}{2n-3}$, cooperative contracts undergo a qualitative change. Achieving personal success is no longer sufficient for compensation. It isn't necessary either: agents might be additionally rewarded with a consolation prize when a critical mass $n^* = \frac{1}{2} \left[\frac{1-\alpha(n-2)}{\alpha} + n - 1 \right]$ or more perform well.¹⁵

B. Participation Constraints. Our analysis can be easily extended to include participation constraints of the form

$$\left[\int_{\Lambda} u(m(\lambda^{[i]})) d\gamma - c \right] + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq i} \left[\int_{\Lambda} u(m(\lambda^{[j]})) d\gamma - c \right] \geq v^0$$

¹⁵The threshold n^* can be computed as the solution to a “zero- Ψ ” condition, under the assumption that i fails while exactly n^* succeed: $\lambda(0) + \frac{\alpha}{1-\alpha(n-2)} [n^* \lambda(1) + (n-1-n^*) \lambda(0)] = 0$, and using equal odds.

for every agent i , where v^0 is a (renormalized) outside-option payoff, common to every agent. By the same change of variables in Observation 1 that we used to arrive at the incentive constraint (10), we can easily restate this participation constraint for agent 1 as

$$(34) \quad \int_{\Lambda} u(m(\lambda)) d\gamma - c \geq \frac{v^0[1 - \alpha(n-2)]}{1 + \alpha} \equiv w \geq 0.$$

How does the additional constraint (34) affect our results? Recall the Lagrangian used in the argument leading up to Proposition 1 and modify it in the obvious way:

$$\mathcal{L}(m) \equiv - \int_{\Lambda} m(\lambda) d\gamma + \nu \left[\int_{\Lambda} u(m(\lambda)) \Psi(\lambda) d\gamma - c \right] + \nu^P \left[\int_{\Lambda} u(m(\lambda)) d\gamma - w \right],$$

where ν and ν^P are multipliers attached to the incentive constraint (10) and the participation constraint (34) respectively. The first order conditions for this problem are given by

$$(35) \quad -1 + \nu u'(m(\lambda)) \Psi(\lambda) + \nu^P u'(m(\lambda)) \leq 0, \text{ with equality if } m(\lambda) > 0,$$

for each λ . It can be seen that the incentive constraint always binds, so $\nu > 0$.¹⁶ Now define a “joint multiplier” $\nu^J \equiv \nu^P / \nu \geq 0$; then the first order condition above reduces to

$$(36) \quad u'(m(\lambda)) [\Psi(\lambda) + \nu^J] \text{ is constant in } \lambda \text{ if } \Psi(\lambda) > -\nu^J, \text{ and } m(\lambda) = 0 \text{ otherwise,}$$

where we’ve used $u'(0) = \infty$. Of course, the participation constraint may or may not be binding, so that ν^P could be positive or zero depending on the parameters of the problem. (In the latter case we’re back to the baseline without participation constraints.)

The optimal contract has a common property as we range over the realizations of different $\mathbf{y}$ -values. In any situation with $m(\lambda) > 0$, any change in λ that raises $\Psi(\lambda)$ must create a larger reward for our agent. This gives us an immediate analogue of Propositions 1 and 2:

Observation 3 (Extension of Propositions 1 and 2). *There exists a Lagrange multiplier on participation ν^J such that the reward to an agent is positive whenever $\Psi(\lambda) > -\nu^J$, and zero when $\Psi(\lambda) \leq -\nu^J$. Moreover, when $\Psi(\lambda) > -\nu^J$, an agent’s reward moves with outputs as summarized in (17). In particular,*

$$(37) \quad \xi(\alpha, n) = \frac{\alpha}{1 - \alpha(n-2)}$$

describes the relative cross-impact of the contract.

¹⁶Suppose on the contrary that $\nu = 0$, then, by $w \geq 0$ and $c > 0$, we have $m(\lambda) > 0$ for some subset of λ -realizations. That implies $\nu^P > 0$. But then by (35), $m(\lambda)$ equals a constant r^* for all λ , which reduces (10) to $\int_{\Lambda} \Psi(\lambda) d\gamma \geq c/u(r^*) > 0$, a contradiction, because the left hand side of this inequality is zero.

With a participation constraint, the boundary for the payment zone is attained on a lower isoquant, $\Psi|_{=-v^I}$, instead of $\Psi|_{=0}$. So the payment zone expands but within it the isoquants for contract $m(\lambda)$ remain aligned with the corresponding curves for Ψ . Therefore, the introduction of an outside option does not alter the own- and cross-effects in the contract. However, payment *levels* adjust. (For any λ , $m(\lambda)$ rises if $v^I > 0$.)

C. Proofs and Other Details. We collect technical arguments and related details here.

Proof of Observation 1. Consider an arbitrary incentive-compatible contract $(\hat{m}_1, \dots, \hat{m}_n)$. Following the steps from Section 4.1, the constraint can be written as

$$(38) \quad \int_Y \left[u(\hat{m}_i(\mathbf{y}))\lambda_i(y_i) + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq i} u(\hat{m}_j(\mathbf{y}))\lambda_i(y_i) \right] d\mu^h \geq c.$$

For any i and λ , let $m_i(\lambda)$ be the *conditional certainty equivalent* of the payment $\hat{m}_i$:

$$(39) \quad u(m_i(\lambda)) \equiv \mathbb{E}(u(\hat{m}_i(\mathbf{y})) \mid \lambda(\mathbf{y}) = \lambda).$$

By construction, the contract $(m_1, \dots, m_n)$ is measurable with respect to λ . Iterating expectations conditional on $\lambda(\mathbf{y}) = \lambda$,

$$\int_{\Lambda} \left[u(m_i(\lambda))\lambda_i + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq i} u(m_j(\lambda))\lambda_i \right] d\gamma = \int_Y \left[u(\hat{m}_i(\mathbf{y}))\lambda_i(y_i) + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq i} u(\hat{m}_j(\mathbf{y}))\lambda_i(y_i) \right] d\mu^h \geq c.$$

(Similar equalities hold verbatim for the robustness and participation constraint when $\hat{m}$ satisfies them.) Because u is strictly increasing and strictly concave, u^{-1} is strictly increasing and strictly convex, so Jensen's inequality applied conditionally on λ yields

$$(40) \quad m_i(\lambda) = u^{-1}(\mathbb{E}[u(\hat{m}_i(\mathbf{y})) \mid \lambda]) \leq \mathbb{E}(\hat{m}_i(\mathbf{y}) \mid \lambda(\mathbf{y}) = \lambda),$$

with strict inequality whenever $\hat{m}_i$ is nondegenerate given λ . Integrating (40) over λ ,

$$\int_{\Lambda} m_i(\lambda) d\gamma \leq \int_Y \hat{m}_i(\mathbf{y}) d\mu^h,$$

again, with strict inequality when $\hat{m}_i$ is nondegenerate given λ . So, when $(\hat{m}_1, \dots, \hat{m}_n)$ is optimal, the corresponding $(m_1, \dots, m_n)$ as defined here is feasible and costs weakly less, strictly so if $\hat{m}_i$ is nondegenerate at λ for any positive γ -measure of performances.

Proof of Observation 2. A principal offers each agent i a limited liability contract $m_i(\lambda)$, measurable with respect to λ . The principal seeks to

$$\text{minimize}_{(m_1, \dots, m_n) \geq 0} \int_{\Lambda} \left[\sum_{i=1}^n m_i(\lambda) \right] d\gamma,$$

where the integral assumes high effort and so is taken over the (product) distribution γ of λ . Using arguments from the main text, we can write the incentive constraints as

$$\int_{\Lambda} \left[u(m_i(\lambda))\lambda_i + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq i} u(m_j(\lambda))\lambda_j \right] d\gamma \geq c,$$

for every individual i , where u is a strictly increasing and strictly concave function of the non-negative monetary payoffs.

Write $u_i(\lambda) = u(m_i(\lambda))$ as the promised (material) utility to agent i under λ . Observe that this entire vector of promised utilities, defined for every λ , can be viewed as an equivalent representation of the contractual arrangement. Let $h \equiv u^{-1}$. Then h is strictly increasing and strictly convex. Our minimization problem changes to:

$$(41) \quad \text{minimize}_{(u_1, \dots, u_n) \geq 0} \int_{\Lambda} \left[\sum_{i=1}^n h(u_i) \right] d\gamma,$$

subject to the constraint

$$(42) \quad \int_{\Lambda} \left[u_i \lambda_i + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq i} u_j \lambda_j \right] d\gamma \geq c \text{ for all } i,$$

where it is understood throughout that u_i and all other u_j move with λ . So the optimal contract is the argmin of a strictly convex function of random utilities over a convex polyhedron defined by incentive and limited-liability linear inequality constraints. By standard arguments, such an argmin is unique almost surely (with respect to the measure γ on λ conditional on high effort).

A contract $\mathbf{u} \equiv (u_1, \dots, u_n)$ is *symmetric* if there is a common function u on Λ , invariant to permutations of its last $n-1$ entries, such that for every i , $u_i(\lambda) = w(\lambda^{[i]})$ almost surely, where $\lambda^{[i]} = (\lambda_i, \lambda_{-i})$. We next show that the unique solution $\mathbf{u}$ to the optimization problem is symmetric. For any permutation π of $\{1, \dots, n\}$, write $\lambda^\pi \equiv (\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)})$, and consider the alternative contract $\mathbf{u}' \equiv (u'_1, \dots, u'_n)$ obtained by relabeling *all* components of the original contract at once:

$$u'_i(\lambda) \equiv u_{\pi^{-1}(i)}(\lambda^\pi) \text{ for every } i.$$

Because the components of λ are independent and identically distributed, the map $\lambda \mapsto \lambda^\pi$ preserves the measure γ . This change of variables converts agent i 's incentive constraint under $\mathbf{u}'$ into agent $\pi^{-1}(i)$'s incentive constraint under $\mathbf{u}$ — every component of the contract must be relabeled in concert for the third-party terms in each constraint to map correctly — and it leaves the objective (41) unchanged, because that objective sums

over all agents. So $\mathbf{u}'$ satisfies all the constraints and yields the same cost as $\mathbf{u}$. Hence $\mathbf{u}'$ is optimal. By almost-sure uniqueness of the optimal contract (component by component), $u'_i = u_i$ for every i ; that is, for every permutation π and every i ,

$$u_i(\boldsymbol{\lambda}) = u_{\pi^{-1}(i)}(\boldsymbol{\lambda}^\pi) \text{ almost surely under } \gamma.$$

Now specialize. When π transposes 1 and i , we obtain $u_i(\boldsymbol{\lambda}) = u_1(\boldsymbol{\lambda}^{[1 \leftrightarrow i]})$ almost surely, where $\boldsymbol{\lambda}^{[1 \leftrightarrow i]}$ is the vector $\boldsymbol{\lambda}$ with components λ_1 and λ_i switched: every agent is offered the same contractual form. When π transposes two indices j and k , both distinct from 1, we obtain $u_1(\boldsymbol{\lambda}) = u_1(\boldsymbol{\lambda}^{[j \leftrightarrow k]})$ almost surely; because such transpositions generate all permutations of the components other than the first, u_1 is symmetric in the performances of agents other than 1. Setting $u \equiv u_1$ delivers symmetry in the sense defined above. The uniqueness and symmetry of the optimal contracts $(m_1, \dots, m_n)$ in the space of monetary rewards follows immediately.

Remark on first-order conditions. The reformulation used in this proof also underlies the Lagrangian arguments of Sections 4.4 and 6. In promised-utility space, the principal minimizes the strictly convex objective (41) subject to constraints — incentive, robustness, and participation alike — that are linear in promised utilities, together with the linear limited-liability restriction. The problem is therefore a convex program. Slater's condition holds: scaling any feasible utility assignment up by a common factor exceeding one produces strict slack in every binding constraint, since each has a strictly positive right-hand side. Standard convex-duality arguments then deliver Lagrange multipliers supporting the (finite) value of the program, whether or not that value is attained. Consequently, any optimum — when one exists — must satisfy the first-order conditions stated in the text, and conversely, any contract satisfying them is optimal; so wherever we exhibit a solution to the first-order conditions, optimality and existence are established simultaneously. An optimum can fail to exist only when the first-order system admits no solution; for instance, the extreme situation flagged in Footnote 10 of Section 4.5.

Proof of Proposition 3, Part (a). Recall from (27) that $S(\alpha) = (1 + \alpha) \int_{\Lambda} u(m(\boldsymbol{\lambda})) \lambda_1 \lambda_2 d\gamma$. From (16), we also know that $m(\boldsymbol{\lambda}) = 0$ as long as $\lambda_1 + \alpha \lambda_2 \leq 0$, whereas if $\lambda_1 + \alpha \lambda_2 > 0$, then $m(\boldsymbol{\lambda}) > 0$ and $u'(m(\boldsymbol{\lambda}))[\lambda_1 + \alpha \lambda_2] = k(\alpha)$ for some $k(\alpha) > 0$. Combining all this information, we must conclude that

$$(43) \quad S(\alpha) \equiv (1 + \alpha) \int_{\Lambda} G(\boldsymbol{\lambda}, \alpha) d\gamma, \text{ where } G(\boldsymbol{\lambda}, \alpha) = f\left(\frac{\lambda_1 + \alpha \lambda_2}{k(\alpha)}\right) \lambda_1 \lambda_2,$$

and where $f(z) \equiv u \circ u'^{-1}(1/z)$ for $z > 0$ and $f(z) = 0$ for $z \leq 0$. It can be seen that f is continuous on $\mathbb{R}$, differentiable on $z > 0$, and that $f'(z) > 0$ for all $z > 0$. Set $f'(z) \equiv 0$

for $z < 0$, the derivative of f on that region; the remaining point $z = 0$ corresponds to the frontier $\{\lambda_1 + \alpha\lambda_2 = 0\}$, which is γ -null under our density assumption.

We will prove that $S'(0) > 0$, which, along with $S(0) = 0$, will establish part (a).

The following technical lemma collects some observations that are used in the differentiation of S .

Lemma 1. *Assume that λ has compact support Λ and that its associated measure γ under high effort admits a continuously differentiable density. Then: (i) The map $k(\alpha)$ is C^1 on a neighborhood of 0, with $k(\alpha) > 0$.*

(ii) The map $S(\alpha)$ is continuously differentiable in a neighborhood of 0, and $S'(\alpha)$ may be computed by differentiating (43) under the integral sign.

The proof of Lemma 1 (using the implicit function theorem for (i) and an elementary differentiation-under-the-integral argument for (ii)) is omitted but is available on request from the authors.

Now, for each λ off the γ -null frontier $\{\lambda_1 + \alpha\lambda_2 = 0\}$, G is differentiable in α with

$$\partial_\alpha G(\lambda, \alpha) = f' \left(\frac{\lambda_1 + \alpha\lambda_2}{k(\alpha)} \right) \left[\frac{\lambda_2}{k(\alpha)} - \frac{(\lambda_1 + \alpha\lambda_2) k'(\alpha)}{k(\alpha)^2} \right] \lambda_1 \lambda_2,$$

where we've used Lemma 1(i). Evaluated at $\alpha = 0$, this expression reduces to

$$(44) \quad \partial_\alpha G(\lambda, 0) = \frac{1}{k(0)} f' \left(\frac{\lambda_1}{k(0)} \right) \lambda_1 \lambda_2^2 - \frac{k'(0)}{k(0)^2} f' \left(\frac{\lambda_1}{k(0)} \right) \lambda_1^2 \lambda_2.$$

Using $\int \lambda_2 d\gamma_2 = 0$ and the independence of λ_1 and λ_2 , it is easy to see that $S(0) = 0$. Therefore, using (43) and Lemma 1(ii),

$$(45) \quad S'(0) = \int_\Lambda \partial_\alpha G(\lambda, 0) d\gamma.$$

Combining (44) and (45) and recalling again the independence of λ_1 and λ_2 , we see that:

$$(46) \quad S'(0) = \frac{1}{k(0)} \left[\int_\Lambda f' \left(\frac{\lambda_1}{k(0)} \right) \lambda_1 d\gamma \right] \left[\int_\Lambda \lambda_2^2 d\gamma \right] - \frac{k'(0)}{k(0)^2} \left[\int_\Lambda f' \left(\frac{\lambda_1}{k(0)} \right) \lambda_1^2 d\gamma \right] \left[\int_\Lambda \lambda_2 d\gamma \right].$$

Because $\int \lambda d\gamma = 0$, the second term vanishes. The first term is strictly positive, by Lemma 1(i) and the fact that $f'(z) > 0$ for $z > 0$. Hence $S'(0) > 0$.

Because $S(0) = 0$, the fact that $S'(0) > 0$ means that there is $\epsilon > 0$ such that $S(\alpha) < 0$ for $\alpha \in (-\epsilon, 0)$ and $S(\alpha) > 0$ for $\alpha \in (0, \epsilon)$, which completes part (a) of the proof.

Part (b). Recall $S(\alpha)$ from (43) and write it more explicitly as a weighted integral over λ_2 :

$$(47) \quad S(\alpha) = (1 + \alpha) \int_{-\infty}^1 \underbrace{\left[\int_{-\alpha\lambda_2}^1 f\left(\frac{\lambda_1 + \alpha\lambda_2}{k}\right) \lambda_1 d\gamma_1 \right]}_{\text{Weighting } W(\lambda_2)} \lambda_2 d\gamma_2,$$

where the highlighted term is the weighting function $W(\lambda_2)$ for the integration. This weighting function itself has weights within it, and can be written as

$$W(\lambda_2) \equiv \int_{-\infty}^1 \underbrace{f\left(\frac{\lambda_1 + \alpha\lambda_2}{k}\right)}_{\text{Weights for } \lambda_1} \lambda_1 d\gamma_1,$$

which is a weighted integral of λ_1 with f as the weights. Note that $f(z) = 0$ for $z \leq 0$, continuous at $z = 0$, and is strictly increasing for $z \geq 0$. So, using the fact that $\int \lambda_1 d\gamma_1 = 0$, we see that $W(\lambda_2) \geq 0$ for every λ_2 , and that $W(\lambda_2) > 0$ whenever $-\alpha\lambda_2$ lies below the maximum of the support of λ_1 ; elsewhere the integrand vanishes identically.

Differentiation with respect to λ_2 tells us that

$$(48) \quad W'(\lambda_2) \equiv \frac{\alpha}{k} \int_{-\alpha\lambda_2}^1 f'\left(\frac{\lambda_1 + \alpha\lambda_2}{k}\right) \lambda_1 d\gamma_1,$$

where we use the fact that $f(0) = 0$. Now, by assumption, f is strictly convex so f' is strictly increasing in z , and therefore strictly increasing in λ_1 (for given λ_2) in (48) above. Using $\int \lambda_1 d\gamma_1 = 0$, we must therefore conclude from (48) that $W'(\lambda_2) \geq 0$ for every λ_2 when $\alpha > 0$ (strictly so wherever $W(\lambda_2) > 0$), with the inequalities reversed for $\alpha < 0$.

Now we return to $S(\alpha)$ in (47). Because $\int \lambda_2 d\gamma_2 = 0$, we must conclude that for $\alpha > 0$ we have $S(\alpha) > 0$ (because $W(\lambda_2)$ is nondecreasing, and nonconstant since it is strictly increasing wherever positive — as it is for λ_2 near the top of the support) and for $\alpha < 0$ we have $S(\alpha) < 0$ (by the mirror-image argument).

Part (c). Assume $\alpha > 0$ (the arguments for negative α are symmetric). Under our assumptions for this part and the fact that $\int \lambda d\gamma(\lambda) = 0$, the support of λ can be normalized without further loss of generality to $[-1, 1]$. Then γ has density $\frac{1}{2}$ on $[-1, 1]$, and the domain of integration is $\Lambda \equiv [-1, 1] \times [-1, 1]$. Let $k > 0$ be the normalizing constant in the optimal contract under standard implementation. Then

$$(49) \quad \frac{4}{1 + \alpha} S(\alpha) = \int_{\Lambda} f\left(\frac{\lambda_1 + \alpha\lambda_2}{k}\right) \lambda_1 \lambda_2 d\lambda_1 d\lambda_2.$$

Because f is increasing, with $f(z) = 0$ for all $z \leq 0$, it can be represented as a “layered” combination of threshold functions:

$$(50) \quad f(z) = \int_{(0,\infty)} \mathbb{1}\{z \geq \tau\} d\nu(\tau) \text{ for every } z,$$

where ν is the Lebesgue-Stieltjes measure induced by f on $(0, \infty)$.¹⁷ Using (50) in (49) and applying Fubini’s theorem,¹⁸

$$(51) \quad \frac{4}{1+\alpha} S(\alpha) = \int_{(0,\infty)} I(k\tau) d\nu(\tau), \text{ where } I(s) \equiv \int_{\{\lambda \in \Lambda: \lambda_1 + \alpha\lambda_2 \geq s\}} \lambda_1 \lambda_2 d\lambda_1 d\lambda_2.$$

The region of integration in $I(s)$ is depicted in Figure 6: it is the portion of Λ that lies beyond a line parallel to the boundary $\lambda_1 + \alpha\lambda_2 = 0$, displaced by the threshold s (the shaded region in the figure). *We claim that $I(s) > 0$ for every $s \in (0, 1 + \alpha)$.* Because f is strictly increasing on $(0, \infty)$, ν assigns strictly positive mass to $(0, (1 + \alpha)/k)$, so the claim establishes $S(\alpha) > 0$ via (51), as desired.

To prove the claim, fix $s \in (0, 1 + \alpha)$ and integrate out λ_1 :

$$I(s) = \int_{\underline{\lambda}}^1 \frac{1 - (s - \alpha\lambda_2)^2}{2} \lambda_2 d\lambda_2,$$

where $\underline{\lambda} \equiv \max\{-1, (s - 1)/\alpha\}$ marks the smallest value of λ_2 for which the inner region is nonempty. If $s \leq 1 - \alpha$, then $\underline{\lambda} = -1$, and direct integration (all even powers of λ_2 integrate to zero over $[-1, 1]$) yields $I(s) = \frac{2}{3}\alpha s > 0$. If $s > 1 - \alpha$, then $\underline{\lambda} = (s - 1)/\alpha$, and the substitution $w = s - \alpha\lambda_2$ gives

$$I(s) = \frac{1}{2\alpha^2} \int_{s-\alpha}^1 (s - w)(1 - w^2) dw.$$

If $s \geq 1$, every w in the domain of integration satisfies $s - w > 0$, while $1 - w^2 \geq 0$, so $I(s) > 0$. If $1 - \alpha < s < 1$, pair each $w = s + \delta \in (s, 1]$ with its mirror image $w = s - \delta$; both lie in the domain of integration, because $s - \delta \geq 2s - 1 \geq s - \alpha$. Each such pair contributes $\delta \{[1 - (s - \delta)^2] - [1 - (s + \delta)^2]\} = 4s\delta^2 > 0$ to the integral, and the residual unpaired interval $[s - \alpha, 2s - 1)$ also contributes positively, because $s - w > 0$ and $1 - w^2 \geq 0$ there. Hence $I(s) > 0$ in all cases, which establishes the claim and completes the proof of part (c).

¹⁷With differentiable f , $d\nu(\tau) = f'(\tau)d\tau$, and (50) is just the fundamental theorem of calculus combined with $f(0) = 0$.

¹⁸The relevant double integral is absolutely convergent; it is bounded by $4\nu((0, (1 + \alpha)/k]) = 4f((1 + \alpha)/k) < \infty$.

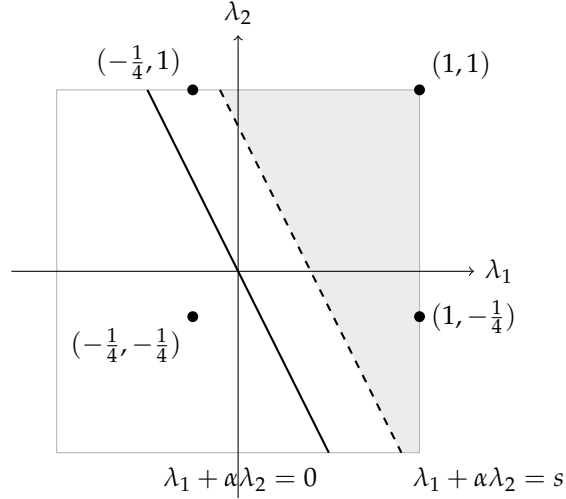


Figure 6. The domain of integration. The solid line is the boundary $\lambda_1 + \alpha\lambda_2 = 0$, drawn for $\alpha = \frac{1}{2}$; the shaded region beyond the dashed line is the region of integration for $I(s)$ in the proof of part (c) of Proposition 3. The four dots mark the possible realizations of (λ_1, λ_2) in the example of Section 6.2.

(λ_1, λ_2) -values	(1, 1)	(1, -1/4)	(-1/4, 1)	(-1/4, -1/4)
Probability	1/25	4/25	4/25	16/25
$\Psi = \lambda_1 + \alpha\lambda_2$	3/2	7/8	1/4	-3/8
$m = (\max\{\Psi, 0\}/k)^{1/\theta}$	1.569	0.862	0.214	0
$f\left(\frac{\lambda_1 + \alpha\lambda_2}{k}\right)$	1.046	0.985	0.8572	0
$f\left(\frac{\lambda_1 + \alpha\lambda_2}{k}\right) - f\left(\frac{\lambda_1}{k}\right)$	0.046	-0.015	0.8572	0
$[f\left(\frac{\lambda_1 + \alpha\lambda_2}{k}\right) - f\left(\frac{\lambda_1}{k}\right)]\lambda_1\lambda_2$	0.05	0.004	-0.21	0

Table 4. Information for the Example in Section 6.2.

Details for the Example in Section 6.2. Note that under high effort, λ takes on two distinct values: $\lambda(0) = -\frac{1}{4}$ and $\lambda(1) = 1$, with probabilities $\frac{4}{5}$ and $\frac{1}{5}$ respectively. That yields four potential outcomes within the Λ -space of a two-agent team, as illustrated in Figure 6. Because λ_1 and λ_2 are independent and $\int \lambda_2 d\gamma_2 = 0$, the term $f(\lambda_1/k) \lambda_1 \lambda_2$ integrates to zero, so that $S(\alpha) = (1 + \alpha) \int_{\Lambda} \left[f\left(\frac{\lambda_1 + \alpha\lambda_2}{k}\right) - f\left(\frac{\lambda_1}{k}\right) \right] \lambda_1 \lambda_2 d\gamma$, which is the *centered* form recorded in the bottom rows of Table 4. Two of these outcomes — (1, 1) and (1, -1/4) — make positive contributions to this centered representation of $S(\alpha)$, whereas the outcome (-1/4, 1) decreases its value (the fourth outcome, (-1/4, -1/4), contributes zero). Table 4 presents the probabilities, payments, payoffs, and the resulting contributions to $S(\alpha)$ for each of the four potential realizations. Although the two positively contributing outcomes

yield higher utility $f\left(\frac{\lambda_1 + \alpha\lambda_2}{k}\right)$ compared to $(-1/4, 1)$, owing to the strict concavity of the realized utility in Ψ , their absolute deviations from $f\left(\frac{\lambda_1}{k}\right)$ are smaller. Integrating the term $[f\left(\frac{\lambda_1 + \alpha\lambda_2}{k}\right) - f\left(\frac{\lambda_1}{k}\right)]\lambda_1\lambda_2$ over the three relevant realizations and multiplying by $(1 + \alpha)$ results in the negative value $S(\alpha) \approx -0.048$ when $\alpha = 0.5$.

Proof of Proposition 4. Fix some contract m . For each index j , define a function σ by

$$\sigma(\lambda_j) \equiv \int_{\Lambda^{n-1}} u(m(\boldsymbol{\lambda})) d\gamma_{-j}$$

describing the expected felicity of agent 1 as a function of λ_j for any given $j \neq 1$, after integrating out over $\boldsymbol{\lambda}_{-j}$. (By Observation 2, σ has the same functional form for all j , though our arguments do not rely on that information.) We claim that if m is an optimal contract, then for every $j \neq 1$,

$$(52) \quad \int_{\Lambda} \lambda_j \sigma(\lambda_j) d\gamma_j \geq 0 \text{ as } \alpha \geq 0.$$

To establish (52), suppose first that $\alpha > 0$. By Proposition 1, σ is nondecreasing, and it is strictly increasing for every $\lambda = \lambda_j$ such that $\Psi(\lambda_j, \boldsymbol{\lambda}_{-j}) > 0$ under a positive $d\gamma_{-j}$ measure of $\boldsymbol{\lambda}_{-j}$ realizations. (In such cases the payment to agent 1 is sensitive to λ_j .) Because these conditions always hold for a positive γ -measure of λ_j , we conclude that σ is nondecreasing, and it is *strictly* increasing for a positive γ -measure of λ_j . Using the fact that $\int_{\Lambda} \lambda d\gamma(\lambda) = 0$, we must conclude that condition (52) must hold for $\alpha > 0$. A parallel argument with σ decreasing applies when $\alpha < 0$.

To complete the proof, recall the incentive constraint from (10) and (11):

$$\int_{\Lambda} \left[\lambda_1 u(m(\boldsymbol{\lambda})) + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq 1} \lambda_j u(m(\boldsymbol{\lambda})) \right] d\boldsymbol{\gamma} \geq c.$$

Because outputs are independent, we can use σ to rewrite this inequality as

$$(53) \quad \int_{\Lambda} \lambda_1 u(m(\boldsymbol{\lambda})) d\boldsymbol{\gamma} + \frac{\alpha}{1 - \alpha(n-2)} \sum_{j \neq 1} \int_{\Lambda} \lambda_j \sigma(\lambda_j) d\gamma_j \geq c.$$

Now suppose that $\alpha > 0$. Increase its value to $\alpha' \in (\alpha, 1/(n-1))$. It is obvious that

$$\frac{\alpha'}{1 - \alpha'(n-2)} > \frac{\alpha}{1 - \alpha(n-2)}.$$

Combining this information with (52), so that $\sum_{j \neq 1} \int_{\Lambda} \lambda_j \sigma(\lambda_j) d\gamma_j > 0$, we must conclude that if m is optimal under α , then

$$(54) \quad \int_{\Lambda} \lambda_1 u(m(\boldsymbol{\lambda})) d\boldsymbol{\gamma} + \frac{\alpha'}{1 - \alpha'(n-2)} \sum_{j \neq 1} \int_{\Lambda} \lambda_j \sigma(\lambda_j) d\gamma_j > c,$$

which means that the earlier contract is now “strictly feasible” under α' . Because the incentive constraint is always binding at the optimum, this must mean that the principal can strictly gain by adjusting the contract. A parallel argument holds when $\alpha < 0$. If $\alpha' < \alpha$ (with $|\alpha'|(n-1) < 1$), then

$$\frac{\alpha'}{1 - \alpha'(n-2)} < \frac{\alpha}{1 - \alpha(n-2)}.$$

Because (52) holds with negative sign, (54) holds again, completing the proof.