Team Formation with Altruism

VAKHTANG ABASHIDZE*

This paper studies the strategic formation of teams within organizations with moral hazard, focusing on the impact of agents' altruism on incentive structures. The study contrasts teams composed of agents with similar preferences with teams that integrate diverse preferences. The analysis demonstrates that, for N identical projects and $2N$ agents, the principal's optimal strategy is to form anti-assortative teams: pairing the most and least altruistic agents, and iterating on the remaining agents. The analysis implies that optimal assignment balances altruism across teams, minimizing expected incentive payments.

KEYWORDS. Team formation, Other-regarding preferences, Moral hazard in teams, Contract design, Teamwork.

1. INTRODUCTION

The composition of teams is a central topic in economics and organizational research, influencing cooperation, incentives, and performance. Foundational works by Marschak and Radner (1972) and Alchian and Demsetz (1972) established models of team production. In recent years, substantial research has examined the composition of teams, focusing on how diversity and implicit incentives affect cooperation, performance, and managerial outcomes.[1]

Over the past three decades, experimental research has provided robust evidence that individuals exhibit other-regarding preferences and do not act solely out of self-interest, particularly when their decisions affect others.[2] Complementing these findings, studies in organizational and field settings show that such preferences play a significant role within firms, shaping worker behaviour, incentive provision, and firm performance.[3]

*Humboldt University of Berlin, Spandauer Str. 1, 10178 Berlin, Germany: Contact: vakhtang.abashidze@hu-berlin.de

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 403041268 – TRR 266 Accounting for Transparency. I would like to thank my supervisors, Anja Schöttner and Roland Strausz. I am also grateful to Helmut Bester, Harvey Upton, Tobias Gamp, Sebastian Schweighofer-Kodritsch, Victor Schauer, and Christian Lukas for valuable comments and suggestions, as well as to audiences at the HU Berlin Microeconomic Theory Colloquium, the TRR 266 Annual Conference 2024, and the GEABA Conference 2024.

[1] See Glover and Kim (2021), Imhof and Kräkel (2023), Dong et al. (2024), Dong and Mayskaya (2024), Blanco-Fernández et al. (2025).

[2] See Levine (1998), Fehr and Schmidt (1999), Abbink et al. (2000), Dufwenberg and Kirchsteiger (2000), Andreoni and Miller (2002).

[3] See Bandiera et al. (2005), Weber and Hertel (2007), Hamilton et al. (2012), Hattori and Yamada (2025).

Among the various forms of other-regarding preferences, altruism is commonly associated with Organizational Citizenship Behavior.[4] It is especially relevant for organizational environments with incentive provision, as altruistic agents derive utility from their coworkers' monetary payoffs. In team-based production settings characterized by moral hazard, this feature implies that incentives offered to one agent affect not only that agent's effort choice but also the incentive requirements of coworkers, thereby directly linking team composition to optimal contract design.

Given this direct link between altruism, incentives, and team composition, managers should account for these preferences when forming teams and designing incentive contracts. In many organizational environments—such as consulting, engineering, and software companies—managers are responsible for forming teams to complete projects while individual employee effort remains unobservable. This setting naturally gives rise to the following research question: How should a principal optimally form teams and design incentives when agents differ in their other-regarding preferences, particularly in their degree of altruism?

To study this question, consider a firm in which a manager must assemble two teams, each consisting of two agents, from a pool of four agents. Agents' effort choices are unobservable, leading to a moral hazard problem for the manager. To identify the precise effect of altruistic preferences, we assume that the agents are identical except for altruistic preferences.

These differences in altruism directly shape agents' incentive requirements. Their altruistic preferences affect agents' utility functions: highly altruistic agents are willing to exert effort for lower incentives than less altruistic agents, as they place greater weight on the team's success and well-being. The manager strategically forms two teams with the aim of minimizing the incentive costs needed to implement high effort. By leveraging heterogeneity in agents' altruistic preferences, the manager seeks to improve financial efficiency while maintaining a productive work environment.

Strategic team formation can be categorized into two principal approaches: homogeneous and heterogeneous team configurations. Homogeneous matching concentrates incentive costs: pairing two low-altruism agents requires high bonuses, whereas pairing two high-altruism agents may allow bonuses to be very low. This occurs because the willingness of highly altruistic agents to exert effort for lower bonuses in favor of team success significantly reduces the overall compensation costs of their team, whereas the team composed of less altruistic agents concentrates higher incentive requirements by grouping them together. In contrast, the heterogeneous (anti-assortative) pairing approach explores whether the tendency of a highly altruistic agent to prioritize collective benefits and work for lower bonuses can counterbalance the financial strain generated by the higher incentive demands of a less altruistic agent. This approach introduces an incentive structure in which compensation extremes may be moderated, thereby offering an intermediate outcome that can lead to more efficient financial results.

[4] See Barghouti et al. (2023), Rioux and Penner (2001), Wagner and Rush (2000)

Building on this intuition, the present study formally analyzes the principal's optimal team formation problem under moral hazard. The principal's goal is to maximize expected profit by exploiting heterogeneity in agents' altruistic preferences. Accordingly, it examines the economic and behavioural dynamics of all potential team formations and assesses their implications for wage management and financial efficiency. By integrating principles of behavioural economics into team formation, this study provides insights into optimizing human resource management within typical organizational settings.

Extending the intuition outlined above, we show that when a principal faces two identical projects, the optimal strategy is to form heterogeneous teams. The key driver of this result is the asymmetric response of incentives to altruistic preferences: *Ceteris paribus,* altruistic agents derive greater utility from their coworkers' bonuses, reducing their need for incentives. In contrast, the least altruistic agents derive only a small positive effect from their coworkers' rewards and, therefore, require higher bonuses. Pairing these two types creates incentive-based complementarities: the most altruistic agents are willing to exert effort for lower bonuses when their teammates are well-compensated, while the least altruistic agents require high bonuses despite receiving little utility from coworkers' bonuses. This preference alignment reduces total incentive costs, making anti-assortative team formation efficient under moral hazard.

Consequently, when projects are identical, the principal can focus exclusively on how agents' altruistic preferences interact in team formation to achieve optimal outcomes. We establish that this result generalizes to environments with N identical projects and $2N$ agents, in which the principal must form N teams.

In the discussion, we examine the impact of self-organized team formation—where agents choose teammates before contracts are set—on the principal's strategy. By analyzing this alternative timing, we illustrate the robustness of the optimal team formation results in decentralized and collaborative settings characterized by altruistic preferences. Importantly, there is no conflict between the agents' and the principal's interests regarding team formation.

Next, we discuss optimal agent selection. Managers often face the challenge of selecting a subset of employees for specific tasks or projects due to budget or resource constraints. This common organizational problem involves choosing the most suitable combination of workers from a limited pool to optimize project outcomes, and we show that it is optimal to assemble a team with the most altruistic agents from the pool. Another aspect explored in the paper is the presence of non-zero outside options for agents, where we demonstrate that the main results remain robust regardless of the nature of these outside options. Next, we analyze the uniqueness of the equilibrium and show that the results remain robust when the principal enforces a unique equilibrium outcome.

We further show that the results continue to hold when agents may exhibit not only altruistic but also spiteful preferences. Finally, we discuss the fairness of the results within the organization and provide several possible extensions for future research.

2. Related Literature

Our research is related to the literature on teamwork, free-riding, the impact of other-regarding preferences on team effectiveness, and strategies for optimal team formation among diverse agents.

First, our work contributes to the growing literature on team formation and diversity. Glover and Kim (2021) show that heterogeneous teams strengthen relational incentives, while Imhof and Kräkel (2023) analyze how diversity interacts with promotion schemes to shape performance. Dong et al. (2024) examine how communication costs influence diversity in team formation, and Dong and Mayskaya (2024) investigate how team objectives determine optimal cognitive diversity. Franco et al. (2011) highlight that team incentive problems can lead to monotone matching and that negative sorting may arise even in the presence of strong production complementarities. Bel et al. (2015) study how to optimally combine familiar (incumbent) and unfamiliar (newcomer) workers when project outcomes depend on team production. Finally, Blanco-Fernández et al. (2025) use agent-based modeling to explore how dynamic team composition and coordination choices affect performance. Similar to this literature, we consider heterogeneous agents with respect to altruistic preferences and show that anti-assortative matching can be optimal.

Second, our paper relates to the literature on other-regarding preferences and teamwork. After establishing the importance of other-regarding preferences in decision making,[5] subsequent studies theoretically examine their role in teamwork, including Itoh (2004), Rey-Biel (2008), Bartling (2011), Livio and De Chiara (2019), and Sarkisian (2021). This literature advances the analysis of optimal incentives within teams characterized by heterogeneous other-regarding preferences. These studies focus on inequity aversion and reciprocity, offering insights into how principals could take advantage of these preferences to structure more effective incentive schemes. Unlike our approach, which models team formation, this body of work takes teams as given and highlights the complexity of agent interactions, as well as the potential for principals to strategically leverage other-regarding preferences to enhance organizational outcomes.

Demougin and Fluet (2003) explore how workers' envy affects the cost-effectiveness of group versus individual bonus schemes. They observe that the choice between the two depends on the agents' outside options and available performance measures. In our analysis, we investigate how variations in agents' outside options impact the principal's strategy, thereby examining the robustness of our results when this assumption is relaxed.

Daske (2017) is closely related to our work. The paper shows that successful staffing requires knowledge of interpersonal relationships and that no universal mechanism guarantees efficiency for team formation. In Daske (2017), similar to us, the author considers static settings and looks for optimal team formation, given that four identical agents have different other-regarding preferences. However, our approach diverges significantly regarding the preference dynamics between agents. Daske's model allows

[5] See Levine (1998), Fehr and Schmidt (1999), Dufwenberg and Kirchsteiger (2000), Abbink et al. (2000), Andreoni and Miller (2002), Engelmann and Strobel (2004), Bandiera et al. (2005), Hamilton et al. (2012)

agents to have distinct preferences toward various coworkers, yet maintains consistent other-regarding preferences between any two specific agents. Consequently, within Daske's framework, agents cannot differ in their other-regarding preferences toward their teammates: while preferences may vary across outsiders, within-team preferences are assumed to be identical.

Another notable distinction lies in the analytical treatment of the cost function. Daske's outcomes are influenced by the convexity of the marginal cost function, particularly the behaviour of its third-degree derivative. In contrast, our study simplifies this aspect by assuming a constant cost of effort, eliminating the influence of higher-order derivatives on our results. This difference in handling the cost function introduces a distinct methodological divergence between the two studies.

Existing work on other-regarding preferences in teams typically takes team composition as given, while work on team formation under moral hazard often abstracts from heterogeneity in other-regarding preferences. We study a setting where a principal designs incentives and chooses the assignment when agents differ in their other-regarding preferences. Under identical projects, the cost-minimizing assignment is anti-assortative matching.

The structure of the paper is as follows. Section 3 introduces the theoretical model and its assumptions. Section 4 presents the main results. Section 5 examines the robustness of these findings and discusses possible extensions. Section 6 concludes. Appendix A contains the proofs of the main results, and Appendix B provides the proofs for the discussion.

3. Model

3.1 *Setup*

Consider a principal–multi-agent model in which agents exhibit altruistic preferences. The principal seeks to implement two projects, indexed by $n \in \{1,2\}$, each requiring two agents' effort, and either succeeds or fails. Upon success, project n generates a revenue of $R_n > 0$.

The set of agents is denoted by $A = \{a,b,c,d\}$, agent $i \in A$ chooses effort $e_i \in \{0,1\}$, where $e_i = 1$ means work and $e_i = 0$ means shirk. Exerting effort entails a cost $k > 0$, identical across agents, whereas shirking is costless. The agent's effort is not observable to the principal.

The probability of success of project n is $p^n : \{0,1\}^2 \to [0,1]$; let $p^n_{e_i e_j} := p^n(e_i, e_j)$. The probability function of each project increases with the effort of the agents, thus satisfying the inequality $1 > p^n_{11} > p^n_{01} = p^n_{10} > p^n_{00} \geq 0$. We assume that all agents have identical productivity and that the revenue from each project is sufficient for the principal to prefer high effort over low effort.[6]

For both projects, the principal will optimally select team formation $(\{i,j\}\{g,h\})$ from all potential team formations T,

$$T = \{(\{a,b\}\{c,d\}), (\{a,c\}\{b,d\}), (\{a,d\}\{b,c\}), (\{c,d\}\{a,b\}), (\{b,d\}\{a,c\}), (\{b,c\}\{a,d\})\}.$$

[6] Implementing $(work, work)$ via the minimal-cost contract yields a higher expected profit than any low-effort outcome.

In $(\{i,j\}\{g,h\})$, $\{i,j\}$ represents the first project team and $\{g,h\}$ represents the second project team, if projects are identical, then the matchings $\{i,j\}\{g,h\}$ and $\{g,h\}\{i,j\}$ are equivalent.

It is assumed that all parties involved are risk-neutral. Additionally, all agents are subject to limited liability, such that the principal cannot impose negative transfers. The outside options of the principal and the agents are normalized to zero.[7]

The principal's *ex-ante* payoff from each project is:

$$\pi_n(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, e_i, e_j) = p_{e_i e_j}^n (R_n - b_{ij}^n - b_{ji}^n) - w_{ij}^n - w_{ji}^n \tag{1}$$

(w_{ij}^n, b_{ij}^n) is offered to agent i working on project n with agent j by the principal; w_{ij}^n represents the fixed wage and b_{ij}^n the bonus.

Agent i's *ex-ante* utility on project n is:

$$u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, e_i, e_j) = v_{ij}(w_{ij}^n, b_{ij}^n) - e_i k + \alpha_i v_{ji}(w_{ji}^n, b_{ji}^n) \tag{2}$$

$v_{ij}(w_{ij}^n, b_{ij}^n) = w_{ij}^n + p_{e_i e_j}^n b_{ij}^n$ is the expected monetary payoff of agent i. In agent i's expected utility, $v_{ji}(w_{ji}^n, b_{ji}^n)$ represents the expected monetary payoff of the other agent working on project n. $\alpha_i \in [0,1)$ expresses agent i's degree of altruism.[8] α_i is *common knowledge*.[9]

If α_i is positive, agent i is altruistic and the agent's utility increases with the coworker's expected monetary payoff; if α_i is zero, agent i is selfish. The restriction $0 \leq \alpha_i < 1$ ensures that agent i places more weight on the agent's direct payoff compared to the payoff of the coworker.

The only difference among agents lies in their altruistic preferences, which are fixed and independent of the identity of their coworkers. Without loss of generality, the agents are ordered according to their preference parameters as follows: $\alpha_a \leq \alpha_b \leq \alpha_c \leq \alpha_d$ so agent a is least altruistic, agent d most altruistic.

The sequence of moves in the game proceeds as follows: First, the principal forms teams and offers contracts to the agents. Then, agents observe their own and their coworker's contracts. If all agents accept the contracts, they simultaneously choose their effort levels. Next, the outcomes of the projects are observed, and the agents receive payments according to their contracts. If a contract is rejected, the principal and the agents resort to their outside options.

In the game, we seek a subgame perfect Nash equilibrium where agents accept the contract, and in each project, both agents choose to work.

3.2 *Principal's problem*

Given that the revenue from each project is sufficient for the principal to prefer high effort over low effort, the principal's objective is to minimize total expenditures over group

[7] Section 5.3 relaxes this normalization by permitting a nonzero common outside option.

[8] Section 5.6 extends to $\alpha_i \in (-1,1)$.

[9] If altruistic preferences were based on the coworker's net material payoff — that is, their total monetary gain minus their costs — the results of the paper would remain unchanged.

formation and the wage schedule. This is subject to the agents' incentive compatibility, participation, and limited liability constraints.

$$\min_{\substack{(\{i,j\}\{g,h\})\in T \\ w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n \in R}} \quad \sum_{n=1}^{2} p_{11}^n (b_{ij}^n + b_{ji}^n) + w_{ij}^n + w_{ji}^n$$

$$\text{s.t. } u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, 1, 1) \geq u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, 0, 1), \forall i \in A, n \in \{1,2\}, \textbf{ (IC)}$$

$$u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, 1, 1) \geq 0, \forall i \in A, n \in \{1,2\}, \textbf{ (PC)}$$

$$w_{ij}^n + b_{ij}^n \geq 0, w_{ij}^n \geq 0, \forall i \in A, n \in \{1,2\}, \textbf{ (LLC)}$$

The incentive compatibility constraint requires that agent i prefers to work rather than shirk, given that the other agent works. The participation constraint indicates that an agent i's expected utility must be at least as high as the agent's outside option. This means that agent i should prefer to accept the contract, assuming that both agents exert effort. The limited liability constraint states that the agent's ex-post payoff must be nonnegative.

The optimal team formation is defined as the selection of team formation and contracts that are designed to minimize the total expected compensation costs associated with implementing high effort by all agents.

4. Main Results

4.1 *Analysis*

We solve the optimization problem in two steps. First, for any fixed team formation, we characterize the cost-minimizing wage contract that implements high effort. Second, given these optimal contracts, we compare total compensation costs across team formations and select the cost-minimizing assignment.

On project n, when the wage schedule of agent i meets the incentive constraint, it implies that if the coworker works, agent i prefers to work rather than shirk. Simplifying agent i's incentive constraint, we obtain:

$$p_{11}^n b_{ij}^n - k + \alpha_i p_{11}^n b_{ji}^n \geq p_{01}^n b_{ij}^n + \alpha_i p_{01}^n b_{ji}^n \tag{3}$$

The term for a fixed wage was omitted because it appears on both sides of the expression. Rearranging yields,

$$b_{ij}^n \geq \frac{k}{p_{11}^n - p_{01}^n} - \alpha_i b_{ji}^n \tag{4}$$

This expression shows that the lowest bonus, which ensures that agent i chooses to work, decreases in α_i. Put differently, the principal prefers higher values of α_i, enabling her to induce agent i to exert effort for a lower incentive.

Note that the incentive constraint is similar for agent j; only subscripts are swapped.

$$b_{ji}^n \geq \frac{k}{p_{11}^n - p_{01}^n} - \alpha_j b_{ij}^n \tag{5}$$

We now relax the problem by dropping (PC) and (LLC), later ensuring that they are satisfied by the optimal solution to the relaxed problem. If the incentive constraint is nonbinding, the principal has the flexibility to reduce the bonus schedule, thereby reducing expected compensation cost. This reduction can continue until the constraint becomes binding, ensuring that the agent still has sufficient incentive to exert the effort. Indeed, suppose that one of the incentive constraints is slack, say for agent j. Then the principal can reduce b_{ji} by $\varepsilon > 0$ without violating (IC_j). This decreases agent i's utility from working by $\alpha_i \varepsilon$. To keep (IC_i) binding, the principal increases b_{ij} by $\alpha_i \varepsilon$. Because $\alpha_i < 1$, the reduction in b_{ji} exceeds the required increase in b_{ij}, so the principal's expected compensation costs reduce by $p_{11}(1-\alpha_i)\varepsilon > 0$. Hence, both incentive constraints (4) and (5) must bind at the optimum. Substituting (5) into (4) yields the optimal bonus for agent i

$$b_{ij}^{n\,*} = \frac{k(1-\alpha_i)}{(p_{11}^n - p_{01}^n)(1-\alpha_i\alpha_j)} \tag{6}$$

Let us denote $b_0^n \coloneqq \frac{k}{(p_{11}^n - p_{01}^n)}$; hence

$$b_{ij}^{n\,*} = b_0^n \frac{(1-\alpha_i)}{(1-\alpha_i\alpha_j)}$$

For any $\alpha_i, \alpha_j \in [0,1)$, the bonus is nonnegative. We know that the project can fail; therefore, the (LLC) constraint will be satisfied if and only if w_{ij}^n is weakly positive. Hence, the contract in which the fixed wage is equal to zero, $(w_{ij}^n, b_{ij}^n) = (0, b_{ij}^{n\,*})$, satisfies the limited liability constraint.[10]

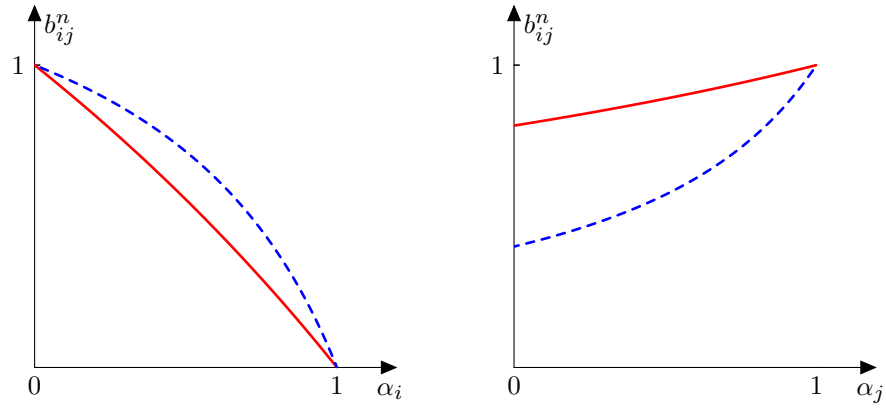
LEMMA 1. *For given pair* $\{i,j\}$, $(w_{ij}^n, b_{ij}^n) = (0, b_{ij}^{n\,*})$ *is the optimal contract. It satisfies incentive compatibility, limited liability, and participation constraints.*

Detailed proofs of all lemmas and propositions are provided in the appendix.

4.2 *Comparative statics*

First, observe that the optimal bonus for agent i decreases with his degree of other-regarding preferences, $\frac{\partial b_{ij}^{n\,*}}{\partial \alpha_i} < 0$. An increase in the agent's altruism raises the weight placed on the coworker's wage in the agent's utility function, thereby increasing the utility derived from the coworker's monetary compensation. As a consequence, the agent requires lower monetary incentives to exert effort (see Figure 1(a)).

[10] Note that if agents do not exert effort, their expected payoffs are nonnegative under the optimal contract; however, $(work, work)$ is the payoff-dominant equilibrium for the agents. The uniqueness of the equilibrium is addressed in Section 5.4.



(a) Effect of own altruism

(b) Effect of coworker altruism.

FIGURE 1. Notes: In both graphs $b_0^n = 1$. Panel (a) shows how agent i's optimal bonus behaves when agent i's altruistic preference changes while the coworker's is fixed. In the blue dashed curve, $\alpha_j = 0.6$, and in the red curve, $\alpha_j = 0.2$. Panel (b) illustrates how agent i's optimal bonus behaves when agent i's altruistic preference is fixed while the coworker's varies. The blue dashed curve shows the optimal bonus for an agent with $\alpha_i = 0.6$. The red curve represents an agent's optimal bonus when the agent is less altruistic $\alpha_i = 0.2$.

Furthermore, we show that for altruistic agent i, the optimal bonus increases with agent j's degree of altruism, $\frac{\partial b_{ij}^{n\,*}}{\partial \alpha_j} > 0$. Thus, for altruistic agents, the optimal bonus rises as their coworker becomes more altruistic. The intuition is that, when agent i's coworker becomes more altruistic, the coworker requires a lower bonus to exert effort, which reduces the coworker's expected monetary payoff. As a result, the altruistic agent derives less utility from the coworker's monetary payoff. Consequently, to sustain the altruistic agent's effort incentives, the principal must offer a higher bonus (see Figure 1(b)).

To further clarify the relationship between these comparative statics, we highlight the connection between the two panels in Figure 1. In panel (a), agent i consistently requires a lower bonus when matched with a less altruistic coworker, as reflected in the comparative statics discussed in panel (b). Conversely, panel (b) shows that agent i requires a lower bonus when he himself is more altruistic, consistent with the relationship illustrated in panel (a). Together, the two panels emphasize the role of own and coworker altruism in shaping optimal incentive provision.

Altruism affects the optimal contract through two channels: (i) a direct effect, where an agent's utility depends on the coworker's wage, and (ii) an indirect effect, which captures how the interaction between the agents' altruistic preferences influences the incentive structure.

Let us illustrate how these comparative statics highlight two distinct channels through which altruism affects optimal incentive provision. *Ceteris paribus*, an increase in an agent's own altruistic preferences reduces the bonus required to induce effort, as

the agent places greater weight on the coworker's monetary payoff. By contrast, *ceteris paribus*, an increase in the coworker's altruistic preferences affects incentives only indirectly: it lowers the coworker's required bonus and thus the coworker's expected monetary payoff, which reduces the utility that an altruistic agent derives from the coworker's compensation. To maintain effort incentives for such an agent, the principal must therefore increase the bonus offered to the altruistic agent.

4.3 *Identical projects*

For the following analysis, we assume that the success probability functions are identical across projects ($p^1_{e_ie_j} = p^2_{e_ie_j}$); accordingly, we omit the superscript indexing the project.

Two selfish agents and two altruistic agents: $\alpha_a = \alpha_b = 0 < \alpha_c \leq \alpha_d$.

In this case, the principal has to choose between two team-forming options: placing two selfish agents in the same group and two altruistic agents in the other group, or forming mixed teams. There are three possible team formations. However, in this particular case, two of these options coincide, namely $(\{a,c\}\{b,d\})$ and $(\{a,d\}\{b,c\})$ which yield identical optimal contracts because, in both formations, an altruistic agent is matched with a selfish agent.

The total bonus payment under the team formations $(\{a,d\}\{b,c\})$ and $(\{a,c\}\{b,d\})$ is equal and given by:

$$b_{ad} + b_{da} + b_{bc} + b_{cb} = b_0 + b_0(1-\alpha_d) + b_0 + b_0(1-\alpha_c) = 2b_0 + b_0(1-\alpha_d) + b_0(1-\alpha_c)$$

While the total bonus payment for $(\{a,b\},\{c,d\})$ team formation is equal to:

$$b_{ab} + b_{ba} + b_{cd} + b_{dc} = 2b_0 + \frac{b_0(1-\alpha_d)}{1-\alpha_c\alpha_d} + \frac{b_0(1-\alpha_c)}{1-\alpha_c\alpha_d}$$

Since $0 < \alpha_c, \alpha_d < 1$, the total bonus payment for the homogeneous teams $(\{a,b\},\{c,d\})$ is higher compared to the heterogeneous team formations. Thus, we can conclude with the following lemma.

LEMMA 2. *When presented with two selfish and two altruistic agents, the principal's optimal team formation strategy is to form heterogeneous teams.*

This result follows from heterogeneity in agents' utility functions. Altruistic agents benefit from higher coworker wages, whereas selfish agents are unaffected by them. Consequently, altruistic agents require lower bonuses when paired with selfish agents than when paired with other altruistic agents, making such pairings cost-minimizing for the principal.

We now transition from the introductory example to the general model. Accordingly, each of the four agents may exhibit any degree of altruistic preferences. For each agent $i \in A$, the preference parameter α_i can take any value in the interval $[0,1)$, as defined in the setup. We proceed to examine whether the team formation principle identified in Lemma 2 continues to hold when agents' altruistic preferences vary.

In particular, the central implication of this principle is that pairing the most altruistic agent with the least altruistic agent reduces the bonus required for the most altruistic agent.

When all agents are altruistic, the agent with the highest degree of altruism derives the greatest utility from the coworker's compensation. Consequently, this agent is willing to accept the lowest bonus when paired with the least altruistic agent, who in turn requires the highest bonus among all agents when matched with the most altruistic agent. Since the least altruistic agent derives the smallest utility gain from the coworker's compensation, it is efficient to pair this agent with the most altruistic agent, who requires the lowest bonus. This pairing not only minimizes total compensation costs for the team but also exploits the remaining agents' altruistic preferences. Specifically, the remaining agents—being more altruistic than the least altruistic agent—derive greater mutual utility when paired together than if either were matched with the most altruistic agent. Therefore, the optimal strategy is to form one team consisting of the most and the least altruistic agents and another team comprising the remaining two agents.

The following proposition formalizes this intuition.

PROPOSITION 1. *In scenarios involving any four agents with $\alpha_a \leq \alpha_b \leq \alpha_c \leq \alpha_d$, the principal's best strategy is to pair the most altruistic agent with the least altruistic agent. The remaining two agents form the second team.*

Next, we characterize optimal team formation by providing alternative conditions that yield the same prediction. First, we introduce a statistic, which plays a key role in this characterization.

DEFINITION 1. For any team formation $(\{ij\},\{gh\}) \in T$, $\mu_{(\{i,j\}\{g,h\})}$ is the absolute difference between the two teams' average altruism levels, thus

$$\mu_{(\{i,j\}\{g,h\})} \equiv \left|\frac{\alpha_i + \alpha_j}{2} - \frac{\alpha_g + \alpha_h}{2}\right|.$$

The statistic μ measures how unevenly altruism is distributed across the two teams: a smaller μ means the teams are more balanced in terms of altruism.

PROPOSITION 2. *In scenarios involving any four agents with $0 \leq \alpha_a, \alpha_b, \alpha_c, \alpha_d < 1$, the principal's optimal team formation minimizes the absolute difference between the two teams' average altruism levels.*

Intuition — Lemma 1 implies that the minimum bonuses required to implement high effort for a pair $\{i,j\}$ depend on both α_i and α_j, because each agent's incentive constraint includes the utility effect of the coworker's compensation. Moreover, an increase in altruism reduces incentive costs *more* when it is paired with a low-altruism coworker than when the coworker's altruism is already high. Hence total expected wage costs are minimized by balancing altruism across pairs (minimizing μ), which yields anti-assortative matching.

Appendix A shows that among two-team partitions, the cost ordering coincides with the ordering induced by μ.

A natural and intuitive question arises concerning the principal's optimal strategy in a setting with $N \in \mathbb{N}$ projects and $2N$ agents. Specifically, how should the principal efficiently structure team formation across projects while simultaneously determining optimal contract offers for the agents? Addressing this question requires a systematic analysis of team composition strategies that accounts for agents' heterogeneous preferences and incentive constraints. Building on the result of Proposition 1, our objective is to derive strategic insights that characterize optimal team formation and contract allocation in environments with N projects.

PROPOSITION 3. *When the principal faces N identical projects and $2N$ agents—so that each project is assigned to two agents—the principal's optimal strategy is as follows:*

1. *Pair the least altruistic agent with the most altruistic agent. If agents do not differ in their altruistic preferences, pair agents arbitrarily.*
2. *For the remaining agents, repeat step 1 until all agents are paired.*

Proof sketch: Order agents by their degree of altruism. Suppose, for contradiction, that an optimal team formation is not anti-assortative. Then there exist two teams forming a non-extreme pair[11] among four agents, in which the least altruistic agent is matched with an intermediate agent while the most altruistic agent is matched with another intermediate agent. By Proposition 1, swapping agents across these two teams so as to pair the least and most altruistic agents (weakly) reduces the principal's total incentive cost (ties arise only when some α's coincide). This contradicts the optimality of the original team formation. Hence, anti-assortative matching is optimal.

Note that when α's coincide, multiple optimal matchings exist; all satisfy anti-assortative pairing up to permutation.

5. DISCUSSION

We now examine whether the optimal team formation result is sensitive to institutional features commonly observed in organizations. First, we examine whether there is a conflict between the optimal team formation of the principal, aimed at minimizing incentive costs, and the preferences of the agents regarding the composition of the team. We show that, under our assumptions, the agents' team choices align with the principal's cost-minimizing configuration, and thus no conflict emerges. Next, we consider how the principal should optimally form a team when only three agents are available by identifying which agent should be excluded to achieve the most efficient outcome.

We further investigate the robustness of our results along several additional dimensions. We then explore the effect of varying outside options, assuming that all agents have the same outside option, which may nevertheless take arbitrary values. This allows

[11] A pairing contains a non-extreme pair if there exist agent indices $g < h < i < j$ for which the matching does not assign agents g to j and agents h to i.

us to assess how external opportunity costs influence the structure of optimal contracts. We also analyze the uniqueness of equilibrium and confirm that our results hold under conditions that eliminate multiple equilibria. Finally, we revisit the assumption that each team consists of two members, question its optimality, and provide support for this structure within our framework. In addition, we extend our analysis to an environment in which agents may exhibit not only altruistic and selfish preferences but also spiteful preferences, and find that the main results continue to hold in that setting. Taken together, these extensions broaden the scope of the analysis and confirm that the main results remain robust.

5.1 *Stability*

In real-world environments, particularly in decentralized organizations or collaborative settings, agents often form teams before formal contracts or incentive schemes are determined. This endogenous team formation may influence the overall efficiency of the principal's strategy, especially when agents exhibit altruistic preferences. To assess whether the optimal team formation identified in the results remains valid under an alternative timing structure, we consider a scenario where agents select their teammates before the principal designs contracts. This allows us to test the robustness of the main results to institutional variations in the sequencing of decisions.

We consider a setting with two identical projects and four agents. While the underlying model remains unchanged, the timing is modified as follows:

First, agents form teams, and after observing these team formations, the principal offers them contracts. The agents then review the contract terms. If the contract is accepted, the agents work simultaneously on the project and are compensated based on the project's success. If the contract is rejected, both the agents and the principal revert to their respective outside options.

Therefore, the principal's minimization problem for a given team formation becomes:

$$\min_{w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n \in R} \quad \sum_{n=1}^{2} p_{11}^n (b_{ij}^n + b_{ji}^n) + w_{ij}^n + w_{ji}^n$$

$$\text{s.t. } u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, 1, 1) \geq u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, 0, 1), \forall i \in A, n \in \{1,2\}, \textbf{ (IC)}$$

$$u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, 1, 1) \geq 0, \forall i \in A, n \in \{1,2\}, \textbf{ (PC)}$$

$$w_{ij}^n + b_{ij}^n \geq 0, w_{ij}^n \geq 0, \forall i \in A, n \in \{1,2\}, \textbf{ (LLC)}$$

Agents anticipate that the principal will offer a contract derived from binding (IC) and (LLC) constraints that satisfy PC constraints. Given the results in the subsection Analysis, agent i, paired with agent j, will receive the contract offer $(w_{ij}^*, b_{ij}^*) = (0, b_0^n \frac{(1-\alpha_i)}{(1-\alpha_i\alpha_j)})$. Therefore, the expected utility of the agent is

$$u_i(w_{ij}^*, w_{ji}^*, b_{ij}^*, b_{ji}^*, 1, 1) = p_{11} b_0^n \frac{(1-\alpha_i)}{(1-\alpha_i\alpha_j)} - k + \alpha_i p_{11} b_0^n \frac{(1-\alpha_j)}{(1-\alpha_i\alpha_j)} = p_{11} b_0^n - k \qquad (7)$$

Note that the equation (7) implies that agents' expected utilities are independent of team composition. Hence, agents are indifferent over coworkers, and any team formation can arise in equilibrium; therefore, the principal can implement her preferred team formation via a simple coordination mechanism (e.g., a public recommendation/assignment device that agents follow when indifferent).

This outcome is derived from the design of the contracts so that each agent's expected payoff is equal to that of a purely self-interested agent. This ensures that the expected utility of an agent is directly related to the agent's contribution to project success. Consequently, under the cost-minimizing contracts characterized above, equilibrium expected utility does not vary with agents' altruism level or teammate identity.[12]

In the case where the principal offers contracts before agents form teams, the same outcome can be achieved by designing a tailored contract schedule for each agent, contingent on their prospective team composition. Under this scheme, both the agent and their teammate receive bonus payments that reflect their own and their partner's other-regarding preferences. Specifically, the principal provides each agent with a contract schedule comprising three distinct bonus levels, one for each potential pairing, given that there are four agents in total and each agent has three possible teammates.

Formally, the principal first announces a menu of contracts indexed by partner identity; after teams are formed, the corresponding contract is implemented without renegotiation.

Therefore, there is no conflict between the agents' and the principal's interests with respect to team formation. Under the cost-minimizing contracts, agents are indifferent among potential teammates in equilibrium.

5.2 *Which agent should be excluded from team formation?*

Situations where a manager must select a subset of employees for a particular task or project are common in organizational settings. For instance, consider a project manager at a consulting firm who has three equally qualified agents available for an assignment, but can only select two due to budget or project constraints. The manager must decide which two agents to include and which to exclude from the project. Such situations frequently occur in practice when making staffing decisions, which involve selecting the most suitable combination of workers from a limited pool.

We now adjust our model to the scenario. Consider a case in which the principal has to select two out of three agents $\{a,b,c\}$. Agents are ordered with their altruistic preferences $0 \leq \alpha_a \leq \alpha_b \leq \alpha_c < 1$. Therefore, the principal has to choose among three possible team formations $T = \{\{a,b\},\{a,c\},\{b,c\}\}$ for the given optimal incentive contracts, derived in (6). Thus, the expected total cost for the (ij) team is

$$TC_{\{i,j\}} = b_0\big(\frac{1-\alpha_i}{1-\alpha_i\alpha_j} + \frac{1-\alpha_j}{1-\alpha_i\alpha_j}\big) = b_0\big(\frac{2-\alpha_i-\alpha_j}{1-\alpha_i\alpha_j}\big)$$

[12] Note that the result is not driven by the specific form of agents' altruistic preferences: it continues to hold when altruism depends on the teammate's identity.

Let us examine the derivative of the total cost with respect to a given variable α_i (note that the expression of the derivative is the same if instead we look at another variable).

$$\frac{\partial TC_{\{i,j\}}}{\partial \alpha_i} = -b_0 \frac{1 - 2\alpha_j + \alpha_j^2}{(1 - \alpha_i \alpha_j)^2} = -b_0 \frac{(1 - \alpha_j)^2}{(1 - \alpha_i \alpha_j)^2} < 0 \tag{8}$$

The total cost of the team is decreasing in any agent's altruistic preferences. Therefore, we can conclude that when we need to form a team of two members, it is optimal to select the two most altruistic agents. The number of available members from which we choose the agents is irrelevant.

This result follows from the observation that, in any team formation, the principal prefers to include the most altruistic agent, as this agent is willing to exert effort for the lowest bonus. The selection of the second team member then involves a trade-off that depends on the nature of the remaining agents' altruistic preferences.

When paired with agent a, agent c requires a lower bonus than when paired with agent b, due to agent a's smaller altruistic preferences. Conversely, agent a demands a higher bonus than agent b when paired with the altruistic agent c. In this case, the optimal team comprises the two most altruistic agents. This outcome arises because the marginal gain from selecting a more altruistic agent outweighs the potential cost savings from choosing an agent with weaker altruistic preferences.

We now generalize the analysis to a broader setting where the principal must form multiple teams. Suppose the principal needs to implement N projects and, for this purpose, must form N teams, each consisting of two agents, selected from a pool of $2N + 1$ available agents. In this scenario, the principal faces an additional decision: one agent must be excluded from team formation. Our analysis shows that it is always optimal for the principal to exclude the least altruistic agent. This strategy minimizes the total incentive costs across teams by ensuring that the remaining agents are paired in a way that explores the complementarities of their preferences, following the same principles established in the earlier two-team case.

5.3 *Outside options*

In the previous analysis, we assumed that agents' outside options were normalized to zero, which simplified the incentive design problem. However, in many real-world settings, employees have alternative opportunities that affect their participation decisions. To capture this more realistic environment, we relax this assumption and consider the case in which agents have non-zero outside options, that is, when $\bar{U} \neq 0$ while maintaining identical projects. This extension allows us to examine how outside options influence team formation and incentive structures.

It is important to note that, if the outside option is such that the incentive compatibility constraint governs the agent's optimal contract, the result remains unchanged: the principal continues to tailor the contract to the agent's effort in order to provide optimal incentives. We summarize this observation in the following claim.

LEMMA 3. *If* $\bar{U} \leq \frac{p_{01}^n k}{p_{11}^n - p_{01}^n}$, *then* $w_{ij}^{n\,*} = 0$ *and all previous results hold.*

In other words, the principal withholds any fixed wage when bonus payments alone suffice to satisfy the agents' participation and limited liability constraints and to induce effort.

We now proceed to examine optimal team formation in the case where the optimal contract is not determined by the incentive compatibility constraint alone. In this setting, the bonus payment implied by incentive compatibility is insufficient to satisfy the participation constraint. Consequently, we solve the principal's problem starting from the participation constraint. The principal's problem is therefore as follows:

$$\min_{\substack{(\{i,j\}\{g,h\})\in T \\ w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n \in R}} \quad \sum_{n=1}^{2} p_{11}^n (b_{ij}^n + b_{ji}^n) + w_{ij}^n + w_{ji}^n$$

$$\text{s.t. } u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, 1, 1) \geq u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, 0, 1),\ \forall i \in A, n \in \{1,2\},\ \textbf{(IC)}$$

$$u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, 1, 1) \geq \bar{U},\ \forall i \in A, n \in \{1,2\},\ \textbf{(PC)}$$

$$w_{ij}^n + b_{ij}^n \geq 0, w_{ij}^n \geq 0,\ \forall i \in A, n \in \{1,2\},\ \textbf{(LLC)}$$

From binding (PC) we have $w_{ij}^n = \bar{U} + k - p_{11} b_{ij}^n - \alpha_i (p_{11} b_{ji}^n + w_{ji}^n)$, solving simultaneously, we obtain $w_{ij}^{n\,*} = \frac{(1-\alpha_i)(\bar{U}+k)}{1-\alpha_i\alpha_j} - p_{11} b_{ij}^n$. Let us use the result from (PC) and restate the problem, and let us drop the (LLC) constraint for a while and check it in the end. Thus,

$$\min_{\substack{(\{i,j\}\{g,h\})\in T \\ b_{ij}^n, b_{ji}^n \in R}} \quad \sum_{n=1}^{2} \frac{(1-\alpha_i)(\bar{U}+k)}{1-\alpha_i\alpha_j} + \frac{(1-\alpha_j)(\bar{U}+k)}{1-\alpha_i\alpha_j}$$

$$\text{s.t.} \quad u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, 1, 1) \geq u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, 0, 1), \forall i \in A, n \in \{1,2\},\ \textbf{(IC)}$$

In the incentive constraint, a fixed wage and an outside option appear on both sides of the inequality and consequently cancel each other out. Substituting the binding participation constraints into the principal's expected payment shows that the objective is independent of the bonuses. Thus, among the set of (IC)-satisfying bonuses, the principal can, without loss of generality, pick the minimal bonuses (the IC-binding) and adjust fixed wages to satisfy PC.

Let us examine the expected payoff of the agents, where $b_{ij}^{n\,*}$ is any bonus which satisfies (IC), $w_{ij}^{n\,*} + p_{11} b_{ij}^{n\,*} = \frac{(1-\alpha_i)(\bar{U}+k)}{1-\alpha_i\alpha_j} > 0$, and this is strictly positive. Thus, (LLC) is satisfied.

Denote $w' \equiv \bar{U} + k$, hence $w_{ij}^{n\,*} + p_{11} b_{ij}^{n\,*} = w' \frac{1-\alpha_i}{1-\alpha_i\alpha_j}$. We observe that the principal's contract has a structure similar to the one derived in Subsection 4.1, "Analysis." The primary difference lies in the fixed part of the contract. Unlike the previous scenario, where the fixed wage was zero, the principal now includes a non-zero fixed wage. This adjustment effectively removes the bonus component from the expected payoffs of both parties.

Accordingly, the procedure for constructing teams under the contract structures specified above closely parallels the approach developed in the section presenting the

main results. This consistency arises from the principal's strategic use of contracts that incorporate a fixed wage component. By embedding a fixed component within the contract, the principal, as in the earlier analysis, exploits the altruistic preferences of team members. This mechanism allows the principal to minimize total wage expenditures, while still satisfying the agents' participation and incentive compatibility constraints. As a result, the logic of optimal team formation remains fundamentally unchanged, despite modifications to the contract structure.

PROPOSITION 4. *Whenever implementation is feasible, in scenarios involving any four agents and any outside option $\bar{U} \in \mathbb{R}$, the principal optimally pairs the least altruistic agent with the most altruistic agent. The remaining two agents form the second team.*

Consequently, Proposition 3 also holds.

5.4 *Uniqueness of equilibrium*

The structure of the production technology determines whether the optimal contract results in a unique equilibrium. When the production function is supermodular, the marginal productivity of an agent increases with the effort level of their teammate ($p_{11} - p_{01} > p_{01} - p_{00}$). This complementarity may lead to multiple equilibria. In such scenarios, if one agent chooses to shirk, the reduced productivity lowers the other agent's expected benefit from exerting effort. This makes shirking a best response and reinforces the non-cooperative outcome, $(shirk, shirk)$ as an equilibrium.

In contrast, when the production function is submodular, the marginal productivity of an agent decreases as the effort of their coworker increases. When incentives lead to an equilibrium of $(work, work)$, an agent derives a higher payoff from exerting effort when their partner shirks. This makes *work* the dominant strategy. This strategic substitutability guarantees a unique equilibrium in which the outcome of both agents working is the only equilibrium.

Since supermodularity can generate multiple equilibria, we consider the *optimal incentive-inducing contract* as defined in Winter (2004).[13] Such a contract induces $(work, work)$ as the unique equilibrium at minimal cost.

To derive an optimal incentive-inducing contract when the probability function is supermodular, the principal must ensure that, within each team, one agent has *work* as a dominant strategy, while the other agent's best response is also *work*, given the belief that the coworker will exert effort with certainty.

To characterize the optimal incentive-inducing contract and the corresponding team formation strategy for the principal, we first identify which agent within a team should be assigned a dominant strategy. The agent assigned the dominant strategy receives a higher bonus than the coworker. Accordingly, when a team consists of two agents with different degrees of altruism levels, the key question becomes which agent should receive the higher bonus.

[13] A contract is an *optimal incentive-inducing contract* if (i) there exists no alternative contract with a lower total reward that implements a unique equilibrium, and (ii) for any $\varepsilon > 0$, the contract with incentive payments increased by ε implements a unique equilibrium.

LEMMA 4. *Suppose agents i and j work on the same project and $0 \leq \alpha_j \leq \alpha_i < 1$. In the optimal incentive-inducing contract, the principal offers a higher bonus to agent j; thus, agent j should have "work" as a dominant strategy.*

The central insight is that when agent j has a dominant strategy, any additional compensation of magnitude Δb provided to agent j, when agent i is altruistic, allows the principal to reduce the bonus offered to agent i by $\alpha_i \Delta b$. Consequently, the principal's net cost of implementing an incentive-inducing bonus becomes $(1 - \alpha_i)\Delta b$. This indicates that it is more cost-effective for the principal to allocate higher incentives to the agent with weaker altruistic preferences (i.e., lower α), thereby minimizing overall contractual expenditures.

ASSUMPTION 1. $\alpha_d \leq \frac{p_{01} - p_{00}}{p_{11} - p_{01}}$.

Assumption 1 is introduced to restrict the model to economically meaningful incentive schemes. In particular, it excludes parameter values under which—due to very high altruism—an agent would exert effort even when his own success payment is nonpositive, solely because of the positive externality generated by the coworker's incentive.

Based on Lemma 4, we can characterize which agent within a team must receive the higher incentive. We use this insight to determine the most efficient team formation from the principal's perspective. In particular, consider agents a and d, who are the least and most altruistic agents, respectively. Lemma 4 implies that, under the incentive-inducing contract, agent a always requires a higher bonus than their teammate to exert effort, whereas agent d can be induced to exert effort with a lower bonus.

We derive a weak ordering over the set of feasible team formations from the principal's perspective. With respect to minimal total incentive costs, pairwise comparisons show that $(\{a,d\},\{b,c\})$ weakly dominates $(\{a,c\},\{b,d\})$, which in turn weakly dominates $(\{a,b\},\{c,d\})$, thereby characterizing the cost-minimizing allocation.

When agent c is matched with agent a, agent c requires a lower incentive to exert effort than agent b would require when matched with agent a. Similarly, agent d requires a lower incentive when matched with agent b than when matched with agent c. These differences reflect that agents with stronger altruistic preferences internalize their coworker's compensation more strongly and therefore need weaker direct incentives.

Moreover, the least altruistic agent a always receives the highest bonus within the team. Consequently, changing agent a's coworker primarily affects the coworker's required bonus, while agent a's own bonus responds less strongly, since α_a is the smallest. Because the coworker also receives a lower bonus than agent a, the induced change in agent a's incentive requirement is weaker in magnitude. As a result, the increase in the bonus required for agent a when he is matched with agent c rather than agent b is weakly smaller than the corresponding decrease in the bonus required for his coworker. By contrast, for agent d, the reduction in the required bonus when he is matched with agent b instead of agent c outweighs the associated increase in the coworker's bonus. Taken together, these effects imply that the team formation $(\{a,c\},\{b,d\})$ weakly dominates $(\{a,b\},\{c,d\})$ in terms of minimal total incentive costs.[14]

[14]When $\alpha_b = \alpha_c$, these two formations have the same total cost.

Next, consider the comparison between $(\{a,c\},\{b,d\})$ and $(\{a,d\},\{b,c\})$. When agent d is matched with agent a, agent d requires a lower incentive than agent c would require when matched with agent a. Similarly, agent b requires a lower incentive when paired with agent c than when paired with agent d.

Furthermore, since agent b is weakly more altruistic than agent a, the reduction in incentive costs achieved by matching agent b with agent c is weakly larger than the reduction achieved by matching agent a with agent c. Consequently, allocating agent c to agent b, rather than to agent a, yields greater overall cost savings. Taken together, these effects imply that the team formation $(\{a,d\},\{b,c\})$ weakly dominates $(\{a,c\},\{b,d\})$.

Accordingly, the cost-minimizing team composition is to pair agent a with agent d, and agent b with agent c. Therefore, the principal's optimal strategy is to form heterogeneous teams by pairing the least altruistic agent with the most altruistic agent in one team and assigning the remaining two agents to the other team.

The rationale behind this formation is simple. Altruistic individuals derive additional satisfaction from their teammates' expected payoffs. By assigning the agent with the highest bonus requirement to work alongside the most altruistic agent, the principal can substantially reduce the second agent's incentive payment without compromising their motivation. Consequently, the least altruistic coworker's utility is less affected by the most altruistic agent's expected payoff. This benefits the principal because other agents derive higher utility from each other than from being paired with the most altruistic agent. Consequently, this pairing minimizes the total compensation required to induce agents to work, making it the most efficient team configuration.

This intuition aligns with the insights presented in the analysis section. We summarize in the following proposition.

PROPOSITION 5. *The principal's optimal team formation is anti-assortative among all team formations that implement* $(work, work)$ *as the unique equilibrium in each team under optimal incentive-inducing contracts.*

5.5 *Optimal team size*

This section presents a stylized extension that isolates the incentive channel; results may differ in the presence of production complementarities that vary with team size. We examine whether focusing on two-agent teams is overly restrictive. To address this question, we consider a setting that isolates the effect of agents' altruism on total wage expenditure. In this subsection, we impose the following assumptions:

ASSUMPTION 2. *Constant marginal contribution across team sizes*

$$p_1 - p_0 = p_{01} - p_{00} = p_{11} - p_{01} = p_{001} - p_{000} = ... = p_{111} - p_{011}.$$

ASSUMPTION 3.

$$u_i(w_L, b_L, e_L) = w^n_{iL} + p^n_{e_L} b^n_{iL} - e_i k + \alpha_i \frac{1}{|L|-1} \sum_{j \in L \setminus i} p^n(e_L) b^n_{jL} + w^n_{jL}$$

where $L \in \mathcal{P}(A)$ denotes the team assigned to project $n \in \{1,2,3,4\}$, and w_L, b_L, and e_L denote the vectors of wages, bonuses, and effort levels of agents in team L, respectively.[15]

Consider a setting with a principal and four agents, where the principal can implement two, three, or four projects. A project is implementable if at least one agent works on that project. Consequently, teams may consist of a single agent or of two, or three agents. In particular, if one team consists of three agents, the remaining agent works alone on the second project.

Assumption 2 states that each project's production function is neither submodular nor supermodular; thus, teamwork does not affect production. Assumption 3 posits that agents' altruistic preferences are adjusted to the size of the team: an agent exhibits altruistic preferences toward all team members.[16]

First, we argue that the case with four projects is suboptimal whenever at least one altruistic agent is present, since forming a two-agent team allows the principal to provide weaker incentives to the altruistic agent than when the agent works alone.

Second, consider the case in which one project is staffed by three agents, while the remaining project is assigned to a single agent. If the agent working alone is selfish, the principal can reduce incentive costs by pairing the most altruistic agent with this selfish agent rather than placing the most altruistic agent in the three-person team. In such a two-agent team, the altruistic agent requires relatively weak incentives, which allows the principal to weakly lower the bonus paid to the altruistic agent.

By contrast, removing the most altruistic agent from the three-agent team would weakly decrease the bonuses required for the remaining team members. The reason is that the most altruistic agent requires the lowest compensation among all agents. When this agent is present in the team, the remaining agents derive relatively low utility from that coworker's compensation, which increases their incentive requirements. Removing the most altruistic agent eliminates this low-compensation coworker, thereby increasing the utility that the remaining agents derive from their coworkers' pay and allowing the principal to reduce the bonuses needed to induce effort. As a result, concentrating agents into two equal-sized teams allows the principal to exploit altruistic preferences more efficiently.

Consequently, from the principal's perspective, it is optimal to form two projects, each staffed by two agents rather than a three-agent team and a singleton. Analogous arguments apply when the altruism level of the lone agent is different.

PROPOSITION 6. *Under Assumptions 2 and 3, the principal's optimal strategy is to implement two projects and form teams in an anti-assortative manner.*

This result arises from the principal's ability to effectively harness agents' other-regarding preferences when they are paired. As team size increases, the influence of these preferences becomes diluted, since agents must now consider the well-being of multiple teammates rather than focusing on a single partner.

[15] When $|L| = 1$ the agent works in isolation and derives no utility from coworkers' compensation; accordingly, the agent's utility equals his net material payoff.

[16] Note that when the team size equals two, agents' preferences coincide with those defined in the baseline setup.

5.6 *Spiteful and altruistic agents*

In the literature, analyses of altruistic agents are often complemented by considerations of spiteful or envious preferences.[17] To align our framework with this broader literature, we extend the analysis to allow agents' other-regarding preferences to be negative, thereby capturing the possibility of spiteful behaviour. We define an agent as spiteful when his utility decreases with the coworker's monetary payoff.

Formally, we extend the support of the other-regarding preference parameter to $\alpha_i \in (-1, 1)$, thereby permitting both altruistic and spiteful preferences within a unified framework. The lower bound ($\alpha_i > -1$) ensures that agents continue to value increases in their own compensation more than reductions in their coworkers' compensation. This restriction rules out extreme forms of spite that would induce agents to sacrifice their own payoffs solely to reduce others' earnings, thereby preserving well-behaved incentive responses and ensuring the tractability of the optimal contracting problem.

This extension has direct implications for incentive provision. *Ceteris paribus*, an altruistic agent derives greater utility when the coworker's bonus is higher, which reduces the external incentives needed to induce effort. In contrast, a spiteful agent's utility decreases when the coworker's bonus increases. Hence, raising the coworker's bonus erodes the spiteful agent's incentives, and the principal must provide stronger incentives to ensure effort.

These opposing incentive effects generate complementarities in team composition. When altruistic and spiteful agents are paired, both agents can be motivated at lower cost: altruistic agents' incentives are eased by higher coworker compensation, while spiteful agents require weaker incentives when coworkers receive lower pay. As a result, heterogeneous teams composed of agents with opposing other-regarding preferences minimize total incentive costs.

This intuition leads directly to the following result.

PROPOSITION 7. *In scenarios involving any type of four agents, the principal's best strategy is to pair the most altruistic (least spiteful) agent with the most spiteful (least altruistic) agent. The remaining two agents form the second team.*

Appendix A provides a unified proof for Propositions 1 and 7.

Proposition 7 is a natural extension of Proposition 1 to a broader support of agents' other-regarding preference parameters. Consequently, Proposition 3 continues to hold.

5.7 *Fairness*

Fairness is an essential aspect of organizations, especially when it comes to decisions involving compensation and team formation. From comparative statics in Section 4.2, we observe that agents with stronger other-regarding preferences receive lower bonus offers compared to agents with weaker such preferences. This outcome implies that the principal designs contracts that differentiate among agents based on their social preferences. As a result, even though the agents possess identical production capabilities,

[17] See Levine (1998), Sobel (2005).

they receive different bonuses if the projects are successful. This aligns with the type of discrimination highlighted in Winter (2004). Unlike in Winter (2004), in the baseline, we do not require uniqueness. However, Section 5.4 shows that the main assignment result is robust when the principal enforces uniqueness.

At the same time, as shown in the first subsection of this section, the principal's contract design ensures that agents' expected payoffs remain equal, regardless of their own or their teammates' other-regarding preferences. Therefore, despite differences in bonus levels, all agents receive the same expected payoff in equilibrium.

In summary, while all agents have identical technological skills and receive the same expected payoffs—suggesting fairness from an ex-ante perspective—the principal's strategy leads to lower bonus offers for agents with stronger other-regarding preferences. This creates a trade-off between fairness in expected outcomes and apparent inequity in realized bonus payments.

5.8 *Extensions*

A natural extension of the current analysis is to examine environments in which projects are heterogeneous. In particular, it would be relevant to identify the conditions under which the optimal team formation result in Proposition 1 continues to hold when project characteristics differ, for example, when production functions vary across projects.

Another potential direction for extension is to analyze cases where agents differ not only in their other-regarding preferences but also in their productivity levels, particularly in settings where project technologies exhibit supermodularity. While our analysis shows that anti-assortative matching (pairing agents with opposing social preferences) is optimal when agents have identical productivity levels, prior research suggests that in standard models with heterogeneous productivity, the principal typically favors assortative matching.[18] Therefore, it would be interesting to explore the trade-off between assortative and anti-assortative matching when both productivity differences and social preferences are present, and to characterize the conditions under which each type of matching is optimal.

6. Conclusion

Our model generates predictions about the influence of other-regarding preferences on team composition. It suggests that an agent's altruistic tendencies can significantly affect the optimal formation of teams within an organization. Specifically, the model predicts how different combinations of agents, based on their other-regarding preferences, can minimize incentive costs while maintaining or enhancing productivity.

Furthermore, we treat α as exogenous. If firms can influence α through culture, training, or selection, our model suggests that shaping and exploiting heterogeneity in other-regarding preferences can lower the incentive costs of implementing effort by guiding team assignment. For example, companies could promote altruism among their

[18] See Becker (1991).

employees by organizing team-building activities or cultivating a corporate culture that prioritizes collective success. This approach aligns with the findings of Hattori and Yamada (2025), who demonstrate that social interaction among team members significantly improves team performance. By shaping other-regarding preferences, firms can foster more cooperative team dynamics, reduce wage demands, and ultimately enhance organizational efficiency.

In conclusion, organizations can design more effective human resource strategies by understanding the role of other-regarding preferences in team formation and performance. This involves selecting appropriate combinations of employees for different projects and actively shaping the social dynamics of the workforce to align with organizational goals. Through these strategic interventions, firms can leverage the insights provided by our model to enhance both employee and organizational performance.

References

ABBINK, KLAUS, BERND IRLENBUSCH, AND ELKE RENNER (2000): "The moonlighting game: An experimental study on reciprocity and retribution," *Journal of Economic Behavior & Organization*, 42 (2), 265–277.

ALCHIAN, ARMEN A AND HAROLD DEMSETZ (1972): "Production, information costs, and economic organization," *The American Economic Review*, 62 (5), 777–795.

ANDREONI, JAMES AND JOHN MILLER (2002): "Giving according to GARP: An experimental test of the consistency of preferences for altruism," *Econometrica*, 70 (2), 737–753.

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.

BARGHOUTI, ZINA, JACOB GUINOT, AND RICARDO CHIVA (2023): "Compassion and altruism in organizations: a path for firm survival," *International Journal of Manpower*, 44 (9), 1–19.

BARTLING, BJÖRN (2011): "Relative performance or team evaluation? Optimal contracts for other-regarding agents," *Journal of Economic Behavior & Organization*, 79 (3), 183–193.

BECKER, GARY S. (1991): *4. Assortative Mating in Marriage Markets*, Cambridge, MA and London, England: Harvard University Press, 108–134.

BEL, ROLAND, VLADIMIR SMIRNOV, AND ANDREW WAIT (2015): "Team composition, worker effort and welfare," *International Journal of Industrial Organization*, 41, 1–8.

BLANCO-FERNÁNDEZ, DARÍO, STEPHAN LEITNER, AND ALEXANDRA RAUSCH (2025): "Interactions between dynamic team composition and coordination: an agent-based modeling approach," *Review of Managerial Science*, 19 (1), 1–37.

DASKE, THOMAS (2017): "Friends and Foes at Work: Assigning Teams in a Social Network," Working paper, Technical University of Munich, Munich, Germany.

DEMOUGIN, DOMINIQUE AND CLAUDE FLUET (2003): "Group vs. individual performance pay when workers are envious," *Cahier de recherche/Working Paper*, 3, 18.

DONG, MIAOMIAO AND TATIANA MAYSKAYA (2024): "Does reducing communication barriers promote diversity?" *Journal of Economic Theory*, 222, 105932.

DONG, MIAOMIAO, TATIANA MAYSKAYA, VLADIMIR SMIRNOV, OLIVIA TAYLOR, AND ANDREW WAIT (2024): "Optimal Skill Diversity in Teams," *Available at SSRN 4343715*.

DUFWENBERG, MARTIN AND GEORG KIRCHSTEIGER (2000): "Reciprocity and wage undercutting," *European Economic Review*, 44 (4-6), 1069–1078.

ENGELMANN, DIRK AND MARTIN STROBEL (2004): "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments," *American Economic Review*, 94 (4), 857–869.

FEHR, ERNST AND KLAUS M SCHMIDT (1999): "A theory of fairness, competition, and cooperation," *The Quarterly Journal of Economics*, 114 (3), 817–868.

FRANCO, APRIL MITCHELL, MATTHEW MITCHELL, AND GALINA VERESHCHAGINA (2011): "Incentives and the structure of teams," *Journal of Economic Theory*, 146 (6), 2307–2332.

GLOVER, JONATHAN AND EUNHEE KIM (2021): "Optimal team composition: Diversity to foster implicit team incentives," *Management Science*, 67 (9), 5800–5820.

HAMILTON, BARTON H, JACK A NICKERSON, AND HIDEO OWAN (2012): "Diversity and productivity in production teams," in *Advances in the Economic Analysis of Participatory and Labor-managed Firms*, Emerald Group Publishing Limited.

HATTORI, KEISUKE AND MAI YAMADA (2025): "Closing the Psychological Distance: Effect of Social Interaction on Team Performance," *Available at SSRN 5204747*.

IMHOF, LUKAS A. AND MICHAEL KRÄKEL (2023): "Team Diversity and Incentives," *Management Science*, 69 (4), 2497–2516.

ITOH, HIDESHI (2004): "Moral hazard and other-regarding preferences," *The Japanese Economic Review*, 55 (1), 18–45.

LEVINE, DAVID K (1998): "Modeling altruism and spitefulness in experiments," *Review of Economic Dynamics*, 1 (3), 593–622.

LIVIO, LUCA AND ALESSANDRO DE CHIARA (2019): "Friends or foes? Optimal incentives for reciprocal agents," *Journal of Economic Behavior & Organization*, 167, 245–278.

MARSCHAK, JACOB AND ROY RADNER (1972): *Economic Theory of Teams*, New Haven, CT: Yale University Press.

REY-BIEL, PEDRO (2008): "Inequity aversion and team incentives," *Scandinavian Journal of Economics*, 110 (2), 297–320.

RIOUX, SHEILA M AND LOUIS A PENNER (2001): "The causes of organizational citizenship behavior: a motivational analysis." *Journal of Applied Psychology*, 86 (6), 1306.

SARKISIAN, ROBERTO (2021): "Optimal Incentives Schemes under Homo Moralis Preferences," *Games*, 12 (1), 28.

SOBEL, JOEL (2005): "Interdependent preferences and reciprocity," *Journal of Economic Literature*, 43 (2), 392–436.

WAGNER, SHARON L AND MICHAEL C RUSH (2000): "Altruistic organizational citizenship behavior: Context, disposition, and age," *The Journal of Social Psychology*, 140 (3), 379–391.

WEBER, BERNHARD AND GUIDO HERTEL (2007): "Motivation gains of inferior group members: a meta-analytical review." *Journal of Personality and Social Psychology*, 93 (6), 973–993.

WINTER, EYAL (2004): "Incentives and discrimination," *American Economic Review*, 94 (3), 764–773.

APPENDIX A: MAIN RESULTS

Roadmap. We first characterize the cost-minimizing incentive contract for an arbitrary pair under limited liability and moral hazard (Lemma 1). We then use the resulting bonus expressions to compare total expected compensation across feasible partitions and establish the optimal two-team assignment (Propositions 1, 2 and 7). Finally, we extend the assignment argument to $2N$ agents and N identical projects via an exchange argument, showing that iteratively pairing the least and most altruistic agents is (weakly) optimal, with non-uniqueness only under ties (Proposition 3).

Proof of Lemma 1: The derived optimal bonus ${b_{ij}^n}^*$ is positive. Since there is always a positive probability that the project fails, the limited liability constraint is equivalent to

$$w_{ij}^n \geq 0, i \in A, n \in \{1,2\}, \quad \textbf{(LLC)}$$

To satisfy the limited liability constraint, it is therefore optimal to set ${w^n}_{ij}^* = 0$.

We show that whenever a contract satisfies the incentive compatibility and limited liability constraints, it also satisfies the participation constraint. Specifically, we derive the minimal bonus that satisfies the participation constraint and show that it is smaller than the optimal bonus implied by the incentive compatibility constraint.

$$u_i(w_{ij}^n, w_{ji}^n, b_{ij}^n, b_{ji}^n, 1, 1) = w_{ij}^n + p_{11}^n b_{ij}^n - k + \alpha_i(p_{11}^n b_{ji}^n + w_{ji}^n) \geq 0, i \in A, n \in \{1,2\}, \quad \textbf{(PC)}$$

Rearranging the terms and inserting the optimal bonus, we obtain:

$$b_{ij}^n \geq \frac{k}{p_{11}^n} - \alpha_i b_{ji}^n \tag{A.1}$$

Let us note that, as for the incentive constraint, the participation constraint is identical for agent j; only subscripts are swapped

$$b_{ji}^n \geq \frac{k}{p_{11}^n} - \alpha_j b_{ij}^n \tag{A.2}$$

Considering binding (PC), we identify a minimal bonus that satisfies it; thus, we have the following

$$b_{ij}^n = \frac{k(1-\alpha_i)}{p_{11}^n(1-\alpha_i\alpha_j)} \tag{A.3}$$

Comparing the bonus satisfying (IC) derived in (6) and satisfying (PC) in (A.3)

$$\frac{k(1-\alpha_i)}{(p_{11}^n - p_{01}^n)(1-\alpha_i\alpha_j)} \geq \frac{k(1-\alpha_i)}{p_{11}^n(1-\alpha_i\alpha_j)}$$

Since $p_{11}^n - p_{01}^n < p_{11}^n$, the inequality is satisfied; therefore, (PC) holds whenever (IC) and (LLC) hold.

□

First, let us make two simple mathematical observations that will be used in our proof of Propositions 1 and 7.

OBSERVATION 1. *For all* $x \in (-1,1)$ *the following is true:* $x(2-x) < 1$.

OBSERVATION 2. *For all* $x \in (-1,1)$ *the following is true:* $x + x^2 - x^3 < 1$.

Here, we provide the general structure of the proof of Proposition 7, which also encompasses the proof of Proposition 1.

Note: This proof is written for $\alpha_i \in (-1,1)$ and therefore implies Proposition 1 as a special case when $\alpha_i \in [0,1)$.

Proof of Propositions 1 and 7: In the proof, we use the following steps:

1. Derive optimal wage contract for any team composition (derived in (6)).
2. Given the optimal wage contract, select the team formation from the set T of all possible team formations that has minimal total cost.

Given that $-1 < \alpha_a \leq \alpha_b \leq \alpha_c \leq \alpha_d < 1$.

Step one: for any team containing agents i and j we derived in (6) optimal contracts $(w_{ij}^{n}{}^*, b_{ij}^{n}{}^*)$ and $(w_{ji}^{n}{}^*, b_{ji}^{n}{}^*)$, where $w_{ij}^{n}{}^* = w_{ji}^{n}{}^* = 0$ and:

$$b_{ij}^{n}{}^* = b_0^n \frac{1-\alpha_i}{1-\alpha_i\alpha_j} \text{ and } b_{ji}^{n}{}^* = b_0^n \frac{1-\alpha_j}{1-\alpha_i\alpha_j}$$

When two teams are formed with agents $\{ij\}$ and $\{gh\}$, the total cost is:

$$TC_{(\{i,j\}\{g,h\})} = b_0^n \left(\frac{1-\alpha_i}{1-\alpha_i\alpha_j} + \frac{1-\alpha_j}{1-\alpha_i\alpha_j} \right) + b_0^m \left(\frac{1-\alpha_g}{1-\alpha_g\alpha_h} + \frac{1-\alpha_h}{1-\alpha_g\alpha_h} \right) \tag{A.4}$$

The first set of parentheses represents the sum of the team's bonuses $\{ij\}$ working on project n, while the second set represents the sum of the team's bonuses $\{gh\}$ working on project m. Therefore

$$TC_{(\{i,j\}\{g,h\})} = b_0^n \left(\frac{2-\alpha_i-\alpha_j}{1-\alpha_i\alpha_j} \right) + b_0^m \left(\frac{2-\alpha_g-\alpha_h}{1-\alpha_g\alpha_h} \right) \tag{A.5}$$

In the scenario where two projects are identical, meaning that $p^1_{e_ie_j}$ equals $p^2_{e_ie_j}$. Therefore, we can conclude that the values of b_0^n and b_0^m are equal, and for simplicity, set $b_0^n = b_0^m = b_0$.

Step two: The principal must choose one optimal team formation out of three possible cases[19]

$$\min\{TC_{(\{a,b\}\{c,d\})}, TC_{(\{a,c\}\{b,d\})}, TC_{(\{a,d\}\{b,c\})}\}$$

whereby (A.5)

$$TC_{(\{a,b\}\{c,d\})} = b_0 \left(\frac{2-\alpha_a-\alpha_b}{1-\alpha_a\alpha_b} + \frac{2-\alpha_c-\alpha_d}{1-\alpha_c\alpha_d} \right), TC_{(\{a,c\}\{b,d\})} = b_0 \left(\frac{2-\alpha_a-\alpha_c}{1-\alpha_a\alpha_c} + \frac{2-\alpha_b-\alpha_d}{1-\alpha_b\alpha_d} \right.$$

[19]When the probability functions of the projects are identical, the total cost $TC_{(\{i,j\}\{g,h\})}$ is equivalent to $TC_{\{g,h\}\{i,j\})}$, which is why we have three possible cases instead of six.

$$TC_{(\{a,d\}\{b,c\})} = b_0 \left(\frac{2-\alpha_a-\alpha_d}{1-\alpha_a\alpha_d} + \frac{2-\alpha_b-\alpha_c}{1-\alpha_b\alpha_c} \right)$$

Let us simplify the total cost in general terms.

$$TC_{(\{i,j\}\{g,h\})} = b_0 \left(\frac{2-\alpha_i-\alpha_j}{1-\alpha_i\alpha_j} + \frac{2-\alpha_g-\alpha_h}{1-\alpha_g\alpha_h} \right) =$$

$$= b_0 \left(\frac{(2-\alpha_i-\alpha_j)(1-\alpha_g\alpha_h)+(2-\alpha_g-\alpha_h)(1-\alpha_i\alpha_j)}{(1-\alpha_i\alpha_j)(1-\alpha_g\alpha_h)} \right) =$$

$$= b_0 \left(\frac{4-\alpha_i-\alpha_j-\alpha_g-\alpha_h+\alpha_i\alpha_g\alpha_h+\alpha_j\alpha_g\alpha_h+\alpha_i\alpha_j\alpha_g+\alpha_i\alpha_j\alpha_h-2(\alpha_i\alpha_j+\alpha_g\alpha_h)}{1+\alpha_i\alpha_j\alpha_g\alpha_h-\alpha_i\alpha_j-\alpha_g\alpha_h} \right).$$

Let us denote $A \coloneqq 4-\alpha_i-\alpha_j-\alpha_g-\alpha_h+\alpha_i\alpha_g\alpha_h+\alpha_j\alpha_g\alpha_h+\alpha_i\alpha_j\alpha_g+\alpha_i\alpha_j\alpha_h$ and $B \coloneqq 1+\alpha_i\alpha_j\alpha_g\alpha_h$. Hence,

$$TC_{(\{i,j\}\{g,h\})} = b_0 \left(\frac{A-2(\alpha_i\alpha_j+\alpha_g\alpha_h)}{B-\alpha_i\alpha_j-\alpha_g\alpha_h} \right) \tag{A.6}$$

Now, we show that $TC_{(\{a,d\}\{b,c\})}$ is minimum. First, we compare the total costs of the two following team formations, $TC_{(\{a,d\}\{b,c\})}$ and $TC_{(\{a,c\}\{b,d\})}$

$$TC_{(\{a,d\}\{b,c\})} \leq TC_{(\{a,c\}\{b,d\})} \Leftrightarrow b_0 \left(\frac{A-2(\alpha_a\alpha_d+\alpha_b\alpha_c)}{B-\alpha_a\alpha_d-\alpha_b\alpha_c} \right) \leq b_0 \left(\frac{A-2(\alpha_a\alpha_c+\alpha_b\alpha_d)}{B-\alpha_a\alpha_c-\alpha_b\alpha_d} \right) \Leftrightarrow$$

$$\Leftrightarrow (A-2(\alpha_a\alpha_d+\alpha_b\alpha_c))(B-\alpha_a\alpha_c-\alpha_b\alpha_d) \leq (A-2(\alpha_a\alpha_c+\alpha_b\alpha_d))(B-\alpha_a\alpha_d-\alpha_b\alpha_c) \Leftrightarrow$$

$$\Leftrightarrow AB-A(\alpha_a\alpha_c+\alpha_b\alpha_d)-2B(\alpha_a\alpha_d+\alpha_b\alpha_c)+2(\alpha_a\alpha_d+\alpha_b\alpha_c)(\alpha_a\alpha_c+\alpha_b\alpha_d) \leq$$

$$\leq AB-A(\alpha_a\alpha_d+\alpha_b\alpha_c)-2B(\alpha_a\alpha_c+\alpha_b\alpha_d)+2(\alpha_a\alpha_c+\alpha_b\alpha_d)(\alpha_a\alpha_d+\alpha_b\alpha_c) \Leftrightarrow$$

$$\Leftrightarrow (A-2B)(\alpha_a\alpha_d+\alpha_b\alpha_c) \leq (A-2B)(\alpha_a\alpha_c+\alpha_b\alpha_d) \Leftrightarrow$$

$$\Leftrightarrow (A-2B)(\alpha_a\alpha_d+\alpha_b\alpha_c-\alpha_a\alpha_c-\alpha_b\alpha_d) \leq 0 \Leftrightarrow$$

$$\Leftrightarrow (A-2B)(\alpha_d(\alpha_a-\alpha_b)-\alpha_c(\alpha_a-\alpha_b) \leq 0 \Leftrightarrow$$

$$\Leftrightarrow (A-2B)((\alpha_d-\alpha_c)(\alpha_a-\alpha_b)) \leq 0^{20} \tag{A.7}$$

Let us show that the second factor is nonpositive $(\alpha_d-\alpha_c)(\alpha_a-\alpha_b) \leq 0$. It is given that $\alpha_c \leq \alpha_d$, therefore, the first term $(\alpha_d-\alpha_c)$ is either positive or zero. Also, we know that $\alpha_a \leq \alpha_b$, thus the second factor $(\alpha_a-\alpha_b)$ is either negative or zero. Hence, the multiplication is zero or negative.

[20] In the case analysis where $\alpha_a = \alpha_b = 0$ and $\alpha_c = \alpha_d = \alpha$, the difference in total costs between the two cases is zero, indicating that the formations $(\{a,c\}\{b,d\})$ and $(\{a,d\}\{b,c\})$ have the same total cost.

Next, we do similar operations and compare $TC_{(\{a,c\}\{b,d\})}$ and $TC_{(\{a,b\}\{c,d\})}$.

$$TC_{(\{a,c\}\{b,d\})} \leq TC_{(\{a,b\}\{c,d\})} \Leftrightarrow b_0\left(\frac{A-2(\alpha_a\alpha_c+\alpha_b\alpha_d)}{B-\alpha_a\alpha_c-\alpha_b\alpha_d}\right) \leq b_0\left(\frac{A-2(\alpha_a\alpha_b+\alpha_c\alpha_d)}{B-\alpha_a\alpha_b-\alpha_c\alpha_d}\right) \Leftrightarrow$$

$$\Leftrightarrow (A-2(\alpha_a\alpha_c+\alpha_b\alpha_d))(B-\alpha_a\alpha_b-\alpha_c\alpha_d) \leq (A-2(\alpha_a\alpha_b+\alpha_c\alpha_d))(B-\alpha_a\alpha_c-\alpha_b\alpha_d) \Leftrightarrow$$

$$\Leftrightarrow AB-A(\alpha_a\alpha_b+\alpha_c\alpha_d)-2B(\alpha_a\alpha_c+\alpha_b\alpha_d)+2(\alpha_a\alpha_c+\alpha_b\alpha_d)(\alpha_a\alpha_b+\alpha_c\alpha_d) \leq$$

$$\leq AB-A(\alpha_a\alpha_c+\alpha_b\alpha_d)-2B(\alpha_a\alpha_b+\alpha_c\alpha_d)+2(\alpha_a\alpha_b+\alpha_c\alpha_d)(\alpha_a\alpha_c+\alpha_b\alpha_d) \Leftrightarrow$$

$$\Leftrightarrow (A-2B)(\alpha_a\alpha_c+\alpha_b\alpha_d) \leq (A-2B)(\alpha_a\alpha_b+\alpha_c\alpha_d) \Leftrightarrow$$

$$\Leftrightarrow (A-2B)(\alpha_a\alpha_c+\alpha_b\alpha_d-\alpha_a\alpha_b-\alpha_c\alpha_d) \leq 0 \Leftrightarrow$$

$$\Leftrightarrow (A-2B)(\alpha_a(\alpha_c-\alpha_b)-\alpha_d(\alpha_c-\alpha_b) \leq 0 \Leftrightarrow$$

$$\Leftrightarrow (A-2B)((\alpha_c-\alpha_b)(\alpha_a-\alpha_d)) \leq 0 \tag{A.8}$$

The second factor $(\alpha_c-\alpha_b)(\alpha_a-\alpha_d)$ is nonpositive as in the previous comparison. Indeed, the first term is positive or zero because $\alpha_b \leq \alpha_c$; the second term $(\alpha_a-\alpha_d)$ is either zero or negative since $\alpha_a \leq \alpha_d$.

Thus, if $A-2B \geq 0$, then

$$TC_{(\{a,d\}\{b,c\})} \leq TC_{(\{a,c\}\{b,d\})} \leq TC_{(\{a,b\}\{c,d\})}, \tag{A.9}$$

and therefore $TC_{(\{a,d\}\{b,c\})}$ attains the minimum among the three team formations. [21]

$$A-2B = 4-\alpha_a-\alpha_b-\alpha_c-\alpha_d+\alpha_a\alpha_c\alpha_d+\alpha_b\alpha_c\alpha_d+\alpha_a\alpha_b\alpha_d+\alpha_a\alpha_b\alpha_c-2(1+\alpha_a\alpha_b\alpha_c\alpha_d) \tag{A.10}$$

In general, we know that $-1 < \alpha_a \leq \alpha_b \leq \alpha_c \leq \alpha_d < 1$. To prove the proposition, we must demonstrate that $A-2B \geq 0$, for given constraints.

First, let us consider the case $\alpha_a = 0$; therefore, $\alpha_b, \alpha_c, \alpha_d \in [0,1)$. Thus, we have

$$A-2B = 4-\alpha_b-\alpha_c-\alpha_d+\alpha_b\alpha_c\alpha_d-2 \tag{A.11}$$

Here we use mathematical manipulation; we add a nonpositive term: $(\alpha_c-\alpha_b)\alpha_c\alpha_d - (\alpha_c-\alpha_b) = -(\alpha_c-\alpha_b)(1-\alpha_c\alpha_d) \leq 0$ to $A-2B$. The term is nonpositive because both expressions within the parentheses are nonnegative. The first expression is nonnegative due to the condition $\alpha_c \geq \alpha_b$, and the expression in the second parenthesis is positive. Thus:

$$A-2B = 2-\alpha_b-\alpha_c-\alpha_d+\alpha_b\alpha_c\alpha_d \geq 2-\alpha_b-\alpha_c-\alpha_d+\alpha_b\alpha_c\alpha_d+((\alpha_c-\alpha_b)\alpha_c\alpha_d-(\alpha_c-\alpha_b)) =$$

$$= 2-2\alpha_c-\alpha_d+\alpha_c^2\alpha_d = 2(1-\alpha_c)-\alpha_d(1-\alpha_c^2) = (1-\alpha_c)(2-\alpha_d(1+\alpha_c)) \geq 0 \tag{A.12}$$

[21] If the minimum is not singleton, then $TC_{(\{a,d\}\{b,c\})}$ belongs to this set.

The first term is positive since $\alpha_c < 1$, and the second term is positive because $1 + \alpha_c$ is less than 2 and α_d is less than one.

Therefore, for any $\alpha_b, \alpha_c, \alpha_d \in [0, 1)$, the expression (A.12) is positive.

Let us consider a scenario when $\alpha_a < 0$:

$$A - 2B = 4 - \alpha_a - \alpha_b - \alpha_c - \alpha_d + \alpha_a\alpha_c\alpha_d + \alpha_b\alpha_c\alpha_d + \alpha_a\alpha_b\alpha_d + \alpha_a\alpha_b\alpha_c - 2(1 + \alpha_a\alpha_b\alpha_c\alpha_d) =$$

$$= 4 - 2(1 + \alpha_a\alpha_b\alpha_c\alpha_d) + \alpha_a(\alpha_c\alpha_d - 1) + \alpha_b(\alpha_c\alpha_d - 1) + \alpha_c(\alpha_a\alpha_b - 1) + \alpha_d(\alpha_a\alpha_b - 1) \quad \text{(A.13)}$$

Let us prove case by case. First consider when $\alpha_d < 0$, therefore, all alphas are negative. The sum of the first two elements is positive because $\alpha_a\alpha_b\alpha_c\alpha_d < 1$, all other elements of the sum are also positive because the term inside the parentheses is always negative, and the coefficient is nonpositive. Therefore, in the case where $\alpha_d < 0$ expression A.10 is nonnegative.

In the second case, consider when $\alpha_b, \alpha_c \leq 0$ and $0 \leq \alpha_d$

$$A - 2B = 2 - 2(1 + \alpha_a\alpha_b\alpha_c\alpha_d) + \alpha_a(\alpha_c\alpha_d - 1) + \alpha_b(\alpha_c\alpha_d - 1) + \alpha_c(\alpha_a\alpha_b - 1) + \alpha_d(\alpha_a\alpha_b - 1) + 2$$

Again, the sum of the first two elements is positive because $\alpha_a\alpha_b\alpha_c\alpha_d \leq 0$. The next three elements are also nonnegative because each coefficient of each parentheses is nonpositive, and the terms in the parentheses are negative. Finally, the sum of the last two terms is positive because $2 > \alpha_d(\alpha_a\alpha_b - 1)$. Thus, in the case where $0 \leq \alpha_d$ but $\alpha_b, \alpha_c \leq 0$ the expression (A.10) is nonnegative.

Next, consider the case where $\alpha_b \leq 0$ and $0 \leq \alpha_c, \alpha_d$. First, let us show that the derivatives of $A - 2B$ with respect to α_d and α_c are negative.

$$\frac{\partial(A - 2B)}{\partial\alpha_d} = -2\alpha_a\alpha_b\alpha_c + \alpha_a\alpha_c + \alpha_b\alpha_c + (\alpha_a\alpha_b - 1) \quad \text{(A.14)}$$

$$\frac{\partial(A - 2B)}{\partial\alpha_c} = -2\alpha_a\alpha_b\alpha_d + \alpha_a\alpha_d + \alpha_b\alpha_d + (\alpha_a\alpha_b - 1) \quad \text{(A.15)}$$

All elements in both expressions are nonpositive. Let us demonstrate one of them. $-2\alpha_a\alpha_b\alpha_c$ is nonpositive because the two terms are nonpositive, hence the negative coefficient holds the sign. $\alpha_a\alpha_c$ and $\alpha_b\alpha_c$ are clearly nonpositive. $(\alpha_a\alpha_b - 1)$ is negative. Therefore, $A - 2B$ decreases (does not increase) in α_c and α_d. To minimize the expression, set $\alpha_c = \alpha_d = 1$. If the expression is still a nonnegative, then it will be positive for any positive values of α_c and α_d.

$$A - 2B\Big|_{\alpha_c=\alpha_d=1} = 2 - 2(1 + \alpha_a\alpha_b) + 2(\alpha_a\alpha_b - 1) + 2 = 0 \quad \text{(A.16)}$$

Therefore, when $\alpha_b \leq 0$ and $0 \leq \alpha_c, \alpha_d$ the expression A.10 is nonnegative.

Finally, the scenario in which $\alpha_a < 0$ concludes with the case in which $0 \leq \alpha_b, \alpha_c, \alpha_d$.

$$A - 2B = 4 - \alpha_a - \alpha_b - \alpha_c - \alpha_d + \alpha_a\alpha_c\alpha_d + \alpha_b\alpha_c\alpha_d + \alpha_a\alpha_b\alpha_d + \alpha_a\alpha_b\alpha_c - 2 - 2\alpha_a\alpha_b\alpha_c\alpha_d =$$

$$= 2 - \alpha_b - \alpha_c - \alpha_d + \alpha_b\alpha_c\alpha_d - \alpha_a(1 - \alpha_b\alpha_c - \alpha_b\alpha_d - \alpha_c\alpha_d + 2\alpha_b\alpha_c\alpha_d) =$$

$$= 2 - \alpha_b - \alpha_c - \alpha_d + \alpha_b\alpha_c\alpha_d - \alpha_a(1 - \alpha_c\alpha_d(1 - \alpha_b) - \alpha_b\alpha_c(1 - \alpha_d) - \alpha_b\alpha_d) \quad \text{(A.17)}$$

As demonstrated in equation (A.12), when $\alpha_a = 0$, the sum of the first five elements is positive. Based on this result, if we can show that the entire expression decreases in α_a, we can infer that the expression is lower for $\alpha_a = 0$ compared to $\alpha_a < 0$. This deduction will suffice to prove that $A - 2B > 0$ when $\alpha_a < 0$. Therefore, we need to examine the derivative of $A - 2B$ with respect to α_a and demonstrate that this derivative is negative[22].

$$\frac{\partial(A - 2B)}{\partial\alpha_a} = -(1 - \alpha_c\alpha_d(1 - \alpha_b) - \alpha_b\alpha_c(1 - \alpha_d) - \alpha_b\alpha_d) < 0$$

$$\Leftrightarrow \alpha_c\alpha_d(1 - \alpha_b) + \alpha_b\alpha_c(1 - \alpha_d) + \alpha_b\alpha_d < 1$$

$$\Leftrightarrow \alpha_d(\alpha_c(1 - \alpha_b) + \alpha_b) + \alpha_b\alpha_c(1 - \alpha_d) < 1 \quad \text{(A.18)}$$

Let us note that the second term of the first multiplication $\alpha_c(1 - \alpha_b) + \alpha_b = \alpha_c + \alpha_b(1 - \alpha_c)$ is increasing in α_b because $\alpha_c < 1$. Therefore, if we assume that $\alpha_b = \alpha_c$ we only increase the coefficient; hence

$$\alpha_c - \alpha_c^2 + \alpha_c = \alpha_c(2 - \alpha_c) < 1 \quad \text{(A.19)}$$

The last inequality is correct by Observation 1, as the maximum of the expression is one and attained when $\alpha_c = 1$. Thus, we can write the following:

$$\alpha_d(\alpha_c(1 - \alpha_b) + \alpha_b) + \alpha_b\alpha_c(1 - \alpha_d) < \alpha_d + \alpha_b\alpha_c(1 - \alpha_d) \leq \alpha_d + \alpha_d^2(1 - \alpha_d) < 1 \quad \text{(A.20)}$$

The second inequality is true because $\alpha_c\alpha_b \leq \alpha_d^2$. The last inequality is correct by Observation 2, as the expression on the left-hand side has local maxima when $\alpha_d = 1$ and the maxima equals 1. Thus, it is strictly less than one when $\alpha_d \in (0, 1)$. We know that the $A - 2B$ is positive when α_a equals zero and we have just shown that it decreases in α_a. Therefore, we can conclude that $A - 2B$ is positive when $\alpha_a < 0$.

We proved the proposition when $\alpha_a \leq 0$.

To complete the proof of the proposition, it is necessary to consider the scenario when $\alpha_a > 0$

$$A - 2B = 4 - \alpha_a - \alpha_b - \alpha_c - \alpha_d + \alpha_a\alpha_c\alpha_d + \alpha_b\alpha_c\alpha_d + \alpha_a\alpha_b\alpha_d + \alpha_a\alpha_b\alpha_c - 2 - 2\alpha_a\alpha_b\alpha_c\alpha_d =$$

$$= 2 - \alpha_a - \alpha_b - \alpha_c - \alpha_d + \alpha_a\alpha_c\alpha_d + \alpha_b\alpha_c\alpha_d + \alpha_a\alpha_b\alpha_d + \alpha_a\alpha_b\alpha_c - 2\alpha_a\alpha_b\alpha_c\alpha_d \quad \text{(A.21)}$$

To prove that the expression is positive, we do the following steps: 1) we look at the partial derivatives of $A - 2B$ with respect to α_b and α_c; if the derivative is positive, this implies that $A - 2B$ increases in α_i, then to minimize the expression, we will take the minimal possible value for α_i; otherwise, if the derivative is negative, we will take maximal possible value for α_i. 2) Using step (1), we chose $\alpha_a, \alpha_b, \alpha_c$ such that $A - 2B$ is as

[22] We can express $A - 2B$ as a function of $\alpha_a, \alpha_b, \alpha_c$ and α_d.

small as possible to contradict the inequality; if the inequality still holds, then Proposition 1 is validated.

$$\frac{\partial(A-2B)}{\partial\alpha_b} = -1+\alpha_c\alpha_d+\alpha_a\alpha_d+\alpha_a\alpha_c-2\alpha_a\alpha_c\alpha_d = \alpha_c\alpha_d(1-\alpha_a)+\alpha_a\alpha_c(1-\alpha_d)+\alpha_a\alpha_d-1 \tag{A.22}$$

Let us demonstrate that the expression is nonpositive. Rearranging yields:

$$\alpha_c\alpha_d(1-\alpha_a)+\alpha_a\alpha_c(1-\alpha_d)+\alpha_a\alpha_d-1\leq 0 \Leftrightarrow \alpha_c\alpha_d(1-\alpha_a)+\alpha_a\alpha_c(1-\alpha_d)+\alpha_a\alpha_d\leq 1 \Leftrightarrow$$

$$\Leftrightarrow \alpha_d(\alpha_c(1-\alpha_a)+\alpha_a)+\alpha_a\alpha_c(1-\alpha_d)\leq 1 \Leftrightarrow \alpha_d(\alpha_c+\alpha_a(1-\alpha_c))+\alpha_a\alpha_c(1-\alpha_d)\leq 1 \tag{A.23}$$

To contradict the hypothesis, let us maximize the left-hand side of the last inequality as much as possible. First, let us focus on the coefficient of α_d in the first term is increasing in α_a as $1-\alpha_c$ is positive, thus

$$\alpha_c+\alpha_a(1-\alpha_c)\leq \alpha_c+\alpha_c(1-\alpha_c)=\alpha_c(2-\alpha_c)<1 \tag{A.24}$$

The last inequality is correct because the expression's $\alpha_c(2-\alpha_c)$ maximum is one and attains it when $\alpha_c=1$. Thus, we have

$$\alpha_d(\alpha_c+\alpha_a(1-\alpha_c))+\alpha_a\alpha_c(1-\alpha_d)<\alpha_d+\alpha_a\alpha_c(1-\alpha_d) \tag{A.25}$$

Let us show that the right-hand side is less than one. The expression is increasing in α_a and α_c, therefore

$$\alpha_d+\alpha_a\alpha_c(1-\alpha_d)\leq \alpha_d+\alpha_d^2(1-\alpha_d)<1 \tag{A.26}$$

The last inequality is correct by Observation 2. Therefore, the total expression is decreasing in α_b.

Let us look at the derivative of $A-2B$ with respect to α_c and check the same steps.

$$\frac{\partial(A-2B)}{\partial\alpha_c} = -1+\alpha_a\alpha_d+\alpha_b\alpha_d+\alpha_a\alpha_b-2\alpha_a\alpha_b\alpha_d = \alpha_a\alpha_d(1-\alpha_b)+\alpha_a\alpha_b(1-\alpha_d)+\alpha_b\alpha_d-1 \tag{A.27}$$

Let us demonstrate that the expression is nonpositive. Thus, let us rearrange the terms:

$$\alpha_a\alpha_d(1-\alpha_b)+\alpha_a\alpha_b(1-\alpha_d)+\alpha_b\alpha_d-1\leq 0 \quad \Leftrightarrow \quad \alpha_a\alpha_d(1-\alpha_b)+\alpha_a\alpha_b(1-\alpha_d)+\alpha_b\alpha_d\leq 1$$

$$\Leftrightarrow \alpha_d(\alpha_a(1-\alpha_b)+\alpha_b)+\alpha_a\alpha_b(1-\alpha_d)\leq 1 \tag{A.28}$$

The coefficient of α_d in the first term is increasing in α_a, and $\alpha_b(1-\alpha_b)+\alpha_b\leq 1$. Hence:

$$\alpha_d(\alpha_a(1-\alpha_b)+\alpha_b)+\alpha_a\alpha_b(1-\alpha_d)\leq \alpha_d+\alpha_a\alpha_b(1-\alpha_d)\leq \alpha_d+\alpha_d^2(1-\alpha_d)\leq 1 \tag{A.29}$$

The second inequality is correct because the expression is increasing in α_a and α_b, and the last inequality is correct because of (A.20).

Therefore, we have shown that the entire expression decreases in α_a, α_b, and α_c. To contradict our prediction, to make $A-2B$ as small as possible, let us take the maximum possible values of α_a, α_b, and α_c. Thus, $\alpha_a = \alpha_b = \alpha_c = \alpha_d$. Therefore,

$$A-2B = 4-4\alpha_d+4\alpha_d^3-2-2\alpha_d^4 = 2-4\alpha_d+4\alpha_d^3-2\alpha_d^4 \geq 0 \quad \Leftrightarrow$$

$$\Leftrightarrow (1-\alpha_d)-\alpha_d+\alpha_d^3+\alpha_d^3(1-\alpha_d) = (1-\alpha_d)-\alpha_d(1-\alpha_d)(1+\alpha_d)+\alpha_d^3(1-\alpha_d) \geq 0$$

$$\Leftrightarrow \quad (1-\alpha_d)(1-\alpha_d(1+\alpha_d)+\alpha_d^3) \geq 0 \tag{A.30}$$

The first term in the multiplication is positive as $\alpha_d < 1$. We show that the second term is nonnegative.

$$0 \leq 1-\alpha_d-\alpha_d^2+\alpha_d^3 \Leftrightarrow \alpha_d+\alpha_d^2(1-\alpha_d) \leq 1 \tag{A.31}$$

By Observation 2, the inequality is correct. Therefore, this completes the proof of Proposition 7.

□

Proof of Proposition 2. We consider three feasible team formations: $(\{ab\},\{cd\})$, $(\{ac\},\{bd\})$, and $(\{ad\},\{bc\})$. Each induces a corresponding value of μ, the absolute difference between the two teams' average altruism levels. By Proposition 1, the associated total costs satisfy

$$TC_{(\{ad\},\{bc\})} \leq TC_{(\{ac\},\{bd\})} \leq TC_{(\{ab\},\{cd\})}. \tag{A.32}$$

Therefore, it suffices to show that $\mu_{(\{a,d\}\{b,c\})}$ is weakly minimal. We begin by comparing it to the team formation $(\{a,b\}\{c,d\})$, whose absolute difference in average altruism levels is:

$$\mu_{(\{a,c\}\{b,d\})} = |\frac{\alpha_c+\alpha_d}{2}-\frac{\alpha_a+\alpha_b}{2}| \geq |\frac{\alpha_a+\alpha_d}{2}-\frac{\alpha_b+\alpha_c}{2}| = \mu_{(\{ad\},\{bc\})} \tag{A.33}$$

Rearranging terms yields $2\alpha_c \geq 2\alpha_a$, which is valid since $\alpha_c \geq \alpha_a$. Next, we compare the absolute difference between the two teams' average altruism levels of $(\{a,c\}\{b,d\})$ and $(\{a,b\}\{c,d\})$, and show that $\mu_{(\{a,c\}\{b,d\})}$ is weakly smaller:

$$\mu_{(\{a,b\}\{c,d\})} = |\frac{\alpha_a+\alpha_b}{2}-\frac{\alpha_c+\alpha_d}{2}| \geq |\frac{\alpha_a+\alpha_c}{2}-\frac{\alpha_b+\alpha_d}{2}| = \mu_{(\{a,c\}\{b,d\})} \tag{A.34}$$

Rearranging terms yields $2\alpha_b \geq 2\alpha_a$, which is valid since $\alpha_b \geq \alpha_a$. Therefore, we show that $\mu_{(\{ad\}\{bc\})} \leq \mu_{(\{a,c\}\{b,d\})} \leq \mu_{(\{a,b\}\{c,d\})}$, hence $\mu_{(\{ad\}\{bc\})}$ is weakly minimal.

□

Proof of Proposition 3: Without loss of generality, assume that agents are ordered regarding their other-regarding preferences such that $0 \leq \alpha_1 \leq \alpha_2 \leq ... \leq \alpha_{2N-1} \leq \alpha_{2N} < 1$. Following the structure of the steps of the proposition, each project l is assigned agents $(l, 2N-l+1)$.

Suppose, for contradiction, that this construction is not optimal. Thus, a different pairing strategy exists, which is optimal.

The procedure is as follows: Step 1: Starting from agent 1, check if his pair is $2N$; if yes, check the next agent and his pair and follow the process until we find different pairs from $\{l, 2N-l+1\}$. Step 2: If for some agent j the pair is (j,i) where $j<i<2N-j+1$, first check whether $\alpha_i = \alpha_{2N-j+1}$, if the equality is correct, continue from Step one and look at the next agent $j+1$.

If there are no different pairs from $\{l, 2N-l+1\}$, the pairing strategy is identical to the constructed strategy in the proposition.

If $\alpha_i \neq \alpha_{2N-j+1}$ then choose the pair (j,i), and the pair of agent $2N-j+1$ call it h, thus the pair $\{h, 2N-j+1\}$. Due to our process $j<i, h<2N-j+1$; hence $\alpha_j < \alpha_i, \alpha_h < \alpha_{2N-j+1}$. By Proposition 1 we know that $TC_{(\{j,2N-j+1\}\{i,h\})} \leq TC_{(\{j,i\}\{h,2N-j+1\})}$. Therefore, transforming the pairs $(\{j,i\}\{h,2N-j+1\})$ into $(\{j,2N-j+1\}\{i,h\})$ is (weakly) beneficial to the principal, and therefore the strategy is not optimal. □

APPENDIX B: DISCUSSION

We first derive the minimal compensation required to satisfy participation for a given outside option and relate it to the incentive-compatible contract (Lemma 3 and Proposition 4). Next, we provide the formal proofs used in the uniqueness discussion by establishing the relevant within-team bonus ranking and comparing total payments across candidate partitions under the maintained supermodularity restriction (Lemma 4 and Proposition 5). Finally, we report the algebra supporting the team-size comparison used in the discussion (Proposition 6).

Proof of Lemma 3: In this proof, we derive a minimal bonus that satisfies the participation constraint for a given outside option $\bar{U}$, and compare it to a bonus satisfying the incentive compatibility constraint derived in (6).

$$u_i(w^n_{ij}, w^n_{ji}, b^n_{ij}, b^n_{ji}, 1, 1) = w^n_{ij} + p^n_{11} b^n_{ij} - k + \alpha_i (p^n_{11} b^n_{ji} + w^n_{ji}) \geq \bar{U},\ i \in A, n \in \{1,2\}, \quad \textbf{(PC)}$$

Following the steps in the proof of Lemma 1, we obtain:

$$b^{n\,*}_{ij} = \frac{(\bar{U}+k)(1-\alpha_i)}{p^n_{11}(1-\alpha_i\alpha_j)}$$

Let us compare the optimal bonuses satisfying (IC) and (PC) we have:

$$\begin{aligned} \frac{k(1-\alpha_i)}{(p^n_{11}-p^n_{01})(1-\alpha_i\alpha_j)} &\geq \frac{(\bar{U}+k)(1-\alpha_i)}{p^n_{11}(1-\alpha_i\alpha_j)} \\ \Leftrightarrow \quad p^n_{01} k &\geq \bar{U}(p^n_{11}-p^n_{01}) \end{aligned} \tag{B.1}$$

□

Proof of Proposition 4: The proof follows directly from the analysis provided in the text and the proof of Proposition 1. The only difference relative to Proposition 1 lies in the agents' expected utilities through w' and b_0, which are constants and therefore unaffected by agents' other-regarding preferences. □

Consider two identical projects and suppose that the production function is supermodular: $p_{11}-p_{01}\geq p_{01}-p_{00}$. The objective of this study is to derive a contract and team formation that achieves a unique equilibrium. In order to establish a unique equilibrium, the principal must ensure that one agent adopts a dominant strategy. This means that the agent will work regardless of the actions of their coworker. The other agent's best response is to work whenever the first agent works.

$$\min_{\substack{((ij),(gh))\in T\\ w^n_{ij},w^n_{ji},b^n_{ij},b^n_{ji}\in R}} \quad \sum_{n=1}^{2} p^n_{11}(b^n_{ij}+b^n_{ji})+w^n_{ij}+w^n_{ji}$$

$$\begin{aligned}
\text{s.t } & u_i(w^n_{ij},w^n_{ji},b^n_{ij},b^n_{ji},1,0)\geq u_i(w^n_{ij},w^n_{ji},b^n_{ij},b^n_{ji},0,0), i\neq j, n\in\{1,2\}, \textbf{ (IC)}_i\\
& u_j(w^n_{ji},w^n_{ij},b^n_{ji},b^n_{ij},1,1)\geq u_j(w^n_{ji},w^n_{ij},b^n_{ji},b^n_{ij},1,0), j\neq i, n\in\{1,2\}, \textbf{ (IC)}_j\\
& u_i(w^n_{ij},w^n_{ji},b^n_{ij},b^n_{ji},1,1)\geq 0, \forall i\in A, n\in\{1,2\}, \textbf{ (PC)}_i\\
& u_j(w^n_{ji},w^n_{ij},b^n_{ji},b^n_{ij},1,1)\geq 0, \forall j\in A, n\in\{1,2\}, \textbf{ (PC)}_j\\
& w^n_{ij}+b^n_{ij}\geq 0, w^n_{ij}\geq 0, \forall i\in A, n\in\{1,2\}, \textbf{ (LLC)}
\end{aligned}$$

The incentive constraint for agent i says that he prefers to exert effort rather than shirk given that his coworker shirks. The incentive constraint for agent j states that agent j prefers to exert effort rather than shirk given his coworker works. From (IC)s, we obtain

$$b^n_{ij}\geq \frac{k}{(1-\alpha_i\alpha_j)(p^n_{11}-p^n_{01})}\left(\frac{p^n_{11}-p^n_{01}}{p^n_{01}-p^n_{00}}-\alpha_i\right) \tag{B.2}$$

$$b^n_{ji}\geq \frac{k}{(1-\alpha_i\alpha_j)(p^n_{01}-p^n_{00})}\left(\frac{p^n_{01}-p^n_{00}}{p^n_{11}-p^n_{01}}-\alpha_j\right) \tag{B.3}$$

To solve the linear program, we drop the (PC)s and derive the optimal bonus payments from the (IC)s and (LLC)s. The incentive constraints for agents i and j are binding. Thus, $b^{n}_{ij}{}^{*}=\frac{k}{(1-\alpha_i\alpha_j)(p^n_{11}-p^n_{01})}\left(\frac{p^n_{11}-p^n_{01}}{p^n_{01}-p^n_{00}}-\alpha_i\right)$ and $b^{n}_{ji}{}^{*}=\frac{k}{(1-\alpha_i\alpha_j)(p^n_{01}-p^n_{00})}\left(\frac{p^n_{01}-p^n_{00}}{p^n_{11}-p^n_{01}}-\alpha_j\right)$; To satisfy (LLC)s and minimize the principal's total expenditure, the fixed wage for both agents' contracts is set to zero. Let us demonstrate that the derived contracts $(w^*_{ij},b^*_{ij}),(w^*_{ji},b^*_{ji})$ satisfy the participation constraint. Indeed, for agent j

$$w^*_{ji}+p^n_{11}b^*_{ji}-k+\alpha_j(w^*_{ij}+p^n_{11}b^*_{ij})=p^n_{11}b^*_{ji}+\alpha_j p^n_{11}b^*_{ij}-k\geq 0$$

$$\Leftrightarrow \frac{p^n_{11}k}{p^n_{11}-p^n_{01}}-k=\frac{p^n_{01}k}{p^n_{11}-p^n_{01}}\geq 0 \tag{B.4}$$

Since agent i receives a higher bonus than agent j, the participation constraint for agent i is also satisfied.

Proof of Lemma 4: We compare the principal's total expected costs across two scenarios. In the first scenario, agent i receives a higher bonus, whereas in the second scenario, agent i receives a lower bonus. We assume $0 \leq \alpha_j \leq \alpha_i < 1$. For each team formation, we compute the principal's total expected cost. The subscript "h" denotes the case in which the agent receives the higher bonus.

$$b_{ji}^{*\,h} + b_{ij}^{*} = \frac{k}{1-\alpha_i\alpha_j}\Big(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_i}{p_{01}-p_{00}} + \frac{1}{p_{01}-p_{00}} - \frac{\alpha_j}{p_{11}-p_{01}}\Big) \tag{B.5}$$

$$b_{ij}^{*\,h} + b_{ji}^{*} = \frac{k}{1-\alpha_i\alpha_j}\Big(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_j}{p_{01}-p_{00}} + \frac{1}{p_{01}-p_{00}} - \frac{\alpha_i}{p_{11}-p_{01}}\Big) \tag{B.6}$$

We show that the first expression, corresponding to the case in which agent j receives the higher bonus, is lower than the second. To see this, observe that

$$b_{ij}^{*\,h} + b_{ji}^{*} \leq b_{ji}^{*\,h} + b_{ij}^{*}$$

$$\Leftrightarrow \frac{1}{p_{11}-p_{01}} - \frac{\alpha_i}{p_{01}-p_{00}} + \frac{1}{p_{01}-p_{00}} - \frac{\alpha_j}{p_{11}-p_{01}} \leq \frac{1}{p_{11}-p_{01}} - \frac{\alpha_j}{p_{01}-p_{00}} + \frac{1}{p_{01}-p_{00}} - \frac{\alpha_i}{p_{11}-p_{01}}$$

$$\Leftrightarrow \alpha_j\Big(\frac{1}{p_{01}-p_{00}} - \frac{1}{p_{11}-p_{01}}\Big) \leq \alpha_i\Big(\frac{1}{p_{01}-p_{00}} - \frac{1}{p_{11}-p_{01}}\Big)$$

$$\Leftrightarrow \alpha_j \leq \alpha_i. \tag{B.7}$$

The last equivalence comes from the supermodularity of the production function.

□

Proof of Proposition 5: The proof proceeds as follows. For each candidate team formation, we first determine the bonus assigned to each agent. We then compare two team formations, say T and T', by computing, agent by agent, the change in bonuses when moving from T' to T. We separate these changes into positive and negative components. If the total increase in bonuses under T (relative to T') exceeds the total decrease, then the principal's objective is higher under T than under T'. In that case, we conclude that team formation T is superior for the principal.

By Lemma 4, agent a, whose altruism parameter α_a is minimal, receives a higher bonus under any team formation, whereas agent d, whose altruism parameter α_d is maximal, receives a lower bonus. Thus, there are three possible team formations: $T_{(\{a,d\}\{b,c\})}$, $T_{(\{a,c\}\{b,d\})}$, and $T_{(\{a,b\}\{c,d\})}$.

First, we show that $T_{(\{a,c\}\{b,d\})} \leq T_{(\{a,b\}\{c,d\})}$. We then complete the proof by showing that $T_{(\{a,d\}\{b,c\})} \leq T_{(\{a,c\}\{b,d\})}$.

$$T_{(\{a,c\}\{b,d\})} \leq T_{(\{a,b\}\{c,d\})} \quad \Longleftrightarrow \quad b_{ac} + b_{ca} + b_{bd} + b_{db} \leq b_{ab} + b_{ba} + b_{cd} + b_{cb}. \tag{B.8}$$

Above in (B.2) and (B.3), we have derived bonuses for any possible team formation. In particular, bonuses under $T_{(\{a,c\}\{b,d\})}$ and $T_{(\{a,b\}\{c,d\})}$ are:

$$b_{ac} = \frac{1}{1-\alpha_a\alpha_c}\left(\frac{1}{p_{01}-p_{00}} - \frac{\alpha_a}{p_{11}-p_{01}}\right), \quad b_{ca} = \frac{1}{1-\alpha_a\alpha_c}\left(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_c}{p_{01}-p_{00}}\right),$$
$$b_{bd} = \frac{1}{1-\alpha_b\alpha_d}\left(\frac{1}{p_{01}-p_{00}} - \frac{\alpha_d}{p_{11}-p_{01}}\right), \quad b_{db} = \frac{1}{1-\alpha_b\alpha_d}\left(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_d}{p_{01}-p_{00}}\right),$$
$$b_{ab} = \frac{1}{1-\alpha_a\alpha_b}\left(\frac{1}{p_{01}-p_{00}} - \frac{\alpha_a}{p_{11}-p_{01}}\right), \quad b_{ba} = \frac{1}{1-\alpha_a\alpha_b}\left(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_b}{p_{01}-p_{00}}\right),$$
$$b_{cd} = \frac{1}{1-\alpha_c\alpha_d}\left(\frac{1}{p_{01}-p_{00}} - \frac{\alpha_c}{p_{11}-p_{01}}\right), \quad b_{dc} = \frac{1}{1-\alpha_c\alpha_d}\left(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_d}{p_{01}-p_{00}}\right),$$

Inequality (B.8) can be restated as:

$$\underbrace{(b_{ba}-b_{ca})}_{(i)} - \underbrace{(b_{ac}-b_{ab})}_{(ii)} \geq \underbrace{(b_{bd}-b_{cd})}_{(iii)} - \underbrace{(b_{dc}-b_{db})}_{(iv)} \tag{B.9}$$

$$\begin{aligned}
(i): \quad b_{ba}-b_{ca} &= \frac{1}{1-\alpha_a\alpha_b}\left(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_b}{p_{01}-p_{00}}\right) - \frac{1}{1-\alpha_a\alpha_c}\left(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_c}{p_{01}-p_{00}}\right) \\
&= \frac{(\alpha_c-\alpha_b)}{(1-\alpha_a\alpha_b)(1-\alpha_a\alpha_c)}\left(\frac{1}{p_{01}-p_{00}} - \frac{\alpha_a}{p_{11}-p_{01}}\right), \\
(ii): \quad b_{ac}-b_{ab} &= \frac{1}{1-\alpha_a\alpha_c}\left(\frac{1}{p_{01}-p_{00}} - \frac{\alpha_a}{p_{11}-p_{01}}\right) - \frac{1}{1-\alpha_a\alpha_b}\left(\frac{1}{p_{01}-p_{00}} - \frac{\alpha_a}{p_{11}-p_{01}}\right) \\
&= \frac{\alpha_a(\alpha_c-\alpha_b)}{(1-\alpha_a\alpha_b)(1-\alpha_a\alpha_c)}\left(\frac{1}{p_{01}-p_{00}} - \frac{\alpha_a}{p_{11}-p_{01}}\right), \\
(iii): \quad b_{bd}-b_{cd} &= \frac{1}{1-\alpha_b\alpha_d}\left(\frac{1}{p_{01}-p_{00}} - \frac{\alpha_b}{p_{11}-p_{01}}\right) - \frac{1}{1-\alpha_c\alpha_d}\left(\frac{1}{p_{01}-p_{00}} - \frac{\alpha_c}{p_{11}-p_{01}}\right) \\
&= \frac{(\alpha_c-\alpha_b)}{(1-\alpha_a\alpha_d)(1-\alpha_c\alpha_d)}\left(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_d}{p_{01}-p_{00}}\right). \\
(iv): \quad b_{dc}-b_{db} &= \frac{1}{1-\alpha_c\alpha_d}\left(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_d}{p_{01}-p_{00}}\right) - \frac{1}{1-\alpha_b\alpha_d}\left(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_d}{p_{01}-p_{00}}\right) \\
&= \frac{\alpha_d(\alpha_c-\alpha_b)}{(1-\alpha_b\alpha_d)(1-\alpha_c\alpha_d)}\left(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_d}{p_{01}-p_{00}}\right).
\end{aligned}$$

First, let us take pairwise differences $(i)-(ii)$ and $(iii)-(iv)$, and next compare these two differences.

$$\begin{aligned}
(i)-(ii) &= \frac{(\alpha_c-\alpha_b)(1-\alpha_a)}{(1-\alpha_a\alpha_b)(1-\alpha_a\alpha_c)}\left(\frac{1}{p_{01}-p_{00}} - \frac{\alpha_a}{p_{11}-p_{01}}\right), \\
(iii)-(iv) &= \frac{(\alpha_c-\alpha_b)(1-\alpha_d)}{(1-\alpha_b\alpha_d)(1-\alpha_c\alpha_d)}\left(\frac{1}{p_{11}-p_{01}} - \frac{\alpha_d}{p_{01}-p_{00}}\right).
\end{aligned}$$

Next we demonstrate that $(i)-(ii) \geq (iii)-(iv)$, indeed:

$$\frac{(\alpha_c-\alpha_b)(1-\alpha_a)}{(1-\alpha_a\alpha_b)(1-\alpha_a\alpha_c)}\left(\frac{1}{p_{01}-p_{00}}-\frac{\alpha_a}{p_{11}-p_{01}}\right) \geq \frac{(\alpha_c-\alpha_b)(1-\alpha_d)}{(1-\alpha_b\alpha_d)(1-\alpha_c\alpha_d)}\left(\frac{1}{p_{11}-p_{01}}-\frac{\alpha_d}{p_{01}-p_{00}}\right) \Leftrightarrow$$

$$\begin{aligned}\Leftrightarrow (1-\alpha_a)(1-\alpha_b\alpha_d)(1-\alpha_c\alpha_d)\left(\frac{1}{p_{01}-p_{00}}-\frac{\alpha_a}{p_{11}-p_{01}}\right) \geq \\ \geq (1-\alpha_d)(1-\alpha_a\alpha_b)(1-\alpha_a\alpha_c)\left(\frac{1}{p_{11}-p_{01}}-\frac{\alpha_d}{p_{01}-p_{00}}\right). \qquad \text{(B.10)}\end{aligned}$$

Rearranging terms:

$$\begin{aligned}\Big((1-\alpha_a)(1-\alpha_b\alpha_d)(1-\alpha_c\alpha_d)+\alpha_d(1-\alpha_a\alpha_b)(1-\alpha_a\alpha_c)\Big)(p_{11}-p_{01}) \geq \\ \geq \Big(\alpha_a(1-\alpha_b\alpha_d)(1-\alpha_c\alpha_d)+(1-\alpha_a\alpha_b)(1-\alpha_a\alpha_c)(1-\alpha_d)\Big)(p_{01}-p_{00}). \qquad \text{(B.11)}\end{aligned}$$

Left polynomial (the part multiplying $p_{11}-p_{01}$)

$$\begin{aligned}(1-\alpha_a)(1-\alpha_b\alpha_d)(1-\alpha_c\alpha_d)+\alpha_d(1-\alpha_a\alpha_b)(1-\alpha_a\alpha_c) = \\ = 1+(\alpha_d-\alpha_a)\big(1-\alpha_a\alpha_b\alpha_c\alpha_d\big)-(\alpha_b\alpha_d+\alpha_c\alpha_d)\big(1-\alpha_a\alpha_d\big)+\alpha_b\alpha_c\alpha_d^2(1-\alpha_a^2). \qquad \text{(B.12)}\end{aligned}$$

Right polynomial (the part multiplying $p_{01}-p_{00}$)

$$\begin{aligned}\alpha_a(1-\alpha_b\alpha_d)(1-\alpha_c\alpha_d)+(1-\alpha_a\alpha_b)(1-\alpha_a\alpha_c)(1-\alpha_d) = \\ = 1-(\alpha_d-\alpha_a)\big(1-\alpha_a\alpha_b\alpha_c\alpha_d\big)-(\alpha_a\alpha_b+\alpha_a\alpha_c)\big(1-\alpha_a\alpha_d\big)+\alpha_a^2\alpha_b\alpha_c(1-\alpha_d^2). \qquad \text{(B.13)}\end{aligned}$$

By supermodularity, $p_{11}-p_{01} \geq p_{01}-p_{00}$, it suffices to show that the difference between the left-hand polynomial and the right-hand polynomial is positive.

$$\begin{aligned}2(\alpha_d-\alpha_a)\big(1-\alpha_a\alpha_b\alpha_c\alpha_d\big)+\alpha_b\alpha_c\alpha_d^2(1-\alpha_a^2)- \\ -\alpha_a^2\alpha_b\alpha_c(1-\alpha_d^2)-(\alpha_b\alpha_d+\alpha_c\alpha_d-\alpha_a\alpha_b-\alpha_a\alpha_c)\big(1-\alpha_a\alpha_d\big) = \\ = 2(\alpha_d-\alpha_a)\big(1-\alpha_a\alpha_b\alpha_c\alpha_d\big)+\alpha_b\alpha_c(\alpha_d^2-\alpha_a^2)-\big(1-\alpha_a\alpha_d\big)(\alpha_d-\alpha_a)(\alpha_b+\alpha_c) = \\ = (\alpha_d-\alpha_a)(2-(1-\alpha_a\alpha_d)(\alpha_b+\alpha_c)+\alpha_a\alpha_b\alpha_c(1-\alpha_d)+\alpha_b\alpha_c\alpha_d(1-\alpha_a)) \geq 0 \qquad \text{(B.14)}\end{aligned}$$

The first part is nonnegative, in the second part last two terms are positive, and the difference between $2-(1-\alpha_a\alpha_d)(\alpha_b+\alpha_c)$ is also positive since $0 \leq \alpha_a,\alpha_b,\alpha_c,\alpha_d \leq 1$. Therefore, $T_{(\{a,c\}\{b,d\})} \leq T_{(\{a,b\}\{c,d\})}$.

Next, let us compare the following

$$T_{(\{a,d\}\{b,c\})} \leq T_{(\{a,c\}\{b,d\})} \quad \Longleftrightarrow \quad b_{ad}+b_{da}+b_{bc}+b_{cb} \leq b_{ac}+b_{ca}+b_{bd}+b_{db}. \qquad \text{(B.15)}$$

Bonuses under team formations $T_{(\{a,c\}\{b,d\})}$ and $T_{(\{a,d\}\{b,c\})}$ are:

$$b_{ac}=\frac{1}{1-\alpha_a\alpha_c}\left(\frac{1}{p_{01}-p_{00}}-\frac{\alpha_a}{p_{11}-p_{01}}\right),\quad b_{ca}=\frac{1}{1-\alpha_a\alpha_c}\left(\frac{1}{p_{11}-p_{01}}-\frac{\alpha_c}{p_{01}-p_{00}}\right),$$
$$b_{ad}=\frac{1}{1-\alpha_a\alpha_d}\left(\frac{1}{p_{01}-p_{00}}-\frac{\alpha_a}{p_{11}-p_{01}}\right),\quad b_{da}=\frac{1}{1-\alpha_a\alpha_d}\left(\frac{1}{p_{11}-p_{01}}-\frac{\alpha_d}{p_{01}-p_{00}}\right),$$
$$b_{bc}=\frac{1}{1-\alpha_b\alpha_c}\left(\frac{1}{p_{01}-p_{00}}-\frac{\alpha_b}{p_{11}-p_{01}}\right),\quad b_{cb}=\frac{1}{1-\alpha_b\alpha_c}\left(\frac{1}{p_{11}-p_{01}}-\frac{\alpha_c}{p_{01}-p_{00}}\right),$$
$$b_{bd}=\frac{1}{1-\alpha_b\alpha_d}\left(\frac{1}{p_{01}-p_{00}}-\frac{\alpha_b}{p_{11}-p_{01}}\right),\quad b_{db}=\frac{1}{1-\alpha_b\alpha_d}\left(\frac{1}{p_{11}-p_{01}}-\frac{\alpha_d}{p_{01}-p_{00}}\right).$$

Equivalently to (B.15) we compare:

$$\underbrace{(b_{db}-b_{da})}_{(i')}+\underbrace{(b_{bd}-b_{bc})}_{(ii')}\geq\underbrace{(b_{cb}-b_{ca})}_{(iii')}+\underbrace{(b_{ad}-b_{ac})}_{(iv')}. \tag{B.16}$$

$$(i'):\quad b_{db}-b_{da}=\frac{1}{1-\alpha_b\alpha_d}\left(\frac{1}{p_{11}-p_{01}}-\frac{\alpha_d}{p_{01}-p_{00}}\right)-\frac{1}{1-\alpha_a\alpha_d}\left(\frac{1}{p_{11}-p_{01}}-\frac{\alpha_d}{p_{01}-p_{00}}\right)$$
$$=\frac{\alpha_d(\alpha_b-\alpha_a)}{(1-\alpha_a\alpha_d)(1-\alpha_b\alpha_d)}\left(\frac{1}{p_{11}-p_{01}}-\frac{\alpha_d}{p_{01}-p_{00}}\right),$$
$$(ii'):\quad b_{bd}-b_{bc}=\frac{1}{1-\alpha_b\alpha_d}\left(\frac{1}{p_{01}-p_{00}}-\frac{\alpha_b}{p_{11}-p_{01}}\right)-\frac{1}{1-\alpha_b\alpha_c}\left(\frac{1}{p_{01}-p_{00}}-\frac{\alpha_b}{p_{11}-p_{01}}\right)$$
$$=\frac{\alpha_b(\alpha_d-\alpha_c)}{(1-\alpha_b\alpha_d)(1-\alpha_b\alpha_c)}\left(\frac{1}{p_{01}-p_{00}}-\frac{\alpha_b}{p_{11}-p_{01}}\right),$$
$$(iii'):\quad b_{cb}-b_{ca}=\frac{1}{1-\alpha_b\alpha_c}\left(\frac{1}{p_{11}-p_{01}}-\frac{\alpha_c}{p_{01}-p_{00}}\right)-\frac{1}{1-\alpha_a\alpha_c}\left(\frac{1}{p_{11}-p_{01}}-\frac{\alpha_c}{p_{01}-p_{00}}\right)$$
$$=\frac{\alpha_c(\alpha_b-\alpha_a)}{(1-\alpha_a\alpha_c)(1-\alpha_b\alpha_c)}\left(\frac{1}{p_{11}-p_{01}}-\frac{\alpha_c}{p_{01}-p_{00}}\right),$$
$$(iv'):\quad b_{ad}-b_{ac}=\frac{1}{1-\alpha_a\alpha_d}\left(\frac{1}{p_{01}-p_{00}}-\frac{\alpha_a}{p_{11}-p_{01}}\right)-\frac{1}{1-\alpha_a\alpha_c}\left(\frac{1}{p_{01}-p_{00}}-\frac{\alpha_a}{p_{11}-p_{01}}\right)$$
$$=\frac{\alpha_a(\alpha_d-\alpha_c)}{(1-\alpha_a\alpha_d)(1-\alpha_a\alpha_c)}\left(\frac{1}{p_{01}-p_{00}}-\frac{\alpha_a}{p_{11}-p_{01}}\right).$$

First, let us take pairwise differences $(i')-(iii')$ and $(ii')-(iv')$, and next look at total sum of these two differences.

$$(i')-(iii')=$$
$$=\frac{(\alpha_b-\alpha_a)(\alpha_d-\alpha_c)}{(1-\alpha_a\alpha_c)(1-\alpha_b\alpha_c)(1-\alpha_a\alpha_d)(1-\alpha_b\alpha_d)}\left(\frac{1-\alpha_a\alpha_b\alpha_c\alpha_d}{p_{11}-p_{01}}-\frac{\alpha_d+\alpha_c-\alpha_c\alpha_d(\alpha_b+\alpha_a)}{p_{01}-p_{00}}\right),$$
$$(ii')-(iv')=$$

$$= \frac{(\alpha_d - \alpha_c)(\alpha_b - \alpha_a)}{(1-\alpha_a\alpha_c)(1-\alpha_b\alpha_c)(1-\alpha_a\alpha_d)(1-\alpha_b\alpha_d)} \left(\frac{1-\alpha_a\alpha_b\alpha_c\alpha_d}{p_{01}-p_{00}} - \frac{\alpha_a+\alpha_b-\alpha_a\alpha_b(\alpha_d+\alpha_c)}{p_{11}-p_{01}} \right),$$

Denote:

$$x \equiv (1-\alpha_a\alpha_c)(1-\alpha_b\alpha_c)(1-\alpha_a\alpha_d)(1-\alpha_b\alpha_d) > 0$$

Therefore, taking the sum of the two differences identified above, should be nonnegative, thus $(i') + (ii') - (iii') - (iv') \geq 0$ if the following is nonnegative

$$\frac{(\alpha_b - \alpha_a)(\alpha_d - \alpha_c)}{x} \left[\frac{1-\alpha_a-\alpha_b+\alpha_a\alpha_b(\alpha_c+\alpha_d)-\alpha_a\alpha_b\alpha_c\alpha_d}{p_{11}-p_{01}} + \right.$$
$$\left. + \frac{1-\alpha_d-\alpha_c+\alpha_c\alpha_d(\alpha_a+\alpha_b)-\alpha_a\alpha_b\alpha_c\alpha_d}{p_{01}-p_{00}} \right]. \quad \text{(B.17)}$$

The first term is nonnegative because $\alpha_a \leq \alpha_b \leq \alpha_c \leq \alpha_d$. And the second factor is nonnegative, as implied by Proposition 1 and Assumption 1. Therefore, we show that $T_{(\{a,d\}\{b,c\})} \leq T_{(\{a,c\}\{b,d\})} \leq T_{(\{a,b\}\{c,d\})}$, by transitivity, $T_{(\{a,d\}\{b,c\})} \leq T_{(\{a,b\}\{c,d\})}$, which concludes our proof. □

Proof of Proposition 6: First, note that if at least one agent is altruistic, implementing four projects with each agent working alone is suboptimal. In that case, an altruistic agent works in isolation, and the positive incentive effect arising from coworkers' monetary payoffs is lost.

Moreover, if at most one agent is selfish, implementing three projects is also not optimal. With three projects, one project must be staffed by two agents while the remaining two projects are staffed by single agents. Since there are at least three altruistic agents in this case, at least one altruistic agent must work alone, again eliminating the positive incentive effect from coworkers' monetary payoffs.

Next, consider the configuration in which one agent works alone, and the remaining three agents form a team. We show that this configuration is also not optimal. To this end, we first derive the optimal wages for a three-agent team. Under binding incentive-compatibility constraints, we obtain the following system:

$$\begin{cases} p_{111}b_{ijg} - k + p_{111}\dfrac{\alpha_i}{2}(b_{jig}+b_{gij}) = p_{011}b_i + p_{011}\dfrac{\alpha_i}{2}(b_{jig}+b_{gij}), \\ p_{111}b_{jig} - k + p_{111}\dfrac{\alpha_j}{2}(b_{ijg}+b_{gij}) = p_{011}b_i + p_{011}\dfrac{\alpha_k}{2}(b_{ijg}+b_{gij}), \\ p_{111}b_{gij} - k + p_{111}\dfrac{\alpha_g}{2}(b_{ijg}+b_{jig}) = p_{011}b_i + p_{011}\dfrac{\alpha_i}{2}(b_{ijg}+b_{jig}). \end{cases}$$

Rearranging terms yields:

$$\begin{cases} b_{ijg} = \dfrac{k}{p_{111} - p_{011}} - \dfrac{\alpha_i}{2}\,(b_{jig} + b_{gij}), \\ b_{jig} = \dfrac{k}{p_{111} - p_{011}} - \dfrac{\alpha_j}{2}\,(b_{ijg} + b_{gij}), \\ b_{gij} = \dfrac{k}{p_{111} - p_{011}} - \dfrac{\alpha_k}{2}\,(b_{ijg} + b_{jig}). \end{cases}$$

Note that a higher average bonus paid to an agent's coworkers raises the utility the agent derives from their payoffs; therefore, the principal can induce the agent to work with a lower own bonus.

By Assumption 2 $b_0 = \dfrac{k}{p_{11} - p_{01}} = \dfrac{k}{p_{111} - p_{011}}$. Given this, the solution of the system is as follows:

$$\begin{cases} b^*_{ijg} = b_0 \dfrac{\big(4 - 4a_i + a_i a_j + a_i a_g - a_j a_g\big)}{\big(4 - a_i a_j - a_i a_g - a_j a_g + a_i a_j a_g\big)}, \\ b^*_{jig} = b_0 \dfrac{\big(4 - 4a_j + a_i a_j - a_i a_g + a_j a_g\big)}{\big(4 - a_i a_j - a_i a_g - a_j a_g + a_i a_j a_g\big)}, \\ b^*_{gij} = b_0 \dfrac{\big(4 - 4a_g - a_i a_j + a_i a_g + a_j a_g\big)}{\big(4 - a_i a_j - a_i a_g - a_j a_g + a_i a_j a_g\big)}. \end{cases}$$

The team structure in which agent a works alone while the remaining agents form a team is not optimal. If agent a is instead teamed up with agent d to the team, the required bonus for a decreases. Moreover, since α_a is minimal, including a raises the team's average bonus and thereby increases the benefit that the other agents derive from their coworkers' bonuses; this, in turn, reduces the bonuses required to induce effort from the remaining agents.

Next, consider the configuration in which agent b works alone and the remaining agents form a team. We show that the team formation $(\{a,c\}\{b,d\})$ is weakly less costly.

When agent b is paired with agent a, the bonus required to induce c to work is weakly lower than when c is grouped with agents a and d. The reason is that agent d requires a lower bonus than agent a, so the average bonus of $\{a,d\}$ generates less utility for b than the bonus of a alone. An analogous argument applies to agent a.

Therefore, it suffices to show that the sum of the bonuses for agents c and d when c works alone exceeds the sum of their bonuses when c and d work together.

$$b^*_{dac} + 1 \geq b^*_{bd} + b^*_{db} \quad \Leftrightarrow$$

$$\Leftrightarrow \quad b_0\big(\frac{4(1-\alpha_d) + 2\alpha_a\alpha_d + \alpha_c\alpha_d - \alpha_a\alpha_c}{4 - \alpha_a\alpha_d - \alpha_c\alpha_d - \alpha_a\alpha_c(1-\alpha_d)} + 1\big) \geq b_0 \frac{2 - \alpha_b - \alpha_d}{1 - \alpha_b\alpha_d} \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \frac{4(1-\alpha_d) + 2\alpha_a\alpha_d + \alpha_c\alpha_d - \alpha_a\alpha_c + \Big(4 - \alpha_a\alpha_d - \alpha_c\alpha_d - \alpha_a\alpha_c(1-\alpha_d)\Big)}{4 - \alpha_a\alpha_d - \alpha_c\alpha_d - \alpha_a\alpha_c(1-\alpha_d)} \geq \frac{2 - \alpha_b - \alpha_d}{1 - \alpha_b\alpha_d} \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \frac{(2-\alpha_d)(4-\alpha_a\alpha_c)}{4-\alpha_a\alpha_c-\alpha_a\alpha_d-\alpha_c\alpha_d+\alpha_a\alpha_c\alpha_d} \geq \frac{2-\alpha_b-\alpha_d}{1-\alpha_b\alpha_d} \tag{B.18}$$

Rearranging terms, we obtain:

$$(2-\alpha_d)(4-\alpha_a\alpha_c)(1-\alpha_b\alpha_d) \geq \big(2-\alpha_b-\alpha_d\big)\big(4-\alpha_a\alpha_c-\alpha_a\alpha_d-\alpha_c\alpha_d+\alpha_a\alpha_c\alpha_d\big) \Leftrightarrow$$

$$\Leftrightarrow \quad \alpha_b(4-\alpha_a\alpha_c)(1-2\alpha_d+\alpha_d^2) \geq (2-\alpha_b-\alpha_d)\big(-\alpha_a\alpha_d-\alpha_c\alpha_d+\alpha_a\alpha_c\alpha_d\big) \Leftrightarrow$$

$$\Leftrightarrow \quad \alpha_b(4-\alpha_a\alpha_c)(1-\alpha_d)^2 \geq -(2-\alpha_b-\alpha_d)\big(\alpha_a\alpha_d+\alpha_c\alpha_d(1-\alpha_a)\big) \tag{B.19}$$

The left-hand side is nonnegative because $\alpha_b \geq 0$. The right-hand side is nonpositive because $0 \leq \alpha_b, \alpha_d < 1$ and $0 < \alpha_a < 1$. Hence, the team formation $(\{a,c\}\{b,d\})$ is weakly less costly than the configuration in which agent b works alone and the remaining agents form a team.

We proceed analogously by considering the case in which agent c works alone and the remaining agents form a team, and we show that $(\{a,c\}\{b,d\})$ is weakly less costly.

As before, when agent c is paired with agent a, the bonus needed to induce c to work is weakly lower than when c is in the three-agent team $\{a,d\}$. Since agent d requires a lower bonus than agent a, the average bonus in $\{a,d\}$ is lower, so c derives less utility from coworkers' payoffs and must be compensated more. The same reasoning applies to agent a.

Now, it suffices to show that the sum of the bonuses for agents c and d when c works alone exceeds the sum of their bonuses when c and d work together.

$$b_{dab}+1 \geq b_{cd}+b_{dc} = b_0\frac{2-\alpha_c-\alpha_d}{1-\alpha_c\alpha_d}, \quad \Leftrightarrow$$

$$\Leftrightarrow \quad b_0\Big(\frac{4(1-\alpha_d)+2\alpha_a\alpha_d+\alpha_b\alpha_d-\alpha_a\alpha_b}{4-\alpha_a\alpha_d-\alpha_b\alpha_d-\alpha_a\alpha_b(1-\alpha_d)}+1\Big) \geq b_0\frac{2-\alpha_c-\alpha_d}{1-\alpha_c\alpha_d} \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \frac{4(2-\alpha_d)-\alpha_a\alpha_b(2-\alpha_d)}{4-\alpha_a\alpha_d-\alpha_b\alpha_d-\alpha_a\alpha_b(1-\alpha_d)} \geq \frac{2-\alpha_c-\alpha_d}{1-\alpha_c\alpha_d} \tag{B.20}$$

Rearranging terms, we obtain:

$$(4-\alpha_a\alpha_b)(2-\alpha_d)(1-\alpha_c\alpha_d) \geq (2-\alpha_c-\alpha_d)\big(4-\alpha_a\alpha_b-\alpha_a\alpha_d-\alpha_b\alpha_d+\alpha_a\alpha_b\alpha_d\big) \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \alpha_c(4-\alpha_a\alpha_b)-\alpha_c\alpha_d(4-\alpha_a\alpha_b)(2-\alpha_d) \geq -(2-\alpha_c-\alpha_d)\big(\alpha_a\alpha_d+\alpha_b\alpha_d(1-\alpha_a)\big) \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \alpha_c(4-\alpha_a\alpha_b)\,(1-2\alpha_d+\alpha_d^2) \geq -(2-\alpha_c-\alpha_d)\big(\alpha_a\alpha_d+\alpha_b\alpha_d(1-\alpha_a)\big) \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \alpha_c(4-\alpha_a\alpha_b)\,(1-\alpha_d)^2 \geq -(2-\alpha_c-\alpha_d)\big(\alpha_a\alpha_d+\alpha_b\alpha_d(1-\alpha_a)\big). \tag{B.21}$$

The left-hand side is nonnegative because $\alpha_c \geq 0$. The right-hand side is nonpositive because $0 \leq \alpha_a, \alpha_c < 1$ and $0 \leq \alpha_a < 1$. Hence, the team formation $(\{a,c\}\{b,d\})$ is weakly

less costly than the configuration in which agent b works alone and the remaining agents form a team.

The last case is analogous to the previous one. Consider the configuration where agent d works alone and the remaining agents form a three-person team. We show that the team formation $(\{a,b\}\{c,d\})$ is less costly. The same argument is relevant regarding agents a and b's bonuses.

Thus, it suffices to show that the sum of the bonuses for agents c and d when d works alone exceeds the sum of their bonuses when c and d work together.

$$b_{cab}+1 \geq b_{cd}+b_{dc} = b_0\frac{2-\alpha_c-\alpha_d}{1-\alpha_c\alpha_d}, \quad \Leftrightarrow$$

$$\Leftrightarrow \quad b_0\big(\frac{4(1-\alpha_c)+2\alpha_a\alpha_c+\alpha_b\alpha_c-\alpha_a\alpha_b}{4-\alpha_a\alpha_c-\alpha_b\alpha_c-\alpha_a\alpha_b(1-\alpha_c)}+1\big) \geq b_0\frac{2-\alpha_c-\alpha_d}{1-\alpha_c\alpha_d} \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \frac{4(2-\alpha_c)-\alpha_a\alpha_b(2-\alpha_c)}{4-\alpha_a\alpha_c-\alpha_b\alpha_c-\alpha_a\alpha_b(1-\alpha_c)} \geq \frac{2-\alpha_c-\alpha_d}{1-\alpha_c\alpha_d} \tag{B.22}$$

Rearranging terms yields:

$$(4-\alpha_a\alpha_b)(2-\alpha_c)(1-\alpha_c\alpha_d) \geq (2-\alpha_c-\alpha_d)\big(4-\alpha_a\alpha_b-\alpha_a\alpha_c-\alpha_b\alpha_c+\alpha_a\alpha_b\alpha_c\big) \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \alpha_c(4-\alpha_a\alpha_b)-\alpha_c\alpha_d(4-\alpha_a\alpha_b)(2-\alpha_d) \geq -(2-\alpha_c-\alpha_d)\big(\alpha_a\alpha_c+\alpha_b\alpha_c(1-\alpha_a)\big) \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \alpha_d(4-\alpha_a\alpha_b)\,(1-2\alpha_c+\alpha_c^2) \geq -(2-\alpha_c-\alpha_d)\big(\alpha_a\alpha_c+\alpha_b\alpha_c(1-\alpha_a)\big) \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \alpha_d(4-\alpha_a\alpha_b)\,(1-\alpha_c)^2 \geq -(2-\alpha_c-\alpha_d)\big(\alpha_a\alpha_c+\alpha_b\alpha_c(1-\alpha_a)\big). \tag{B.23}$$

The left-hand side is positive because $\alpha_d > 0$ and $\alpha_c < 1$. The right-hand side is non-positive since $0 < \alpha_c, \alpha_d < 1$ and $0 < \alpha_a < 1$. Hence, the team formation $(\{a,c\}\{b,d\})$ is weakly less costly than the configuration in which agent b works alone and the remaining agents form a team.

Therefore, it is always optimal to implement two projects, each staffed by a two-agent team. Moreover, by Proposition 1, the cost-minimizing team formation for two projects with two agents per project is anti-assortative. □