games

Article

# Optimal Incentives Schemes under *Homo Moralis* Preferences

**Roberto Sarkisian**

Department of Economics and Finance, Tor Vergata University of Rome, Via Columbia, 2 00133 Rome, Italy; roberto.sarkisian@uniroma2.eu

**Abstract:** This study focuses on the optimal incentive schemes in a multi-agent moral hazard model, where each agent has other-regarding preferences and an individual measure of output, with both being observable by the principal. In particular, the two agents display homo moralis preferences. I find that, contrary to the case with purely selfish preferences, tournaments can never be optimal when agents are risk averse, and as the degree of morality increases, positive payments are made in a larger number of output realizations. Furthermore, I extend the analysis to a dynamic setting, in which a contract is initially offered to the agents, who then repeatedly choose which level of effort to provide in each period. I show that the optimal incentive schemes in this case are similar to the ones obtained in the static setting, but for the role of intertemporal discounting.

**Keywords:** moral hazard in teams; optimal contracts; homo moralis preferences

**Citation:** Sarkisian, R. Optimal Incentives Schemes under *Homo Moralis* Preferences. *Games* **2021**, *12*, 28. https://doi.org/10.3390/g12010028

Academic Editors: Luís Santos-Pinto and Ulrich Berger

Received: 25 February 2021
Accepted: 15 March 2021
Published: 19 March 2021



## 1. Introduction

While most of the traditional economic literature on moral hazard has focused on agents' heterogeneous skills [1,2] and task allocation [3,4], it is crucial to also take into account social preferences in the context of incentive provision ([4] explore the notion of a mission-oriented production of collective goods, emphasizing the role of matching between the mission preferences of principals and agents, since the former economizes on the need for high-powered incentives). As pointed out in [5], a considerable fraction of the agents participating in their workplace experiment do not behave as selfishly as standard theory would predict. Fehr and et al. [6,7] show that fairness concerns may drastically impact contractual designs in principal agent environments. Dohmen et al. [8] survey experimental evidence of reciprocity, both in stylized labor markets as well as in other decision settings. The survey [9] finds evidence that explicit economic incentives can either reinforce or weaken prosocial behavior, and that the latter is more common, due to explicit incentives adversely affecting the individual's other-regarding preferences.

Here, I study the optimal incentives schemes a principal can offer to a team of two agents characterized by a novel class of other-regarding preferences, namely homo moralis preferences. The concept of Kantian ethical rules in economic interactions was first introduced by [10], while [11,12] build upon the ideas of assortativity and evolutionaty stability presented in [13] to derive a class of preferences that would be favored by evolution in settings with which individuals carrying rare mutant preferences get to interact (recent experimental evidence supporting homo moralis preferences can be found in [14,15], while [16] proposes a wider discussion on modeling prosocial preferences). Using a multi-agent moral hazard environment, as first proposed in [17,18], I show that the optimal contracts offered to the teams of agents have to balance three different aspects: the agents' prosocial behavior, here characterized by their degree of morality, risk aversion and incentive provision (Section 2 explores in more depth the concept and the utility function representing moral preferences). I also consider the possibility of repeated interactions between the agents, as in [19], and show that the optimal incentive scheme in the dynamic setting largely maintains the structure of its static counterpart but for the effects of discounting in the wages paid by the principal.

More closely related to this paper are the theoretical contributions identifying the effects of other-regarding preferences in contract design and incentives provision. Many of those study inequity aversion, following the seminal work of [20]. While [21] considers inequity-averse agents in tournaments, [22–24] look for the optimal incentive schemes under such preferences. While the first focus on binary effort choices by the agent (as in [20]), the latter two allow for continuous effort choice, and in [23], incomplete contracts are considered. In general, the results in this literature show that team incentives may outperform both individual and relative performance schemes when agents sufficiently dislike inequity ([24] shows that similar results hold for status-seeking agents as well).

In a similar vein to [20] as well, [25] derives optimal incentive schemes for reciprocal agents, a class of preferences first modeled in normal form games by [26] (a wider discussion on different classes of prosocial preferences can be found in [16]). As a result, [25] finds that the optimal incentive scheme depends on the interplay between risk aversion and the degree of reciprocity. More precisely, a relative performance scheme, which induces negative reciprocity, is optimal when agents are not very risk averse, while a joint performance scheme inducing positive reciprocity is better when agents become more risk averse. A different form of reciprocity between agents is altruism ([27] studies a model where agents have heterogeneous degrees of altruism (and greed). Their construction differs from [28] notion of altruism, because on the latter, it is the agents' concern about each other's wellbeing rather than their concern about own social reputation that induces prosocial behavior). Meanwhile, [29,30] study conditions under which explicit incentives can improve or damage altruism between co-workers (see [31,32] for more on altruism). In contrast to inequity aversion, and closer to the results in reciprocity, they find that both team performance and relative performance schemes can reinforce altruism in the workplace.

Differently than the literature above, I find that in most cases, relative performance is the optimal scheme for incentivising moral agents. In one particular case, team performance is also optimal, but it is so because all other schemes are not available, since limited liability constraints rule them out. Moreover, I also show that tournaments are never optimal, in stark contrast to the studies of optimal incentive schemes with purely selfish individuals.

The choice of homo moralis preferences comes from the realization that, in all the literature listed above, other-regarding preferences are assumed based only on psychological and experimental results. Although in most cases assuming a certain type of preferences have an intuitive appeal, as in the intra-household models based on forms of altruism, a theoretical foundation for the choice of one or other preference representation was lacking. The missing link, then, is a specification of preferences that is robust in a general setting, or one that evolves endogenously over time in a population. Alger and Weibull [11,12] provide such a link. They show that under incomplete information (agents' preferences are privately observed) and assortative matching, homo moralis preferences emerge as the evolutionarily stable ones, and that the degree of morality is given by the degree of assortativity of the matching process in which the individuals participate. Moreover, [11,12] argue that the utility function representing homo moralis preferences is the only one that proves to be robust against invasion in monomorphic populations in the class of continuous utility functions. As described in their paper, these preferences can be understood as a convex combination of the well-known selfish homo oeconomicus preferences and [33]'s concept of Kantian morality.

The paper continues in the following way. Section 2 introduces the model and the homo moralis utility function. Section 3 then analyses the problem faced by the principal in the static setting, while Section 4 extends the results to the dynamic environment. Section 5 concludes. For ease of exposition, all proofs are collected in the Appendix A.

## 2. The Model

Consider a firm composed of one manager (principal) and two employees (agents), denoted by $i \in \{A, B\}$. Each agent produces an observable output $x_i \in \{x^H, x^L\}$, with $x^H > x^L$, which is stochastically determined by the agent's choice of either exerting effort or shirking, i.e., $e_i \in \{0,1\}$. This production technology is characterized by the probability of achieving a high output conditional on the effort supplied:

$$Prob\Big(x_i = x^H | e_i = 1\Big) = p \in (0,1), \tag{1}$$

$$Prob\Big(x_i = x^H | e_i = 0\Big) = q \in (0,p). \tag{2}$$

This formulation assumes that the observable outputs $x_A$ and $x_B$ depend only on the corresponding agent's choice of effort and are independently drawn, and the production technology is symmetric. The cost of exerting effort is given by

$$C(e_i) = ce_i, \quad c > 0, i \in \{A,B\}.$$

The principal is assumed to be risk-neutral, and can use a remuneration scheme $\mathbf{w} = (\mathbf{w}_A, \mathbf{w}_B)$ to compensate her employees, which possibly depends on the output realizations $x_A$ and $x_B$. Thus, the principal's expected payoff can be written as

$$V(x_A, x_B, \mathbf{w}) = \sum_i \mathbb{E}[x_i - w_i].$$

Each agent's material payoff is assumed to be additively separable in wages and effort, i.e.,

$$\pi_i(w_i, e_i) = u_i(w_i) - C(e_i).$$

For ease of exposition, I assume that employees $A$ and $B$ value wages identically: $u_A(w) = u_B(w) = w^{1-\rho}$, for $\rho \in [0,1)$, thus allowing one to examine the behavior under risk neutrality ($\rho = 0$) as a limiting case of risk-averse agents ($\rho \in (0,1)$). Therefore, their material payoffs can be rewritten as

$$\pi(w_i, e_i) = w_i^{1-\rho} - ce_i. \tag{3}$$

For any pair of effort choices $(e_A, e_B)$, the space of possible output realizations is $\mathcal{S} = \{(x^H, x^H), (x^H, x^L), (x^L, x^H), (x^L, x^L)\}$, where each element $s \in \mathcal{S}$ is an ordered pair $s = (x_A, x_B)$. The principal can offer compensation schemes determining wages after each possible realization of output, namely

$$\mathbf{w_i} = (w_{iHH}, w_{iHL}, w_{iLH}, w_{iLL}),$$

where $w_{iHH}$ specifies, for instance, the wage received by agent $i$ when both output realizations are high and $w_{iHL}$ denotes the same agent's wage when his realized output is high while his partner's output realization is low. The agents' expected material payoff, conditional on efforts, is

$$\begin{aligned}\mathbb{E}\big[\pi(w_i, e_i)|e_i, e_j\big] = {} & P(e_i)P(e_j)w_{iHH}^{1-\rho} + P(e_i)\big[1 - P(e_j)\big]w_{iHL}^{1-\rho} \\ & + [1 - P(e_i)]P(e_j)w_{iLH}^{1-\rho} + [1 - P(e_i)]\big[1 - P(e_j)\big]w_{iLL}^{1-\rho} - ce_i,\end{aligned}$$

for $i, j \in \{A,B\}, j \neq i$.

Up to this moment, the preferences of the employees have not been fully described. In particular, I assume that the agents have homo moralis preferences (see [11,12]), represented by the (expected) utility function

$$U_i(\mathbf{w}_i, e_i, e_{-i}; \kappa_i) = (1 - \kappa_i)\mathbb{E}[\pi(w_i, e_i)|e_i, e_{-i}] + \kappa_i\mathbb{E}[\pi(w_i, e_i)|e_i, e_i], \tag{4}$$

where $\kappa_i \in [0,1]$ denotes agent $i$'s degree of morality. Inspection of the above expression shows that this specification is the convex combination between the usual representation of selfish preferences (the first term) and agent $i$'s material payoff if agent $j$ were to choose the same action (second term). Moreover, the limiting cases are interesting: while taking $\kappa_i = 0$

reduces the utility function to the standard selfish preferences, $\kappa_i = 1$ captures a situation where agent $i$ doesn't behave strategically: indeed, the problem, in that case, reduces to a single decision where $j \neq i$ choice of effort has not effect on agent $i$'s utility.

Throughout the exposition, I assume that the difference $x^H - x^L > 0$ is large enough for the principal to always prefer to induce both agents not to shirk. Furthermore, in order to focus on incentives provision, I assume that the workers are already employed by the firm, that contracts are bound by limited liability constraints and that preferences and costs are common information. Thus, the only private information is the agents' choices of effort. Timing is as follows: the principal sets her preferred incentive schemes (possibly contingent on both performance indicators $(x_A, x_B)$). The agents then simultaneously choose whether or not to exert effort. Finally, $(x_A, x_B)$ is realized and payments are made according to the incentives schemes proposed by the employer.

Some remarks must be made. First, given any incentive scheme, agents $A$ and $B$ play a static game with complete information. Not only do they know the proposed incentive scheme, they also know their partner's degree of morality, and thus his preferences. Moreover, since this is a one-shot game, it is irrelevant whether the agents can observe each other's choice of effort after the outputs are realized or not, and thus discussions about commitment are outside the scope of this model. Second, assuming that the agents are already employed by the firm somewhat relaxes the problem that will be solved by the principal, since participation constraints will not be considered. I will consider, however, limited liability on wages. Thus, if the outside option on the participation constraint would be set to zero, then limited liability would imply the former.

## 3. The Principal's Problem in the Static Framework

The principal's problem is

$$\begin{array}{lll} \max_{\mathbf{w}} & V(x_A, x_B, \mathbf{w}) & \\ s.t. & U_i(\mathbf{w}_i, 1, 1; \kappa_i) \geq U_i(\mathbf{w}_i, 0, 1; \kappa_i) & (IC_i) \\ & w_{iHH}, w_{iHL}, w_{iLH}, w_{iLL} \geq 0 & (LL_i) \end{array}$$

for $i \in \{A, B\}$. Given the risk neutrality and the linearity of the expectation operator, and assuming both agents will exert effort, the principal's expected profits can be rewritten as.

$$V(x_A, x_B, \mathbf{w}) = \mathbb{E}[x_A + x_B] - \left[ p^2 \sum_i w_{iHH} + p(1-p) \sum_i (w_{iHL} + w_{iLH}) + (1-p)^2 \sum_i w_{iLL} \right].$$

Since the principal maximizes over the incentives schemes, the problem above is equivalent to

$$\begin{array}{lll} \min_{\mathbf{w}} & p^2 \sum_i w_{iHH} + p(1-p) \sum_i (w_{iHL} + w_{iLH}) + (1-p)^2 \sum_i w_{iLL} & \\ s.t. & U_i(\mathbf{w}_i, 1, 1; \kappa_i) \geq U_i(\mathbf{w}_i, 0, 1; \kappa_i) & (IC_i) \\ & w_{iHH}, w_{iHL}, w_{iLH}, w_{iLL} \geq 0 & (LL_i) \end{array}$$

Let's focus now on the incentive compatibility constraint. On the left-hand side, both agents are exerting effort, so that $\mathbb{E}[\pi(\mathbf{w}_i, e_i^*) | e_i^*, e_j^*] = \mathbb{E}[\pi(\mathbf{w}_i, e_i^*) | e_i^*, e_i^*]$. Therefore, one obtains

$$\begin{aligned} U_i(\mathbf{w}_i, 1, 1; \kappa_i) &= \mathbb{E}[\pi(\mathbf{w}_i, 1) | e_i^* = 1, e_j^* = 1] \\ &= p^2 w_{iHH}^{1-\rho} + p(1-p) w_{iHL}^{1-\rho} + (1-p) p w_{iLH}^{1-\rho} + (1-p)^2 w_{iLL}^{1-\rho} - c, \end{aligned}$$

while the right-hand side writes

$$U_i(\mathbf{w}_i, 0, 1; \kappa_i) = (1-\kappa_i)\Big[qpw_{iHH}^{1-\rho} + q(1-p)w_{iHL}^{1-\rho} + (1-q)pw_{iLH}^{1-\rho} + (1-q)(1-p)w_{iLL}^{1-\rho}\Big] + \kappa_i\Big[q^2w_{iHH}^{1-\rho} + q(1-q)w_{iHL}^{1-\rho} + (1-q)qw_{iLH}^{1-\rho} + (1-q)^2w_{iLL}^{1-\rho}\Big].$$

Due the limited liability constraints and an implicit assumption of a normalized outside option to zero, if $q = 0$ the principal can set $w_{iLL} = 0$ and the incentive compatibility constraints for the moral agents become identical to the one for a purely selfish agent.

Plugging in the above equations into the incentive compatibility constraint and rearranging the terms around the wages yields

$$\begin{aligned} &w_{iHH}^{1-\rho}\Big[p^2 - (1-\kappa_i)qp - \kappa_i q^2\Big] \\ &\quad + w_{iHL}^{1-\rho}[p(1-p) - (1-\kappa_i)q(1-p) - \kappa_i q(1-q)] \\ &\quad + w_{iLH}^{1-\rho}[(1-p)p - (1-\kappa_i)(1-q)p - \kappa_i(1-q)q] \\ &\quad + w_{iLL}^{1-\rho}\Big[(1-p)^2 - (1-\kappa_i)(1-q)(1-p) - \kappa_i(1-q)^2\Big] \geq c. \end{aligned}$$

This form of writing the incentive compatibility constraint is very convenient to observe how the degree of morality affects the incentives of agent $i$ to exert effort. To start, take the term multiplying $w_{iHH}$, and suppose $\kappa_i = 0$. In this case, one obtains $p \cdot p - q \cdot p = (p-q) \cdot p$, which exactly describes the decrease in the probability of achieving the output realization $(x^H, x^H)$ that would be observed under selfish preferences: agent $i$ would take the action $e_{-i} = 1$ as a given, and would only consider the effects caused by his own shirking. On the other hand, for $\kappa_i = 1$, the term would become $p \cdot p - q \cdot q = (p-q) \cdot (p+q) > (p-q) \cdot p$: everything else fixed, the principal would need a smaller wage $w_{iHH}$ to incentivise agent $i$, since now agent $i$ would evaluate his payoff as if both him and his partner were shirking. Similar reasoning can be applied to the remaining terms.

One interesting remark is in order at this point. Under standard homo oeconomicus preferences, both agents are characterized by the same degree of morality $\kappa_i = 0$, and thus each multiplicative term is identical for employees $A$ and $B$. However, if $\kappa_A \neq \kappa_B$, these terms may not be the same any longer, and the workers would behave as if they possess heterogeneous beliefs (see [34] for moral hazard problems with heterogenous beliefs) about the realizations of output. This would, therefore, give a rationale for different wages being proposed (and accepted in the case where participation constraints are included in the model) by agents facing the same disutility of effort and attitude towards risk. Observe, however, the two approaches are radically different at heart: while [34] assumes agents have heterogeneous beliefs about the probability of success, thus implying that at least one of them have incorrect beliefs, in my model I assume both agents have correct beliefs about the probability of success, but differ only on their degree of morality.

For ease of exposition, the analysis will be divided into two parts: first, the risk-neutral case ($\rho = 0$) will be tackled. Then, I proceed to characterize the optimal incentive schemes when the agents are risk averse ($\rho \in (0,1)$).

### *3.1. Optimal Incentive Schemes for Risk-Neutral Agents ($\rho = 0$)*

For now, focus is channeled towards risk-neutral agents ($\rho = 0$). Under this additional assumption, the principal's problem is a linear programming problem with five inequality constraints: the incentive compatibility and the four limited liability constraints. The first result states that the principal's problem accepts three widely known solution candidates, namely an individual incentive scheme, where the principal remunerates each agent $i$ according to his observable measure of output $x_i$ alone; a team incentive scheme, in which the basis for remuneration is the sum of the individual observable measures; and a tournament scheme, such that agent $i$ receives a bonus if his output measurement has the highest value.

**Lemma 1.** *When agents are risk neutral with respect to wealth and have homo moralis preferences, the following two solution candidates implement $e_i = 1$, $\forall \kappa_i \in [0,1]$, $i \in \{A, B\}$:*

1. *an individual incentive scheme, with*

$$w_{iHH} = w_{iHL} = \frac{c}{p-q} > w_{iLH} = w_{iLL} = 0;$$

2. *a team incentive scheme, such that*

$$w_{iHH} = \frac{c}{(p-q)(p+\kappa_i q)} > w_{iHL} = w_{iLH} = w_{iLL} = 0.$$

*For $\kappa_i < \frac{1-p}{q}$, a tournament scheme also implements $e_i = 1$:*

$$w_{iHL} = \frac{c}{(p-q)(1-p-\kappa_i q)} > w_{iHH} = w_{iLH} = w_{iLL} = 0.$$

**Proof.** all proofs are in the Appendix A. □

Inspection of the remuneration structures reveals two interesting insights. First, under the individual incentive schemes, the wage paid following a high realization of the observable measure of output does not depend on the agents' degrees of morality, in contrast with the remaining schemes. Intuitively, this is a consequence of the independence assumptions on the production technology and its stochastic measurement: together with an incentive scheme that relies solely on individual performance; this environment reduces to zero the effect of Kantian morality in the incentives provision; it is as if the employees are purely selfish.

Second, the tournament is only feasible if agent $i$ does not exhibit a high degree of morality. The mechanism behind this is the asymmetric nature of this particular incentive scheme: an employee can only receive the bonus if he outperforms his colleague, thus conflicting the agent's urge to do the right thing. However, if $p + q \leq 1$, a tournament is feasible for all $\kappa_i \in [0,1]$. In this case, since the probability of realizing a high output measure is sufficiently small, the incentives provided by the asymmetric scheme may overpower the agents' morality in order to induce both to exert effort.

In order to determine which scheme among the ones mentioned above is the most profitable for the principal, one must simply compare the expected payments made under each alternative structure.

**Lemma 2.** *When agents are risk neutral with respect to wealth and have homo moralis preferences, the principal is indifferent among the alternative schemes if $\kappa_i = 0$. If $\kappa_i \in (0,1]$, the principal strictly prefers the team incentive scheme over the individual and tournament structures.*

The statement considers two distinct cases: one for $\kappa = 0$ and another for $\kappa > 0$. In the first case, the analysis boils down to standard homo oeconomicus preferences with risk-neutral agents. Thus, since the agents are identical and risk-sharing is not an issue, all three structures provide exactly the same expected payments to the employees and, therefore, have the same expected cost for the principal. One concludes that the principal is indifferent among the alternative compensation schemes.

The interesting case, however, lies in $\kappa > 0$. When the employees display a concern with doing the right thing, the principal is strictly better off implementing a team incentive scheme. Such a scheme implies that the desired outcome is a high output realization for agents 1 and 2, which transforms exerting a high effort into being the right thing. Since both agents now display a positive degree of morality, the total expected cost of explicitly incentivising the agents is reduced.

Although Lemma 2 rules out individual performance and tournaments as the optimal incentive schemes (for $\kappa_i > 0$), it does not fully characterize the solution to the principal's problem. This is done in Proposition 1 below.

**Proposition 1.** *When agents are risk neutral with respect to wealth and have homo moralis preferences, the optimal incentive scheme for the principal is team performance.*

Proposition 1 strengthens Lemma 2: team incentives are the best scheme a principal can use to incentivise a team of moral and risk-neutral agents, among all schemes that satisfy the incentive compatibility and limited liability constraints.

The proof of Proposition 1 is constructed in four steps. First, I show that any optimal incentive scheme always has $w_{iLL} = 0$ for $i \in \{A, B\}$. Then, it is easy to show that the incentive compatibility constraint must be satisfied with equality. The third step uses Lemma 2, thus eliminating any incentive scheme such that $w_{iHL} > 0$. Then, the fourth and last step must only consider schemes with $w_{iHH}, w_{iLH} \geq 0$; finally, I show that the principal's expected transfers to the agents are minimized with a team incentive scheme for any $\kappa_i \in [0, 1]$.

Closer inspection of the optimal incentive scheme shows that the principal is better off with teams of highly moral agents. The mechanism behind this is that a larger degree of morality slackens the incentive compatibility constraint, thus demanding a smaller transfer from the employer to the employees. This is stated formally below.

**Corollary 1.** *Under the optimal incentive scheme with risk-neutral agents (team performance), the principal's expected profit is strictly increasing in the agents' degrees of morality.*

### *3.2. Optimal Incentive Schemes for Risk-Averse Agents ($\rho \in (0, 1)$)*

Studying the risk-neutral case allows an understanding of the effects that homo moralis preferences have on designing the optimal incentive scheme, without having to take into consideration the trade-off between incentive provision and risk sharing. In particular, the agents' urge to do the right thing makes team performance scheme the most profitable for the principal in that case. In this section, the risk neutrality assumption is relaxed, and the optimal incentive scheme will have to balance morality, incentive provision and risk aversion.

The assumption on a functional form for the utility function over wealth, namely $u(w) = w^{1-\rho}$ for $\rho \in [0, 1)$, comes in handy in this section, since the results under risk neutrality can be treated as a particular case of this more general framework. Thus, at least for sufficiently high degrees of morality and low risk aversion, one expects team performance to be the optimal incentive scheme. The analysis below aims to specify the conditions for that claim to hold.

First, it is noteworthy that the usual incentive schemes (team, individual performance and tournaments) can be used by the principal to elicit effort. However, one other scheme must also be considered here: relative performance. In such a scheme, payments to agent $i$ are made whenever his output realization is high, but it differs from an individual incentive scheme in allowing different wages following good or bad realizations of output from agent $j$. Under risk neutrality, both schemes are identical because of the linearity of the utility function. However, under risk aversion, the concavity of $u$ allows the principal to induce high effort by offering such a compensation scheme, since now any scheme must balance the trade-off between incentive provision and risk sharing.

**Lemma 3.** *When agents are risk averse with respect to wealth and have homo moralis preferences, the following incentive schemes implement $e_i = 1$ for $i \in \{A, B\}$:*

1. *an individual incentive scheme, for any* $\kappa_i \in [0,1]$, *with*

$$w_{iHH} = w_{iHL} = \left(\frac{c}{p-q}\right)^{\frac{1}{1-\rho}} > w_{iLH} = w_{iLL} = 0;$$

2. *a team incentive scheme, for any* $\kappa_i \in [0,1]$, *such that*

$$w_{iHH} = \left(\frac{c}{(p-q)(p+\kappa_i q)}\right)^{\frac{1}{1-\rho}} > w_{iHL} = w_{iLH} = w_{iLL} = 0;$$

3. *a tournament scheme, for* $\kappa_i < \frac{1-p}{q}$, *in which*

$$w_{iHL} = \left(\frac{c}{(p-q)(1-p-\kappa_i q)}\right)^{\frac{1}{1-\rho}} > w_{iHH} = w_{iLH} = w_{iLL} = 0;$$

4. *a relative performance scheme, for* $\kappa_i < \frac{1-p}{q}$

$$w_{iHH} = \left(\frac{c}{(p-q)(p+\kappa_i q) + A(\kappa_i,\rho)^{1-\rho}(p-q)(1-p-\kappa_i q)}\right)^{\frac{1}{1-\rho}} \geq$$
$$w_{iHL} = w_{iHH} \cdot A(\kappa_i,\rho) > w_{iLH} = w_{iLL} = 0$$

*where* $A(\kappa_i,\rho) = \left(\frac{p(1-p-\kappa_i q)}{(1-p)(p+\kappa_i q)}\right)^{\frac{1}{\rho}} \in [0,1]$.

For the first three schemes, taking $\rho = 0$ yields exactly the same expressions shown in Lemma 1, which characterized such schemes for risk-neutral agents. Now, taking the limit as $\rho \to 0$ on the relative performance scheme yields the same expression as in the team performance: lacking the need for risk sharing, both schemes are identical. Once again, if $q = 0$, all the incentive schemes above become independent of the prosociality degree $\kappa$, since the principal sets $w_{iLL} = 0$ in equilibrium and therefore the incentive compatibility constraint for the homo moralis agent becomes identical to the constraint for a purely selfish one.

Before characterizing the optimal incentive scheme for the principal, the following intermediate results deserves a few remarks.

**Lemma 4.** *For any* $\rho \in (0,1)$ *and* $\kappa_i \in [0,1]$, $i = 1,2$ , *the principal prefers an individual incentive scheme over a tournament.*

The intuition for Lemma 4 is very simple: since a tournament imposes more risk on the agent than an individual incentive scheme, it must remunerate the agent for the increase in the riskiness of the contract. However, this compensation for risk is not profitable for the principal, for any degree of morality of the agent. Moreover, if the degree of morality is sufficiently high, such a scheme does not even satisfy the incentive compatibility constraint.

In contrast to the risk-neutral case, the optimality of a team performance scheme no longer holds for all values of $\kappa_i$, $p$ and $q$. In particular, when compared to the individual performance scheme, the principal will only prefer the former if the agents' degrees of morality are very high, or if their coefficient of risk aversion is sufficiently low.

**Lemma 5.** *The principal strictly prefers team performance over individual performance schemes if* $\kappa_i > \overline{\kappa}(\rho) = \frac{p(1-p^\rho)}{qp^\rho}$.

Again, observing this result extends the findings under risk neutrality: for $\rho = 0$, the right-hand side of the necessary and sufficient condition becomes 0, and thus any positive degree of morality will imply the optimality of team incentives over individual performance as was seen before. However, the right-hand side is strictly increasing in $\rho$, which implies that only a very high degree of morality can offset an increase in the degree of risk aversion in order for the principal to profit from the team incentive scheme. As $\rho \to 1$, the condition becomes $\kappa_i > \frac{1-p}{q}$, which can never be satisfied if $p + q \leq 1$. Counterintuitively, as the employee becomes more risk averse, the principal can benefit from a high degree of morality by offering the agent a contract associating positive payments with a larger number of possible output realizations. On the other hand, if the agent is not very risk averse but has a very high degree of morality, remunerating solely on the case where both agents are successful in obtaining the high output is optimal given the beliefs held by the moral agent.

Lemmas 4 and 5 rank the principal's preferences over team, individual and tournament schemes, but refrain from comparing them to relative performance schemes. Proposition 2 below strengthens the comparison, by determining the optimal incentive scheme for the principal depending on the probabilities of attaining the high output, the agent's risk aversion and degree of morality.

**Proposition 2.** *Suppose agents are risk averse with respect to wealth and have homo moralis preferences. Then, for any $\rho \in (0, 1)$:*

1. *If $p + q > 1$ and*
   - $\kappa \in \left[0, \frac{1-p}{q}\right]$*: a relative performance scheme, with $w_{iHH}, w_{iHL} \geq 0$ and $w_{iLH} = w_{iLL} = 0$, is optimal;*
   - $\kappa \in \left[\frac{1-p}{q}, 1\right]$*: a team performance scheme, with $w_{iHH} \geq 0$ and $w_{iHL} = w_{iLH} = w_{iLL} = 0$, is optimal.*
2. *If $p + q \leq 1$ and*
   - $\kappa \in \left[0, \frac{p}{1-q}\right]$*: a relative performance scheme, with $w_{iHH}, w_{iHL} \geq 0$ and $w_{iLH} = w_{iLL} = 0$, is optimal;*
   - $\kappa \in \left[\frac{p}{1-q}, 1\right]$*: a performance scheme with $w_{iHH}, w_{iHL}, w_{iLH} \geq 0$ and $w_{iLL} = 0$ is optimal.*

The interplay between risk aversion and morality leads to the optimality of relative performance schemes in most cases: it is profitable for the principal to offer compensation schemes that induce positive payments in as many output realizations as possible. One case, however, does the exact opposite by proposing an incentive scheme where the only positive payment comes only if both agents are successful in their tasks: if $p + q > 1$ and $\kappa_i \geq \frac{1-p}{q}$, the principal can profit by exploring the agent's high degree of morality and, thus, belief in the realization of high outcomes to concentrate transfer to that particular realization instead of promising positive transfers, even when outputs are low.

**Corollary 2.** *Under the optimal incentive scheme with risk-averse agents, the principal's expected profit is non-decreasing in the agents' degrees of morality.*

As was the case under risk neutrality, the principal benefits from hiring agents with large degrees of morality, since they will need less explicit incentives embedded in the optimal compensation scheme in order to exert effort. However, the interplay of employees' morality and risk sharing demands compensation schemes that spread out payments more evenly across the possible realizations of output, in particular when the probability of realizing a high output is not very large (i.e., when $p + q \leq 1$).

## 4. Repeated Interactions

In what follows, I consider a repeated setting where the agents are expected to either exert effort ($e = 1$) or shirk ($e = 0$) in each period. As in [19], this arrangement is openended and can be terminated at the end of each period $t = 0, 1, \ldots$ with probability $1 - \delta \in (0, 1)$, where $\delta$ can also be thought of as the common discount factor for all three parties. A history at time $t$ is a sequence of effort choices made by the employees until period $t - 1$, and thus a strategy profile is a sequence of functions mapping from any possible history at each period into actions (more precisely, into a probability distribution over effort choices).

In this section, I will show that the optimal incentive schemes derived from the static model and stated in Proposition 2 are also capable of providing the incentives for both agents to exert effort in the repeated setting. Firstly, note that a dynamic incentive compatibility constraint, for any incentive scheme $\mathbf{w}^*$, will be written as

$$U_i(\mathbf{w}^*, 1, 1; \kappa_i) \geq (1 - \delta) U_i(\mathbf{w}^*, 0, 1; \kappa_i) + \delta \min\{U_i(\mathbf{w}^*, 0, 1; \kappa_i), U_i(\mathbf{w}^*, 0, 0; \kappa_i)\}. \tag{5}$$

While the left-hand side of inequality 5 yields the average present-discounted expected utility if both agents exert effort, the right-hand side represents the average present-discounted expected utility if one agent shirks and is sub-sequentially punished with the worst equilibrium payoff, which can be either $U_i(\mathbf{w}^*, 0, 1; \kappa_i)$ or $U_i(\mathbf{w}^*, 0, 0; \kappa_i)$ depending on the incentive scheme implemented by the principal (this point is explored further in [19], by showing that relative performance schemes in which $\pi(w, 0, 1) < \pi(w, 0, 0)$ will not be implemented by a profit maximizing principal in a repeated setting with purely selfish agents. The incentive compatibility constraint in inequality 5 reflects the findings of [19] in the setting with homo moralis agents).

Since $1 > p > q > 0$ by assumption, and together with the limited liability constraints, it is the case that $U_i(\mathbf{w}^*, 0, 1; \kappa_i) \geq U_i(\mathbf{w}^*, 0, 0; \kappa_i)$ under the three optimal incentive schemes in Proposition 2, so the relevant incentive constraint is

$$U_i(\mathbf{w}^*, 1, 1; \kappa_i) \geq (1 - \delta) U_i(\mathbf{w}^*, 0, 1; \kappa_i) + \delta U_i(\mathbf{w}^*, 0, 0; \kappa_i), \tag{6}$$

which holds whenever the static incentive compatibility constraint is satisfied; indeed, for any of optimal static schemes and $\delta \in [0, 1]$,

$$\begin{aligned} U_i(\mathbf{w}^*, 1, 1; \kappa_i) &\geq U_i(\mathbf{w}^*, 0, 1; \kappa_i) \\ &= (1 - \delta) U_i(\mathbf{w}^*, 0, 1; \kappa_i) + \delta U_i(\mathbf{w}^*, 0, 1; \kappa_i) \\ &\geq (1 - \delta) U_i(\mathbf{w}^*, 0, 1; \kappa_i) + \delta U_i(\mathbf{w}^*, 0, 0; \kappa_i). \end{aligned}$$

Moreover, one can easily check that $U_i(\mathbf{w}, 1, 1; \kappa_i) \geq U_i(\mathbf{w}, 0, 0; \kappa_i)$, so collusion in shirking is deterred by use of any of the three optimal incentive schemes in Proposition 2. However, the argument built until now does not imply that $e = 0$ is a symmetric Nash equilibrium of the stage-game. If it is not, then the trigger-strategy here considered does not induce both agents to exert effort in the repeated game. Such an issue does not arise if $U_i(\mathbf{w}^*, 0, 0; \kappa_i) \geq U_i(\mathbf{w}^*, 1, 0; \kappa_i)$, which can be written as

$$\begin{aligned} & q^2 w_{iHH}^{1-\rho} + q(1-q) w_{iHL}^{1-\rho} + (1-q) q w_{iLH}^{1-\rho} \geq \\ & \quad (1 - \kappa_i) \left[ pq w_{iHH}^{1-\rho} + p(1-q) w_{iHL}^{1-\rho} + (1-p) q w_{iLH}^{1-\rho} \right] \\ & \quad + \kappa_i \left[ p^2 w_{iHH}^{1-\rho} + p(1-p) w_{iHL}^{1-\rho} + (1-p) p w_{iLH}^{1-\rho} \right] - c. \end{aligned}$$

Let $\bar{c}(\mathbf{w}^*, \kappa_i)$ denote the value of $c$ that satisfies the condition above with equality for some optimal scheme $\mathbf{w}^*$ and degree of morality $\kappa_i$. I can now state the following result.

**Proposition 3.** *Consider an incentive scheme $\mathbf{w}^*$ characterized in Proposition 2. If $c \geq \max\{\bar{c}(\mathbf{w}^*, \kappa_A), \bar{c}(\mathbf{w}^*, \kappa_B), 0\}$, then the static optimal incentive scheme $\mathbf{w}^*$ induces both agents to cooperate in the repeated setting.*

An important point of Proposition 3 is that it holds for any value of the discount factor $\delta$. That is, as long as the cost of exerting effort is sufficiently high to avoid $e = 1$ being a (weakly) dominant strategy for any of the employees, the optimal static incentive schemes of Proposition 2 also generate implicit incentives deterring shirking in the dynamic case irrespective of how patient the agents are. This is a consequence of the dynamic incentive compatibility constraint 6 being automatically satisfied by the schedules respecting its static version. Therefore, tournaments and individual performance schemes can also sustain effort in the dynamic game.

**Corollary 3.** *Tournaments ($\mathbf{w}^{Tourn}$) and individual performance scheme ($\mathbf{w}^{Ind}$) induce both agents to exert effort in the repeated setting if $c \geq \max\{\bar{c}(\mathbf{w}, \kappa_A), \bar{c}(\mathbf{w}, \kappa_B), 0\}$.*

Now, I want to focus on the more general principal's problem

$$\begin{array}{lll} \min_{\mathbf{w}} & p^2 w_{iHH} + p(1-p)(w_{iHL} + w_{iLH}) + (1-p)^2 w_{iLL} & \\ s.t. & U_i(\mathbf{w}^*, 1, 1; \kappa_i) \geq (1-\delta) U_i(\mathbf{w}^*, 0, 1; \kappa_i) & \\ & \quad +\delta \min\{U_i(\mathbf{w}^*, 0, 1; \kappa_i), U_i(\mathbf{w}^*, 0, 0; \kappa_i)\} & (DIC_i) \\ & w_{iHH}, w_{iHL}, w_{iLH} w_{iLL} \geq 0 & (LL_i) \end{array}$$

If $U_i(\mathbf{w}^*, 0, 1; \kappa_i) \leq U_i(\mathbf{w}^*, 0, 0; \kappa_i)$, the principal's problem is identical to the one in the static case, then the optimal incentive schemes described in Proposition 2 apply to the repeated setting. The more interesting case happens if $U_i(\mathbf{w}^*, 0, 1; \kappa_i) > U_i(\mathbf{w}^*, 0, 0; \kappa_i)$: for a large discount factor $\delta$, the unique optimal incentive scheme will be either a team incentive scheme or a complete incentive scheme if $p + q < 1$ or $p + q > 1$, respectively. If, however, $p + q = 1$, then a relative performance scheme is uniquely optimal. The formal statement is given below.

**Proposition 4.** *Let $\underline{\kappa}(\delta)$ and $\overline{\kappa}(\delta)$ be such that*

$$\overline{\kappa}(0) = \frac{1-p}{q}, \quad \underline{\kappa}(0) = \frac{p}{1-q},$$

*and*

$$\frac{\partial \overline{\kappa}(\delta)}{\partial \delta} \begin{cases} > 0 & \text{if } p + q < 1 \\ = 0 & \text{if } p + q = 1 \\ < 0 & \text{if } p + q > 1 \end{cases}, \frac{\partial \underline{\kappa}(\delta)}{\partial \delta} \begin{cases} < 0 & \text{if } p + q < 1 \\ = 0 & \text{if } p + q = 1 \\ > 0 & \text{if } p + q > 1 \end{cases}.$$

*Then, for any $\rho \in (0, 1)$ and $\delta \in (0, 1)$, the optimal incentive scheme for a risk-averse agent characterized by homo moralis preferences is:*

1. *if $p + q > 1$ and*
   - *$\kappa \in [0, \overline{\kappa}(\delta))$: a relative performance scheme, with $w_{iHH}, w_{iHL} > 0$ and $w_{iLH} = w_{iLL} = 0$;*
   - *$\kappa \in [\overline{\kappa}(\delta), 1]$: a team performance scheme, with $w_{iHH} > 0$ and $w_{iHL} = w_{iLH} = w_{iLL} = 0$.*
2. *if $p + q < 1$ and*
   - *$\kappa \in [0, \underline{\kappa}(\delta)]$: a relative performance scheme, with $w_{iHH}, w_{iHL} > 0$ and $w_{iLH} = w_{iLL} = 0$;*
   - *$\kappa \in (\underline{\kappa}(\delta), 1]$: a performance scheme with $w_{iHH}, w_{iHL}, w_{iLH} > 0$ and $w_{iLL} = 0$.*
3. *if $p + q = 1$, then $\underline{\kappa}(\delta) = \overline{\kappa}(\delta) = 1$ and a relative performance scheme, with $w_{iHH}, w_{iHL} > 0$ and $w_{iLH} = w_{iLL} = 0$, is optimal.*

An increase in the discount factor has two effects. The first one is the shifts in the thresholds $\underline{\kappa}(\delta)$ and $\overline{\kappa}(\delta)$. As $\delta$ approaches one, the values of the thresholds escape the interval $[0, 1]$ that characterizes the degree of morality of the agents, and thus only one incentive scheme is optimal for each case (in the proof of Proposition 4 in the Appendix A, I show that the limits go to plus and minus infinity depending on whether $p+q>1$ or $p+q<1$).

The second effect is that an increase in the discount factor decreases the wage that must be paid to the agents, in particular if both output measures are high. This is true because the incentive schemes satisfying the dynamic incentive compatibility constraint carry implicit incentives for both agents to exert effort, by the existing threat of everlasting punishment in case of an unilateral deviation. Therefore, the principal benefits rest on how moral and patient his employees are, as intuition suggests.

## 5. Concluding Remarks

Studying optimal incentive schemes with other-regarding preferences highlights the fact that the traditional trade-off between risk sharing and incentive provision is not the only one to influence the characterization of optimal contracts, and thus, may provide a better understanding of why the contracts observed in reality are not as high-powered as the ones predicted in the theory. In this line, [20,23,25], among others, explore the effects that altruism, inequity aversion and reciprocity have on compensation schemes.

Using the recent results on the evolution of preferences provided by [11,12], I study the problem of optimal incentive provision when agents display other-regarding preferences and different attitudes toward risk. In particular, I have shown that the optimal incentive scheme for moral agents may exhibit more risk than for selfish agents, in the sense that compensation is spread among more possible outcomes rather than aggregated around only one agent's output realization, as a consequence of the implicit incentives generated by morality. Furthermore, in contrast to [20,25], I show that tournaments are never optimal for positive degrees of morality.

Following [19], I extend the analysis to a dynamic environment and show that the optimal incentive schemes derived in the static case are also optimal when the agents are engaged in repeated interactions. The only difference between the former and the latter is intertemporal discounting, which affects the amount but not the underlying structure of the compensation schemes in the dynamic setting.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** I thank Ingela Alger, François Salanié, Leonard Wantchekon, Jörgen Weibull, and two anonymous referees and the editor for their very valuable feedback. I also thank seminar audiences at Toulouse School of Economics.

**Conflicts of Interest:** The author declares no conflict of interest.

## Appendix A. Proofs

### *Appendix A.1. Proof of Lemma 1*

As a starting point, I claim that any optimal contract must satisfy the incentive compatibility constraint with equality, and must be such that $w_{iLL}=0$. To see this, I rewrite the $(IC_i)$ constraint assuming $\rho=0$ as follows:

$$w_{iHH}[p+\kappa_i q]+w_{iHL}[(1-p)-\kappa_i q]-w_{iLH}[p-\kappa_i(1-q)]-w_{iLL}[(1-p)+\kappa_i(1-q)]\geq\frac{c}{p-q}.$$

Note that for any $0<q<p<1$ and $\kappa_i$, $w_{iLL}$ is multiplied by a strictly negative term, while $w_{iHH}$ is multiplied by a strictly positive term. For $w_{iHL}$, the term multiplying it is strictly positive for $\kappa_i<\frac{1-p}{q}$ (and negative otherwise), and $w_iLH$ is multiplied by a strictly positive term whenever $\kappa_i>\frac{p}{1-q}$ (and negative otherwise).

Therefore, suppose, by contradiction, that $\mathbf{w}_i=(w_{iHH},w_{iHL},w_{iLH},w_{iLL})$ is an optimal contract that satisfies the $(IC_i)$ with some slack and also satisfies the limited liability (non-negativity) constraints. If $\mathbf{w}_i$ is such that $w_{iLH} > 0$, the principal can offer a new contract $\mathbf{w}_i' = (w_{iHH}, w_{iHL}, w_{iLH}, 0)$ which doesn't violate any of the constraints and make him better off. Indeed, note that

$$\begin{aligned} & w_{iHH}[p+\kappa_i q] + w_{iHL}[(1-p)-\kappa_i q] - w_{iLH}[p-\kappa_i(1-q)] > \\ & \quad w_{iHH}[p+\kappa_i q] + w_{iHL}[(1-p)-\kappa_i q] - w_{iLH}[p-\kappa_i(1-q)] - w_{iLL}[(1-p)+\kappa_i(1-q)], \end{aligned}$$

and

$$p^2 w_{iHH} + p(1-p)(w_{iHL}+w_{iLH}) < p^2 w_{iHH} + p(1-p)(w_{iHL}+w_{iLH}) + (1-p)^2 w_{iLL}.$$

Thus, any optimal contract must have $w_{iLL} = 0$. Moreover, a similar argument shows that $w_{iHL} = 0$ and $w_{iLH} = 0$ whenever their multiplying terms are strictly negative, i.e., whenever $\kappa_i \geq \frac{1-p}{q}$ and $\kappa_i \leq \frac{p}{1-q}$, respectively.

Now, see that the principal can reduce the expected transfers to the agents by offering a contract $\mathbf{w}_i'' = (w_{iHH}-\varepsilon_1, w_{iHL}-\varepsilon_2, w_{iLH}-\varepsilon_3, 0)$ where $\varepsilon_1 > 0$ and

$$\varepsilon_2 \begin{cases} > 0, & \text{if } \kappa_i < \frac{1-p}{q} \\ = 0, & \text{otherwise.} \end{cases}$$

and

$$\varepsilon_3 \begin{cases} > 0, & \text{if } \kappa_i > \frac{p}{1-q} \\ = 0, & \text{otherwise.} \end{cases}$$

For $\varepsilon_1, \varepsilon_2, \varepsilon_3 \approx 0$, the incentive compatibility constraint is still satisfied, while

$$p^2(w_{iHH}-\varepsilon_1) + p(1-p)[(w_{iHL}-\varepsilon_2)+(w_{iLH}-\varepsilon_3)] < p^2 w_{iHH} + p(1-p)(w_{iHL}+w_{iLH}).$$

Thus, the principal reduces $w_{iHH}$, $w_{iHL}$ and $w_{iLH}$, when the latter are not already zero, until the incentive compatibility constraint is satisfied with equality.

Given the argument above, attention can be restricted to incentives schemes such that $w_{iHH}, w_{iHL}, w_{iLH} \geq 0$ and $w_{iLL} = 0$. Thus, it is easy to see the three common incentive schemes, namely individual incentive scheme, team incentive scheme and tournament scheme, satisfy the conditions above. I analyze each in turn.

First, consider the individual incentive scheme, such that $w_{iHH} = w_{iHL} = w_{iH}, w_{iLH} = 0$. Substituting the first equality in $(IC_i)$ yields

$$\begin{aligned} & w_{iH}\Big[p^2 - (1-\kappa_i)qp - \kappa_i q^2 + p(1-p) - (1-\kappa_i)q(1-p) - \kappa_i q(1-q)\Big] - c = 0 \\ & w_{iH} = \frac{c}{p-q} > 0, \end{aligned}$$

since $p > q$ by assumption.

A team incentive scheme would have $w_{iHH} \geq 0, w_{iHL} = w_{iLH} = 0$. By force of $(IC_i)$, one obtains

$$w_{iHH} = \frac{c}{p^2 - (1-\kappa_i)qp - \kappa_i q^2} = \frac{c}{(p-q)(p+\kappa_i q)} > 0.$$

A tournament scheme consists of $w_{iHL} \geq 0, w_{iHH} = w_{iLH} = 0$. Again, using the incentive compatibility constraint yields

$$w_{iHL} = \frac{c}{p(1-p) - (1-\kappa_i)q(1-p) - \kappa_i q(1-q)} = \frac{c}{(p-q)(1-p-\kappa_i q)}.$$

Observe that $w_{iHL} > 0$ here if, and only if, $1-p-\kappa_i q > 0$, i.e., $\kappa_i < \frac{1-p}{q}$. Therefore, a tournament is a candidate solution if, and only if, agent $i$'s degree of morality is not very high.

*Appendix A.2. Proof of Lemma 2*

The proof follows directly from the comparison of expected payments. For ease of exposition, they are written:

1. Individual incentive scheme: $\sum_i \left[p^2 + p(1-p)\right] w_{iH} = 2p\frac{c}{p-q}$;
2. Team incentive scheme: $p^2 \sum_i w_{iHH} = p^2 \sum_i \frac{c}{(p-q)(p+\kappa_i q)}$;
3. Tournament: $p(1-p)\sum_i w_{iHL} = p(1-p)\sum_i \frac{c}{(p-q)(1-p-\kappa_i q)}$, for $\kappa_i < \frac{1-p}{q}$.

First, compare the individual incentive scheme against the team incentive scheme. For $\kappa_i = 0$, both generate the same expected payment for the principal. However, for $\kappa_i > 0$, one has

$$\frac{pc}{(p-q)} \cdot \frac{p}{p+\kappa_i q} < \frac{pc}{(p-q)},$$

since $\frac{p}{p+\kappa_i q} < 1$. Therefore, for all $\kappa_i \in [0,1]$ and $i \in \{A, B\}$, the principal is weakly better off implementing a team incentives scheme.

Now, it is only left to compare a team incentives scheme with a tournament. To do so, suppose $\kappa_i < \frac{1-p}{q}$; otherwise, the latter scheme does not satisfy the non-negativity constraints. Then, one can see that

$$\begin{aligned}
\frac{p^2 c}{(p_q)(p+\kappa_i q)} &\leq \frac{p(1-p)c}{(p-q)(1-p-\kappa_i q)} && \Leftrightarrow \\
\frac{p}{p+\kappa_i q} &\leq \frac{1-p}{1-p-\kappa_i q} && \Leftrightarrow \\
p - p^2 - \kappa_i pq &\leq p + \kappa_i q - p^2 - \kappa_i pq && \Leftrightarrow \\
\kappa_i q &\geq 0,
\end{aligned}$$

which is always satisfied, since $\kappa_i \in [0,1]$ and $q \in (0,1)$ by assumption. Therefore, a team incentives scheme is also weakly preferred by a principal over a tournament scheme.

*Appendix A.3. Proof of Proposition 1*

Building on the proof of Lemma 1, I restrict attention to schemes in which $w_{iLL} = 0$. As a first step, I show that it is never optimal for the principal to offer a contract with $w_{iHL} > 0$ (given that $\kappa_i < \frac{1-p}{q}$). Indeed, suppose $\mathbf{w}_i = (w_{iHH}, w_{iHL}, w_{iLH}, 0)$ satisfy the $(IC_i)$ with equality and the limited liability constraints; now, consider the alternative scheme $\mathbf{w}'_i = (w'_{iHH}, 0, w'_{iLH}, 0)$ such that

$$\begin{aligned}
w'_{iLH} &= w_{iLH} \\
w'_{iHH} &= w_{iHH} + \frac{1-p-\kappa_i q}{p+\kappa_i q} w_{iHL}.
\end{aligned}$$

Note that this scheme also satisfies the incentive compatibility constraint with equality. Indeed,

$$\begin{aligned}
&[p+\kappa_i q]w'_{iHH} - [p-\kappa_i(1-q)]w'_{iLH} \\
&= [p+\kappa_i q]\left(w_{iHH} + \frac{1-p-\kappa_i q}{p+\kappa_i q} w_{iHL}\right) - [p-\kappa_i(1-q)]w_{iLH} \\
&= [p+\kappa_i q]w_{iHH} + [1-p-\kappa_i q]w_{iHL} - [p-\kappa_i(1-q)]w_{iLH} \\
&= \frac{c}{p-q}.
\end{aligned}$$

Now, observe the principal's expected transfers under $\mathbf{w}'_i$ are less or equal than under $\mathbf{w}_i$ if, and only if

$$\begin{aligned} p^2 w'_{iHH} + p(1-p) w'_{iLH} &\leq p^2 w_{iHH} + p(1-p)(w_{iHL} + w_{iLH}) &\Leftrightarrow \\ p^2 w_{iHH} + p^2 \tfrac{1-p-\kappa_i q}{p+\kappa_i q} w_{iHL} + p(1-p) w_{iLH} &\leq p^2 w_{iHH} + p(1-p)(w_{iHL} + w_{iLH}) &\Leftrightarrow \\ p^2 w_{iHH} + p(1-p)(w_{iHL} + w_{iLH}) &\leq p(1-p) &\Leftrightarrow \\ (p+\kappa_i q)(1-p) &\geq p(1-p-\kappa_i q) &\Leftrightarrow \\ \kappa_i q &\geq 0, \end{aligned}$$

which is always satisfied, since $\kappa_i \in [0,1]$ and $q \in (0,1)$ by assumption (equality will only hold for $\kappa_i = 0$).

Therefore, any optimal contract must have $w_{iHL} = w_{iLL} = 0$, which rules out individual performance and tournament schemes. Note, however, a team incentive scheme may still be optimal. Therefore, the optimal incentive schemes must be such that

$$\begin{aligned} w_{iHH} &= \max\left\{0, \frac{1}{p+\kappa_i q}\left(\frac{c}{p-q} + [p - \kappa_i(1-q)] w_{iLH}\right)\right\} \\ w_{iLH} &\in \left[0, \frac{c}{(p-q)(\kappa_i(1-q)-p)}\right], \end{aligned}$$

for $\kappa_i > \frac{p}{1-q}$.

Given the contract described above, the principal's problem can be equivalently written as

$$\begin{aligned} \min_{w_{iLH}} p^2 \frac{1}{p+\kappa_i q}\left(\frac{c}{p-q} + [p-\kappa_i(1-q)] w_{iLH}\right) + p(1-p) w_{iLH} \\ = \frac{p^2 c}{(p-q)(p+\kappa_i q)} + \left[p^2 \frac{p-\kappa_i(1-q)}{p+\kappa_i q} + p(1-p)\right] w_{iLH} \\ = \frac{p^2 c}{(p-q)(p+\kappa_i q)} + \left[p - \frac{p^2 \kappa_i}{p+\kappa_i q}\right] w_{iLH}, \end{aligned}$$

where I assume $w_{iLH} \geq 0$. Observe that it is optimal for the principal to choose $w_{iLH} > 0$ if

$$\begin{aligned} p - \tfrac{p^2 \kappa_i}{p+\kappa_i q} &< 0 &\Leftrightarrow \\ p + \kappa_i q &< p\kappa_i &\Leftrightarrow \\ \kappa_i &> \tfrac{p}{p-q}. \end{aligned}$$

However, since $0 < q < p < 1$, the last inequality demands $\kappa_i > 1$, violating the assumption about the agents' degrees of morality. Thus, the principal optimally chooses the team incentives scheme, given by $\mathbf{w}_i^{opt} = (w_{iHH}^{opt}, 0, 0, 0)$, with

$$w_{iHH}^{opt} = \frac{c}{(p-q)(p+\kappa_i q)} > 0,$$

for all $\kappa_i \in [0,1]$, $i \in \{A, B\}$ and $0 < q < p < 1$.

*Appendix A.4. Proof of Corollary 1*

Simply take the derivative of $w_{iHH}$ under a team incentive scheme with respect to $\kappa_i$, and note its sign is strictly negative.

*Appendix A.5. Proof of Lemma 3*

Consider the principal's problem described in the main text. The KKT conditions are necessary and sufficient to characterize the candidate solutions, and are given by

$$p^2 - \mu_i(1-\rho) w_{iHH}^{-\rho}[(p-q)(p+\kappa_i q)] - \lambda_{iHH} = 0 \tag{A1}$$

$$p(1-p) - \mu_i(1-\rho) w_{iHL}^{-\rho}[(p-q)(1-p-\kappa_i q)] - \lambda_{iHL} = 0 \tag{A2}$$

$$(1-p)p - \mu_i(1-\rho)w_{iLH}^{-\rho}[-(p-q)(p-\kappa_i(1-q))] - \lambda_{iLH} = 0 \tag{A3}$$

$$(1-p)^2 - \mu_i(1-\rho)w_{iLL}^{-\rho}[-(p-q)(1-p+\kappa_i(1-q))] - \lambda_{iLL} = 0 \tag{A4}$$

$$\begin{aligned} &w_{iHH}(p+\kappa_i q) + w_{iHL}(1-p-\kappa_i q) \\ &\quad -w_{iLH}(p-\kappa_i(1-q)) - w_{iLL}((1-p)+\kappa_i(1-q)) \geq \tfrac{c}{p-q} \end{aligned} \tag{A5}$$

$$\begin{aligned} &\mu_i\{w_{iHH}(p+\kappa_i q) + w_{iHL}(1-p-\kappa_i q) \\ &\quad -w_{iLH}(p-\kappa_i(1-q)) - w_{iLL}((1-p)+\kappa_i(1-q)) - \tfrac{c}{p-q}\Big\} = 0 \end{aligned} \tag{A6}$$

$$w_{iHH} \geq 0 \tag{A7}$$

$$w_{iHL} \geq 0 \tag{A8}$$

$$w_{iLH} \geq 0 \tag{A9}$$

$$w_{iLL} \geq 0 \tag{A10}$$

$$\lambda_{iHH} w_{iHH} = 0 \tag{A11}$$

$$\lambda_{iHL} w_{iHL} = 0 \tag{A12}$$

$$\lambda_{iLH} w_{iLH} = 0 \tag{A13}$$

$$\lambda_{iLL} w_{iLL} = 0 \tag{A14}$$

$$\lambda_{iHH} \geq 0 \tag{A15}$$

$$\lambda_{iHL} \geq 0 \tag{A16}$$

$$\lambda_{iLH} \geq 0 \tag{A17}$$

$$\lambda_{iLL} \geq 0 \tag{A18}$$

$$\mu_i \geq 0 \tag{A19}$$

for all $i \in \{1,2\}$, where $\mu_i$ is the Lagrange multiplier associated with the incentive compatibility constraint, while $\lambda_i$s are the ones associated with the non-negativity constraint.

As was the case under risk aversion, $(IC_i)$ must bind. If that was not the case, $\mu_i = 0$ would imply through Equations (A1)–(A4) that $\lambda_{iHH}, \lambda_{iHL}, \lambda_{iLH}, \lambda_{iLL} > 0$, and thus, by force of the complementary slackness conditions (A11)–(A14), that $w_{iHH} = w_{iHL} = w_{iLH} = w_{iLL} = 0$. However, substituting into (A5), one obtains $0 \geq \frac{c}{p-q} > 0$, a contradiction.

An argument similar to the one used in the risk-neutral case could be employed here as well, and would fit the more general case of a utility function of wealth satisfying $u' > 0$, $u'' \leq 0$, $u(0) = 0$.

The first three incentive schemes described in the text are obtained by using Equation (A5), the incentive compatibility constraint, with equality and considering each case in turn:

1. Individual incentive scheme: $w_{iHH} = w_{iHL} > 0 = w_{iLH} = w_{iLL}$;
2. Team incentive scheme: $w_{iHH} > 0 = w_{iHL} = w_{iLH} = w_{iLL}$
3. Tournament scheme: $w_{iHL} > 0 = w_{iHH} = w_{iLH} = w_{iLL}$

For the relative performance scheme, assume $1 - p - \kappa_i q > 0$ and compute the ratio of Equations (A1) and (A2),

$$\begin{aligned} \frac{p^2}{p(1-p)} &= \frac{\mu_i(1-\rho)w_{iHH}^{-\rho}(p-q)(p+\kappa_i q)}{\mu_i(1-\rho)w_{iHL}^{-\rho}(p-q)(1-p-\kappa_i q)} && \Leftrightarrow \\ \left(\frac{w_{iHL}}{w_{iHH}}\right)^{\rho} &= \frac{p(1-p-\kappa_i q)}{(1-p)(p+\kappa_i q)} && \Leftrightarrow \\ w_{iHL} &= w_{iHH} \underbrace{\left(\frac{p(1-p-\kappa_i q)}{(1-p)(p+\kappa_i q)}\right)^{\frac{1}{\rho}}}_{=A(\kappa_i,\rho)} \end{aligned}$$

Since I assume $1 - p - \kappa_i q > 0$, $\kappa_i \in [0,1]$ and $0 < q < p < 1$, note that $A(\kappa_i, \rho) > 0$. Moreover, $A(0,\rho) = 1$ and

$$\frac{\partial A(\kappa_i,\rho)}{\partial \kappa_i} \propto -pq(1-p)(p+\kappa_i q) - q(1-p)p(1-p-\kappa_i q) < 0,$$

so that $A(\kappa_i,\rho) \in (0,1]$ for all $\kappa_i \in [0,1]$ and $\rho \in (0,1)$. Plugging $w_{iHH}$, $w_{iHL} = w_{iHH}A(\kappa_i,\rho)$ and $w_{iLH} = w_{iLL} = 0$ in (A5) yields the result, taking into consideration the non-negativity constraint as well.

*Appendix A.6. Proof of Lemma 4*

Suppose $1 - p - \kappa_i q > 0$, so that a tournament is a candidate solution to the principal's problem. For $\rho \in (0,1)$, the principal prefers a tournament over an individual performance scheme if, and only if, the expected transfers under the former are smaller than under the latter, that is, if

$$\begin{array}{ll} p(1-p)\left(\frac{c}{(p-q)(1-p-\kappa_i q)}\right)^{\frac{1}{1-\rho}} < [p^2 + p(1-p)]\left(\frac{c}{p-q}\right)^{\frac{1}{1-\rho}} & \Leftrightarrow \\ (1-p)^{1-\rho}\frac{c}{(p-q)(1-p-\kappa_i q)} < \frac{c}{p-q} & \Leftrightarrow \\ \kappa_i < \frac{1-p}{q}[1-(1-p)^{-\rho}] & \end{array}$$

Since $\kappa_i \in [0,1]$ by assumption, the inequality above holds only if $1 - (1-p)^{-\rho} \geq 0$, which is equivalent to

$$1 \geq \frac{1}{(1-p)^{\rho}} > 1,$$

a contradiction.

*Appendix A.7. Proof of Lemma 5*

The principal's expected payments under team incentives are smaller than under individual performance if

$$\begin{array}{ll} p^2\left(\frac{c}{(p-q)(p+\kappa_i q)}\right)^{\frac{1}{1-\rho}} < [p^2 + p(1-p)]\left(\frac{c}{p-q}\right)^{\frac{1}{1-\rho}} & \Leftrightarrow \\ p^{1-\rho}\frac{c}{(p-q)(p+\kappa_i q)} < \frac{c}{p-q} & \Leftrightarrow \\ \kappa_i q > p^{1-\rho} - 1 & \Leftrightarrow \\ \kappa_i > \underbrace{\frac{p}{q} \cdot \frac{1-p^{\rho}}{p^{\rho}}}_{=\overline{\kappa}(\rho)}. & \end{array}$$

*Appendix A.8. Proof of Proposition 2*

Using the KKT conditions obtained in the proof of Lemma 3, I will look for the optimal incentive scheme. As argued before, such a scheme must satisfy the incentive compatibility constraint with equality (i.e. $\mu_i > 0$ for all $i \in \{1,2\}$). Moreover, it must be such that $w_{iLL} = 0$. Indeed, on equation (A4), note that $-(p-q)[(1-p) + \kappa_i(1-q)] < 0$ for all $0 < q < p < 1$ and $\kappa_i \in [0,1]$; therefore, if $w_{iLL} > 0$, the complementary slackness condition implies that $\lambda_{iLL} = 0$, and thus the left-hand side of Equation (4) is strictly positive, contradicting the first-order condition.

A similar argument can be used on Equations (A2) and (A3): whenever the term multiplying the wage is negative, a solution must have the nonnegativity constraint binding. Therefore,

$$\kappa_i \geq \frac{1-p}{q} \Rightarrow w_{iHL} = 0, \tag{A20}$$

and

$$\kappa_i \leq \frac{p}{1-q} \Rightarrow w_{iLH} = 0. \tag{A21}$$

One can easily check that

$$\frac{1-p}{q} < 1 < \frac{p}{1-q} \Leftrightarrow p+q > 1, \quad \frac{1-p}{q} \geq 1 \geq \frac{p}{1-q} \Leftrightarrow p+q \leq 1,$$

so the analysis can be conveniently divided in two cases, namely $p+q > 1$ and $p+q \leq 1$.

Suppose first that $p+q > 1$. If $\kappa_i \in \left[\frac{1-p}{q}, 1\right]$, conditions (A20) and (A21) imply that $w_{iHL} = w_{iLH} = 0$, and the only solution candidate is the team incentive scheme described in Lemma 3. On the other hand, for $\kappa_i \in \left[0, \frac{1-p}{q}\right)$, the two conditions above imply that $w_{iHH}, w_{iHL} \geq 0$ and $w_{iLH} = w_{iLL} = 0$, so the four incentive schemes in Lemma 3 are candidate solutions.

It is easy to see that the relative performance scheme performs at least as good as any of the other three schemes in this case. Indeed, let $\mathcal{C} = \{\mathbf{w} \in \mathbb{R}^4_+ : w_{iHH}, w_{iHL} \geq 0, w_{iLH} = w_{iLL} = 0\}$ denote the set of contracts than can be offered if $p+q > 1$ and $\kappa_i \in \left[0, \frac{1-p}{q}\right)$. In a similar fashion, let

$$\begin{aligned}
\mathcal{C}^{Team} &= \left\{\mathbf{w} \in \mathbb{R}^4_+ : w_{iHH} \geq 0, w_{iHL} = w_{iLH} = w_{iLL} = 0\right\} \\
\mathcal{C}^{Ind} &= \left\{\mathbf{w} \in \mathbb{R}^4_+ : w_{iHH} = w_{iHL} \geq 0, w_{iLH} = w_{iLL} = 0\right\} \\
\mathcal{C}^{Tour} &= \left\{\mathbf{w} \in \mathbb{R}^4_+ : w_{iHL} \geq 0, w_{iHH} = w_{iLH} = w_{iLL} = 0\right\} \\
\mathcal{C}^{Rel} &= \left\{\mathbf{w} \in \mathbb{R}^4_+ : w_{iHH}, w_{iHL} \geq 0, w_{iLH} = w_{iLL} = 0\right\},
\end{aligned}$$

denote the set of contracts satisfying the conditions for the performance schemes described in Lemma 3. One can readily note that $\mathcal{C}^{Team}, \mathcal{C}^{Ind}, \mathcal{C}^{Tour} \subset \mathcal{C}$ and $\mathcal{C}^{Rel} = \mathcal{C}$. Therefore, team, individual or tournament schemes add more constraints to the set of contracts under which the principal can maximize his profits, and must not yield a strictly higher profit than the one obtained under the more relaxed constraint set $\mathcal{C}$.

If $p+q \leq 1$ and $\kappa < \frac{p}{1-q}$, the optimal scheme is the same as in the previous paragraph, i.e., the relative performance scheme with $w_{iHH}, w_{iHL} \geq 0$ and $w_{iLH} = w_{iLL} = 0$. However, if $p+q \leq 1$ and $\kappa \in \left[\frac{p}{1-q}, 1\right]$, the principal can maximize over the set $\widetilde{\mathcal{C}} = \{\mathbf{w} \in \mathbb{R}^4_+ : w_{iHH}, w_{iHL}, w_{iLH} \geq 0, w_{iLL} = 0\}$. Now, the contract sets defined by the four schemes presented above are strict subsets of $\widetilde{\mathcal{C}}$ and cannot, thus, yield a strictly higher payoff to the principal.

*Appendix A.9. Proof of Corollary 3*

Follows from the observation that the proposed incentive schemes satisfy the static incentive compatibility constraint and, thus, the dynamic version considered in Proposition 3.

*Appendix A.10. Proof of Proposition 4*

The proof follows closely the argument developed in Lemma 3 and Proposition 2. Suppose that $U_i(\mathbf{w}^*, 0, 1; \kappa_i) > U_i(\mathbf{w}^*, 0, 0; \kappa_i)$. The principal's problem becomes

$$\begin{aligned}
\min_{\mathbf{w}} \quad & p^2 w_{iHH} + p(1-p)(w_{iHL} + w_{iLH}) + (1-p)^2 w_{iLL} \\
s.t. \quad & p^2 w_{iHH}^{1-\rho} + p(1-p)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-p)^2 w_{iLL}^{1-\rho} \geq \\
& \quad (1-\delta)\Big[(1-\kappa_i)(qpw_{iHH}^{1-\rho} + q(1-p)w_{iHL}^{1-\rho} + (1-q)pw_{iLH}^{1-\rho} + (1-q)(1-p)w_{iLL}^{1-\rho}) \\
& \quad +\kappa_i(q^2 w_{iHH}^{1-\rho} + q(1-q)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-q)^2 w_{iLL}^{1-\rho})\Big] \\
& \quad +\delta(q^2 w_{iHH}^{1-\rho} + q(1-q)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-q)^2 w_{iLL}^{1-\rho}) && (DIC_i) \\
& w_{iHH}, w_{iHL}, w_{iLH} w_{iLL} \geq 0 && (LL_i)
\end{aligned}$$

whose KKT conditions are given by

$$\begin{aligned}&p^2 - \lambda_{iHH}\\ &\quad -\mu_i(1-\rho)w_{iHH}^{-\rho}[p^2 - (1-\delta)((1-\kappa_i)pq + \kappa_i q^2) - \delta q^2] = 0\end{aligned} \tag{A22}$$

$$\begin{aligned}&p(1-p) - \lambda_{iHL}\\ &\quad -\mu_i(1-\rho)w_{iHL}^{-\rho}[p(1-p) - (1-\delta)((1-\kappa_i)q(1-p) + \kappa_i q(1-q)) - \delta q(1-q)] = 0\end{aligned} \tag{A23}$$

$$\begin{aligned}&(1-p)p - \lambda_{iLH}\\ &\quad -\mu_i(1-\rho)w_{iLH}^{-\rho}[p(1-p) - (1-\delta)((1-\kappa_i)(1-q)p + \kappa_i q(1-q)) - \delta q(1-q)] = 0\end{aligned} \tag{A24}$$

$$\begin{aligned}&(1-p)^2 - \lambda_{iLL}\\ &-\mu_i(1-\rho)w_{iLL}^{-\rho}[(1-p)^2 - (1-\delta)((1-\kappa_i)(1-q)(1-p) + \kappa_i(1-q)^2) - \delta(1-q)^2] = 0\end{aligned} \tag{A25}$$

$$\begin{aligned}&p^2 w_{iHH}^{1-\rho} + p(1-p)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-p)^2 w_{iLL}^{1-\rho} \geq\\ &\quad (1-\delta)\Big[(1-\kappa_i)(qpw_{iHH}^{1-\rho} + q(1-p)w_{iHL}^{1-\rho} + (1-q)pw_{iLH}^{1-\rho} + (1-q)(1-p)w_{iLL}^{1-\rho})\\ &\quad +\kappa_i(q^2 w_{iHH}^{1-\rho} + q(1-q)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-q)^2 w_{iLL}^{1-\rho})\Big]\\ &\quad +\delta(q^2 w_{iHH}^{1-\rho} + q(1-q)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-q)^2 w_{iLL}^{1-\rho})\end{aligned} \tag{A26}$$

$$\begin{aligned}&\mu_i\Big\{p^2 w_{iHH}^{1-\rho} + p(1-p)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-p)^2 w_{iLL}^{1-\rho}\\ &\quad (1-\delta)\Big[(1-\kappa_i)(qpw_{iHH}^{1-\rho} + q(1-p)w_{iHL}^{1-\rho} + (1-q)pw_{iLH}^{1-\rho} + (1-q)(1-p)w_{iLL}^{1-\rho})\\ &\quad +\kappa_i(q^2 w_{iHH}^{1-\rho} + q(1-q)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-q)^2 w_{iLL}^{1-\rho})\Big]\\ &\quad +\delta(q^2 w_{iHH}^{1-\rho} + q(1-q)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-q)^2 w_{iLL}^{1-\rho})\Big\} = 0\end{aligned} \tag{A27}$$

$$w_{iHH} \geq 0 \tag{A28}$$

$$w_{iHL} \geq 0 \tag{A29}$$

$$w_{iLH} \geq 0 \tag{A30}$$

$$w_{iLL} \geq 0 \tag{A31}$$

$$\lambda_{iHH} w_{iHH} = 0 \tag{A32}$$

$$\lambda_{iHL} w_{iHL} = 0 \tag{A33}$$

$$\lambda_{iLH} w_{iLH} = 0 \tag{A34}$$

$$\lambda_{iLL} w_{iLL} = 0 \tag{A35}$$

$$\lambda_{iHH} \geq 0 \tag{A36}$$

$$\lambda_{iHL} \geq 0 \tag{A37}$$

$$\lambda_{iLH} \geq 0 \tag{A38}$$

$$\lambda_{iLL} \geq 0 \tag{A39}$$

$$\mu_i \geq 0 \tag{A40}$$

By assumption, $1 > p > q > 0$, and thus

$$\begin{aligned}&(1-\delta)\Big[(1-\kappa_i)(1-q)(1-p) + \kappa_i(1-q)^2\Big] + \delta(1-q)^2\\ &\quad > (1-\delta)\Big[(1-\kappa_i)(1-q)(1-q) + \kappa_i(1-q)^2\Big] + \delta(1-q)^2\\ &\quad = (1-q)^2\\ &\quad > (1-p)^2,\end{aligned}$$

so Equation (A25) can only be satisfied if $w_{iLL} = 0$. Otherwise, the complementary slackness condition (A35) would imply $\lambda_{iLL} = 0$ and Equation (A25) would be violated for any $\mu_i \geq 0$. Moreover, there exists no solution such that $\lambda_{iHH}, \lambda_{iHL}, \lambda_{iLH} > 0$: if that was true, then $w_{iHH} = w_{iHL} = w_{iLH} = w_{iLL} = 0$, and (A26) would be reduced to $-c \geq 0$, a contradiction.

Notice that $w_{iHH} > 0$ or $w_{iHL} > 0$ or $w_{iLH} > 0$, only if $\mu_i > 0$ and the terms in brackets in Equations (A22)–(A24), respectively, are strictly positive. Thus, in any solution, the dynamic incentive compatibility constraint must be binding.

In Equation (A22), it is easy to see that $(1-\delta)((1-\kappa_i)pq + \kappa_i q^2) + \delta q^2 < (1-\delta)((1-\kappa_i)pp + \kappa_i q^2) + \delta q^2 < (1-\delta)((1-\kappa_i)pq + \kappa_i p^2) + \delta p^2 = p^2$, so that $w_{iHH} > 0$ for any values of $\delta$ and $\kappa_i$. In Equation (A23), $p(1-p) > (1-\delta)((1-\kappa_i)q(1-p) + \kappa_i q(1-q)) + \delta q(1-q)$ if

$$\kappa_i < \overline{\kappa}(\delta) = \frac{p(1-p) - \delta q(1-q)}{(1-\delta)q(p-q)} - \frac{1-p}{p-q},$$

and, in Equation (A24), $p(1-p) > (1-\delta)((1-\kappa_i)(1-q)p + \kappa_i q(1-q)) + \delta q(1-q)$ if

$$\kappa_i > \underline{\kappa}(\delta) = \frac{\delta q(1-q) - p(1-p)}{(1-\delta)(1-q)(p-q)} + \frac{p}{p-q}.$$

Notice that

$$\overline{\kappa}(0) = \frac{1-p}{q}, \quad \underline{\kappa}(0) = \frac{p}{1-q},$$

and

$$\frac{\partial \overline{\kappa}(\delta)}{\partial \delta} \begin{cases} > 0 & \text{if } p+q < 1 \\ = 0 & \text{if } p+q = 1 \\ < 0 & \text{if } p+q > 1 \end{cases}, \frac{\partial \underline{\kappa}(\delta)}{\partial \delta} \begin{cases} < 0 & \text{if } p+q < 1 \\ = 0 & \text{if } p+q = 1 \\ > 0 & \text{if } p+q > 1 \end{cases}.$$

Moreover,

$$\lim_{\delta \to 1} \overline{\kappa}(\delta) = \begin{cases} +\infty & \text{if } p+q < 1 \\ -\infty & \text{if } p+q > 1 \end{cases}, \lim_{\delta \to 1} \underline{\kappa}(\delta) = \begin{cases} -\infty & \text{if } p+q < 1 \\ +\infty & \text{if } p+q > 1 \end{cases}.$$

As was the case in Proposition 2, if $p+q > 1$, then $\underline{\kappa}(0) > 1 > \overline{\kappa}(0) > 0$. Thus, for $\kappa_i \geq \overline{\kappa}(\delta)$, a team performance scheme $\mathbf{w}_i^{Team} = (w_{iHH}^{Team}, 0, 0, 0)$ such that

$$w_{iHH}^* = \left( \frac{c}{p^2 - (1-\delta)q[(1-\kappa_i)q + \kappa_i q] - \delta q^2} \right)^{\frac{1}{1-\rho}}$$

is optimal. For $\kappa_i < \overline{\kappa}(\delta)$, the relative performance scheme $\mathbf{w}_i^{Rel} = (w_{iHH}^{Rel}, w_{iHL}^{Rel}, 0, 0)$ is optimal, with

$$w_{iHL}^{Rel} = w_{iHH}^{Rel} \times \underbrace{\left( \frac{p}{1-p} \cdot \frac{p(1-p) - (1-\delta)q[(1-\kappa_i)(1-p) + \kappa_i(1-q)] - \delta q(1-q)}{p^2 - (1-\delta)q[(1-\kappa_i)p + \kappa_i q] - \delta q^2} \right)^{\frac{1}{\rho}}}_{=\mathcal{A}(\kappa_i, \delta, \rho)},$$

$$w_{iHH}^{Rel} = \left( \frac{c}{p[p + (1-p)\mathcal{A}] - (1-\delta)q\{(1-\kappa_i)[p + (1-p)\mathcal{A}] + \kappa_i[q + (1-q)\mathcal{A}]\} - \delta q[q + (1-q)\mathcal{A}])} \right)^{\frac{1}{1-\rho}}.$$

If $p+q < 1$, then $0 < \underline{\kappa}(0) < 1 < \overline{\kappa}(0)$. For $\kappa_i \leq \underline{\kappa}(\delta)$, the optimal incentive scheme is the relative performance described in the last paragraph. On the other hand, for $\kappa_i > \underline{\kappa}(\delta)$, the optimal incentive scheme is $\mathbf{w}_i^{Comp} = (w_{iHH}^{Comp}, w_{iHL}^{Comp}, w_{iLH}^{Comp}, 0)$ such that

$$w_{iLH}^{Comp} = w_{iHH}^{Comp} \times \underbrace{\left( \frac{p}{1-p} \cdot \frac{p(1-p) - (1-\delta)(1-q)[(1-\kappa_i)(1-p) + \kappa_i(1-q)] - \delta(1-q)^2}{p^2 - (1-\delta)q[(1-\kappa_i)p + \kappa_i q] - \delta q^2} \right)^{\frac{1}{\rho}}}_{=\mathcal{B}(\kappa_i,\delta,\rho)},$$

$$w_{iHL}^{Comp} = w_{iHH}^{Comp} \times \underbrace{\left( \frac{p}{1-p} \cdot \frac{p(1-p) - (1-\delta)q[(1-\kappa_i)(1-p) + \kappa_i(1-q)] - \delta q(1-q)}{p^2 - (1-\delta)q[(1-\kappa_i)q + \kappa_i q] - \delta q^2} \right)^{\frac{1}{\rho}}}_{=\mathcal{A}(\kappa_i,\delta,\rho)},$$

$$w_{iHH}^{Comp} = \left( \frac{c}{D(\kappa_i, \delta, \rho)} \right)^{\frac{1}{1-\rho}},$$

where

$$\begin{aligned} D(\kappa_i, \delta, \rho) = {} & p[p + (1-p)(\mathcal{A} + \mathcal{B})] \\ & - (1-\delta)\{(1-\kappa_i)[pq + q(1-p)\mathcal{A} + (1-q)p\mathcal{B}] + \kappa_i q[q + (1-q)(\mathcal{A} + \mathcal{B})]\} \\ & - \delta q[q + (1-q)(\mathcal{A} + \mathcal{B})]. \end{aligned}$$

## References


1. Laffont, J.J.; Tirole, J. Using Cost Observation to Regulate Firms. *J. Political Econ.* **1986**, *94*, 614–641. [CrossRef]
2. Bolton, P.; Dewatripont, M. Contract Theory. In *MIT Press Books*; The MIT Press: Cambridge, MA, USA, 2005; Volume 1.
3. Crawford, V.P.; Knoer, E.M. Job Matching with Heterogeneous Firms and Workers. *Econometrica* **1981**, *49*, 437–450. [CrossRef]
4. Besley, T.; Ghatak, M. Competition and Incentives with Motivated Agents. *Am. Econ. Rev.* **2005**, *95*, 616–636. [CrossRef]
5. Nagin, D.S.; Rebitzer, J.B.; Sanders, S.; Taylor, L.J. Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment. *Am. Econ. Rev.* **2002**, *92*, 850–873. [CrossRef]
6. Fehr, E.; Schmidt, K.M. Fairness, Incentives, and Contractual Choices. *Eur. Econ. Rev.* **2000**, *44*, 1057–1068. [CrossRef]
7. Fehr, E.; Klein, A.; Schmidt, K.M. Fairness and Contract Design. *Econometrica* **2007**, *75*, 121–154. [CrossRef]
8. Dohmen, T.; Falk, A.; Huffman, D.; Sunde, U. Homo Reciprocans: Survey Evidence on Behavioural Outcomes. *Econ. J.* **2009**, *119*, 592–612. [CrossRef]
9. Bowles, S.; Polania-Reyes, S. Economic Incentives and Social Preferences: Substitutes or Complements? *J. Econ. Lit.* **2012**, *50*, 368–425. [CrossRef]
10. Bergstrom, T.C. On the Evolution of Altruistic Ethical Rules for Siblings. *Am. Econ. Rev.* **1995**, *85*, 58–81.
11. Alger, I.; Weibull, J.W. Homo Moralis-Preference Evolution Under Incomplete Information and Assortative Matching. *Econometrica* **2013**, *81*, 2269–2302.
12. Alger, I.; Weibull, J.W. Evolution and Kantian Morality. *Games Econ. Behav.* **2016**, *98*, 56–67. [CrossRef]
13. Bergström, T.C. The Algebra of Assortative Encounters and the Evolution of Cooperation. *Int. Game Theory Rev. (IGTR)* **2003**, *5*, 211–228. [CrossRef]
14. Miettinen, T.; Kosfeld, M.; Fehr, E.; Weibull, J. Revealed preferences in a sequential prisoners' dilemma: A horse-race between six utility functions. *J. Econ. Behav. Organ.* **2020**, *173*, 1–25. [CrossRef]
15. Van Leeuwen, B.; Alger, I. Estimating Social Preferences and Kantian Morality in Strategic Interactions. TSE Working Papers. No. 19-1065. 2021. Available online: https://www.tse-fr.eu/publications/estimating-social-preferences-and-kantian-morality-strategic-interactions-0 (accessed on 18 March 2021).
16. Sobel, J. Interdependent Preferences and Reciprocity. *J. Econ. Lit.* **2005**, *43*, 392–436. [CrossRef]
17. Alchian, A.A.; Demsetz, H. Production , Information Costs, and Economic Organization. *Am. Econ. Rev.* **1972**, *62*, 777–795.
18. Holmström, B. Moral Hazard in Teams. *Bell J. Econ.* **1982**, *13*, 324–340. [CrossRef]
19. Che, Y.K.; Yoo, S.W. Optimal Incentives for Teams. *Am. Econ. Rev.* **2001**, *91*, 525–541. [CrossRef]
20. Itoh, H. Moral Hazard and Other-Regarding Preferences. *Jpn. Econ. Rev.* **2004**, *55*, 18–45. [CrossRef]
21. Demougin, D.; Fluet, C. Inequity Aversion in Tournaments. CIRPEE Working Paper. No. 03-22. 2003, pp. 1–24. Available online: http://www.cirpee.org/fileadmin/documents/Cahiers_2003/CIRPEE03-22.pdf (accessed on 18 March 2021).
22. Rey-Biel, P. Inequity Aversion and Team Incentives. *Scand. J. Econ.* **2008**, *110*, 297–320. [CrossRef]
23. Englmaier, F.; Wambach, A. Optimal incentive contracts under inequity aversion. *Games Econ. Behav.* **2010**, *69*, 312–328. [CrossRef]
24. Bartling, B. Relative performance or team evaluation? Optimal contracts for other-regarding agents. *J. Econ. Behav. Organ.* **2011**, *79*, 183–193. [CrossRef]
25. Livio, L.; De Chiara, A. Friends or foes? Optimal incentives for reciprocal agents. *J. Econ. Behav. Organ.* **2019**, *167*, 245–278. [CrossRef]

26. Rabin, M. Incorporating Fairness into Game Theory and Economics. *Am. Econ. Rev.* **1993**, *83*, 1281–1302.
27. Bénabou, R.; Tirole, J. Incentives and Prosocial Behavior. *Am. Econ. Rev.* **2006**, *96*, 1652–1678. [CrossRef]
28. Becker, G.S. A Theory of Social Interactions. *J. Political Econ.* **1974**, *82*, 1063–1093. [CrossRef]
29. Dur, R.; Sol, J. Social Interaction, Co-worker Altruism, and Incentives. *Games Econ. Behav.* **2010**, *69*, 293–301. [CrossRef]
30. Dur, R.; Tichem, J. Altruism and Relational Incentives in the Workplace. *J. Econ. Manag. Strategy* **2015**, *24*, 485–500. [CrossRef]
31. Kolm, S.; Ythier, J.M. (Eds.) *Handbook of the Economics of Giving, Altruism and Reciprocity*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2006; Volume 1.
32. Kolm, S.; Ythier, J.M. (Eds.) *Handbook of the Economics of Giving, Altruism and Reciprocity*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2006; Volume 2.
33. Laffont, J.J. Macroeconomic Constraints, Economic Efficiency and Ethics: An Introduction to Kantian Economics. *Economica* **1975**, *42*, 430–437. [CrossRef]
34. De la Rosa, L.E. Overconfidence and Moral Hazard. *Games Econ. Behav.* **2011**, *73*, 429–451. [CrossRef]