games

MDPI

Team Incentives under Moral and Altruistic Preferences: Which Team to Choose?

Roberto Sarkisian

Toulouse School of Economics, 21 Allee de Brienne, MF 003, Toulouse 31000, France; roberto.sarkisian@tse-fr.eu

Abstract: This paper studies incentives provision when agents are characterized either by homo moralis preferences, i.e., their utility is represented by a convex combination of selfish preferences and Kantian morality, or by altruism. In a moral hazard in a team setting with two agents whose efforts affect output stochastically, I demonstrate that the power of extrinsic incentives decreases with the degrees of morality and altruism displayed by the agents, thus leading to increased profits for the principal. I also show that a team of moral agents will only be preferred if the production technology exhibits decreasing returns to efforts; the probability of a high realization of output conditional on both agents exerting effort is sufficiently high; and either the outside option for the agents is zero or the degree of morality is sufficiently low.

1. Introduction

Teamwork permeates economic activities. In some cases, different skills are needed complementing each other to complete a project; in other cases, division of labor plays a crucial role in timely delivery of a product. Partnerships, group projects and team sports are but a few examples in which individuals team up to achieve a certain goal. With the exception of partnerships, it falls upon an employer to hire the team of employees to fulfill the task at hand. In particular, as expenditure in recruitment and assessment surges[1], with U.S. companies spending on average around US$4000 per hire, an increase of nearly 15% in the last four years, it is clear that recruitment and talent research divisions have turned their attention to more than the job applicants' professional abilities. As a matter of fact, common practice includes the analysis of criminal[2] and credit histories[3], and more recently, social networks as well[4].

While employees' technical skills are important, interest in their personal characteristics other than job-relevant skills may be related to the now widespread knowledge that economic agents are not purely selfish, often displaying other-regarding preferences [3–7]. The literature in behavioral and experimental economics strongly suggests that social preferences affect outcomes in standard economic models [8,9]. The works in [10–13] analyze, in particular, the role other-regarding preferences play in interactions among employees in models of the workplace. The main findings in this literature show that the employees' concerns towards one another affect not only the provision of effort, but also the compensation schemes that are offered. Thus, it is only natural to wonder what an ideal team

---

[1] See [1].
[2] See the Society for Human Resource Management survey results at https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Pages/criminalbackgroundcheck.aspx.
[3] See the Society for Human Resource Management survey results at https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Pages/creditbackgroundchecks.aspx
[4] See, for instance, [2].

would look like, for a given set of skills: Would it be a team composed of selfish individuals, whose only concern is their own gains, or maybe a group of altruistic agents, who would be content with increasing their workmates' wellbeing? Perhaps a crew of moral employees deriving satisfaction in choosing actions they think are the right ones? These are the questions I address in this paper.

In view of the overwhelming experimental evidence of behaviors that are incompatible with purely selfish preferences [4,5,14,15], it is important to understand how prosocial preferences affect behavior in the workplace and, by extension, the design of contracts in the workplace. I propose a model to address this question.

Specifically, I focus on the optimal compensation schemes that should be used in a standard moral hazard setting to incentivize the employees to fulfill their tasks. In doing so, I am able to compare the profits obtained by the employer from a team composed of individuals with different kinds of prosocial preferences. Although I do not study the recruitment process per se, I am able to make predictions about which preferences the principal would prefer.

The framework utilized is the multiagent moral hazard model, as first proposed by [16,17], where a risk-neutral principal hires a team of two risk-averse agents. The agents can exert costly effort in order to stochastically affect the realization of output. By assumption, efforts are simultaneously and independently chosen by the agents and cannot be observed by the principal. On the other hand, output is observable by third parties after being realized and can thus be contracted upon.

In the behavioral economics literature, several classes of prosocial preferences have been proposed. I analyze two of them. The first one is altruism [18][5], a class that has been extensively used in the literature on the voluntary contribution of public goods. This is natural since one can think of efforts made in the context of teamwork in a firm as contributions to a public good (the firm's profit). Second, in light of recent results by [20,21], who show that a particular, novel, class of preferences stands out as being favored by evolution, I compare the optimal contract under altruism with the optimal contract under this class of homo moralis preferences, a convex combination of selfishness and morality. In sum, my model allows addressing the following questions: if an employer could choose between a team of two moral agents and a team of two altruists, which team would he/she prefer and why[6]?

I characterize the optimal contract for a team of equally altruistic agents and a team of equally moral agents and compare them. First, I find that the trade-off between risk-sharing and incentive provision is present, as in the case with standard selfish preferences. However, as intuition would suggest, I find that high-powered incentives are less needed to induce effort as the agents become more concerned about the right thing to do or about each other's material payoff, and that the principal's expected profit obtained from the interaction with each team is increasing in the team's degree of morality or altruism. Second, if efforts are symmetric and could be contracted upon, the principal would be better off hiring a team of altruistic agents over the other ones, for any degrees of morality and altruism, because altruism towards one's partner reduces the payment necessary to induce participation, one effect that is not present with selfish or moral preferences. On the other hand, when efforts are not observable, which team is going to be preferred depends on the production technology: in particular, if the stochastic production technology displays increasing returns to efforts, the altruistic team is the least expensive to hire. This is a consequence of the different nature of each class of preferences. While altruistic agents derive benefits from increased material payoffs of their fellows, moral agents take satisfaction in doing the right thing. Intuitively, a higher effort under increasing returns drastically increases the expected material payoff of the agents, on which altruism is based. Meanwhile, the choice of the right thing to do depends only on the contract offered by the principal, and not on the production's underlying technology. Therefore, under increasing returns, altruistic

---

[5]　For a broader debate on the definition of altruism from an evolutionary biology view, see [19].
[6]　Reference [20] shows that homo moralis and altruistic preferences are behaviorally alike in many situations, and a similar point can be found in [22]; however, this is not the case in this exposition, as will be seen later on.

agents possess higher intrinsic motivation to exert the high level of effort, thus demanding a less high-powered contract and saving costs for the principal.

This paper is closely related to the moral hazard literature, in particular to two of its strands: moral hazard in teams and moral hazard with prosocial preferences. References [17,23] characterize the basic results on moral hazard in teams that are used to build the model below[7]. References [25–28] study optimal incentive schemes under different prosocial preferences: the first three focus on inequity aversion, while the last models agents exhibiting reciprocity concerns towards each other. References [29–31] consider moral hazard in team problems where the agents can monitor one another and, therefore, study the effects of peer pressure on the optimal contract design. Furthermore, [32–34] study problems where the agents' prosocial preferences are private information and give conditions for separating equilibria and self-selection by different types. None of them, however, raises the question of which preferences yield the least cost to the principal.

The analysis below differs from the previous literature in three crucial points: first, it considers homo moralis preferences, which has not, to the best of my knowledge, been done before in a contracting setting, thus presenting a simple environment where the principal can profitably explore idiosyncrasies generated by those and altruistic preferences. Second, it does not allow for monitoring, nor private information about the agents' preferences, so that I can focus solely on the effect of the prosocial preferences on the optimal contract design. Last, and more importantly, the analysis contrasts the optimal contracts under each class of preferences, and derives conditions under which the principal would prefer hiring one team over the other, therefore providing a rationale for firms to collect soft information on potential employees to compose teams that will minimize the total payments to be made.

I proceed as follows. Section 2 presents the environment, while Sections 3 and 4 study the optimal contract assuming efforts are contractible and non-contractible, respectively. Section 5 concludes. For the ease of exposition, all proofs are relegated to Appendix C.

2. The Model

I analyze the interaction between a principal and two agents, denoted by $i \in \{A, B\}$. The principal hires the two agents to work on a joint task, which generates revenue $x \in \{x^H, x^L\}$ for the principal, where $x^H > x^L$. Each agent can exert either a low or a high effort level $e_i \in \{0, 1\}$. Efforts determine revenues stochastically, according to the following probability distribution:

$$
\begin{array}{cc}
 & e_B = 1 \quad e_B = 0 \\
\begin{array}{c} e_A = 1 \\ e_A = 0 \end{array} & \left(\begin{array}{cc} p_2 & p_1 \\ p_1 & p_0 \end{array} \right)
\end{array}
$$

Throughout, I assume that revenue is never certain and that the probability of achieving a high outcome is increasing in the total effort exerted by the agents: $1 > p_2 > p_1 > p_0 > 0$.

If effort is costless, the assumption above indicates a preference of the principal for both agents to exert effort. However, effort is costly to each agent; for each $i = A, B$[8],

$$
c(e_i) = \begin{cases} c > 0 & \text{if } e_i = 1, \\ 0 & \text{if } e_i = 0. \end{cases}
$$

---

[7] Reference [24] study optimal incentives for teams in a repeated setting.
[8] in reality, agents may differ in their respective cost of effort, but this is not pursued in this paper because it does not qualitatively change the results, and at the same time, it adds a more cumbersome notation.

The principal offers the agents contracts $w_i(x)$, $i = A, B$, specifying payments that will follow each realization of revenues. The principal is assumed to be risk neutral, and his/her payoff is given by:

$$V(x, w_A(x), w_B(x)) = x - w_A(x) - w_B(x).$$

Denote by $\pi(e_i, e_j, w_i(x))$ the expected material payoff accruing to agent i from the effort choices (e_i, e_j) and wage schedule $w_i(x)$, for $i, j \in \{A, B\}, i \neq j$. I restrict attention to wage schedules pairs $\mathbf{w}_i = (w_i^H, w_i^L)$ determining the payments following good and bad realizations of revenues. In what follows, the material payoff function takes the expected additively separable form:

$$\pi(e_i, e_j, \mathbf{w}_i) = p_{e_i + e_j} \left[u(w_i^H) - c(e_i) \right] + (1 - p_{e_i + e_j}) \left[u(w_i^L) - c(e_i) \right], \tag{1}$$

where $u : \mathbb{R}_+ \to \mathbb{R}$ is the function that associates the agent's consumption utility with each amount of money. The dependence of i's expected material payoff on e_j comes from the effect of the other agent's effort on the probability distribution of revenues. The agents are risk averse towards wages: $u(w)$ is assumed to be twice-continuously differentiable, strictly increasing and strictly concave[9].

The principal faces either a team consisting of two agents characterized by homo moralis preferences with degree of morality $\kappa_i \in [0, 1]$, represented by the utility function:

$$U^{HM}(e_i, e_j, \mathbf{w}_i, \kappa_i) = (1 - \kappa_i)\pi(e_i, e_j, \mathbf{w}_i) + \kappa_i \pi(e_i, e_i, \mathbf{w}_i), \tag{2}$$

a team comprised of two altruistic agents, whose preferences are summarized by the utility function:

$$U^{Alt}(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \alpha_i) = \pi(e_i, e_j, \mathbf{w}_i) + \alpha_i \pi(e_j, e_i, \mathbf{w}_j), \tag{3}$$

for $\alpha_i \in [0, 1]$ and $i, j \in \{A, B\}, i \neq j$. Both specifications take the standard selfish preferences as a special case ($\kappa_i = \alpha_i = 0$), and this will allow comparisons between the results to be presented below and the benchmark moral hazard problem.

As pointed in [20,22], this specification of preferences for altruistic agents gives rise to the behavioral equivalence between homo moralis preferences and altruism for $\alpha_i = \kappa_i$ in many classes of games. With that in mind, I will make the following assumption for the rest of the exposition.

Assumption 1. $\alpha_A = \alpha_B = \kappa_A = \kappa_B = \theta$.

Thus, the agents' utility functions are simplified to:

$$U^{HM}(e_i, e_j, \mathbf{w}_i, \theta) = (1 - \theta)\pi(e_i, e_j, \mathbf{w}_i) + \theta\pi(e_i, e_i, \mathbf{w}_i), \tag{4}$$

$$U^{Alt}(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \theta) = \pi(e_i, e_j, \mathbf{w}_i) + \theta\pi(e_j, e_i, \mathbf{w}_j). \tag{5}$$

The relationship among the three parties unfolds as follows. First, the principal offers each agent a contract $\mathbf{w_i}$, which can be either accepted or rejected by the agents. If at least one agent rejects the contract, the game ends, and every party receives his/her own reservation utility. If both agents accept the principal's offers, they play a normal form game[10]: both of them must simultaneously and independently choose an effort level, from which revenues will be realized according to the probability distribution given by the production technology above. Payments are made according to the schedules proposed by the firm, and the agents' payoffs in the normal form game are given by their expected

---

9   The assumption that $u(\cdot)$ is strictly concave can be relaxed, and the same model below can be solved in a setting with risk-neutral agents and limited liability constraints, where the qualitative results are not changed from the analysis below.

10   The normal form game here is comprised of the set of players $\{A, B\}$, the common set of pure strategies $S = \{0, 1\}$ and payoff function $U(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \theta)$.

utilities with regard to received wages, efforts and preferences. While each agent's effort choice is private information, revenues and wages are publicly observable. It is also assumed that the agents' preferences are common knowledge[11].

3. Studying the Benchmark: The Contractible Effort Case

As a starting point, I derive the optimal contract assuming efforts are observable and contractible by the principal, to serve as a benchmark for later results. In what follows, I assume that each agent possesses an outside option that gives him/her utility $\overline{u} \geq 0$ if he/she does not accept the principal's contract offer. Therefore, agent $i \in \{A, B\}$ is willing to participate in the proposed relationship iff:

$$U(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \theta) \geq \overline{u}. \tag{IR}$$

As discussed in the previous section, standard selfish preferences are a particular case of both homo moralis and altruistic preferences, and for the ease of exposition, I begin this and the next section by analyzing the optimal contract for that instance. Thus, under contractible efforts, the standard Borch rule:

$$\frac{p(e_i, e_j)}{u'(w_i^H) p(e_i, e_j)} = \frac{1 - p(e_i, e_j)}{u'(w_i^L)[1 - p(e_i, e_j)]}$$

gives:

$$\frac{u'(w_i^H)}{u'(w_i^L)} = 1, \tag{BR_S}$$

which implies $w_i^H = w_i^L = w_i = u^{-1}(\overline{u} + c(e_i))$ for $e_i \in \{0, 1\}$. The intuition here is the same as in the classical moral hazard problem with one principal and one agent: if effort is contractible, the principal optimally offers a constant wage schedule remunerating the agent according to his/her reservation utility and the cost of the principal's desired level of effort.

When the principal faces a team of altruistic agents, he/she solves:

$$\max_{\mathbf{w}_A, \mathbf{w}_B} \quad p(e_A, e_B)\left(x_H - w_A^H - w_B^H\right) + (1 - p(e_A, e_B))\left(x_L - w_A^L - w_B^L\right)$$

$$s.t. \quad \left[p(e_A, e_B)u(w_A^H) + (1 - p(e_A, e_B))u(w_A^L) - c(e_A)\right]$$

$$+ \theta\left[p(e_A, e_B)u(w_B^H) + (1 - p(e_A, e_B))u(w_B^L) - c(e_B)\right] \geq \overline{u} \quad (IR_A)$$

$$\left[p(e_A, e_B)u(w_B^H) + (1 - p(e_A, e_B))u(w_B^L) - c(e_B)\right]$$

$$+ \theta\left[p(e_A, e_B)u(w_A^H) + (1 - p(e_A, e_B))u(w_A^L) - c(e_A)\right] \geq \overline{u} \quad (IR_B)$$

An interior solution is characterized by the KKT first-order conditions:

$$- p(e_i, e_j) + \lambda_i p(e_i, e_j)u'(w_i^H) + \lambda_j \theta p(e_i, e_j)u'(w_i^H) = 0$$

$$- [1 - p(e_i, e_j)] + \lambda_i[1 - p(e_i, e_j)]u'(w_i^L) + \lambda_j \theta[1 - p(e_i, e_j)]u'(w_i^L) = 0,$$

so the Borch rule becomes[12]:

$$\frac{u'(w_i^H)}{u'(w_i^L)} = 1, \tag{BR_{Alt}}$$

---

[11]　This is the reason why contracts are indexed by i.

[12]　Since $1 \geq p_2 > p_1 > p_0 \geq 0$ and $u' > 0$ by assumption, the first-order conditions imply that $\lambda_i + \lambda_j \theta > 0$ for $i, j \in \{A, B\}, i \neq j$.

for any choices of effort $(e_A, e_B) \in \{0,1\}^2$ and $i, j \in \{A, B\}$, $i \neq j$. Therefore, if agents are altruistic, the optimal contract under verifiable efforts proposes a constant wage schedule, just as was the case in the benchmark selfish preferences, given by:

$$w^{Alt} = u^{-1}\left(\frac{1}{1+\theta}\overline{u} + c(e_i)\right), \tag{6}$$

which is well-defined for all $\theta \in [0,1]$. It is easy to see that for $\theta = 0$, this is exactly the same expression as for the optimal contract under verifiable efforts in the benchmark case, while that for any positive degree of altruism is lower than it would be for selfish agents. Intuitively, each altruistic agent recognizes that by participating in the relationship, his partner's material payoff increases, and therefore, the wage demanded for participation declines when the degree of altruism becomes larger.

Finally, consider the team with homo moralis preferences, in which agent $i \in \{A, B\}$ is willing to participate in the proposed relationship iff:

$$U^{HM}(e_i, e_j, \mathbf{w}_i, \theta) = (1 - \theta)\pi(e_i, e_j, \mathbf{w}_i) + \theta\pi(e_i, e_i, \mathbf{w}_i) \geq \overline{u},$$

i.e., iff

$$u(w_i^H)[(1-\theta)p(e_i, e_j) + \theta p(e_i, e_i)] + u(w_i^L)[(1-\theta)(1 - p(e_i, e_j)) + \theta(1 - p(e_i, e_i))] - c(e_i) \geq \overline{u}.$$

Some points are noteworthy. First, as mentioned before, in the case where $\theta = 0$, this participation constraint reduces to the usual (IR) constraint in the benchmark moral hazard problem, since the selfish preference is a particular case of this framework. Second, for $\theta = 1$, $U^\kappa(e_i, e_j, \mathbf{w}_i, 1) = \pi(e_i, e_i, \mathbf{w}_i)$. In this case, agent i's choice of effort does not depend on agent j's effort choice, and choosing e_i becomes an individual decision problem. Third, if $e_i = e_j = e \in \{0,1\}$, the participation constraint collapses into:

$$p(e, e)u(w_i^H) + (1 - p(e, e))u(w_i^L) - c(e) \geq \overline{u},$$

Note here that the agents' degrees of morality are irrelevant, and the participation constraints are exactly the same as those that would be obtained in a symmetric equilibrium in the benchmark moral hazard problem: by imposing $e_i = e_j$, both expected material payoffs terms are identical, and since the utility function is constructed as a convex combination of these functions, the expressions above are obtained.

By Assumption 1, every agent in each team is identical to his/her partner, since the only source of heterogeneity in the general formulation was given by the preferences. Therefore, I will restrict attention to symmetric choices of effort $e_A = e_B$ in the rest of the discussion[13].

Proposition 1. *Suppose Assumption 1 holds. Then, there exists $c^* > 0$ such that for all $c \in (0, c^*)$, the principal induces agents in the teams of moral or altruistic agents to exert high effort by means of a constant wage.*

This result is not surprising: if efforts are contractible, the principal compensates the agents with a fixed transfer in case they exert the desired level of effort or punishes them if there is a deviation. Furthermore, if the cost of exerting effort is small, then the amount the principal has to transfer back to the agents in order to have an increased chance of obtaining a high realization of revenues is also small and, thus, profitable to implement. Moreover, since I restrict attention to symmetric equilibria,

---

[13]　In Appendix B, I show that relaxing both these assumptions leads to a Borch rule for moral agents that demands nonconstant wages when efforts are observable, in stark contracts to the literature with selfish and altruistic agents.

the degree of morality plays no role when I consider a team of homo moralis agents: the compensation schedule and effort choices are exactly the same as those obtained in the benchmark problem.

Now, I can focus on the central question of the paper: given the optimal contracts that induce the desired level of effort, which team should the principal hire?

Proposition 2. *Suppose Assumption 1 holds, efforts are verifiable and the principal wants both agents to make the high effort. For any $\theta \in [0, 1]$, the principal prefers hiring the team of altruistic agents over the team of selfish or moral agents.*

Intuitively, in situations where someone fails to do the right thing, a moral agent derives part of his/her utility from contemplating what would happen if everyone did the right thing. If both agents do in fact exert high effort, the contemplation in question does not add utility beyond the material utility that the agents thus obtain. By contrast, for altruistic agents, any choice but high effort decreases the material payoff of both agents and, consequently, all of the utility of each altruistic employee. Therefore, intrinsic motivation is larger for altruistic agents, and a team comprised of such employees is less costly for the principal.

4. Moving to the Second Best: Non-Contractible Efforts

Throughout the rest of the exposition, I focus on contracts that induce both agents to participate in the relationship and also exert the high level of effort ($e = 1$).

As a benchmark, focus first on standard selfish preferences. If efforts are non-contractible and the principal wishes to induce both agents to exert effort, he/she must solve, for $i, j \in \{A, B\}, i \neq j$,

$$\max_{\mathbf{w}_A, \mathbf{w}_B} \quad p_2(x^H - w_A^H - w_B^H) + (1 - p_2)(x^L - w_A^L - w_B^L)$$

$$\text{s.t.} \quad p_2\left[u(w_i^H) - c\right] + (1 - p_2)\left[u(w_i^L) - c\right] \geq p_1 u(w_i^H) + (1 - p_1)u(w_i^L) \quad (IC_i)$$

$$p_2\left[u(w_i^H) - c\right] + (1 - p_2)\left[u(w_i^L) - c\right] \geq \overline{u}. \quad (IR_i)$$

Manipulating the incentive compatibility constraint yields:

$$u(w_i^H) - u(w_i^L) \geq \tfrac{c}{p_2 - p_1}. \quad (IC_i)$$

By assumption, $c > 0$ and $p_2 > p_1$. Thus, the incentive compatibility constraint implies a monotonicity constraint on the wages following a good and a bad realization of output, since $u(\cdot)$ is assumed to be strictly increasing. Standard arguments show that both the incentive compatibility and the individual rationality constraints must bind at the optimum, so that the solution to the principal's problem is a contract $\mathbf{w}_S = (w_S^H, w_S^L)$ such that:

$$u(w_S^L) = \overline{u} - \frac{p_1 c}{p_2 - p_1},$$

$$u(w_S^H) = \overline{u} + \frac{(1 - p_1)c}{p_2 - p_1}.$$

Given the incentive compatibility constraint, it is clear that $\Delta w_S \equiv w_S^H - w_S^L > 0$.

Of course, if the principal wishes to induce the agents not to exert effort, a constant wage schedule $w^H = w^L = w = u^{-1}(\overline{u})$ would be optimal. Comparison of the principal's profits when agents exert effort and shirk show that the former is preferred by the employer for any $c \leq \overline{c}_S, 0 < \overline{c}_S < c^*$.

Under altruistic preferences for the agents, the principal's problem is:

$$\max_{w^H, w^L} \quad p_2(x^H - 2w^H) + (1 - p_2)(x^L - 2w^L)$$

$$\text{s.t.} \qquad (1 + \theta)[p_2 u(w^H) + (1 - p_2)u(w^L) - c] \geq \bar{u} \qquad (IR)$$

$$(1 + \theta)\left[p_2 u(w^H) + (1 - p_2)u(w^L) - c\right] \geq$$
$$\left[p_1 u(w^H) + (1 - p_1)u(w^L)\right] + \theta\left[p_1 u(w^H) + (1 - p_1)u(w^L) - c\right]. \quad (IC)$$

Rewrite the incentive compatibility constraint as:

$$u(w^H) - u(w^L) \geq \frac{c}{(1 + \theta)(p_2 - p_1)}, \tag{7}$$

and notice that the right-hand side is strictly decreasing in the degree of altruism α. The intuition behind this is the tradeoff between explicit and intrinsic incentives. Indeed, as the agent cares less about his/her own material payoff relative to that of his/her teammate, the intrinsic incentive derived from an increase in the probability of a high realization of output (and a consequent increase in the expected material benefit of his/her partner) becomes larger than the explicit incentives given by a high powered contract in inducing the agent to exert the high level of effort.

The proposition below characterizes the optimal contract[14].

Proposition 3. *Suppose Assumption 1 holds. There exists $\bar{c}_{Alt} > \bar{c}_S$ such that, for all $c < \bar{c}_{Alt}$, it is optimal for the principal to induce both altruistic agents to exert effort, $e_A = e_B = 1$, by means of a contract* $\mathbf{w_{Alt}}^* = (w_{Alt}^H, w_{Alt}^L)$ *such that:*

$$\Delta w_{Alt} \equiv w_{Alt}^H - w_{Alt}^L \leq \Delta w_S$$

$\forall \theta \in [0, 1]$, *with strict inequality for any $\theta > 0$.*

Close inspection of the incentive compatibility constraint shows that any contract that would induce a selfish agent to exert the high effort would also induce an altruistic employee to do the same. Furthermore, for any given contract, an increase in θ would increase the utility of each agent. Hence, the principal can profit by reducing both wages[15]. This argument is formally stated below.

Corollary 1. *Suppose Assumption 1 holds. Then, the principal's expected profits are strictly increasing in θ.*

Last, I consider homo moralis preferences. The principal must choose wage vectors $\mathbf{w}_A, \mathbf{w}_B$ to solve:

$$\max_{\mathbf{w}_A, \mathbf{w}_B} \quad p_2\left(x_H - w_A^H - w_B^H\right) + (1 - p_2)\left(x_L - w_A^L - w_B^L\right)$$

$$\text{s.t.} \qquad (1 - \theta)\left[p_2(u(w_i^H) - c) + (1 - p_2)(u(w_i^L) - c)\right]$$
$$+ \theta\left[p_2(u(w_i^H) - c) + (1 - p_2)(u(w_i^L) - c)\right] \geq$$
$$(1 - \theta)\left[p_1 u(w_i^H) + (1 - p_1)u(w_i^L)\right]$$
$$+ \theta\left[p_0 u(w_i^H) + (1 - p_0)u(w_i^L)\right] \qquad (IC_i)$$
$$(1 - \theta)\left[p_2(u(w_i^H) - c) + (1 - p_2)(u(w_i^L) - c)\right]$$
$$+ \theta\left[p_2(u(w_i^H) - c) + (1 - p_2)(u(w_i^L) - c)\right] \geq \bar{u} \quad (IR_i)$$

---

[14] In Appendix A, I show that the optimal contracts for both moral and altruistic agents may lead to a multiplicity of equilibria in their effort choices, as in [17].
[15] See Appendix C for the proof.

for $i = A, B$. Note that the individual rationality constraint can be rewritten in the simpler form:

$$p_2 u(w_i^H) + (1 - p_2)u(w_i^L) - c \geq \bar{u},$$

since, in equilibrium, the principal's offer induces the symmetric effort choice $e_A = e_B = 1$.

The incentive compatibility constraint also simplifies to:

$$p_2 u(w_i^H) + (1 - p_2)u(w_i^L) - c \geq$$
$$(1 - \theta)\left[p_1 u(w_i^H) + (1 - p_1)u(w_i^L) \right] + \theta \left[p_0 u(w_i^H) + (1 - p_0)u(w_i^L) \right].$$

The right-hand side of this inequality highlights an interesting fact: a positive degree of morality implies the agent internalizes the cost of choosing a low effort by evaluating what would happen if the other agent also were to make the same decision. This is very different in nature to how an altruistic agent evaluates any deviation: while the latter considers only the effects of his/her own deviation on his/her own material payoff and on his/her partner's, the former would consider the effect of the same deviation being made by his/her partner on his/her own payoff. Besides, by force of the assumptions presented above, further manipulation of (IC_i) yields:

$$u(w_i^H) - u(w_i^L) \geq \frac{c}{(p_2 - p_1) + \theta(p_1 - p_0)}, \tag{8}$$

where $\frac{c}{(p_2 - p_1) + \theta(p_1 - p_0)} > 0$. Because of the strict concavity of $u(\cdot)$, the incentive compatibility constraint for moral agents also implies a monotonicity condition on the optimum compensation schedules offered by the principal, even when the agents display the highest degree of morality. This last remark implies that the intrinsic incentives of the most moral agent are not sufficiently large to overcome the need to provide him/her with explicit incentives to exert the high level of effort.

Proposition 4. *Suppose Assumption 1 holds. There exists $\bar{c}_{HM} > \bar{c}_S$ such that, for all $c < \bar{c}_{HM}$, it is optimal for the principal to induce both moral agents to exert effort, $e_A = e_B = 1$, by means of a contract* $\mathbf{w_{HM}}^* = (w_{HM}^H, w_{HM}^L)$, *such that:*

$$\Delta w_{HM} \equiv w_{HM}^H - w_{HM}^L \leq \Delta w_S$$

$\forall \theta \in [0, 1]$, *and strict inequality for $\theta > 0$.*

The intuition behind the monotonicity constraint is the same as in the benchmark model: if wages following a low realization of revenues were larger than their counterpart after a good realization, then agents would prefer to exert low effort in order to receive this higher compensation and save in the cost of exerting effort.

The novelty in the results relates to how the compensation schedules vary with respect to the degree of morality θ. Keeping in mind that $u' > 0$, one can see that as θ increases, the right-hand side of (IC_i) becomes ever smaller, albeit positive. This implies that the gap in wages following good and bad realizations of revenues must decrease, since the incentive compatibility constraint binds, but monotonicity still holds. Intuitively, the principal can reduce the compensation over high realizations of revenues given to an agent who is very concerned about doing the right thing. However, at the same time, he/she must increase wages after bad outcomes in order to satisfy the participation constraint.

Because of this diminishing wage gap, intuition would suggest that the first-best result is obtained for a sufficiently high degree of morality. However, this is not the case. To see this, take $\theta = 1$, where agent i's preferences are purely Kantian, and thus, his/her utility is completely characterized by the expected material payoff $\pi(e_i, e_i)$. Although the participation constraint does not vary with

the agent's degree of morality[16], the same is not true for the incentive compatibility constraint. Now, when considering the pros and cons of a deviation in terms of effort choice, agent i internalizes what would happen if agent j were to do the same. Specifically, this entails a reduction in the probability of the good revenue being realized from p_2 to p_0, instead of the reduction to p_1 in the selfish term. This internalization is reflected in the (IC_i) constraint, which becomes:

$$u(w_i^H) - u(w_i^L) = \frac{c}{p_2 - p_0} > 0.$$

The denominator on the right-hand side is exactly the difference in the probabilities discussed above. Taking $\theta = 1$ makes the incentive compatibility constraint for a team of moral agents as easy to satisfy as possible, but it still binds, thus pushing the optimal contract away from the first-best one (constant wage schedule).

Given this behavior of wage schedules with respect to the degree of morality, a natural question to be asked is whether the principal is better off with highly moral agents or not. The answer is unconditional and presented in the following result.

Corollary 2. *The principal's expected profit is strictly increasing in θ.*

Corollary 2 contrasts with the contractible effort case, where the principal's profits were identical when hiring a team of selfish agents or a team of moral agents, for any degree of morality the last would display. Mathematically, the result is a consequence of the individual rationality constraints being identical in both cases, while the incentive compatibility constraint has a smaller right-hand side under moral agents than under selfish ones. Intuitively, the principal exploits the agents' morality, as he/she did with altruistic employees, as well, to induce high effort by means of less high-powered incentives, while inducing participation with a slightly increased payment after a bad realization of output. Thus, one concludes that the expected savings in wages after a good realization made by the principal by choosing a high morality agent offsets the expected increase in payments after low revenues.

One remark is in order here. Because of the assumption that $1 > p_2 > p_1 > p_0 > 0$ and the monotonicity condition implied by the incentive compatibility constraints for each preferences, it is the case that $U(0, 1, \mathbf{w}; \theta) \geq U(0, 0, \mathbf{w}; \theta)$, where $\mathbf{w}$ is the optimal contract offered by the principal. Thus, using the (IC) constraints again, I have that $U(1, 1, \mathbf{w}; \theta) \geq U(0, 0, \mathbf{w}; \theta)$, so the agents have no incentives to jointly deviate to shirking. The same is also true for altruistic agents.

So far, I have showed that the principal can attain higher profits by exploiting the agents' morality or altruism, thus reducing high-powered explicit incentives in the optimal contract in such a way that participation and incentives to exert high effort are still satisfied. Therefore, from the employer's perspective, knowing which class of preferences demands the least amount of explicit incentives is crucial. Lemma 1 tells us that the answer to that question depends on the stochastic production technology.

Lemma 1. *Under Assumption 1, $\left[u(w_{HM}^H) - u(w_{HM}^L)\right] - \left[u(w_{Alt}^H) - u(w_{Alt}^L)\right]$ has the same sign as $(p_2 - p_1) - (p_1 - p_0)$.*

In other words, if the stochastic technology presents increasing returns on aggregate efforts, the optimal contract under homo moralis preferences is (weakly) more high-powered than its counterpart under altruism when $\kappa = \alpha$: any contract inducing moral agents to exert high effort would do the same to altruistic employees. The converse is true if the technology has decreasing returns on

---

16   For symmetric equilibrium choices of effort.

efforts. This can be seen in Figure 1. The middle (green) line represents the incentive compatibility constraint for altruistic agents, whose format is not affected by the production technology. The top (red) and the bottom (blue) lines are the graphic representations of the (IC) constraint for moral agents when $p_2 - p_1 > p_1 - p_0$ and $p_2 - p_1 < p_1 - p_0$, respectively[17].
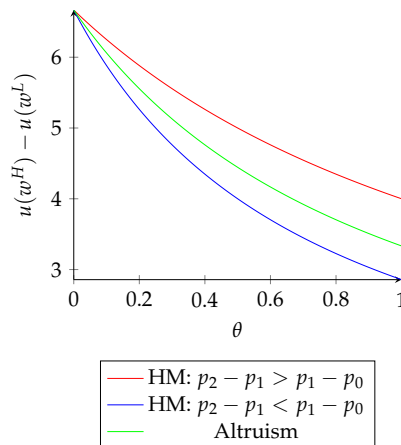

Figure 1. Comparing the power of optimal contracts.

Given the result above, one would expect that the principal's expected payoff will be uniformly higher if the agents are altruistic rather than moral when the production technology presents increasing returns to efforts, while the opposite would be true if decreasing returns are present. The flaw with such a logic is not considering the effects of the binding individual rationality constraints, which implied under contractible efforts that the team of altruistic agents was always the least expensive to hire. In particular, remember that for altruistic agents, the outside option $\bar{u}$ is divided by $1 + \theta$ in the participation constraint, a factor that is not present under selfish and moral preferences. This implies that w_{Alt}^L should also be smaller than w_{HM}^L[18]. However, the principal has clear preferences over the composition of the team, and the result below precisely states when one team is preferred over the other.

Theorem 1. *Assume the principal offers contracts* $\mathbf{w}_{HM}$ *and* $\mathbf{w}_{Alt}$ *to homo moralis and altruistic agents, respectively, inducing them to exert the high level of effort. Furthermore, assume Assumption 1 holds. Then, if the stochastic production technology exhibits:*

1. *increasing returns to efforts ($p_2 - p_1 \geq p_1 - p_0$), the principal is better off hiring a team of altruistic agents over a team of moral agents,*
2. *decreasing returns to efforts ($p_2 - p_1 < p_1 - p_0$) and*

    - *the outside option is zero ($\bar{u} = 0$), the principal prefers a team of moral agents; or*
    - *the outside option is positive, and the degree of morality is sufficiently low ($\bar{u} > 0$, $\kappa \to 0$); the principal prefers a team of moral agents only if $p_2 > \bar{p}_2 \in (0, 1)$.*

Under increasing returns to efforts, an altruistic team is less expensive for the principal because of two reasons. First, the wage that must be paid after a bad realization of output is smaller than its counterparts under selfish or moral preferences, and this is a consequence of the fact that the former's consideration with regards to the payoff of his/her partner slackens the participation constraint.

17　If $p_2 - p_1 = p_1 - p_0$, both lines coincide with the (IC) for altruistic agents.
18　This intuition in indeed right and integrates the proof of Theorem 1 in the Appendix.

On he other hand, such a concern also slackens the incentive compatibility constraint in this case, because exerting efforts drastically increases the probability of being successful, thus providing implicit incentives for the altruistic worker to exert effort and requiring a less high-powered contract to be proposed by the employer.

Such a difference in the intrinsic incentives to exert effort disappear when the production technology has constant returns, so that the power of the contract remains the same for both teams. However, it is still the case that the principal exploits the fact that altruistic agents derive utility from each other's material payoff and can thus pay them less.

The third case, with decreasing returns to efforts, is the most interesting, because the preference of the principal results from the net effect of two opposing forces. While it is still true that $w_{Alt}^L \leq w_{HM}^L$, Lemma 1 states that now, the power of the contract required by moral agents is smaller than the one for altruistic agents. In the range where such a reduction is the most drastic, the principal will prefer the team of moral agents rather than altruistic ones. The first condition for this to happen is that the probability of a success when both agents are exerting effort is sufficiently high, as can be seen from the incentive compatibility constraints. The second condition is that either the outside option for the agents is zero or if it is positive, the degree of morality or altruism is close to zero. If both cases, the participation constraints for moral and altruistic agents become arbitrarily close (identical if $\overline{u} = 0$) so that the exploitability of altruistic preferences, described in the preceding paragraph, becomes small, and the principal profits by hiring the agents demanding the least powered contracts: the moral agents in this case.

Figure 2a–c provides an example of Theorem 1 for $u(w) = \sqrt{w}$. Figure 2a represents the case where the production technology exhibits increasing returns to efforts. With the exception of $\theta = 0$, where both teams are identical to the selfish agents, the principal's profit is higher with a team of altruistic agents (V^{Alt}) than with a team of moral agents (V^{HM}).
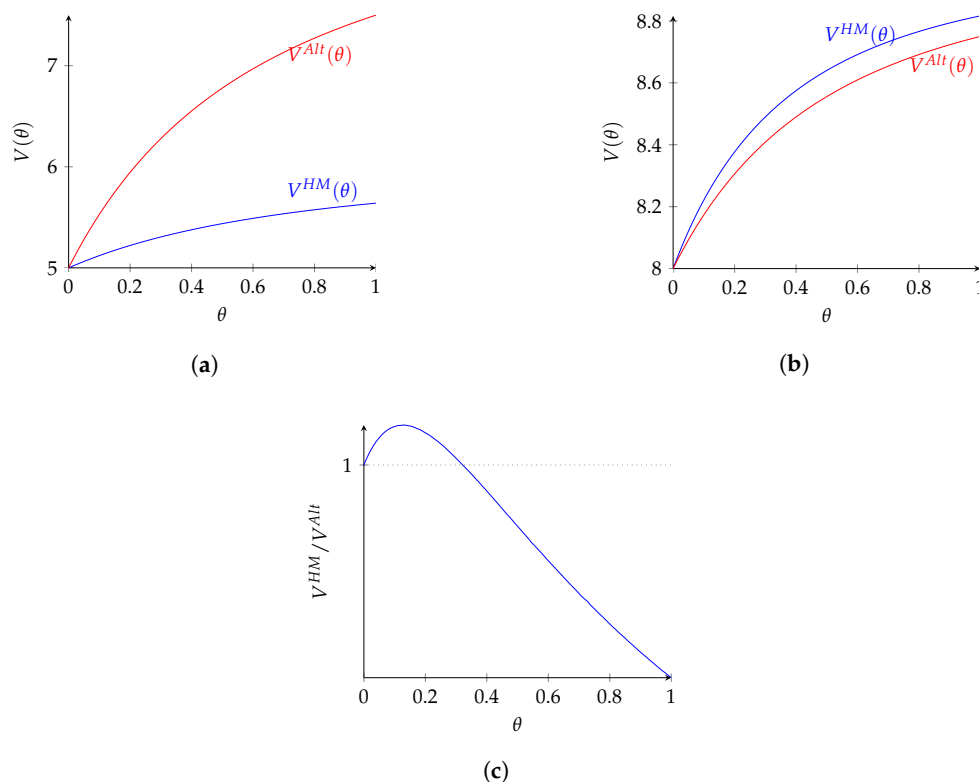
(a)

(b)

(c)

Figure 2. (**a**) Comparing principal's profits for $p_2 - p_1 \geq p_1 - p_0$; (**b**) comparing principal's profits for $p_2 - p_1 < p_1 - p_0$ and $\overline{u} = 0$; (**c**) comparing the ratio of principal's profits for $p_2 - p_1 < p_1 - p_0$ and $\overline{u} > 0$.

Figure 2b exemplifies the case with decreasing returns to efforts, the zero outside option for the agents and a high probability of success if both agents exert effort (namely, I set $p_2 = 0.9$). As Theorem 1 states, under these conditions, $V^{HM}(\theta) \geq V^{Alt}(\theta)$ for all equal degrees of morality and altruism.

Finally, Figure 2c plots the ratio V^{HM}/V^{Alt} for decreasing returns to efforts and $\bar{u} = 0.2$. Contrary to the previous case where $\bar{u} = 0$, the difference in the participation constraints for moral and altruistic agents makes it unprofitable for the employer to hire the moral team if θ becomes larger, since the decrease in w^L_{Alt} would be, in expected terms, sufficient to compensate the savings related to the power of the contract. This is represented by the region in the figure in which $V^{HM}/V^{Alt} \leq 1$.

Therefore, a rationale in terms of the principal's expected profits is given for trying to sort employees with respect to their preferences. If the production technology exhibits increasing returns with respect to efforts, the principal's choice is straightforward: always choose to employ altruistic agents. However, if the condition does not hold, employing moral individuals may lead to higher profits in comparison to both altruistic and purely selfish agents.

5. Concluding Remarks

This paper presents a comparison between the optimal contracts offered to teams of agents, who may be characterized by either homo moralis preferences or altruism towards each other. These contracts were explored in situations where the teams have only two agents with binary choices of efforts, affecting stochastically the revenues accrued by the principal.

Under contractible efforts, I show that altruistic agents are more exploitable by the principal, in the sense that the employers need to pay a smaller wage to induce the participation of the those agents when compared to the case where he/she would hire a team of selfish or moral employees. When efforts are no longer contractible, this exploitability also shows up for moral agents, and I show that the larger the degree of altruism or morality displayed by the members of each team, the higher the expected profits for the principal. Then, the natural question is which class of preferences would require smaller wages to exert effort and participate in the contractual relationship?

The main finding is that the principal obtains a higher expected profit hiring a team composed of moral agents under restrictive conditions: first, the stochastic technology exhibits decreasing returns with respect to efforts; second, the outside option of the agents yields zero utility; or third, the degree of morality is sufficiently low.

It is noteworthy that even in such a simple environment, prosocial preferences affect the contractual design, by adding a third channel to the traditional trade-off between risk-sharing and incentive provision. In effect, the principal will be better off employing a team of either altruistic or moral agents instead of a team composed solely of selfish employees, since a higher degree of morality and altruism decreases the amount of explicit incentives provided by the optimal contracts to induce the agents to exert effort. However, this additional channel is not enough to completely extinguish the need for explicit incentives even when the agents are purely moral or altruistic. Because it is more costly to the principal to hire a team of selfish agents, the exploitability of prosocial preferences can thus explain the costly acquisition of job applicants' soft information in the labor market.

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

Appendix A. Multiplicity of Equilibria

The optimal contracts derived in the main text are such that the principal can induce both agents to exert the high level of effort $e = 1$, both for the homo moralis or altruistic teams. However, the strategy

profile $(e_A, e_B) = (1, 1)$ may not be the unique Nash equilibrium of the simultaneous game played by the pair agents for a given contract. Indeed, let $U(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j; \theta)$ denote agent i's expected utility under this contract and degree of morality or altruism θ[19]. Then, the stage game played by Agents 1 and 2 is represented by

	$e_B{=}1$	$e_B{=}0$
$e_A = 1$	$U(1, 1, \mathbf{w}_A, \mathbf{w}_B; \theta)$	$U(1, 0, \mathbf{w}_A, \mathbf{w}_B; \theta)$
$e_A = 0$	$U(0, 1, \mathbf{w}_A, \mathbf{w}_B; \theta)$	$U(0, 0, \mathbf{w}_A, \mathbf{w}_B; \theta)$

From the principal's problem, the incentive compatibility constraint implies that:

$$U(1, 1, \mathbf{w}_A, \mathbf{w}_B; \theta) \geq U(0, 1, \mathbf{w}_A, \mathbf{w}_B; \theta),$$

i.e., given that the other agent is already exerting the high level of effort, it is not profitable for agent A to shirk when his/her compensation follows the optimal contract $\mathbf{w}^*$. Since this is true for both agents, it follows that $(e_A, e_B) = (1, 1)$ is a pure strategy Nash equilibrium for $\mathbf{w}^*$.

However, the comparison between $U(1, 0, \mathbf{w}_A, \mathbf{w}_B; \theta)$ and $U(0, 0, \mathbf{w}_A, \mathbf{w}_B; \theta)$ is not clear. In particular, if the latter is greater than the former, then $(e_A, e_B) = (0, 0)$ would constitute another pure strategy Nash equilibrium for $\mathbf{w}^*$.

Under homo moralis preferences, note that:

$U(0, 0, \mathbf{w}^*; \theta) \geq U(1, 0, \mathbf{w}^*; \theta) \Leftrightarrow$

$p_0 u(w^H) + (1 - p_0) u(w^L) \geq (1 - \theta)[p_1 u(w^H) + (1 - p_1) u(w^L) - c] + \theta[p_2 u(w^H) + (1 - p_2) u(w^L) - c] \Leftrightarrow$

$c \geq u(w^H)[(1 - \theta) p_1 + \theta p_2 - p_0] + u(w^L)[(1 - \theta)(1 - p_1) + \theta(1 - p_2) - (1 - p_0)] \Leftrightarrow$

$c \geq u(w^H)[(1 - \theta) p_1 + \theta p_2 - p_0] - u(w^L)[(1 - \theta) p_1 + \theta p_2 - p_0] \Leftrightarrow$

$c \geq [u(w^H) - u(w^L)][(1 - \theta) p_1 + \theta p_2 - p_0] \Leftrightarrow$

$c \geq \dfrac{c}{(p_2 - p_1) + \kappa(p_1 - p_0)}[(1 - \kappa) p_1 + \kappa p_2 - p_0] \Leftrightarrow$

$(p_2 - p_1) + \theta(p_1 - p_0) \geq (p_1 - p_0) + \theta(p_2 - p_1) \Leftrightarrow$

$(p_2 - p_1) - (p_1 - p_0) \geq \theta[(p_2 - p_1) - (p_1 - p_0)] \Leftrightarrow$

$(1 - \theta)[(p_2 - p_1) - (p_1 - p_0)] \geq 0,$

while for altruistic preferences:

$U(0, 0, \mathbf{w}_{\mathbf{Alt}}^*; \theta) \geq U(1, 0, \mathbf{w}_{\mathbf{Alt}}^*; \theta) \Leftrightarrow$

$(1 + \theta)[p_0 u(w_{Alt}^H) + (1 - p_0) u(w_{Alt}^L)] \geq (1 + \theta)[p_1 u(w_{Alt}^H) + (1 - p_1) u(w_{Alt}^L)] - c \Leftrightarrow$

$c \geq (1 + \theta)(p_1 - p_0)[u(w_{Alt}^H) - u(w_{Alt}^L)] = (1 + \theta)(p_1 - p_0)\dfrac{c}{(1 + \theta)(p_2 - p_1)} \Leftrightarrow$

$p_2 - p_1 \geq p_1 - p_0.$

Observe that the inequality for moral agents is always satisfied if $\theta = 1$. On the other hand, if $\kappa \in [0, 1)$, that inequality holds iff $p_2 - p_1 \geq p_1 - p_0$. Therefore, the following result holds.

Lemma A1. *Suppose Assumption 1 holds and that the principal offers the optimal contracts $\mathbf{w}_{HM}$ and $\mathbf{w}_{Alt}$ for the teams of moral and altruistic agents, respectively. Then, $e_A = e_B = 1$ is the unique symmetric Nash*

---

[19] Given the symmetry of the problem, I focus attention on agent A and drop the subscripts. The same results would hold for agent B by simply reversing the effort choices e_A and e_B.

equilibrium of the simultaneous choice of effort game played by the agents iff $p_2 - p_1 < p_1 - p_0$ and $\theta < 1$ for homo moralis agents. Otherwise, $e_A = e_B = 0$ is also a symmetric Nash equilibrium.

One remark about asymmetric equilibria must be made here. If $p_2 - p_1 = p_1 - p_0$, both an altruistic and a moral agent will be indifferent between shirking and exerting effort when their partners are shirking. Moreover, since the optimal contract satisfies the incentive compatibility constraint with equality for both types of pro-social preferences, the workers are also indifferent between shirking or not when their partner is exerting the high effort. Therefore, in this case, the asymmetric efforts $(e_A = 1, e_B = 0)$ and $(e_A = 0, e_B = 1)$ are also pure strategy Nash equilibria of the simultaneous choice of effort game.

Appendix B. Obtaining the Borch Rule for Asymmetric Efforts under Homo Moralis Preferences

Relax Assumption 1, and consider the contracting problem of a team of moral agents when efforts are observable. If the principal wishes to induce asymmetric choices of effort, constant wages are not optimal for moral agents, as they were for altruistic and selfish agents[20]. Note first that the Borch rule for teams of selfish and altruistic agents is derived for an arbitrary pair (e_A, e_B), and the ratio of marginal utilities with high or low wages is equal to one whether $e_A = e_B$ or not. For homo moralis preferences, suppose, without loss of generality, that agent A exerts high effort, while agent B exerts low effort. In this case, the principal solves:

$$\max_{\mathbf{w}_A, \mathbf{w}_B} \quad p_1 \left(x_H - w_A^H - w_B^H \right) + (1 - p_1) \left(x_L - w_A^L - w_B^L \right)$$

$$s.t. \quad (1 - \kappa_A) \left[p_1 u(w_A^H) + (1 - p_1)u(w_A^L) \right]$$

$$+ \kappa_A \left[p_2 u(w_A^H) + (1 - p_2)u(w_A^L) \right] - c \geq \overline{u} \quad (IR_A)$$

$$(1 - \kappa_B) \left[p_1 u(w_B^H) + (1 - p_1)u(w_B^L) \right]$$

$$+ \kappa_B \left[p_0 u(w_B^H) + (1 - p_0)u(w_B^L) \right] \geq \overline{u} \quad (IR_B)$$

Close observation of the constraints reveals two differences between them. First, only (IR_A) contains the cost of effort, since agent A is the only one to exert high effort. Second, and more important, the probabilities of high and low realizations of revenues in the Kantian morality terms of the two constraints are different, but the same in the other term. This is true because each agent evaluates the consequence of his/her own effort should both agents choose this particular effort.

The wages must satisfy the Borch rule, given by:

$$\frac{p_1}{u'(w_A^H) \left[(1 - \kappa_A)p_1 + \kappa_A p_2 \right]} = \frac{1 - p_1}{u'(w_A^L) \left[(1 - \kappa_A)(1 - p_1) + \kappa_A(1 - p_2) \right]},$$

$$\frac{p_1}{u'(w_B^H) \left[(1 - \kappa_B)p_1 + \kappa_B p_0 \right]} = \frac{1 - p_1}{u'(w_B^L) \left[(1 - \kappa_B)(1 - p_1) + \kappa_B(1 - p_0) \right]}.$$

---

[20] As shown by the Borch rules (BR_S), (BR_{Alt}).

Observe that the usual finding that $w_i^H = w_i^L = w_i^{FB}$ is only obtained if $\kappa_i = 0$, that is only if both agents display the standard selfish preferences. If the degree of morality is not zero, the marginal utility ratios must be such that:

$$\frac{u'(w_A^H)}{u'(w_A^L)} = \frac{(1-\kappa_A)p_1(1-p_1) - \kappa_A p_1 p_2 + \kappa_A p_1}{(1-\kappa_A)p_1(1-p_1) - \kappa_A p_1 p_2 + \kappa_A p_2} < 1,$$

$$\frac{u'(w_B^H)}{u'(w_B^L)} = \frac{(1-\kappa_B)p_1(1-p_1) - \kappa_B p_1 p_0 + \kappa_B p_1}{(1-\kappa_B)p_1(1-p_1) - \kappa_B p_1 p_0 + \kappa_B p_0} > 1,$$

which implies that the optimal contract satisfies $w_A^H > w_A^L$ and $w_B^H < w_B^L$.

Therefore, should the principal want to induce the moral agents to undertake different efforts, two differences arise in comparison to the selfish and altruistic preferences cases. First is the general argument that the principal should pay a constant wage (that satisfies the participation constraint) in case the appropriate level of effort is exerted by the agent no longer holds. Indeed, for the agent exerting high effort, a monotonicity result similar to the one obtained in the second-best cases of the traditional moral hazard problems is observed. On the other hand, agent B, who is not supposed to exert effort, is paid according to a reverse monotonicity result: that wage after a good realization of revenue must be lower than its counterpart after a bad realization. These results are summarized in the next proposition.

Proposition A1. *Suppose the principal restricts attention to asymmetric equilibria of the kind $e_i = 1 > e_j = 0$ for $i, j \in \{A, B\}$, $i \neq j$ when the agents exhibit homo moralis preferences. Then, there does not exist a constant contract ($w_i^H = w_i^L$ and $w_j^H = w_j^L$) that maximizes the principal's profits and satisfies the agents participation constraints.*

Appendix C. Proofs

Proof of Proposition 1. Assume $e_i = e_j = e \in \{0,1\}$ for $i \in \{A, B\}$. The discussion in the main text shows that the optimal contract under contractible efforts for teams of selfish or moral agents is:

$$w^{FB} = u^{-1}\left(\overline{u} + c(e)\right),$$

while altruistic agents must be compensated according to:

$$w_{i,Alt}^{FB} = u^{-1}\left(\frac{1}{1+\theta}\overline{u} + c(e)\right).$$

Denote by V_{11}^{FB} and V_{00}^{FB} the principal's expected profits in the cases where both agents exert high and low effort, respectively, and the teams are comprised of either selfish or moral agents. Plugging in the optimal wages obtained above yields:

$$V_{11}^{FB} = p_2 x^H + (1 - p_2)x^L - 2u^{-1}(\overline{u} + c)$$
$$V_{00}^{FB} = p_0 x^H + (1 - p_0)x^L - 2u^{-1}(\overline{u})$$

and thus $V_{11}^{FB} \geq V_{00}^{FB}$ if and only if:

$$u^{-1}(\overline{u} + c) - u^{-1}(\overline{u}) \leq \frac{(x^H - x^L)(p_2 - p_0)}{2}.$$

By assumption, $p_2 > p_0$ and $x^H > x^L$, so the right-hand side is strictly positive, while $u' > 0$ implies that the left-hand side is also positive since $c > 0$. By continuity of u, there exists $c' > 0$ such that $2\left(u^{-1}(\overline{u} + c') - u^{-1}(\overline{u})\right) = (x^H - x^L)(p_2 - p_0)$, and the inequality above holds for all $c \in (0, c']$.

Now, doing the same for altruistic agents, write:

$$V_{11,Alt}^{FB} = p_2 x^H + (1 - p_2) x^L - 2u^{-1}\left(\frac{1}{1+\theta}\bar{u} + c\right)$$

$$V_{00,Alt}^{FB} = p_0 x^H + (1 - p_0) x^L - 2u^{-1}\left(\frac{1}{1+\theta}\bar{u}\right)$$

where an argument similar to the paragraph above implies that there exists $c'' > 0$ such that, $\forall c \in [0, c'')$, $V_{11,Alt}^{FB} > V_{00,Alt}^{FB}$.

Letting $c^* = \min\{c', c''\}$ concludes the proof. □

Proof of Proposition 2. Since u is a strictly increasing strictly concave function by assumption, its inverse u^{-1} is, in turn, a strictly increasing strictly convex function. Therefore, since:

$$\bar{u} \geq \bar{u}\frac{1}{1+\theta}$$

For all $\theta \in [0, 1]$, the amount of compensation dispensed by the principal is larger under homo moralis or selfish preferences. □

Proof of Proposition 3. Under Assumption 1 and $e_A = e_B = 1$, the optimal symmetric contract $\mathbf{w_{Alt}}^* = (w_{Alt}^H, w_{Alt}^L)$ offered by the principal must solve:

$$\max_{w^H, w^L} \quad p_2(x^H - 2w^H) + (1 - p_2)(x^L - 2w^L)$$

$$s.t. \quad (1+\theta)[p_2 u(w^H) + (1 - p_2)u(w^L) - c] \geq \bar{u} \quad (IR)$$

$$u(w^H) - u(w^L) \geq \frac{c}{(1+\theta)(p_2 - p_1)} \quad (IC)$$

where the KKT conditions are necessary and sufficient for optimality and given by:

$$u'(w_{Alt}^H)(1+\theta)[\lambda p_2 + \mu(p_2 - p_1)] = p_2 \quad (A2a)$$

$$u'(w_{Alt}^L)(1+\theta)[\lambda(1 - p_2) - \mu(p_2 - p_1)] = (1 - p_2) \quad (A2b)$$

$$\lambda\left\{(1+\theta)[p_2 u(w_{Alt}^H) + (1 - p_2)u(w_{Alt}^L) - c] - \bar{u}\right\} = 0 \quad (A2c)$$

$$\mu\left\{(1+\theta)(p_2 - p_1)[u(w_{Alt}^H) - u(w_{Alt}^L)] - c\right\} = 0 \quad (A2d)$$

$$(1+\theta)[p_2 u(w_{Alt}^H) + (1 - p_2)u(w_{Alt}^L) - c] - \bar{u} \geq 0 \quad (A2e)$$

$$(1+\theta)(p_2 - p_1)[u(w_{Alt}^H) - u(w_{Alt}^L)] - c \geq 0 \quad (A2f)$$

$$\lambda \geq 0 \quad (A2g)$$

$$\mu \geq 0 \quad (A2h)$$

Note that $\lambda = 0$ cannot be a solution since it violates Equation (A2b), because $u' > 0$ and $1 > p_2 > p_1 > p_0 > 0$ by assumption. Moreover, $\mu > 0$; otherwise, Equations (A2a) and (A2b) would imply:

$$u'(w_{Alt}^H) = \frac{1}{(1+\theta)\lambda} = u'(w_{Alt}^L),$$

which yields $w_{Alt}^L = w_{Alt}^H$ for all $\theta \in [0, 1]$ since $u'' < 0$, thus violating the incentive compatibility constraint in (A2b). Therefore, any solution must have $\lambda, \mu > 0$ such that:

$$\begin{cases} \lambda p_2 + \mu(p_2 - p_1) > 0 \\ \lambda(1 - p_2) - \mu(p_2 - p_1) > 0 \end{cases}$$

Since the Lagrange multipliers are strictly positive, the optimal contract is fully characterized by the bindings *IC* and *IR*, which rearranged result in:

$$u(w_{Alt}^H) = \frac{\bar{u}}{1+\theta} + c\frac{[(1-p_1) + \theta(p_2 - p_1)]}{(1+\theta)(p_2 - p_1)},$$

$$u(w_{Alt}^L) = \frac{\bar{u}}{1+\theta} - c\frac{[p_1 - \theta(p_2 - p_1)]}{(1+\theta)(p_2 - p_1)}.$$

Differentiation of the incentive compatibility constraint with respect to θ leads to:

$$\frac{\partial(w_{Alt}^H - w_{Alt}^L)}{\partial\theta} < 0.$$

Given the optimal contract, one must again wonder whether the principal will induce both agents to exert high effort or not. If not, then the principal can offer the constant wage $w = u^{-1}\left(\frac{\bar{u}}{1+\theta}\right)$ as before, since this satisfies the participation constraint, but not the incentive compatibility constraint. Thus, the principal's expected payoff in this case is again given by:

$$V_{00}(\theta) = p_0 x^H + (1 - p_0)x^L - 2u^{-1}\left(\frac{\bar{u}}{1+\theta}\right),$$

while inducing high effort yields expected profits:

$$V_{11}(\theta) = p_2\left[x^H - 2w_{Alt}^H\right] + (1 - p_2)\left[x^L - 2w_{Alt}^L\right].$$

Consequently, it is only beneficial to the principal demanding high effort from both agents if $V_{11}(\theta) \geq V_{00}(\theta)$, that is,

$$(p_2 - p_0)(x^H - x^L) + 2u^{-1}\left(\frac{\bar{u}}{1+\theta}\right) \geq 2\left[p_2 u^{-1}\left(\frac{\bar{u}}{1+\theta} + \frac{c[(1-p_1) + \theta(p_2 - p_1)]}{(1+\theta)(p_2 - p_1)}\right)\right.$$
$$\left. + (1 - p_2)u^{-1}\left(\frac{\bar{u}}{1+\theta} - c\frac{[p_1 - \theta(p_2 - p_1)]}{(1+\theta)(p_2 - p_1)}\right)\right].$$

Take $c = 0$. The right-hand side reduces to $2[p_2 u^{-1}\left(\frac{\bar{u}}{1+\theta}\right) + (1 - p_2)u^{-1}\left(\frac{\bar{u}}{1+\theta}\right)] = 2u^{-1}\left(\frac{\bar{u}}{1+\theta}\right)$, and the inequality is automatically satisfied, since $1 > p_2 > p_1 > p_0 > 0$ and $x^H > x^L$ by assumption.

Therefore, by continuity, $\exists \bar{c}_{Alt} > 0$ such that $\forall c \in (0, \bar{c}_{Alt}]$, $V_{11}(\theta) \geq V_{00}(\theta)$. □

Proof of Corollary 1. Take $\theta_0, \theta_1 \in [0, 1]$ such that $\theta_0 < \theta_1$, and let $\mathbf{w}_{Alt}^*(\theta) = (w_{Alt}^H(\theta), w_{Alt}^L(\theta))$ denote the optimal wage offered by the principal when agents display the degree of morality $\theta \in [0, 1]$. Moreover, for any $\theta \in [0, 1]$, let $\mathcal{C}(\theta)$ denote the set of contracts satisfying both the *IR* and *IC* constraints for the degree of altruism θ, so that $\mathbf{w}_{Alt}^*(\theta) \in \mathcal{C}(\theta)$.

Then, using the constraints, one can check that:

$$u(w_{Alt}^H(\theta_0)) - u(w_{Alt}^L(\theta_0)) = \frac{c}{(1+\theta_0)(p_2 - p_1)} > \frac{c}{(1+\theta_1)(p_2 - p_1)}$$

and:

$$\bar{u} = (1 + \theta_0)[p_2 u(w_{Alt}^H(\theta_0)) + (1 - p_2)u(w_{Alt}^L(\theta_0)) - c] < (1 + \theta_1)[p_2 u(w_{Alt}^H(\theta_0)) + (1 - p_2)u(w_{Alt}^L(\theta_0)) - c],$$

so that $\mathbf{w}_{Alt}^*(\theta_0) \in \mathcal{C}(\theta_1)$. However, the KKT conditions imply that $\mathbf{w}_{Alt}^*(\theta_1)$ is the unique solution to the principal's problem for $\theta = \theta_1$. Then, it must be the case that $V_{11}^{Alt}(\mathbf{w}_{Alt}^*(\theta_1); \theta_1) > V_{11}^{Alt}(\mathbf{w}_{Alt}^*(\theta_0); \theta_1)$.

Now, observe that:

$$\frac{dw_{Alt}^{H}(\theta)}{d\theta} = -\frac{1}{u'(w_{Alt}^{H}(\theta_0))}\left[\frac{(1-p_2)c}{(1+\theta)^2(p_2-p_1)} + \frac{\bar{u}}{(1+\theta)^2}\right] < 0.$$

Thus, keeping $w_{Alt}^{L}(\theta_1) = w_{Alt}^{L}(\theta_0)$ and taking $w_{Alt}^{H}(\theta_1) = w_{Alt}^{H}(\theta_0) - \varepsilon, \varepsilon \approx 0$, the principal satisfies both constraints while increasing his/her payoff by $2p_2\varepsilon > 0$. Therefore, the principal's expected profit is strictly increasing in θ. $\square$

Proof of Proposition 4. The principal's problem is given by:

$$\begin{aligned}
\mathbb{L} = \ & p_2\left[x_H - w_i^H - w_j^H\right] + (1-p_2)\left[x_L - w_i^L - w_j^L\right] \\
& + \sum_{i=1}^{2}\lambda_i\left[p_2u(w_i^H) + (1-p_2)u(w_i^L) - c - \bar{u}\right] \\
& + \sum_{i=1}^{2}\mu_i\left[\left(u(w_i^H) - u(w_i^L)\right)\left((p_2-p_1) + \kappa_i(p_1-p_0)\right) - c\right]
\end{aligned}$$

for $i = A, B$ and $j \neq i$. Then, the KKT conditions are given by the system of equations:

$$-p_2 + \lambda_i p_2 u'(w_i^H) + \mu_i u'(w_i^H)\left((p_2-p_1) + \kappa_i(p_1-p_0)\right) = 0 \tag{A3a}$$

$$-(1-p_2) + \lambda_i(1-p_2)u'(w_i^L) - \mu_i u'(w_i^L)\left((p_2-p_1) + \kappa_i(p_1-p_0)\right) = 0 \tag{A3b}$$

$$p_2 u(w_i^H) + (1-p_2)u(w_i^L) - c \geq \bar{u} \tag{A3c}$$

$$\left(u(w_i^H) - u(w_i^L)\right)\left((p_2-p_1) + \kappa_i(p_1-p_0)\right) - c \geq 0 \tag{A3d}$$

$$\lambda_i\left[p_2 u(w_i^H) + (1-p_2)u(w_i^L) - c - \bar{u}\right] = 0 \tag{A3e}$$

$$\mu_i\left[\left(u(w_i^H) - u(w_i^L)\right)\left((p_2-p_1) + \kappa_i(p_1-p_0)\right) - c\right] = 0 \tag{A3f}$$

$$\lambda_i \geq 0 \tag{A3g}$$

$$\mu_i \geq 0 \tag{A3h}$$

Equations (A3a) and (A3b) clearly show that $\lambda_i = \mu_i = 0$ is not a possibility. Indeed, if that were the case, then $p_2 = 0$, which contradicts our initial assumption. Furthermore, I cannot have $\mu_i > 0 = \lambda_i$, because this would imply that Equation (A3b) is not satisfied. Therefore, I must either have $\lambda_i > 0 = \mu_i$ or $\lambda_i, \mu_i > 0$. Solving for the multipliers in Equations (A3a) and (A3b) yields:

$$\lambda_i = \frac{(1-p_2)u'(w_i^H) + p_2 u'(u_i^L)}{u'(w_i^H)u'(w_i^L)} > 0$$

$$\mu_i = \frac{p_2(1-p_2)(u'(w_i^L) - u'(w_i^H))}{u'(w_i^H)u'(w_i^L)((p_2-p_1) + \kappa_i(p_1-p_0))} > 0$$

so both the (IC_i) and (IR_i) constraints bind. Thus, using Equations (A3c) and (A3d), one finds that the optimal schedule must satisfy:

$$u(w_i^L) = \bar{u} - \frac{c[(1-\kappa_i)p_1 + \kappa_i p_0]}{(p_2-p_1) + \kappa_i(p_1-p_0)}$$

$$u(w_i^H) = \bar{u} + \frac{c[(1-p_1) + \kappa_i(p_1-p_0)]}{(p_2-p_1) + \kappa_i(p_1-p_0)}.$$

Differentiating this expressions with respect to κ_i yields:

$$\frac{dw_i^H}{d\kappa_i} = -\frac{(1-p_2)(p_1-p_0)}{(p_2-p_1)+\kappa_i(p_1-p_0)}\frac{c}{u'(w_i^H)} < 0$$

$$\frac{dw_i^L}{d\kappa_i} = \frac{p_2(p_1-p_0)}{(p_2-p_1)+\kappa_i(p_1-p_0)}\frac{c}{u'(w_i^L)} > 0.$$

Given the optimal contract, one must again wonder whether the principal will induce both agents to exert high effort or not. If not, then the principal can offer the constant wage $w = u^{-1}(\overline{u})$ as before, since this satisfies the participation constraint, but not the incentive compatibility constraint. Thus, the principal's expected payoff in this case is again given by:

$$V_{00}(\check{}) = p_0 x^H + (1-p_0)x^L - 2u^{-1}(\overline{u}),$$

while inducing high effort yields expected profits:

$$V_{11}(\kappa) = p_2\left[x^H - \sum_{i=A,B} u^{-1}\left(\overline{u} + \frac{c[(1-p_1)+\kappa_i(p_1-p_0)]}{(p_2-p_1)+\kappa_i(p_1-p_0)}\right)\right]$$
$$+ (1-p_2)\left[x^L - \sum_{i=A,B} u^{-1}\left(\overline{u} - \frac{c[(1-\kappa_i)p_1+\kappa_i p_0]}{(p_2-p_1)+\kappa_i(p_1-p_0)}\right)\right].$$

Consequently, it is only beneficial to the principal demanding high effort from both agents if $V_{11}(\kappa) \geq V_{00}(\kappa)$, that is,

$$(p_2-p_0)(x^H - x^L) + 2u^{-1}(\overline{u}) \geq \sum_{i=A,B}\left[p_2 u^{-1}\left(\overline{u} + \frac{c[(1-p_1)+\kappa_i(p_1-p_0)]}{(p_2-p_1)+\kappa_i(p_1-p_0)}\right)\right.$$
$$\left.+ (1-p_2)u^{-1}\left(\overline{u} - \frac{c[(1-\kappa_i)p_1+\kappa_i p_0]}{(p_2-p_1)+\kappa_i(p_1-p_0)}\right)\right].$$

Take $c = 0$. The the right-hand side reduces to $\sum_{i=A,B}[p_2 u^{-1}(\overline{u}) + (1-p_2)u^{-1}(\overline{u})] = 2u^{-1}(\overline{u})$, and the inequality is automatically satisfied, since $1 > p_2 > p_1 > p_0 > 0$ and $x^H > x^L$ by assumption.

Therefore, by continuity, $\exists \overline{c}_{HM} > 0$ such that $\forall c \in (0, \overline{c}_{HM}]$, $V_{11}(\kappa) \geq V_{00}(\kappa)$. □

Proof of Corollary 2. Denote again the principal's indirect expected profits by $V_{11}(\kappa_A, \kappa_B)$. Then,

$$\frac{\partial V_{11}(\kappa_A, \kappa_B)}{\partial \kappa_i} = -p_2 \frac{dw_i^H}{d\kappa_i} - (1-p_2)\frac{dw_i^L}{d\kappa_i}$$

$$= \frac{c(p_1-p_0)}{[(p_2-p_1)+\kappa_i(p_1-p_0)]^2}\left[\frac{p_2(1-p_2)}{u'(w_i^H)} - \frac{(1-p_2)p_2}{u'(w_i^L)}\right]$$

$$= \underbrace{\frac{c(p_1-p_0)(1-p_2)p_2}{[(p_2-p_1)+\kappa_i(p_1-p_0)]^2}}_{>0} \times \frac{u'(w_i^L) - u'(w_i^H)}{u'(w_i^L)u'(w_i^H)} > 0$$

since $w_i^H > w_i^L$ by the monotonicity implied by the (IC_i) and $u'' < 0$. Therefore, the principal's expected payoff is strictly increasing in each degree of morality κ_i.

In a similar fashion, let $U(\kappa_i)$ denote the agent's indirect utility under the optimal contract. Then,

$$\frac{\partial U(\kappa_i)}{\partial \kappa_i} = p_2 u'(w_i^H)\frac{dw_i^H}{d\kappa_i} + (1 - p_2)u'(w_i^L)\frac{dw_i^L}{d\kappa_i}$$

$$= \frac{c(p_1 - p_0)}{[(p_2 - p_1) + \kappa_i(p_1 - p_0)]^2}\left[-p_2(1 - p_2) + (1 - p_2)p_2\right] = 0.$$

□

Proof of Lemma 1. Let $\kappa = \alpha = \theta \in [0, 1]$, $c > 0$ and $1 > p_2 > p_1 > p_0 > 0$ by assumption. Then:

$$\frac{c}{(p_2 - p_1) + \theta(p_1 - p_0)} \geq \frac{c}{(1 + \theta)(p_2 - p_1)} \Leftrightarrow$$
$$(p_2 - p_1) + \theta(p_2 - p_1) \geq (p_2 - p_1) + \theta(p_1 - p_0) \Leftrightarrow$$
$$p_2 - p_1 \geq p_1 - p_0.$$

□

Proof of Theorem 1. Let $h^H = u(w^H)$ and $h^L = u(w^L)$, which are uniquely determined for any values of w since u is strictly increasing by assumption. The principal's problem can thus be rewritten as:

$$\max_{h^L, h^H} \quad p_2\left(x^H - 2u^{-1}(h^H)\right) + (1 - p_2)\left(x^L - 2u^{-1}(h^L)\right)$$

subject to

$$(h^L, h^H) \in \mathcal{C}_{HM}(\theta) = \left\{(h_1, h_2) \in \mathcal{R}^2 : p_2 h_2 + (1 - p_2)h_1 - c \geq \overline{u}, h_2 - h_1 \geq \frac{c}{k_2 + \theta k_1}\right\}$$

if the principal is hiring a team of moral agents, and:

$$(h^L, h^H) \in \mathcal{C}_{Alt}(\theta) = \left\{(h_1, h_2) \in \mathcal{R}^2 : p_2 h_2 + (1 - p_2)h_1 - c \geq \frac{\overline{u}}{1 + \theta}, h_2 - h_1 \geq \frac{c}{(1 + \theta k_2)}\right\},$$

where $k_1 = p_1 - p_0 > 0$ and $k_2 = p_2 - p_1 > 0$, if he/she considers a team of altruistic agents. The sets $\mathcal{C}_{HM}(\theta)$ and $\mathcal{C}_{Alt}(\theta)$ collect all of the values of h^L and h^H satisfying the participation and incentive compatibility constraints for a given degree of morality or altruism $\theta \in [0, 1]$.

First, notice that for any value of $\theta \in [0, 1]$, a pair $(h_L, h_L) \in \mathcal{C}_{HM}(\theta)$ also satisfies the (IR) constraint in $\mathcal{C}_{Alt}(\theta)$: indeed, $p_2 h_2 + (1 - p_2)h_1 - c \geq \overline{u} \geq \frac{\overline{u}}{1+\theta}$ for $\overline{u} \geq 0$.

Suppose that $p_2 - p_1 \geq p_1 - p_0$, i.e., $k_2 \geq k_1$. Then, by Lemma 1, $h_{HM}^H - h_{HM}^L \geq h_{Alt}^H - h_{Alt}^L$, which implies that the optimal contract under homo moralis preferences also satisfies the incentive compatibility of altruistic agents, so that $(h_{HM}^L(\theta), h_{HM}^H(\theta)) \in \mathcal{C}_{Alt}(\theta)$. This implies that $\mathcal{C}_{HM}(\theta) \subset \mathcal{C}_{Alt}(\theta)$, and one can conclude that $V_{11}^{Alt}(\theta) \geq V_{11}^{HM}(\theta)$. This can be graphically seen in Figure A1.

Suppose now that $p_2 - p_1 < p_1 - p_0$, i.e., $k_2 < k_1$. In this case, the incentive compatibility constraint for moral agents is below the one for altruistic agents, as can be seen in Figure A2 and implied by Lemma 1, but because the reverse holds for the participation constraint, one cannot say that $\mathcal{C}_{Alt}(\theta) \subset \mathcal{C}_{HM}(\theta)$. Using the results in Propositions 2 and 3, one can check that:

$$h_{HM}^L - h_{Alt}^L = \overline{u} - \frac{\overline{u}}{1 + \theta} + c - c + p_2 c\left(\frac{1}{(1 + \theta)k_2} - \frac{1}{k_2 + \theta k_1}\right)$$

$$= \frac{\theta \overline{u}}{1 + \theta} + \frac{p_2 c\theta(k_1 - k_2)}{(1 + \theta)k_2(k_2 + \theta k_1)} \geq 0$$

for all $\theta \in [0,1]$, $\bar{u} \geq 0$, $c > 0$ and $0 < p_0 < p_1 < p_2 < 1$. Thus, the wage paid after a bad realization of output for a moral agent is larger than the corresponding wage paid to an altruistic agent if $k_1 > k_2$. Therefore, the principal can only be better off hiring a team of moral agents if the wage paid after a good realization of output to the latter is sufficiently smaller than the one paid for altruistic agents and the isoprofit curve is sufficiently flat. The former holds only if:

$$h_{HM}^H - h_{Alt}^H = h_{HM}^L - h_{Alt}^L + c \left(\frac{1}{k_2 + \theta k_1} - \frac{1}{(1+\theta)k_2} \right)$$
$$= \frac{\theta \bar{u}}{1 + \theta} - \frac{(1-p_2)c\theta(k_1 - k_2)}{(1+\theta)k_2(k_2 + \theta k_1)} < 0.$$

For $\theta \in (0,1]$, $c > 0$ and $0 < p_0 < p_1 < p_2 < 1$, the inequality above holds iff:

$$\bar{u}(k_2 + \theta k_1) < \frac{1 - p_2}{p_2 - p_1}(k_1 - k_2),$$

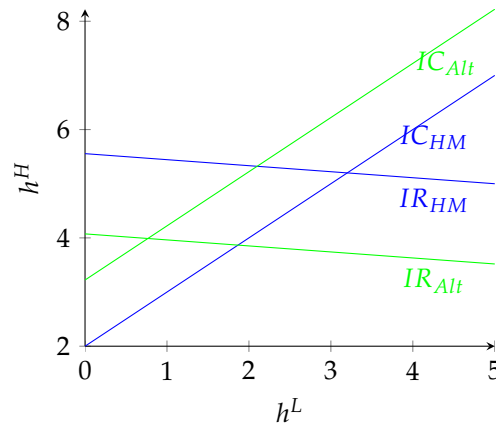that is, for $\bar{u} = 0$ or small values of θ for $\bar{u} > 0$.

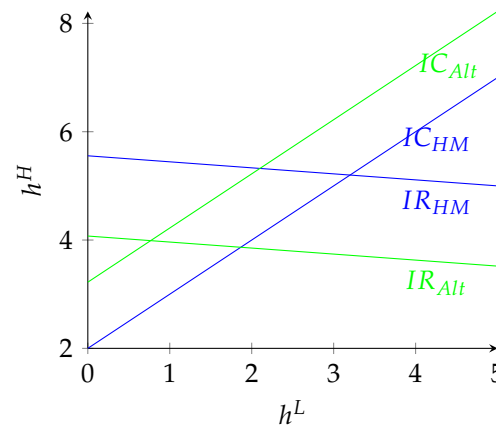**Figure A1.** Optimal contracts for $p_2 - p_1 > p_1 - p_0$.

Figure A2. Optimal contracts for $p_2 - p_1 < p_1 - p_0$.

Now, remember that the slope of the isoprofit curve for the principal is given by $\frac{dh^H}{dh^L} = -\frac{1-p_2}{p_2} \frac{u'(w^H)}{u'(w^L)} < 0$, which becomes flatter as p_2 approaches one. Thus, if $k_1 > k_2$, the principal is better off with a team of moral agents if p_2 is close to one and either $\bar{u} = 0$ or $\bar{u} > 0$ and $\theta \to 0$. $\square$

Proof of Proposition 5. The existence of the contract follows from the KKT conditions written in the main text. The same goes for the inequalities on the wage schedules. $\square$

References

1.　O'Leonard, K.; Erickson, R.; Krider, J. *Talent Acquisition Factbook 2015: Benchmarks and Trends in Spending, Staffing, and Key Recruiting Metrics*; Bersin: Oakland, CA, USA, 2015.

2.　Brown, V.R.; Vaughn, E.D. The Writing on the (Facebook) Wall: The Use of Social Networking Sites in Hiring Decisions. *J. Bus. Psychol.* **2011**, *26*, 219.

3.　Fehr, E.; Schmidt, K.M. The Economics of Fairness, Reciprocity and Altruism—Experimental Evidence and New Theories. In *Handbook on the Economics of Giving, Reciprocity and Altruism*; Kolm, S., Ythier, J.M., Eds.; Elsevier: Amsterdam, The Netherlands, 2006; Volume 1, Chapter 8, pp. 615–691.

4.　Kolm, S.; Ythier, J.M. (Eds.) *Handbook of the Economics of Giving, Altruism and Reciprocity*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2006; Volume 1.

5.　Kolm, S.; Ythier, J.M. (Eds.) *Handbook of the Economics of Giving, Altruism and Reciprocity*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2006; Volume 2.

6.　Kahneman, D.; Knetsch, J.L.; Thaler, R.H. Fairness and the Assumptions of Economics. *J. Bus.* **1986**, *59*, S285–S300.

7.　Kahneman, D.; Knetsch, J.L.; Thaler, R.H. Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *J. Econ. Perspect.* **1991**, *5*, 193–206.

8.　Dufwenberg, M.; Heidhues, P.; Kirchsteiger, G.; Riedel, F.; Sobel, J. Other-Regarding Preferences in General Equilibrium. *Rev. Econ. Stud.* **2011**, *78*, 613–639.

9.　Bowles, S.; Polania-Reyes, S. Economic Incentives and Social Preferences: Substitutes or Complements? *J. Econ. Lit.* **2012**, *50*, 368–425.

10.　Bandiera, O.; Barankay, I.; Rasul, I. Social Preferences and the Response to Incentives: Evidence from Personnel Data. *Q. J. Econ.* **2005**, *120*, 917–962.

11.　Bandiera, O.; Barankay, I.; Rasul, I. Social Incentives in the Workplace. *Rev. Econ. Stud.* **2010**, *77*, 417–458.

12.　Barr, A.; Serneels, P. Reciprocity in the Workplace. *Exp. Econ.* **2009**, *12*, 99–112.

13.　Rotemberg, J.J. Altruism, Reciprocity and Cooperation in the Workplace. In *Handbook on the Economics of Giving, Reciprocity and Altruism*; Kolm, S., Ythier, J.M., Eds.; Elsevier: Amsterdam, The Netherlands, 2006; Volume 1, Chapter 21, pp. 1371–1407.

14.　Thaler, R.H. Anomalies: The Winner's Curse. *J. Econ. Perspect.* **1988**, *2*, 191–202.

15.　Tversky, A.; Thaler, R.H. Anomalies: Preference Reversals. *J. Econ. Perspect.* **1990**, *4*, 201–211.

16.　Alchian, A.A.; Demsetz, H. Production, Information Costs, and Economic Organization. *Am. Econ. Rev.* **1972**, *62*, 777–795.

17.　Holmström, B. Moral Hazard in Teams. *Bell J. Econ.* **1982**, *13*, 324–340.

18.　Becker, G.S. A Theory of Social Interactions. *J. Polit. Econ.* **1974**, *82*, 1063–1093.

19.　Wilson, D.S.; Kniffin, K. Altruism from an Evolutionary Perspective. In *Research on Altruism and Love: An Annotated Bibliography of Major Studies in Socionlogy, Evolutionary Biology, and Theology*; Post, S., Johnson, B., McCullough, M., Schloss, J., Eds.; Templeton Foundation Press: West Conshohocken, PA, USA, 2003; Volume 1, Chapter 3, pp. 117–136.

20.　Alger, I.; Weibull, J.W. Homo Moralis—Preference Evolution Under Incomplete Information and Assortative Matching. *Econometrica* **2013**, *81*, 2269–2302.

21.　Alger, I.; Weibull, J.W. Evolution and Kantian Morality. *Games Econ. Behav.* **2016**, *98*, 56–67.

22.　Bergström, T.C. On the Evolution of Altruistic Ethical Rules for Siblings. *Am. Econ. Rev.* **1995**, *85*, 58–81.

23.　Mookherjee, D. Optimal Incentive Schemes with Many Agents. *Rev. Econ. Stud.* **1984**, *51*, 433–446.

24.　Che, Y.K.; Yoo, S.W. Optimal Incentives for Teams. *Am. Econ. Rev.* **2001**, *91*, 525–541.

25.　Itoh, H. Moral Hazard and Other-Regarding Preferences. *Jpn. Econ. Rev.* **2004**, *55*, 18–45.

26.　Englmaier, F.; Wambach, A. Optimal incentive contracts under inequity aversion. *Games Econ. Behav.* **2010**, *69*, 312–328.

27.　Rey-Biel, P. Inequity Aversion and Team Incentives. *Scand. J. Econ.* **2008**, *110*, 297–320.

28.　Livio, L. Friend or Foes? Optimal Incentives for Reciprocal Agents. 2016. Available online: https://lucalivio85.files.wordpress.com/2016/05/reciprocity-november-2015-mgmm.pdf (accessed on 4 September 2017).

29.　Dur, R.; Sol, J. Social Interaction, Co-worker Altruism, and Incentives. *Games Econ. Behav.* **2010**, *69*, 293–301.

30.　Barron, J.M.; Gjerde, K.P. Peer Pressure in an Agency Relationship. *J. Labor Econ.* **1997**, *15*, 234–254.

31. Kandel, E.; Lazear, E.P. Peer Pressure and Partnerships. *J. Polit. Econ.* **1992**, *100*, 801–817.

32. Kosfeld, M.; von Siemens, F.A. Worker Self-Selection and the Profits from Cooperation. *J. Eur. Econ. Assoc.* **2009**, *7*, 573–582.

33. Kosfeld, M.; von Siemens, F.A. Competition, cooperation, and corporate culture. *RAND J. Econ.* **2011**, *42*, 23–43.

34. Delfgaauw, J.; Dur, R. Signaling and screening of workers' motivation. *J. Econ. Behav. Organ.* **2007**, *62*, 605–624.