

Sarkisian, Roberto

**Article**

## Screening teams of moral and altruistic agents

Games

**Provided in Cooperation with:**

MDPI – Multidisciplinary Digital Publishing Institute, Basel

*Suggested Citation:* Sarkisian, Roberto (2021) : Screening teams of moral and altruistic agents, Games, ISSN 2073-4336, MDPI, Basel, Vol. 12, Iss. 4, pp. 1-11, <https://doi.org/10.3390/g12040077>

This Version is available at:

<https://hdl.handle.net/10419/257559>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

Article

# Screening Teams of Moral and Altruistic Agents

Roberto Sarkisian

Department of Economics and Finance, University of Rome Tor Vergata, 00133 Rome, Italy;  
roberto.sarkisian@uniroma2.eu; Tel.: +39-347-883-3299

**Abstract:** This paper studies the problem of screening teams of either moral or altruistic agents, in a setting where agents choose whether or not to exert effort in order to achieve a high output for the principal. I show that there exists no separating equilibrium menu of contracts that induces the agents to reveal their types unless the principal either (i) excludes one group from the productive relationship, or (ii) demands different efforts from different preference groups. I also characterize the contract-inducing pooling equilibria in which all agents are incentivized to exert a high level of effort.

**Keywords:** screening; homo moralis; altruism

**JEL Classification:** D82; D86; D03



**Citation:** Sarkisian, R. Screening Teams of Moral and Altruistic Agents. *Games* **2021**, *12*, 77. <https://doi.org/10.3390/g12040077>

Academic Editors: Heinrich H. Nax and Ulrich Berger

Received: 30 August 2021

Accepted: 13 October 2021

Published: 20 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Can an employer distinguish between moral and altruistic employees? Ref. [1] explores a moral hazard in teams problem where an employer has to choose between hiring a team of altruistic agents or a team of moral agents (as in [2–4]). The key finding is that the principal sometimes prefers the team of moral agents over the team of altruistic ones depending on the production technology and the common degree of morality or altruism. The author then argues that firms may have incentives to collect information about their prospective employees' preferences in order to benefit from offering less costly contracts.

This last point, however, is not developed there. In particular, [1] assumes that the agents' preferences are common knowledge, i.e. the principal knows not only which kind of prosocial preferences the prospective employees have, but also what is the common degree of morality or altruism displayed by the agents.

The objective of this paper is to relax that strong assumption: in what follows, it is assumed that the degree of altruism or morality is known to all parties, but the utility function specification is private knowledge of the agents. The principal then seeks to distinguish the two groups by offering menus of contracts that induce participation, effort provision, and revelation of private information by the employees.

This class of adverse selection followed by moral hazard problems has been analyzed before. Ref. [5,6] have dedicated sections to this broad class of problems, and provide more references to the literature. To cite but a few articles exploring screening preferences, Ref. [7] considers the problem of screening risk-averse agents under moral hazard under the strong assumption that the utility function satisfies single-crossing and CARA properties. As a result, they find that the power of incentives is decreasing with respect to risk-aversion. Ref. [8] study a two-output model with risk-neutral agent protected by limited liability and ex-post participation constraints, and find that a fully pooling contract is optimal. Ref. [9] build upon the previous model by assuming that the agent is risk-averse, and also finds that pooling contracts are difficult to avoid.

All the papers cited above differ from the environment studied here in one important way: they assume that preferences are common knowledge, but that either the degree of risk-aversion or a productivity parameter is private information of the single agent. Here,

as stated before, the utility function rather than the common degree of altruism or morality is private information of the agents. The main results, however, are in line with [8,9]: separation is difficult to achieve by the principal if she desires the agents to exert effort in equilibrium. Intuitively, this is a consequence of the utility functions not displaying a single-crossing-like property, an assumption that is imposed in [7].

Screening prosocial preferences has been the central issue in some studies, both theoretically and empirically. Ref. [10] studies an environment with a single principal screening a continuum of workers that have private information about their ability and preferences over social comparisons. In particular, Ref. [10] contrasts the optimal employment contracts for selfish and inequity-averse agents, and finds that it is impossible to screen workers of similar ability with respect to their social preferences within the firm, a result that is line with the ones found here. The main difference between [10] and the model in this study is that the former considers only the adverse selection problem faced by the principal when hiring a single agent, while the latter assumes teamwork and moral hazard. Closer in essence to this paper are the works of [11,12], who consider screening followed by moral hazard when agents' prosocial preferences are characterized by inequity aversion. Their results also suggest that screening agents according to their social preferences is not feasible.

The approach here developed is close in essence to [13]. In the first step, I search for the set of feasible contracts, that is, those that satisfy the participation constraints plus the moral hazard incentive compatibility constraints for each preference group. In the second step, the principal selects the least costly menu of feasible contracts that also induce the preference groups to truthfully reveal their types. As a consequence, I obtain a *no-distortion-at-the-top*-like result: the principal offers the second-best moral hazard contract to the least costly to hire preference group, and distorts the allocation to the costlier one either by excluding it from the relationship or by demanding a different level of effort.

The paper goes as follows. The next section presents the environment and the concept of separating equilibrium to be considered. Section 3 discusses screening and existence of separating equilibria, while Section 4 characterizes contracts that support pooling equilibria. Section 5 concludes. For ease of exposition, all proofs are relegated to Appendix A.

## 2. The Model

This model builds upon the model in [1]. Consider a single risk-neutral principal (she/firm) who faces a continuum of potential employees with total mass normalized to one. Alternatively, the model can be restated by considering  $n$  pairs of potential employees, without loss. The firm seeks to hire a pair of agents to work on a common task that yields output  $x \in \{x^H, x^L\}$  to the principal, with  $x^H > x^L$ . The probability  $p$  of the high outcome being achieved depends on the binary choices of effort made by the agents employed in the firm,  $e_i = 0, 1$  for  $i \in \{A, B\}$ . In particular,

$$Pr(x = x^H | e_A, e_B) = p_{e_A + e_B}, \quad (1)$$

where I assume that  $1 > p_2 \geq p_1 \geq p_0 > 0$ . The cost of exerting effort is identical to every agent,  $C(e) = ce$ , for  $c > 0$ .

Output is contractible, and the principal posts wage schedules  $w_i(x)$  in order to attract the teams of agents. If the firm successfully attracts a pair of employees, her realized profit is

$$V(x, w_A, w_B) = x - w_A(x) - w_B(x). \quad (2)$$

Denoted by  $\pi_i(e_i, e_j, w_i(x))$ , 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. This is in line with [8,9], where the schedules are composed of a

fixed plus a variable part. In what follows, the *material* payoff function takes the expected additively separable form

$$\pi(e_i, e_j, \mathbf{w}_i) = p_{e_i+e_j} [u(w_i^H) - c(e_i)] + (1 - p_{e_i+e_j}) [u(w_i^L) - c(e_i)], \tag{3}$$

where  $u : \mathbb{R}_+ \rightarrow \mathbb{R}$  is the function that associates the agent’s consumption utility to each wage realization  $w$ . The agents are risk averse towards wages:  $u(w)$  is assumed to be twice-continuously differentiable, strictly increasing and strictly concave.

Each pair of agents belongs to one class of preference group: altruistic or moral. More precisely, each team is composed of two agents drawn from the same preference group, as in [1]. The principal only knows the proportion of the population that each group corresponds to:  $\lambda \in (0, 1)$  for altruist, and  $1 - \lambda$  for moral. The agents’ preferences in each group are represented by the utility functions

$$U^{Alt}(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \alpha_i) = \pi(e_i, e_j, \mathbf{w}_i) + \alpha_i \pi(e_j, e_i, \mathbf{w}_j) \tag{4}$$

for the altruists and

$$U^{HM}(e_i, e_j, \mathbf{w}_i, \kappa_i) = (1 - \kappa_i) \pi(e_i, e_j, \mathbf{w}_i) + \kappa_i \pi(e_i, e_i, \mathbf{w}_i) \tag{5}$$

for the moral agents, where  $\alpha \in [0, 1]$  and  $\kappa \in [0, 1]$  represent the agents’ degrees of altruism and morality, respectively. While subtle, the difference in the utility functions is crucial. For both preference groups, the first term captures the individual’s material payoff, given the strategy profile effectively used. For altruistic agents, the second term captures the impact of agent  $j$ ’s material payoff given the strategy profile  $(e_i, e_j)$  on agent  $i$ ’s utility, weighted by the degree of altruism  $\alpha_i$ . Meanwhile, for moral agents, the second term captures the Kantian moral concern. In other words, each moral individual considers what his material payoff would be if, hypothetically, the other agent were to mimic him. Extended discussions on the interpretations of moral preferences can be found in [2–4].

In what follows, as discussed in [1], I assume that  $\alpha_A = \alpha_B = \kappa_A = \kappa_B = \theta$ , and focus on the comparable functions

$$U^{Alt}(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \theta) = \pi(e_i, e_j, \mathbf{w}_i) + \theta \pi(e_j, e_i, \mathbf{w}_j), \tag{6}$$

$$U^{HM}(e_i, e_j, \mathbf{w}_i, \theta) = (1 - \theta) \pi(e_i, e_j, \mathbf{w}_i) + \theta \pi(e_i, e_i, \mathbf{w}_i). \tag{7}$$

As pointed out in [2] and also explored in [14], this is the formulation that gives rise to the behavioral equivalence between *homo moralis* and altruistic preferences. Under an appropriate change of variables, the altruistic utility function could be rewritten as  $U^{Alt}(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \theta) = (1 - \tilde{\theta}) \pi(e_i, e_j, \mathbf{w}_i) + \tilde{\theta} \pi(e_j, e_i, \mathbf{w}_j)$ , for  $\tilde{\theta} \in [0, 1/2]$ .

Throughout the exposition, I use the superscripts *HM* to denote equations and variables relevant to moral agents, and *Alt* to refer to altruistic ones.

The timing of the game is depicted in Figure 1.

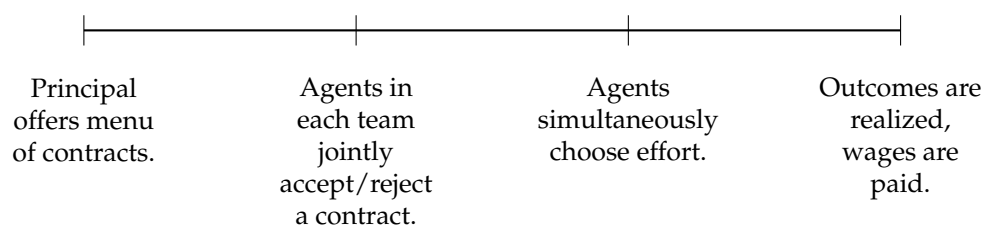


Figure 1. Timing of the game.

### 3. Screening

Due to the assumption of a common degree of morality or altruism, I restrict attention to symmetric contracts offered to each team. These assumptions simplify the problem in

the sense that both the incentive compatibility and individual rationality constraints are similar to the ones studied in the literature with a single agent, save for their dependence on the common degree of morality/altruism. For the pure moral hazard problem, these constraints are

$$(1 + \theta) [p_2 u(w^H) + (1 - p_2) u(w^L) - c] \geq \bar{u}_{Alt}, \quad (8)$$

$$u(w^H) - u(w^L) \geq \frac{c}{(1 + \theta)(p_2 - p_1)} \quad (9)$$

for altruistic agents, and

$$p_2 u(w^H) + (1 - p_2) u(w^L) - c \geq \bar{u}_{HM}, \quad (10)$$

$$u(w^H) - u(w^L) \geq \frac{c}{(p_2 - p_1) + \theta(p_1 - p_0)} \quad (11)$$

for moral agents.

In contrast to [1], I allow the different groups to have different reservation utilities. Two particular cases deserve a special mention. First, as in [1], agents in each group may have exactly the same reservation utility  $\bar{u}_{Alt} = \bar{u}_{HM} = \bar{u}$ , which generates different utility levels for the participating agents whenever  $\bar{u} > 0$  due to the utility function representing each prosocial preference. The second particular case is  $\bar{u}_{Alt} = (1 + \theta)\bar{u}_{HM}$ , so that the participation constraints for both moral and altruistic agents are identical for any common degree of prosociality  $\theta \in (0, 1]$ . If  $\theta = 0$ , then both moral and altruistic agents behave as purely selfish individuals, and the screening problem becomes irrelevant.

Let  $\mathcal{C}_{Alt}$  denote the set of contracts that satisfy the participation and incentive compatibility constraints of altruistic agents, and similarly define the set  $\mathcal{C}_{HM}$  for moral agents. The principal's screening problem is to choose  $\mathbf{w}_{Alt} \in \mathcal{C}_{Alt}$  and  $\mathbf{w}_{HM} \in \mathcal{C}_{HM}$  such that neither group has an incentive to pick the contract designed for the other group. This is akin to the incentive compatibility constraint in the adverse selection problem, and it can be seen as an additional set of constraints in the principal's maximization program. The issue, however, is that the intersection between these two sets of feasible contracts is not empty, and thus one can always construct a separating equilibrium by selecting two contracts,  $\mathbf{w}$  and  $\mathbf{w}'$ , in  $\mathcal{C}_{Alt} \cap \mathcal{C}_{HM}$ , and arguing that each group will self-select into one, and only one, of these contracts.

I will, therefore, focus on a stronger form of separation: I will require that a menu of contracts has at most one element in the intersection of the feasible sets. This will ensure that at least one group has no incentives to deviate and accept the contract designed for the other group.

Let  $h = u(w)$ , which is uniquely defined for each  $w \in \mathbb{R}$  since  $u$  is strictly increasing by assumption. I can therefore rewrite the sets of feasible contracts using the linear constraints

$$(1 + \theta) [p_2 h^H + (1 - p_2) h^L - c] \geq \bar{u}_{Alt}, \quad (12)$$

$$h^H - h^L \geq \frac{c}{(1 + \theta)(p_2 - p_1)} \quad (13)$$

for altruistic agents and

$$p_2 h^H + (1 - p_2) h^L - c \geq \bar{u}_{HM}, \quad (14)$$

$$h^H - h^L \geq \frac{c}{(p_2 - p_1) + \theta(p_1 - p_0)} \quad (15)$$

for moral agents. I can easily draw the sets of feasible contracts for the cases in which the production technology displays decreasing or increasing returns to efforts, by appropriately choosing the reservation utilities  $\bar{u}_{HM}$  and  $\bar{u}_{Alt}$ . In Figures 2 and 3, I assume that  $\bar{u}_{HM} = \bar{u}_{Alt} > 0$ . Notice, in Figure 2, that  $\mathcal{C}_{HM} \subset \mathcal{C}_{Alt}$ , which implies that any feasible contract offered to moral agents is also accepted by altruistic ones at the same time that it also

provides the latter with incentives to exert high effort. Meanwhile, in Figure 3, a contract in  $C_{HM}$  is also accepted by altruistic agents, but it may not necessarily induce them to exert the high effort.

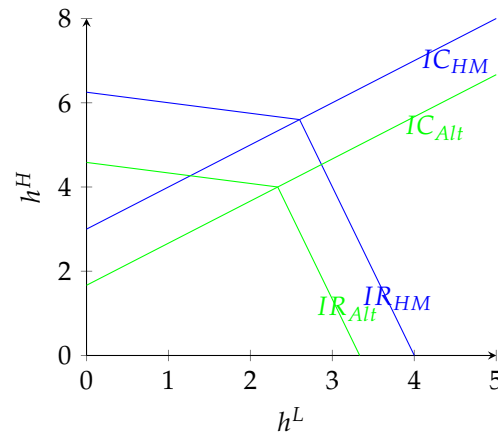


Figure 2. Feasible sets of contracts for  $p_2 - p_1 > p_1 - p_0$ .

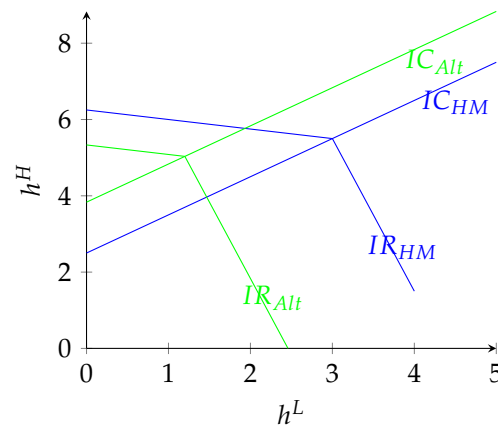


Figure 3. Feasible sets of contracts for  $p_2 - p_1 < p_1 - p_0$ .

**Lemma 1.** Suppose that  $\bar{u}_{Alt} < (1 + \theta)\bar{u}_{HM}$ . There exists no strictly separating menu of contracts that induces both agents to exert high effort. Similarly, if  $\bar{u}_{Alt} > (1 + \theta)\bar{u}_{HM}$ , no strictly separating equilibrium exists that induces all agents to exert effort.

Lemma 1 states that a separating equilibrium does not exist if the reservation utility of both groups is such that the participation constraints are never identical and contracts incentivize agents to exert the high effort. Then, for each case, one can find a profitable deviation for a group, i.e. either the moral agents are better off taking the contract designed for the altruistic teams, or altruists like the moral contracts better than their own.

**Lemma 2.** Suppose that  $\bar{u}_{Alt} = (1 + \theta)\bar{u}_{HM} \geq 0$ . There exists no strictly separating menu of contracts that induces both agents to exert high effort.

The negative results in Lemmas 1 and 2 can be linked to the fact that the isoutility curves of moral and altruistic agents never cross in the region where both groups are incentivized to exert the high effort. Indeed, under the assumptions of each statement, the indifference curves of each kind of prosocial agent are either identical to one another, or they are parallel. It is this violation of a single-crossing-like property that prevents the principal from finding schedules that elicit the agents' preferences.

**Proposition 1.** *There exists no strictly separating menu of contracts that induces both types of agents to accept a contract and exert high effort for any  $\bar{u}_{Alt}, \bar{u}_{HM} \in \mathbb{R}_+$ .*

### 3.1. Separating Equilibria with Low Effort

A separating equilibrium also does not exist if the principal requires both types of agents to exert low effort. Indeed, due to risk aversion by the agents, the principal can induce participation by offering the constant schedules  $\bar{w}_{Alt}$  and  $\bar{w}_{HM}$  for the altruistic and moral groups, respectively, satisfying the individual rationality constraints

$$u(\bar{w}_{Alt}) \geq \frac{\bar{u}_{Alt}}{1+\theta}, \quad (16)$$

$$u(\bar{w}_{HM}) \geq \bar{u}_{HM}. \quad (17)$$

By standard arguments, these constraints must bind in an equilibrium. However, if  $\bar{u}_{HM} \neq \frac{\bar{u}_{Alt}}{1+\theta}$ , one preference group always has incentives to deviate and accept the contract designed for the second group. On the other hand, if  $\bar{u}_{HM} = \frac{\bar{u}_{Alt}}{1+\theta}$ , then  $\bar{w}_{Alt} = \bar{w}_{HM}$  since  $u$  is strictly increasing, which implies that all workers accept exactly the same contract, and thus picking them apart is impossible for the principal. This argument is collected in the following result.

**Proposition 2.** *No separating equilibrium exists if the principal wishes to induce both preference groups to accept the contract and exert low effort.*

### 3.2. Screening Preference Groups through Exclusion

Propositions 1 and 2 have shown that the principal cannot screen moral agents from altruistic ones when she must induce both participation and high effort. However, the principal might be able to screen the different preference groups by offering a single (non-null) contract.

Turn once more to Figure 2, by assuming identical reservation utilities and increasing returns to effort. If the principal offers a menu with a single contract that satisfies both the participation and incentive compatibility constraint of the altruistic agent with equality (the intersection of the green lines), she will ensure that: (i) altruistic agents accept the offer and exert high effort, and (ii) moral agents choose not to participate in the relationship with the principal. The same can be achieved under decreasing returns to effort by offering a similar contract (Figure 3).

More generally, the principal can screen the preference groups by offering a singleton menu, where the contract offered necessarily satisfies with equality the participation constraint of the preference group with the lowest reservation utility.

**Proposition 3.** *Suppose that  $\bar{u}_{HM} \neq \frac{\bar{u}_{Alt}}{1+\theta}$ . The principal can screen different preference groups by offering a single contract that excludes the agents with the highest reservation utility.*

Proposition 3 holds either when the principal wishes to induce high or low effort. For the latter case, the argument behind Proposition 3 is even more compelling since the principal will offer a constant wage schedule to the risk-averse agents to exert zero effort, and therefore she can simply choose to employ the cheapest of the preference groups in terms of reservation utilities.

### 3.3. Screening with Different Efforts

So far, my analysis has focused on the case where both groups of agents are required by the principal to exert the same level of effort, either high or low. The negative results are basically a consequence of the indifference curves for the two groups being parallel to one another when efforts are the same: this implies that the contract offered to the group with the highest outside option also attracts the other team.

Although excluding one preference group from participating in the relationship with the principal is one way to screen agents, a second one exists, namely, requiring that only one group exerts high effort.

If only one group is expected to exert effort, the incentive compatibility constraint with respect to effort can be neglected for that group. Moreover, a constant schedule should be offered to that same group due to the agents' risk-aversion. In what follows, I will denote by 1 the preference group that should exert effort, and by 2 the preference group who should not exert effort. The feasible set of contracts for the principal will be given by all values of  $\mathbf{w} = ((w_1^H, w_1^L), w_2)$  satisfying the incentive compatibility and individual rationality constraints for group 1, and the participation constraint for group 2.

One must, however, notice an important difference between the participation constraints for both groups. For group 1, which is bound by the incentive compatibility constraint, the (IR) is given by

$$p_2u(w_1^H) + (1 - p_2)u(w_1^L) \geq \bar{u}_1 \quad \text{for all } (w^H, w^L) \text{ that satisfies (IC),} \quad (18)$$

$$p_0u(w_1^H) + (1 - p_2)u(w_1^L) \geq \bar{u}_1 \quad \text{otherwise.} \quad (19)$$

On the other hand, for group 2, participation must satisfy

$$u(w_2) = p_0u(w_2^H) + (1 - p_0)u(w_2^L) \geq \bar{u}_2. \quad (20)$$

If the individual rationality curves never intersect, i.e. if either  $\bar{u}_2 < \bar{u}_1$  or  $\bar{u}_2 \gg \bar{u}_1$ , then a separating equilibrium does not exist, for the simple reason that the contract offered to the group with the highest outside option also attracts the agents of the other group, in much a similar manner to the case where the principal induces no group to high effort.

This is not true if the participation constraints intersect (which requires that  $\bar{u}_2 \geq \bar{u}_1$ ). Using the linearization  $h = u(w)$ , the feasible set of contracts can be represented as in Figure 4 below.

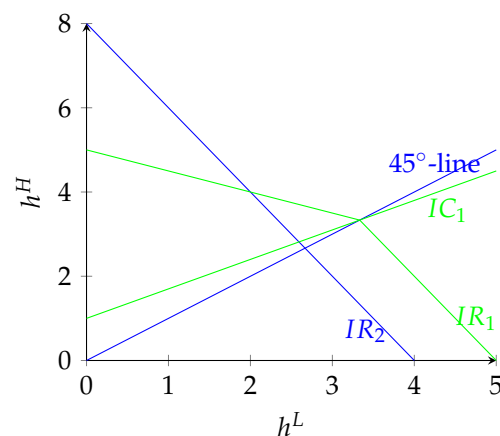


Figure 4. Feasible sets of contracts for different efforts.

The contract offered to the agents in group 2 is given by the intersection of the 45° line with the participation constraint for said group, since such a point has the principal proposing a constant schedule to the agents who are not expected to exert effort. On the other hand, agents in group 1 are offered the contract lying in the intersection between the two participation constraints, which they strictly prefer to the constant schedule of group 1 (while the latter is indifferent between the two contracts). The assumption that the participation constraints intersect also implies that the incentive compatibility constraint for group 1 is satisfied.

One remark is in order here: the principal leaves group 1 agents some rent for exerting the high effort, in the sense that the incentive compatibility constraint is not necessarily satisfied with equality (i.e. the pair  $(w_1^H, w_1^L)$  does not lie in  $IC_1$ ). This can be interpreted as



a *no distortion at the top* result: the principal's offer does not distort (downward) the effort demanded from the least costly group, but she must still pay a rent to that group.

**Proposition 4.** *Suppose that  $\bar{u}_{HM}$  and  $\bar{u}_{Alt}$  are such that  $\bar{u}_{HM} \neq \frac{\bar{u}_{Alt}}{1+\theta}$  and the individual rationality constraints from both groups cross each other once. Then, a separating equilibrium exists if the principal induces only one preference group to exert the high effort.*

#### 4. Pooling Equilibria

Let  $\mathbf{w}_{HM}$  be the contract that satisfies both conditions in  $\mathcal{C}_{HM}$  with equality, and similarly, define  $\mathbf{w}_{Alt}$ . Additionally, denote by  $\mathbf{w}_P$  the contract that satisfies both the participation constraint for moral agents and the incentive compatibility constraint for altruistic agents with equality. The following proposition states the result formally.

**Proposition 5.** *Suppose that  $\bar{u}_{Alt} = \bar{u}_{HM} = \bar{u} > 0$ .  $\mathbf{w}_{HM}$  constitutes a pooling equilibrium with both groups of agents exerting the high effort under increasing returns to efforts, while  $\mathbf{w}_P$  constitutes such an equilibrium under decreasing returns to efforts.*

I do not claim in Proposition 5 that  $\mathbf{w}_{HM}$  and  $\mathbf{w}_P$  are the unique pooling equilibrium contracts under increasing and decreasing returns to efforts, respectively. Indeed, in the former case, any contract in  $\mathcal{C}_{HM}$  indeed constitutes a pooling equilibrium. These two contracts, however, are completely characterized by a simple linear system of two equations. They also characterize one pooling equilibrium when  $\bar{u}_{Alt} = (1 + \theta)\bar{u}_{HM}$ : in this case, the participation constraints for both groups are identical, and characterizing the feasible sets for the contracts depends only on comparisons of the incentive compatibility constraints. Moreover, they are the least costly for the principal to offer among all pooling contracts.

However, one must stress that the principal is better off by offering a menu of contracts that screens different preference groups through exclusion. Indeed, since the principal always favors one preference group over the other for any degree of prosociality different than zero (where both moral and altruistic agents would be identical to purely selfish ones), the principal maximizes expected profits by hiring only the cheapest group he can incentivize to participate and exert the high effort, while excluding the more expensive group from partaking in any relationship with himself.

#### 5. Discussion

The results presented above, in line with the literature on screening prosocial preferences, imply that the principal may be unable to construct a menu of contracts that is successful in screening teams of agents belonging to different preference groups. As a consequence, developing experiments to infer agents' preferences in a static environment would present the same difficulties.

However, one possible strategy would be to offer the contracts sequentially. To fix ideas, suppose that the production technology exhibits increasing returns to efforts, and that  $\bar{u}_{Alt} = \bar{u}_{HM}$ . Under these circumstances,  $\mathcal{C}_{HM} \subset \mathcal{C}_{Alt}$ , as was argued in the proof of Lemma 1. If agents are perfectly patient, then the principal could offer  $\mathbf{w}_{Alt}$  in the first period, which would be accepted by all the altruistic agents but not by the moral ones, and only then offer  $\mathbf{w}_{HM}$  to the remaining agents. Such sequential mechanism would make use of time to screen the agents, a channel that is not available in the static model described above.

There are two main issues with such an approach, at least from a theoretical viewpoint. First, if all the potential employees are aware that the employer would utilize the sequential offer mechanism above, altruistic agents would not accept  $\mathbf{w}_{Alt}$  in the first period in order to contract under  $\mathbf{w}_{HM}$  in the second period and therefore enjoy a higher utility. Clearly, such deviation by altruistic agents would again leave the principal unable to screen between the two preference groups.

Secondly, the sequential approach relies on the agents being infinitely patient and the two preference groups displaying the same reservation utility. The mechanism could still be employed in the situation where  $\bar{u}_{Alt} < (1 + \theta)\bar{u}_{HM}$  and  $\delta_{Alt} < \delta_{HM}$ , where  $\delta_j \in (0, 1)$  denotes group  $j = \{Alt, HM\}$  discount factor. In this case, if the altruistic group discounts the future much more than its moral counterpart, the mechanism could indeed lead to full screening. Unfortunately, to the best of my knowledge, I do not know any research establishing conditions under which different prosocial preferences lead to heterogeneous discount factors.

## 6. Concluding Remarks

This paper extends the analysis in [1] by relaxing the assumption that the agents' preferences are common knowledge in the contractual relationship. In particular, the interest lies in characterizing a separating equilibrium in which moral and altruistic individuals reveal their type and exert a high level of effort in the task proposed by the principal.

In effect, the results are negative, but in line with the literature of adverse selection followed by moral hazard: screening prosocial preferences is not possible unless the principal distorts the allocation provided to the least preferred group by either excluding them from the relationship or by inducing a different level of effort. The empirical implication follows naturally: one cannot distinguish groups of agents characterized by the two classes of preferences described above when the degree of prosociality is the same for the two groups, without relaxing either the participation or the moral hazard incentive compatibility constraint.

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

**Acknowledgments:** I thank Ingela Alger, Renee Bowen, Leandro de Magalhaes, Roger Myerson, François Salanié, Jörgen Weibull, and three anonymous referees for very valuable feedback. Finally, I also thank seminar audiences at Toulouse School of Economics, as well as conference participants at the 2018 Africa Meeting of the Econometric Society.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Proofs

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

The assumption that  $\bar{u}_{Alt} < (1 + \theta)\bar{u}_{HM}$  implies that the participation constraint for one group is different than the one for the other. In particular, setting  $h = u^{-1}(w)$ , and drawing the participation constraints on the plane  $(h^L, h^H)$  allows us to see that the participation constraint for the altruistic group is always below its counterpart for moral agents.

The proof then considers two cases in turn. Suppose first that the production technology is characterized by decreasing returns to efforts, i.e.  $p_2 - p_1 < p_1 - p_0$ . Then, any contract  $\mathbf{w} \in \mathcal{C}_{HM}$  satisfies the participation constraint of altruistic agents with slackness, and thus it is profitable for this group of agents to deviate and take the contract designed for moral agents. Thus, no separating equilibrium exists in this case.

Under increasing returns to efforts,  $p_2 - p_1 \geq p_1 - p_0$ , on the other hand, one can readily check that  $\mathcal{C}_{HM} \subset \mathcal{C}_{Alt}$ . Then, again, all altruistic agents would deviate and choose the contract designed for moral agents, since this contract would satisfy the former group's participation constraint with slackness.

For  $\bar{u}_{Alt} > (1 + \theta)\bar{u}_{HM}$ , the proof is similar. If  $p_2 - p_1 < p_1 - p_0$ , then  $\mathcal{C}_{Alt} \subset \mathcal{C}_{HM}$  and thus every altruistic agent has an incentive to deviate and take the contract designed for moral agents. Conversely, if  $p_2 - p_1 > p_1 - p_0$ , every contract in  $\mathcal{C}_{Alt}$  satisfies the participation constraint of the moral agents with slackness, and therefore such agents have incentives to deviate and take the contract designed for the former group.

### Appendix A.2. Proof of Lemma 2

Suppose first that  $p_2 - p_1 \geq p_1 - p_0$ , so that the incentive compatibility constraint of moral agents is always above the one for altruists. Then,  $\mathcal{C}_{HM} \subseteq \mathcal{C}_{Alt}$  and the latter always prefer a contract designed for the former. If  $p_2 - p_1 < p_1 - p_0$ , the reverse holds:  $\mathcal{C}_{Alt} \subseteq \mathcal{C}_{HM}$  and moral agents always prefer the contract designed for altruists rather than their own.

### Appendix A.3. Alternative Proof of Proposition 1

Proposition 1 is a direct consequence of Lemma 1 and Lemma 2. Alternatively, the same result can be reached by the following reasoning. Suppose that  $\mathbf{h}_{Alt} = (h_{Alt}^H, h_{Alt}^L)$  and  $\mathbf{h}_{HM} = (h_{HM}^H, h_{HM}^L)$  are contracts that constitute a strictly separating equilibrium. Then, no team of agents has incentives to deviate and take the contract designed for the other group. In particular, this means that for altruistic agents the following inequality must hold

$$(1 + \theta)[p_2 h_{Alt}^H + (1 - p_2)h_{Alt}^L - c] > (1 + \theta)[p_2 h_{HM}^H + (1 - p_2)h_{HM}^L - c], \quad (\text{A1})$$

while

$$p_2 h_{HM}^H + (1 - p_2)h_{HM}^L - c > p_2 h_{Alt}^H + (1 - p_2)h_{Alt}^L - c \quad (\text{A2})$$

must hold for moral agents. Since  $\theta \in [0, 1]$ , condition (A1) reduces to

$$p_2 h_{Alt}^H + (1 - p_2)h_{Alt}^L - c > p_2 h_{HM}^H + (1 - p_2)h_{HM}^L - c, \quad (\text{A3})$$

which together with (A2) imply that

$$p_2 h_{Alt}^H + (1 - p_2)h_{Alt}^L > p_2 h_{HM}^H + (1 - p_2)h_{HM}^L > p_2 h_{Alt}^H + (1 - p_2)h_{Alt}^L, \quad (\text{A4})$$

a contradiction.

### Appendix A.4. Proof of Proposition 3

For  $\bar{u}_{HM} \neq \frac{\bar{u}_{Alt}}{1+\theta}$ , the proof is analogous to that of Lemma 1. The impossibility of screening through exclusion when  $\bar{u}_{HM} = \frac{\bar{u}_{Alt}}{1+\theta}$  comes from the argument of Proposition 2 if the principal does not wish to induce high effort, and generalizes straightforwardly to the case when she wishes to induce effort.

### Appendix A.5. Proof of Proposition 5

Under increasing returns to efforts,  $\mathcal{C}_{HM} \subset \mathcal{C}_{Alt}$ , and thus  $\mathbf{w}_{HM} \in \mathcal{C}_{Alt}$ . In particular, as shown in Sarkisian (2017), this contract is the least costly one the principal can offer to moral agents in order to induce both participation and effort.

Under decreasing returns to efforts,  $\frac{c}{(1+\theta)(p_2-p_1)} > \frac{c}{(p_2-p_1)+\theta(p_1-p_0)}$ , and thus any contract satisfying the incentive compatibility constraint for altruistic agents automatically satisfies, with slackness, its counterpart for moral agents. Therefore, let us take the most restrictive set of constraints, namely the participation constraint for moral agents and the participation constraint for altruistic agents. Any contract satisfying both constraints, in particular,  $\mathbf{w}_P$  will necessarily belong to both  $\mathcal{C}_{HM}$  and  $\mathcal{C}_{Alt}$ , so moral and altruistic agents alike accept such contract and exert the high level of effort.

## References

1. Sarkisian, R. Team Incentives under Moral and Altruistic Preferences: Which Team to Choose? *Games* **2017**, *8*, 37.
2. Alger, I.; Weibull, J.W. Homo Moralistic-Preference Evolution Under Incomplete Information and Assortative Matching. *Econometrica* **2013**, *81*, 2269–2302.
3. Alger, I.; Weibull, J.W. Evolution and Kantian Morality. *Games Econ. Behav.* **2016**, *98*, 56–67.
4. Alger, I.; Weibull, J.W. Strategic Behavior of Moralists and Altruists. *Games* **2017**, *8*, 38.
5. Laffont, J.J.; Martimort, D. (Eds). *The Theory of Incentives: The Principal-Agent Model*, 1st ed.; Princeton University Press: Princeton, NJ, USA, 2002; Volume 1.

6. Bolton, P.; Dewatripont, M. *Contract Theory*; The MIT Press: Cambridge, MA, USA, 2005; Volume 1.
7. Jullien, B.; Salanié, B.; Salanié, F. Screening risk-averse agents under moral hazard: single-crossing and the CARA case. *Econ. Theory* **2007**, *30*, 151–169. doi:10.1007/s00199-007-0299-3.
8. Ollier, S.; Thomas, L. Ex post participation constraint in a principal-agent model with adverse selection and moral hazard. *J. Econ. Theory* **2013**, *148*, 2383–2403.
9. Maréchal, F.; Thomas, L. The Optimal Contract under Adverse Selection in a Moral-Hazard Model with a Risk-Averse Agent. *Games* **2018**, *9*, 12.
10. von Siemens, F. Heterogeneous social preferences, screening, and employment contracts. *Oxf. Econ. Pap.* **2011**, *63*, 499–522.
11. Cabrales, A.; Charness, G. Optimal contracts with team production and hidden information: An experiment. *J. Econ. Behav. Organ.* **2011**, *77*, 163–176.
12. Demougin, D.; Fluet, C.; Helm, C. Output and Wages with Inequality Averse Agents. *Can. J. Econ. Rev. Can. D'Economique* **2006**, *39*, 399–413.
13. Grossman, S.J.; Hart, O.D. An Analysis of the Principal-Agent Problem. *Econometrica* **1983**, *51*, 7–45.
14. Bergström, T.C. On the Evolution of Altruistic Ethical Rules for Siblings. *Am. Econ. Rev.* **1995**, *85*, 58–81.