Research in Economics 68 (2014) 27–38

Contents lists available at ScienceDirect

Research in Economics

journal homepage: www.elsevier.com/locate/rie

Optimal contracting with altruism and reciprocity

Matteo Bassi [a], Marco Pagnozzi [a], Salvatore Piccolo [b]

[a] Università di Napoli Federico II and CSEF, Italy
[b] Università Cattolica del Sacro Cuore, Milan and CSEF, Italy

ARTICLE INFO

Article history:
Received 10 May 2013
Accepted 8 December 2013
Available online 18 December 2013

Keywords:
Adverse selection
Altruism
Reciprocity
Optimal contracting

ABSTRACT

Motivated by the recent experimental evidence on altruistic behavior, we study a simple principal–agent model where each player cares about other players' utility, and may reciprocate their attitude towards him. We show that, relative to the selfish benchmark, efficiency improves when players are altruistic. Nevertheless, in contrast to what may be expected, an increase in the degree of the agent's altruism as well as a more reciprocal behavior by players has ambiguous effects on efficiency. We also consider the effects of the presence of spiteful players and discuss how monetary transfers between players depend on their degrees of altruism and spitefulness.

1. Introduction

Standard economic theory assumes that agents are selfish and only care about their own monetary utility. In practice, however, elements such as fairness, altruism and reciprocity seem to play a crucial role in individual and collective decision making—see, e.g., Becker (1976), Kahneman et al. (1986), and Berg et al. (1995). The experimental evidence amply supports this view. For example, Thaler (1988) finds that, when playing the ultimatum game, proposers (who should make arbitrarily small offers in theory) typically offer equal divisions with responders, who frequently reject ungenerous offers. Similarly, Dawes and Thaler (1988) find that participants in public good contribution games typically make positive contributions, although (in theory) they should not.[1] This suggests that, in real life, individuals are altruistic (i.e., they care about each other's utility) and act reciprocally (i.e., they are good to other good people, and hurt those who hurt them).

Motivated by this evidence, we introduce behavioral elements in a simple principal–agent relationship where the agent is privately informed about his marginal cost of production in order to analyze the effects of altruistic and reciprocal motives in a standard adverse selection model à la Baron and Myerson (1982). We show that the presence of reciprocal and altruistic motives affects not only the enforcement of incentive contracts (as shown by Fehr et al., 1997), but also their design and efficiency properties. Specifically, we derive the optimal incentive compatible contract and show how the degrees of altruism and reciprocity affect the standard trade-off between efficiency and rent extraction. The predictions of the model apply to a wide range of standard contracting environment, such as employer–employee relationships, manufacturer–retailer deals, regulatory policies, etc., which are usually analyzed under the hypothesis that the contracting parties are selfish.

Following Levine (1998), we model altruism by introducing a positive weight assigned by a player to his opponents' monetary payoff, and we model reciprocity by assuming that this weight depends on how altruistic the opponents are.[2]

E-mail addresses: matteo.bassi@unina.it (M. Bassi), pagnozzi@unina.it (M. Pagnozzi), salvapiccolo@gmail.com (S. Piccolo).

[1] See also Kahneman et al. (1986), Fehr et al. (1997), and Fehr and List (2004).

[2] By contrast, in Rabin (1993), Segal and Sobel (1999), and Falk and Fischbacher (2006), a player's degree of altruism depends on his own utility (with respect to a "fair" utility level). In Battigalli and Dufwenberg (2009), instead, higher-order beliefs, beliefs of others, and plans of action influence motivation and behavioral concerns, so as to capture dynamic psychological effects (such as sequential reciprocity, psychological forward induction, and regret).

1090-9443/$ - see front matter
http://dx.doi.org/10.1016/j.rie.2013.12.001

Hence, we distinguish between a player's intrinsic altruistic attitude toward opponents, which is an innate characteristic, and his global attitude, which also depends on the interaction between the opponents' intrinsic attitude and the degree of reciprocity.

If the principal and the agent are globally altruistic, the inefficiency due to asymmetric information is lower than with selfish players. Moreover, the more altruistic is the principal, the closer the level of production is to the first-best outcome. The reason is that the principal's global altruism relaxes the trade-off between rents and efficiency and allows players to exploit production opportunities that, with selfish players, were ruled out by asymmetric information. Surprisingly, though, an increase in the agent's global altruism decreases efficiency (i.e., reduces output) because a relatively more altruistic agent is less responsive to monetary incentives, which makes it more costly for the principal to induce an efficient type not to mimic an inefficient one, thus worsening the standard ‘distortion at the bottom’ result and leading to higher distortions for the quantity produced by inefficient types. Therefore, although altruistic players trade more efficiently than selfish ones, more altruistic players do not necessarily trade more efficiently than less altruistic ones (in contrast to what may be expected).

We also determine the impact of changes in players' intrinsic attitude on efficiency. While improvements in the principal's intrinsic altruism always increase efficiency, changes in the agent's intrinsic altruism generate efficiency gains only under specific conditions on the degree of reciprocity between players. If the level of reciprocity is high, the principal rewards a more altruistic agent by reducing output distortions. If reciprocity is low, the principal has a weaker incentive to limit distortions to reward the agent, and hence he reduces the output further. Moreover, the effect of increasing reciprocity between players depends on the difference between the agent's and the principal's intrinsic attitudes: if the agent has a more (resp. less) altruistic attitude than the principal, the principal rewards (resp. punishes) him by increasing (resp. decreasing) output and information rents. This non-monotone comparative statics stems from the opposite impact of players' global altruistic attitude on efficiency, and it offers a set of new testable implications on the link between optimal contracting, efficiency and behavioral concerns under asymmetric information.

Players' altruism also has interesting effects on the monetary transfer paid by the principal to the agent. When the agent is sufficiently altruistic, the transfer may be negative, so that the agent pays the principal in order to be able to produce. Moreover, contrary to what may be expected, a more altruistic principal may manage to induce the agent to produce a higher quantity (thus increasing total surplus) and, at the same time, obtain a lower transfer. In our model, this “paradox of gift” happens arises when the agent is sufficiently altruistic and inefficient.

Finally, if players are globally spiteful—i.e., they assign a negative weight to their opponent's monetary payoff—the inefficiency due to asymmetric information is higher than with selfish players because the principal always increases the output distortion to reduce the agent's rent. Contrary to the case of altruistic players, a reduction in the degree of global or intrinsic spitefulness always reduces this inefficiency. The reason is that a less spiteful principal cares less about reducing the agent's rent, while a less spiteful agent cares more about total surplus and less about the transfer. In both cases the incentive problem is relaxed, so that the principal needs to distort output relatively less. As with altruistic players, the effect of increasing reciprocity depends on the difference between the agent's and the principal's intrinsic attitudes.

Our findings contribute to the literature on optimal contracting with altruistic and motivated agents. Shchetinin (2009) analyzes optimal contracting in a principal–agent model where the agent is altruistic only if the principal is also altruistic and there is asymmetric information on the degree of altruism. By contrast, we allow both the principal and the agent to be altruistic and we assume that the asymmetric information is on the agent's production cost. Siciliani (2009), Chone and Ma (2004), and Jack (2004) analyze the role of altruism in designing physician's contracts, under the assumption that the physician displays intrinsic altruism toward the patient and is privately informed about his health conditions. Similarly, Delfgaauw and Dur (2008) study how the intrinsic motivation of privately informed workers affect efficiency in a perfectly competitive market, while Delfgaauw and Dur (2009) also consider the case of a monopolistic principal that is only interested in minimizing costs. Shchetinin (2010), Akerlof and Yellen (1990) and Dur (2009), instead, study the effect of employers' intrinsic altruism on workers' effort levels. All these models only consider altruism on one side of the contractual relationship and, differently from us, do not distinguish between altruistic and reciprocal behavior.

In moral hazard environments, Dufwenberg and Kirchsteiger (2000), Netzer and Schmutzler (2010) and Immordino and De Marco (2013) show that, when a selfish principal interacts with reciprocal agents, efficiency generally increases in symmetric equilibria. Similarly, Dur and Tichem (2012) show that the presence of altruistic players who induce good social relationships in the workplace improves the capacity of relational contracts to induce workers' high effort, while bad social relationships might undermine it. With adverse selection, however, we show that the beneficial effect of reciprocal and altruistic concerns may be outweighed by the effects of these concerns on information rents, even in a single principal–agent relationship.

The rest of the paper is structured as follows. In Section 2 we present the model. Section 3 develops two benchmarks: one where there is asymmetric information but players are selfish, the other where players have altruistic and reciprocal concerns but there is complete information. In Section 4 we characterize the optimal contract with altruistic and reciprocal behavior and perform the relevant comparative statics. Section 5 considers spiteful players. Section 6 concludes.

2. The model

Environment: Consider a principal–agent relationship under adverse selection—see, e.g., Baron and Myerson (1982) and Laffont and Martimort (2002). A risk-neutral principal (P) contracts with a risk-neutral agent (A) who produces output q at cost θq in exchange for a monetary transfer t. Production generates a surplus $S(q)$ for the principal, with $S'(\cdot) \geq 0$, $S'(0) = +\infty$, $\lim_{q \to +\infty} S'(q) = 0$, and $S''(\cdot) < 0$.

A is privately informed about θ (the marginal cost of production), which is distributed on the compact support $\Theta \equiv [\underline{\theta}, \overline{\theta}]$ according to the (commonly known) continuously differentiable and atomless c.d.f. $F(\theta)$, with density $F'(\theta) = f(\theta)$ and increasing inverse hazard rate $F(\theta)/f(\theta)$. Players' direct (monetary) utilities from contracting are $u_P = S(q) - t$ and $u_A = t - \theta q$.

Altruism and reciprocity: Following Levine (1998), we assume that each player maximizes an adjusted utility, which depends both on his own direct utility, and on the other player's direct utility. Specifically, player i obtains an adjusted utility equal to $v_i = u_i + \phi_i u_j$—i.e., the principal's utility is

$$v_P = S(q) - t + \phi_P(t - \theta q),$$

and the agent's utility is

$$v_A = t - \theta q + \phi_A(S(q) - t),$$

where the coefficient

$$\phi_i \equiv \frac{\alpha_i + \lambda \alpha_j}{1 + \lambda}$$

measures player i's *global* attitude toward player j. In particular, when $\phi_i > 0$, player i has a global altruistic attitude (or is *globally altruistic*).

The parameter $\alpha_i \geq 0$ is an index of player i's *intrinsic* altruism, while the parameter $\lambda \geq 0$ is the (common) measure of players' reciprocity, or attitude for fairness. For simplicity, we assume that α_A, α_P and λ are common knowledge. (See Remark 5 for a discussion of the additional complexities that emerge when this hypothesis is relaxed.)

When $\alpha_i > 0$ we refer to player i as intrinsically altruistic, as such a player has a positive regard for his opponent and his adjusted utility is increasing in player j's direct utility. If $\alpha_i = 0$ we refer to the player as *selfish*. We assume that $\alpha_i < 1$, so that no player has a higher regard for his opponent than for himself. In Section 5, we also consider intrinsically spiteful players with $\alpha_i < 0$.

The parameter $\lambda \in [0, 1]$ reflects the fact that a player may want to reciprocate his opponent's attitude, and hence weighs more the utility of an (intrinsically) altruistic opponent than of a selfish one.[3] If $\lambda = 0$ then $\phi_i = \alpha_i$ and there is pure altruism as in Ledyard (1995). If $\lambda = 1$ then $\phi_i = \phi_j$ and there is maximal reciprocity.

Contracts and timing: We use the Revelation Principle to characterize the optimal contract. Hence, P offers a direct revelation mechanism $\mathcal{M} \equiv \{q(\hat{\theta}), t(\hat{\theta})\}_{\hat{\theta} \in \Theta}$ to A where, given A's report $\hat{\theta}$, $q(\hat{\theta})$ is the output produced by A and $t(\hat{\theta})$ is the transfer paid to A. If A rejects the contract, players' utility is normalized to zero.

The timing of the game is as follows:

1. The agent learns his type.
2. The principal offers a mechanism $\mathcal{M}$.
3. If the agent accepts mechanism $\mathcal{M}$, he makes a report $\hat{\theta}$, and the output and the transfer are selected according to the mechanism.

3. Benchmarks

In this section we briefly analyze the two useful benchmarks of selfish players and altruistic players with complete information.

Selfish players: First, consider the case of selfish players—i.e., $\alpha_A = \alpha_P = \lambda = 0$. The agent produces the second-best output $q^{SB}(\theta)$ that solves the standard Baron-Myerson (1982) rule:

$$S'\left(q^{SB}(\theta)\right) = \theta + \frac{F(\theta)}{f(\theta)} \;\Rightarrow\; q^{SB}(\theta) \leq q^{FB}(\theta),$$

where $q^{FB}(\theta)$ is the first-best output such that $S'(q^{FB}(\theta)) = \theta$ (see, e.g., Laffont and Martimort, 2002, Ch. 2). The principal chooses an inefficiently low output to minimize the agent's information rent—i.e., the second-best output equalizes the marginal benefit from production to the virtual marginal cost.

Complete information: Second, assume that players feature altruistic and reciprocal behavior, but that the realization of θ is common knowledge, so that there is no adverse selection. In this case, the optimal output is equal to the first-best level $q^{FB}(\theta)$ regardless of α_A, α_P, and λ. Hence, altruism and reciprocity have no effect on efficiency. This *neutrality result* arises because, with complete information, the principal fully internalizes the effect of altruism and reciprocity through the choice of a transfer that extracts the agent's whole surplus. Hence, first-best efficiency is achieved.

[3] Experimental evidence in Gill et al. (2012), among others, show that agents reciprocate their principal's attitude.

4. Optimal contract

In contrast to the complete information benchmark, when players have altruistic and reciprocal concerns and the agent is privately informed about his production cost, the information rent paid by the principal to the agent (in order to induce truthful information disclosure) enters with weight different than 1 into the principal's objective function. This is for two reasons. First, since the principal cares about the agent's utility, reducing the information rent harms the principal. Second, since the agent also cares about the principal's utility, the former is less eager to extract a rent from the latter; hence, the principal may wish to distort more this rent in order to make mimicking less appealing for the agent. These two effects, which stem from the information rent that the agent enjoys thanks to his private information, have an opposite effect on the principal's objective function. The analysis of the impact of the trade-off between these two effects on the optimal contract is the objective of this section.

In order to characterize the incentive feasible allocations, let

$$v_A(\theta,\hat{\theta}) \equiv t(\hat{\theta})-\theta q(\hat{\theta})+\phi_A(S(q(\hat{\theta}))-t(\hat{\theta})) \quad \forall(\theta,\hat{\theta}) \in \Theta^2$$

be the agent's utility when his θ and he reports $\hat{\theta}$, and let $v_A(\theta) \equiv v_A(\theta,\theta)$ be the agent's rent. The principal solves

$$\max_{\{q(\cdot),t(\cdot)\}} \int_\theta [S(q(\theta))-t(\theta)+\phi_P(t(\theta)-\theta q(\theta))]\, dF(\theta),$$

subject to

$$v_A(\theta) \geq 0 \quad \forall \theta \in \Theta, \tag{1}$$

$$v_A(\theta) \geq v_A(\theta,\hat{\theta}) \quad \forall(\theta,\hat{\theta}) \in \Theta^2. \tag{2}$$

Condition (1) is A's participation constraint, while (2) is A's incentive compatibility constraint.[4]

Using the expression of the agent's utility, the transfer $t(\theta)$ as a function of the rent $v_A(\theta)$ is

$$t(\theta) = \frac{v_A(\theta)+\theta q(\theta)-\phi_A S(q(\theta))}{1-\phi_A}. \tag{3}$$

Substituting (3) into the principal's objective function, standard techniques (see the Appendix) allow to rewrite P's optimization program as

$$\max_{\{q(\cdot),v_A(\cdot)\}} \int_\theta [(S(q(\theta))-\theta q(\theta))(1-\phi_A\phi_P)-(1-\phi_P)v_A(\theta)]\, dF(\theta),$$

subject to (1) and

$$v_A(\theta) = v_A(\overline{\theta})+\int_\theta^{\overline{\theta}} q(x)\, dx.$$

Hence, with altruism, the objective function assigns different weights to A's rent, $v_A(\theta)$, and to the total surplus, $S(q(\theta))-\theta q(\theta)$. An increase in the degree of global altruism of the principal ϕ_P has two effects. First, it reduces the loss suffered by the principal for giving up a rent to the agent, because the principal cares more about the agent's utility. Second, it reduces the weight assigned to the total surplus in the principal's objective function, since $(1-\phi_A\phi_P)$ is decreasing in ϕ_P. As we will explain below, this second effect is larger the more altruistic the agent is—i.e., the larger is ϕ_A.

The next proposition characterizes the optimal contract chosen by the principal.

Proposition 1. *In the optimal contract, the output $q^*(\theta) > q^{SB}(\theta)$ satisfies the first-order necessary and sufficient condition*

$$S'(q^*(\theta)) = \theta + \underbrace{\frac{1-\phi_P}{1-\phi_A\phi_P}\frac{F(\theta)}{f(\theta)}}_{adjusted\ distortion}, \tag{4}$$

and the optimal transfer is

$$t^*(\theta) = \frac{1}{1-\phi_A}\left[\theta q^*(\theta)+\int_\theta^{\overline{\theta}} q^*(x)\, dx\right] - \frac{\phi_A}{1-\phi_A}S(q^*(\theta)). \tag{5}$$

[4] We do not impose a limited liability constraint on A's direct utility—i.e., $t(\theta)-\theta q(\theta) \geq 0$—because, with this constraint, altruism increases efficiency by creating countervailing incentives—see, e.g., Siciliani (2009) who considers a model where $\alpha_P = \lambda = 0$ and $\alpha_A \neq 0$.

The optimal contract depends on α_A, α_P and λ. Compared to the case of selfish players, efficiency increases with altruistic players since

$$\frac{1-\phi_P}{1-\phi_A\phi_P}\frac{F(\theta)}{f(\theta)} < \frac{F(\theta)}{f(\theta)},$$

and hence the distortion in output induced by asymmetric information is lower. The intuition is that, when the principal's adjusted utility assigns a positive weight to the agent's utility, the principal has a lower incentive to reduce the agent's information rent (because the weight assigned to the transfer is lower than in the selfish case). This relaxes the standard trade-off between efficiency and rent extraction, and hence induces the principal to increase the output produced.

Remark 1. In a model where α_A and α_P are not common knowledge, the contracting problem becomes much more complex. The reason is that, in this case, the contract offered by P to A may be contingent on: (i) A's reports about θ and α_A; (ii) P's claim about α_P. This requires additional incentive constraints both for the agent (which induces P to solve a multi-dimensional screening problem) and for the principal, who needs to offer menus of contracts that credibly signal his type to the agent, yielding an informed principal problem with common values à la Maskin and Tirole (1992).[5] We conjecture that, as the dimension of the information asymmetry increases, our qualitative results hold if the cost uncertainty is relatively large compared to the uncertainty about α_A and α_P. When this is not the case, new distortions may emerge and create more complex effects of altruism on optimal contracts. Notice however that, when α_A is common knowledge but α_P is not, and $\lambda=0$, our model is a special case of the informed principal problem with private values and risk neutrality analyzed in Maskin and Tirole (1990). These authors show that, because of quasi-linear utilities, in this case the optimal contract is the one characterized in Proposition 1.

In the next two sections, we discuss the effects of changes in players' altruism and reciprocity on the optimal output and the optimal transfer.

4.1. *Effects of altruism on efficiency*

What is the effect of the strength of players' global altruism on the optimal output? Since the adjusted distortion characterized in Proposition 1 that reduces $q^*(\theta)$ is decreasing in ϕ_P and increasing in ϕ_A, we have the following result.

Corollary 1. *$\partial q^*(\theta)/\partial\phi_P > 0$ and $\partial q^*(\theta)/\partial\phi_A < 0$. As $\phi_P \to 1$ output converges to the first-best level for any ϕ_A; as $\phi_A \to 1$ output converges to the second-best level.*

Therefore, an increase in the principal's degree of global altruism increases efficiency, since the principal cares relatively less about reducing the agent's rent and hence increases output. By contrast, efficiency decreases as the agent's degree of global altruism rises. To see this, notice that the marginal effect of an increase in the transfer on the agent's utility is decreasing in ϕ_A.[6] Hence, when ϕ_A increases the agent is less responsive to monetary incentives and it is more costly for the principal to induce an efficient agent not to mimic an inefficient one. So an altruistic principal prefers to reduce the output when trading off rent and efficiency. In other words, since (ceteris paribus) a more altruistic agent cares more about total surplus and less about the transfer, to minimize rents the principal assigns a lower weight to total surplus maximization, which in turn induces a higher output distortion.

Both ϕ_P and ϕ_A depend on the parameters measuring intrinsic altruism (α_A and α_P) and reciprocity (λ). The next proposition shows how changes in these parameters affect efficiency.

Proposition 2. *The optimal output $q^*(\theta)$ is:*

- *increasing in α_P;*
- *increasing in λ if $\alpha_A > \alpha_P$, and decreasing in λ if $\alpha_A < \alpha_P$;*
- *decreasing in α_A if $\lambda < \lambda^*$, and increasing in α_A if $\lambda > \lambda^*$, where λ^* is a unique threshold $\in (0,1)$.*

An increase in the principal's intrinsic altruism always increases efficiency, since this reduces the principal's incentive to reduce the agent's rent by distorting output downward. Notice that a higher α_P increases both ϕ_P and ϕ_A, which affect efficiency in opposite ways (as seen in Corollary 1). However, for $\lambda \in (0,1)$, the effect on ϕ_P prevails.

By contrast, an increase in the level of reciprocity increases efficiency if and only if the agent is more intrinsically altruistic than the principal. In fact, a higher λ implies that both players tend to reciprocate more the intrinsic attitude of their opponent, thus acting more like he does. When $\alpha_P > \alpha_A$, an increase in λ reduces the principal's global altruistic attitude

[5] Common values arise because, when $\lambda \neq 0$, α_A and α_P affect both players' utilities.

[6] Notice that $\partial^2 v_A/\partial t \partial\phi_A = -1$.

and increases the agent's global altruistic attitude—i.e.,

$$\frac{\partial \phi_P}{\partial \lambda} < 0 \quad \text{and} \quad \frac{\partial \phi_A}{\partial \lambda} > 0 \Leftrightarrow \alpha_P > \alpha_A.$$

By Corollary 1, both these effects reduce efficiency. By contrast, when the agent is more intrinsically altruistic than the principal, the effects of an increase in λ are reversed: the principal's global altruistic attitude increases while the agent's global altruistic attitude decreases. By Corollary 1, both these effects increase efficiency.

Finally, when the agent's intrinsic altruism rises, efficiency increases if and only if players' reciprocity is sufficiently high. To interpret this result, consider the two extreme cases. When $\lambda = 0$, there is no reciprocity and the effect of a change in α_A is equivalent to the effect of a change in ϕ_A in Corollary 1. When $\lambda = 1$, there is maximal reciprocity and both players have exactly the same global degree of altruism—i.e., $\phi_P = \phi_A$. In this case, the adjusted distortion is equal to $F(\theta)/(1+\phi_A)f(\theta)$, and an increase in the altruism of any player increases output and efficiency, because it makes both players care less about the monetary transfer, and more about the total surplus created.

Summing up, our analysis suggests that, even though globally altruistic players trade more efficiently than selfish ones as expected, an increase in the degree of intrinsic altruism or reciprocity of players does not necessarily yield higher efficiency. This point, which has not been made in the earlier literature on the effects of the presence of altruistic players, should be taken into account in order to properly evaluate the social desirability of public policies that tend to induce agents either to act more altruistically or to reciprocate more the attitude of their opponents.

4.2. Effects of altruism on transfers

How does the optimal monetary transfer $t^*(\theta)$ varies with the players' degrees of global altruism? It may be expected that, when the degree of global altruism of the principal increases, the principal chooses to pay a higher transfer to the agent. This is not necessarily the case, however. In fact, we show that a "paradox of gift" arises in our model: a more altruistic principal pays a lower transfer to agents that are sufficiently altruistic and inefficient,[7] although it induces them to produce more (relatively to less altruistic principals). The next proposition provides sufficient conditions for this result.

Proposition 3. *If θ is sufficiently close to $\overline{\theta}$ and*

$$\frac{(1-\phi_P)\phi_A}{(1-\phi_P\phi_A)(1-\phi_A)} \geq \overline{\theta} f(\overline{\theta}), \tag{6}$$

$t^*(\theta)$ is decreasing in ϕ_P.

Notice that the left-hand side of condition (6) is increasing in ϕ_A, and the condition cannot be satisfied when $\phi_A = 0$. Hence, the transfer paid to inefficient types is more likely to be decreasing in ϕ_P when the agent is sufficiently altruistic. Intuitively, two effects influence the responsiveness of the optimal transfer to the principal's global altruism. On the one hand, by Corollary 1, the optimal output increases as ϕ_P increases, which tends to raise $t^*(\theta)$. On the other hand, a higher output increases total surplus and, ceteris paribus, this reduces $t^*(\theta)$, since an altruistic agent cares relatively more about total surplus than the transfer. This second effect dominates when ϕ_A is large.

Notice also that the left-hand side of condition (6) is decreasing in ϕ_P, and the condition cannot be satisfied when $\phi_P = 1$: the paradox of gift arises only when the principal is not too altruistic. The reason is that, when $\phi_P \to 1$, the output tends to the first-best level, so that the second effect discussed above is negligible.

So far, we have only provided sufficient conditions under which the optimal transfer may decrease with ϕ_P. In order to fully characterize the comparative statics of $t^*(\theta)$ with respect to players' altruism (including ϕ_A),[8] we now consider the uniform-quadratic case—i.e., we assume that $S(q) = aq - \frac{1}{2}q^2$, with $a > 2$, and $\theta \sim U[0,1]$.

Lemma 1. *In the uniform-quadratic case, the optimal transfer $t^*(\theta)$ is decreasing in ϕ_A. Moreover, there are two unique thresholds $\phi_A^* < 1$ and $\phi_A^{**} < 1$ such that the optimal transfer $t^*(\theta)$ is*

- *negative if $\phi_A > \phi_A^*$;*
- *increasing in ϕ_P if $\phi_A < \phi_A^{**}$ or $\theta < \frac{1}{\sqrt{3}}$, and decreasing in ϕ_P if $\phi_A > \phi_A^{**}$ and $\theta > \frac{1}{\sqrt{3}}$.*

This result confirms the intuitions stated above: the dominant effect of a higher ϕ_P on $t^*(\theta)$ depends on the degree of A's global altruism and on his marginal cost. Moreover, an agent with a very high degree of global altruism may compensate the principal for being part of the relationship.

[7] Clearly for agent's types close to the most efficient one, the optimal transfer is increasing in ϕ_P.

[8] In general, the comparative statics of $t^*(\theta)$ with respect to ϕ_A is more complex than the comparative static with respect to ϕ_P. In fact, in addition to the indirect effects through $q^*(\theta)$, ϕ_A also has a direct effect on $t^*(\theta)$ (see Proposition 1). For the same reasons, it is difficult to determine the effects on the optimal transfer of changes in the parameters measuring intrinsic altruism and reciprocity.

Finally, increasing A's degree of global altruism has contrasting effects on the optimal transfer. First, by Corollary 1 a more altruistic agent induces P to reduce $q^*(\theta)$, and this tends to decrease $t^*(\theta)$. Second, ceteris paribus, a higher ϕ_A increases the weight attached by A to the total surplus relative to the transfer. This direct effect tends to reduce $t^*(\theta)$. Third, a lower optimal quantity reduces the total surplus, and this requires a higher $t^*(\theta)$ to compensate the agent. Under our assumptions, the first two effects prevail so that $t^*(\theta)$ is decreasing in ϕ_A.

5. Spitefulness

In this section, we extend the analysis to the case where both players are intrinsically spiteful—i.e., $\alpha_i < 0$, $i = A, P$—and, hence, have a global spiteful attitude—i.e., $\phi_i < 0$, $i = A, P$. In this case, each player has a negative regard for his opponent and his adjusted utility is decreasing in the other player's direct utility. A reduction in α_i implies an increase in the degree of spitefulness of player i. We assume that $|\alpha_i| < 1$, so that no player has a higher regard for his opponents than for himself.

5.1. Effects of spitefulness on efficiency

With spiteful players, the optimal contract is the same as the one characterized in Proposition 1. In contrast to the case of altruistic players, however, the presence of a globally spiteful principal always reduces efficiency compared to the case of selfish players since, when $\phi_P < 0$:

$$\frac{1-\phi_P}{1-\phi_A\phi_P}\frac{F(\theta)}{f(\theta)} > \frac{F(\theta)}{f(\theta)},$$

and hence $q^*(\theta) < q^{SB}(\theta)$. This holds regardless of whether the agent is globally spiteful or not. In fact, when the principal's adjusted utility assigns a negative weight to the agent's direct utility, the principal always increases the output distortion to reduce the agent's information rent.

Since with spiteful players the adjusted distortion due to asymmetric information characterized in Proposition 1 is decreasing in both ϕ_P and ϕ_A (because $\phi_i < 0$, $i = A, P$), we have the following result.

Corollary 2. *When $\alpha_i < 0$, $i = A$, P: $\partial q^*(\theta)/\partial \phi_P > 0$ and $\partial q^*(\theta)/\partial \phi_A > 0$.*

Hence, decreasing either the principal's or the agent's degree of global spitefulness increases efficiency. On the one hand, when ϕ_P increases the principal cares less about reducing the agent's rent and hence increases output. On the other hand, when ϕ_A increases, the agent cares more about total surplus and less about the transfer, so that the incentive problem is relaxed and the principal can distort output relatively less.

The next proposition shows how the parameters measuring intrinsic altruism (α_A and α_P) and reciprocity (λ) affect efficiency when players are intrinsically spiteful.

Proposition 4. *When $\alpha_i < 0$, $i = A, P$, the optimal output $q^*(\theta)$ is increasing in α_P and α_A. Moreover, there are two unique thresholds $\lambda^{**} \in (0,1)$ and $\underline{\alpha}_A \in (-1,0)$ such that the optimal output $q^*(\theta)$ is*

- *decreasing in λ if (i) $\alpha_P > \alpha_A$ or (ii) $\alpha_A > \underline{\alpha}_A$ and $\lambda < \lambda^{**}$;*
- *increasing in λ if (i) $\alpha_P < \alpha_A < \underline{\alpha}_A$ or (ii) $\alpha_A > \underline{\alpha}_A$ and $\lambda > \lambda^{**}$.*

When the principal becomes less intrinsically spiteful, efficiency increases, since this reduces the principal's incentive to reduce the agent's rent by distorting output downward. This is because a higher α_P increases both ϕ_P and ϕ_A and, by Corollary 2, these effects unambiguously increase efficiency. In contrast to the case of altruistic players, however, increasing α_A also increases output and efficiency, since a higher α_A increases both ϕ_P and ϕ_A.

The effects of a higher degree of reciprocity λ on efficiency are more interesting. If the agent is intrinsically more spiteful than the principal—i.e., $\alpha_P > \alpha_A$—an increase in the degree of reciprocity decreases efficiency. By the analysis of Section 4, when $\alpha_P > \alpha_A$ a higher λ increases P's global spitefulness ($\partial\phi_P/\partial\lambda < 0$) but decreases A's global spitefulness ($\partial\phi_A/\partial\lambda > 0$) and, by Corollary 2, these two changes have opposite effects on efficiency. On balance, the effect on P's global attitude is stronger and optimal output decreases if reciprocity increases: the principal's interest in decreasing the agent's rent prevails over the agent's reduced interest in the transfer.

By contrast, when the principal is intrinsically more spiteful than the agent—i.e., $\alpha_P < \alpha_A$—the effect on efficiency of a change in λ depends on the degree of the agent's intrinsic spitefulness. First, if α_A is relatively low, output and efficiency increase with λ. To see why, recall that $\partial\phi_P/\partial\lambda > 0$ and $\partial\phi_A/\partial\lambda < 0$ when $\alpha_P < \alpha_A$, but the effect on the principal's global attitude dominates when α_A is relatively low. Second, if α_A is relatively high, there is a large difference in players' degrees of spitefulness and efficiency increases with λ if and only if reciprocity is sufficiently high. To see why, consider the two extreme cases. When λ is small, the effect on the agent's global attitude prevails: the output distortion increases because inducing truthtelling is more costly for the principal. When λ is high, the effect on the principal's global attitude prevails: the output distortion decreases because the principal is less interested in reducing the agent's rent.

Summing up, considering players' intrinsic spitefulness provides new interesting results. First, the presence of a globally spiteful principal creates more allocative distortions compared to the second-best outcome with selfish players, while the agent's global attitude has no effect. Second, although changes in the degree of intrinsic spitefulness go in the expected direction, a change in reciprocity has non-trivial effects on efficiency and output.

5.2. Effects of spitefulness on transfers

How does the optimal transfer $t^*(\theta)$ vary with spitefulness? As in Section 4.2, to analyze this issue we consider the uniform-quadratic case and assume that $S(q) = aq - \frac{1}{2}q^2$, with $a > 2$, and $\theta \sim U[0,1]$.

Lemma 2. *In the uniform-quadratic case, when $\phi_i < 0$, $i = A, P$, the optimal transfer $t^*(\theta)$ is*

- *always positive*;
- *increasing in ϕ_P*;
- *decreasing in ϕ_A if either ϕ_P or ϕ_A are large enough*;
- *increasing in ϕ_A if both ϕ_P and ϕ_A are not too large.*

Hence, with globally spiteful players, the principal always pays to the agent a positive transfer. Moreover, when P's global spitefulness decreases—i.e., ϕ_P increases—the two effects of Lemma 1 go in the same direction, and the optimal transfer increases with ϕ_P. This is because, by Corollary 2, a less spiteful principal increases $q^*(\theta)$ and this increases rents and total production costs, which tends to increase $t^*(\theta)$. Furthermore, a higher $q^*(\theta)$ increases total surplus, and this also tends to increase $t^*(\theta)$.

Finally, when A's global spitefulness decreases—i.e., ϕ_A increases—there are three contrasting effects. First, by Corollary 2, a less spiteful agent induces P to increase $q^*(\theta)$, which increases rents and total costs and, hence, $t^*(\theta)$. Second, a higher $q^*(\theta)$ rises total surplus and this tends to increase $t^*(\theta)$. Third, a less spiteful agent attaches, ceteris paribus, a higher weight to total surplus relative to the transfer, and this tends to decrease $t^*(\theta)$. If either the principal or the agent is relatively selfish, the last effect dominates and the transfer decreases in ϕ_A. By contrast, when both players are relatively spiteful, the first two effects prevail and the transfer increases in ϕ_A.

Remark 2. When one of the players is altruistic while the other is spiteful, the qualitative insights of our results remain the same. Noteworthy, the comparative statics of the optimal transfer with respect to ϕ_A and ϕ_P depends on whether the principal or the agent is the altruistic player. Specifically, it can be shown that, if $\phi_P > 0$ and $\phi_A < 0$, the optimal transfer is increasing in ϕ_P and decreasing in ϕ_A. By contrast, when $\phi_P < 0$ and $\phi_A > 0$, again there is a sort of paradox of gift when the principal becomes less spiteful, while comparative statics with respect to ϕ_A has the same qualitative features as those in Lemma 2.

6. Conclusions

This paper contributes to the behavioral contracting literature by analyzing the effects of altruism and reciprocity on the design of optimal contracts in a simple principal–agent relationship with adverse selection. We have considered the effects both of the intrinsic attitude of players towards opponents, which is an innate characteristic, and of their global attitude, which also takes into account their willingness to reciprocate the opponents' behavior. Although global altruistic behavior allows to sustain more efficient outcomes than with selfish players, the (marginal) effect of an increase in players' intrinsic altruistic and reciprocal attitudes has ambiguous effects on efficiency. In particular, we have shown that a more reciprocal behavior improves efficiency if and only if the agent is intrinsically more altruistic than the principal. Similarly, dealing with an intrinsically more altruistic agent does not necessarily improves efficiency, and it actually leads to a higher distortion when both players feature a low reciprocal attitude.

By contrast, the presence of a globally spiteful principal always reduces efficiency compared to the case of selfish players, regardless of whether the agent is globally spiteful or not. In fact, when the principal assigns a negative weight to the agent's direct utility, the principal always increases the output distortion to reduce the agent's rent. As expected, efficiency increases when the players become less spiteful. By contrast, changes in the degree of reciprocity generate efficiency gains only under specific conditions on the agent's degree of intrinsic spitefulness.

The comparative statics on the optimal transfer also offers an interesting result. Contrary to what may be expected, a more altruistic principal may manage to induce the agent to produce a higher quantity (thus increasing total surplus) and, at the same time, obtain a lower transfer. In our model, this “paradox of gift” arises when the agent is sufficiently altruistic and inefficient.

These findings offer new testable implications on the link between optimal contracting, efficiency and behavioral concerns under asymmetric information: our model predicts that, in environments with good social relationships between players, the inefficiency due to asymmetric information is less severe, while it becomes more severe in environments with bad social relationships where players are likely to be spiteful. The relationship between efficiency and behavioral concerns could be tested through laboratory experiments. For example, subjects may first play an ultimatum game, in order to evaluate their degree of altruism and reciprocity, and then participate in a contracting experiments where they are asked to choose among a menu of contracts, that differ in terms of how they trade-off efficiency and rent extraction.

Acknowledgments

We would like to thank the associate editor and an anonymous referee for extremely helpful comments and suggestions.

Appendix A

Proof of Proposition 1. The incentive compatibility constraint (2) implies the following first- and second-order local incentive constraints

$$\left.\frac{\partial v_A(\theta,\hat\theta)}{\partial\hat\theta}\right|_{\hat\theta=\theta}=0 \Leftrightarrow (1-\phi_A)\dot t(\theta)-\theta\dot q(\theta)+\phi_A S'(q(\theta))\dot q(\theta)=0 \quad \forall\theta\in\Theta, \tag{A.1}$$

$$\left.\frac{\partial^2 v_A(\theta,\hat\theta)}{\partial\hat\theta^2}\right|_{\hat\theta=\theta}\le 0 \Leftrightarrow -\dot q(\theta)\ge 0 \quad \forall\theta\in\Theta, \tag{A.2}$$

which yield the envelope condition

$$\dot v_A(\theta)=-q(\theta) \quad \forall\theta\in\Theta. \tag{A.3}$$

Using the definition of $v_A(\theta)$ to solve for $t(\theta)$—i.e., Eq. (3)—P's objective function is

$$\int_\theta [(S(q(\theta))-\theta q(\theta))(1-\phi_A\phi_P)-(1-\phi_P)v_A(\theta)]\,dF(\theta). \tag{A.4}$$

Integrating (A.3) yields the standard expression for the agent's rent:

$$v_A(\theta)=v_A(\overline\theta)+\int_\theta^{\overline\theta} q(x)\,dx. \tag{A.5}$$

To solve P's program, we first ignore (A.2) and then check that it is satisfied in the solution obtained. Hence, substituting (A.5) into (A.4) and integrating by parts, P's (relaxed) optimization program is

$$\max_{q(\theta)}\int_\theta\left[S(q(\theta))-\theta q(\theta)-\frac{1-\phi_P}{1-\phi_A\phi_P}\frac{F(\theta)}{f(\theta)}q(\theta)\right]dF(\theta). \tag{A.6}$$

Since this objective function is strictly concave under our assumptions, the first-order condition (4) that defines $q^*(\theta)$ is also sufficient for an internal optimum. Moreover, by the implicit function theorem,

$$\dot q^*(\theta)=\frac{1+\dfrac{1-\phi_P}{1-\phi_A\phi_P}\dfrac{\partial}{\partial\theta}\dfrac{F(\theta)}{f(\theta)}}{S''(q^*(\theta))}.$$

This is negative since $S''(\cdot)<0$ and $F(\theta)/f(\theta)$ is increasing in θ—i.e., $q^*(\theta)$ satisfies (A.2).

Finally, to check that the global incentive compatibility constraint is satisfied, notice that

$$\begin{aligned}
&v_A(\theta)\ge v_A(\theta,\theta') \quad \forall(\theta,\theta')\in\Theta^2\\
&\quad\Rightarrow t^*(\theta)-\theta q^*(\theta)+\phi_A(S(q^*(\theta))-t^*(\theta))\ge t^*(\theta')-\theta q^*(\theta')+\phi_A(S(q^*(\theta'))-t^*(\theta'))\\
&\quad\Leftrightarrow (1-\phi_A)\int_{\theta'}^{\theta}\dot t^*(x)\,dx-\theta\int_{\theta'}^{\theta}\dot q^*(x)\,dx+\phi_A\int_{\theta'}^{\theta}S'(q^*(x))\dot q^*(x)\,dx\ge 0.
\end{aligned}$$

Using the first-order incentive compatibility constraint (A.1), this yields

$$\begin{aligned}
&\int_{\theta'}^{\theta}[x\dot q^*(x)-\phi_A S'(q^*(x))\dot q^*(x)-\theta\dot q^*(x)+\phi_A S'(q^*(x))\dot q^*(x)]\,dx\ge 0\\
&\quad\Leftrightarrow \int_{\theta'}^{\theta}\dot q^*(x)[x-\theta]\,dx\ge 0,
\end{aligned}$$

which is always satisfied since $\dot q^*(x)<0$.

Notice that

$$q^*(\theta)>q^{SB}(\theta)\Leftrightarrow S'\left(q^*(\theta)\right)<S'\left(q^{SB}(\theta)\right)$$

$$\Leftrightarrow \frac{1-\phi_P}{1-\phi_A\phi_P} < 1 \Leftrightarrow \phi_P(1-\phi_A) > 0,$$

which is satisfied since $\phi_A < 1$. □

Proof of Proposition 2. Let the adjusted distortion be

$$\Delta \equiv \frac{1-\phi_P}{1-\phi_A\phi_P}\frac{F(\theta)}{f(\theta)}.$$

The output $q^*(\theta)$ is decreasing in Δ for every θ, since $S(\cdot)$ is concave and $\phi_i \in (0,1)$, $i = A, P$.

First, consider the effect of a change in α_P on Δ. Differentiating and rearranging terms yields

$$\frac{\partial \Delta}{\partial \alpha_P} < 0 \Leftrightarrow \lambda(1-\alpha_P)^2 + \left[1-\lambda^3\alpha_A + \lambda(1-\lambda\alpha_P) + \lambda^2(1-\alpha_P)\right](1-\alpha_A) > 0. \tag{A.7}$$

This inequality is always satisfied. Therefore, $\partial q^*(\theta)/\partial \alpha_P > 0$.

Second, consider the effect of a change in λ on Δ. Differentiating and rearranging terms yields

$$\frac{\partial \Delta}{\partial \lambda} < 0 \Leftrightarrow (\alpha_A - \alpha_P)\underbrace{[(1+\lambda^2\alpha_A)(1-\alpha_A) + (\lambda^2+\alpha_P)(1-\alpha_P) + 2\lambda(1-\alpha_A\alpha_P)]}_{>0} > 0. \tag{A.8}$$

Hence, $\partial q^*(\theta)/\partial \lambda > 0 \Leftrightarrow \alpha_A > \alpha_P$.

Finally, consider the effect of a change in α_A on Δ. Differentiating and rearranging terms yields

$$\frac{\partial \Delta}{\partial \alpha_A} < 0 \Leftrightarrow \left(\lambda^3 + \lambda(1-2\alpha_A) - \alpha_P\right)(1-\alpha_P) + \lambda^2(2-\alpha_P-\alpha_A(2-\alpha_A)) > 0. \tag{A.9}$$

The left-hand-side of this inequality is a strictly concave function of λ, it is strictly positive if $\lambda \to 1$ (when either $\alpha_A \neq 0$ or $\alpha_P \neq 0$), and it is strictly negative when $\lambda \to 0$. Hence, there exists a unique threshold $\lambda^* \in (0,1)$ such that $\partial q^*(\theta)/\partial \alpha_A < 0$ if $\lambda < \lambda^*$, and $\partial q^*(\theta)/\partial \alpha_A > 0$ if $\lambda > \lambda^*$. □

Proof of Proposition 3. Using a first-order Taylor approximation of $t^*(\theta)$ around $\theta \to \overline{\theta}$, the optimal transfer can be rewritten as

$$t^*(\theta) \simeq t^*(\overline{\theta}) + \dot{t}^*(\overline{\theta})(\overline{\theta}-\theta),$$

where $\dot{t}^*(\overline{\theta}) = \overline{\theta} - \phi_A S'(q^*(\overline{\theta}))$. Hence, for θ close enough to $\overline{\theta}$,

$$t^*(\theta) \simeq \frac{[\overline{\theta}q^*(\overline{\theta}) - \phi_A S(q^*(\overline{\theta}))] + [\overline{\theta} - \phi_A S'(q^*(\overline{\theta}))]\dot{q}^*(\overline{\theta})(\overline{\theta}-\theta)}{1-\phi_A}.$$

Taking the derivative of this expression with respect to ϕ_P and using the first-order condition (4), when $\theta \to \overline{\theta}$ we have

$$\frac{\partial t^*(\theta)}{\partial \phi_P} = \left[\frac{\overline{\theta}(1-\phi_A)}{\phi_A} - \frac{(1-\phi_P)}{(1-\phi_A\phi_P)f(\overline{\theta})}\right]\frac{\partial q^*(\overline{\theta})}{\partial \phi_P}.$$

Since $\partial q^*(\overline{\theta})/\partial \phi_P > 0$ by Corollary 1, it follows that in a neighborhood of $\overline{\theta}$

$$\frac{\partial t^*(\theta)}{\partial \phi_P} > 0 \Leftrightarrow \frac{\overline{\theta}(1-\phi_A)}{\phi_A} > \frac{(1-\phi_P)}{(1-\phi_A\phi_P)f(\overline{\theta})}.$$

Rearranging yields the result. □

Proof of Lemma 1. In the uniform-quadratic case, the optimal quantity is

$$q^*(\theta) = q^{FB}(\theta) - \frac{1-\phi_P}{1-\phi_A\phi_P}\theta,$$

where $q^{FB}(\theta) = a - \theta$. Hence, using Eq. (5) the optimal transfer is

$$t^*(\theta) = \frac{1}{1-\phi_A}\left[\theta\left(q^{FB}(\theta) - \frac{1-\phi_P}{1-\phi_A\phi_P}\theta\right) + \left(a - \frac{1}{2}\left(1 + \frac{1-\phi_P}{1-\phi_A\phi_P}\right)\right)(1-\theta)\right]$$
$$+ \frac{\phi_A}{1-\phi_A}\left(a - \frac{1}{2}\left(q^{FB}(\theta) - \frac{1-\phi_P}{1-\phi_A\phi_P}\theta\right)\right)\left(q^{FB}(\theta) - \frac{1-\phi_P}{1-\phi_A\phi_P}\theta\right).$$

First, notice that the sign of $t^*(\theta)$ is equal to the sign of

$$[(a^2-\theta^2)\phi_P\phi_A + (\theta^2(4-\phi_P) - 2a(\phi_P+a) + \phi_P)]\phi_P\phi_A^2 + (\phi_P^2 + (\theta^2+4a-3)\phi_P + (a^2-4\theta^2))\phi_A + (\theta^2+1)(2-\phi_P) - 2a,$$

which, for $\phi_A \in (0,1)$, is a strictly concave function of ϕ_A, is positive for $\phi_A \to 0$, and is negative for $\phi_A \to 1$. Hence, there exists a unique threshold $\phi_A^* \in (0,1)$ such that $t^*(\theta) > 0$ if $\phi_A < \phi_A^*$ and $t^*(\theta) < 0$ otherwise.

Next, differentiating $t^*(\theta)$ with respect to ϕ_P yields

$$\frac{\partial t^*(\theta)}{\partial \phi_P} > 0 \Leftrightarrow \theta \frac{\partial q^*(\theta)}{\partial \phi_P} + \int_\theta^1 \frac{\partial q^*(x)}{\partial \phi_P} dx - \phi_A S'\left(q^*(\theta)\right) \frac{\partial q^*(\theta)}{\partial \phi_P} > 0$$
$$\Leftrightarrow 2\theta^2 \phi_P \phi_A^2 + (\theta^2 \phi_P - 4\theta^2 - \phi_P)\phi_A + (\theta^2+1) > 0. \tag{A.10}$$

The left-hand side of condition (A.10) is a strictly decreasing function of ϕ_A, and is positive when $\phi_A \to 0$. Moreover:

- If $\theta < \frac{1}{\sqrt{3}}$, this function is always positive so that $\partial t^*(\theta)/\partial \phi_P > 0$.
- if $\theta > \frac{1}{\sqrt{3}}$, this function is negative when $\phi_A \to 1$. Therefore, in this case, there exists a unique threshold

$$\phi_A^{**} = \frac{4\theta^2 + \phi_P(1-\theta^2) - \sqrt{\phi_P^2(1-\theta^2)^2 + 16\theta^4(1-\phi_P)}}{4\theta^2 \phi_P} \in (0,1)$$

  such that $\partial t^*(\theta)/\partial \phi_P > 0$ when $\phi_A < \phi_A^{**}$, and $\partial t^*(\theta)/\partial \phi_P < 0$ otherwise.

Finally, differentiating $t^*(\theta)$ with respect to ϕ_A yields

$$\frac{\partial t^*(\theta)}{\partial \phi_A} > 0 \Leftrightarrow M\phi_A^2 - (R+M)\phi_A - (N-R) > 0, \tag{A.11}$$

where

$$R \equiv \theta \frac{\partial q^*(\theta)}{\partial \phi_A} + \int_\theta^1 \frac{\partial q^*(x)}{\partial \phi_A}\, dx < 0,$$
$$M \equiv S'\left(q^*(\theta)\right) \frac{\partial q^*(\theta)}{\partial \phi_A} < 0,$$
$$N \equiv S(q^*(\theta)) - \theta q^*(\theta) - \int_\theta^1 q^*(x)\, dx.$$

It can be shown that the sign of $\partial t^*(\theta)/\partial \phi_A$ is equal to the sign of

$$[((a-1)^2\phi_P + 2\theta^2(1-\phi_P))\phi_P\phi_A + (1-3\theta^2)\phi_P^2]\phi_P\phi_A^2 + [(3a(2-a)+7\theta^2-4)\phi_P - 4\theta^2]\phi_P\phi_A^2 + a(2-a) + 2(\theta^2-1)$$
$$+((\theta^2(\phi_P+2)-2)\phi_P + (3(a(a-2)-\theta^2)+5))\phi_P\phi_A + (2\phi_P-5)\theta^2\phi_P^2, \tag{A.12}$$

which is strictly increasing in ϕ_A and negative when $\phi_A \to 1$. Hence, $\partial t^*(\theta)/\partial \phi_A < 0$. □

Proof of Proposition 4. First, consider the effect of changes in α_P and α_A on the adjusted distortion

$$\Delta \equiv \frac{1-\phi_P}{1-\phi_A\phi_P} \frac{F(\theta)}{f(\theta)}.$$

Since inequalities (A.7) and (A.9) hold even when $\alpha_P < 0$ and $\alpha_A < 0$, $\partial q^*(\theta)/\partial \alpha_P > 0$ and $\partial q^*(\theta)/\partial \alpha_A > 0$.

Second, consider the effect of a change in λ on Δ. Differentiating and rearranging terms yields

$$\frac{\partial \Delta}{\partial \lambda} < 0 \Leftrightarrow (\alpha_A - \alpha_P)\left[\underbrace{2\lambda(1-\alpha_A\alpha_P) + \lambda^2(1-\alpha_A^2) + (1-\alpha_P)^2}_{>0} - (\alpha_A-\alpha_P)\underbrace{(1-\lambda^2)}_{>0}\right] > 0. \tag{A.13}$$

When $\alpha_A < \alpha_P$, condition (A.13) is never satisfied (since $\alpha_P < 0$ and $\alpha_A < 0$) and, hence, $\partial q^*(\theta)/\partial \lambda < 0$. When, $\alpha_A > \alpha_P$, condition (A.13) yields

$$\frac{\partial \Delta}{\partial \lambda} < 0 \Leftrightarrow -\left(\alpha_A - \alpha_P + 1 - \alpha_A^2\right)\lambda^2 - 2(1-\alpha_A\alpha_P)\lambda - (\alpha_P(1-\alpha_P) + 1 - \alpha_A) < 0. \tag{A.14}$$

The left-hand side of inequality (A.14) is a strictly concave function of λ (for $\lambda > 0$) and is strictly negative when $\lambda \to 1$. Letting $\underline{\alpha}_A \equiv 1 + \alpha_P - \alpha_P^2$, there are two possible cases:

1. If $\alpha_A < \underline{\alpha}_A$, the left-hand side of (A.14) is strictly negative for any $\lambda \in (0,1)$. Hence, $\partial q^*(\theta)/\partial \lambda > 0$.
2. If $\alpha_A > \underline{\alpha}_A$, the left-hand side of (A.14) is strictly positive when $\lambda \to 0$. Hence, in this case, there exists a unique threshold $\widehat{\lambda} \in (0,1)$ such that $\partial q^*(\theta)/\partial \lambda < 0$ if $\lambda < \widehat{\lambda}$, and $\partial q^*(\theta)/\partial \lambda > 0$ otherwise. □

Proof of Lemma 2. First, Proposition 1 implies that $t^*(\theta) > 0$ when $\phi_i < 0$, $i = A, P$. Second, $\partial t^*(\theta)/\partial \phi_P$ (which is derived in the proof of Lemma 1) is positive when $\phi_i < 0$, $i = A, P$. Notice that these two results hold for any specification of $S(q)$ and any distribution of θ.

Finally, the sign of $\partial t^*(\theta)/\partial \phi_A$ (which is derived in the proof of Lemma 1) is equal to the sign of (A.12), which is a strictly decreasing function of ϕ_A and is negative when $\phi_A \to 0$. Moreover, letting

$$\phi_P^* \equiv -1 + \frac{\sqrt{2\theta^2(2(a-\theta)^2+1)+1} - (1+\theta^2)}{a(a-2)} \in (-1,0),$$

we have that:

- If $\phi_P > \phi_P^*$, (A.12) is negative when $\phi_A \to -1$. Hence, in this case, $\partial t^*(\theta)/\partial \phi_A < 0$.
- If $\phi_P < \phi_P^*$, (A.12) is positive when $\phi_A \to -1$. Hence, there exists a unique threshold $\phi_A^{***} \in (-1,0)$ such that $\partial t^*(\theta)/\partial \phi_A > 0$ if $\phi_A < \phi_A^{***}$ and $\partial t^*(\theta)/\partial \phi_A < 0$ otherwise. □

References

Akerlof, G., Yellen, J., 1990. The fair wage-effort hypothesis and unemployment. Q. J. Econ. 105 (2), 255–283.
Baron, P.D., Myerson, R.D., 1982. Regulating a monopolist with unknown costs. Econometrica 50, 911–930.
Battigalli, P., Dufwenberg, M., 2009. Dynamic psychological games. J. Econ. Theory 144, 1–35.
Becker, G., 1976. Altruism, egoism, and genetic fitness: economics and sociobiology. J. Econ. Lit. 14, 817–826.
Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games Econ. Behav. 10 (1), 122–142.
Chone, P., Ma, C.A., 2004. Asymmetric information from physician agency: optimal payment and healthcare quantity. Ann. d'Economie et de Stat. 101–102, 229–256.
Dawes, R., Thaler, R., 1988. Anomalies: cooperation. J. Econ. Perspect. 2, 187–198.
Delfgaauw, J., Dur, R., 2008. Incentives and workers' motivation in the public sector. Econ. J. 118, 171–191.
Delfgaauw, J., Dur, R., 2009. From public monopsony to competitive market: more efficiency but higher prices. Oxf. Econ. Pap. 61, 586–602.
Dufwenberg, M., Kirchsteiger, G., 2000. Reciprocity and wage undercutting. Eur. Econ. Rev. 44, 1069–1078.
Dur, R., Tichem, J., 2012. Social Relations and Relational Incentives. Tinbergen Institute Discussion Paper.
Dur, R., 2009. Gift exchange in the workplace: money or attention? J. Eur. Econ. Assoc. 7, 550–560.
Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games Econ. Behav. 54, 293–315.
Fehr, E., Gachter, S., Kirchsteiger, G., 1997. Reciprocity as a contract enforcement device: experimental evidence. Econometrica 65, 833–860.
Fehr, E., List, J., 2004. The hidden costs and returns of incentives: trust and trustworthiness among CEOs. J. Eur. Econ. Assoc. 2 (5), 743–771.
Gill, D., Prowse, V., Vlassopoulos, M., 2012. Cheating in the Workplace: an experimental study of the impact of bonuses and productivity. IZA DP No. 6725.
Immordino, G., De Marco, G., 2013. Partnership, reciprocity and team design. Res. Econ. 67, 39–58.
Jack, W., 2004. Purchasing health care services from providers with unknown altruism. J. Health Econ. 24, 73–93.
Kahneman, D., Knetsch, J., Thaler, R., 1986. Fairness as a constraint on profit-seeking: entitlements in the market. Am. Econ. Rev. 76 (4), 728–741.
Laffont, J.J., Martimort, D., 2002. The Theory of Incentives: The Principal–Agent Model. Princeton University Press.
Ledyard, J., 1995. Public goods: a survey of experimental research. In: Kagel, J., Roth, A. (Eds.), Handbook of Experimental Economics, Princeton University Press.
Levine, D., 1998. Modeling altruism and spitefulness in experiments. Rev. Econ. Dyn. 1, 593–622.
Maskin, E., Tirole, J., 1990. The principal–agent relationship with an informed principal: the case of private values. Econometrica 58 (2), 379–409.
Maskin, E., Tirole, J., 1992. The principal–agent relationship with an informed principal, II: common values. Econometrica 60 (1), 1–42.
Netzer, N., Schmutzler, A., 2010. Rotten Kids with Bad Intentions. CEPR Working Paper No. 7667.
Rabin, M., 1993. Incorporating fairness into game theory and economics. Am. Econ. Rev. 83 (5), 1281–1302.
Segal, U., Sobel, J., 1999. Tit for tat: foundations of preferences for reciprocity in strategic settings. J. Econ. Theory 137 (1), 197–216.
Shchetinin, O., 2009. Contracting Under Reciprocal Altruism. MPRA Paper No. 15430.
Shchetinin, O., 2010. Altruism and Career Concerns. Mimeo.
Siciliani, L., 2009. Paying for performance and motivation crowding out. Econ. Lett. 103, 68–71.
Thaler, R., 1988. Anomalies: the ultimatum game. J. Econ. Perspect. 2, 195–207.