Social Exchange and Common Agency in Organizations

Robert Dur¹
Hein Roelfsema²

¹ Erasmus Universiteit Rotterdam, and Tinbergen Institute;
² Utrecht School of Economics, Utrecht.
Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at http://www.tinbergen.nl.
Social Exchange and Common Agency in Organizations
Robert Dur* and Hein Roelfsema†
December 15, 2006

Abstract

We study the relation between formal incentives and social exchange in organizations where employees work for several managers and reciprocate to a manager’s attention with higher effort. To this end we develop a common agency model with two-sided moral hazard. We show that when effort is contractible and attention is not, the first-best can be achieved through bonus pay for both managers and employees. When neither effort nor attention are contractible, an ‘attention race’ arises, as each manager tries to sway the employee’s effort his way. While this may result in too much social exchange, the attention race may also be a blessing because it alleviates managers’ moral-hazard problem in attention provision. Lastly, we derive the implications of these contract imperfections for optimal organizational design.

Keywords: social exchange, reciprocity, incentive contracts, common agency, organizational design.

JEL-codes: D86, J41, M50, M54, M55.

*Department of Economics, Erasmus University Rotterdam, Tinbergen Institute, and CESifo. Address: Erasmus University, FEW/H 8-15, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: dur@few.eur.nl
†Utrecht School of Economics, Utrecht University. E-mail: h.j.roelfsema@econ.uu.nl
1 Introduction

Managers in organizations often use common human resources. For example, they share secretary services, the IT unit, and the personnel department. People who provide these supporting roles typically face financial incentives, be they explicit or implicit. However, most of us have the experience that social interaction to establish a good ‘working relation’ really helps to get your computer fixed in time, to speed up formalities in last-minute hiring, or to crucially improve the lay-out of your document in Powerpoint. Although (obviously) we are nice people, niceness is sometimes also used strategically to trigger reciprocal feelings in others to get things done.

Such social exchange in organizations has been studied extensively in the organizational sociology and management literature, where worker reciprocity in response to managerial or organizational support is a common theme (see, e.g., Baron and Kreps 1999, chapter 5). The economic literature on manager-subordinate reciprocity has so far mainly focussed on how generous financial compensation is interpreted as ‘kindness’ by the employer, triggering effort and loyalty in the employment relation (Akerlof 1982, Fehr and Gächter 2000). Relatively little has been written in economics about social interactions between managers and workers in firms.

This paper develops an economic model of social exchange within firms. Our aim is to get more insight into the relation between workers’ formal employment contract (the wage, financial incentives) and social interaction between managers and workers at the workplace. Further, we investigate the consequences of social interaction and workers’ reciprocity for optimal organizational design.

We focus on a situation where multiple principals compete for the effort of an agent with reciprocal feelings. The game consists of three stages. In the first stage, the principals (or the superior of the principals) design a contract that specifies the relationship between each principal and the agent. In addition, in this first stage the principals decide on whether and, if so, with how many of them to share an agent. For example, we may think of professors (or the department chair) deciding on with how many to share one secretary, how effort by this secretary is rewarded financially, and possibly how they treat the secretary while on the job (buying flowers on her birthday). In the second stage, each principal decides on his ‘attention level’ for the agent, taking into account that in the third stage the agent responds positively to attention by providing more effort for him. In the third stage, the agent decides on her effort level for each of the principals, taking account of her contract and the attention provided by the principals.
In the second stage.

In the first-best – effort and attention are contractible – the marginal benefits from effort for each principal equal the marginal costs of effort for the agent minus a reciprocity discount. Further, when designing the contract, the principals trade off the costs of giving attention against the wage compensation that is necessary to attract or retain the agent. In the optimum, the marginal cost of giving attention of each principal equal the marginal utility that the agent derives from attention. Lastly, the optimal number of principals that share an agent equates the average costs of effort to the marginal costs of effort.

Certainly, effort and bilateral social interaction may not be contracted upon, for instance because enforcement costs are high (effort) or because the outcomes can not be verified by a third party (attention). We show that when attention by managers can not be contracted, but the agent’s effort is verifiable, the first-best outcome can still be achieved. With bonus pay for the agent, attention of each principal can be inferred from the effort of the agent for each of them. Hence, in the first stage of the game, the principals can contract to punish those among them for which effort is too low, implicitly punishing ‘bad management’. By contrast, when the agent’s effort is specified in a forcing contract that stipulates a fixed wage, in the second stage principals have no incentive to provide a positive level of attention. This moral-hazard problem is anticipated in the first stage by the agent: she expects little attention from the principals and therefore demands a high fixed wage. We thus identify a new benefit of granting flexibility to workers: it enables the organization to measure and reward managers’ performance.

A similar moral-hazard problem in attention provision exists when neither effort nor attention are contractible. However, due to an externality problem, we show that an ‘attention race’ among the principals arises, which mitigates or even reverses the underprovision of attention. The reason for the attention race is that in a non-cooperative setting each principal has an incentive to sway the agent’s effort his way. This incentive is stronger for a larger number of principals and the more reciprocal the worker is.

Whether the attention race on balance is good or bad for the firm depends on the preferences of the agent for attention. If the worker cares much for attention, then the attention race is beneficial for the firm, for it creates a positive working environment that is much appreciated by the agent. Without the attention race, the principals’ inability to commit to attention would be very costly. The reverse happens when the worker puts a low value on attention. In that case the attention race is costly to the firm, for it has
little effect on the participation constraint of the worker. Consequently, the firm may want to limit the number of principals that share an agent.

The remainder of the paper is structured as follows. The next section provides some further empirical motivation and briefly discusses related literature. Section 3 presents the model. Section 4, 5, and 6 examine the first-best, the case of noncontractible attention, and the case of noncontractible attention and effort, respectively. Section 7 concludes.

2 Empirical motivation and related literature

While the economic analysis of workers’ reciprocity has so far been mainly confined to the role of the formal employment relation (the wage contract) in triggering reciprocity and loyalty, in other fields there is scepticism towards such overemphasis on pay as the motivator (see e.g. Pfeffer 1998). The observation that social exchange provides non-financial incentives within organizations has for long been recognized in the management literature and in organizational sociology. Following a seminal paper in this field by Gouldner (1960) – who argues that reciprocity is needed for the stability of social systems, including the firm –, Blau (1964) distinguishes two types of exchange in organizations: economic and social. Economic exchange typically is defined as formal ‘in-role’ behavior (the wage and contracted effort in our model). Social exchange includes various ‘extra-role’ activities such as giving attention in our model or providing non-specified effort for a particular manager. According to Blau (1964, p. 94) “only social exchange tends to engender feelings of personal obligations, gratitude, and trust; purely economic exchange as such does not”.

Empirical research provides substantial support for a performance enhancing effect of social exchange between managers and workers. For example, using data from surveys among managers and workers, many studies have found that a higher quality of so-called ‘leader-member exchange relationships’ (LMX) and ‘perceived supervisor support’ (PSS) are associated

---

span style="color: #7f8c8d; font-size: 8pt;"">1Reciprocity is only recently gaining prominence in organizational and managerial economics. For example, in Prendergast’s (1999) review of incentive provision in firms, the word reciprocity is not mentioned once. In a traditional economic model, a higher wage only induces more effort if the wage is, in some or another way, linked to the worker’s effort. By contrast, as argued by Akerlof (1982), when workers are reciprocal a more generous wage induces additional efforts, for a worker increases his utility by reciprocating such a gift. While many laboratory studies provide support for this gift-exchange relation (e.g. Fehr and Gächter 2000), in a recent field experiment Gneezy and List (2006) find only a temporary effect of high wages on effort.

---
with better performance of the worker, both in required duties as well as in those beyond the formal employment contract (see e.g. Settoon et al. 1996, Wayne et al. 1997, Uhl-Bien and Maslyn 2003, Dabos and Rousseau 2004, and Shanock and Eisenberger 2006).smith
2 Nagin et al. (2002) study cheating in call centers and find that employees who feel that they are treated well by their employer cheat less and are less responsive to changes in monitoring by the employer. Their data also show that many employees (70% in their sample) have the impression that the company cares about their personal well-being. In line with this, US survey evidence indicates that many managers are willing to listen to personal problems of their subordinates.3

Social exchange has also been shown to affect workers’ wage compensation. A famous example is given by Gittell (2003) who evaluates pay conditions at Southwest Airlines and finds that the positive corporate culture and strong relational contracts make employees willing to accept a lower wage than their industry counterparts. A number of examples from other industries are provided in Pfeffer (1998).

Our model incorporates these findings on how social exchange matters for productivity and wage cost in the following two ways. First, social exchange relaxes the participation constraint of the workers: attention by their superiors makes workers feel good, which makes them willing to work for a lower wage. Second, we model reciprocity by assuming that social interaction with a superior reduces the costs of effort for that superior at the margin, so that attention induces the employee to work harder for a given wage schedule.4

Our paper builds on the common agency literature with moral hazard,

---

2 A closely related and partially overlapping body of research examines the effects of ‘perceived organizational support’ (POS), that is, an employee’s belief about how much the organization as a whole provides support to or cares about the worker. Generally, empirical studies find strong effects of POS on commitment to the organization (loyalty, turnover), but – in contrast to LMX and PSS – only weak effects on performance (see Pazy and Ganzach 2006 and the references therein).

3 34% of workers state that their boss is “very much” willing to listen to their personal problems. Only 11% report that their boss is not willing to listen at all. See the Wisconsin Longitudinal Study, 2002-2003 Wave, http://www.ssc.wisc.edu/wlsresearch/.

4 We thus assume that reciprocating provides immediate utility for the worker. Such emotional and expressive reactions that stem from the evolution of the human brain are often stressed in psychology (see Frank 1988) where, in general, reciprocity is not considered a cognitive process. Instead, by configuration, people can not help responding to the impulse of reciprocating a gift. To illustrate this, a famous (and true) story of how reciprocal emotions can be used is that of car salesman Joe Gerard, known as “the greatest car salesman”. His secret was to write each of his 13,000 former customers a card every month with only the words “I like you” on it.
initiated by Bernheim and Whinston (1985, 1986). The ‘attention race’ among principals that we identify in Section 6 echoes Bernheim and Whinston’s (1986) result that inefficiencies may arise from coordination problems between principals when coordination among principals would not achieve the first-best. Our results on the optimal number of principals that share an agent are close in spirit to Bernheim and Whinston (1985). They analyze manufacturer’s incentives to share a marketing agent, which serves as a facilitating device for product market collusion among manufacturers (see also Gal-Or 1991 and Martimort 1996, among others). An important difference between our set-up and the standard common-agency model is that principals in our model have two instruments at their disposal to motivate the agent: contracts and attention. Another distinguishing feature of our analysis is that we allow for two-sided moral hazard in the relation between the agent and the principals: after contracts have been signed, both the agent and the principals take actions that are imperfectly contractible. One of the results emanating from this is that the attention race among principals may be a blessing in disguise, for it may resolve a moral hazard problem on the side of the principals. Lastly, we depart from the existing literature on common agency by focusing on social exchange and reciprocity in organizations.

3 The model

Our model revolves around homogenous principals, indexed by $i$, whose payoffs depend on the services provided for them by one or more homogenous agents. Agents are available in unlimited supply, but must be compensated for foregoing their outside option ($U > 0$) and for the (net) disutility from working. A principal can hire his own agent, but he can also decide to share an agent with other principals (that is, to hire a common agent together with other principals). Since principals are homogenous, they have identical preferences over the number of principals to share an agent with; this number

\footnote{There is also a large literature on common agency with adverse selection, see Laffont and Martimort (1997) and Martimort (2006) for surveys.}

\footnote{See Dixit (1997) for an interesting analysis of common agency problems and incentive provision in public organizations. Tirole (1988) and Holmstrom (1999) also discuss common agency problems within organizations.}

\footnote{None of the results change when a superior of the principals makes these decisions, provided that the superior’s and joint principals’ interests are the same.
is denoted by \( n \). Each group of \( n \) principals collectively offers a contract to an agent which specifies the agent’s wage compensation, \( w \), as well as the way in which wage costs are shared among the principals. If the agent’s effort and/or the principals’ attention are contractible, the contract can also contain other provisions, e.g. bonus pay to the agent depending on her effort or discounts on a principal’s contribution to the agent’s wage costs depending on the attention provided. After signing the contract, each principal independently chooses the level of attention he gives to the agent, which we denote by \( a_i \). Subsequently, the agent chooses the level of effort she exerts for each of the principals involved in the contract; effort for principal \( i \) is denoted by \( e_i \).

The payoff of principal \( i \) is described by

\[
\pi_i = Q(e_i) - H(a_i) - s_i, 
\]

where \( Q(e_i) \) is a strictly concave and increasing function reflecting the value for the principal of the agent’s effort for him, \( H(a_i) \) is a strictly convex and increasing function representing the principal’s costs of giving attention, and \( s_i \) is the contribution of principal \( i \) to the agent’s wage compensation, \( w \). The function \( Q(e_i) \) satisfies the Inada conditions and the function \( H(a_i) \) satisfies \( H(0) = H_a(0) = 0 \), where (throughout the paper) subscripts of functions denote partial derivatives. The budget constraint for the set of \( n \) principals is \( \int_0^n s_i = w \).

The agent’s payoff \( U \) is:

\[
U = w - C(\int_0^n e_i) + \int_0^n F(e_i, a_i). 
\]

Since the wage enters linearly in the payoff function, the agent is risk-neutral in income. The function \( C \) represents the agent’s cost of effort and is strictly convex and increasing in the total effort exerted for all principals. It satisfies \( C(0) = 0 \) and \( C_e(0) = 0 \). The function \( F(a_i, e_i) \) describes both the value of attention to the agent and her reciprocity. It has properties \( F_{e_i} > 0 \) when \( a_i > 0 \) (the agent is reciprocal), \( F_{e_i a_i} > 0 \) (the willingness to reciprocate effort to principal \( i \) increases with the attention given by principal \( i \)), and

---

8 We will abstract from the problem that may be relevant in relatively small organizations that not all of the existing principals may find sufficiently many other principals to optimally share an agent with. Also, for ease of exposition and without significant loss of generality, when determining the optimal level of \( n \), we will treat \( n \) as a continuous variable.

9 It should be noted that, since a principal’s ‘output’ \( Q \) depends only on the agent’s effort for him, contractibility of the principal’s output and contractibility of the agent’s effort yield identical results.
$F_{ai} > 0$ (the agent enjoys attention). Further, we assume that $F_{aa} < 0$ and $F_{ee} < 0$, so as to rule out corner solutions where infinite or zero attention is optimal or where the agent exerts effort for only one of the principals. Throughout the paper, the third derivatives of all functions are assumed to be negligibly small. The participation constraint of the agent is:

$$w - C(\int_0^1 e_i) + \int_0^n F(e_i, a_i) \geq U. \quad (3)$$

The timing of the game is as follows. In stage one, the principals cooperatively decide on the number of principals that shares one agent and on the content of the contract. The agent decides whether to accept or reject the contract. Once the organizational structure and the contract are in place, in stage two each principal independently chooses the level of attention he gives to the agent. In stage 3 the agent chooses her effort level for each principal.

4 Complete contracts

We start by considering the first-best case where both effort of the agent and attention of the principals are contractible. The first-best contract stipulates the levels of effort, attention, the number of principals that share an agent, the agent’s wage, and the way wage costs are shared by the principals. Our results are summarized in the following proposition.

**Proposition 1** In the first-best:

1. the marginal benefit from effort for each principal equals the marginal cost of effort for the agent minus a reciprocity discount: $Q_{e_i}(e_i) = C_e(\int_0^n e_i) - F_e(e_i, a_i)$,

2. the marginal cost of attention for each principal equals the marginal benefit of attention for the agent: $H_a(a_i) = F_a(e_i, a_i)$,

3. the optimal number of principals that share one agent equates the average cost of effort to the marginal cost of effort: $[\overline{U} + C(\int_0^n e_i)] / ne_i = C_e(\int_0^n e_i)$.

The proof is given in the appendix. Clearly, from the first part of the proposition it follows that if the optimal level of attention is positive, then for a given number of principals the first-best contract specifies an effort level $e_i$ that is higher than the conventional level which ignores reciprocity (i.e.,
ignores the term $F_e > 0$). The intuition is that the principals’ attention reduces the agent’s marginal cost of effort because she feels the need to reciprocate. This reduction in marginal cost raises the optimal level of effort the agent exerts for each principal.

The second part of the proposition describes the optimal levels of attention. This involves a trade-off between the costs of giving attention for the principals and the benefit from attention for the agent, where the optimal level equates the marginal cost and marginal benefit. Obviously, in our setup the benefits from attention ultimately accrue to the principals, as the agent’s wage compensation is such that her participation constraint binds.

The third part of the proposition shows that the optimal number of principals that share one agent equates the agent’s average cost of effort to the agent’s marginal cost of effort. This makes sense: sharing an agent with more principals reduces each principal’s contribution to the fixed cost of the agent (the agent’s compensation for foregoing her outside option utility $\mathcal{U}$), but increases the agent’s compensation for effort costs because these are convex. As usual, the lowest cost per unit of effort is attained when marginal cost equal average cost.

Note that the agent’s reciprocity has no direct effect on the optimal number of principals. This follows from the separability of each principal’s attention in the agent’s utility function, implying that the rents from giving attention are constant per principal. However, the optimal number of principals sharing an agent is indirectly affected by reciprocity through its effect on the optimal level of effort. This is described in the following corollary.

**Corollary 1** In the first-best, the agent’s reciprocity makes it optimal for the principals to commit to give attention to the agent, which increases optimal effort for each principal and, therefore, reduces the optimal number of principals that share one agent.

The proof is in the appendix. The intuition is as follows. As we have seen, when the agent is reciprocal, principals optimally give attention, which raises the agent’s effort for each of them. Consequently, the marginal cost of effort ($C_e$) increase and so sharing an agent with more principals becomes more costly. The resulting decline in the number of principals further increases the optimal effort the agent exerts for each principal, as can been seen from the first part of Proposition 1.

---

10 An alternative assumption would be that the agent’s utility from a principal’s attention decreases with total attention received from all principals, i.e. $F_{a_i} \int_0^{a_i} a_i < 0$. Clearly, this modification would reduce the optimal number of principals who share one agent.
5 Incomplete contracts: unverifiable attention

Consider next a setting where the agent’s effort is verifiable, but attention by the principals is not. As we will see, the first-best can be achieved even though attention cannot be verified. This is summarized in the following proposition.

**Proposition 2** If effort is contractible and attention is not contractible, then the first-best can be achieved by offering an incentive contract that pays the agent the full marginal product of her effort for each principal and that conditions each principal’s contribution to the agent’s wage on the effort provided for him.

We prove the proposition by backward induction, and so start with the agent’s effort choice in stage 3 of the game. Let the agent’s wage consist of a base salary, \( w \), and a set of \( n \) bonuses with each bonus \( b(e_i) \) depending on the effort for one of the principals involved in the contract. The first-order conditions for the agent’s optimal efforts then are:

\[
b_e(e_i) - C_e(\int_0^a e_i) + F_e(e_i, a_i) = 0 \text{ for all } i. \tag{4}
\]

Comparing to the first-best level of effort as described in Proposition 1, it follows that if the agent is shared by the first-best number of principals and principals have given the first-best level of attention in the second stage of the game, then \( b_e(e_i) = Q_e(e_i) \) for all \( e_i \) ensures that the agent provides first-best effort for all principals. So, we have the standard result that, when both principals and agents are risk-neutral in income, optimal bonus pay equals the full marginal product.

Next consider the principals’ choice of attention in the second stage of the game. Let each principal’s contribution to the agent’s wage costs depend on the effort provided for him and take the following form: \( s(e_i) = b(e_i) + t(e_i) \). Thus, each principal covers the agent’s bonus pay arising from the agent’s effort for him and makes an additional contribution to or receives a discount from the organization, \( t(e_i) \), which also depends on the agent’s effort for him. The first-order condition for principal \( i \)’s optimal attention then is:

\[
\frac{de_i}{da_i} [Q_e(e_i) - b_e(e_i) - t_e(e_i)] - H_a(a_i) = 0, \tag{5}
\]
where, by totally differentiating (4), the effect of attention on effort is given by: \[ \frac{de_i}{da_i} = \frac{-F_{ee} [b_{ee} + F_{ee} - (n - 1) C_{ee}]}{(b_{ee} + F_{ee})^2 - nC_{ee} (b_{ee} + F_{ee})} > 0, \] (6)

where the sign follows from our assumptions about the functions \( F \) and \( C \) as described in Section 3 and from \( b_{ee} = Q_{ee} < 0 \), which is implied by the result that the agent receives the full marginal product of effort, as shown in the previous step of this proof. Equation (6) shows that when the agent is reciprocal, the agent exerts more effort for a principal when the principal has given more attention (that is, if \( F_{ea} > 0 \), then \( \frac{de_i}{da_i} > 0 \)).

This positive effect of attention on effort can give principals an incentive to provide attention, as described by first-order condition (5). Note that, since the agent receives her full marginal product \( (b_i(e_i) = Q_e(e_i)) \), the first two terms in square brackets in (5) cancel out. This bears an important message: paying a high bonus for effort to the agent implies weak incentives for principals to give attention. The intuition for this result is that a higher bonus means that more of the returns to effort translate into higher agent’s pay rather than into a higher payoff for the principal. Hence, principals have less to gain from increasing the agent’s effort by giving attention.

In order to induce the first-best level of attention, the principals need to be incentivized through the other part of the contribution scheme, \( t(e_i) \). Comparing (5) to the condition for first-best attention as described in Proposition 1, and denoting the first-best levels of attention and effort by \( a^* \) and \( e^* \) respectively, it follows that if

\[ t_e(e_i) = -\left( \frac{de_i}{da_i} \right)^{-1} F_a(a^*, e^*) \]

for all \( e_i \), then each principal has an incentive to provide the first-best level of attention.\(^{12}\) Since \( \frac{de_i}{da_i} > 0 \) and \( F_a > 0 \), the optimal \( t_e \) is negative and so principals receive a discount on their contribution to wage costs, which increases with the effort provided for them. This way, principals fully internalize the positive effects of their attention on the agent’s payoff.\(^{13}\)

\(^{11}\)Note that (4) represents \( n \) first-order conditions describing optimal effort for all \( n \) principals. Expression (6) follows from differentiating all these first-order conditions to \( a_i \), \( e_i \), and all \( e_j \neq i \), and then combining them so as to end up with an expression for \( \frac{de_i}{da_i} \). Since the third derivatives of all functions are assumed to be negligibly small, we have suppressed the arguments of all functions in (6).

\(^{12}\)Note that \( \frac{de_i}{da_i} \) is independent of effort and attention by our assumption about the third derivatives.

\(^{13}\)First-best attention can also be achieved by a forcing contract for the principals which punishes those principals for whom effort deviates from the first-best. An advantage of
Lastly, since first-best effort and first-best attention can be sustained, the optimal number of principals that share an agent is also equal to its first-best level. The base salary of the agent is such that her participation constraint binds:

$$\bar{w} = \bar{U} + C(\int_0^n e^* - \int_0^n F(e^*, a^*) - nb(e^*),$$

and the lump-sum part of the principals’ contributions must be such that their budget constraint binds:

$$ns(e^*) = nb(e^*) + nt(e^*) = \bar{w} + nb(e^*) \Rightarrow nt(e^*) = \bar{w}.$$  

Summarizing, the noncontractibility of attention does neither affect the players’ payoffs nor organizational design. However, it has important consequences for the optimal design of the contract. While in the previous section a simple forcing contract stipulating the agent’s effort and the principals’ attention was sufficient to attain the first-best, with noncontractible attention incentives need to be fine-tuned. First, the agent must be offered flexibility in effort choice rather than face binding effort constraints. Suppose the agent’s contract would dictate her effort level, as in the previous section. Then, principals would have no incentive to give attention, since the agent’s effort would be unresponsive to social exchange.\(^{14}\) Anticipating this lack of attention, the agent would then demand a higher wage for the same (or a lower) level of effort. Second, to avoid underprovision of effort, the organization can not rely on social exchange alone. To attain first-best effort, the agent must be offered incentive pay equal to the full marginal product of her effort. Third, principals must be given financial incentives to give attention. Even though attention can not be directly rewarded, principals can be induced to give sufficient attention by making their contribution to the wage costs dependent on the agent’s effort for him. This way, ‘bad management’ by principals is implicitly punished and ‘good management’ is implicitly rewarded. Last, we have found that higher performance pay for the agent strengthens the agent’s incentives to exert effort but weakens the principals’ incentives to give attention, because principals have less to gain from higher effort when agents keep more of the fruits of their effort. To the incentive contract studied in the main text is that it better insulates principals from trembling hands of their colleagues, the agent, and themselves.\(^{14}\)A related argument is given by Bernheim and Whinston (1998) in the context of a dynamic model without social exchange. They show that, when an employee’s effort is not fully contractible, the employer has an incentive to leave some of his (potentially contractible) obligations ambiguous, so as to strengthen repeated game incentives.

---

\(^{14}\) A related argument is given by Bernheim and Whinston (1998) in the context of a dynamic model without social exchange. They show that, when an employee’s effort is not fully contractible, the employer has an incentive to leave some of his (potentially contractible) obligations ambiguous, so as to strengthen repeated game incentives.
keep principals motivated to provide attention to the agent, they must be sufficiently rewarded by the organization (the group of principals or their superior) when the agent works hard for them.

6 Incomplete contracts: unverifiable effort and attention

When neither effort nor attention are contractible, the contract only stipulates the number of principals the agent will work for and her base salary, which is equally shared by the principals. We solve the game by backward induction.

In the third stage, the agent chooses effort for each principal to maximize her payoff (2). The first-order conditions are:

$$-C_e(\int_0^n e_i) + F_e(e_i, a_i) = 0 \text{ for all } i. \quad (7)$$

Notice that, in general, we cannot say whether effort is higher or lower than in the first-best case. While the agent has neither a contractual obligation (Section 4) nor a financial incentive (Section 5) to provide effort, the principals may give more attention or share the agent with fewer principals than in the first-best as substitutes for the lack of formal incentives. For later use, we derive some comparative static results for the agent’s choice of effort, which are summarized in the following lemma.

Lemma 1 If neither effort nor attention are contractible, then:

1. for a given level of attention by each principal, effort for each principal decreases with the number of principals that share an agent (holding \(a_i\) constant, \(de_i/dn < 0\));

2. for a given number of principals that share an agent, attention by principal \(i\) increases the agent’s effort for principal \(i\), decreases the agent’s effort for all other principals, and increases the agent’s total effort (holding \(n\) constant, \(de_i/da_i > 0\), \(de_{j\neq i}/da_i < 0\), and \(d(\int_0^n e_i/da_i) > 0\));

3. the effect of principal \(i\)’s attention on the agent’s effort for him increases with the number of principals that share an agent (\(d(de_i/da_i)/dn > 0\)).

The proof is in the appendix, the intuition follows. First, for a given level of attention, working for more principals reduces effort for each principal, because the accompanying increase in total effort raises the agent’s
marginal cost of effort for each principal. Second, when a principal gives more attention to the agent, the agent feels a need to reciprocate and so provides more effort for him. As this raises her marginal cost of effort, she provides less effort for the other principals. Hence, attention by a principal imposes a negative externality on other principals. Third, the effectiveness at the margin of a principal’s attention in raising the agent’s effort for him increases with the number of principals. The intuition is as follows. When a principal gives more attention, the agent reciprocates by increasing effort for that principal, which comes partly at the cost of increasing total effort, and partly at the cost of reducing effort for the other principals. When the agent works for more principals, the latter cost is lower, because a given reduction in effort for the other principals can be spread over a larger number of principals. This implies that the agent is more responsive to a principal’s attention as the number of principals increases.

In the second stage principals independently choose their level of attention. The first-order condition for optimal attention of principal $i$ is:

$$\frac{de_i}{da_i} Q_e(e_i) - H_a(a_i) = 0.$$ (8)

Each principal trades off the cost of giving attention and the benefit that the agent exerts more effort for him. Note the difference with attention in the first-best case described by Proposition 1: when effort and attention are noncontractible, the provision of attention depends on the responsiveness of the agent to attention and not on how valuable attention is to the agent. As we will see, this may result in too high or too low attention compared to the first-best. Note also that each principal disregards the effect of his attention on the efforts provided by the agent for the other principals. As Lemma 1 showed, if principal $i$ gives more attention, the agent responds by increasing effort for principal $i$ and decreasing effort for the other $(n - 1)$ principals. This, in turn, induces these other principals to give more attention to the agent as well, because the decrease in effort makes additional efforts more valuable. In equilibrium, the negative externality that principals impose on one another makes that each principal’s attention is above the level that is in their joint interest during the second stage of the game.\footnote{The principals’ joint surplus maximizing level of attention in the second stage is described by:

$$\frac{de_i}{da_i} Q_e(e_i) + (n - 1) \frac{de_j}{da_i} Q_e(e_{j\neq i}) - H_a(a_i) = 0 \text{ for all } i.$$}
The following lemma describes the effect on attention of the number of principals that share an agent.

**Lemma 2** If neither effort nor attention are contractible, then an increase in the number of principals that share an agent increases each principal’s attention \((da_i/dn > 0)\).

The proof is in the appendix. Increasing the number of principals has two effects working in the same direction. First, for a given level of attention, the agent exerts less effort for each principal when the number of principals increases. Since the marginal product of effort is decreasing in effort, additional effort becomes more valuable to each principal. Hence, each principal increases attention so as to increase effort for him. Second, when the number of principals increases, the agent becomes more responsive to a principal’s attention (as we have seen in Lemma 1), and so each principal has a stronger incentive to provide attention.

Lastly, in the first stage the principals decide on the number of principals to share an agent with, taking into account the effects on attention and effort we derived above as well as the effect on the wage costs. As in the first-best, the wage costs are equally shared among the homogenous principals. The following proposition describes the optimal number of principals that share an agent when neither effort nor attention are contractible and reiterates the levels of effort and attention that arise.

**Proposition 3** If neither effort nor attention are contractible, then:

1. the agent’s effort is described by 
   \[-Ce(e_i) + Fe(e_i, a_i) = 0\] for all \(i\),
2. the principal’s attention is given by 
   \[da_i / dn \cdot Qe(e_i) - Ha(a_i) = 0\] for all \(i\),
3. the number of principals that share an agent is described by:
   \[1/2 \cdot [U + C(\int_0^n e_i)] = 1/n \cdot c_i C(\int_0^n e_i) - d a_i / dn [H_a(a_i) - F_a(e_i, a_i)] = 0.\]

The proof is in the appendix. As in the first-best case described in Proposition 1, the first two terms in (9) show that principals strive to minimize the costs of the agent’s services by setting the number of principals that share an agent such that the agent’s average and marginal cost of effort are equal. Importantly, the level of \(n\) where the first two terms in (9) are zero need not coincide with the first-best level of \(n\). The reason is that the level of effort will generally be different in the two cases, see the first parts
of Propositions 1 and 3. When effort is lower than in the first-best (because the lack of financial incentives is not made up for by increased attention), principals have an incentive to share an agent with more principals than in the first-best case, because sharing an agent is less costly when the agent provides lower effort for each principal. When effort is higher than in the first-best (when the attention race is severe, inducing high effort for each principal), the reverse is true: principals have an incentive to share an agent with fewer principals.

The third term in (9) describes the effect of the number of principals that share an agent on the principal’s payoff through its effect on attention provision. Note that the terms in brackets sum to zero when both attention and effort are at their first-best level. However, principals’ attention will likely not match with the first-best level because, as we have seen, the motives for giving attention differ between the two cases (compare the second parts of Propositions 1 and 3). While in the first-best case attention is set so as to maximize the agent’s and principals’ joint surplus from attention provision, in the case of noncontractibility, attention is chosen noncooperatively and only given to evoke reciprocal feelings in the agent. If the resulting attention is low so that \( F_a \) exceeds \( H_a \) (e.g., because the agent values attention a lot but the agent’s effort is only weakly responsive to attention), then the third term in (9) is positive and so principals have an incentive to increase the number of principals that share an agent. By doing so, they make the attention race more severe, which is a blessing, since it mitigates the principals’ underprovision of attention. Conversely, if attention is high so that \( F_a \) is lower than \( H_a \) (e.g. because the attention race is severe, while attention is not valued that much by the agent), the third term in (9) is negative and so principals have an incentive to limit the number of principals that share an agent so as to reduce wasteful attention provision.

7 Conclusion

This paper has explored social interaction in the workplace between managers and workers. We have investigated the conditions under which inefficient social exchange in organization can occur, and how this may be avoided. We have used a common agency setting to allow for multiple post-hiring loyalties in organizations on top of the formal employment contract. We think that such a setting captures the essences of life in the modern firm, for many employees effectively work to satisfy the demands of more than one superior. On top of the examples of employees in support de-
partments mentioned in the introduction, we may add the many employees in matrix organizations (who report to a functional manager as well as to a project manager), internal service managers (who provide an interface between parts of the organization), and procurement managers (who coordinate and communicate the requests of various branches of the organization towards external suppliers).

Our paper is just one piece in a growing body of literature on behavioral personnel economics that stresses the importance of feelings in the workplace. Classic early examples are the papers by Rotemberg (1994) on altruism among workers and Rotemberg and Saloner (1993) on the role of emphatic managers in alleviating the hold-up problem. In Rob and Zemsky (2002) workers derive social utility from cooperative tasks that have no monetary rewards. Akerlof and Kranton (2005) discuss the implications of social identity at work. Delfgaauw and Dur (2005) and Francois (2006) consider how monetary rewards affect the sorting of motivated workers into firms. In addition, there is a growing literature that studies manager-subordinate interaction in the post-hiring phase. For example, using a model with imperfect information and a single principal, Benabou and Tirole (2003) and Ellingsen and Johannesson (2006) discuss how task assignment and pay structure may reveal the manager’s private judgement of the worker’s ability or his beliefs about the abilities of workers in general. The communication of these beliefs in turn affects the worker’s motivation. In our view, these papers have considerably increased our understanding of human relations in the workplace.

We have taken a noncognitive approach to modeling reciprocity. The economic literature offers two alternative ways to model reciprocal behavior. First, the intention-based approach stresses motives rather than actions (Rabin 1993, Falk and Fischbacher 2006). A recent application of this view to manager-worker relations is Arbak and Kranich (2005) who derive conditions under which a high wage offer is informative about the employer’s true intentions. A similar reasoning may hold for manager’s attention-giving. Second, agents may reciprocate out of inequity aversion (Fehr and Schmidt 1999, Bolton and Ockenfels 2000). Englmayer and Wambach (2005) and Fehr et al. (2006) study optimal contract design when agents care for the relative payoff. Our treatment of reciprocity need not conflict with this approach, except when the wealth of the principal and the agent are far apart. Recently, Maximiano et al. (2006a, 2006b) show in experiments that agents reciprocate to a generous wage, irrespective of the number of agents the principal employs and also irrespective of whether the principal is residual claimant.
What should managers take away from our paper? First, granting more flexibility to employees in their choice of effort makes it easier for organizations to measure and reward the performance of middle managers. When employees are easily motivated by these managers, then the organization gets hold of a powerful set of motivational practices. Second, we have seen that increasing performance pay for workers should be complemented by stronger incentives for managers to support their subordinates. Third, one should be cautious with sharing human resources, for it results in rivalry among managers, absorbing precious corporate energy. Hence, in addition to worrying about underprovision of support by managers, executives should also be concerned about limiting excessive influence activities by managers in the workplace.16 Last, selecting and recruiting employees on assumed reciprocity, flexibility and loyalty can have its cost, as it induces managers higher on the corporate ladder to devote too much time and energy to social exchange.

Appendix

Proof of Proposition 1

With verifiable effort and attention, principals design a forcing contract, specifying both the effort exerted by the agent for each principal, and the attention given by each principal to the agent. Since principals are homogeneous and we have decreasing returns to scale in production and convex cost of effort and attention, we can safely assume that the first-best contract has effort and attention being the same for all principals,17 and that the wage of the agent is shared equally among the \( n \) principals that share the agent:

\[
s_i = \frac{w}{n}.
\]

Using this and the principal’s payoff function (1), the first-best optimal levels of effort, attention, and number of principals who share an agent solves:

\[
\max_{a_i, e_i, n} Q(e_i) - H(a_i) - \frac{1}{n} \left[ U + C(\int_0^n e_i) - \int_0^n F(e_i, a_i) \right],
\]

16 A similar concern can apply to post-hiring influence activities by workers, see Milgrom (1988), Milgrom and Roberts (1988), Fairburn and Malcomson (2001), and Inderst et al. (2005) for insightful analyses.

17 Alternatively, we can proceed in two steps. First, for a given \( n \), determine the jointly optimal levels of \( e_i \) and \( a_i \) for all \( i \), where \( e_i \) and \( a_i \) are allowed to be different across principals. By our assumptions about the shape of the functions \( Q, H, \) and \( F \), we find that optimal \( e_i \) and \( a_i \) are the same for all \( i \). Second, using these results, determine the optimal \( n \) from the perspective of an individual principal. The results are the same.
where \( a_i \) and \( e_i \) are restricted to be the same for all \( i \) and the term in brackets is the agent’s wage which follows from the agent’s participation constraint (3). The first-order conditions are:

\[
Q_e(e_i) - C_e(\int_0^n e_i) + F_e(e_i, a_i) = 0 \quad \text{for all } i, \tag{A1}
\]

\[
-H_a(a_i) + F_a(e_i, a_i) = 0 \quad \text{for all } i, \tag{A2}
\]

\[
\frac{1}{n^2} \left[ \bar{U} + C(\int_0^n e_i) - \int_0^n F(e_i, a_i) \right] \frac{1}{n} \left[ e_i C_e(\int_0^n e_i) - F(e_i, a_i) \right] = 0. \tag{A3}
\]

The conditions (A1) and (A2) are identical to those in the first and second part of Proposition 1. Since optimal effort and attention are the same for all \( i \), condition (A3) can be simplified to:

\[
\frac{1}{ne_i} \left[ \bar{U} + C(\int_0^n e_i) \right] = C_e(\int_0^n e_i), \tag{A4}
\]

which is identical to the condition in the third part of Proposition 1. □

**Proof of Corollary 1**

For convenience, we introduce a weight, \( \gamma \), on the term \( \int_0^n F(e_i, a_i) \) in the agent’s utility function (2). Differentiating the adapted first-order conditions to \( e_i, a_i, n, \) and \( \gamma \) and solving the differential equations gives:

\[
\frac{de_i}{d\gamma} = -\frac{(H_{aa} - \gamma F_{aa}) \gamma F_e - \gamma^2 F_{ea} F_a}{(Q_{ee} + \gamma F_{ee}) (H_{aa} - \gamma F_{aa}) + \gamma^2 F_{ea} F_a} > 0,
\]

\[
\frac{da_i}{d\gamma} = \frac{1}{(H_{aa} - \gamma F_{aa})} \left( \gamma F_{ae} \frac{de_i}{d\gamma} + \gamma F_a \right) > 0,
\]

\[
\frac{dn}{d\gamma} = -\frac{n}{e} \frac{de_i}{d\gamma} < 0,
\]

where we have suppressed the arguments of all functions for brevity, and the denominator of \( de_i/d\gamma \) is negative by the second-order conditions. □

**Proof of Lemma 1**

Note that (7) represents \( n \) first-order conditions describing optimal effort for all \( n \) principals. Applying the implicit function theorem gives:
\[
\frac{de_i}{dn} = \frac{e_i C_{ee}}{F_{ee} - nC_{ee}} < 0, \\
\frac{de_i}{da_i} = \frac{-F_{ea} [F_{ee} - (n - 1)C_{ee}]}{F_{ee}^2 - nC_{ee}F_{ee}} > 0, \\
\frac{de_{j\neq i}}{da_i} = \frac{-F_{ea}C_{ee}}{F_{ee}^2 - nC_{ee}F_{ee}} < 0, \\
\frac{d f^n_{0} e_i}{da_i} = \frac{de_i}{da_i} + (n - 1) \frac{de_{j\neq i}}{da_i} = \frac{-F_{ea}}{F_{ee} - nC_{ee}} > 0, \\
\frac{d (de_i/da_i)}{dn} = \frac{-F_{ea}F_{ee}C_{ee}^2}{(F_{ee}^2 - nC_{ee}F_{ee})^2} > 0,
\]

where the signs follow from our assumptions about the functions \(F\) and \(C\) as described in Section 3. \(\square\)

**Proof of Lemma 2**

Applying the implicit function theorem to (8), while treating \(e_i\) as an endogenous variable and noting that \(de_i/da_i\) depends on \(n\) but not on any of the other endogenous variables (see the proof of Lemma 1 and recall that all third derivatives are negligibly small), gives:

\[
\frac{da_i}{dn} = \frac{1}{H_{aa}} \left[ \frac{d (de_i/da_i)}{dn} Q_e + \frac{de_i}{dn} Q_e \right] = \frac{1}{H_{aa}} \left( \frac{-F_{ea}F_{ee}C_{ee}^2 Q_e}{(F_{ee}^2 - nC_{ee}F_{ee})^2} + \frac{e_i C_{ee} Q_e}{F_{ee} - nC_{ee}} \right) > 0.
\]

\(\square\)

**Proof of Proposition 3**

The proofs to the first and second part of Proposition 3 are given in the main text preceding the proposition. The third part describes the optimal number of principals that share an agent, which follows from:

\[
\max_n Q(e_i) - H(a_i) - \frac{1}{n} \left[ \mathcal{U} + C(\int_0^n e_i) - \int_0^n F(e_i, a_i) \right].
\]

resulting in the first-order condition:

\[
\frac{1}{n^2} \left[ \mathcal{U} + C(\int_0^n e_i) - \int_0^n F(e_i, a_i) \right] - \frac{1}{n} \left[ e_i C_{ee}(\int_0^n e_i) - F(e_i, a_i) \right] + \\
+ \frac{da_i}{dn} \left\{ -H_{aa}(a_i) + F_a(e_i, a_i) + \frac{de_i}{da_i} \left[ Q_e(e_i) - C_e(\int_0^n e_i) + F_e(e_i, a_i) \right] \right\} + \\
+ \frac{de_i}{dn} \left[ Q_e(e_i) - C_e(\int_0^n e_i) + F_e(e_i, a_i) \right].
\]

(10)
Using the first-order conditions (7) and (8) and noting that $a_i$ and $e_i$ are identical for all $i$, the condition can be simplified to the expression in Proposition 3. □
References


