Economic Theory (2006)
DOI 10.1007/s00199-005-0057-3

RESEARCH ARTICLE

Ingela Alger · Régis Renault

Screening ethics when honest agents keep their word

Received: 5 April 2004 / Accepted: 20 October 2005

Abstract Using a principal-agent setting, we introduce honesty that requires pre-commitment. The principal offers a menu of mechanisms to screen ethics. Agents may misrepresent ethics. Dishonest agents may misrepresent the match with the assigned task (good or bad), while honest agents reveal the match honestly if they have pre-committed. Ethics-screening, that allows for match-screening with dishonest agents while leaving a lower rent to honest agents, is optimal if both honesty and a good match are likely. Otherwise the optimal mechanism is the standard second-best or the first-best (where dishonest agents misrepresent the match), if dishonesty is likely or unlikely respectively.

Keywords Ethics · Honesty · Loyalty · Adverse selection · Screening

JEL Classification Numbers D82

1 Introduction

For the past 30 years, a large body of literature has developed on the theme that individuals use their private information in an opportunistic manner. However,

An earlier version of this paper was entitled "Honest Agents and Equilibrium Lies." We are grateful to seminar participants at MIT, Université de Caen, University of St Andrews, University of Virginia, and ESEM 99, and two anonymous referees for useful comments.

I. Alger (✉)
Economics Department, Boston College, 140 Commonwealth Avenue,
Chestnut Hill, MA 02467, USA
E-mail: ingela.alger@bc.edu

R. Renault
THEMA, Université de Cergy-Pontoise, 33 Bd du Port, 95011 Cergy-Pontoise Cedex, France
Institut Universitaire de France, 103 boulevard Saint Michel, 75005 Paris, France
E-mail: regis.renault@u-cergy.fr

evidence suggests that some individuals behave honestly even if they thereby forgo material benefits. For instance, using data from a firm that kept monitoring at a high level while making the employees believe that it had been relaxed, Nagin, Rebitzer, Sanders and Taylor (2002) found that only some employees chose to increase the extent of shirking.[1] Some authors have introduced such an heterogeneity in ethics in various contexts. In Jaffee and Russell (1976), banks may ration credit if borrowers differ in their willingness to default on loans. Assuming that some taxpayers report their income truthfully regardless of economic incentives, Erard and Feinstein (1994) are able to rule out unrealistic equilibria.[2] Analyzing various economic settings, Tirole (1992), Kofman and Lawarrée (1996), and Alger and Ma (2003) find that it is suboptimal to deter collusion if potential dishonesty is sufficiently unlikely. Among these papers, only Alger and Ma allow for screening the agent's ethics. In this paper and our companion work, Alger and Renault (2006), we address this question in the canonical principal-agent model with adverse selection and look at the implications of different specifications of an honest behavior.

We consider a model where a principal contracts over a decision (e.g., output) and a transfer (e.g., a wage) with an agent, whose match with the principal may be either good or bad. Moreover, the agent may be either opportunistic or honest, where an honest agent reveals the match even if he is not given proper incentives. The principal observes neither the agent's ethics, nor his match. If honesty involves revealing any private information at no cost, the principal screens ethics perfectly: she may screen matches whether the agent is honest or dishonest, while only handing out an informational rent to a dishonest agent. This has been shown by Deneckere and Severinov (2001, 2003), who characterize the optimal contract in a related setting with a continuum of matches. By contrast if, like in the early literature, the principal could only resort to contracts where the set of messages is the set of matches, then in the optimal contract the principal either leaves as much rent to an honest and a dishonest agent, or gives up screening matches for a dishonest agent.[3]

These two polar cases illustrate the potential benefits from screening ethics. However, assuming that an honest agent unconditionally reveals any private information at no cost does not receive much support from research in psychology. Hartshorne and May (1928, p. 385) laid out what is still the basic tenet for much of this literature, namely, that moral behavior cannot be viewed as emanating from "an inner entity operating independently of the situations in which the individuals are placed". Instead, the general perception is that behavior is conditional on various factors; in particular, studies have shown that such concerns as financial needs, the fear of getting caught, perceived equity, and loyalty all affect the propensity to behave honestly (see, e.g., Spicer and Becker 1980; Terris and Jones 1982), all of which would translate differently into a formal model. In our companion paper, we focus on honesty as being conditional on the perceived equity of the contract. In this paper, we instead look at loyalty as a trigger of honest behavior.

[1] See the references in Alger and Ma (2003) and Alger and Renault (2006) for further examples.

[2] See Picard (1996) for an equilibrium analysis of insurance markets with heterogeneous ethics.

[3] Strictly speaking this is true in general only when there are two match realizations; if there are more than two the principal may leave less rent to an honest agent while only partially giving up match screening for a dishonest.

The following two examples illustrate situations where a sense of loyalty may lead an agent to follow the prescriptions of some explicit or implicit contract. In subcontracting situations, such as with plumbers and carpenters, there frequently is *ex ante* uncertainty pertaining to the cost of the assignment. Typically the subcontractor would learn the relevant information while completing the job, whereas the customer would remain uninformed. A prior discussion of these uncertainties with the house owner might prevent some subcontractors from inflating the cost. In the workplace, an employee may discover how well he is matched with a given task only in the course of performing it: either because of unforeseen contingencies, or because he has never performed this task before. This may for instance result in different times needed to achieve a given output. In this case an honest employee would reveal this information if he has committed to doing so by signing a pre-specified contract with the employer.

In their work on commitment in the workplace, Meyer and Allen (1991) distinguish between three kinds of commitment: "Affective commitment refers to the employee's emotional attachment to, identification with, and involvement in the organization. (...) Continuance commitment refers to an awareness of the costs associated with leaving the organization. (...) Finally, normative commitment reflects a feeling of obligation to continue employment" (p. 67). In their 1997 book, they devote a chapter to understanding how commitment towards an organization develops. Existing research indicates that individuals vary in their propensity to become committed, but that "relations between demographic variables and affective commitment are neither strong nor consistent" (Meyer and Allen 1997, p. 43). Instead, the degree of affective commitment depends to a large extent on organizational features, such as fairness and supportiveness, as well as work experiences (more or less pleasant experiences with supervisors and co-workers, the degree of challenge, the variety of skills required, etc.). These findings thus suggest that commitment may be important for some individuals only, that an employee's sense of commitment develops after being hired, and that there seems to be no clear relation between commitment and observable differences such as gender or age. Meyer and Allen (1997, pp. 26–35) further report evidence pointing to a link between an employee's sense of commitment and his honesty (or lack of willingness to shirk); numerous studies have shown that an employee's degree of commitment to the firm is (a) negatively correlated with his/her likelihood of "voluntary absence" (absenteeism due to reasons that were under the employee's own control), and (b) positively correlated with his job performance.

We use these insights to formalize honesty in a two-period principal-agent setting. The agent discovers his match in the second period. However, he knows his ethics in the first period: for instance he knows whether or not he will develop a strong sense of commitment in the second period. An honest agent would reveal his match truthfully provided that he has signed a contract specifying which allocation should be implemented as a function of the true match, i.e., a direct mechanism. Any contract specifying more than one message pertaining to a specific match, or using messages pertaining to no true match realization, would be perceived as attempts by the principal to manipulate the honest agent. In the first period the principal may offer a menu of contracts in order to screen ethics. Because the agent has not yet developed a sense of commitment in the first period, he is willing to misrepresent his ethics, should that give him a higher expected surplus.

This setup differs from that of Alger and Renault (2006) and Deneckere and Severinov (2001) where an honest agent behaves honestly even if the contract specifies more allocations than there are match realizations. However, we allow for an *ex ante* screening of ethics by letting the principal offer a menu of contracts before the match realization is known to the agent; in the early literature involving ethics heterogeneity the principal was restricted to offering only one contract specifying a unique allocation for each match realization.

The principal would ideally choose to leave no rent to an honest agent, while leaving a rent to an opportunistic agent in order to screen matches. But since an honest agent is willing to lie about his ethics, inducing truthful revelation of the match by an opportunistic agent requires leaving as large a rent for an honest as for an opportunistic agent. The only way for the principal to leave a smaller rent to an honest agent is to allow for misrepresentation of the match by an opportunistic agent. This means that the revelation principle does not apply. This failure of the revelation principle, that was a key feature of the literature where ethics screening was ruled out, arises despite the potential for screening ethics in the present framework. In fact, we find that the no-screening constraint that was exogenously imposed in that literature may emerge as a property of the optimal mechanism. Then the contract is as if the agent were dishonest with certainty if dishonesty is sufficiently likely, so that the standard analysis is robust to the introduction of honest agents. Otherwise, the contract is as if the agent were honest with certainty, i.e., as if information were complete. In this case, a dishonest agent always claims that the match is bad, which results in an inefficient allocation when the match is actually good: in this sense there is a distortion at the top, in contrast with standard results.

Without ethics screening, the principal is led to choose between screening matches for the dishonest and leaving no rent to the honest. We show that there exists an alternative contract involving ethics screening that enables the principal to screen matches for the dishonest, while leaving a lower rent to an honest than to a dishonest. This is achieved by inducing the dishonest to systematically misrepresent the match, and by exploiting the fact that the honest is not willing to misrepresent the match should he choose the contract meant for the opportunistic agent. Clearly, if the probability that the agent is dishonest is above one half, the principal would like to do the reverse, namely, give a smaller rent to the dishonest. However, this is impossible since the dishonest agent is unrestrained in his willingness to misrepresent the match. As a result, ethics screening is only used when the agent is more likely to be honest than dishonest. It follows that the standard analysis, where there are no honest agents, is robust to the introduction of honest agents.

Whether or not ethics screening is used when honesty is likely depends on how it compares to the contract where the principal does not screen matches for the dishonest. The latter contract involves a cost for the principal only in the event that the match is good, because the allocation for the dishonest is then distorted. By contrast the cost of ethics screening is to leave a rent to the honest. We find that ethics screening dominates the complete information contract when the probability of a good match is sufficiently large. Only then does the benefit from match screening for the opportunistic outweigh the cost of leaving a rent to the honest.

The next section introduces the formal model. In Section 3, we characterize the optimal contract, and contrast results with those of Alger and Ma (2003) and Alger and Renault (2006), and we conclude in Section 4. Proofs are in Appendix.

2 The model

We model an honest behavior triggered by a commitment to reveal some private information. To this end we consider a two-period framework where a contract is signed in period 1 before some piece of private information (the "match") is revealed to the agent.

Although our model applies to many situations, we specify it to suit the employer–employee relationship. An employer (the principal) hires an individual (the agent) to perform a task with output x. This output gives the principal surplus $\pi(x)$, where π is a strictly increasing and concave function, with $\pi(0) = 0$, $\pi'(0) = +\infty$, and $\lim_{x\to+\infty} \pi'(x) = 0$. The principal pays the agent a transfer t, so that her net surplus is $\Pi(x, t) = \pi(x) - t$.

There is uncertainty as to how well the agent's abilities match the requirements of the task he is assigned. Let $v(x, \theta)$ denote the cost for the agent of producing x units, where the parameter θ represents the match.[4] The match may be either good ($\theta = \underline{\theta}$) or bad ($\theta = \bar{\theta}$), with the probability of a bad match being $\alpha \in (0, 1)$. The cost function v is strictly increasing and strictly convex in x, with $v(0, \theta) = 0$. Finally, when the match is good, the production cost incurred by the agent is smaller than when it is bad, for any output level: $\forall x > 0,\ v'(x, \bar{\theta}) > v'(x, \underline{\theta})$ (the prime indicates a partial derivative with respect to x). This is the standard Spence–Mirrlees condition. The employee's net surplus is the transfer net of the production cost $U(x, t, \theta) = t - v(x, \theta)$. The agent's reservation utility is normalized to zero; we assume that the agent may quit at any time. For further reference, the first-best values of x, $\bar{x}^*$ if $\theta = \bar{\theta}$ and $\underline{x}^*$ if $\theta = \underline{\theta}$, are defined by $\pi'(\bar{x}^*) = v'(\bar{x}^*, \bar{\theta})$ and $\pi'(\underline{x}^*) = v'(\underline{x}^*, \underline{\theta})$, respectively. Given the properties of π and v, they are uniquely defined, and $0 < \bar{x}^* < \underline{x}^*$.[5] We will call an output with its corresponding transfer an allocation $y = (x, t)$.

In the standard setting, the principal would discriminate between a good and a bad match using a mechanism, to be at work in the second period.

Definition 1 (Mechanism) A mechanism m defines a space $\mathcal{M}_2$ of messages μ_2 and a mapping $y_m : \mathcal{M}_2 \to \mathbb{R}^+ \times \mathbb{R}^+$, where $\mathbb{R}^+ \times \mathbb{R}^+$ is the set of feasible allocations. A mechanism is said to be direct if the message space is the set of matches, $\mathcal{M}_2 = \{\underline{\theta}, \bar{\theta}\}$; otherwise, it is indirect.

The principal faces an additional uncertainty, regarding the agent's ethics, denoted k: independently of the match realization, with probability $\gamma \in (0, 1)$, the agent is dishonest ($k = d$), and with the complementary probability, he is honest ($k = h$). All probability distributions are common knowledge. By contrast with

[4] The results would not be substantially affected if we allowed for θ to directly affect the principal's surplus π (see further the discussion following Proposition 4); however, the notation and some of the arguments used in the proofs would be more cumbersome.

[5] Additional assumptions about functional forms would be needed to guarantee existence and uniqueness of solutions in the incomplete information cases considered in the paper.

the match realization which is revealed to the agent in period 2, the agent learns his ethics in period 1, before any contract is signed. The agent's ethics is his private information. A dishonest agent always maximizes his surplus, like in any standard principal-agent setting. By contrast, an honest agent may forego material benefits to truthfully announce his match θ (at date 2). In order for him to feel committed to truthful match revelation in period 2, it must be that all allowed messages specify a match realization. If a mechanism specifies messages that are unrelated to the true values of θ, telling the truth is a fuzzy notion and we might as well assume that an honest agent would not be reluctant to announce such messages, so that the principal could not gain anything by using them.[6] We further assume that an honest agent does not feel committed to revealing his match truthfully if there are several messages specifying the same match realization.

With our specification of honesty, the contract offered in the first period may be described as a menu of direct mechanisms:

Definition 2 (Contract) A contract C defines a space $\mathcal{M}_1$ of messages μ_1 that may be sent by the agent at date 1, and a mapping $c : \mathcal{M}_1 \to M_D$ where M_D is the set of direct mechanisms m (see Definition 1).

The timing may be summarized as follows. In period 1, the agent learns his ethics, the principal then offers a contract, and finally the agent selects a message in $\mathcal{M}_1$. This message determines which direct mechanism will be used in period 2. In period 2, the agent learns his match, and chooses a message in $\{\underline{\theta}, \bar{\theta}\}$, and the associated allocation is carried out.

Formally, honesty is equivalent to the following restriction on the message space at date 2:

Assumption 1 (Ethics I) (Messages at date 2)

(i) A dishonest may announce any $\theta \in \{\underline{\theta}, \bar{\theta}\}$.
(ii) An honest with match $\theta \in \{\underline{\theta}, \bar{\theta}\}$ must announce θ if a contract was signed at date 1; otherwise, he may announce any $\theta \in \{\underline{\theta}, \bar{\theta}\}$.

Since no contract has been signed prior to date 1 an honest agent does not feel committed to revealing his ethics. Formally, we have the following assumption

Assumption 2 (Ethics II) Whether the agent is honest or dishonest, he may announce any $\mu_1 \in \mathcal{M}_1$.

For further use below, let $\hat{\theta}^*(\theta, k, m)$ denote the agent's best response in mechanism m if his match is θ, and $\mu_1^*(k)$ the best response at date 1. As a tie-breaking rule, we assume that whenever an agent is indifferent between revealing the truth and lying, he reveals the truth.

As we argue in our companion paper Alger and Renault (2006), this analysis could be phrased in more general terms by allowing for messages that do not necessarily spell out a match realization. What is critical is that for each match realization an honest agent is restricted to a subset of messages, where subsets

[6] This is related to the idea that honest behavior may be linked to a fear of being found out lying. In many cases, one can argue that the value of θ is verifiable information, albeit at a very high cost; a potentially honest agent may underestimate that cost. But if the honest agent sends the message "blue" for instance, there is no information to be verified.

associated with different matches do not intersect. Here it is assumed that an honest behavior is guaranteed only if at most one allocation is associated with each subset, and no allocation is associated with a message outside these restricted sets. By contrast, in Deneckere and Severinov (2001) honest behavior is ensured even with mechanisms specifying allocations for messages that are out of reach of the honest agent.[7]

As a benchmark we describe the optimal contracts if ethics were known. First, if the principal faced an honest agent, she would implement the first-best decisions and extract the whole surplus by offering a contract with one message leading to the first-best mechanism:

Definition 3 (First-best mechanism) The first-best mechanism, denoted m^*, defines $\mathcal{M}_2 = \{\underline{\theta}, \bar{\theta}\}$, and the mapping associating $(\bar{x}^*, \bar{t}^*)$ to message $\mu_2 = \bar{\theta}$, and $(\underline{x}^*, \underline{t}^*)$ to the message $\mu_2 = \underline{\theta}$, where $\bar{t}^* = v(\bar{x}^*, \bar{\theta})$ and $\underline{t}^* = v(\underline{x}^*, \underline{\theta})$.

If the agent is known to be dishonest, the revelation principle applies and the mechanism that maximizes the principal's expected surplus may be found by imposing individual rationality and incentive compatibility constraints. This yields the standard second-best mechanism (Definition 4).

Definition 4 (Standard second-best mechanism) The standard second-best mechanism, denoted m^s, defines $\mathcal{M}_2 = \{\underline{\theta}, \bar{\theta}\}$, and the mapping associating $(\bar{x}^s, \bar{t}^s)$ to message $\mu_2 = \bar{\theta}$, and $(\underline{x}^s, \underline{t}^s)$ to message $\mu_2 = \underline{\theta}$, where

$$\bar{t}^s = v(\bar{x}^s, \bar{\theta}) \qquad \underline{t}^s = v(\underline{x}^*, \underline{\theta}) + [v(\bar{x}^s, \bar{\theta}) - v(\bar{x}^s, \underline{\theta})]$$

$$v'(\bar{x}^s, \bar{\theta}) = \pi'(\bar{x}^s, \bar{\theta}) - \frac{1-\alpha}{\alpha}[v'(\bar{x}^s, \bar{\theta}) - v'(\bar{x}^s, \underline{\theta})], \qquad \underline{x}^s = \underline{x}^*.$$

In equilibrium, the agent receives a rent if the match is good, and the allocation is sub-optimal if the match is bad.

If ethics is the agent's private information, and if the first-period menu offered by the principal is comprised of the first-best and the standard second-best mechanism, the agent always selects the wrong mechanism from the principal's viewpoint. This shows that the solution with complete information on ethics is not incentive compatible in a very strong sense since both honest and dishonest agents would deviate. We next turn to determining the optimal contract.

3 Analysis

3.1 Preliminaries

The standard approach to the present asymmetric information problem would rely on the revelation principle, whereby for any contract there exists a direct incentive-compatible contract that implements the same allocation in every state of nature.

[7] Deneckere and Severinov (2001) use "password mechanisms" where the agent announces his match together with the set of matches he could have announced, and they assume that an honest agent may not lie about the set of matches he could have announced. Similar mechanisms are also used in Forges and Koessler (2005) for communication equilibria with certifiable information.

Here it is straightforward to prove that this principle does not hold. Consider the following contract: $(m_h, m_d) = (m^*, m^*)$. In equilibrium, the honest agent reveals θ truthfully and gets no rent. The dishonest agent always announces $\bar{\theta}$; he gets no rent if $\theta = \bar{\theta}$, and the rent $v(\bar{x}^*, \bar{\theta}) - v(\bar{x}^*, \underline{\theta})$ if $\theta = \underline{\theta}$. According to the revelation principle, there should exist a menu of direct mechanisms (m'_h, m'_d) yielding the same allocations as the original ones in every state of nature, and inducing the agent to tell the truth. Therefore, m'_d must specify the allocation $(\bar{x}^*, \bar{t}^*)$ whether the agent announces $\underline{\theta}$ or $\bar{\theta}$. As a result, for the honest agent to reveal h truthfully, m'_h must give him an expected rent $(1-\alpha)[v(\bar{x}^*, \bar{\theta}) - v(\bar{x}^*, \underline{\theta})] > 0$, which yields a contradiction with the requirement that the allocation should be the same in every state of nature, in which case the honest agent would get no rent. Hence the revelation principle does not hold in our framework.[8] At first sight, the task to determine the optimal contract thus seems daunting. The following observation facilitates the determination of the optimal contract.

Lemma 1 *Without loss of generality, one may focus on contracts C that define the message space* $\mathcal{M}_1 = \{h, d\}$ *and that induce truth-telling at date 1, i.e,* $\mu_1^*(h, C) = h$ *and* $\mu_1^*(d, C) = d$.

Lemma 1 implies that we may impose two date 1 incentive compatibility constraints ensuring that an honest agent announces h while a dishonest agent announces d. We may now use a lighter notation: a mechanism m_k, where $k \in \{h, d\}$, specifies the allocation $(\underline{x}_k, \underline{t}_k)$ if the date 2 message is $\underline{\theta}$, and $(\bar{x}_k, \bar{t}_k)$ if the date 2 message is $\bar{\theta}$. We will say that ethics screening occurs if $m_d \neq m_h$.

Given that the equilibrium behavior of the agent is not known a priori, we have to determine the contracts that are optimal conditional on assumed behavior, before being able to determine the equilibrium contract. We start by analyzing mechanisms that do induce truth-telling at date 2 by the dishonest agent, before turning to mechanisms inducing lies.

3.2 Incentive-compatible contracts

Here we assume that in equilibrium a dishonest agent reveals θ at date 2: $\forall \theta \in \{\underline{\theta}, \bar{\theta}\}$ we have $\hat{\theta}^*(\theta, d, m_d) = \theta$. Thus the principal chooses the pair of mechanisms $\{m_d, m_h\} = \{\{\bar{x}_d, \bar{t}_d, \underline{x}_d, \underline{t}_d\}, \{\bar{x}_h, \bar{t}_h, \underline{x}_h, \underline{t}_h\}\}$ so as to maximize:

$$\gamma[\alpha(\pi(\bar{x}_d) - \bar{t}_d) + (1-\alpha)(\pi(\underline{x}_d) - \underline{t}_d)]$$

$$+(1-\gamma)[\alpha(\pi(\bar{x}_h, \bar{\theta}) - \bar{t}_h) + (1-\alpha)(\pi(\underline{x}_h, \underline{\theta}) - \underline{t}_h)].$$

[8] This is similar, although not formally equivalent, to the result in Green and Laffont (1986) that in a one-stage information revelation problem, when an agent's message space is restricted, the revelation principle may fail to apply. See Deneckere and Severinov (2001), Forges and Koessler (2005), and Alger and Renault (2006) for further discussions of this issue in one-stage models. Alger and Ma (2003) have the same informational and contractual structure as we do; as a result this observation is true in their setting as well.

For the agent to reveal his ethics at date 1, the following incentive compatibility constraints must hold:

$$\begin{aligned}
&\alpha[\bar{t}_h - v(\bar{x}_h, \bar{\theta})] + (1-\alpha)[\underline{t}_h - v(\underline{x}_h, \underline{\theta})] \\
&\quad \geq \alpha[\bar{t}_d - v(\bar{x}_d, \bar{\theta})] + (1-\alpha)[\underline{t}_d - v(\underline{x}_d, \underline{\theta})] \qquad (1) \\
&\alpha[\bar{t}_d - v(\bar{x}_d, \bar{\theta})] + (1-\alpha)[\underline{t}_d - v(\underline{x}_d, \underline{\theta})] \\
&\quad \geq \alpha \max\{\bar{t}_h - v(\bar{x}_h, \bar{\theta}), \underline{t}_h - v(\underline{x}_h, \bar{\theta})\} \\
&\qquad +(1-\alpha)\max\{\underline{t}_h - v(\underline{x}_h, \underline{\theta}), \bar{t}_h - v(\bar{x}_h, \underline{\theta})\} \qquad (2)
\end{aligned}$$

An honest agent reveals θ truthfully irrespectively of the mechanism chosen at date 1. Therefore, the incentive constraint for the honest agent (1) is straightforward. In contrast, we do not know the dishonest agent's behavior if he were to claim to be honest at date 1: therefore, the right-hand side of constraint (2) is not based on truth-telling. However, by assumption, the dishonest agent truthfully reveals θ in equilibrium, yielding the expression on the left-hand side of the constraint. Second, for the dishonest agent to reveal his match θ in equilibrium, the two following incentive compatibility constraints must hold, but as is standard only the first one binds:

$$\underline{t}_d - v(\underline{x}_d, \underline{\theta}) \geq \bar{t}_d - v(\bar{x}_d, \underline{\theta}) \qquad (3)$$

$$\bar{t}_d - v(\bar{x}_d, \bar{\theta}) \geq \underline{t}_d - v(\underline{x}_d, \bar{\theta}). \qquad (4)$$

Finally, the participation constraints are:

$$\bar{t}_k - v(\bar{x}_k, \bar{\theta}) \geq 0 \quad k = h, d \qquad (5)$$

$$\underline{t}_k - v(\underline{x}_k, \underline{\theta}) \geq 0 \quad k = h, d. \qquad (6)$$

The principal maximizes his expected utility subject to the constraints (1)-(6), yielding:

Proposition 1 *Suppose that the principal is restricted to contracts inducing the dishonest to reveal his true θ. Then the optimal contract is to offer the standard second-best mechanism independent of the announced ethics.*

The intuition is as follows. The principal must give the dishonest agent a rent to make him reveal θ at date 2. But since the honest agent behaves opportunistically at date 1, she must give him the same rent. As a result, the optimal contract is as if the agent were dishonest with certainty. Proposition 1 indicates that the only way to leave a smaller rent to the honest than to the dishonest agent is to let the dishonest agent manipulate the information at date 2.

3.3 Contracts inducing lies

From Lemma 1 we may without loss of generality impose date 1 incentive compatibility constraints, which ensure truthful ethics revelation. Which kind of second period incentive compatibility constraints should be imposed, depends on the assumed equilibrium and out-of-equilibrium behavior of the dishonest agent; the exact formulation of all constraints also depends on this. Given that the dishonest

lies in equilibrium we need to consider only two cases: the revealing lies case where he always lies in equilibrium (the principal can then infer θ), and the non-revealing lies case where the dishonest agent's equilibrium announcement is independent of his match.

In the revealing lies case the dishonest lies systematically, by announcing $\bar{\theta}$ when the match is $\underline{\theta}$, and vice versa. The principal therefore chooses the pair of mechanisms $\{m_d, m_h\} = \{(\bar{x}_d, \bar{t}_d, \underline{x}_d, \underline{t}_d), (\bar{x}_h, \bar{t}_h, \underline{x}_h, \underline{t}_h)\}$, so as to maximize:

$$\begin{aligned}&\gamma[\alpha(\pi(\underline{x}_d) - \underline{t}_d) + (1-\alpha)(\pi(\bar{x}_d) - \bar{t}_d)]\\ &\quad +(1-\gamma)[\alpha(\pi(\bar{x}_h) - \bar{t}_h) + (1-\alpha)(\pi(\underline{x}_h) - \underline{t}_h)].\end{aligned} \tag{7}$$

In the Appendix we show that both of the following ethics incentive compatibility constraints are binding,

$$\begin{aligned}&\alpha[\bar{t}_h - v(\bar{x}_h, \bar{\theta})] + (1-\alpha)[\underline{t}_h - v(\underline{x}_h, \underline{\theta})]\\ &\quad \geq \alpha \max\{\bar{t}_d - v(\bar{x}_d, \bar{\theta}), 0\} + (1-\alpha)\max\{\underline{t}_d - v(\underline{x}_d, \underline{\theta}), 0\}\end{aligned} \tag{8}$$

$$\begin{aligned}&\alpha[\underline{t}_d - v(\underline{x}_d, \bar{\theta})] + (1-\alpha)[\bar{t}_d - v(\bar{x}_d, \underline{\theta})]\\ &\quad \geq \alpha \max\{\bar{t}_h - v(\bar{x}_h, \bar{\theta}), \underline{t}_h - v(\underline{x}_h, \bar{\theta})\}\\ &\qquad +(1-\alpha)\max\{\bar{t}_h - v(\bar{x}_h, \underline{\theta}), \underline{t}_h - v(\underline{x}_h, \underline{\theta})\},\end{aligned} \tag{9}$$

as are the participation constraints corresponding to a bad match, which is given by (5) with $k = h$ for an honest and by

$$\underline{t}_d - v(\underline{x}_d, \bar{\theta}) \geq 0 \tag{10}$$

for a dishonest. Note that because of the tie-breaking rule that an agent who is indifferent between lying and not lying tells the truth, the constraint ensuring that the dishonest with a good match announces a bad match must hold as a strict inequality:

$$\bar{t}_d - v(\bar{x}_d, \underline{\theta}) > \underline{t}_d - v(\underline{x}_d, \underline{\theta}). \tag{11}$$

Even though this constraint cannot be binding, it turns out to be relevant, because it leads to non-existence of an optimal contract with revealing lies for some parameter values, as is specified in the following proposition.

Proposition 2 (Revealing lies) *Suppose that the principal is restricted to contracts inducing the dishonest to always lie in equilibrium. Then, an optimal contract exists only if $\gamma < 1/2$, in which case it is such that $\bar{x}_d^R = \underline{x}_h^R = \underline{x}^*$,*

$$v'(\underline{x}_d^R, \bar{\theta}) = \pi'(\underline{x}_d^R) - \frac{1-\gamma}{\gamma}\frac{1-\alpha}{\alpha}[v'(\underline{x}_d^R, \bar{\theta}) - v'(\underline{x}_d^R, \underline{\theta})]$$

$$v'(\bar{x}_h^R, \bar{\theta}) = \pi'(\bar{x}_h^R) - \frac{\gamma}{1-\gamma}\frac{1-\alpha}{\alpha}[v'(\bar{x}_h^R, \bar{\theta}) - v'(\bar{x}_h^R, \underline{\theta})]$$

and

$$\underline{t}_d^R = v(\underline{x}_d^R, \bar{\theta}) \qquad \bar{t}_d^R = v(\bar{x}_d^R, \underline{\theta}) + [v(\bar{x}_h^R, \bar{\theta}) - v(\bar{x}_h^R, \underline{\theta})]$$

$$\bar{t}_h^R = v(\bar{x}_h^R, \bar{\theta}) \qquad \underline{t}_h^R = v(\underline{x}_h^R, \underline{\theta}) + [v(\underline{x}_d^R, \bar{\theta}) - v(\underline{x}_d^R, \underline{\theta})].$$

For $\gamma < 1/2$ both the honest and the dishonest obtain a rent, however it is lower for the honest thanks to the message reversal. Here the rent-efficiency trade-off is implied by the ethics incentive compatibility constraints being binding. If a dishonest claimed to be honest he would earn a rent by pretending that the match is bad when it is good, while if an honest agent claimed to be dishonest he would earn a rent by telling the truth when the match is good. Hence the rent of an honest depends positively on $\underline{x}_d$, whereas that of a dishonest depends positively on $\bar{x}_h$. This in turn implies that the quantities $\bar{x}_h$ and $\underline{x}_d$ are distorted downwards from $\bar{x}^*$, and depend on the probability that the agent is dishonest. Since for $\gamma < 1/2$ it is more likely that the agent is honest, the downward distortion is more pronounced for $\underline{x}_d$ than for $\bar{x}_h$; moreover, we have $\underline{x}_d < \bar{x}^s < \bar{x}_h$, $\bar{x}_h$ is decreasing and $\underline{x}_d$ increasing in γ, and both tend to $\bar{x}^s$ as γ tends to one half. For $\gamma \geq 1/2$ the downward distortion for $\underline{x}_d$ would be less pronounced than for $\bar{x}_h$ (with equality for $\gamma = 1/2$). The rent of a dishonest agent with a good match being determined by $\bar{x}_h$ through the ethics incentive compatibility constraint, it would not be high enough to prevent him from telling the truth and earn a rent that would be determined by $\underline{x}_d$. In other words (11) would be violated.[9]

Since $\underline{x}_d$ is increasing in γ, the honest agent's rent falls as the probability that the agent is dishonest becomes smaller. Furthermore, the mechanism intended for the honest agent tends toward the first-best mechanism as that probability tends to zero.

In the non-revealing lies case the principal chooses the contract that maximizes her expected surplus, given that the dishonest obtains the same allocation independent of his match.

Proposition 3 (Non-revealing lies) *Suppose that the principal is restricted to contracts such that the dishonest obtains the same allocation (x_d, t_d) independent of his match. Then, in any optimal contract the honest agent obtains the first-best mechanism, and $(x_d, t_d) = (\bar{x}^*, v(\bar{x}^*, \bar{\theta}))$. Furthermore, a dishonest agent always announces $\bar{\theta}$.*

An obvious way to implement the optimal allocations is to offer the first-best mechanism independent of ethics.[10] It is critical that the dishonest agent should announce a bad match in equilibrium. If instead the dishonest agent announced a good match, the honest agent with a good match would derive a rent from the allocation chosen by the dishonest, since that allocation must satisfy the participation constraint for a bad match.

In any contract inducing lies, the optimal mechanism for an honest is such that the dishonest agent would announce a bad match if he were to choose that mechanism. It would clearly not be optimal to design m_h so as to induce the dishonest to reveal θ truthfully, as that would simply bring us back to the incentive-compatible case. It is also intuitive that the principal should not specify m_h so that a dishonest

[9] The non-existence for $\gamma \geq 1/2$ is therefore due to the tie-breaking rule that the agent announces the truth if indifferent. For $\gamma \geq 1/2$, if this tie-breaking rule were not imposed, it can be shown that the optimal contract would implement the standard second-best allocations for both honest and dishonest agents. However, with the standard second-best allocations, the dishonest with a good match is indifferent between the two allocations specified in m_d so that with the tie-breaking rule he would announce a good match.

[10] The allocation corresponding to a good match for the dishonest is in fact indeterminate, since it is never chosen in equilibrium.

with match $\bar{\theta}$ would announce $\underline{\theta}$ in m_h; then there would be no benefit to screening matches for honest agents, since an agent with a good match should be paid as much as an agent with a bad match for a given output.

3.4 The equilibrium contract

Having determined the contracts that are optimal conditional on the three relevant message structures, we can characterize the unconditional equilibrium. The standard second-best and revealing-lies contracts together define the largest surplus that the principal can achieve while screening matches for the dishonest: only the former is relevant for $\gamma \geq \frac{1}{2}$, while the latter is preferred otherwise. Hence to derive the equilibrium contract we need to compare this surplus with that obtained with the first-best contract. Figure 1, where the two relevant surpluses (first-best and match screening) are drawn as a function of γ, depicts the three situations that may arise. We show in the proof of Proposition 4 that the surplus curves are as shown in the picture (Π^S denotes the surplus with the standard second-best contract, Π^R the surplus derived from the revealing lies contract, and Π^* the surplus associated to the first-best contract).

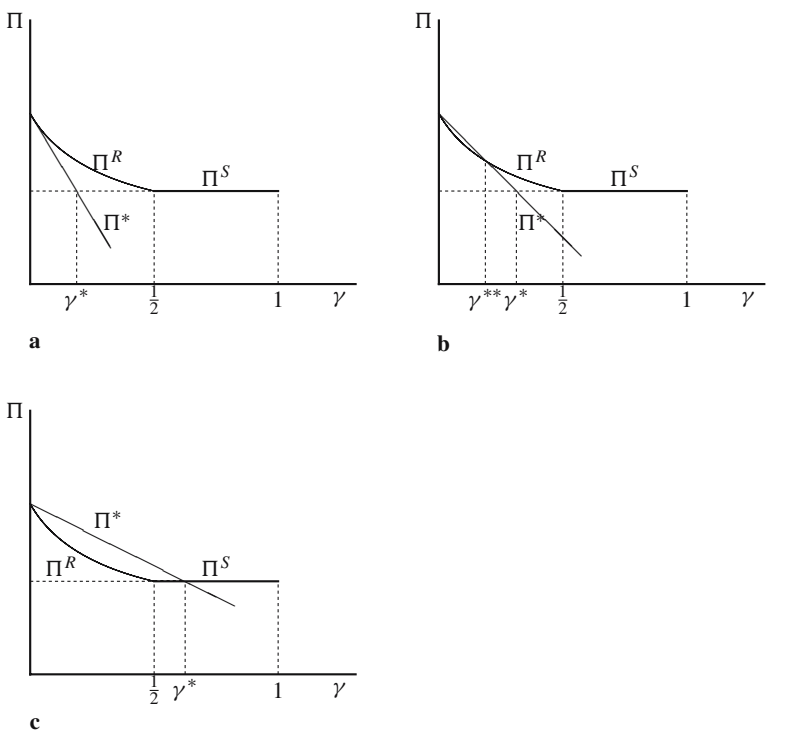

Fig. 1 **a** Case 1 in Proposition 4. **b** Case 2 in Proposition 4. **c** Case 3 in Proposition 4

In panel (a) the first-best surplus curve always lies beneath the match screening surplus curve, so that the equilibrium contract is whichever match screening contract is relevant. As we will see in the proposition below, this occurs when the probability α of a bad match is small. Intuitively this makes sense, since resorting to the first-best contract is costly only when the match is good. By contrast, when this probability is large, offering the first-best contract is always optimal when dishonesty is sufficiently unlikely, as shown in panels (b) and (c). Then the equilibrium contract depends on whether the first-best surplus crosses the second-best surplus below or above $\gamma = \frac{1}{2}$. In the former case, the revealing lies contract is optimal for intermediate values of γ (panel b), whereas in the latter case it is never optimal (panel c). In the proof of the following proposition we show that there exists a unique threshold value γ^* such that the second-best dominates the first-best contract if and only if $\gamma > \gamma^*$; we also find that γ^* is strictly increasing in α, the probability of a bad match, and that it is bounded above by

$$\hat{\gamma} \equiv \frac{v(\bar{x}^*, \bar{\theta}) - v(\bar{x}^*, \underline{\theta})}{[\pi(\underline{x}^*) - v(\underline{x}^*, \underline{\theta})] - [\pi(\bar{x}^*) - v(\bar{x}^*, \bar{\theta})]} < 1. \tag{12}$$

Proposition 4 *There exists a unique* $\alpha_1 \in (0, 1)$ *and a unique* $\alpha_2 \in (\alpha_1, 1]$ *such that:*

1. *If* $\alpha \leq \alpha_1$*: the principal offers the incentive-compatible contract if* $\gamma \geq 1/2$*, and the revealing-lies contract otherwise.*
2. *If* $\alpha \in [\alpha_1, \alpha_2)$*, there exists a unique* $\gamma^{**}(\alpha) \in (0, 1/2)$ *such that the principal offers the incentive-compatible contract if* $\gamma \geq 1/2$*, the revealing-lies contract if* $\gamma \in [\gamma^{**}, 1/2)$*, and the non-revealing lies contract if* $\gamma < \gamma^{**}$*.*
3. *If* $\alpha \geq \alpha_2$*, there exists a unique* $\gamma^*(\alpha) \in (1/2, 1)$ *such that the principal offers the incentive-compatible contract if* $\gamma \geq \gamma^*$*, and the non-revealing lies contract otherwise.*

These results are visualized in Figure 2.[11]

If $\alpha < \alpha_1$, the benefits of screening matches for the dishonest are so large that the honest receives a positive rent for any value of the probability γ that the agent is dishonest: the principal offers the standard second-best contract if $\gamma \geq 1/2$, and the revealing-lies contract otherwise. For intermediary values of α, the benefits from screening matches for the dishonest are smaller; therefore, when γ is sufficiently small the principal forgoes them altogether and leaves no rent to the honest, by offering the first-best contract. However, ethics screening then still occurs for intermediary values of γ, and the standard second-best contract is offered whenever dishonesty is more likely than honesty. As α increases beyond α_2, though, the benefits from leaving no rent to an honest outweigh the benefits from screening matches for a dishonest even if dishonesty is more likely than honesty ($\gamma^* \geq 1/2$): then, ethics screening does not occur at all, since the principal switches from the first-best to the standard second-best contract as the probability of dishonesty increases.

[11] In the figure we have assumed that γ^{**} is a monotonic function of α, and that both γ^* and γ^{**} are linear in α.

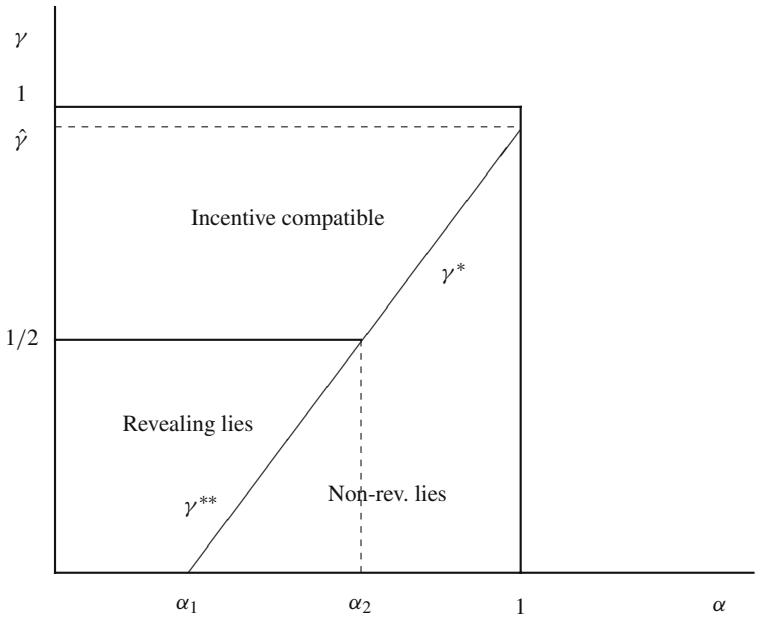

Fig. 2 The optimal contract

Surpluses are affected in intuitive ways. Both the principal and the dishonest agent gain from the presence of honest agents if and only if the probability of honesty rises above the threshold at which the principal drops the standard second-best contract. Whereas the surplus of the principal increases in a continuous manner (see Fig. 1), it involves some discontinuities for the opportunistic agent whenever the principal switches to the first-best contract. By contrast, an honest is penalized by an increase in the likelihood of honesty. In the second-best contract he is treated as well as a dishonest agent, but otherwise he receives a smaller or no rent; if he receives a rent, which is the case in a revealing lies contract, it is decreasing in the likelihood of honesty.

Our results indicate that ruling out ethics screening altogether as in most of the previous literature implies a loss of generality. Here ethics screening arises whenever the principal finds it optimal to use the revealing lies contract. This contract allows the principal to screen among matches for dishonest agents while leaving a smaller rent to the honest than with the second-best contract. In Alger and Renault (2006) we also find that ethics screening may arise, although it requires that there be more than two match realizations. There we study ethics screening when honest behavior is conditional on the fairness of the contract. There is only one period and the principal attempts to simultaneously screen both ethics and matches. The honest behaves honestly only if he expects a low enough probability that a dishonest misrepresents his match in equilibrium. The allocations in the optimal revealing lies contract of Proposition 2 would not be acceptable for an honest agent with fairness concerns: this is because there are two allocations associated with claims that the match is bad that yield different surpluses for an agent whose

match is indeed bad. The allocation yielding the lower surplus would therefore only be chosen by a dishonest agent misrepresenting his match. Thus, with two matches the ability of the principal to screen ethics is greater when honesty is conditional on commitment than when it is conditional on fairness.[12]

Alger and Ma (2003) also find that ethics screening is optimal: they actually find that it is always optimal whenever honesty is sufficiently likely. In their model an insurer (the principal) contracts with a patient and a physician in the first period. A contract specifies the amount of treatment (the decision in our model) and transfers from the patient to the insurer as well as from the insurer to the physician. In the second period, the physician submits a claim to the insurer about the patient's health state; if the physician is opportunistic, he may collude with the patient against the insurer by submitting a false claim. In their setup an exogenous restriction on the contract effectively rules out the possibility of using a revealing-lies contract.[13] As a result, it is clearly the non-revealing lies contract which is optimal whenever γ is sufficiently small. By contrast with our model, the non-revealing lies contract involves ethics screening; furthermore, it never specifies the first-best decision, even for an honest physician. This is due to the insurance motive: the risk aversion of the patient makes it optimal to reduce the difference between the allocations associated with an honest and an opportunistic physician.

Another difference is that, using our notation, in their model the principal's surplus π is directly affected by the match θ. Since in their model the insurance market is competitive, the principal's objective is to maximize the patient's expected utility. As a result, even if the physician always claims that the patient is ill, the objective function still mirrors the fact that the patient is healthy with some probability. We analyzed a version of the present model where the match affects the principal's surplus directly and found that qualitatively, the results are the same as here, except for the decision x specified for an opportunistic agent in the non-revealing lies contract: this decision would then be an increasing function of the probability that the match is good, and it would be equal to $\bar{x}^*$ (as here) only if that probability is zero. However, and this clearly points out the role played by risk aversion in Alger and Ma, the principal would still offer the first-best contract to an honest agent.

4 Concluding remarks

As we saw in the Introduction, modern psychology has taught us that it would not be appropriate to view honesty as an unrestrained desire to reveal hidden information. Rather, it should be dependent upon conditions under which the individual

[12] Nevertheless, the revealing lies contract is a bit unsettling. In a workplace context, if an honest agent were to choose the mechanism meant for the dishonest, he would be asked to commit to producing more with a bad than with a good match. It is questionable whether an honest agent would feel compelled to commit to such an absurd scheme. But maybe more importantly, the ability of the principal to commit to this type of mechanism strongly depends upon her expectation that no honest agent would ever select it; if she assigned a positive probability to such an event, she would have an incentive to renegotiate and offer the standard second-best mechanism instead. This would be accepted by an honest agent since his surplus is larger than what he would obtain by telling the truth in the original mechanism.

[13] The amount of treatment is assumed to be nil whenever the physician claims that the patient is healthy. Hence, if the physician and patient prefer to claim that the patient is ill when in fact he is healthy, they also prefer that claim when the patient is ill.

is led to choose whether to give up the benefits of an opportunistic behavior. Here we have focused on the role of loyalty in inducing honesty. *Ex ante* opportunistic behavior regarding ethics revelation enables an honest agent to sometimes garner a rent which would be denied to him if honesty was unconditional. It also implies that the principal must let the dishonest misrepresent matches if she wants to leave a smaller rent to an honest than to an opportunistic agent. As in Alger and Ma (2003) who use an information structure similar to ours, we find that the principal may choose to screen ethics. However, in contrast with their results, we find that ethics screening is not used if the match between the agent and the task is likely to be bad. Still the potential for screening ethics is greater in the present setup than in Alger and Renault (2006). When honesty is conditional on fairness ethics screening occurs only if there are more than two match realizations.

The results in our companion paper (Alger and Renault 2006) pertaining to situations with more than two matches suggest that increasing the number of match realizations enhances the principal's ability to screen on the basis of ethics. We expect that this would be the case here as well. Furthermore, introducing a larger set of matches may mitigate our result that the standard analysis, where the agent is assumed to be opportunistic with certainty, is robust to the introduction of honest agents. In our companion paper with three match realizations if the intermediate match is sufficiently close to the best match, the standard second-best contract would not be optimal even if honesty is very unlikely.

Appendix

Proof of Lemma 1 For the purpose of this proof, let y denote an allocation (x, t). There is no loss of generality in assuming participation at date 2, since no participation is equivalent to the allocation $(0, 0)$.

Consider some contract which associates a direct mechanism to each message μ_1, through the function m. Now, consider another contract, in which the message space at date 1 is $\{h, d\}$. Let $\hat{k}$ denote the message sent at date 1 in this contract. To each $\hat{k}$, the contract associates a direct mechanism through the function $\tilde{m}$. Each mechanism $\tilde{m}(\hat{k})$ in turn defines an allocation $\tilde{y}$ to each message $\hat{\theta} \in \{\underline{\theta}, \bar{\theta}\}$ sent by the agent at date 2, through the function $\tilde{y}_{\tilde{m}}$. Let this function be such that:

$$\tilde{y}_{\tilde{m}(\hat{k})}(\hat{\theta}) = y_{m(\mu_1^*(\hat{k}))}(\hat{\theta}) \quad \forall \hat{k},\ \forall \hat{\theta}.$$

This implies that:

$$\hat{\theta}^*(\theta, \hat{k}, \tilde{m}(\hat{k})) = \hat{\theta}^*(\theta, \hat{k}, m(\mu_1^*(\hat{k}))) \quad \forall \hat{k},\ \forall \theta.$$

Then, we can show that truth-telling at date 1 is an equilibrium of the new mechanism, by applying the standard technique.

Proof of Proposition 1 Add constraints (1) and (2) to get:

$$\begin{aligned} &\alpha[\bar{t}_h - v(\bar{x}_h, \bar{\theta})] + (1-\alpha)[\underline{t}_h - v(\underline{x}_h, \underline{\theta})] \\ &\geq \alpha \max\{\bar{t}_h - v(\bar{x}_h, \bar{\theta}), \underline{t}_h - v(\underline{x}_h, \bar{\theta})\} \\ &\quad + (1-\alpha)\max\{\underline{t}_h - v(\underline{x}_h, \underline{\theta}), \bar{t}_h - v(\bar{x}_h, \underline{\theta})\} \end{aligned}$$

This implies that the following constraints must hold:

$$\bar{t}_h - v(\bar{x}_h, \bar{\theta}) \geq \underline{t}_h - v(\underline{x}_h, \bar{\theta}) \tag{13}$$

$$\underline{t}_h - v(\underline{x}_h, \underline{\theta}) \geq \bar{t}_h - v(\bar{x}_h, \underline{\theta}) \tag{14}$$

We can add these two constraints to the program, since they are implied by constraints (1) and (2). Let us now omit constraints (1) and (2), and find the solution to the relaxed program: maximize the expected utility subject to the constraints (3)–(6), (13), and (14). By inspection, we note that we can split this program into the following two programs: for $k = d$ and $k = h$, respectively, find $m_k = \{\bar{x}_k, \bar{t}_k, \underline{x}_k, \underline{t}_k\}$ so as to maximize $\alpha(\pi(\bar{x}_k, \bar{\theta}) - \bar{t}_k) + (1 - \alpha)(\pi(\underline{x}_k, \underline{\theta}) - \underline{t}_k)$, subject to the constraint (3) and the relevant part of constraints (5) and (6) for $k = d$, and subject to the constraints (13), (14) and the relevant part of constraints (5) and (6) for $k = h$. These programs are identical in structure, and therefore have the same solution, so that $m_h = m_d$. The omitted constraints (1) and (2) are satisfied. The solution is then trivial, since the maximization problem is well-known from standard principal-agent models.

Proof of Proposition 2 Consider the relaxed program where the expected surplus of the principal is maximized subject to individual rationality constraints (5) with $k = h$ and (10), as well as

$$\alpha[\bar{t}_h - v(\bar{x}_h, \bar{\theta})] + (1 - \alpha)[\underline{t}_h - v(\underline{x}_h, \underline{\theta})] \geq (1 - \alpha)[\underline{t}_d - v(\underline{x}_d, \underline{\theta})], \tag{8′}$$

which is implied by the honest agent's ethics incentive constraint (8), and

$$\alpha[\underline{t}_d - v(\underline{x}_d, \bar{\theta})] + (1 - \alpha)[\bar{t}_d - v(\bar{x}_d, \underline{\theta})] \geq \bar{t}_h - \alpha v(\bar{x}_h, \bar{\theta}) - (1 - \alpha)v(\bar{x}_h, \underline{\theta}), \tag{9′}$$

which is implied by the dishonest agent's ethics incentive constraint (9). We check *ex post* that the solution to the relaxed program satisfies (8) and (9) as well as all other omitted constraints.

We prove that the ethics incentive constraints (8′) and (9′) are binding. First, if (8′) were slack, the principal could increase her expected surplus by decreasing $\underline{t}_h$ (since we have omitted the participation constraint for the honest with a good match). Similarly, if (9′) were slack, the principal could increase her surplus by decreasing $\bar{t}_d$.

We now show that the individual rationality constraint (5) for $k = h$ is binding. Suppose that it were not binding. Then the principal should decrease $\bar{t}_h$ and increase $\underline{t}_h$ so as to keep the expected transfer unchanged; this decreases the right-hand side of (9′) implying that the principal may decrease $\bar{t}_d$.

We now show that the individual rationality constraint (10) is binding. If it were not binding, the principal should decrease $\underline{t}_d$ and increase $\bar{t}_d$ so as to keep the expected transfer unchanged; this decreases the right-hand side of (8′) implying that the principal may decrease $\underline{t}_h$.

The transfers are therefore as in the proposition; replacing them in the objective function and taking first-order conditions yields the quantities in the proposition.

It is easy to verify that this solution satisfies the omitted constraints other than (11) for all the parameter values. In particular, given the specified allocations the

out-of-equilibrium behavior of the agent would be such that the ethics incentive compatibility constraints (8) and (9) could be written as (8′) and (9′).

Constraint (11), however, is satisfied if and only if $\gamma < \frac{1}{2}$. Because the constraint (11) may not be binding any optimal contract must be the solution to a program where it has been ignored: this solution is precisely the one described in the proposition. Hence, whenever the solution does not satisfy the constraint, there is no optimal contract. Note that in this case the set of revealing lies contracts is not a closed subset of $\mathbb{R}^8$, and is therefore not compact; this is the source of the non-existence result.

Proof of Proposition 3 Consider the program where the principal chooses (x_d, t_d), $(\underline{x}_h, \underline{t}_h)$, and $(\bar{x}_h, \bar{t}_h)$ so as to maximize

$$\gamma[\pi(x_d) - t_d] + (1-\gamma)[\alpha(\pi(\bar{x}_h) - \bar{t}_h) + (1-\alpha)(\pi(\underline{x}_h) - \underline{t}_h)],$$

subject to participation constraints for the honest agent, and the participation constraint for the dishonest agent with match $\bar{\theta}$:

$$t_d - v(x_d, \bar{\theta}) \geq 0.$$

The solution to this program yields an upper bound on what can be achieved if only one allocation is implemented for a dishonest. It involves binding all the participation constraints, and setting $x_d = \bar{x}_h = \bar{x}^*$ and $\underline{x}_h = \underline{x}^*$. Besides (x_d, t_d), the mechanism meant for the dishonest agent specifies another allocation; that allocation does not affect the principal's expected surplus as long as it is not chosen by the dishonest over (x_d, t_d). For instance it may be the allocation $(\underline{x}^*, v(\underline{x}^*, \underline{\theta}))$.

This solution satisfies the omitted ethics incentive compatibility constraints if and only if (x_d, t_d) is associated to message $\bar{\theta}$.

Proof of Proposition 4 Let $\Pi^S(\alpha, \gamma)$, $\Pi^*(\alpha, \gamma)$, and $\Pi^R(\alpha, \gamma)$ denote the principal's expected surplus given the optimal incentive-compatible, non-revealing lies, and revealing lies contracts, respectively. Letting $S(x, \theta) = \pi(x) - v(x, \theta)$ denote total surplus at x given match θ, we have:

$$\begin{aligned}
\Pi^S(\alpha, \gamma) &= \alpha S(\bar{x}^s, \bar{\theta}) + (1-\alpha) S(\underline{x}^*, \underline{\theta}) - (1-\alpha)[v(\bar{x}^s, \bar{\theta}) - v(\bar{x}^s, \underline{\theta})] \\
\Pi^*(\alpha, \gamma) &= \gamma \left[\alpha S(\bar{x}^*, \bar{\theta}) + (1-\alpha) S(\bar{x}^*, \underline{\theta}) - (1-\alpha)[v(\bar{x}^*, \bar{\theta}) - v(\bar{x}^*, \underline{\theta})]\right] \\
&\quad + (1-\gamma)[\alpha S(\bar{x}^*, \bar{\theta}) + (1-\alpha) S(\underline{x}^*, \underline{\theta})] \\
\Pi^R(\alpha, \gamma) &= \gamma \left[\alpha S(\underline{x}^R_d, \bar{\theta}) + (1-\alpha) S(\underline{x}^*, \underline{\theta}) - (1-\alpha)[v(\bar{x}^R_h, \bar{\theta}) - v(\bar{x}^R_h, \underline{\theta})]\right] \\
&\quad + (1-\gamma) \Big[\alpha S(\bar{x}^R_h, \bar{\theta}) + (1-\alpha) S(\underline{x}^*, \underline{\theta}) \\
&\quad - (1-\alpha)[v(\underline{x}^R_d, \bar{\theta}) - v(\underline{x}^R_d, \underline{\theta})]\Big].
\end{aligned}$$

First, note that Π^S is constant in γ. Second, Π^* is linear in γ; the slope is equal to

$$\begin{aligned}
&\left[\alpha S(\bar{x}^*, \bar{\theta}) + (1-\alpha) S(\bar{x}^*, \underline{\theta}) - (1-\alpha)[v(\bar{x}^*, \bar{\theta}) - v(\bar{x}^*, \underline{\theta})]\right] \\
&\quad - [\alpha S(\bar{x}^*, \bar{\theta}) + (1-\alpha) S(\underline{x}^*, \underline{\theta})],
\end{aligned}$$

which is strictly negative. Finally consider Π^R. By the envelope theorem:

$$\frac{\partial \Pi^R}{\partial \gamma} = (1-\alpha)\Big[[v(\underline{x}_d^R, \bar{\theta}) - v(\underline{x}_d^R, \underline{\theta})] - [v(\bar{x}_h^R, \bar{\theta}) - v(\bar{x}_h^R, \underline{\theta})]\Big] + \alpha[S(\underline{x}_d^R, \bar{\theta}) - S(\bar{x}_h^R, \bar{\theta})].$$

Here we assume that $\gamma < \frac{1}{2}$, since otherwise there is no optimal revealing lies contract. Thus, $\underline{x}_d^R < \bar{x}_h^R < \bar{x}^*$; moreover, $\underline{x}_d^R$ and $\bar{x}_h^R$ tend to $\bar{x}^s$ as γ tends to $\frac{1}{2}$. Therefore the above expression is strictly negative for $\gamma < \frac{1}{2}$ and it tends to zero as γ tends to $\frac{1}{2}$. Furthermore, $\underline{x}_d^R$ is increasing in γ while $\bar{x}_h^R$ is decreasing in γ. Therefore $S(\underline{x}_d^R, \bar{\theta}) - S(\bar{x}_h^R, \bar{\theta})$ is strictly increasing in γ, and the term multiplying $1-\alpha$ is also increasing in γ. Therefore Π^R is a strictly decreasing and strictly convex function of γ.

A second step is to note a few relations between the expressions Π^S, Π^* and Π^R, and their implications. First, Π^R tends to Π^S as γ tends to $\frac{1}{2}$. Together with the above, this implies that $\Pi^R > \Pi^S$ for $\gamma < \frac{1}{2}$. Second, Π^* is equal to the first-best principal's surplus for as γ tends to 0, and falls below the second-best expected surplus Π^S as γ tends to 1 (because the second-best mechanism is optimal for $\gamma = 1$). Together with the above, this means that for every α, there is a unique $\gamma^* \in (0,1)$ such that $\Pi^* = \Pi^S$, above which $\Pi^* < \Pi^S$ and below which $\Pi^* > \Pi^S$. Finally, note that $\lim_{\gamma\to 0} \Pi^* = \lim_{\gamma\to 0} \Pi^R$.

We now show that the three cases of the proposition may arise. Since Π^* is decreasing and linear in γ, whereas Π^R is decreasing and convex in γ, a necessary and sufficient condition for the non-revealing lies contract to be optimal for some values of γ is that Π^* be less steep than Π^R for γ close to 0. Since we only need to compare $\frac{\partial \Pi^*}{\partial \gamma}$ and $\frac{\partial \Pi^R}{\partial \gamma}$ for γ close to 0, we compare $\frac{\partial \Pi^*}{\partial \gamma}$ to:

$$\lim_{\gamma\to 0} \frac{\partial \Pi^R}{\partial \gamma} = -(1-\alpha)[v(\bar{x}^*, \bar{\theta}) - v(\bar{x}^*, \underline{\theta})] - \alpha S(\bar{x}^*, \bar{\theta}).$$

Π^* is less steep than Π^R for γ close to 0 if $\frac{\partial \Pi^*}{\partial \gamma} > \lim_{\gamma\to 0} \frac{\partial \Pi^R}{\partial \gamma}$, i.e., if:

$$\alpha S(\bar{x}^*, \bar{\theta}) + (1-\alpha) S(\bar{x}^*, \underline{\theta}) + (1-\alpha)[v(\bar{x}^*, \bar{\theta}) - v(\bar{x}^*, \underline{\theta})] - (1-\alpha) S(\underline{x}^*, \underline{\theta}) - (1-\alpha)[v(\bar{x}^*, \bar{\theta}) - v(\bar{x}^*, \underline{\theta})] > 0$$

It is easy to verify that this holds for α close to 1, but not for α close to 0. We now check whether the left-hand side is monotonic in α, so that there exists a unique threshold value α_1 such that the left-hand side is equal to zero. The inequality may be written:

$$S(\bar{x}^*, \bar{\theta}) - (1-\alpha) S(\underline{x}^*, \underline{\theta}) + (1-\alpha)[v(\bar{x}^*, \bar{\theta}) - v(\bar{x}^*, \underline{\theta})] > 0.$$

The derivative of the left-hand side with respect to α is positive if

$$S(\underline{x}^*, \underline{\theta}) - [v(\bar{x}^*, \bar{\theta}) - v(\bar{x}^*, \underline{\theta})] > 0,$$

or, equivalently, if:

$$S(\underline{x}^*, \underline{\theta}) + \pi(\bar{x}^*) - v(\bar{x}^*, \bar{\theta}) > \pi(\bar{x}^*) - v(\bar{x}^*, \underline{\theta}).$$

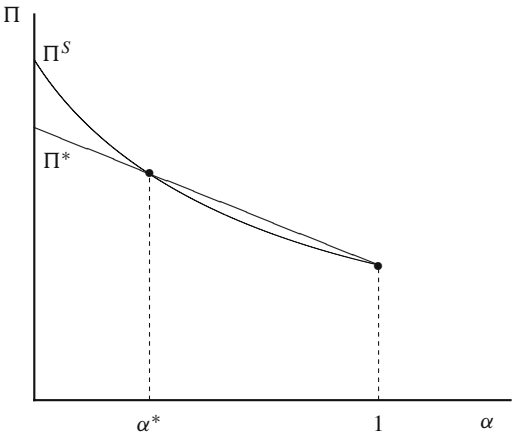

Fig. 3 Monotonicity of γ^*

This is true since the right-hand side is smaller than $S(\underline{x}^*, \underline{\theta})$, by the definition of $\underline{x}^*$.

To show that there exists a unique α_2 as specified in the proposition, it is sufficient to show that γ^* is strictly increasing in α. First note that γ^* tends to 0 as α tends to 0, and that

$$\lim_{\alpha \to 1} \gamma^* \equiv \hat{\gamma} = \frac{v(\bar{x}^*, \bar{\theta}) - v(\bar{x}^*, \underline{\theta})}{[\pi(\underline{x}^*) - v(\underline{x}^*, \underline{\theta})] - [\pi(\bar{x}^*) - v(\bar{x}^*, \bar{\theta})]},$$

which is strictly positive. Therefore it is sufficient to show that γ^* is an invertible function mapping $(0, 1)$ onto $(0, \hat{\gamma})$, with the inverse being strictly increasing. To this end we show that for every $\gamma \in (0, \hat{\gamma})$ there exists a unique $\alpha^* \in (0, 1)$ such that $\Pi^S > \Pi^*$ for $\alpha < \alpha^*$, and $\Pi^S < \Pi^*$ for $\alpha > \alpha^*$, and that α^* is strictly increasing in γ (the arguments for the proof are illustrated in Fig. 3).

Ceteris paribus, Π^S and Π^* are decreasing in α:

$$\frac{\partial \Pi^*}{\partial \alpha} = (1 - \gamma)[\pi(\bar{x}^*) - v(\bar{x}^*, \bar{\theta}) - \pi(\underline{x}^*) + v(\underline{x}^*, \underline{\theta})] < 0,$$
$$\frac{\partial \Pi^S}{\partial \alpha} = \pi(\bar{x}^s) - v(\bar{x}^s, \underline{\theta}) - \pi(\underline{x}^*) + v(\underline{x}^*, \underline{\theta}) < 0,$$

where the second expression is derived using the envelope theorem. Moreover, Π^* is linear, whereas Π^S is convex:

$$\frac{\partial^2 \Pi^S}{\partial \alpha^2} = [\pi'(\bar{x}^s) - v'(\bar{x}^s, \bar{\theta}) + v'(\bar{x}^s, \bar{\theta}) - v'(\bar{x}^s, \underline{\theta})] \frac{\partial \bar{x}^s(\alpha)}{\partial \alpha}.$$

This is positive since both terms are positive ($\bar{x}^s$ being smaller than $\bar{x}^*$). Furthermore, for any $\gamma \in (0, 1)$, $\lim_{\alpha \to 0} \Pi^S(\alpha, \gamma) > \lim_{\alpha \to 0} \Pi^*(\alpha, \gamma)$, whereas

$\lim_{\alpha\to 1} \Pi^S(\alpha, \gamma) = \lim_{\alpha\to 1} \Pi^*(\alpha, \gamma)$. Thus, a necessary and sufficient condition for there to exist a unique α^*, is that

$$\lim_{\alpha\to 1} \frac{\partial \Pi^S}{\partial \alpha} > \lim_{\alpha\to 1} \frac{\partial \Pi^*}{\partial \alpha},$$

which reduces to $\gamma < \hat{\gamma}$. Finally, since Π^S is constant in γ for any value of α, whereas Π^* is decreasing, α^* is strictly increasing.

References

Alger, I., Ma, C.-t. A.: Moral hazard, insurance, and some collusion. J Econ Behav Organ **50**, 225–247 (2003)

Alger, I., Renault, R.: Screening ethics when honest agents care about fairness. Int Econ Rev **47**, 59–85 (2006)

Deneckere, R., Severinov, S.: Mechanism design and communication costs. Mimeo, University of Wisconsin (2001)

Erard, B., Feinstein, J.S.: Honesty and evasion in the tax compliance game. RAND J Econ **25**, 1–19 (1994)

Forges, F.: Koessler, F.: Communication equilibria with partially verifiable types. J Math Econ **41**, 793–936 (2005)

Green, J.R., Laffont, J.-J. : Partially verifiable information and mechanism design. Rev Econ Stud **53**, 447–456 (1986)

Hartshorne, H., May, M.: Studies in the nature of character. New York: Macmillan 1928

Jaffee, D.M., Russell, T.: Imperfect information, uncertainty, and credit rationing. Quart J Econ **90**, 651–666 (1976)

Kofman, F., Lawarrée, J.: On the optimality of allowing collusion. J Public Econ **61**, 383–407 (1996)

Meyer, J.P., Allen, N.J.: A three-component conceptualization of organizational commitment. Hum Resour Manage Rev **1**, 61–89 (1991)

Meyer, J.P., Allen, N.J.: Commitment in the workplace. Thousand Oaks CA: SAGE Publications 1997

Nagin, D., Rebitzer, J., Sanders, S., Taylor, L.: Monitoring, motivation and management: the determinants of opportunistic behavior in a field experiment. Am Econ Rev **92**, 850–873 (2002)

Picard, P.: Auditing claims in the insurance market with fraud: the credibility issue. J Public Econ **63**, 27–56 (1996)

Severinov, S., Deneckere, R.: Does the monopoly need to exclude? Mimeo, University of Wisconsin and Duke University (2003)

Spicer, M.W., Becker, L.A.: Fiscal inequity and tax evasion: an experimental approach. Nat Taxation J **33**, 171–175 (1980)

Terris, W., Jones, J.: Psychological factors related to employees' theft in the convenience store industry. Psychol Rep **51**, 1219–1238 (1982)

Tirole, J.: Collusion and the theory of organizations. In: Laffont, J.-J. (ed.) Advances in economic theory: Proceedings of the Sixth World Congress of the Econometric Society. Cambridge: Cambridge University Press 1992