Information Extraction and Norms of Mutual Protection*

Alberto Bisin†
New York University and NBER

Danilo Guaitoli‡
Universitat Rovira i Virgili

Revised draft: September 2010

*Financial support from the Spanish Ministry of Education and Science and FEDER through grant SEJ2006-03879, and from the Generalitat de Catalunya through grants 2005SGR00447 and 2005SGR00949 is gratefully acknowledged. The paper greatly owes to comments from the Editor and an anonymous referee for substantial improvements in the presentation. Thanks to Vittorio Bisin for his help with American slang.

†Dept. of Economics, New York University; and NBER E-mail: alberto.bisin@nyu.edu

‡Dept. d'Economia, Universitat Rovira i Virgili, Reus (Tarragona), Spain. E-mail: danilo.guaitoli@urv.cat

ABSTRACT

We study a class of moral hazard economies in which a principal interacts with several agents. In these economies first best allocations can be implemented via full information extraction when side-trades between agents can be restricted. When instead side-trades cannot be restricted, the ability of the principal to extract information from the agents is severely hampered. In this context, side-trades take the form of informal contracts which can be directly interpreted as norms of mutual protection, which are indeed quite common among extended family members as well as inside various social, political, and religious groups.

1 Introduction

We study the organization of the activities and information flows in teams. More specifically, we concentrate on moral hazard economies in which a principal interacts with several agents, who in turn participate in a collaborative activity. In such environment, an agent's private effort contributes to the probability distribution of each agent's outcome. Furthermore, each agent observes a signal about other agents' choice of effort. In these economies, any correlation among agents' outputs may be used on the part of the principal to design mechanisms for the extraction of the information about their effort choices. In some particular cases it is well known that first best allocations can be implemented via full information extraction.[1]

We address in particular the robustness of information extraction mechanisms when the principal cannot observe, monitor or contract upon any side-trades among agents.[2] Side-trades among the agents can be formally contracted upon, typically in monetary form, through an intermediary. But, perhaps most importantly, side-trades be supported informally, inside families, social clubs, and institutions. In this case, side-trades take the form of accepted social norms of mutual protection. The adoption of such a norm, with its associated social (non-monetary) punishments for deviations, is in fact very difficult to observe or monitor on the part of the principal.

Consider a team of two agents involved in an activity for a principal. Each individual output is uncertain, subject to external (idiosyncratic) shocks, but certainly affected (in expected terms) not only by the agent's own effort, but also by his/her teammate's effort. The principal can observe agents' outputs but not efforts (not fully, at least). The agents in the team, on the other hand, do observe each other's effort. If the agents receive a fixed compensation regardless of performance, they will not have any incentive to put more than minimum effort. Compensation could instead be made contingent on observed output. It would then provide incentives, at a cost to risk-averse agents and/or to the principal, however. Alternatively, we can imagine a mechanism by which the principal is able to elicit relevant information, rewarding agents for reporting their teammates' effort. In some instances, the most efficient allocation can be supported by means of a compensation scheme of this kind: despite receiving a fixed compensation, either agents would put a high enough effort for fear of being reported to the principal by his/her teammate. Of course the compensation scheme must be such that each agent has an incentive to report truthfully the other agent's effort.

We will show however that the compensation scheme just described is not sustainable in equilibrium unless the principal can restrict side-trades between the agents.

[1] Ma (1988) shows that full information extraction can be implemented as a Perfect Nash Equilibrium of general mechanisms (e.g. sequential mechanisms); see also Ma-Moore-Turnbull (1988) for Nash implementation.

[2] These contractual relationships are called *nonexclusive* in the contract theory literature. In standard principal agent economies (that is, with a single agent) they have been studied by Arnott-Stiglitz (1991), Allen (1985), Hellwig (1983), Kahn-Mookherjee (1998), Cole-Kocherlakota (2001), Bisin-Guaitoli (2004), Bisin-Gottardi-Rampini (2008) and many others.

Suppose the agents face in fact this compensation scheme. There exist then side-trades which act as a pre-commitment on the part of both agents never to report the other's deviation from the effort prescribed by the principal. By adding the side trades to the principal's compensation scheme either agent is able to choose the minimal effort, without his/her choice being revealed to the principal. Therefore, both agents enjoy the fixed compensation provided by the mechanism while saving on the effort cost. We show that, in general, a side-trade which renders information extraction unsustainable has two fundamental properties: i) it punishes agents for revealing the non-prescribed (effort) choice of the teammate, ii) it insures agents against the possibility that the teammate truthfully reveals his/her effort.

Do we actually observe agents entering side-trade agreements of this kind in reality? We propose that norms of mutual protection which (promise to) punish their members for 'telling the principal on each other' and insure them by means of tight social group-protection mechanisms represent in fact such agreements in many social and economic environments of interest. Norms of mutual protection from 'outsiders' are quite common e.g., among extended family members, among members of various social, political, religious, or intellectual groups, etc.

Although hard evidence is difficult to collect, at least anecdoctically we often hear e.g., of people complaining about police officers' cover ups about their colleagues' misbehavior.[3] Similarly, norms of mutual protection are often prevalent e.g., among students with respect to teachers. In fact, the psychological punishment mechanism pertaining to such norms is implicit in the fact that many languages have disparaging words for responsible students who tend to side with teachers and tell on the other students; e.g., *tattletale* in the U.S. and *spione* in Italian. The same norms hold e.g., between trade union members, against industry or firm officials. A revealing, though extreme, example is the case of communist terrorists groups during the 1970's in Italy: they effectively took advantage of the 'code of silence' prevalent in labor unions' culture (they were referred to as "mistaken comrades" and never turned in to the police). The first member of the unions to break this code, Guido Rossa, was killed by the Red Brigades in 1979. Finally, norms of mutual protection are obviously prevalent among members of criminal organizations. These norms include severe punishment for those who become informant of the police or of elected officials. They also typically include various insurance mechanisms to help the members who maintain the code of silence and end up being sentenced as a consequence. A masterly representation of the power of such norms is in the F.F. Coppola movie, *The Godfather II,* when an FBI informant retracts his accusations against the mob during the Senate hearings because, seeing his old brother at the hearings, he is confronted with the dishonor that his behavior would dispel on his whole family. Once again languages have developed derogatory words those who break norms of mutual protection in criminal organizations, e.g., *dobbers* and *snitches* in the American slang, *infami* in Italian. Finally, the recently discovered

[3] On October 8th 2008 a Google search for "police cover up" generated about 1,820,000 results, with stories of alleged cover ups in many countries, from China to the U.S.

behavior of the Catholic church which has apparently protected for years its members from sex-abuse scandals also fits well the profile of a social protection norm.[4]

More generally, informants within teams are exceptions, usually ostracized, harassed and subjected to social sanctions by their teammates. For this reason, in many private and public institutions whistleblowers rarely come forward and need to be formally protected when they do. The *Public Interest Disclosure Act 1998* in the U.K and the *Whistleblower Protection Act 2007* for U.S. federal employees are typical examples.

In summary, social norms penalizing informants and compensating their victims seem widespread. Three main points arise from our analysis. First, we show that such norms of mutual protection can be represented (in a stylized way) in terms of informal contracts which imply (monetary or non-monetary) transfers among agents contingent on agents' reports to the principal. Second, we show that the possibility of side-trades are really the key factor that allows such norms to offset information extraction mechanisms. If the main compensation scheme could be made contingent on the absence of any side-trades, information extraction would obtain. But it is nearly impossible for a principal to observe or verify the occurrence of side-trades when these are informal expressions of a social norm. Third, collusion (in the sense of agents acting cooperatively or in a coordinated way as a coalition) is not necessary for this behavior to take place. Even if each agent decides voluntarily, unilaterally, whether to join or not the informal social contract, taking all other decisions non-cooperatively, and social norms are not forced upon agents who do not want to be part of it (so that an agent who did not enter the informal side-contract will be left alone even if he/she behaves as an informant), side-contracts can be made attractive enough so that each agent will unilaterally choose to join. In this paper in fact we require that side-trades constitute a Nash-equilibrium for the agents. A related literature[5] shows that, full information extracting contracts are not sustainable if agents can collude. Brusco (1997), the closest to our approach, allows for joint deviations of the agents (collusion), but requires them not to contract on reports to the principal, therefore limiting the class of colluding contracts. In this paper, we instead allow agents to contract on reports to the principal, but we do not allow for joint deviations.

Our results are limited in one fundamental respect: we are not able to develop a complete equilibrium characterization in the case side-trades cannot be monitored. Equilibria in these environments are typically sustained by a complex menu of contracts, including so-called *latent contracts* which are issued but are not active in equilibrium, as studied by Hellwig (1983) and Bisin-Guaitoli (2004) in a principal-agent economy with a single agent. We offer several insights about the properties of equilibrium, however, in the Concluding remarks.

[4] See e.g., the documents collected in http://en.wikipedia.org/wiki/Catholic_sex_abuse_cases

[5] See e.g., Itoh (1993), Brusco (1997), Miller (1997), Baliga-Sjostrom (1998) and many others.

2 The economy

The economy we study is as follows. Agents value consumption (positively) and effort (negatively): $u(c) - e$, with $u : \Re_+ \rightarrow \Re$ continuous, strictly increasing, strictly concave, and unbounded below. Effort can take two values, $e \in \{a, b\}$, with $a > b$. Agent i's output w_{s^i} takes values w_H or w_L, with $w_H > w_L$. Agents are grouped in teams $(i, X(i))$ (X is a one-to-one and onto map).[6] The probability distribution of agent i's output is affected by both i's and $X(i)$'s effort. We denote with $\pi_{e^i, e^{X(i)}}$ the probability that agent i has high output w_H, given that he chose effort e^i and his teammate $X(i)$ chose effort $e^{X(i)}$. The probability of high output for agent i is increasing in e^i as well as in $e^{X(i)}$. Agents' effort is private information, in particular it is not observed by any principal. But agent i receives a private information signal which fully reveals his teammate's effort $e^{X(i)}$ (and symmetrically for agent $X(i)$). Agent i then sends a message m^i with value in $\{a, b\}$ about $e^{X(i)}$ (and symmetrically for agent $X(i)$). Finally, contract j's payoff to agent i is denoted $d^j = \{d^j_{H, m^i, m^{X(i)}}, d^j_{L, m^i, m^{X(i)}}\}$: it depends on the agent's realized output and on both agents' messages regarding their teammate's effort.[7] The principal is risk-neutral: the expected profit of a contract j, for given effort choices $(e^i, e^{X(i)})$ and messages $(m^i, m^{X(i)})$, is $-[\pi_{e^i, e^{X(i)}} d^j_{H, m^i, m^{X(i)}} + (1 - \pi_{e^i, e^{X(i)}}) d^j_{L, m^i, m^{X(i)}}]$ (note the usual abuse of the law of large number adopted here).

3 Information extraction

We are now ready to characterize the incentive constrained optimum for the information extraction economy when side-traits can be restricted by the principal. In this case, without loss of generality, we consider optimal allocations that give agents all the surplus and zero profit to the principal. Besides being the most common reference point in the literature, it is also, intuitively, the allocation attained in equilibrium when principals compete in a market with free entry.[8]

Define a *Pareto optimum* as an allocation (c^i_L, c^i_H, e^i), $\forall i$, that maximizes agents' expected utility subject to non-negative expected profits for the principal. It obviously implies full insurance at fair prices for agents' consumption at the efficient effort levels $E = (e^i, e^{X(i)})^{PO}$, i.e. $c_H = c_L = \pi_E w_H + (1 - \pi_E) w_L$.[9]

[6] Agent i and $X(i)$ are symmetric. For simplicity we report the notation only for agent i.

[7] In general payoffs should be allowed to depend also on the teammate's realized outcome. Here we omit this dependence for simplicity, since (as we prove) it is not needed for the optimal contract to reach the first best (Proposition 1), while it would not change qualitatively the main consequence of nonexclusivity, i.e. the impossibility of full information extraction (Proposition 2).

[8] If side-trades cannot be monitored, however, the whole costrained efficient frontier becomes relevant, as equilibria typically provide the principal with positive profits; see Bisin and Guaitoli (2004).

[9] If agents have some alternative opportunities, we should take into account also a participation constraint. But according to our definition, the Pareto optimum gives the highest possible expected utility to agents: if that does not satisfy the participation constraint, then trivially no transaction will take place. So we assume that the participation constraint is satisfied at the first best, in which case

Define an *incentive constrained optimum* as an allocation (c^i_L, c^i_H, e^i), $\forall i$, that maximizes agents' expected utility subject to: i) non-negative expected profits for the principal, ii) incentive constraints that guarantee truth-telling in messages and choice of (constrained) efficient effort by each agent. That is, given the contingent payoffs, no agent has an incentive to lie about their teammate's effort and no agent has an incentive to unilaterally deviate from the prescribed effort.

The following proposition characterizes the incentive constrained optimum showing how, under certain conditions, it attains the Pareto optimal (first-best) allocation.

Proposition 1 *In the information extraction economy with no side-trades, assume*

$$\frac{(1-\pi_{aa})}{\pi_{aa}} < \frac{(1-\pi_{ba})}{\pi_{ba}} < \frac{(1-\pi_{ab})}{\pi_{ab}} < \frac{(1-\pi_{bb})}{\pi_{bb}}. \tag{1}$$

Then the incentive constrained optimal allocation is unique and achieves the Pareto optimum:

$$c_H = c_L = \pi_E w_H + (1-\pi_E) w_L$$

where $E = (e^i, e^{X(i)})^{PO}$ *is the effort choice at the Pareto optimum.*

Proof. See Appendix.

< *Figure 1* >

The proof in the Appendix constructs a contract which implements full information extraction (the only non-trivial case is when at the Pareto optimum both agents choose the high effort, $E = (a, a)$). Figure 1 shows the structure of the game where agents first simultaneously choose effort, then (again simultaneously) messages on their teammate's effort. Payoffs are determined by the optimal contract so to induce truth-telling in the message subgame and they depend on messages in a non-trivial way: $d_{s,M} \neq d_{s,M'}$, $\forall s \in \{H, L\}$, $M, : M' \in \{a, b\}^2$, where $M = (m^i, m^{X(i)})$. Each agent has a dominant strategy in the message subgame, so that each subgame has a unique equilibrium; by backward induction, we get an effort game (Figure A1 in the Appendix) which, for an appropriate choice of the contract's payoffs, has the efficient outcome as the unique equilibrium.

< *Figure 2* >

We can get the idea of how the contract works with the help of Figure 2, which shows agents' consumption (and output w) in the two states, low (L) and high (H). There are four indifference curves going through each point, corresponding to the possible pairs

we don't have to worry about it even in the incentive constrained optimum which, as we show below, gives agents the same first best payoff.

of agents' effort (which affect the probabilities of the two states). In equilibrium agents get the fair full-insurance consumption allocation c^{FI} under high effort. An agent reported as having put low effort sees those transfers reduced by a certain penalty in both states (so that his or her utility is reduced by P). Then there must be an incentive for the agent to report a teammate's low effort if and only if that teammate actually put low effort: this is achieved by adding a state contingent transfer for the agent who reports low effort which increases his expected utility (compared to c^{FI}) if and only if he is telling the truth, otherwise decreasing his expected utility. These transfers take the agent to a point like $w + d_{ba}$ (resp. $w + d_{bb}$) in Figure 2 which, compared with c^{FI} (resp. c^P), lies above the indifference curves for efforts (a, b) or (b, b) and below the indifference curves for efforts (a, a) or (b, a). Condition (1) is sufficient to guarantee that such contract payoffs, giving agent i a positive expected value when $X(i)$ has chosen the low effort, and a negative expected amount when $X(i)$ has chosen the high effort, do in fact exist. Moreover the reward for truthfully reporting the low effort of the other agent in the group is never paid by the principal in equilibrium, since both agents choose the high effort.

4 Side-trades

We now consider the information extraction economy under the condition that the principal is unable to monitor agents' side-trades, or that information about such trades is not verifiable. For expositional purpose it is convenient to think of side-trades as formal contracts (pooled, e.g., by an intermediary, to exploit the law of large numbers). This should not distract the reader from our preferred interpretation of side-trades as commonly represented by social norms of mutual protection.

Proposition 2 shows that there exists an additional contract (which we construct in the proof) inducing both i and $X(i)$ never to report the other's low effort. By entering both this contract and the incentive constrained optimal contract, agents are able to choose low effort without being caught, and hence to enjoy full insurance at the (better than fair) price $(1 - \pi_{aa})/\pi_{aa}$, while saving on the effort cost.

Referring to the contract which implements the Pareto optimum (as in the proof of Proposition 1) as the full information extraction contract d, we have the following result.

Proposition 2 *In the information extraction economy with side-trades, assume condition (1) and $E = (a, a)$ at the Pareto optimum. There exists then a contract d' such that, if the set of contracts $\{d, d'\}$ is available:*

- *all agents accept both contracts*
- *all agents choose low effort, i.e. $(e^i, e^{X(i)}) = (b, b)$*

- *each agent i reports high effort for agent $X(i)$ and viceversa, i.e.* $(m^i, m^{X(i)}) = (a, a)$.

Proof. See Appendix.

< *Figure 3* >

We interpret contract d' as a side-trade between agents. Figure 3 shows the structure of the game when all agents accept both contracts (analogously one can define the game when only one of the agents accepts both contracts, the other entering only the first contract - see the proof in the Appendix). Contract d' essentially insures agent i against the possibility of being reported by agent $X(i)$ as having chosen low effort, and punishes agent i for revealing the low effort choice of $X(i)$. Referring again to Figure 2, contract d''s contingent transfers can be seen as the sum of two parts: the first part exactly offsets the corresponding transfers of contract d (the ones that go from c^{FI} (resp. c^P) to $w + d_{ba}$ (resp. $w + d_{bb}$)); the second part takes the agent to a point like $c^{FI} + d''$ (resp. $c^P + d''$) which, compared with c^{FI} (resp. c^P), lies below the indifference curve for efforts (b, b). The result is that agents will not report their teammate's low effort, thereby having the opportunity of choosing low effort without losing the first best consumption allocation c^{FI}.

Notice that we are not introducing any collusion or cooperative behavior among agents. Each agent always behaves non-cooperatively, not only in the choice of messages and effort, but also in the portfolio choice of contracts. So after analizing the games corresponding to all portfolio choices, we look at the contract game where each agent non-cooperatively chooses whether or not to enter also the second contract. It turns out that for each agent it is a dominant strategy to enter both contracts, regardless of what the other does. If an agent did not subscribe voluntarily to d', no penalty or transfer other than those implied by d would be imposed on him.

Whenever contracts d and d' are simultaneously available, then, the incentive constrained optimal contract d will make negative profits, while contract d' will make zero profits (d' can obviously be perturbed to generate small positive profits).

Finally, we stress that our characterization of contract d' satisfies the two fundamental properties we have identified in the Introduction (which allow us to interpret it as a norm of mutual protection): i) it punishes agents for revealing the non-prescribed (effort) choice of the teammate, ii) it insures agents against the possibility that the teammate truthfully reveals his/her effort.

5 Concluding remarks

The implications of these results go beyond the specific economy we have considered here. First of all, the results are not sensitive to the details of the extensive form of the game. As shown in the Appendix, for a suitable choice of the contracts' payoffs, each agent has a dominant strategy in the message subgame, in the effort game and even in the contract choice game. Hence we would obtain exactly the same results derived here if the game were one in which agents chose sequentially.

A second element suggesting that the results are likely to be robust with respect to other information extraction mechanisms is the following. The proof of Proposition 2 shows that the transfers of the implicit contract among agents can be decomposed in two parts: one exactly offsets the principal's transfers, the other gives agents the opposite incentives (i.e., to shirk and lie). Now the first component is obviously always possible; the second component is subject to conditions, but such conditions (condition (1) in our economy) are exactly the same under which the proposed information extraction mechanism is constructed. In other words, the existence of a second (possibly implicit) contract that undermines the incentives of the principal's contract does not require any additional condition besides the ones that make an information extraction mechanism possible with no side-trades. This suggests that many other potential mechanisms will be vulnerable to nonexclusivity in a way similar to the one described here.

No full characterization of equilibrium allocations with side-trades is available for economies with interacting agents like the one discussed here. Nonetheless, we conjecture that such equilibria will rely on contracts like the ones we constructed which break information extraction (and which we interpret as norms of mutual protection). This insight is motivated by the observation that all studies on economies with side-trades (that is, nonexclusive contracts, like our Bisin-Guaitoli (2004) as well as others) include the following results: 1) sometimes it is not possible to support high effort even if it would be efficient under exclusivity conditions; 2) when it is possible, it necessarily requires additional contracts that are available but not active in equilibrium (these so-called *latent contracts* serve the purpose of deterring entry of other contracts which would undermine the incentives of the active contracts); 3) even when this is done, the resulting allocation is inefficient compared with the exclusive case (e.g. it leaves more risk to agents).

In conclusion, we argue that social norms penalizing informants and compensating their victims seem widespread in contexts of agents working in teams; that those norms can be represented in terms of informal (implicit) contracts among agents; that the possibility of such norms naturally creates conditions where monitoring of (nonmonetary) side-trades cannot be enforced; that in these conditions information extraction mechanisms will tipically fail to be sustainable; that these results are obtained in a purely non-cooperative way, i.e. even if agents cannot collude or establish cooperative agreements. Empirically, we believe that these implications are relevant in a variety of contexts where agents interact, in particular in relatively symmetric and homogeneous conditions, where individual outcomes are significantly affected by other agents' actions and individual behaviour is common information among agents. Examples include public and private institutions like the police, the military, burocratic administrations, firms based on team production, unions, etc. These problems may explain on one hand why it is not so common to see agents informing on their teammates misbehaviour, and on the other why specific laws protecting whistleblowers have been enacted in some countries.

References

Allen, Franklin (1985). "Repeated Principal-Agent Relationships with Lending and Borrowing." *Economics Letters*, 17, 27-31.
Arnott, Richard, and Joseph Stiglitz (1991). "Equilibrium in Competitive Insurance Markets with Moral Hazard." *NBER Working Paper* 3588.
Baliga, S. and T. Sjostrom (1998): "Decentralization and Collusion," *Journal of Economic Theory,* 83, 196-232.
Bisin, A., P. Gottardi, and A. Rampini (2008): "Managerial Hedging and Portfolio Monitoring," *Journal of the European Economic Association*, 6, 158-209.
Bisin, A., and D. Guaitoli (2004): "Moral Hazard and Nonexclusive Contracts," *Rand Journal of Economics*, 35(2), 306-28.
Bizer, David S., and Peter M. DeMarzo (1992). "Sequential Banking." *Journal of Political Economy*, 100, 41-61.
Brusco, S. (1997): "Implementing Action Profiles when Agents Collude," *Journal of Economic Theory,* 73, 395-424.
Cole, Harold L., and Narayana R. Kocherlakota (2001): "Efficient Allocations with Hidden Income and Hidden Storage," *Review of Economic Studies*, 68, 523-542.
Hellwig, M. (1983): "On Moral Hazard and Non-price Equilibria in Competitive Insurance Markets," *mimeo*, University of Bonn.
Itoh, H. (1993): "Coalitions, Incentives, and Risk Sharing," *Journal of Economic Theory,* 60, 410-27.
Kahn, Charles M., and Dilip Mookherjee (1998). "Competition and Incentives with Nonexclusive Contracts." *RAND Journal of Economics* 29, 443-465.
Ma, C.-t. A. (1988): "Unique Implementation of Incentive Contracts with Many Agents", *Review of Economic Studies*, 55, 555-71.
Ma, C.-t. A., J. Moore, and S. Turnbull (1988): "Stopping Agents from Cheating,", *Journal of Economic Theory,* 46, 355-72.
Miller, N.H. (1997): "Efficiency in Partnership with Joint Monitoring,", *Journal of Economic Theory,* 77, 285-99.

Appendix

Proof of Proposition 1. The proof is by construction of an optimal contract d (we concentrate on the only non-trivial case in which, at the Pareto optimum, $E = (a, a)$). Take $d_{s,a,a} = d_s^{FI}$, $\forall s \in S$, such that

$$w_H + d_H^{FI} = w_L + d_L^{FI} \quad \text{and} \quad \pi_{a,a} d_H^{FI} + (1 - \pi_{a,a}) d_L^{FI} = 0$$

i.e. d_s^{FI} are the payoffs associated with the (full insurance) Pareto optimal contract. Also take $d_{s,a,b} = d_s^{FI} - P$, $d_{s,b,a} = d_s^{FI} + \hat{d}_s$, $d_{s,b,b} = d_s^{FI} + \hat{d}_s - P$, where $\hat{d} = (\hat{d}_H, \hat{d}_L)$ satisfies:

$$\begin{aligned}
\pi_{a,a} u^i(w_H + d_H^{FI} + \hat{d}_H) + (1 - \pi_{a,a}) u^i(w_L + d_L^{FI} + \hat{d}_L) &= U^{FI} - \epsilon_{a,a} \\
\pi_{a,b} u^i(w_H + d_H^{FI} + \hat{d}_H) + (1 - \pi_{a,b}) u^i(w_L + d_L^{FI} + \hat{d}_L) &= U^{FI} + \epsilon_{a,b} \\
\pi_{b,a} u^i(w_H + d_H^{FI} + \hat{d}_H) + (1 - \pi_{b,a}) u^i(w_L + d_L^{FI} + \hat{d}_L) &= U^{FI} - \epsilon_{b,a} \\
\pi_{b,b} u^i(w_H + d_H^{FI} + \hat{d}_H) + (1 - \pi_{b,b}) u^i(w_L + d_L^{FI} + \hat{d}_L) &= U^{FI} + \epsilon_{b,b}
\end{aligned}$$

with $\epsilon_{hk} > 0$, $h, k \in \{a, b\}$; and $U^{FI} = u^i(w_H + d_H^{FI}) = u^i(w_L + d_L^{FI})$. A sufficient condition for $\hat{d}$ to exist (i.e. for the equations above to be satisfied) is

$$\frac{(1 - \pi_{a,a})}{\pi_{a,a}} < \frac{(1 - \pi_{b,a})}{\pi_{b,a}} < \frac{(1 - \pi_{a,b})}{\pi_{a,b}} < \frac{(1 - \pi_{b,b})}{\pi_{b,b}}$$

(construct $\hat{d}$ to have negative expected value if $(e, e) = (a, a)$ or if $(e, e) = (b, a)$, and positive expected value if $(e, e) = (a, b)$ or $(e, e) = (b, b)$). For $\hat{d}$ small enough there exist then $\omega_{hk} > 0$, $h, k \in \{a, b\}$, such that with messages (b, b)

$$\begin{aligned}
\pi_{a,a} u^i(w_H + d_H^{FI} + \hat{d}_H - P) + (1 - \pi_{a,a}) u^i(w_L + d_L^{FI} + \hat{d}_L - P) &= U^P - \omega_{a,a} \\
\pi_{a,b} u^i(w_H + d_H^{FI} + \hat{d}_H - P) + (1 - \pi_{a,b}) u^i(w_L + d_L^{FI} + \hat{d}_L - P) &= U^P + \omega_{a,b} \\
\pi_{b,a} u^i(w_H + d_H^{FI} + \hat{d}_H - P) + (1 - \pi_{b,a}) u^i(w_L + d_L^{FI} + \hat{d}_L - P) &= U^P - \omega_{b,a} \\
\pi_{b,b} u^i(w_H + d_H^{FI} + \hat{d}_H - P) + (1 - \pi_{b,b}) u^i(w_L + d_L^{FI} + \hat{d}_L - P) &= U^P + \omega_{b,b}
\end{aligned}$$

with $U^P = u^i(w_H + d_H^{FI} - P) = u^i(w_L + d_L^{FI} - P)$.
We do not need to impose market clearing conditions on $\hat{d}$ since in equilibrium, we will show that, given d, $E = (e^i, e^{X(i)}) = (a, a)$. The game played by agents is represented in Figure 1.
Each agent has a dominant strategy in the message subgame, so that each subgame has a unique equilibrium; the reader can check that the unique Nash Equilibria of the four simultaneous message games (from top to bottom) are respectively: (a, a); (b, a); (a, b); (b, b). By backward induction, we have then the following first-stage simultaneous effort game:

< Figure A1 >

For $(e^i, e^{X(i)}) = (a,a)$ to be a Nash equilibrium, we need $U^{FI} - a > U^P - b$, i.e. $U^{FI} - U^P > a - b$, which is satisfied with a penalty P large enough (with a utility function unbounded below, we don't have to worry about the non-negativity constraints on consumption); it will be unique if $U^{FI} - a + \epsilon_{a,b} > U^P - b + \omega_{b,b}$, i.e. $U^{FI} - U^P > a - b + (\omega_{b,b} - \epsilon_{a,b})$, which is satisfied for $\hat{d}$ small enough (and therefore ω's and ϵ's small enough).
This contract then induces all agents to choose high effort and send truthfull messages, thereby attaining the first best allocation with payoff $U^{FI} - a$. $\diamondsuit$

Proof of Proposition 2. The proof is by construction. Construct a contract d' as follows: $d'_{s,a,a} = d'_{s,a,b} = 0$ and $d'_{s,b,a} = d'_{s,b,b} = -\hat{d}_s + d''_s$, where the d''_s are such that

$$\begin{aligned}
\pi_{a,a} u^i(w_H + d^{FI}_H + d''_H) + (1-\pi_{a,a}) u^i(w_L + d^{FI}_L + d''_L) &= U^{FI} + \delta_{a,a} \\
\pi_{a,b} u^i(w_H + d^{FI}_H + d''_H) + (1-\pi_{a,b}) u^i(w_L + d^{FI}_L + d''_L) &= U^{FI} + \delta_{a,b} \\
\pi_{b,a} u^i(w_H + d^{FI}_H + d''_H) + (1-\pi_{b,a}) u^i(w_L + d^{FI}_L + d''_L) &= U^{FI} + \delta_{b,a} \\
\pi_{b,b} u^i(w_H + d^{FI}_H + d''_H) + (1-\pi_{b,b}) u^i(w_L + d^{FI}_L + d''_L) &= U^{FI} - \delta_{b,b}.
\end{aligned}$$

Such d''_s exist if

$$\frac{(1-\pi_{a,a})}{\pi_{a,a}} < \frac{(1-\pi_{b,a})}{\pi_{b,a}} < \frac{(1-\pi_{a,b})}{\pi_{a,b}} < \frac{(1-\pi_{b,b})}{\pi_{b,b}}.$$

Again, for d''_s small enough there exist $\gamma_{hk} > 0$, $h,k \in \{a,b\}$, such that

$$\begin{aligned}
\pi_{a,a} u^i(w_H + d^{FI}_H + d''_H - P) + (1-\pi_{a,a}) u^i(w_L + d^{FI}_L + d''_L - P) &= U^P + \gamma_{a,a} \\
\pi_{a,b} u^i(w_H + d^{FI}_H + d''_H - P) + (1-\pi_{a,b}) u^i(w_L + d^{FI}_L + d''_L - P) &= U^P + \gamma_{a,b} \\
\pi_{b,a} u^i(w_H + d^{FI}_H + d''_H - P) + (1-\pi_{b,a}) u^i(w_L + d^{FI}_L + d''_L - P) &= U^P + \gamma_{b,a} \\
\pi_{b,b} u^i(w_H + d^{FI}_H + d''_H - P) + (1-\pi_{b,b}) u^i(w_L + d^{FI}_L + d''_L - P) &= U^P - \gamma_{b,b}.
\end{aligned}$$

Consider first the case in which all agents $(i, X(i))$ enter both contracts (d, d') (see the game described in Figure 3). As in the proof of Proposition 1, we can proceed by backward induction to show that the unique Nash equilibria of the message subgames are respectively: (b,b); (b,b); (b,b); (a,a). This gives us the following effort game:

< Figure A2 >

$(e^i, e^{X(i)}) = (b,b)$ is a Nash equilibrium if $U^{FI} - b > U^P - a + \gamma_{a,b}$, i.e. $\gamma_{a,b} < (U^{FI} - U^P) + (a - b)$; moreover it is unique if $U^P - b + \gamma_{b,a} > U^P - a + \gamma_{a,a}$, i.e. $\gamma_{a,a} - \gamma_{b,a} < a - b$. Both conditions are satisfied for d''_s (and therefore γ's) small enough.
Similarly, we can look at the games where only one of the agents enters both contracts, while the other only has contract d (these can be defined by giving the agent with both contracts the payoffs from Figure 3

and the agent with only the first contract the payoffs from Figure 1).

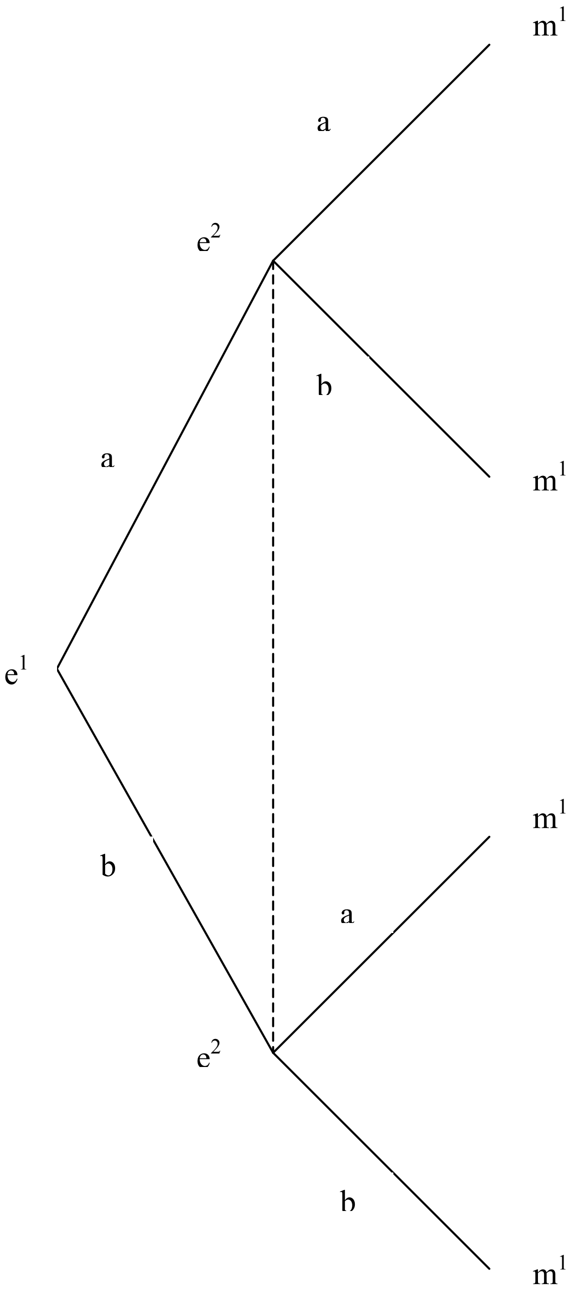
It can be checked that when only one agent has both contracts, the equilibria in the message subgames are respectively (b, a), (b, a), (b, b), (a, b). The Nash equilibrium of the corresponding effort game is $(e^i, e^{X(i)}) = (a, b)$ (unique if $\gamma_{b,a} - \delta_{a,a} < (U^{FI} - U^P) - (a - b)$, which is again satisfied for d''_s, and therefore γ's and δ's, small enough). Simmetrically, when only the other agent has both contracts, the equilibria in the message subgames are respectively (a, b), (b, b), (a, b), (b, a) and the equilibrium efforts are $(e^i, e^{X(i)}) = (b, a)$. Note that the agent with only contract d shirks to low effort, since the other (subscribing to d' as well as d) will not report him, while the agent with both contracts does not shirk (otherwise his teammate, not subscribing to d', would report him).

Now we can determine the contract choice when each agent non-cooperatively chooses whether or not to enter also the second contract. Using the equilibrium payoffs of the corresponding games, the contract game is the following:

< *Figure A3* >

The unique equilibrium is $(d + d', d + d')$, since for each agent it is a dominant strategy to enter both contracts, regardless of what the other does. Hence agents will play $(e^i, e^{X(i)}) = (b, b)$ and get a payoff $U^{FI} - b$. Contract d' makes zero profits (no transfers in equilibrium), while contract d makes negative expected profits. ♢

Figure 1



	m^2: a	m^2: b
m^1: a	U^{FI} - a U^{FI} - a	U^P - a U^{FI} - a - ε_{aa}
m^1: b	U^{FI} - a - ε_{aa} U^P - a	U^P - a - ω_{aa} U^P - a - ω_{aa}

	m^2: a	m^2: b
m^1: a	U^{FI} - a U^{FI} - b	U^P - a U^{FI} - b - ε_{ba}
m^1: b	U^{FI} - a + ε_{ab} U^P - b	U^P - a - ω_{ab} U^P - b - ω_{ba}

	m^2: a	m^2: b
m^1: a	U^{FI} - b U^{FI} - a	U^P - b U^{FI} - a + ε_{ab}
m^1: b	U^{FI} - b - ε_{ba} U^P - a	U^P - b - ω_{ba} U^P - a + ω_{ab}

	m^2: a	m^2: b
m^1: a	U^{FI} - b U^{FI} - b	U^P - b U^{FI} - b + ε_{bb}
m^1: b	U^{FI} - b + ε_{bb} U^P - b	U^P - b + ω_{bb} U^P - b + ω_{bb}

Figure 2

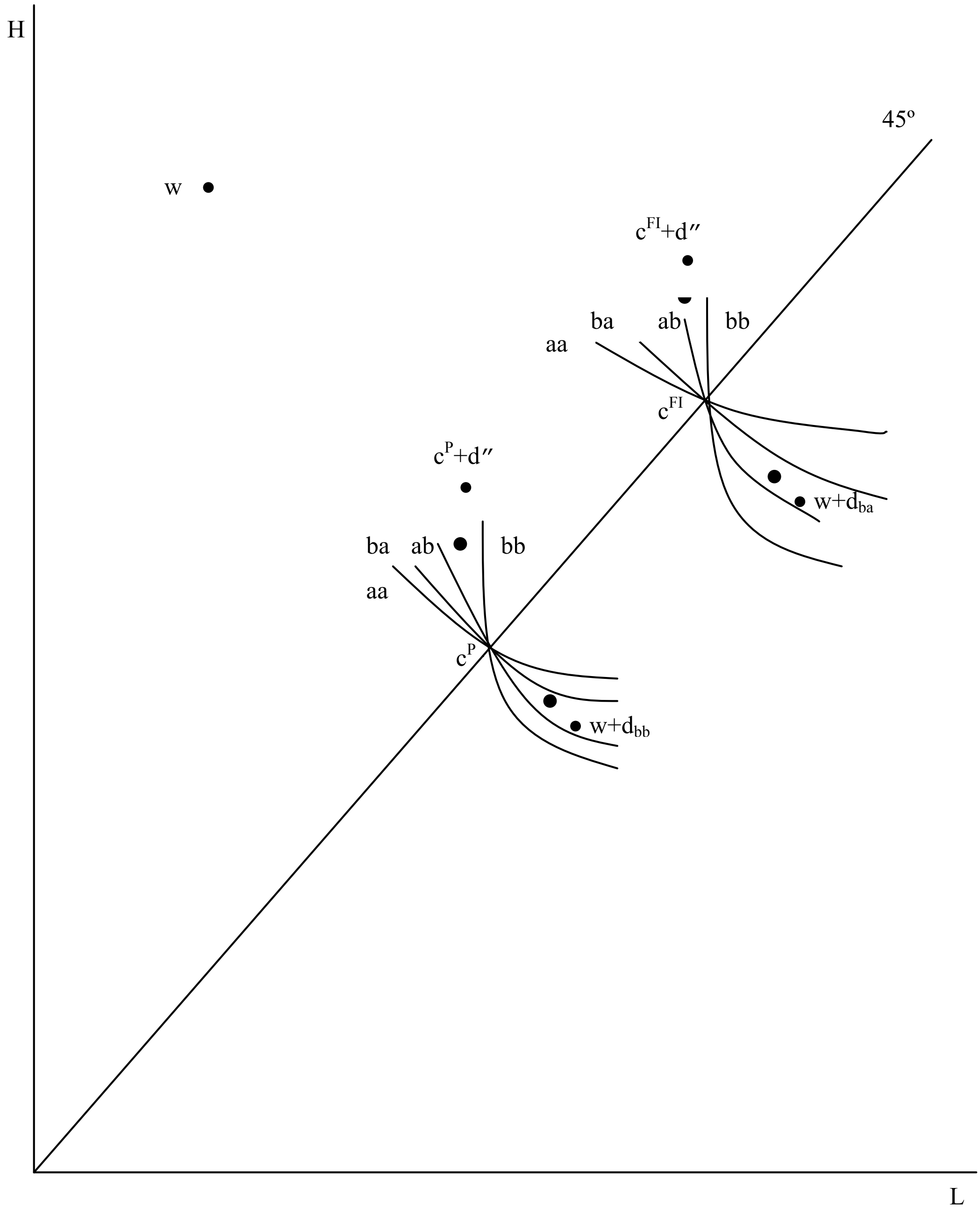
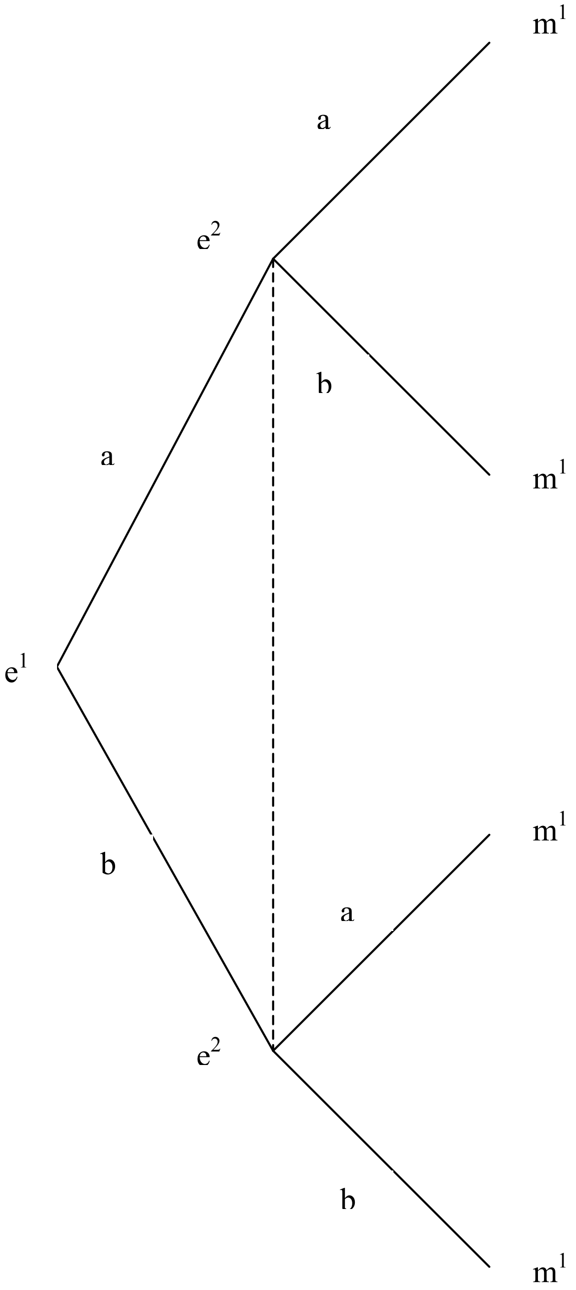


Figure 3



m^1 \ m^2	a	b
a	U^{FI} - a U^{FI} - a	U^{P} - a U^{FI} - a + δ_{aa}
b	U^{FI} - a + δ_{aa} U^{P} - a	U^{P} - a + γ_{aa} U^{P} - a + γ_{aa}

m^1 \ m^2	a	b
a	U^{FI} - a U^{FI} - b	U^{P} - a U^{FI} - b - δ_{ba}
b	U^{FI} - a + δ_{ab} U^{P} - b	U^{P} - a + γ_{ab} U^{P} - b + γ_{ba}

m^1 \ m^2	a	b
a	U^{FI} - b U^{FI} - a	U^{P} - b U^{FI} - a + δ_{ab}
b	U^{FI} - b + δ_{ba} U^{P} - a	U^{P} - b + γ_{ba} U^{P} - a + γ_{ab}

m^1 \ m^2	a	b
a	U^{FI} - b U^{FI} - b	U^{P} - b U^{FI} - b - δ_{bb}
b	U^{FI} - b- δ_{bb} U^{P} - b	U^{P} - b - γ_{bb} U^{P} - b - γ_{bb}

Figure A1

e^1 \ e^2	a	b
a	U^{FI}- a U^{FI}- a	U^{FI}- a + ε_{ab} U^{P}- b
b	U^{P}- b U^{FI}- a + ε_{ab}	U^{P}- b + ω_{bb} U^{P}- b + ω_{bb}

Figure A2

		e^2: a	e^2: b
e^1	a	$U^P - a + \gamma_{aa}$ $U^P - a + \gamma_{aa}$	$U^P - a + \gamma_{ab}$ $U^P - b + \gamma_{ba}$
	b	$U^P - b + \gamma_{ba}$ $U^P - a + \gamma_{ab}$	$U^{FI} - b$ $U^{FI} - b$

Figure A3

	d	d+d′
d	U^{FI}- a U^{FI}- a	U^{P}- b U^{FI}- a + δ_{ab}
d+d′	U^{FI}- a + δ_{ab} U^{P}- b	U^{FI}- b U^{FI}- b