UNSW
THE UNIVERSITY OF NEW SOUTH WALES

Faculty of Business

School of Economics
UNSW, Sydney 2052
Australia

http://www.economics.unsw.edu.au

The Fragmentation of Reputation

Gautam Bose

School of Economics Discussion Paper: 2007/03

ISSN 1323-8949

ISBN 978 0 7334 2433 5

January 3, 2007

The Fragmentation of Reputation

Gautam Bose*

School of Economics
University of New South Wales
Sydney, NSW 2052, Australia

e-mail: g.bose@unsw.edu.au

Abstract

This paper investigates the use of reputation in an economy where principals hire agents for two different kinds of tasks, in which the agents have differing aptitudes. Principal-agent matches are remade every period, but a principal can acquire some information on the past behavior of her current agent. This allows consideration of two different reputation mechanisms—one in which an agent's past record of defections makes no reference to the kind of task, and another in which information about past defections is available separately for each task. The two kinds of reputation can be interpreted as "personal honor" and performance record (e.g. credit history) respectively.

I first characterise the equilibria under the two mechanisms. I then assume that the economy is in equilibrium under one mechanism when the other becomes available. I find that it may be incentive-compatible for individual agents to use the new mechanism, thus dislodging the existing equilibrium, even when the change ultimately turns out to be efficiency-reducing.

JEL Classifications:

---

* I thank the participants at the NEUDC conference 2000, the Econometric Society summer meetings 2001, The Australian Economic Theory Workshop, and seminar participants at the Universities of NSW, Sydney and Melbourne, Hong Kong University, Simon Fraser University, Indian Statistical Institute Delhi, and the Delhi School of Economics. I also thank Catherine de Fontenay and Hodaka Morita for extensive comments and discussion on an earlier draft. The hospitality of the Economic Theory Centre at the University of Melbourne in September 2002 is warmly acknowledged.

1 Introduction

Institutions play a critical role in facilitating economic activity. Institutions provide incentives, aid in enforcing contracts, and generate information in situations where traditional market forces alone would be insufficient for these purposes. New institutions arise and develop in response to market failures, and existing ones are shaped and reshaped by these needs. With the development of appropriate theoretical tools for strategic and informational analysis, it has become possible to treat institutions as endogenous—if somewhat self-willed—actors in the economic realm.

Reputation is one institution that has received significant recent attention in the context of contract enforcement. In societies where agents interact with different partners at different times, reputation enforces cooperative behavior in one-off encounters. Agents cooperate because defection in one encounter will attract adverse responses from future partners. Such a disciplining mechanism requires that a record of an agent's past behavior be available to his current collaborators, and collaborators act according to that record in a socially agreed-upon manner. To constitute an equilibrium, this social agreement must specify behavior which is individually rational for each agent. Such a mechanism is called a *norm equilibrium* (Okuno-Fujiwara and Postlewaite, 1995).

A powerful example of a norm equilibrium is provided by ancient Hindu society which functioned nearly unchanged for thousands of years, and showed itself remarkably resilient in the face of a succession of assaults and incursions by different rulers over the centuries. Conformity in caste-based Hindu society was enforced by social ostracism. Offenders against social norms were punished by imposing on them various levels of excommunication, which consisted of a denial of complementary exchanges of goods and services with other members in the same (usually village-level) society. Since caste rules severely restricted the activities an individual could undertake—and therefore circumscribed the directions in which he or she may have even the most basic functionality—such excommunication constituted substantial punishment (Freitas 2006, Lal 1988).

Punishment for social deviation in caste-based Hindu society was to a great extent context-insensitive—the same set of restrictions on social and economic intercourse being invoked as reprisal for a variety of offences. Offending agents were blacklisted for a period of time, after which their status would be restored, often following an elaborate ceremony. In the mean time, other agents only needed to know that the individual or caste-group is under punishment, and not the details of the offence.

Blacklisting has been a widespread practice in many societies, including the pre-twentieth century West. This is reflected in the concern that prominent citizens displayed for their "good name" and "family honor". When family honor was tarnished, the consequence was a lack of access to complementary exchanges with one's own social class. Once again, the embargo was not limited to transactions in a specific category defined by the initial offence, but was imposed across the board.

With the enlargement of the sphere of anonymous transactions during the last century, sophisticated institutions have emerged which maintain records on individual agents, and make these records available to potential partners looking to transact with those individuals. Credit bureau records, academic transcripts, and employment references are prominent examples. Depending upon the type of transaction, partners may also access more sensitive information such as criminal histories and health records. [1] With increasing anonymity brought about by urbanization and mobility, and with record-keeping facilitated by technology, it is safe to say that contemporary transactors even in Hindu society rely more heavily on the task-specific reputations of their transacting partners—on credit records and employment references, and, to an alarming extent, on academic transcripts.

The technological capacity to store and disseminate specific information on past actions must have been significantly limited in earlier times, and reputation must therefore largely have been of the holistic kind. Decisions on whether to engage in economic transactions with specific partners may well have largely been based on rumors and hearsay. In recent times, there

[1] Access is typically restricted by law to only that information which is relevant to the transaction at hand. For example, a bank can legally access credit and income information for a potential borrower, but not health data or academic transcripts.

has been an immense increase in the capacity to store and transmit task-specific information, to the extent that elaborate privacy laws are needed to restrict individuals from misusing such information about others. Thus there must have been a period when the dominant practice was to use general reputation, while task-specific information was becoming available in parallel.

The natural questions to ask are, under what circumstances would the task-specific information structure be adopted by individuals, and would the result be an increase in efficiency? These are the questions this paper asks.

In this paper I investigate two possible kinds of reputation in an economy where principals hire agents for two different kinds of tasks, in which the agents have differing aptitudes. Principal-agent matches are remade every period, but a principal can acquire some information on the past behavior of her current agent. This allows consideration of two different reputation mechanisms—one in which an agent's past record of defections makes no reference to the kind of task, and another in which information about past defections is available separately for each task.

I first characterize the equilibria under the two mechanisms. The main questions I then ask are: first, suppose that one mechanism was in use, and the other became available—perhaps as a result of technological innovation in record-keeping. Under what conditions can we expect principals in the economy to endogenously switch from using the 'old' mechanism to using the 'new' one? Secondly, is there reason to think that these conditions are ones in which the new mechanism is in some sense more efficient than the old one? We find that it may be incentive-compatible for individual principals to use the new mechanism even when the current norm is to use the old one, thus triggering a switch to an equilibrium using the new mechanism. Further, this may happen even though the resulting equilibrium is less efficient.

The theory of cooperation in infinitely repeated prisoners dilemma games is well established (see Fudenberg and Maskin 1986, Abreu 1988). The model in this paper is closest to Tirole (1996), which examines reputation-based equilibria in a principal-agent model with random matching in every period.

The most prominent difference is that the present model has multiple tasks, which in turn makes possible the consideration of different reputation mechanisms. There are also differences in the way reputation works, but these are of less consequence. Dixit (2001) presents a model of enforcement using intermediaries, where the intermediary is partly a repository of agents' reputations. Ahn and Suominen (2001) study an economy in which reputation is acquired by partners in the form of signals from a sample of agents that had previously encountered the partner.

Norm equilibria that do not utilize reputation have been investigated by Kandori (1992) and Ellison (1994), among others. In their models, agents who are cheated in an interaction deduce that general cooperation will decline in the future, and themselves cease to cooperate with future partners—thus bringing about the general decline in cooperation. The fear that a small number of defections can lead to a general decline provides the incentives for agents to remain honest. Ghosh and Ray (1996) examine cooperation in an economy with no reputation, but where agents have the option of continuing to interact with each other in successive periods, or to end the relationship. In their model with heterogeneous agents, "good" agents that are matched with each other develop cooperative relationships.

This paper is closest in spirit to Greif (1994), in the sense that I attempt a comparison between two different reputation mechanisms, and trace their efficiency consequences. Greif compares "individualist" and "collectivist" societies which coexisted in the same historical period, and characterizes the corresponding equilibria. He then shows that, with changing historical circumstances of trade—specifically the expansion of trading relationships beyond closed communities—one mechanism turned out to be more flexible and adaptive than the other, with the result that the latter declined and ceased to be used.

In this paper I characterize the equilibria under two different mechanisms that broadly succeeded each other in time. I then assume that the economy is in equilibrium using the earlier mechanism, and examine the conditions under which the availability of the later mechanism is sufficient to cause the demise of the earlier. I find that a switch to an equilibrium using the

later mechanism may occur under conditions where it leads to a decrease in efficiency.

The next section sets out the formal model. Characterization of equilibrium under integrated and fragmented reputation occupies sections 3 and 4. Efficiency comparisons are made in section 5 and the possibilities of transition from one system to the other are explored in 6.

2 The model

2.1 Agents and transactions

The economy operates over an infinite succession of discrete periods. There is a large number of agents and at least as many principals.[2] At the beginning of a period, a principal randomly draws one of two tasks, A or B, with equal probability. Each principal is then randomly matched with an agent. The principal then decides whether to hire the agent to perform the task she has drawn. To aid this decision, she may consult a reputation mechanism (described later) to obtain some information about the agent's past behavior.

Players cannot identify each other as ones they have met in the past. This allows us to focus on public reputation as transmitted by the mechanisms under investigation, and abstract from considerations of private reputations.

If the principal decides not to hire the agent, then both obtain a payoff of zero in that period. If she does hire, then she pays the agent a wage of w. The wage is an economy-wide parameter, and exogenous in this model. The agent then decides whether to exert effort in the task. The agent can exert the effort required to make the project a success, which is inflexible and depends on the type of the agent, as described below. Alternatively the agent may exert no effort, which results in failure. Success results in a gross revenue of z, which accrues to the principal, failure results in zero revenue. If the agent exerts the necessary effort, we will say that he "cooperates", if he does not, we say that he "defects" or "cheats".

[2]We need the numbers to be large enough such that in the matching stage, ex ante probabilities are realized as ex post proportions. This can be achieved by formally assuming a continuum of agents.

If the agent defects, the principal can report the agent and the report is transmitted to all active reputation mechanisms (described below). Reporting is costless. Since players are anonymous to each other, a principal has no incentive to report a cooperating agent, or to not report a defector.

Agents are equally divided between two types, A and B. Each agent's type is private information, and remains constant over time.

A type-A agent is relatively more efficient at A-transactions, with cost $a_0 \geq 0$ of cooperating. His cost of cooperating in type-B tasks is $\alpha \in [\alpha_0, \alpha_1]$, where $\alpha_1 > \alpha_0 \geq a_0$. Similarly, for type-$B$ agents the cost of cooperating in type-B tasks is a_0, and the cost of high effort in type-A tasks, α, lies in the interval $[\alpha_0, \alpha_1]$. The cost of exerting low effort (defecting) is 0 for all agents in all transactions.

We normalise the length of the interval $[\alpha_0, \alpha_1]$ to unity, and assume that the effort costs of type-A agents in task B is distributed uniformly over the interval. Similarly, the effort costs of type-B agents in task A is distributed uniformly on $[\alpha_0, \alpha_1]$. Thus each type-A agent has a cooperation cost of a_0 for type A tasks, and a unique cooperation cost α in type-B tasks, where α is drawn randomly from the uniform distribution with support $[\alpha_0, \alpha_1]$ (similarly for type-B).

For an agent of type i, $(i = A, B)$ we will refer to the type i transaction as his "preferred" transaction, and the type not-i transaction as his "dispreferred" transaction.

Within a period, the subgame that occurs after the principal hires the agent is thus a one-sided prisoners dilemma. The agent can cooperate, in which case the payoffs to the principal and agent are $(z - w, w - a)$, with $a = a_0$ or $a = \alpha \in [\alpha_0, \alpha_1]$ depending on the agent's type. Alternatively the agent can defect, resulting in payoffs $(-w, w)$.

Both principals and agents are infinitely lived, discount the future by a factor $\delta \in (0, 1)$, and maximize presented discounted value of future income.

The only incentive to exert high effort comes from the fact that defections are reported to and recorded by the reputation mechanisms. Potential future employers consult the agent's reputation, hence defection may lead to

the absence of a hire in future periods. We consider two alternative reputation mechanisms, "integrated reputation" (IR) and "fragmented reputation" (FR).

2.2 Reputation mechanisms

A reputation mechanism records reports made by principals against agents who defect (put in low effort) in a task. The record consists of a register in which a mark is put against the agent's name when an adverse report is received. The mark remains on the register for a specified number T of periods, and then disappears. We say that an agent is "under punishment" or "marked" if a mark is currently recorded against his name. If a new adverse report is received while the agent is already marked, then the old mark is removed and the new one is entered, which in turn stays on record for another T periods. Thus an agent's reputation is simply an identifier which indicates that he has defected within the last T periods.

Integrated reputation (IR) refers to a mechanism in which there is a single register for the entire economy, and reports of defection are recorded without reference to the type of transaction in which the defection had occurred. Thus when a principal looks at an agent's record, she either knows that the agent has not defected in any transaction within the previous T periods (no mark), or that he has defected in some transaction within that period (there is a mark). If there is a mark, however, she cannot tell which transaction—A or B—the agent has defected in, nor the period in which the defection occurred.[3]

Fragmented reputation (FR) maintains a separate register for each transaction. When a principal reports an agent for defecting in transaction i, a mark is recorded against the agent's name in the register for transaction i. However, no mark is recorded against the agent's name in the register for transaction not-i. The principal is allowed to check the agent's record only for the task she has drawn.[4] Thus she knows whether the agent has defected

[3]The punishment may unravel if the principal knows when the agent had defected, see Bhaskar (1998).

[4]This is in keeping with privacy laws that are in force in most developed countries. A bank, for example, can check a potential borrower's credit record, but cannot ask for

in that particular transaction within the last T periods.[5]

2.3 Strategies and equilibria

For a principal, there is no cost to consulting the reputation of her assigned agent, and to reporting an agent who defects. We therefore assume that each principal does this in every period. Since she is unlikely to be assigned the same agent again, there is also no benefit to misreporting. The principal's strategy then consists of a decision to hire the agent (or not) depending the agent's reputation status. In section 6, the principal will also have to decide whether to consult the IR or FR mechanism to verify her agent's reputation.

For an agent of type i, his strategy in the stage-game in a given period is conditioned by his reputation status in that period. At the beginning of the period, he is either unmarked, or he has a mark which is $T - \tau$ periods old. Let $\tau = 0, 1, ..., T$ represent the number of periods of punishment remaining, where $\tau = 0$ indicates that he is unmarked.

If he is not hired, he has no choices to make. If he is hired for a task (A or B), the agent can either cooperate or defect. If he cooperates, his punishment status will be reduced to $\max(\tau - 1, 0)$. If he defects, his punishment status will be reset to $\tau = T$ in the next period.

The agent can thus adopt one of four strategies in the stage game. When hired, he can (i) cooperate regardless of the task, (ii) cooperate in task A and defect in task B, (iii) cooperate in task B and defect in task A, and (iv) defect regardless of the task. Thus a complete strategy for an agent of type i is a $T + 1$-tuple of pairs $\{s_\tau = (s^i_\tau, s^j_\tau)_{\tau=0,...,T}\}$ where $i, j \in \{A, B\},\ i \neq j$. $(s^i_\tau, s^j_\tau) \in \{c, d\} \times \{c, d\}$ denotes the agent's plan of action when faced with task i or j with τ periods of punishment remaining. For each agent, we will let the first element of s_τ represent his strategy component in his preferred transaction.

We restrict attention to pure, stationary strategies, and look for subgame-perfect, steady-state equilibria. The degenerate outcome, in which no agent

employment references, or use other "unrelated" information.

[5]It will be clear from what follows that the principal does not gain an advantage from consulting registers for both records, and thus would stick to the relevant one if there was an infinitesimally small positive cost for consulting each register. Such a small cost would not compromise any of the results in this paper.

is hired regardless of mark, and all agents defect when hired, is an equilibrium for all configurations of parameters. The next two sections derive the conditions under which a non-degenerate equilibrium exists for each reputation mechanism, and characterize the equilibria.

3 Integrated Reputation

This section shows that a non-degenerate equilibrium exists for the IR economy if and only if the wage is within a certain viable range, and that this equilibrium is unique. Unmarked agents are hired, marked agents are not, and each agent cooperates when assigned to his preferred transaction. An agent who has sufficiently low cost of cooperation in his dispreferred transaction also cooperates when assigned to that transaction.

3.1 Agents' strategies

Assume that each principal hires her assigned agent if and only if that agent is unmarked. The optimality of this strategy for the principals will be established later.

An agent's strategy specifies his response to a task assignment when he is marked and in the τ-th period of his punishment ($\tau = 1, ..., T$), as well as when he is unmarked ($\tau = 0$).[6] An arbitrary agent of type A has four choices of strategy (s_τ^A, s_τ^B) in the stage game, which are (d, d), (c, d), (d, c), and (c, c). Of these, the third—cooperating in the dispreferred task but defecting in the preferred task, can be shown to be always dominated by the second, and will henceforth be ignored.

First consider the agent's choice of actions when he is unmarked. If the agent cooperates when assigned to a task $j = A, B$, then he gets a payoff equal to w less his effort, and moves into the next period unmarked. If he decides to defect, he gets w, and becomes marked for the next T periods. He does not expect to be hired in those T periods, but will be hired again in the $T + 1$-th period when he becomes unmarked. We derive the agent's

[6] Though a marked agent does not receive contracts in equilibrium, the agent's potential response to out-of-equilibrium offers leads to some restrictions on equilibria.

optimal stage-strategy by directly comparing his corresponding expected payoffs under the different feasible strategies.

To ease notation, we first define an often-recurring expression

$$\beta(T) \ := \ \frac{\delta(1-\delta^T)}{[(1-\delta)+\frac{1}{2}\delta(1-\delta^T)]} \tag{1}$$

and a function $a^I(w)$ which is central to the description of cooperative behavior, in the sense that the unmarked agent will cooperate in a task if and only if his effort-cost does not exceed $a^I(w)$.

$$a^I(w) \ := \ \beta(T)(w-\frac{1}{2}a_0). \tag{2}$$

$a^I(w)$ is clearly linear and increasing in w. Define $w^I(.)$ as the inverse of $a^I(w)$, i.e. $w^I(a)$ is the value of w such that $a^I(w) = a$. The values corresponding to effort levels a_0, α_0 and α_1 will be invoked often, so we name them as follows:

$$w^I \ := \ w^I(a_0), \qquad w_0^I := w^I(\alpha_0), \qquad w_1^I := w^I(\alpha_1) \tag{3}$$

When agents are assigned to dispreferred tasks, the restriction of $a^I(w)$ to the interval $[\alpha_0, \alpha_1]$ is relevant, so define:

$$\alpha^I(w) \ = \ \begin{cases} \alpha_0 & \text{if} \quad a^I(w) \leq \alpha_0 \\ a^I(w) & \text{if} \quad a^I(w) \in (\alpha_0, \alpha_1) \\ \alpha_1 & \text{if} \quad a^I(w) \geq \alpha_1 \end{cases} \tag{4}$$

We can now summarize the unmarked agent's strategy as follows:

Proposition 1 *: If principals hire exactly the unmarked agents, then an unmarked agent's optimal strategy when hired and assigned to a task is as follows:*

If the task is his preferred task, cooperate if and only if $w \geq w^I$.

If it is his dispreferred task, cooperate if and only if $w \geq w^I(\alpha) \Leftrightarrow \alpha \leq \alpha^I(w)$.

(All proofs are in the appendix.)

Clearly, no activity will take place in the economy if $w < w^I$. To ensures that all agents cooperate when assigned to their preferred tasks, we will henceforth assume:

Assumption 1 *: $w \geq w^I$.*

Those agents with $\alpha \leq a^I(w)$ also cooperate in their dispreferred task. No unmarked agent will cooperate in dispreferred tasks if $w < w_0^I$, since α_0 is the lowest cost that any agent must expend in such tasks. If the wage is higher than $w^I(\alpha_1)$, however, even the highest-cost agents will cooperate in dispreferred jobs.

Next we turn to a marked agent with τ periods remaining of his punishment. He does not expect to be hired in the next τ periods. However, suppose he is in fact hired. If he cooperates, then he has the remaining τ periods of punishment to go. If he defects, his punishment starts again, i.e. it is lengthened by $T - \tau$ periods. Thus the punishment for the current defection is effectively shorter.

If the agent has been following an optimal strategy, and is marked, then it is because he has $\alpha > a^I(w)$, and had defected in the dispreferred task when unmarked. Since defection now carries a lighter penalty, he will defect if offered a dispreferred contract; he may also defect in a preferred contract if he has too long a period of punishment remaining (i.e. if τ is large). The threshold period is defined as $\tau^*(w)$ such that

$$\tau^*(w) \quad \text{solves} \quad \frac{\delta^{\tau+1}(1-\delta^{T-\tau})}{\delta(1-\delta^T)}\beta(T)(w - \frac{1}{2}a_0) \;=\; a_0 \tag{5}$$

$\tau^*(w)$ is well-defined. When $\tau = T$, the LHS of (5) reduces to 0, which is less than a_0, and when $\tau = 0$, the LHS reduces to $a^I(w) > a_0$. Since the LHS is continuous and decreasing in τ, there is $\tau^* \in (0, T]$ such that (5) holds with equality. The strict inequality holds for $\tau < \tau^*$.

Proposition 2 *Suppose exactly the unmarked agents are hired. The optimal stage-τ strategy for an agent who is marked as a result of a rational past defection is:*

Defect if hired in the dispreferred task.

If hired in the preferred task, cooperate if and only if $\tau \leq \tau^*(w)$.

3.2 Principals' strategies

Next consider the principal's choice of strategy. When assigned an agent, the principal observes whether the agent is marked or unmarked, and has no other information about that specific agent. However, some of the unmarked agents are potential defectors in the task the principal has drawn. The principal's expected payoff from hiring an agent will be non-negative only if the losses from these defectors are balanced by gains made from the cooperating agents. This requires that the wage is not too high.

Let $\theta^I = \alpha_1 - \alpha^I$ be the fraction of potential defectors, and let ϕ^I be the fraction that is marked at any given time. ϕ^I is computed as follows: of the $\theta^I - \phi^I$ potential defectors who are unmarked, half are assigned to their dispreferred tasks in each period, and therefore acquire marks. At any time there are T such marked cohorts, so in a steady-state we have

$$T.\frac{1}{2}.(\theta^I - \phi^I) = \phi^I \qquad \Rightarrow \quad \phi^I = \frac{T}{T+2}\theta^I \tag{6}$$

Specifically, define w_z^I such that

$$w_z^I \quad \text{solves} \quad w = \frac{(T+2) - (T+1)\theta^I(w)}{(T+2) - T\theta^I(w)}\, z. \tag{7}$$

Proposition 3 *Suppose agents follow the strategies described in propositions 1 and 2, and the economy is in steady-state. It is optimal for an individual principal to hire an unmarked agent if and only if* $w \leq w_z^I$.

If w is any greater, no agents will be hired, regardless of reputation status, and the only equilibrium is degenerate. Thus we assume:

Assumption 2 *:* $w \leq w_z^I$.

Since $\theta^I(w)$ is non-increasing in w, w_z^I is increasing in z. If $w \leq w_0^I$, we have $\theta^I(w) = 1$, and the value of z necessary to make the economy viable is $z \geq 2w$. At the other extreme when $w \geq w_1^I$, i.e. no one defects in any task, it is sufficient that $z = w$. Hence $z \geq \max\{2w_0^I, w_1^I\}$ is a sufficient bound to ensure that the economy is viable over the entire range $[w^I, z]$ of wages.

Finally we turn to the principal who is faced with a marked agent. We know that some of these agents would in fact cooperate if hired in the

preferred task (see proposition 2). If w is small relative to z, then the gain on those agents who do cooperate outweighs the loss on the ones that do not, and the principal will be induced to hire marked agents. But then agents have no incentive to cooperate, and the equilibrium breaks down.

Indeed, of the T cohorts of marked agents, those with $\tau \leq \tau^*(w)$ periods of punishment would cooperate in the preferred task. Since half of the agents are assigned to preferred tasks, the principal's expected gain from hiring marked agents is non-positive only if the following condition is satisfied, which we state as an assumption.

Assumption 3 $w \geq \frac{1}{2}\frac{\tau^*(w)}{T}z$.

Proposition 4 *Suppose agents follow the strategy profile described by propositions 1 and 2, and the economy is in steady-state. Then it is optimal for an individual principal to not hire a marked agent if and only if assumption 3 is satisfied.*

Assumption 3 provides another lower bound on the wage, along with w^I in assumption 1. This bound is inconvenient since $\tau^*(w)$ on the RHS of assumption 3 is itself a function of w, and it is not possible to make the relation explicit. However, it turns out that assumption 1 implies assumption 3 if the length of punishment is large enough.

Observation 1 *: Assumption 3 holds if assumption 1 is satisfied and* $T \geq \frac{1}{2}z\frac{1-\frac{1}{2}\delta}{\delta|\ln\delta|}$.

This condition is not necessary for assumption 3 to be satisfied. However, it provides a more tractable bound, and will be used in section 6.

3.3 Equilibrium

Proposition 5 *: There is a unique non-degenerate equilibrium for the IR economy if and only if the wage satisfies assumptions 1, 2 and 3. In this equilibrium principals only hire unmarked agents, and each such agent cooperates unless he has $\alpha > a^I(w)$ and is assigned to his dispreferred task.*

If any one of the three assumptions is not satisfied, then the only equilibrium is the degenerate one in which no agent is hired.

The proof follows readily from propositions 1, 2, 3 and 4, and is omitted.

4 Fragmented Reputation

This section establishes the conditions that ensure the existence of a non-degenerate equilibrium under FR. Individual components of the equilibrium strategy profile are stated as propositions 6 to 8. The existence result is then stated as proposition 9.

Again, we establish that in the only possible non-degenerate equilibria, unmarked agents are hired and marked agents are not. Further, every agent cooperates in his preferred transaction, and agents with low enough costs $\alpha \leq \alpha^F$ also cooperate in their dispreferred transactions.

For a range of wages $[w_0^F, w_z^F]$ and a value of α^F to be determined below, this strategy profile constitutes the unique non-degenerate equilibrium.

Recall that under FR the principal can only consult the agent's record corresponding to the transaction which she has drawn in the current period.

4.1 Agents' strategies

The considerations for the agent's choice of strategy are simpler, since his reputation in a given transaction only affects his future employment possibilities in that transaction.

First define the function that describes the pattern of cooperation:

$$a^F(w) \;:=\; \frac{1}{2}\beta(T)\; w \tag{8}$$

and its inverse

$$w^F(a) \;=\; \frac{2}{\beta(T)}a \tag{9}$$

which leads to the following benchmark values:

$$w^F \;:=\; w^F(a_0), \qquad w_0^F \;:=\; w^F(\alpha_0), \qquad w_1^F \;:=\; w^F(\alpha_1) \tag{10}$$

Note that $a^F(w)$ is increasing in w, and reduces to a_0 if $w = w^F$. Let:

$$\alpha^F(w) \;=\; \begin{cases} \alpha_0 & \text{if} \quad a^F(w) \leq \alpha_0 \\ a^F(w) & \text{if} \quad a^F(w) \in (\alpha_0, \alpha_1) \\ \alpha_1 & \text{if} \quad a^F(w) \geq \alpha_1 \end{cases} \tag{11}$$

As in the IR economy, these define the ranges of wages and effort costs which separate cooperators from defectors. $\alpha^F(w)$ is the highest effort level which is compatible with cooperation in the dispreferred task.

Proposition 6 *: If principals hire exactly the unmarked agents, then an unmarked agent's optimal strategy when hired and assigned to a task is as follows:*
If the task is the preferred task, cooperate if and only if $w \geq w^F$.
If the task is in the dispreferred task, cooperate if and only if $\alpha \leq \alpha^F(w) \Leftrightarrow w \geq w^F(\alpha)$.

In order to ensure a non-degenerate equilibrium, we assume

Assumption 4 *:* $w \geq w^F$.

Next consider a marked agent i in the τ-th period of his punishment in transaction j, and suppose that he is hired (this is off the equilibrium path). If he defects, his punishment is lengthened by τ periods (since it starts all over again). This punishment for a second defection is lighter than the punishment for the original defection. If he cooperates his remaining punishment is unaffected.

If the agent is under punishment because he had earlier rationally defected in this task, it follows that he will do best to defect now. Consideration of an agent who has earlier made a disequilibrium move is not important since it does not affect the principals' expectations.

Proposition 7 *Suppose principals hire exactly the unmarked agents. An agent who is currently marked in a given task because of an earlier rational defection in that task will find it optimal to defect again if hired in that task.*

The proof is straightforward and is omitted.

4.2 Principals' strategies

Of all the agents assigned to a given task in a given period, one-half are in their preferred task—hence they are unmarked and will not defect. The other half are in their dispreferred task. Some of these have high effort cost,

and are potential defectors. Some potential defectors are marked, and some are unmarked.

If a principal hires to an unmarked agent, the agent will defect with a probability equal to the fraction of unmarked agents who are potential defectors. Let $\theta^F = \alpha_1 - \alpha^F$ represent the fraction of agents that are potential defectors. In order to make it profitable to hire, the principal's gain on the fraction of agents who cooperate must outweigh the loss on those that defect. This requires $w \leq w_z^F$, which is defined by

$$w_z^F \quad \text{solves} \quad w = \frac{2(T+2) - (T+2)\theta^F(w)}{2(T+2) - T\theta^F(w)} z. \tag{12}$$

When a principal is faced with a marked agent, he expects the agent to defect with probability one, and will therefore not offer a contract. The principal's optimal strategy is therefore summarized as follows.

Proposition 8 *: Suppose agents follow the strategy profile described by propositions 6 and 7, and the economy is in steady-state. An individual principal will hire an unmarked agent if and only if $w \in [w^F, w_z^F]$. A principal will never hire a marked agent.*

Assumption 5 $w \leq w_z^F$.

We can verify that, when $w = w^F$, $\theta^F(w) = 1$ and the assumption is satisfied if $z \geq \frac{T+4}{T+2} w$. When $w \geq w_1^F$, $\theta^F(w) = 0$ and $z \geq w$ is sufficient. Note, incidentally, that the highest wage w_z^F which is compatible with activity in the economy under FR is larger than w_z^I, the highest viable wage under IR.

4.3 Equilibrium

Combining propositions 6, 7, and 8, we can state the following result. The proof is straightforward, and is omitted.

Proposition 9 *: There is a unique non-degenerate equilibrium for the FR economy if and only if the wage satisfies assumptions 4 and 5. In this equilibrium principals hire only the agents who are unmarked in the register*

for the assigned task. An agent defects if and only if he has $\alpha > a^F(w)$ and is assigned to his dispreferred task. Otherwise he cooperates.

If either of the two assumptions is not satisfied, then the only equilibrium is the degenerate one in which no agent is hired.

5 Efficiency comparisons

In this section we address the question of comparative efficiency of the two reputation mechanisms, taking as yardstick the surplus net of effort-cost that is generated each period. All calculations are made using the assumption that the economy is in steady-state under the respective reputation mechanism.

The last two sections showed that the degree of cooperation under either mechanism varies with the wage. Hence the amount of surplus also varies accordingly. In comparing the two mechanisms, therefore, we hold the wage constant, and ask whether one or the other generates a greater surplus at a given wage. The answer depends on the level at which the wage is fixed.

To ensure that there is an equilibrium at each relevant wage, assume z is sufficiently large such that assumptions 2 and 5 are satisfied for $w \leq w_1^F$. Also let T be large enough for the condition in observation 1 be satisfied.

The maximum potential surplus S^* is generated when all agents cooperate in either task. Half the agents are assigned to their preferred tasks and generate a net surplus of $(z - a_0)$ each,while the other half are assigned to dispreferred tasks and generate $(z - \alpha)$. Noting that α is uniformly distributed on the interval $[\alpha_0, \alpha_1]$, the length of which has been normalised to unity, we obtain:

$$S^* \;=\; \frac{1}{2}(z - a_0) \;+\; \frac{1}{2}z \;-\; \frac{1}{4}(\alpha_1^2 - \alpha_0^2) \tag{13}$$

To facilitate comparison in the range $[w^F, w_1^I)$, we compute the surplus under the two mechanisms explicitly.

Integrated Reputation: Given a wage w, a fraction $\theta^I(w) = \alpha_1 - \alpha^I(w)$ of agents defect in dispreferred tasks. In steady-state, $\dfrac{T}{T+2}$ of these are marked (see equation(6)), and are therefore not hired. Of the remainder,

half are assigned to preferred tasks and cooperate, generating a surplus of $z-a_0$ each, while the other half are assigned to dispreferred tasks and defect.

The fraction $1-\theta^I(w) = \alpha^I(w) - \alpha_0$ are unmarked and cooperate regardless of task. Half are assigned to preferred tasks and generate $z - a_0$ each, while the other half are assigned to dispreferred tasks and generate $z - \alpha$ each. Combining we obtain the net surplus:

$$\begin{aligned} S^I(w) \quad &= \quad \frac{1}{2}(z-a_0)(\alpha^I(w)-\alpha_0) \;+\; \frac{1}{2}z(\alpha^I(w)-\alpha_0) \\ &\quad + \frac{1}{T+2}(z-a_0)(\alpha_1-\alpha^I(w)) \;-\; \frac{1}{4}[(\alpha^I(w))^2-(\alpha_0)^2] \end{aligned} \tag{14}$$

The net surplus is zero for wages below w^I, where it jumps to $\frac{1}{T+2}(z-a_0)$ and remains constant until the wage reaches w_0^I. Thereafter it increases monotonically to reach S^* at w_1^I, where it becomes constant again.

Fragmented Reputation: Under FR, all agents are unmarked when assigned to their preferred tasks, and hence such agents are always hired. Of the half that are assigned to their dispreferred tasks, a fraction $\theta^F = \alpha_1 - \alpha^F(w)$ are potential defectors. These are either marked and not hired, or they are unmarked, hired, and defect. In either case this fraction generates zero surplus. The remaining $\alpha^F(w) - \alpha_0$ cooperate in their dispreferred tasks, therefore they are unmarked and are hired. The total surplus generated therefore turns out to be:

$$S^F(w) \;=\; \frac{1}{2}(z-a_0) + \frac{1}{2}z(\alpha^F(w)-\alpha_0) - \frac{1}{4}[(\alpha^F(w))^2-(\alpha_0)^2]$$

$S^F(w)$ is zero for $w < w^F$, at which point it jumps up to $\frac{1}{2}(z-a_0)$, where it remains constant until w_0^F. Thereafter it rises monotonically until it reaches S^* at w_1^F, and becomes constant.

Graphs of S^F and S^I are drawn in figure 1 for two sets of parameters. Analytically, the comparison between the two mechanisms is straightforward for $w < w^F$ and $w \geq w_1^I$, and we record these first. From equations (1), (3) and (10) it can readily be ascertained that $w^I < w^F$ and $w_1^I < w_1^F$.

Lemma 1 *The following efficiency comparisons holds for wages outside the interval* $[w^F, w_1^I)$.

For $w < w^I$	$S^I(w) = S^F(w) = 0.$
For $w \in [w^I, w^F)$	$S^I(w) > S^F(w) = 0.$
For $w \in [w_1^I, w_1^F)$	$S^I(w) > S^F(w).$
For $w \geq w_1^F$	$S^I(w) = S^F(w) = S^*.$

Though in the intervals considered above IR dominates FR in terms of surplus generated, this may be reversed in part of the interval $[w^F, w_1^I)$. Note that there is a discontinuity in S^F at w^F, which is the lowest wage for which the FR economy has an active equilibrium. This equilibrium generates a surplus which is larger than that generated by IR at its lowest viable wage w^I.

Thus if the degree of cooperation in the IR economy does not rise to too high a level at $w = w^F$, then S^F will jump up above S^I at this point. In particular this will hold if α_0 is greater than the value attained by $a^I(w)$ at w^F. A less restrictive sufficient condition, which subsumes the above condition, is derived below.

Lemma 2 *The surplus generated by FR at* $w = w^F$ *is greater than that generated by IR if* $\alpha^I(w^F) < \frac{T}{2T+2}\alpha_1 + \frac{T+2}{2T+2}\alpha_0$.

This is a sufficient condition, and is stronger than necessary for the antecedent to hold. The condition in the proposition essentially requires that the interval $[\alpha_0, \alpha_1]$ is located sufficiently high compared to a_0, though even $\alpha_0 = a_0$ may be adequate for the condition to hold.

Even if S^F exceeds S^I at w^F, the latter must overtake the former before w_1^I, since we know that $S^I(w_1^I)$ is strictly greater than $S^F(w_1^I)$. Both functions are continuous for $w > w^F$, and it is straightforward to show that S^I has a steeper positive slope in the interval (w_0^F, w_1^I). These observations, together with the previous propositions, lead to the following proposition. The proof is omitted.

Proposition 10 *IR generates strictly greater surplus than FR for all wages* $w^I \leq w \leq w_1^I$, *except possibly in an interval* $w^F \leq w \leq w^*$, *where* $w^* < w_1^I$.

Efficiency comparisons between the two mechanisms for two sets of parameter values are shown in figures 1a and 1b. The following consequence of proposition 10 will be used in the next section, and is stated here for convenience.

Corollary 1 *There exists a non-empty interval which includes* w_1^I *in its interior on which IR is strictly more efficient than FR.*

6 Transition between regimes

Suppose that the economy operates under an IR mechanism, and is in a steady-state. Now imagine that, exogenously, a parallel FR mechanism becomes available. This may occur as a result of technological change, which allows more detailed information to be acquired and stored. Or it may reflect private initiative in response to a perceived profitable opportunity. However, given the existing equilibrium, agents expect their future principals to consult the IR records and not the FR ones, and incentives are accordingly defined.

In this section I identify conditions under which individual principals may nevertheless prefer to consult the FR records rather than the IR records. It follows that under those conditions, exclusive use of the IR mechanism alone cannot continue to be an equilibrium. The economy must move to a new state in which the FR mechanism is used at least by some principals.

As argued in the introductory section, such a change have indeed occurred in the past century. However, a switch turns out to be incentive-compatible only under specific conditions. What is more interesting, some of these conditions overlap with those under which, in the previous section, we found FR to be less efficient than IR.

Let the economy be in a steady-state where all principals consult the IR mechanism. Suppose the FR mechanism becomes available, so that an individual principal has the option of consulting the FR register instead.

I assume that a principal is only permitted to consult one mechanism in a given period. This assumption would be validated if there is some small cost of acquiring the information, or if she has to 'buy into' one or the other

mechanism. It will be clear from what follows that it is never useful for a principal to use more than one reputation mechanism. To avoid clutter, I have assumed this to be a rule rather than prove it as a proposition.

Let $w < w_1^I$, so that some agents cheat in their dispreferred tasks, and consequently some agents are marked. Consider a principal who has drawn an A-task, and has been assigned an agent. This principal has the option of consulting the IR mechanism, or the A-register under the FR mechanism. Note that her information structure under the second option is not a refinement of the first. The two options will yield different signals under some circumstances. Under other circumstances the signals will be indistinguishable.

The agent may be in one of the following three states. He may be currently unmarked—i.e., he has not cheated in either task in the previous T periods. Both the IR register and the FR register will then show him as unmarked. Alternatively, he may have a mark because he cheated in an A-task within the previous T periods. In this case, both the IR register as well as the FR register for A-tasks will show him as marked. In either of these cases, both registers yield the same signal to the principal.

The third possibility is that the agent may be currently marked because he cheated in a B-task within the past T periods. The IR register will then show him as marked, whereas the FR register for A-tasks will show him as unmarked. Thus the principal would hire this agent if she consulted the FR mechanism, but not the IR mechanism.

This is the only event in which the principal receives different signals from the two mechanisms. Thus she is better off consulting FR if, conditional on this event, her expected payoff from hiring is greater than her expected payoff from not hiring.

The latter payoff is zero. We therefore need to evaluate the principal's payoff from offering a A-task to an agent who is marked on account of having cheated in a B-task, and to find conditions under which this payoff is positive. These then will precisely be the conditions under which the principal will prefer to consult the FR register.

Since the economy is in steady-state and agents are following their op-

timal strategies, the agent who is marked as a result of a B infraction must be a A-type agent. In other words, the task A which has been drawn is the agent's preferred task. Recall from proposition 2 that such an agent will cooperate when hired if he has $\tau \leq \tau^*$ periods of punishment remaining, where τ^* is defined in equation (5).

At any date, there are T equal cohorts of agents of each type that are in punishment, with one such cohort having entered punishment in each of the previous T periods. τ^* of these cohorts have τ^* or less periods remaining, thus the fraction of such agents who will cooperate when hired is $\frac{\tau^*}{T}$. Thus for an arbitrary agent of this type, this is the probability that the agent will cooperate, while with the remaining probability $1 - \frac{\tau^*}{T}$ he will defect.

If the agent cooperates the principal gets $(z - w)$, if he defects she gets $-w$. Thus the principal's expected payoff is $z\frac{\tau^*}{T} - w$, which is positive if

$$\frac{\tau^*(w)}{T} \geq \frac{w}{z} \tag{15}$$

Note that the LHS of equation 15 is independent of z, whereas the RHS decreases with z. Thus for any w such that $\tau^*(w)$ is positive, there is z large enough that 15 is satisfied. For $w < w^I$ the equilibrium in the IR economy is degenerate. For $w \geq w_1^I$ there is full cooperation in all tasks so there are no marked agents. Thus we need to focus attention on the interval (w^I, w_1^I).

$\frac{w}{z}$ is linear in w with a slope $\frac{1}{z}$ which diminishes as z increases. The shape of $\frac{\tau^*(w)}{T}$ is established in the following lemma. The two curves are graphed in figure 2 for two sets of parameter values.

Lemma 3 *$\frac{\tau^*(w)}{T}$ attains a value of zero at $w = w^I$, is increasing and concave in w, and asymptotically approaches unity as $w \longrightarrow \infty$.*

Clearly the parameters may be such that condition (15) is never satisfied, and an individual principal never has the incentive to consult the FR register. However, it is interesting to note that, if the economy is sufficiently productive (i.e. z is large enough), then for *any* $w \in]w^I, w_1^I[$ the individual principal will prefer to consult the FR register, even though the norm in the economy is to use IR. Further, this is in spite of the fact that, by corollary 1 if w is sufficiently close to w_1^I then the equilibrium using the FR register is less efficient than that using the IR register.

Proposition 11 *: If z is sufficiently large, then there is $w^* \in (w^F, w_1^I)$ such that, for $w \in (w^*, w_1^I)$, principals prefer to consult the FR mechanism, but the FR mechanism is less efficient than the IR.*

In other words, when the economy is highly productive and workers are relatively prosperous, individual incentives may lead to the socially inferior institutional choice being endogenously made.

Figures 1a and 2a, drawn for the same parameter values, illustrate a case where IR dominates FR everywhere, and the FR mechanism will not be used at any wage. In contrast, the economy in figures 1b and 2b is one in which FR dominates in a small interval (about 0.7 to 1.1), but transition will take place at a large range of wages (all wages between about 0.8 and 2).

In figure 3, transition possibilities from IR to FR are offset against the efficiency comparison in $T - w$ space. For given values of the other parameters, figure 3a shows the combinations of T and w for which condition (15) is satisfied—i.e., conditions under which a transition from IR to FR may be expected. Figure 3b shows combinations for which FR generates (weakly) more surplus than does IR.[7] It is clear that transition from IR to FR may happen in a large area of the parameter space where IR is in fact the more efficient mechanism. For completeness, we note that transition in the reverse direction will not occur.

Proposition 12 *If the economy is in equilibrium under FR, and an IR register becomes available, no principal will individually have an incentive to consult the IR register instead of the FR register.*

We do not provide a formal proof because the argument is straightforward. The only event in which the IR and FR registers provide different signals is when the agent has recently defected in the task other than the one the principal has currently drawn. The FR register shows the agent as unmarked in the current task, prompting the principal to offer a contract, whereas the IR register will show him as marked, leading to the denial of a contract.

[7] The hatched area in figure 3b corresponds to values where both mechanisms generate zero or S^*.

However, this is exactly the case where the currently drawn task is the agent's preferred task, and we know that under FR the agent will not defect in this task (proposition 6). Hence the IR register provides an inferior signal and will not be used.

7 Conclusion

This paper had two complementary objectives. One was to produce a model of reputation mechanisms which structure incentives for cooperation, and compare two different such mechanisms which have in fact been observed to operate at different times and places. The particular question of interest which I have attempted to address is whether the availability of one mechanism may disrupt equilibrium under the other, and generate a movement away from the incumbent to the competing mechanism.

The other aim—within the context of that analysis—was to demonstrate that such a movement may in fact occur because it is profitable for the individual; however, the equilibrium that would result if all individuals adopted the change may nevertheless be inferior.

The analysis here attains these goals, if at all, only on a very modest scale. The model, in particular the punishment rule is extremely simple, as is the assumed distribution of abilities and efforts in different tasks. Nor does it incorporate a true dynamic analysis of the relevant institutional change. The attempt would have been much more attractive if the length of punishment were endogenous, but I have not found a tractable way to overcome that shortcoming.

However, the analysis may have some value since theoretical models of institutional change are still relatively scarce, even though the subject itself is attracting increasing interest.

Of greater enduring interest to economists is the question of whether economic processes evolve towards greater efficiency when left to their own devices. This example suggests that this may not always be the case.

References

Abreu, D. (1988): On the theory of infinitely repeated games with discounting. *Econometrica* **56**, 383-396.

Ahn, I. and M. Suominen (2001): Word-of-mouth communication and community enforcement. *International Economic Review* **42**, 399-415.

Bhaskar, V. (1998):Informational constraints and the overlapping generations model: folk and anti-folk theorems. *Review of Economic Studies*, vol. 65, no. 1, pp. 135-49.

Dixit, A. K.(2003):On modes of economic governance. *Econometrica*, **71**, 449-481.

Ellison, G. (1994): Cooperation in the prisoners' dilemma with anonymous random matching. *Review of Economic Studies* **61**, 567-88.

Freitas, K. (2006): The Indian caste system as a means of contract enforcement. *mimeo* Northwestern University. Paper presented at the NEUDC conference 2006, Cornell University.

Fudenberg, D. and E. Maskin (1986): The folk theorem in repeated games with discounting or with imperfect information. *Econometrica* **54**, 533-554.

Ghosh, P. and D. Ray (1996): Cooperation in community interaction without information flows. *Review of Economic Studies*, 491-519.

Greif, A. (1994): Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies. *Journal of Political Economy*, **102** (5), pp. 912-950.

Kandori, M. (1992): Social norms and community enforcement. *Review of Economic Studies* **59**, 63-80.

Lal, Deepak (1988): *The Hindu Equilibrium.* Oxford University Press.

Okuno-Fujiwara, M. and A. Postlewaite (1995): Social norms and random matching games. *Games and Economic Behavior* **9**, 79-109.

Tirole, Jean (1996): A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *Review of Economic Studies*, vol. 63, no. 1, pp. 1-22.

A Appendix

Proof of proposition 1: Let the agent's expected discounted sum of all future payoffs under (s_0^A, s_0^B) be denoted $v_0(s_0^A, s_0^B)$. If the agent chooses to defect at both tasks, he will get his wage, not spend effort, and go under punishment, returning to the unmarked state to repeat his stage-strategy T periods later. This gives

$$v_0(d,d) \;=\; w \;+\; \delta^{T+1} v_0(d,d) \;\Longrightarrow$$

$$v_0(d,d) \;=\; \frac{1}{1-\delta^{T+1}} w. \tag{16}$$

On the other hand, suppose that he decides to cooperate in his preferred task and defect in the dispreferred task. He is assigned to the former with probability half, gets $w - a_0$, and returns unmarked next period. With probability the other half, he is assigned to the dispreferred task, defects and get w, but becomes marked for the next T periods. Thus his expected payoff is:

$$v_0(c,d) \;=\; \frac{1}{2}[(w-a_0) + \delta v_0(c,d)] \;+\; \frac{1}{2}[w \;+\; \delta^{T+1} v_0(c,d)] \;\Longrightarrow$$

$$v_0(c,d) \;=\; \frac{1}{(1-\delta) + \frac{1}{2}(\delta - \delta^{T+1})}[w - \frac{1}{2}a_0]. \tag{17}$$

Finally, if his strategy is to cooperate in both tasks, he gets $w - a_0$ or $w - \alpha$ with equal probability depending on task, and never enters punishment. His payoff is:

$$v_0(c,c) \;=\; \frac{1}{1-\delta}[w - \frac{1}{2}a_0 - \frac{1}{2}\alpha]. \tag{18}$$

Using (16) and (17), and recalling that $a_0 \leq \alpha_0$ we get:

$$v_0(c,d) \;\geq\; v_0(d,d) \;\Leftrightarrow\; w \;\geq\; w^I$$

which establishes the first item in the proposition. Using (17) and (18) we get:

$$v_0(c,c) \;\geq\; v_0(c,d) \;\Leftrightarrow \alpha \;\geq\; a^I(w)$$

which establishes the second and third items in the proposition. ■

Proof of proposition 2: Since the agent is marked following a rational defection, he has $\alpha > \alpha^I$, and an optimal stage-0 strategy $s_0 = (c, d)$. His payoff to becoming unmarked is therefore $v_0(c, d)$, and he expects zero payoff in each period that he is marked.

Given a contract, his discounted expected payoff is:

$$(w \ - \ a) \ + \ \delta^{\tau+1} v_0(c, d)$$

if he cooperates, and

$$w \ + \ \delta^T v_0(c, d)$$

if he defects, where $a = a_0, \alpha$ depending on the task. Substituting from (17) and simplifying, we find that he will cooperate iff

$$\frac{\delta^{\tau+1}(1 - \delta^{T-\tau}}{(1-\delta) + \frac{1}{2}(\delta - \delta^{T+1})}[w - \frac{1}{2}a_0] \ \geq \ a. \tag{19}$$

A comparison with (2) in the text shows immediately that the left-hand side of (19 is smaller than $a^I(w) \leq \alpha$, thus the agent will defect in the dispreferred task. However, the LHS is continuous in τ and equals $a^I(w) > a_0$ when $\tau = 0$. Thus for $a = a_0$ and τ small enough, the inequality (19) is satisfied, and the agent will cooperate in the preferred task. The threshold level τ^* is the value for which (19) holds with equality for $a = a_0$. Thus his optimal strategy, which is paraphrased in the proposition, is

$$s_\tau \ = \ \begin{cases} (d, d) & \text{if} \quad \tau < \tau^* \\ (c, d) & \text{if} \quad \tau \geq \tau^* \end{cases}$$

■

Proof of proposition 3: From proposition (1) we know that the only agents who will defect are those with $\alpha > \alpha^I$ and are assigned to their dispreferred tasks. So the proportion of potential defectors across the two tasks is $\theta^I(w) = \dfrac{\alpha_1 - \alpha^I}{\alpha_1 - \alpha_0}$. By (6), $\dfrac{T}{T+2}\theta^I(w)$ of these are marked at any given time, and $\dfrac{2}{T+2}\theta^I(w)$ are unmarked.

Hence the total number of unmarked agents is $1 - \theta^I(w) + \dfrac{2}{T+2}\theta^I(w)$. Of these, half of the unmarked defectors, or $\dfrac{1}{T+2}\theta^I(w)$ will be assigned

to their dispreferred tasks and defect when offered a contract. Hence, $1 - \theta^I(w) + \frac{1}{T+2}\theta^I(w)$ will cooperate.

The principal pays w if she offers a contract, and receives z if the agent cooperates. Using these to calculate expected payoff, and simplifying we find that the principal's expected payoff if she offers a contract is non-negative if $w \leq w_z^I$. ■

Proof of proposition 4: A marked agent will cooperate when offered a contract if and only if it is in his preferred transaction, and he is in a period $\tau \leq \tau^*$ of his punishment. Since the economy is in steady state, agents enter punishment at a constant rate in each period, and the number of agents in each stage $\tau = 1, ..., T$ of punishment is $1/T$. Thus the probability that the agent will cooperate is $\frac{1}{2}\frac{\tau^*}{T}$, and the principal's expected net gain from offering a contract is $\frac{1}{2}\frac{\tau^*}{T}z - w$. For the principal to deny contracts to marked agents, this net gain must be non-positive, which yields the condition in assumption (3). ■

Proof of proposition 6: Consider an agent i who has been contracted to perform transaction j, and does not have a mark on his j-record. In any future period, he will be assigned to a principal who has drawn task j with probability $\frac{1}{2}$. If his strategy is to cooperate in transaction j, then his expected discounted payoff from j-assignments is

$$v^j(c) \;=\; [w - \alpha^{ij}] \;+\; \frac{1}{2}\frac{\delta}{1-\delta}[w - \alpha^{ij}] \;=\; \frac{1 - \frac{1}{2}\delta}{1-\delta}[w - \alpha^{ij}] \qquad (20)$$

where the first term between the equality signs is the current period payoff, and the second term is the expected discounted payoff in future periods.

If instead he pursues a strategy of defecting when assigned to transaction j, his payoff is

$$v^j(d) \;=\; w \;+\frac{1}{2}\delta^{T+1}v^j(d) \;+\; (\frac{1}{2})^2\delta^{T+2}v^j(d) \;+\; (\frac{1}{2})^3\delta^{T+3}v^j(d) \;+\; ...$$

$$\Longrightarrow v^j(d) \;=\; \frac{1 - \frac{1}{2}\delta}{(1-\delta) + \frac{1}{2}\delta(1-\delta^T)}\; w \qquad (21)$$

It follows that the agent will cooperate if $v^j(c) \geq v^j(d)$, which reduces to $a^{ij} \leq a^F(w)$. Substituting a_0 for a^{ij} and rearranging yields $w \geq w^F$ which is the condition for cooperation in the preferred task. Similarly substituting α for a^{ij}, and using (11) to restrict its range to $[\alpha_0, \alpha_1]$ yields $\alpha \leq \alpha^F(w)$ as the condition for cooperation in the dispreferred task. ■

Proof of proposition 8: Of the agents assigned to principals who have drawn the agent's dispreferred transaction, a fraction $\theta^F(w) = \alpha_1 - \alpha^F(w)$ have effort costs too high, and are potential defectors. $\frac{T}{T+2}\theta^F(w)$ of these agents are marked, and the remaining $\frac{2}{T+2}\theta^F(w)$ are unmarked. All agents assigned to principals who have drawn the agent's preferred transaction are unmarked, since $w \geq w^F$.

Thus the total number of agents who are unmarked in a given transaction is $[\frac{1}{2}+\frac{1}{2}(1-\theta^F(w))+\frac{1}{2}(\frac{2}{T+2})\theta^F(w)]$. Of these, $\frac{1}{2}(\frac{2}{T+2})\theta^F(w)$ will defect if assigned to that transaction, and $[\frac{1}{2}+\frac{1}{2}(1-\theta^F(w))]$ will cooperate. Since the principal will gain $(z-w)$ if the agent cooperates, and lose w if he defects, the condition for her expected payoff to be positive reduces to $w \leq w_z^F$.

If $w < w^F$, then all agents defect in all transactions, hence the principal will not offer a contract. Finally, if an agent is marked in the given transaction, then this is his dispreferred transaction and he has $\alpha > \alpha^F(w)$, so he will defect. Hence the principal will not offer him a contract. ■

Proof of lemma 1: For $w < w^I$ there is no activity under either IR or FR, hence both mechanisms generate zero surplus.

For $w \in [w^I, w^F)$ there is some activity and cooperation under IR, but no activity under FR, hence the former generates greater surplus.

For $w \in [w_1^I, w_1^F)$, all agents cooperate and none are marked under the IR mechanism, hence the full surplus S^* is realised. However, there is some defection under the FR mechanism, and some agents are marked. Hence IR generates greater surplus.

For $w \geq w_1^F$ both mechanisms generate full cooperation and the full surplus S^*.■

Proof of lemma 2: Use equations (14) and (15) to obtain the following

condition for $S^F > S^I$:

$$\frac{T}{T+2}(z-a_0)(\alpha_1-\alpha^I)+\frac{1}{2}[(\alpha^I)^2-\alpha_0^2] > z[\alpha^I-\alpha_0] \tag{22}$$

where α^I represents $\alpha^I(w^F)$. This condition is clearly satisfied if $\alpha^I(w^F) = \alpha_0$, so let $\alpha^I(w^F) > \alpha_0$. Divide by $\alpha^I - \alpha_0$ and rearrange to obtain:

$$\frac{T}{T+2}\frac{z-a_0}{z}\frac{\alpha_1-\alpha^I}{\alpha^I-\alpha_0} \; + \; \frac{\alpha^I+\alpha_0}{2z} \; > \; 1$$

Note that $\alpha^I > \alpha_0 \geq a_0$, so the numerator of the second term on the LHS can be substituted with $2a_0$ (this is where the condition becomes stronger than necessary). Further rearrangement yields

$$T\alpha_1-(2T+2)\alpha^I+(T+2)\alpha_0 > \frac{a_0}{z}[T\alpha_1-(2T+2)\alpha^I+(T+2)\alpha_0]$$

Since $\frac{a_0}{z} < 1$, the condition holds if and only if the LHS is positive (note that the LHS is identical to the term in square brackets on the RHS). Rearranged, this yields the condition in the proposition. ■

Proof of lemma 3: Rewrite (5) as

$$\delta^{\tau^*} \; = \; \delta^T \; + \; (1-\delta^T)\frac{a_0}{\beta(w-\frac{1}{2}a_0)} \tag{23}$$

Note that the denominator of the second term on the RHS is in fact $a^I(w)$ which reduces to a_0 when $w = w^I$, implying that $\tau^* = 0$. Also note that δ^{τ^*} is greater than δ^T for all finite w, implying that $\tau^* < T$. As w increases, δ^{τ^*} approaches δ^T from above, implying that τ^* approaches T from below since $\delta \in]0,1[$. The concavity of $\tau^*(w)$ can be demonstrated by differentiating twice and establishing that the second derivative is negative.■

Proof of proposition 11: Choose z sufficiently large such that $w_1^I/z < \tau^*(w_1^I)/T$. This is feasible since the function $\tau^*(w)$ is independent of z. Then the line w/z lies below the curve τ^*/T in the neighbourhood of w_1^I, hence there is an interval I_1 of wages with w_1^I as its right-hand limit such that principals will choose to use FR if the wage is in this interval.

We know from proposition 10 that there is an interval I_2 around w_1^I in which IR generates greater surplus than does FR. The intersection between I_1 and I_2 is then a set of wages in which the proposition holds. ■

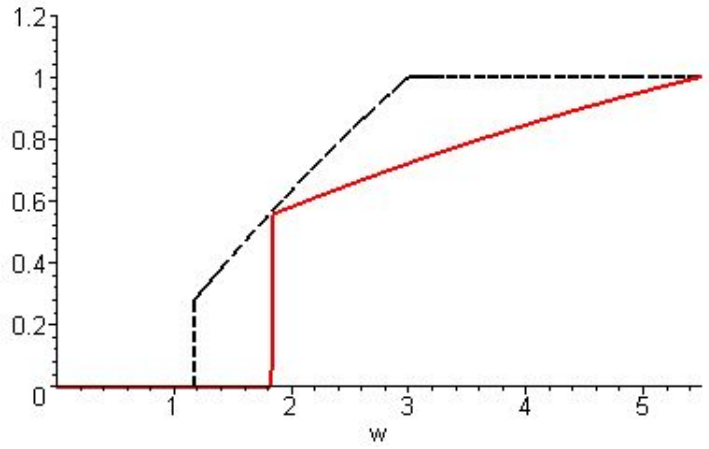

Figure 1.a

Graph of S^I (dashed) and S^F (solid) normalized by S^*
for parameter values
$z = 3,\ T = 2,\ a_0 = 0.5,\ \alpha_0 = 0.5,\ \alpha_1 = 1.5,\ \delta = 0.5.$

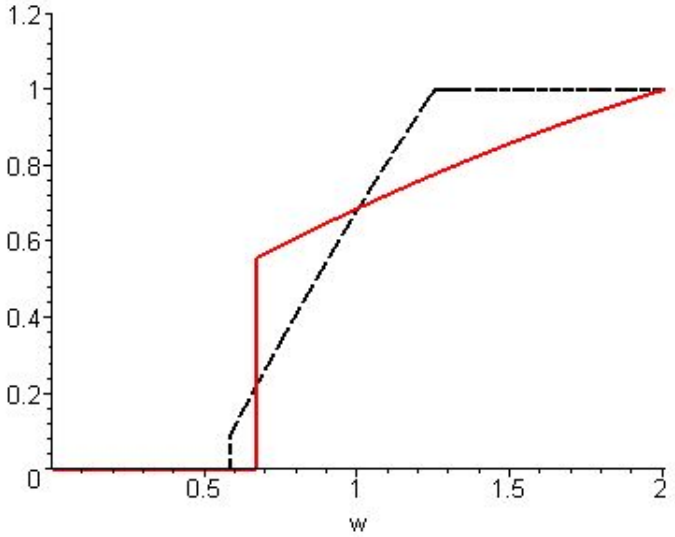

Figure 1.b

Graph of S^I (dashed) and S^F (solid) normalized by S^*
for parameter values
$z = 3,\ T = 10,\ a_0 = 0.5,\ \alpha_0 = 0.5,\ \alpha_1 = 1.5,\ \delta = 0.9.$

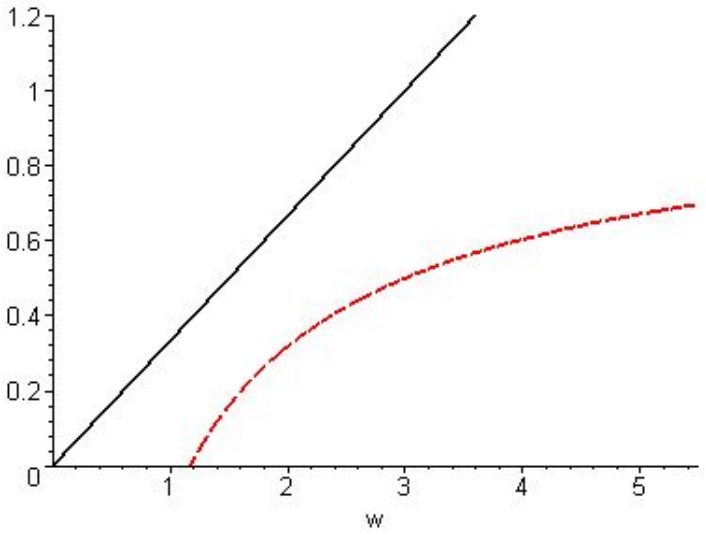

Figure 2.a

Graph of $\frac{\tau^*}{T}$ and $\frac{w}{z}$ for parameter values

$z = 3,\ T = 2,\ a_0 = 0.5,\ \alpha_0 = 0.5,\ \alpha_1 = 1.5,\ \delta = 0.5.$

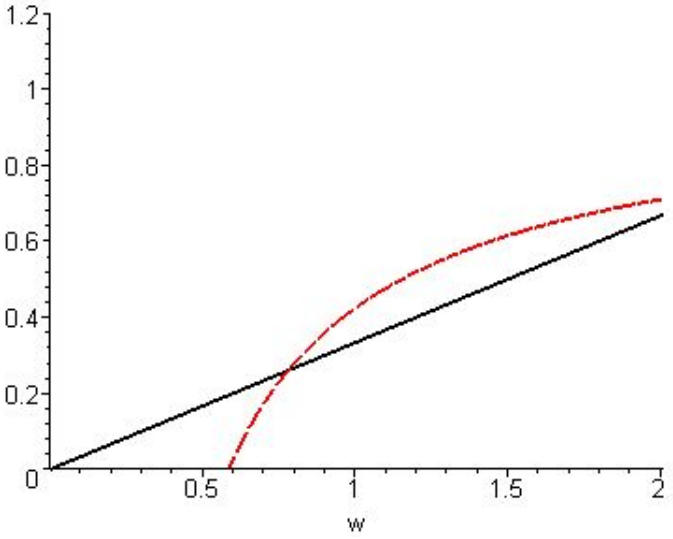

Figure 2.b

Graph of $\frac{\tau^*}{T}$ and $\frac{w}{z}$ for parameter values

$z = 3,\ T = 10,\ a_0 = 0.5,\ \alpha_0 = 0.5,\ \alpha_1 = 1.5,\ \delta = 0.9.$

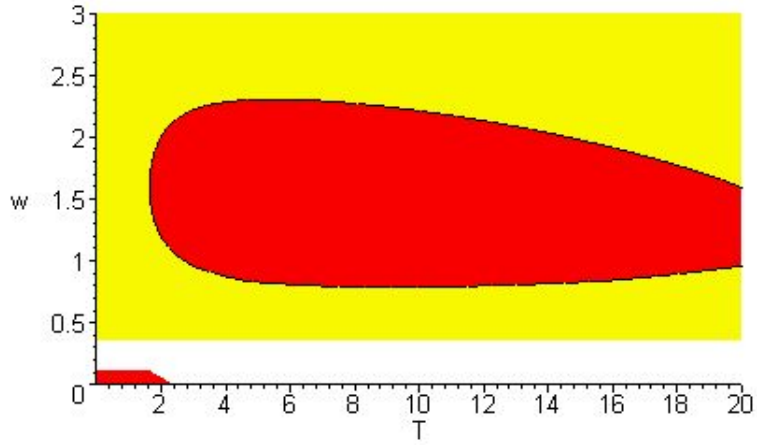

Figure 3.a

Region in $T - w$ space where principals choose to consult FR, given that IR is in common use (dark shaded).

Values: $z = 3,\ a_0 = 0.5,\ \alpha_0 = 0.5,\ \alpha_1 = 1.5,\ \delta = 0.9.$

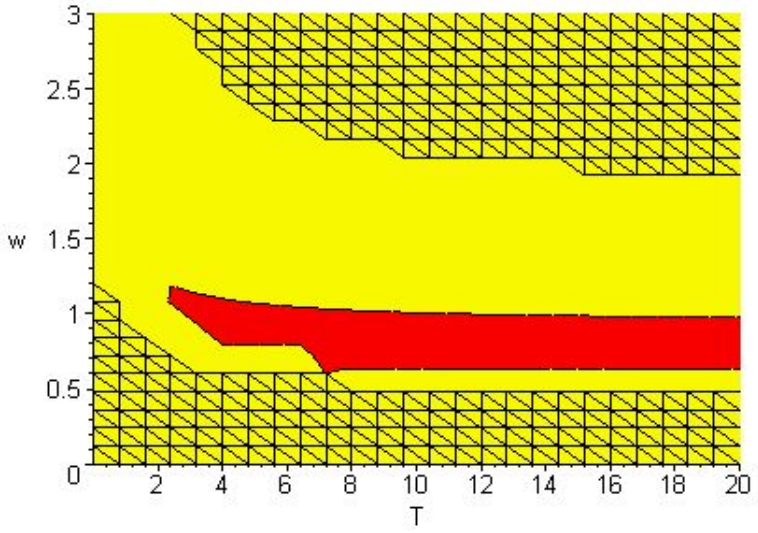

Figure 3.b

Region in $T - w$ space where FR is more efficient than IR (dark shaded).

Values: $, z = 3\ a_0 = 0.5,\ \alpha_0 = 0.5,\ \alpha_1 = 1.5,\ \delta = 0.9.$