



Industriens Utredningsinstitut

THE INDUSTRIAL INSTITUTE FOR ECONOMIC AND SOCIAL RESEARCH

A list of Working Papers on the last pages

No. 341, 1992

THE HONEST SOCIETY: STABILITY AND POLICY CONSIDERATIONS

by

Sten Nyberg

This is a preliminary paper. It is intended for private circulation and should not be quoted or referred to in publications without permission of the author. Comments are welcome.

September 1992

The honest society: Stability and policy considerations

Sten Nyberg¹

*The Industrial Institute for Economic and Social Research
Box 5501, 114 85 Stockholm*

Abstract

Occasionally calls are made for a moral restoration lest we become a society of liars and cheats. Is societal honesty inherently unstable and what is the social value of marginal changes? This paper addresses these issues in an evolutionary game framework with random matching where individuals may use costly safeguards to partially protect their transactions. It is shown that a high level of safeguard subsidies is merited. By contrast, sharply increased safeguard costs, e.g. soaring litigation costs, may initiate a process of disintegration of honesty in society. Simply returning to the initial cost level does not suffice to restore the previous equilibrium. If feasible, reestablishing honesty is likely to be very costly.

¹I am grateful to Ken Burdett and Jörgen Weibull for helpful comments and to Oliver Williamson for originally inspiring me to think about the subject. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

1. Introduction

In a society where people are able to rationally trust one another, cooperative undertakings can be realized without devoting considerable resources to contingency contracting and other precautions. In fact, in the absence of trust many cooperative ventures would not be viable. That societal morals may be important for economic prosperity has long been recognized. [See e.g. Banfield (1958)] However, even if honesty is collectively rational it is far from evident that it is rational for the individual to be honest. Akerlof (1983) discusses equilibrium honesty in a partial model where individual characteristics are observable. If agents interact with each other and moral standing is subject to choice, e.g. through upbringing, the game is likely to be of a prisoners dilemma type.²

More recently some models explaining the emergence of honest behavior as an outcome of an evolutionary process have appeared. [See e.g. Witt (1986), Frank (1987), (1989) and Harrington (1989).] The basic tenet common to all equilibrium stories about honesty is that the returns from being honest must be greater or equal to what can be obtained by being dishonest. If dishonest individuals differ from honest individuals only in that the dishonest are less restricted in their behavior, everyone would be dishonest in an evolutionary equilibrium. To allow for the emergence of trustworthy behavior it is sometimes assumed that there is some probability that other actors can identify a player's type, or that there is some cost involved in adopting the dishonest strategy. By contrast, in this paper evolutionary equilibria featuring honest behavior emerge because, in the spirit of Williamson (1985), honest types may use costly precautions to partially safeguard their transactions.³

The object of this paper is primarily to examine the level and stability of the equilibrium proportion of honest in the society, in an evolutionary framework, and to address the social welfare implications of policy measures like safeguard subsidies. First, in section 2, the basic model is presented and the conditions for existence of an interior equilibrium featuring both honest and "naïvely" dishonest agents are discussed. Section 3 deals with the effect of changes in safeguard costs on equilibrium outcomes and social welfare. In an equilibrium population featuring both honest and dishonest safeguard subsidies are found to

²See Ullman-Margalit (1977) for a treatise on the role of 'norms' in solving PD problems.

³The significance of costly dishonesty for transitions between equilibria is however considered in section 4.

increase social welfare. Conversely, it is shown that a moderate deterioration of societal trust, brought about by increased safeguard costs, can initiate a process of disintegration of societal honesty. Furthermore, simply restoring the conditions previously supporting a honest equilibrium is generally not sufficient to return from a dishonest situation. Finally, in section 4, the implications of some less restrictive assumptions about the types are analyzed. Individuals receive the option of abstaining from interaction if the expected utility falls short of the reservation level. Furthermore, a more sophisticated variety of dishonesty is introduced, allowing for honest behavior should that be more profitable.

2. The Model

The interaction between agents is modelled as a random matching game with a nonatomic population of players. The players can be thought of as engaging in team production where they share the fruits of their joint effort but where the individual effort level is difficult to observe [e.g. Alchian & Demsetz (1972), Holmström (1979)], or they could be viewed as participants in a transaction which involves asset specific investments and is subject to opportunistic behavior.

The players can be of one of two types; honest, who would never renege on a promise, and dishonest who do not feel compelled to honor any agreements. Honesty is viewed as a character trait which is relatively stable over time and not subject to conscious choice by the agent, unlike a strategy. In transactions between honest parties the cooperative outcome is attainable and the proceeds are shared equally yielding an individual payoff α . Whenever dishonest agents are involved the scope for synergies is diminished and the gross value of the interaction is 2β , where $\alpha > \beta$. When a dishonest player meets an honest player the former pockets the entire 2β whereas in an encounter between two dishonest individuals the proceeds are shared equally assuming they are equally skilled in deception. Though, scheming in vain is unproductive it is not as bad as trusting the other party and being cheated. Furthermore, cheating an honest agent, forfeiting synergies, is more profitable than sticking to it and sharing the proceeds, i.e. 2β is greater than α . The situation facing the interacting agents is structurally a prisoners dilemma situation.

FIGURE 1 ABOUT HERE

In an evolutionary framework the most successful types increase in frequency in the population. This could be thought of as parents adapting their upbringing to maximize the payoff of their offspring.⁴ Hence, the composition of the population will change over time so that the type receiving the highest expected payoff smoothly increases in frequency.

Let z be the difference in expected payoffs of the two types, $\pi_h - \pi_d$, to be defined later. Then the change of the proportion of dishonest in the population, p , can be described by any continuous dynamic $\dot{p}(z)$, defined for $p \in [0, 1]$, that is strictly decreasing in z and is zero for $z = 0$.⁵ Population proportions such that the population dynamic has a fixed point constitute dynamic equilibria. Furthermore, an equilibrium is asymptotically stable if there is some neighborhood of $\bar{p}$ such that any trajectory of the population dynamic originating in the neighborhood converges to $\bar{p}$. [van Damme (1987), Friedman (1991)]

In the game outlined above dishonesty dominates honesty and the only feasible equilibrium is a situation where everybody is dishonest. However, the introduction of safeguards may change that. Economic transactions differ greatly in their susceptibility to opportunistic behavior and based on their assessment of the riskiness of the transaction honest players may find it worthwhile to check the other party's credit history or to make provisions for a wider range of contingencies than those covered in a standard contract, etc, before engaging in a business relationship. The term safeguards will be used to denote all the various efforts to reduce exposure to opportunism.⁶

Safeguards are operationalized as the fraction, θ , of the maximum loss, $-\beta$, an honest agent will incur should he encounter a dishonest agent. Low θ s thus correspond to extensive precautions. Although prudent, writing extensive contracts and undertaking other protective

⁴In evolutionary biology models there are compelling genetic arguments for the frequency of a type to depend on its relative fitness. This is not necessarily the case in the social sciences where traits or behaviors are transferred through imitation and learning. [See e.g. Friedman (1991)]

⁵Since the population is at most bimorphic an expansion of one type in the population implies a contraction of the other. Thus the relative rate of change for several different types is of no concern and it suffices to use any continuous replicator dynamic such that the type receiving the highest payoff will increase in frequency.

⁶Carefully crafted contracts facilitates recouping losses in court following a breach of trust. In court proceedings other costs arise, such as litigation costs. Even though these arise after a defection by the other party they correspond to safeguard costs in that increased expected litigation costs essentially renders the taken precautions less effective. To achieve the same level of protection as before the increase more resources must be spent on safeguarding.

In a model featuring risk averse agents, increased uncertainty concerning the outcome of court proceedings will have a similar effect.

measures is certainly costly. The cost of a θ level of precaution is given by a continuous, twice differentiable, cost function $c(\theta)$ defined on $[0, 1]$, reflecting the safeguard technology. Safeguards are assumed to exhibit diminishing returns and enough so to make complete protection undesirable. A zero level of safeguards is however free.

When matching is random the probability of meeting a dishonest player equals their frequency in the population, p . Thus, the payoff accruing to honest players can be written as:

$$\pi_h = (1-p)\alpha + p(1-\theta)\beta - c(\theta) \quad (1)$$

Honest individuals choose the level of safeguards to maximize π_h . For all p greater than zero honest agents wish to take some precautions and the optimal θ is given by;

$$\theta = c'^{-1}(-p\beta), \quad \theta \in [0,1] \quad (2)$$

since $c'' > 0$, c' is one-to-one and thus has an inverse. In this section untrustworthy individuals are assumed to be naïvely dishonest, that is they only know how to cheat and are incapable of behaving honestly even if it would be more profitable to do so.

$$\pi_d = (1-p)(1+\theta)\beta + p\beta \quad (3)$$

Both types prefer to interact with honest counterparts and, as would be expected, the payoffs for both types increase as the proportion of honest in the population increases. This is easily seen by differentiating the payoffs with respect to p , using expression (2). In an almost entirely honest population the probability of being cheated is minuscule warranting only small safeguard expenses and preying on the honest pays off handsomely. Thus, unless safeguards are free, there can never be an equilibrium with only honest individuals. The best we can hope for is asymptotically stable equilibria containing some proportion of honest agents, participating in the interaction. Such equilibria will be referred to as good while equilibria featuring only dishonest types will be called bad.

The feasibility of different equilibrium types is determined by the parameter values in the model, (α, β) . Figure 2 illustrates a situation where both types of equilibria are feasible. There is a good equilibrium in p_1 , where the payoff functions intersect for the first time coming from the left. There is also a dishonest equilibrium in $p=1$. Of course p_2 is also an equilibrium point but it is not stable. Initial p 's in the interval $[0, p_2)$ yields convergence to p_1 , whereas p 's greater than p_2 will result in an asymptotically stable equilibrium where $p = 1$.

FIGURE 2 ABOUT HERE

Lemma 2.1: For all $c(\theta)$ there are (α, β) s.t. there exists a "good" equilibrium which is asymptotically stable.

Proof: Let $\alpha = r\beta$, $r > 1$, and consider p and β , $p\beta = m$, s.t. $c^{-1}(-m) = \bar{\theta}$ satisfies $[(1-p)(r-1) - \bar{\theta}] > 0$. Then, for a sufficiently large β $z(p) > 0$ implying $\dot{p} < 0$. Since $z(p)$ is continuous and $z(0) < 0$ there is at least one p , s.t. $z(p) = 0$, constituting an asymptotically stable equilibrium. $\square$

This means that as long as safeguards are reasonably cheap, compared to the interaction payoffs, then equilibria with some fraction of the population being honest are feasible. However, given a sufficiently high initial proportion of dishonest agents a degenerate dishonest equilibrium will always be reached and if safeguard cost are exorbitantly costly dishonesty may prevail for all initial p .⁷

Lemma 2.2: If $c''(\theta)$ exists and is ≤ 0 then there can only be one good equilibrium.

Proof: $c''(\theta) \leq 0$ ensures that $z(p)$ is concave.

Naturally, studying transitions between good and bad equilibria makes more sense in societies where good equilibria are feasible. Hence, through the remainder of the paper it is assumed that the parameters are such that both equilibrium types are feasible.

3. Safeguard costs and social welfare

In this section the effects of changes in safeguard costs with respect to equilibrium stability and social welfare will be discussed. The cost of safeguards can be affected directly through, for instance, subsidies for individuals seeking legal redress, lowering individual costs, or

⁷Harrington (1989), remarks in a comment to Frank's model that, "cooperative behavior need not arise as part of an evolutionary stable outcome, ..." and argues that with more plausible assumptions, and given payoffs, the decisive factor in this regard is whether the initial population has a sufficiently high proportion of honest agents.

indirectly through policies increasing the uncertainty of the outcome of this process, thereby raising costs.

Apart from strengthening the protection for the honest individual in a transaction additional safeguards also make dishonesty less attractive thus slightly reducing the proportion of dishonest people in equilibrium. This socially beneficial effect is not fully taken into account by honest individuals contemplating the appropriate level of safeguards.⁸ Thus there may be a case for subsidizing safeguards from a social welfare point of view. Welfare is simply assumed to be a population weighed average of the payoffs irrespective of whether individuals are honest or not. Social welfare is given by,⁹

$$S(p, \theta, \gamma) = (1-p)\pi_h + p\pi_d = (1-p)^2\alpha + (2-p)p\beta - (1-p)c(\theta) \quad (4)$$

Now, suppose the government contemplates subsidizing safeguards, financing it by levying a uniform tax on all citizens. Since there is a continuum of agents they do not perceive their choice of precautions to influence the tax and treat it as a fixed cost. Let γ denote the safeguard subsidy. Honest agents thus only pay $(1-\gamma)c(\theta)$ to obtain a θ level of protection. Note that a safeguard subsidy, γ , only affects social welfare indirectly through the propensity to invest in safeguards since the full cost of safeguards, $(1-p)c(\theta)$, still burdens society's resources, leaving expression (4) unchanged.

Proposition 3.1: In a good equilibrium with no safeguard subsidies the introduction of subsidies: (i) improves social welfare (ii) and continues to do so for $\gamma \leq 0.5$ (iii) reduces the investments in safeguards in equilibrium.

Proof: In appendix.

⁸In his analysis of individual precautions to prevent theft Shavell (1990) distinguishes between a diversion effect, where observable precautions make thieves choose other victims, and a theft reduction effect induced by the reduced profitability of theft in general. The former effect, which may cause potential victims as a group to overinvest in precautions, is not considered in this paper. The theft reduction effect, loosely corresponding to a decrease in p in the evolutionary model in this paper, is not fully appropriated by individuals thus leading to under investment.

⁹Social welfare here measures both the extent to which synergies are realized and the amount of resources spent on unproductive safeguards. If α is close to β increasing social welfare becomes a matter of minimizing safeguard costs in which case a degenerate dishonest equilibrium might well be preferable to an interior equilibrium. Furthermore, a stronger emphasis on the well-being of the honest would bias the analysis towards more subsidies.

Not surprisingly, subsidizing safeguards is initially beneficial from a social point of view. The level of subsidies implied is however quite high partly reflecting that subsidies is the only means available in the model to influence the level of honesty in society. Most societies do extend some type of safeguard subsidies. For instance, judicial systems are normally partially state funded, requiring the individual to pay only a fraction of the real litigation costs.

The third result may seem somewhat counter intuitive at first but is quite straightforward. The introduction of safeguard subsidies increases the payoffs of honest individuals, for a given level of safeguards, thereby causing the proportion of dishonest agents to go down which in turn make honest individuals invest less in safeguards.

Policies affecting the cost of safeguards may however have more than just marginal effects. In fact, in any society in a good equilibrium the social level of trust can degenerate when sufficiently undermined.

Proposition 3.2: (i) In any good equilibrium a sufficiently large increase in safeguarding costs will induce a transition to the bad equilibrium. (ii) A loss of trust is irreversible unless safeguards are free.

Proof: (i) Let safeguard costs be given by $\tau c(\theta)$. As τ increases so does the optimal θ given by equation 2. For a large enough τ θ becomes sufficiently large, i.e. close to one, to make $z(p)$ negative for all p . (ii) $z(1-\varepsilon) < 0$ for small ε if $\tau > 0$. $\square$

It could be argued that policies aimed at instilling higher moral standards, creating a stronger capacity for remorse in individuals or perhaps working to uphold commendable behavior through group pressure or ostracism., could restore lost trust. However, these mechanisms are likely to be most effective when deviations are rare and can be expected to increase compliance with an already widely held norm but to be quite ineffective when a majority challenges the "norm", i.e. they are preventive rather than corrective. The value of pushing a bimorphic equilibrium towards more honesty might not be negligible though.

Now, suppose that there is some cost involved in adopting the dishonest strategy, e.g. because deception is more mentally taxing than simply sticking to what was agreed upon.

Attempting to outsmart the other player is assumed to entail a small cost, $k \geq 0$.¹⁰ Thus, for sufficiently low safeguard costs a transition from a bad to a good equilibrium is feasible. However, this does not mean that a good equilibrium can be restored by recreating the conditions that prevailed before the loss of trust.

Proposition 3.3: (i) A loss of trust, in a state admitting both equilibrium types, induced by a cost increase $\Delta\tau$ cannot be reversed by $-\Delta\tau$.

Proof: Both before and after $\Delta\tau$ $z(1) < 0$, $p=1$ being an asymptotic equilibrium. Furthermore, z is monotonously decreasing in τ and thus $z(1) < 0$ for $\tau \in [\tau^0, \tau^0 + \Delta\tau]$. $\square$

Hence, there is an element of hysteresis in the transition and it is necessary to overshoot in order to return to a good equilibrium. This feature, sometimes referred to as a cusp catastrophe, is consonant with arguments cautioning us about the perils of becoming a people of liars and cheats. Although being a quite intuitively appealing property this is not captured in the standard evolutionary equilibrium model. The main point here is that the mere addition of the seemingly innocuous assumption that agents are allowed to undertake costly safeguards generates this feature. Moreover, it is robust in the sense that the analysis is valid for all safeguard technologies with diminishing returns.

If trustworthiness, or honesty, is thought of as a general trait which deterioration is not easily confined to specific aspects of behavior or particular transactions. That is to say that a person that behaves opportunistically in one aspect is not to be trusted in other matters either. Then the level of trustworthiness may be affected by any policy that change the relative payoffs of different types, for instance, a tax policy relying on honesty on the part of the taxpayers, thus favoring the dishonest relatively speaking, would shift a mixed equilibrium toward increased dishonesty. In fact, nice systems that credit its citizens, or users, with being responsible individuals may actually pose a threat to viability of a high level of honesty in a society and should perhaps be eschewed.

¹⁰ As Akerlof (1983) puts it "There is a return to appearing honest, but not to being honest. It pays parents to teach their children to be honest because the individually functional trait of appearing honest is jointly produced with the individually dysfunctional trait of being honest." The rationale for this, for "Fagans" disappointing, hypothesis is that it is costly to train children to be convincingly deceptive. Akerlof mentions daytime TV as anecdotal evidence in support of the scarcity of talent in this area.

4. Sophisticated strategies

The payoffs have thus far, for convenience, been assumed to exceed the individuals' reservation level. Relaxing this assumption leaves the analysis basically unchanged but may admit an interval of bad equilibria. Suppose agents require an expected payoff of at least β to participate. It could be thought of as the autarcic payoff. Thus the payoffs are now given by,

$$\hat{\pi}_i = \max\{\pi_p, \beta\} \quad (5)$$

which depicted graphically typically would look like;

FIGURE 3 ABOUT HERE

The dishonest equilibrium is now to be found in p^* where the honest part of the population prefer not to participate while the dishonest find participation to be weakly dominating. If there is some small fixed cost associated with being dishonest the only behavior that is nash in the participation choice and also constitutes a dynamic equilibrium is nonparticipation on the part of both groups, for all $p \in [p^*, 1]$.¹¹

Apart from the participation decision it could also be argued that it is not plausible that dishonest persons should cheat when it is contrary to their interests to do so. Allowing dishonest players to mimic honest behavior should that be profitable implies that they face the following payoff function,

¹¹There are several papers examining credibility problems on the individual interaction level using a incomplete information framework, notably Sobel (1984) and Dasgupta (1987). Agents meet and interact repeatedly while updating their prior beliefs about their partner's type. To entice dishonest types to reveal themselves defecting must be at least as attractive to them as pretending to be honest and enjoy the benefits of the partner's increased trust in them. Whether it is worthwhile to try to learn more about the other party and build trust or whether it is better not to interact, or to interact only in a risk free way, depends on the prior probability that the other party is honest. Building trust is always optimal from a social point of view and thus the prior constitutes a social capital.

If individual encounters take place randomly the prior probability corresponds to the proportion of the population being honest. The model in this paper, although being symmetric, could be interpreted as being analogous to a collapsed repeated game model endogenizing the prior within an evolutionary framework. The payoffs would then represent the total value of interacting over time with a specific type.

$$\pi_d = \max\{ \pi_h, (1-pq)(1+\theta)\beta \} \quad (6)$$

Proposition 4.1: There exist good equilibria for all $p \in [p^, 1]$ with the same payoff as in p^* .*

Proof: Let q be the proportion of dishonest actually acting dishonest. For $p \in [p^*, 1]$ q s.t. $pq = p^*$ yield perfect nash equilibria, in q , in the "stage game" and support an asymptotically stable equilibrium in the population game. $\square$

As long as π_d is greater than π_h all dishonest agents will naturally prefer to defect. Now consider a p large enough to push the dishonest payoff below that of the honest players. This will induce some dishonest agents to act as honest ones. This in turn implies that, "coming from a low p ", any p is compatible with a "good" equilibrium. However, if there were some cost associated with being dishonest or a lexicographic preference for honesty would make p^* the unique "good" equilibrium. There are of course still initial population proportion supporting degenerate dishonest equilibria. This means that for a range of initial p 's both types of equilibria are feasible.

The introduction of sophisticated dishonesty leaves the analysis basically unchanged although one difference compared to the case involving naïve dishonesty is that while a transition to a good equilibrium in that case requires a substantial change in the proportion of types a transition in the sophisticated case "merely" involves a coordinated change of strategy on the part of a sufficient number of dishonest agents. Even though this distinction is inconsequential in the model it is perhaps plausible that a change in the proportion of types would take considerable time whereas a coordinated change in strategies could be achieved much faster, and substantially cheaper.

These remarks of course simply concerns the properties of the model under these different assumptions. But they serve to point out that while a "loss of trust" most likely is a very serious matter indeed it is perhaps not be the abyss indicated by a naïve modelling approach.

5. Conclusions

The environment in which individual interactions take place determines the riskiness of the transactions and the relative payoffs to honest and dishonest individuals. Important environmental factors are the ease of monitoring, the efficacy of legal redress etc. In the paper all activities reducing the exposure to opportunism are summarized under the term safeguards. The basic assumption about safeguards is that it gets increasingly expensive to move toward complete protection, i.e. the technology is assumed to exhibit diminishing returns.

Both honest and dishonest benefit from a more honest population. Private investments in safeguards promote honesty benefiting everyone in the population. This externality is not taken into account by individuals causing under investment in safeguards. Thus, not surprisingly, safeguard subsidies are found to increase social welfare. More interestingly, the optimal level of subsidies can be shown to exceed 50%, perhaps to some extent reflecting that safeguards constitute the only means in the model by which the level of honesty can be influenced. Furthermore, investments in safeguards turn out to decrease with subsidization because of the decrease in the fraction of dishonest in the population resulting from safeguard subsidies.

Conversely, transient increases in safeguard costs may prompt a transition to a dishonest equilibrium thereby doing lasting damage to the social trust capital. A return to an honest equilibrium generally requires more than just revoking the policy that gave rise to the shift, i.e. it requires some degree of overshooting. This, is in concordance with popular views of pendulum motions in societal evolution, swings from lax to austere regimes etc. Smooth adjustments simply won't do.

References

- Alchian, Armen, A. and Harold Demsetz, 1972, Production, Information, Costs, and Economic Organization, *American Economic Review*, 62, 777-795.
- Akerlof, George, 1983, Loyalty Filters, *American Economic Review*, 73, 54-63.
- Banfield, Edward C., 1958, *The Moral Basis of a Backward Society* (The Free Press).
- Dasgupta, Partha, 1988, Trust as a Commodity, In D. Gambetta, ed. *Trust: Making and Breaking of Cooperative Relationships*, 49-72, (Oxford: Blackwell Publisher).
- Frank, Robert H., 1987, If homo economicus could choose his own utility function, would he want one with a conscience?, *American Economic Review*, 77, 593-604.

- Frank, Robert H., 1989, If homo economicus could choose his own utility function, would he want one with a conscience?: Reply *American Economic Review*, 79, 594-596.
- Friedman, Daniel, 1991, Evolutionary games in economics, *Econometrica* 59, 637-666.
- Harrington, Joseph E., 1989, If homo economicus could choose his own utility function, would he choose one with a conscience?: Comment, *American Economic Review*, 79, 588-593.
- Holmström, Bengt, 1979, Moral hazard and observability, *Bell Journal of Economics* 13, 324-340.
- Shavell, Steven, 1990, Individual Precautions to Prevent Theft: Private versus Socially Optimal Behavior, *NBER working paper* # 3560.
- Sobel, Joel, 1985, A theory of credibility, *Review of Economic Studies*, 52, 557-573.
- Ullman-Margalit, Edna, 1977, *The Emergence of Norms*, (Oxford University Press).
- van Damme, Eric, 1987, *Stability and Perfection of Nash Equilibria*, (Springer-Verlag).
- Witt, Ullrich, 1986, Evolution and stability of cooperation without enforceable contracts, *Kyklos*, 39, 245-266.
- Williamson, Oliver E., 1985, *The Economic Institutions of Capitalism* (The Free Press).

APPENDIX

The marginal effect of subsidizing safeguards in a good equilibrium is given by

$$\frac{dS(\cdot)}{d\gamma} = \frac{\partial S(\cdot)}{\partial p} \frac{\partial p(\cdot)}{\partial \gamma} + \frac{\partial S(\cdot)}{\partial \theta} \frac{\partial \theta(\cdot)}{\partial \gamma} \quad (\text{A1})$$

where,

$$\frac{\partial S(\cdot)}{\partial p} = -2(1-p)(\alpha-\beta) + c(\theta) \quad (\text{A2})$$

and

$$\frac{\partial S(\cdot)}{\partial \theta} = -(1-p)c'(\theta) = (1-p)p \frac{\beta}{1-\gamma} > 0 \quad (\text{A3})$$

The first order condition on investments in safeguards, $G(p, \theta, \gamma) = c'(\theta) + p\beta/(1-\gamma) = 0$, and the equal profit condition, satisfied in a dynamic equilibrium, $F(p, \theta, \gamma) = (1-p)(\alpha-\beta) - \theta\beta - (1-\gamma)c(\theta) = 0$ implicitly defines p and θ in terms of γ . Substituting $\theta = c'^{-1}(-p\beta)$ into $F(p, \theta, \gamma)$, thus making it a function of p and γ alone, and applying the implicit function theorem yields

$$\frac{dp}{d\gamma} = - \frac{c(\cdot) - (\beta + (1-\gamma)c'(\cdot)) \frac{\partial \theta}{\partial \gamma}}{-(\alpha - \beta) - (\beta + (1-\gamma)c'(\cdot)) \frac{\partial \theta}{\partial p}} = - \frac{c(\cdot) - \beta(1-p) \frac{\partial \theta}{\partial \gamma}}{-(\alpha - \beta) - \beta(1-p) \frac{\partial \theta}{\partial p}} \quad (\text{A4})$$

The last equality follows from $c'(\theta) = -p\beta/(1-\gamma)$. The partial derivatives in expression (A4) are derived from $G(p, \theta, \gamma) = 0$ using the implicit function theorem

$$\left. \frac{\partial \theta}{\partial \gamma} \right|_{p=\bar{p}} = - \frac{1}{c''(\cdot)} \frac{p\beta}{(1-\gamma)^2} < 0$$

$$\left. \frac{\partial \theta}{\partial p} \right|_{\gamma=\bar{\gamma}} = - \frac{1}{c''(\cdot)} \frac{\beta}{1-\gamma} < 0$$

Thus, the numerator of (A4) is clearly positive. The denominator can also be demonstrated to be positive by utilizing that in a stable equilibrium it must be true that

$$\frac{\partial \pi_h}{\partial p} > \frac{\partial \pi_d}{\partial p}$$

that is,

$$-\alpha + (1-\theta)\beta > -\theta\beta + (1-p)\beta \frac{\partial\theta}{\partial p}$$

Hence, expression (A4) is strictly negative and subsidies can thus be seen to reduce the proportion of dishonest individuals in the population.

Similarly, differentiating $G(p, \theta, \gamma)$, letting p be given implicitly by $F(p, \theta, \gamma)$, gives us

$$\frac{d\theta}{d\gamma} = - \frac{\frac{p\beta}{(1-\gamma)^2} + \frac{\beta}{1-\gamma} \frac{\partial p}{\partial \gamma}}{c''(\theta) + \frac{\beta}{1-\gamma} \frac{\partial p}{\partial \theta}} \quad (\text{A5})$$

where the partial derivative of p wrt γ is positive making the numerator positive.

$$\left. \frac{\partial p}{\partial \gamma} \right|_{\theta=\bar{\theta}} = \frac{c(\bar{\theta})}{\alpha - \beta} > 0$$

$$\left. \frac{\partial p}{\partial \theta} \right|_{\gamma=\bar{\gamma}} = - \frac{\beta + (1-\gamma)c'(\bar{\theta})}{\alpha - \beta} = - \frac{(1-p)\beta}{\alpha - \beta} < 0$$

The denominator of expression (A5) can be shown to be proportional, with the reverse sign, of the denominator in expression (A4) and is thus negative. Consequently expression (A5) is negative, that is subsidization of safeguards will reduce the investments in safeguards in equilibrium.

Insertion of the derivatives in expressions (A2)-(A5) into expression (A1) yields that social welfare increases with subsidization of safeguards up to the point where (A2) reverses sign and becomes sufficiently positive to dominate expression (A1). Using $F(p, \theta, \gamma)$ expression (A2) can be written

$$\frac{\partial S(\cdot)}{\partial p} = -2\theta\beta - (1-2\gamma)c(\theta)$$

For that to happen the degree of subsidization, γ , must exceed 0.5.

	H	D
H	α, α	$0, 2\beta$
D	$2\beta, 0$	β, β

Figure 1

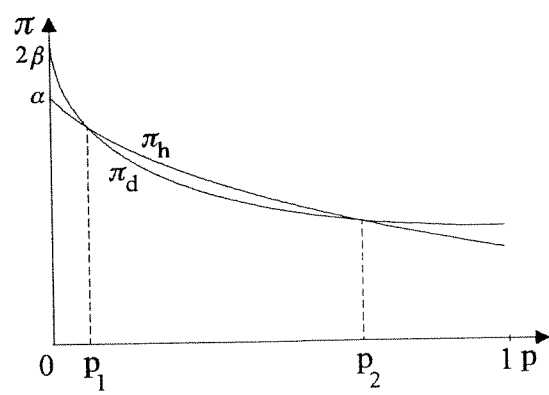


Figure 2

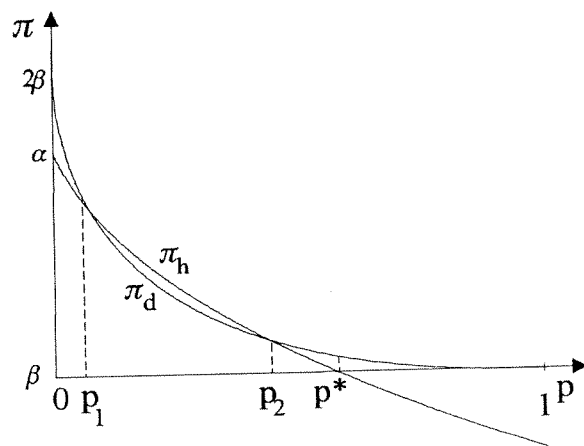


Figure 3

