

Dufwenberg, Martin; Lundholm, Michael

**Working Paper**

## Social Norms and Moral Hazard

Working Paper, No. 1997:28

**Provided in Cooperation with:**

Department of Economics, Uppsala University

*Suggested Citation:* Dufwenberg, Martin; Lundholm, Michael (1997) : Social Norms and Moral Hazard, Working Paper, No. 1997:28, Uppsala University, Department of Economics, Uppsala

This Version is available at:

<https://hdl.handle.net/10419/82925>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Social norms and moral hazard\*

Martin Dufwenberg<sup>†</sup> and Michael Lundholm<sup>‡</sup>

April 20, 1998

## Abstract

The probability of income loss depends on talent and effort. Effort has positive externalities and therefore individuals are awarded status in proportion to their perceived diligence. The social norm requires more effort from individuals perceived as more talented, but talent is private information and individuals cunningly choose effort so as to manipulate the public perception of their talent. We analyze the workings of a social insurance system in this setting. It turns out that social norms may mitigate moral hazard. However, the distribution of social status in society will not be uniform.

JEL: A14, C72, D81, D82, G28, H55, I38, J65.

Keywords: Moral hazard, social insurance, social norms, and status.

---

\*We thank Ernst Fehr, Kari Lantto, Åsa Rosén, and seminar participants at FIEF, Humboldt University, Lund University, Uppsala University, the University of Zürich, and the 1997 Econometric Society and European Economic Association meetings in Toulouse for valuable comments. The research is funded by the Swedish Council for Research in the Humanities and Social Sciences (Grant F212/95 “Social insurance, norms and individual incentives”).

<sup>†</sup>Dept. of Economics, Uppsala University. E-mail: martin.dufwenberg@nek.uu.se .

<sup>‡</sup>Dept. of Economics, Stockholm University. E-mail: mlu@ne.su.se .

# 1 Introduction

A welfare state provides insurance for its citizens. However, the greater the generosity of an insurance scheme, the less an insuree cares if a bad outcome occurs, and the less incentive he has to take adequate precautions. Such *moral hazard* considerations have received ample attention in the insurance economics literature.<sup>1</sup> However, it is seldom recognized that social norms may constrain behavior in the situations analyzed: A careless insuree typically shifts the costs of his behavior onto others and for this reason he may be accorded low social status (see Coleman [7, p.275]).<sup>2</sup> This is perhaps especially true if the insuree is conceived of as a talented fellow who, with high likelihood, could have avoided the bad outcome had he tried. To keep his status high, the insuree may therefore decide not to be so careless after all. On the other hand, if the social norm relates to perceived talent, the insuree may cunningly choose his actions so as to fake a level of talent such that others bestow high status on him.

In this paper we take a first step towards analyzing issues of this kind. The analysis has some potential bearing on many kinds of insurance relationships, but in order to focus matters we phrase the presentation in terms of the following situation: A population of individuals differ in talent, which affects how efficiently they transform effort (e.g., education or search activity) into probability of getting a job. The individuals are risk averse and there is some scope for welfare enhancing unemployment insurance. However, the higher is the unemployment benefit, the less incentive there are for individuals to exert effort, which in turn raises unemployment rates at everyone's expense. "Lazy" behavior is therefore not socially approved, especially if a lazy person is regarded as having high talent. A social norm governs how much effort an individual perceived to have a certain talent should exert, and this target is increasing in perceived talent. Social status is assumed to be an increasing function of the difference between an individuals actual choice of effort and the relevant target.

Individual talent is not easily inferred from outward appearances. We assume that each individual's talent is private information. Hence, others have to draw their conclusions from observing choices of effort only. This adds considerable intricacy to the motivation which affects individual choices of effort. Not only does effort influence employment probability, it also conveys a signal about personal talent, and this signal may be manipulated.

---

<sup>1</sup>See e.g., Arrow [2], Ehrlich & Becker [9], Pauly [19] and Shavell [21]. A general overview on moral hazard is Dutta & Radner [8].

<sup>2</sup>Coleman argues that for social rewards to be effective "[t]he existence of externalities is a necessary condition."

We solve the model using techniques from information economics. We apply an equilibrium refinement, a social equilibrium, which is akin to the D1 refinement of sequential equilibrium and show that there is a unique social equilibrium, which furthermore is fully separating.

In the social equilibrium there is for all levels of talent less moral hazard than would be the case if social status was not important. This finding is not a priori obvious, because although status is positively related to effort for a given level of perceived talent, individuals also have the incentive to imitate less talented individuals by exerting less effort. We also show that when the social insurance benefit is marginally increased individuals with low levels of talent will increase effort, which is in sharp contrast to standard predictions. We conclude, however, that one should be cautious when making welfare judgements based on these results: People with low talent work hard even though the positive externality they create is small. Moreover, despite the social norm requiring more talented individuals to work harder, social status in the model is increasing in talent. Having social norms to reduce moral hazard is not a free lunch.

The paper is structured as follows. The model is introduced in Section 2 and solved for equilibrium behavior in Section 3. In Section 4 we discuss the equilibrium in terms of moral hazard, marginal policy changes, and social status. We also discuss alternative modeling choices. We conclude the paper in Section 5. An Appendix contains the proofs of some lemmata used to prove our main result. All figures are found at the end of the paper.

## 2 The model

We begin by describing in detail the model, discussing in turn its strategic structure (2.1), material payoffs (2.2.), social payoffs (2.3) and individual objectives (2.4).

### 2.1 Effort, talent, unemployment

By exerting *effort*  $x \in X = [0, 1]$  individuals affect their probability of getting employed. This effort may have many interpretations, e.g., time or attention devoted to search activities on the labor market or to acquiring an education. Individuals differ according to their *talent*  $t \in T = [0, 1]$ . There is a continuum of individuals whose talent is distributed on  $T$  according to the distribution function  $F$ , which is strictly increasing, continuously differentiable, and has a density function  $f$ . We focus on the case when each individual's level of talent is private information. However,  $F$  and  $f$  are com-

mon knowledge among all individuals. An individual's probability of getting employed equals  $tx$ , and hence depends on both his effort and his talent.

Figure 1 illustrates the structure of the decision problem faced by an individual of talent  $t$ . First, knowing his talent  $t$ , the individual chooses an effort level  $x \in X$ . Then, a chance move determines whether the individual gets a job or becomes unemployed. The relevant probabilities are given in brackets in Figure 1.

## 2.2 Material payoffs

An individual with a job has positive income from labor, normalized to one.<sup>3</sup> No private unemployment insurance exists but we assume there exists a public unemployment insurance system. The system is financed through general public funds. The administrator of the insurance system *only* observes whether an individual has a job or not but is unable to observe an individual's talent or effort level. Since effort is unobservable for the administrator unemployment insurance is a uniform unemployment benefit  $\beta$  received by an unemployed individual regardless of his talent and his ex ante chosen effort to avoid unemployment.

In standard models that do not include social status concerns the individual's payoff depends only on income (from labor or social insurance) minus the cost of effort. We call this payoff the *material payoff*. Given talent  $t \in T$ , effort  $x \in X$ , and social insurance benefit  $\beta$  the material payoff is given by

$$tx + (1 - tx)u(\beta) - \frac{K}{2}x^2 \quad (1)$$

where the income valuation function  $u : \mathbb{R}_+ \rightarrow \mathbb{R}$  is normalized so that  $u(0) = 0$ ,  $u(1) = 1$ , and satisfies that  $u' > 0$ ,  $u'' < 0$ ,  $u' \rightarrow +\infty$  as  $y \rightarrow 0$ .  $K$  is a positive constant large enough to guarantee any  $x = 1$  is a strictly dominated strategy. (An explicit restriction on  $K$  is specified in equation (5) in section 2.4).

Given our assumptions about  $u$ , for reasonable levels of the marginal cost of public funds, it is easy to find a social welfare function such that some  $\beta \in (0, 1)$  is the optimal benefit level. In the following we consider only such benefits.

We close this subsection by recording how an individual of talent  $t$  motivated solely by his material payoff would behave. He would choose  $x$  so as to maximize (1). Irrespective of the behavior of others, there is a unique such

---

<sup>3</sup>This means that there is only one type of job, unlike in some other models where social concerns play a role, like Fershtman et al [10], Fershtman & Weiss [11], Gottfries & McCormick [14], McCormick [18].

choice. The optimal choice for each  $t \in T$  is described by the function  $x^{STD}$  defined by

$$x^{STD}(t) = \frac{1}{K}(1 - u(\beta))t. \quad (2)$$

Note that for all  $t > 0$  this specification entails moral hazard in the sense that the higher is  $\beta$ , the lower will be  $x^{STD}(t)$ .

### 2.3 Social norms and social payoffs

Individuals care about *social status* which they bestow on one another. An individual's status depends on his choice of effort and on how talented he is perceived to be. Although individual talent is private information it can be inferred from an individual's choice of effort, which is observable to other individuals. The motivation for why effort matters is that, with  $\beta > 0$ , an individual's effort has positive external effects, since effort reduces the probability of income loss and social insurance benefits are paid through public funds. Therefore, others may be inclined to despise a lazy individual and to hold a diligent individual in high esteem.<sup>4</sup>

We assume that individuals are accorded zero social status if they are perceived of as maximizing their material payoff. If they make some other choice, they are awarded status in proportion to the difference between their actual choice of effort and the choice of effort that would maximize their material payoff for the relevant level of perceived talent. More specifically, an individual who chooses effort  $x$  and is perceived to have talent  $t \in T$  is accorded a *social status* of  $x - x^{STD}(t)$  and a *social payoff* of  $\sigma(x - x^{STD}(t))$ , where  $\sigma$  is a non-negative number (common to all individuals) measuring sensitivity to social status and  $x^{STD}(t)$  is given by (2). We shall refer to  $x^{STD}(t)$  as the *social norm* for an individual perceived to have talent  $t$ .<sup>5</sup>

It remains to be explained *how* perceptions of an individual's talent come about. Individuals have private information about their own talent, so others can only draw inferences based on observations of the complete vector of chosen actions across the population.<sup>6</sup> We adapt standard tools of information economics which guarantee that, in equilibrium, all individuals make the

---

<sup>4</sup>If the reduction of government spending is *per se* something all persons desire, the following quote from Coleman [7, p.274] captures the spirit of our modeling: "If a number of persons' interests are satisfied by the same outcome, then each has an incentive to reward the others for working toward that outcome. Each may in fact find it in his interest to establish a norm toward working for that outcome, with negative sanctions for shirking and positive sanctions for working toward the common goal."

<sup>5</sup>A similar approach to modeling social status is used by Kandel & Lazear [16].

<sup>6</sup>This aspect resembles Bernheim's [3] model of conformity, except that in his model

same inferences if an individual with a certain talent takes a certain action. The details are discussed in Section 3. For the time being we postulate the existence of an *inference function* which, for each level of talent and action chosen, specifies a perceived talent. Let  $\tau : X \rightarrow T$  denote such a function;  $\tau(x) \in T$  is the perceived talent of an individual who chooses effort  $x \in X$ .

Given an inference function  $\tau$ , the social payoff of an individual of talent  $t$  choosing  $x$  can now be written as  $\sigma(x - x^{STD}(\tau(x)))$  or, using (2), as

$$\sigma\left(x - \frac{1}{K}(1 - u(\beta))\tau(x)\right). \quad (3)$$

## 2.4 Total payoffs

We assume that each individual's material and social payoffs (1) and (3) add up to his *total payoff*. An individual of talent  $t$  choosing  $x$  who is perceived to be of talent  $\tau(x)$  hence has total payoffs as given by the utility function  $U : X \times T^2 \rightarrow \mathbb{R}$  defined by

$$\begin{aligned} U(x, t, \tau(x)) &= tx + (1 - tx)u(\beta) - \frac{K}{2}x^2 \\ &+ \sigma\left[x - \frac{1}{K}(1 - u(\beta))\tau(x)\right]. \end{aligned} \quad (4)$$

In order to make  $x = 1$  a strictly dominated choice we assume that

$$K > 1 - u(\beta) + \sigma. \quad (5)$$

The interpretation is that no individual finds it worthwhile to spend all his time to reduce the risk of losing income.

We note that these preferences satisfy the so-called single-crossing property. To see this consider the slope of indifference curves given by

$$\left.\frac{dx}{d\tau}\right|_{dU=0} = \frac{\frac{\sigma}{K^2}(1 - u(\beta))}{\frac{1}{K}[(1 - u(\beta))t + \sigma] - x} \quad (6)$$

and the decrease in slope over talents

$$\frac{\partial}{\partial t} \left( \left.\frac{dx}{d\tau}\right|_{dU=0} \right) = -\frac{\frac{\sigma}{K^3}(1 - u(\beta))^2}{\left(\frac{1}{K}[(1 - u(\beta))t + \sigma] - x\right)^2} < 0. \quad (7)$$

---

the counterpart to our social norm is constant over types. Our model furthermore shares with Bernheim's the feature that individuals care *directly* about other's inferences. The resulting signaling game is non-standard; usually the inferring parties take *actions* about which others are concerned. However, as noted by Bernheim [3, footnote 6], the assumption we make has no formal significance and the analysis can be recast in more traditional ways.

### 3 Social equilibrium

We solve the model for equilibrium behavior, requiring that individuals of all talents maximize total payoffs given the equilibrium inference function. Furthermore, the inference function must be reasonable in the following two ways. First, the perceived talent associated with any effort level chosen by some individual in equilibrium should correspond to the expected level of talent of the individuals making that equilibrium choice. This means that if, in some equilibrium, the effort level  $x$  is chosen only by individuals with talent  $t$ , then  $\tau(x) = t$ . If, on the other hand, individuals of many different talents pool at  $x$ , then the inference gives the expected talent of an individual drawn at random from the pool.

Second, we impose a requirement on inferences concerning choices of effort that do not occur in equilibrium. To motivate this suppose for a moment that such a requirement is *not* imposed. Consider the strategy profile where each individual, irrespective of his talent, chooses an effort level of  $\frac{1}{K}(1 - u(\beta) + \sigma)$ . This means that all individuals pool at the level of effort which would maximize the total payoff of the most talented individual had his talent been common knowledge.<sup>7</sup> For  $\sigma$  high enough, this profile is sustained as an equilibrium by the inference function  $\tau$ , defined by  $\tau(\frac{1}{K}(1 - u(\beta) + \sigma)) = \int_0^1 tf(t)dt$  and  $\tau(x) = 1 \forall x \neq \frac{1}{K}(1 - u(\beta) + \sigma)$ . In words, each individual is perceived to have a talent equal to the expected talent of an individual drawn at random from the whole population and no individual dares to deviate because he would automatically be perceived as the individual with the *highest* talent, even for deviation towards *lower* levels of effort!

We find such an equilibrium questionable because it is founded on questionable out-of-equilibrium inferences. The problem is very similar to that which arises in signaling games when the sequential equilibrium concept is applied. In that context, a standard reaction is to invoke an equilibrium refinement which restricts possible inferences concerning out-of-equilibrium actions. The equilibrium refinement we apply is an adaptation of the D1 equilibrium (see Cho & Kreps [6] and cf. Bernheim [3]). In the context of the present model the D1 criterion mandates that the inference concerning any out-of-equilibrium action must be on a talent level of an individual with the *strongest* incentive to deviate under that particular inference. Compare this inference with the previous example where the out-of-equilibrium inference was on the talent level of the individual with the *weakest* incentive to deviate!

We now define the concept of *social equilibrium*, which we shall use to

---

<sup>7</sup>To see this, substitute a 1 for  $\tau(x)$  in (4), and maximize with respect to  $x$ .



solve the model when  $\sigma > 0$  and  $\beta \in (0, 1)$ . Let  $x^{SE} : T \rightarrow X$  be a function which describes which effort levels are chosen by individuals with different talents in the social equilibrium. We refer to the image of  $x^{SE}$  as “equilibrium efforts” and to complementary effort levels as “out-of-equilibrium efforts”. Let  $\tau^{SE} : X \rightarrow T$  describe inferences in the equilibrium.

**Definition.** A *social equilibrium* is a pair of functions  $(x^{SE}, \tau^{SE})$  such that

- (i)  $x^{SE}(t) \in \operatorname{argmax}_{x \in X} U(x, t, \tau^{SE}(x)) \forall t \in T$ ,
- (ii) If  $x \in X$  is an equilibrium effort, then  $\tau^{SE}(x) = \frac{\int_P t f(t) dt}{\int_P f(t) dt}$  where  $P = \{t : x^{SE}(t) = x\}$ , and
- (iii) If  $x$  is an out-of-equilibrium effort, then

$$\tau^{SE}(x) \in \operatorname{argmax}_{t \in T} \left[ U(x, t, \tau^{SE}(x)) - U(x^{SE}(t), t, \tau^{SE}(x^{SE}(t))) \right].$$

Condition (i) in the Definition requires optimality of chosen actions, conditions (ii) and (iii) correspond to the requirements on the inference function discussed before.<sup>8</sup> We now solve the model:

**Theorem.** *There exists a unique social equilibrium  $(x^{SE}, \tau^{SE})$ .  $x^{SE}$  is continuous and strictly increasing (full separation), strictly convex, and satisfies that  $x^{SE}(0) \in (0, \frac{\sigma}{K})$ , and  $x^{SE}(1) = \frac{\sigma}{K}(1 - u(\beta)) < 1$ .  $\tau^{SE}$  satisfies that*

$$\tau^{SE}(x) = \begin{cases} 0 & \forall x \in [0, x^{SE}(0)) \\ t & \forall x \in [x^{SE}(0), x^{SE}(1)] \\ 1 & \forall x \in (x^{SE}(1), 1]. \end{cases}$$

To prove the Theorem we will make use of the following four lemmata, which apply to any social equilibrium  $(x^{SE}, \tau^{SE})$  (proofs in the appendix):

**Lemma 1.** *If  $t < t'$ , then  $x^{SE}(t) \leq x^{SE}(t')$ .*

**Lemma 2.** *If  $x < x'$ , then  $\tau^{SE}(x) \leq \tau^{SE}(x')$ .*

**Lemma 3.** *The set of equilibrium efforts is connected.*

---

<sup>8</sup>A social equilibrium is not formally equivalent to a D1 equilibrium because condition (ii) (of the Definition) is not quite Bayes' rule and (iii) is not the D1 criterion. These changes are made to simplify the presentation of (ii), and because we prefer the economic intuition behind (iii) to that of the D1 criterion. Both changes are inconsequential to equilibrium behavior. In particular, in our game (iii) and the D1 criterion affect behavior in the same way.

**Lemma 4.** *No pool exists, i.e.,  $\nexists t, t' \in T$  s.t.  $t \neq t', x^{SE}(t) = x^{SE}(t')$ .*

These lemmata show in turn that the equilibrium effort and inference functions  $x^{SE}$  and  $\tau^{SE}$  are monotone, that the effort levels chosen in equilibrium form a connected set, and that there are no pools.

*Proof of the Theorem:*<sup>9</sup> Combining Lemmata 1, 3, and 4 one sees that  $x^{SE}$  must be strictly increasing and continuous. Moreover,  $x^{SE}$  must be differentiable; in order to prescribe equilibrium choices for all talents  $x^{SE}$  must for each  $t$  be tangent to an indifference curve as given by (6), and this would be impossible if  $x^{SE}$  had a “kink” (either there would be no point of tangency at the kink and individuals with talent  $t$  would want to deviate, or there would be multiple tangencies at the kink and a pool would be attracted). For equilibrium efforts,  $\tau^{SE}$  is the inverse of  $x^{SE}$ , and hence  $\tau^{SE}$  must there be differentiable too. Hence we can proceed by maximizing (4) with respect to  $x$  to get the first order condition  $\tau'(x) = K \frac{(1-u(\beta))t + \sigma - Kx}{\sigma(1-u(\beta))}$ . Since inferences are correct we can substitute  $\tau(x)$  for  $t$  to get

$$\tau'(x) = K \frac{(1-u(\beta))\tau(x) + \sigma - Kx}{\sigma(1-u(\beta))}. \quad (8)$$

This is a linear first order differential equation which can be solved by standard methods. However, we need an initial condition. Since in a separating equilibrium true talents are revealed, the individual with talent 1 will have that highest conceivable social norm apply to him, and this could never be consistent with equilibrium behavior unless he chooses the effort level that would be optimal for him had his level of talent been commonly know. Substituting a 1 for  $\tau(x)$  in (4), and maximizing with respect to  $x$  one concludes that he chooses an effort of  $\frac{1}{K}(1-u(\beta)+\sigma)$ , so the appropriate initial condition is  $\tau\left(\frac{1}{K}(1-u(\beta)+\sigma)\right) = 1$ . The definite solution to (8) then is

$$\tau(x) = \frac{K}{1-u(\beta)}x - \frac{\sigma}{1-u(\beta)}e^{\frac{K}{\sigma}\left(x - \frac{1}{K}(1-u(\beta)+\sigma)\right)}. \quad (9)$$

Equation (9) describes the equilibrium inference function  $\tau^{SE}$  for equilibrium levels of effort. Since equilibrium efforts increase with talent (Lemma 1 + separation), effort levels above  $x^{SE}(1) = \frac{1}{K}(1-u(\beta)+\sigma)$  are irrelevant in (9).

---

<sup>9</sup>Ramey [20] proves existence, uniqueness, and full separation of D1 equilibria in a wide class of signaling games. However, the game we study does not belong to that class (unlike Ramey, we have bounded strategy sets), so despite the similarity between the equilibrium concept we use and that of D1 equilibrium (cf. footnote 8), the proof cannot be shortened by referring to Ramey’s results.

Moreover, effort levels below  $x^{SE}(0)$  are irrelevant. By Lemma 2,  $\tau^{SE}(x) = 0$  for any  $x \in [0, x^{SE}(0))$  and  $\tau^{SE}(x) = 1$  for any  $x \in (x^{SE}(1), 1]$ .

$x^{SE}$  is the inverse of  $\tau^{SE}$  for equilibrium effort levels. Hence substituting  $t$  for  $\tau(x)$  and  $x^{SE}(t)$  for  $x$  in (9) we get the following expression which implicitly defines equilibrium efforts for all talents in  $T$ :

$$x^{SE}(t) = \frac{(1 - u(\beta))t}{K} + \frac{\sigma e^{\frac{K}{\sigma}(x^{SE}(t) - \frac{1}{K}(1 - u(\beta) + \sigma))}}{K} \quad (10)$$

Letting  $t = 1$  it is straightforward to verify that  $x^{SE}(1) = \frac{1}{K}(1 - u(\beta) + \sigma)$ . Letting  $t = 0$ , and noticing that the exponential expression in (10) (in the following denoted  $e^{\frac{K}{\sigma}(\cdot)}$ ) is less than unity (follows from Lemma 1 + separation), one sees that  $x^{SE}(0) = \frac{\sigma}{K}e^{\frac{K}{\sigma}(\cdot)} \in (0, \frac{\sigma}{K})$ . Finally, since  $\frac{\partial^2 x^{SE}}{\partial t^2} = \frac{1 - u(\beta)}{K\sigma} \left(1 - e^{\frac{K}{\sigma}(\cdot)}\right)^3 > 0$ ,  $x^{SE}$  is strictly convex.  $\square$

The effort choices in the social equilibrium given by (10) are illustrated in Figure 2. The graphs  $I_t^*$  and  $I_1^*$  display the indifference curves for the individuals with talent levels  $t$  and 1 which are tangent to  $x^{SE}$ . The points of tangency are then  $(t, x^{SE}(t))$  and  $(1, x^{SE}(1))$ . The inference function  $\tau^{SE}$  is illustrated by Figure 3.

## 4 Discussion

In this section we analyse the social equilibrium in terms of moral hazard, marginal policy changes, and social status. We also discuss some possible alternative modeling choices.

### 4.1 Moral hazard

We define moral hazard as the change in individual effort caused by the introduction of social insurance. We assume that social status is important only because effort has negative externalities when  $\beta > 0$ . Then, with  $\beta = 0$  social status is unimportant and  $\sigma = 0$ . We refer to the situation with  $\beta = \sigma = 0$  as *autarchy* (*AUT*). In autarchy individuals choose effort so as to maximize (1) with  $\beta = 0$ , in which case effort is given by  $x^{AUT}(t) = \frac{t}{K}$ . Hence our measure of moral hazard for an individual of talent  $t$  in the social equilibrium is  $x^{AUT}(t) - x^{SE}(t)$ . We now compare this measure to two others which would be relevant under different assumptions than those pertaining to the social equilibrium.

First, there is the case described by the standard model where individuals are motivated solely material payoff (so that  $\beta > 0$ ,  $\sigma = 0$ ). In this case efforts are given by  $x^{STD}$  as defined in (2) (*STD* means "standard" and suggests that social status concerns are suppressed) The relevant moral hazard measure in this case is  $x^{AUT}(t) - x^{STD}(t) = \frac{t}{K}u(\beta)$ .

We now compare the moral hazard measures in the social equilibrium and the standard model. A special case, when  $\sigma$  is not too large, is illustrated in Figures 4(a)–(b). In the social equilibrium moral hazard  $x^{AUT}(t) - x^{SE}(t)$  is *negative* at the lower end of the distribution of talents. Moral hazard then increases with talent up to some point, where it tapers off a bit. This is in sharp contrast to the standard prediction where the measure of moral hazard  $x^{AUT}(t) - x^{STD}(t)$  is never negative, and zero only for individuals with zero talent. In the social equilibrium, if  $\sigma$  is not too high, there will be positive moral hazard for some individuals with high enough talent, but never will moral hazard be as high as in the standard model.

Second, there is the case where social status is important but information about talent is common knowledge among individuals (there is *complete information (CI)* about each individual's  $t$ ). Then individuals choose effort so as to maximize (4) with a  $t$  substituted for  $\tau(x)$  and effort is given by  $x^{CI}(t) = \frac{1}{K}((1 - u(\beta))t + \sigma)$ . In this case the moral hazard measure becomes  $x^{AUT}(t) - x^{CI}(t) = \frac{t}{K}u(\beta) - \frac{\sigma}{K}$ .

We now compare the social equilibrium to the model with complete information, and again Figures 4(a)–(b) illustrate a special case when  $\sigma$  is not too large. Moral hazard is never lower in the social equilibrium than in the case with complete information. The difference in moral hazard between the two cases decreases with talent, and for the most talented individual there is no difference at all.

Summing up, there is unambiguously less moral hazard in the social equilibrium than in the standard case. This finding was not a priori obvious, because although effort *per se* tends to increase individuals' status, there was the counter-effect that individuals might try to fake their level of talent by shading effort choices. Such an effect indeed is important, as witnessed by the fact that there is more moral hazard in the social equilibrium than in the case with complete information. Still, social norms may to some extent alleviate free-riding.<sup>10</sup> However, one should be cautious in terms of judging the welfare implications of this finding. In the social equilibrium individuals of low talent work hard even though the positive externality they create is

---

<sup>10</sup>The conclusion that "peer monitoring" may mitigate moral hazard is obtained also by Arnott & Stiglitz [1] though the result in their case has a different cause (e.g., non-market insurance within families).

small.<sup>11</sup> Especially for populations where  $f$  has a fat bottom tail, this may seem socially undesirable.

## 4.2 Marginal policy changes

We now consider the effect of a marginal change in the benefit level  $\beta$  holding constant the social preference parameter  $\sigma$ . The conclusion of the standard model, as well as the model with complete information, is that effort is reduced when the benefit level is increased, except for individuals of talent zero whose effort is unaffected by the social insurance benefit. By contrast, in the social equilibrium there will be a "cutoff level of talent" in  $(0, 1)$  such that all individuals with talents below this level will *increase* effort, and all individuals with talents above will decrease effort when the benefit level is increased. This is illustrated in Figure 5.

With an eye to (10) we (somewhat loosely speaking) explain this as follows. The right hand side of (10) is the sum of two terms, each representing a different "effect". The first term represents the direct incentive to scale down effort as a reaction to reduced material incentives. The strength of this effect increases with talent, and hence by itself tends to make  $x^{SE}$  more flat. The second effect mirrors the altered incentives to adjust effort so as to reach a desired level of social status. Since the first effect tends to make  $x^{SE}$  more flat, a marginal increase of effort will tend to bring about a less drastic increase in the relevant social norm. Therefore, the incentives to increase effort go *up*. By inspection of (10), one sees that the higher is an individual's talent the less important is the second effect. In fact, the second effect is irrelevant to the individual with talent equal to 1. Since the two effects work in opposite directions and apply differently to individuals of different talents the explanation follows.

## 4.3 Social status

In the social equilibrium (where inferences are correct) the social status of an individual with talent  $t$  is endogenously determined and equal to the difference between the actually chosen effort and the choice of effort which would maximize that individual's material payoff. That is, the individual's social status rank is measured by  $x^{SE}(t) - x^{STD}(t)$ . One might have expected

---

<sup>11</sup>Cf. Goffman [13] who argues that the stigmatized individual can "... attempt to correct his condition indirectly by devoting much private effort to the mastery of areas of activity ordinarily left to be closed on incidental and physical grounds to one with his shortcoming."

that social status in equilibrium would be the same for individuals of all talents; after all, the social norms are sensitized to individual talents. However, as illustrated by Figure 4(c), this is not the case. In the social equilibrium, social status *increases* with talent.

Again, this is a result that suggests that one must proceed cautiously if one wants to discuss welfare issues. In our model, social status is systematically lower for people with low talent than for people with high talent, and this may be important if the welfare of an economy is assessed. Having social norms to reduce moral hazard is not a free lunch.<sup>12</sup>

#### 4.4 Alternative approaches

There are several ways in which one might modify or augment the model in this paper and we now mention a few of these: We assume that  $\sigma > 0$  if and only if  $\beta > 0$ , without considering elaborate connections that may seem natural. For example, there may be some specific (presumably monotone) relationship between  $\beta$  and  $\sigma$ . Investigating such a connection could be particularly interesting if government is introduced as a player who *chooses*  $\beta$ . Moreover, there may be some inertia associated with the formation of  $\sigma$  or the social norm, so that if government suddenly changes  $\beta$ , then  $\sigma$  or the social norm adjust only slowly.

The social preference parameter  $\sigma$  is exogenously given and independent of the action profile chosen by the population in the model. As an alternative,  $\sigma$  could be a function of the effort choices in society.<sup>13</sup> Yet another possibility is that  $\sigma$ , as well as the nature of the social norm itself, could be determined by forces of natural selection.<sup>14</sup>

The concept of social equilibrium presumes that individual talent is private information, while effort choice are observable for other individuals. We have not explored other observability assumptions, although we note that there are alternatives: Individuals could also observe each other's average

---

<sup>12</sup>Cf. Besley & Coate [4] who model the take up of welfare where the associated stigma in equilibrium depends on the difference between the average type of all claimants of welfare and the average type of deserving (poor) claimants. With dichotomous choice their model implies statistical discrimination. In our model information about “deservability” is by contrast inferred from observed behavior.

<sup>13</sup>Cf. the models of Kandel and Lazear [16] where average effort matters and Lindbeck et al [17] or Bird [5] where the social payoff associated with some action is inversely related to the number of individuals choosing that action.

<sup>14</sup>Cf. Güth & Yaari [15] who introduce the “indirect evolutionary approach”, in which individuals behave as rational agents for given preferences, but where preferences selected by evolution. Fershtman & Weiss [12] apply such an approach to analyze the evolutionary stability of certain social preferences.

unemployment spells (i.e.,  $xt$ ) or whether an individual is unemployed or employed. What assumption is reasonable depends on the interpretation of effort. At least if effort is education we think our assumption is reasonable, but nevertheless it may be interesting to explore other possibilities.

Finally we mention that techniques related to those used in this paper can probably be applied to many different economic problems. To assume that a social norm depends on perceived individual ability seems to be very reasonable when one analyzes problems of team production, profit sharing, work cooperatives, etc.

We think that it would be interesting to extend the model in any of these directions. However, we leave such tasks for future research.

## 5 Conclusions

We analyze the relation between social norms and moral hazard in an unemployment insurance context. An individual's effort to get a job has positive externalities (by reducing public expenditure on unemployment benefits), and therefore "lazy" behavior is not socially approved. A social norm governs how much effort an individual perceived to have a certain talent should exert, and this norm is increasing in perceived talent. Individuals care about what others think about their observed behavior, and therefore the model contains social incentives in addition to standard material incentives. It is not a priori clear how the social incentives affect behavior. On the one hand, for a given level of perceived talent, an individual's status increases with his effort. On the other hand, perceived talent depends on effort and so individuals may reduce their effort choices in order to manipulate others' perception of their talent.

Using techniques from information economics we solve the model for a unique equilibrium which is fully separating. Despite the counteracting social incentives, in this equilibrium there is unambiguously less moral hazard than without the social norms. Moreover, marginal increases of the social insurance benefit lead all individuals with low enough talents to increase their effort. These conclusions stand in contrast to standard predictions. However, since people with low talents work hard even though the positive externality they create is small, and since social status is still systematically lower for them than for people with high talent, we conclude that one should be cautious when making welfare judgements based on our results.

## Appendix

*Proof of Lemma 1.* This is a standard result for signaling games in which the single-crossing property holds, but we still give a proof for completeness: Suppose  $t < t'$ . Define  $x = x^{SE}(t)$ ,  $x' = x^{SE}(t')$ ,  $s = \tau^{SE}(x)$  and  $s' = \tau^{SE}(x')$ . We want to show that  $x \leq x'$ . Suppose to the contrary that  $x' < x$ . By Definition (i) (incentive compatibility) it holds that

$$U(x, t, s) \geq U(x', t, s') \quad \text{and} \quad U(x', t', s') \geq U(x, t', s) \quad (11)$$

which combined gives

$$Q = \left[ U(x', t, s') - U(x, t, s) \right] - \left[ U(x', t', s') - U(x, t', s) \right] \leq 0. \quad (12)$$

However, using the analytical specification of  $U(x, t, s)$  in (4) we get

$$Q = (1 - u(\beta))(t - t')(x' - x). \quad (13)$$

Since  $t < t'$  and  $x' < x$  by assumption, it must be by (13) that  $Q > 0$ . This is a contradiction and therefore  $x \leq x'$ .  $\square$

*Proof of Lemma 2.* It follows from Lemma 1 and Definition (ii) that  $\tau^{SE}$  is (strictly) increasing over equilibrium efforts. Hence it remains to prove  $\tau^{SE}$  is monotone also out-of-equilibrium. Note two things (a) and (b): (a) If  $t < t'$  and  $x^{SE}(t) < x^{SE}(t')$  then the total payoff of individuals with talent  $t'$  must exceed the total payoff of individuals with talent  $t$  (otherwise talent  $t' \in T$  could improve by choosing  $x^{SE}(t)$ ). (b) Given  $\tau^{SE}$ , higher talents gain more (or lose less) than lower talents by increasing effort from one given level to another given level, and lower talents gain more (or lose less) than higher talents by decreasing effort from one given level to another given level. (To verify (b), study the sign of  $Q$ , in equations (12) and (13) above, for different combinations of  $t, t', x, x'$ .) Combining (a) and (b) and Definition (ii and iii) one sees that  $\tau^{SE}(x) = 0$  for any  $x < x^{SE}(0)$ , and that  $\tau^{SE}(x) = 1$  for any  $x < x^{SE}(1)$ . It now only remains to prove monotonicity of  $\tau^{SE}$  when  $x$  is an out-of-equilibrium effort such that  $x^{SE}(0) < x < x^{SE}(1)$ . Let  $\underline{t}$  be the supremum of the talents making equilibrium effort choices below  $x$ , let  $\bar{t}$  be the infimum of the talents making equilibrium effort choices above  $x$ . Clearly  $\underline{t} = \bar{t}$ , and by (b) and Definition (iii) we get  $\tau^{SE}(x) = \underline{t} = \bar{t}$ . Combining this observation with the previous ones, one sees that  $\tau^{SE}$  is monotonic.  $\square$

*Proof of Lemma 3.* Suppose Lemma 3 is not true. Then there exists an out-of-equilibrium effort  $x$  such that  $x^{SE}(0) < x < x^{SE}(1)$ . Let  $\underline{x}$  be the supremum of the highest equilibrium effort choice below  $x$ , let  $\bar{x}$  be the infimum of



the lowest equilibrium effort choice above  $x$ . By (b) in the proof of Lemma 2 and Definition (iii) one sees that  $\tau^{SE}(x)$  has a uniquely determined value. By continuity of total payoffs in talent, one infers that  $u(\underline{x}, \tau^{SE}(x), \tau^{SE}(\underline{x})) = u(x, \tau^{SE}(x), \tau^{SE}(x)) = u(\bar{x}, \tau^{SE}(x), \tau^{SE}(\bar{x})) = 0$ . These equalities cannot hold unless  $\tau^{SE}(\underline{x}) \neq \tau^{SE}(x) \neq \tau^{SE}(\bar{x})$ , since material payoffs are strictly concave in effort. And then, by Lemma 2,  $\tau^{SE}(\underline{x}) < \tau^{SE}(x) < \tau^{SE}(\bar{x})$ . If  $\bar{x}$  is an equilibrium effort, this is impossible; any individual of talent  $t \in T$  with  $x^{SE}(t) = x$  would gain by choosing  $(\bar{x} - \varepsilon) \in X$  for  $\varepsilon$  small enough. If  $\bar{x}$  is an out-of-equilibrium effort there must exist  $\varphi > 0$  such that  $(\bar{x}, \bar{x} + \varphi)$  is a set of equilibrium efforts for which no pooling occurs. This too is impossible; there would exist  $\gamma$  with  $0 < \gamma < \varphi$  such that  $\tau^{SE}(\bar{x} + \gamma) \in T$  would gain by choosing  $\bar{x}$ . Hence no  $\tau^{SE}$  exist which sustains  $(x^{SE}, \tau^{SE})$  as a social equilibrium, a contradiction.  $\square$

*Proof of Lemma 4.* Suppose Lemma 4 is not true. Then there exists  $t, t' \in T$  such that  $t \neq t'$  and  $x^{SE}(t) = x^{SE}(t')$ . By Lemma 1 we can find a connected set  $\Theta \subseteq T$  of talents pooling at  $x^{SE}(t)$ . Let  $\underline{t}, \bar{t}$  be the infimum and supremum of  $\Theta$ . By Definition (ii),  $\underline{t} < \tau^{SE}(x^{SE}(t)) < \bar{t}$ . It is impossible that  $x^{SE}(t) > 0$  since then we can find  $\varepsilon > 0$  small enough that individuals with talent  $(\underline{t} + \varepsilon) \in \Theta$  would gain by choosing  $x^{SE}(t) - \varepsilon$  (the loss of material payoff is arbitrarily small and outweighed by a substantial social payoff gain). Hence it must be that  $x^{SE}(t) = 0$ . However, this is impossible too. To see this note first that  $\Theta = [0, 1]$  is impossible;  $1 \in T$  would deviate and choose effort level  $\frac{(1-u(b))+\sigma}{K}$ , where he has positive material payoff (instead of zero) and zero social payoff (instead of negative). Hence, if pooling occurs at effort 0, also other choices are made in equilibrium. By Lemma 3 these choices are all connected. But then we can find an equilibrium effort  $\varepsilon \in X$ ,  $\varepsilon > 0$  but small enough that no pooling occurs at any effort level in the set  $(0, \varepsilon]$ . An individual with talent  $t^{SE}(\varepsilon)$  chooses  $\varepsilon$ , but he in fact can gain by deviating to effort 0 (again the loss of material payoff is arbitrarily small and outweighed by a substantial social payoff gain). Hence there can be no pool, a contradiction.  $\square$

## References

- [1] Richard Arnott and Joseph E. Stiglitz. Moral hazard and nonmarket institutions: Dysfunctional crowding out or peer monitoring? *American Economic Review*, 81:179–190, 1991.
- [2] Kenneth J. Arrow. Uncertainty and the welfare economics of medical care. *American Economic Review*, 53:941–971, 1963.

- [3] B. Douglas Bernheim. A theory of conformity. *Journal of Political Economy*, 102:841–877, 1994.
- [4] Timothy Besley and Stephen Coate. Understanding welfare stigma: Taxpayer resentment and statistical discrimination. *Journal of Public Economics*, 48:165–183, 1992.
- [5] Edward J Bird. Culture and social policy: Does the welfare state need new tools? Unpublished paper, 1995.
- [6] In-Koo Cho and David M. Kreps. Signaling games and stable equilibria. *Quarterly Journal of Economics*, 52:179–221, 1987.
- [7] James S. Coleman. *Foundations of Social Theory*. The Belknap Press of Harvard University Press, Cambridge, Massachusetts and London, England, 1990.
- [8] Prajit Dutta and Roy Radner. Moral hazard. In R. J. Aumann and S. Hart, editors, *Handbook of Game Theory*, volume 2, pages 869–903, Amsterdam, 1994. Elsevier Science B.V.
- [9] Isaac Ehrlich and Gary Becker. Market insurance, self-insurance, and self-protection. *Journal of Political Economy*, 80:623–648, 1972.
- [10] Chaim Fershtman, Kevin M. Murphy, and Yoram Weiss. Social status, education, and growth. *Journal of Political Economy*, 104:108–132, 1996.
- [11] Chaim Fershtman and Yoram Weiss. Social status, culture and economic performance. *Economic Journal*, 103:946–959, 1993.
- [12] Chaim Fershtman and Yoram Weiss. Social rewards, externalities and stable preferences. Unpublished manuscript, 1995.
- [13] Erving Goffman. *Stigma: Notes on the Management of Spoiled Identity*. Prentice-Hall, Englewood Cliffs, New Jersey, 1963.
- [14] Nils Gottfries and Barry McCormick. Discrimination and open unemployment in a segmented labour market. *European Economic Review*, 39:1–15, 1995.
- [15] Werner Güth and Menahem E. Yaari. Explaining reciprocal behavior in simple strategic games: An evolutionary approach. In U. Witt, editor, *Explaining Process and Change — Approaches to Evolutionary Economics*, pages 23–34. Ann Arbor, Michigan, 1992.

- [16] Eugene Kandel and Edward P. Lazear. Peer pressure and partnership. *Journal of Political Economy*, 100:801–817, 1992.
- [17] Assar Lindbeck, Sten Nyberg, and Jörgen W. Weibull. Social norms and economic incentives in the welfare state. The Research Institute of Industrial Economics, Working Paper 476, 1997.
- [18] Barry McCormick. A theory of signalling during job search, employment efficiency, and “stigmatised” jobs. *Review of Economic Studies*, 57:299–313, 1990.
- [19] Mark Pauly. Overinsurance and public provision of insurance: The roles of moral hazard and adverse selection. *Quarterly Journal of Economics*, 88:44–62, 1974.
- [20] Garey Ramey. D1 signaling equilibria with multiple signals and a continuum of types. *Journal of Economic Theory*, 69:508–531, 1996.
- [21] Steven Shavell. On moral hazard and insurance. *Quarterly Journal of Economics*, 93:541–562, 1979.

Figure 1: The decision problem facing an individual with talent  $t \in [0, 1]$

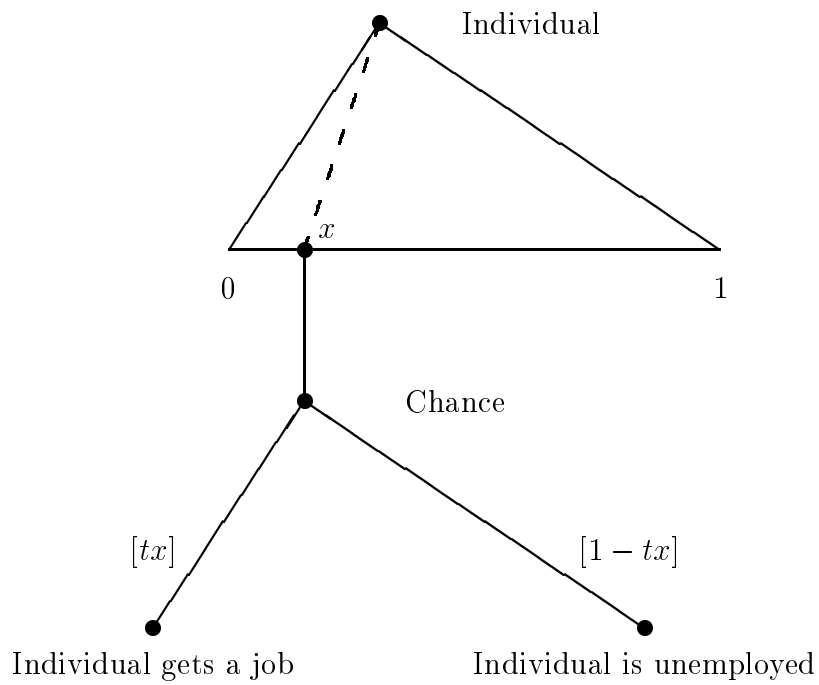


Figure 2: The social equilibrium.

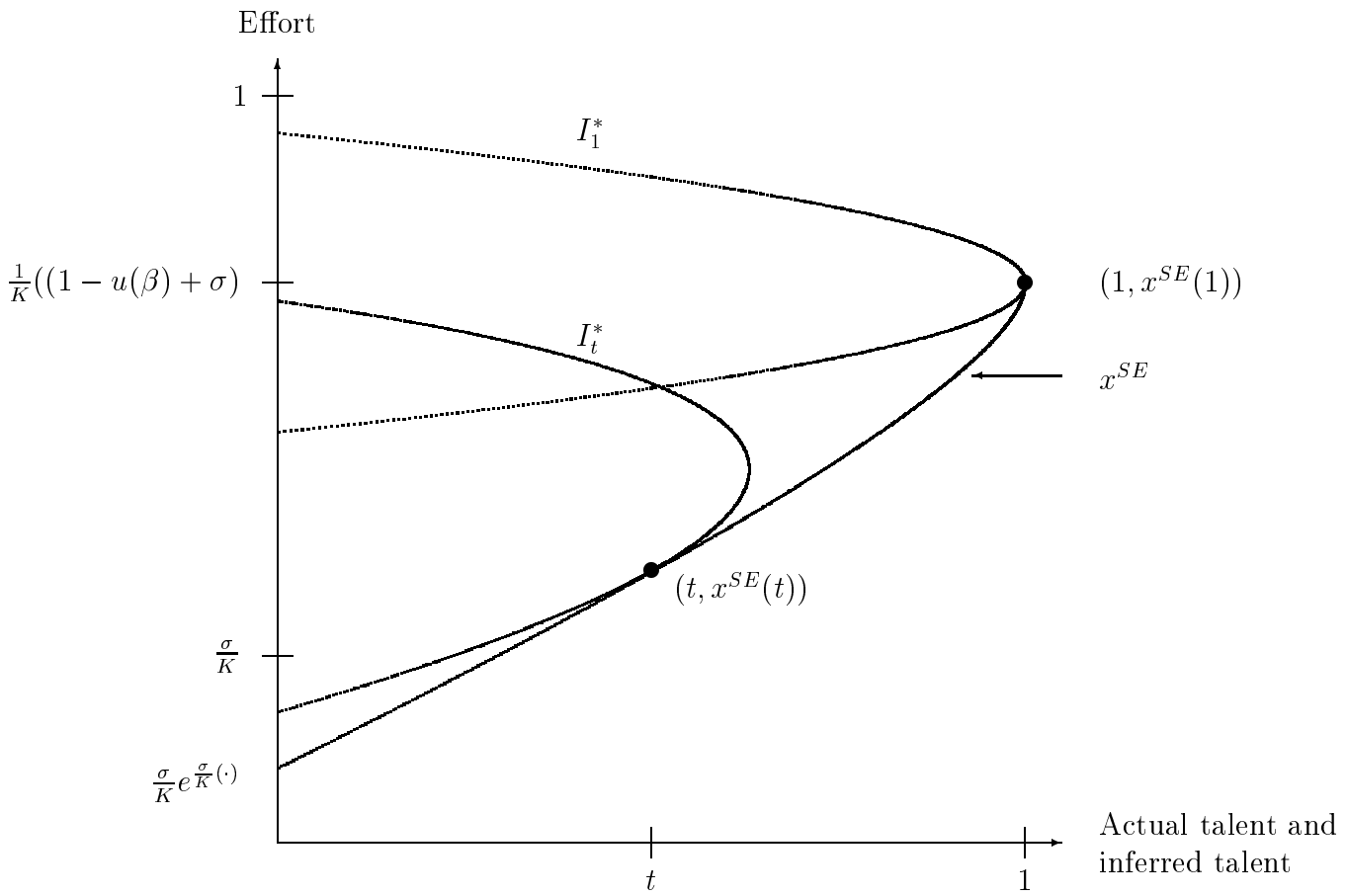


Figure 3: The inference function  $\tau^{SE}$ .

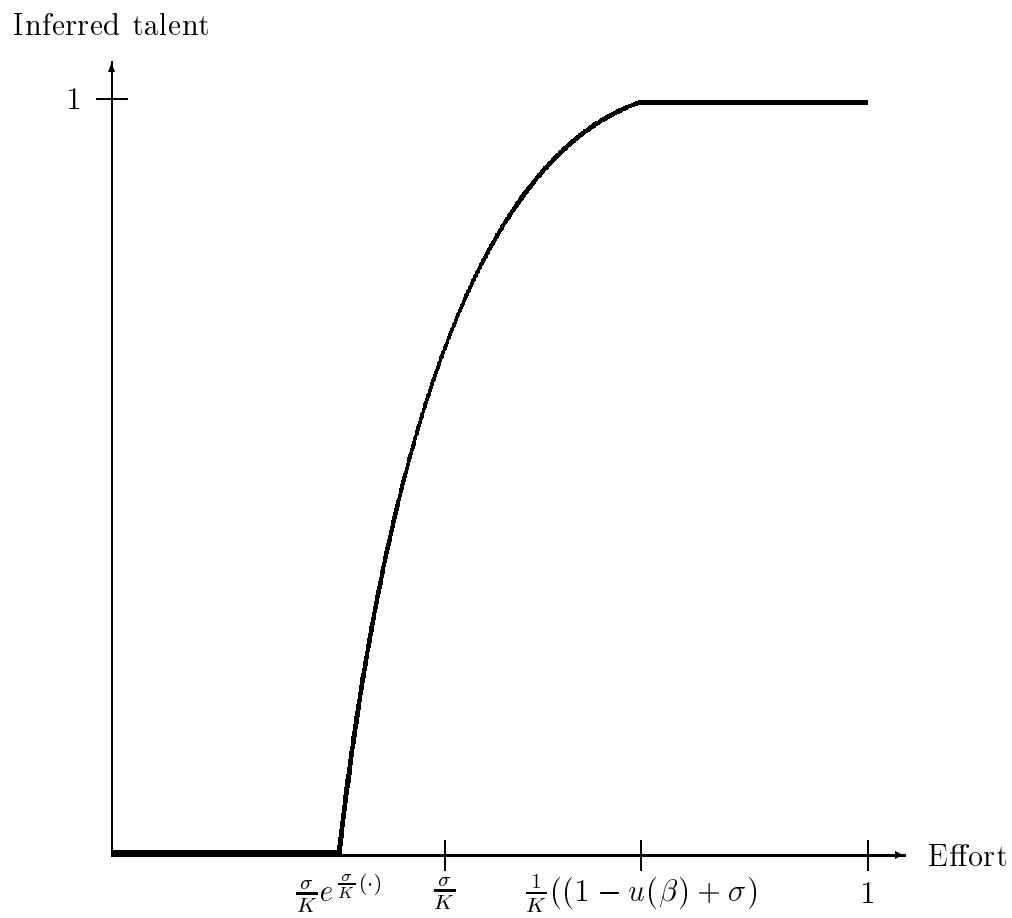


Figure 4: Effort choices (a), Moral hazard (b), Social status (c).

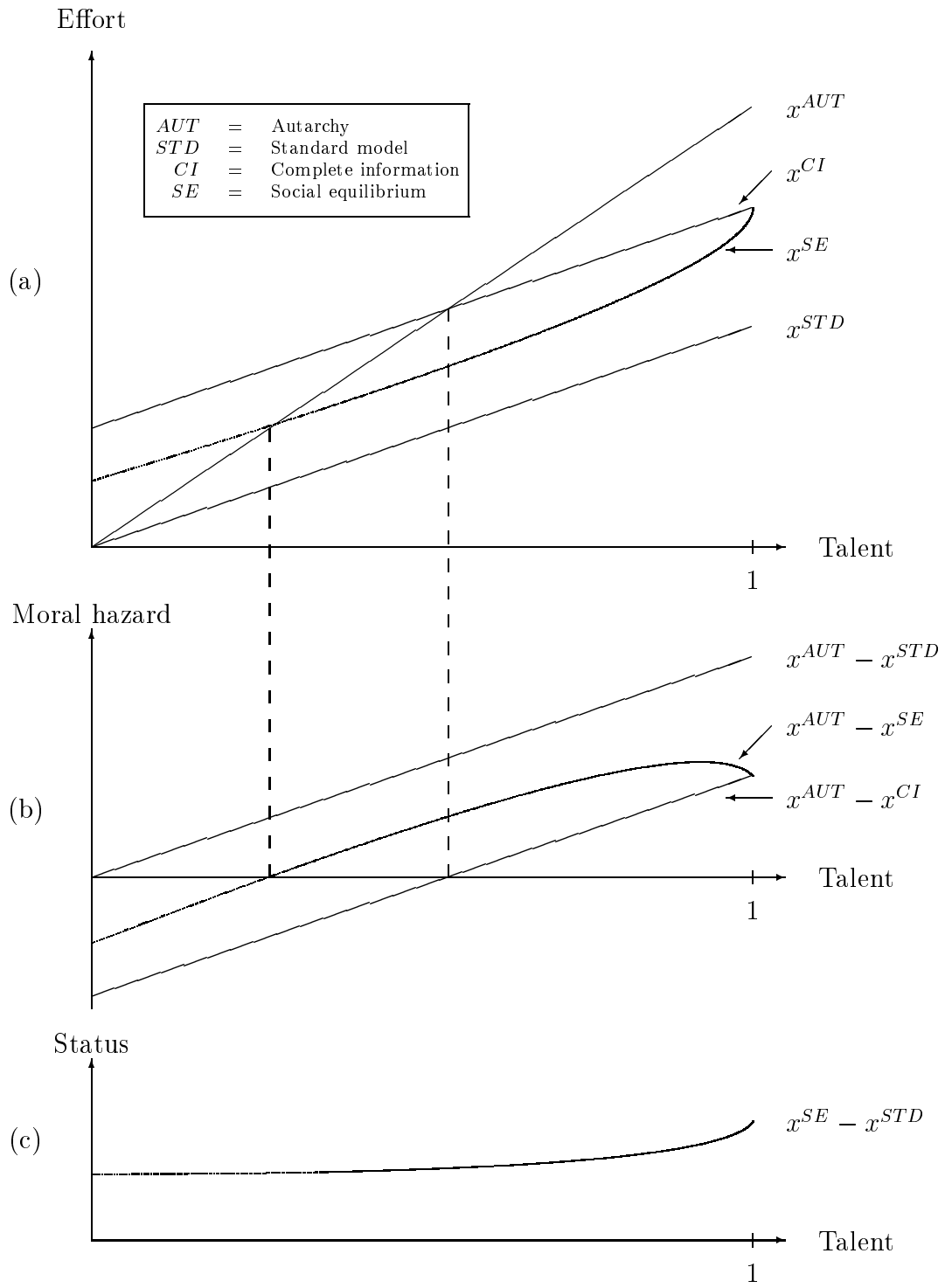


Figure 5: Effect of policy changes in the social equilibrium;  $\beta < \beta'$

