

SOCIAL NORMS AND MORAL HAZARD*

Martin Dufwenberg and Michael Lundholm

We examine the impact of social rewards in an unemployment insurance context. A social norm requires effort in proportion to perceived talent, but individuals cunningly choose effort so as to manipulate the perception of their talent. The model predicts that low talented individuals increase effort in response to more generous unemployment insurance. The welfare consequences of the social rewards are ambiguous. Social rewards boost effort, but for individuals with low talent more than any real economic concern can justify. Moreover, the distribution of social respect is slanted in favour of the more talented.

Some people are concerned that a fundamental pillar of the welfare state is crumbling. This pillar is made up of individual morality and social putty, and the idea is that such values have become less important in recent times. It is suggested that this has led to an insidious process in which generous social insurance systems have become too expensive to operate, given the attenuation of the forces that would curb individual temptations to exploit the system excessively. At least in Sweden, concerns of this nature are very much alive in public debate, and related discussions have surfaced in government sponsored investigations like Bröms *et al.* (1994) (see for example p. 82) as well as in scholarly journals (for example Lindbeck (1995)).

Is this important? To answer this question one must get a grip on how various aspects of economic behaviour and social rewards connect in a social insurance context. However, formal economic theory has had little to say. There are many economic models on social insurance where people care about material rewards, but these models typically do not incorporate social rewards. In this paper we take a step towards filling this gap.

Our approach is inspired by some ideas from sociology. Sociologists widely acknowledge that some individual decisions may be met with social sanctions that typically are taken to have immaterial content. People bestow various degrees of moral opprobrium or respect to one another, and they care about these social values. Coleman (1990, p. 274) argues that social rewards are especially important when individual decisions carry externalities:

‘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.’

* We are very grateful to two referees for their insightful and detailed suggestions. We have also benefited a lot from comments by Timothy Besley, Gary Charness, David de Meza, Ernst Fehr, Bertil Holmlund, Kari Lantto, Åsa Rosén, Jörgen Weibull, and several seminar participants. 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’).

If indeed there is a link between social rewards and how hard individuals try to reduce externalities, then this may be important in the social insurance setting we wish to study. Insuree behaviour often influences the likelihood of various benefits being paid, and so carries externalities.

In this paper we formally connect social rewards to individuals' attempts to reduce externalities, and others' perception of this behaviour. While the analysis has potential bearing on many kinds of insurance relationships, we phrase the presentation in terms of unemployment insurance: A population of individuals exerts costly 'effort' which influences the probability with which they become employed. Effort can be interpreted as (for example) education or search activity. Each individual's choice of effort is observed by some neighbours. The individuals are risk averse and there is scope for welfare enhancing unemployment insurance. Unemployment benefit is paid to any individual who is unemployed, independently of the individual's choice of effort. The interpretation is that effort choices are hidden to the administrator of the insurance system, or that the use of such information has been deemed an illegitimate basis for government action. Effort has externalities, since it influences employment probability and hence the expected unemployment benefit payment. Therefore lazy behaviour (by an individual) is not socially approved (by his neighbours).

We admit heterogeneity of the population with respect to a particular characteristic called 'talent'. An individual's talent measures how efficient he is in transforming effort into an enhanced probability of getting a job. Effort pays off less to individuals with low talent. It is therefore reasonable to assume that the more talented an individual is, the more effort he must exert in order not to be considered lazy. We assume that the social respect bestowed upon an individual by his neighbours depends on their assessment of the difference between his actual choice of effort and the effort he would choose if he maximised his own material well-being. We refer to the latter level of effort as a 'social norm' for this individual. This feature of the model raises an important issue. Who knows an individual's level of talent? We focus primarily on the case where each individual's level of talent is private information. 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.

To solve the model we apply tools of information economics. In equilibrium everyone optimises given the pattern of inference formation which is the basis for social evaluations. Essentially, for any choice made in equilibrium the inference is required to be correct, while for any other choice the inference must select an individual for whom that choice makes most sense given the inference. We prove that the model always has a unique equilibrium with these properties.

We use the model to shed light on several issues. The first is about the impact of moral hazard. In our model individual choices are observed by some neighbours, but the model entails moral hazard in the sense that the payment of unemployment benefits is not conditioned on any individual's choice of

effort.¹ By the impact of moral hazard, we refer to the difference in effort before and after a social insurance system is introduced. The answer to the question of how social rewards influence the impact of moral hazard is not *a priori* obvious. Although respect 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. Nevertheless, we show that in equilibrium individuals of all talents exert more effort than they would if social respect were not important.

Second, we analyse how behaviour changes if the unemployment benefit is changed. The answer is perhaps somewhat counter-intuitive: increases in the benefit level make individuals with low talent work *harder*, while more talented individuals work less hard.

Third, we consider how the introduction of social rewards affects welfare. We conclude that this is not clear. On the one hand, having social rewards mitigates the impact of moral hazard, and this may appear to enhance welfare. On the other hand, people with low talent exert lots of costly effort even though the enhanced probability of becoming employed as well as the positive externality created are small. This seems wasteful. Moreover, the distribution of social respect in society is increasing in talent. This is not obviously welfare improving. Social norms do not provide a free lunch, with respect to alleviating opportunistic behaviour.

The model is introduced in Section 1 and solved for equilibrium behaviour under various assumptions in Section 2. In Section 3 we discuss the results, and comment on some alternative modelling choices. Section 4 presents concluding remarks. An Appendix contains the proof of the equilibrium existence theorem.

1. The Model

We next describe the model, discussing in turn its strategic structure (1.1), material payoffs (1.2), social payoffs (1.3), and total payoffs (1.4).

1.1. *Effort, Talent, Unemployment*

By exerting *effort* $x \in X = [0, 1]$ individuals affect their probability of becoming employed. This effort may have many interpretations, e.g., time or attention devoted to search activities on the labour market or to acquiring an education. Individuals differ according to their *talent* $t \in T = [0, 1]$. Depending on context, this trait may be interpreted as ability in production, proximity to the labour market, or intelligence. 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

¹ Moral hazard considerations have received ample attention in the insurance economics literature. See, e.g., Arrow (1963), Ehrlich and Becker (1972), Pauly (1974) and Shavell (1979). A general overview on moral hazard is Dutta and Radner (1994). This literature does not, however, consider how social rewards may constrain behaviour.

function f . F and f are common knowledge among all individuals. An individual's probability of becoming employed equals tx , and hence depends on both his effort and his talent.

Fig. 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 Fig. 1.

1.2. Material Payoffs

An individual with a job has positive income from labour. Income is normalised to one, since we describe a situation with only one type of job,² and talent does not influence income. One referee suggested that this is one feature which makes our model North European rather than American.

There is no private unemployment insurance but we assume that there is a public unemployment insurance system financed through public funds. A uniform unemployment benefit β is received by an unemployed individual

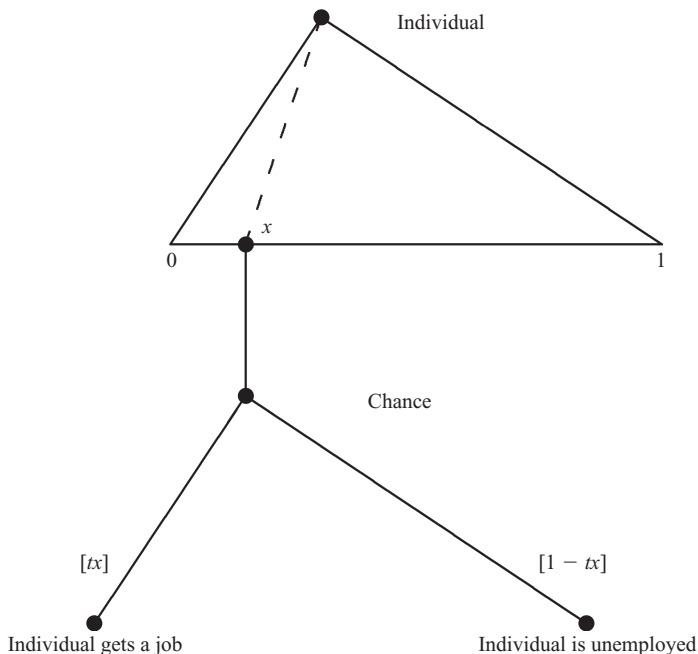


Fig. 1. *The Decision Problem Facing an Individual with Talent $t \in [0, 1]$*

² This is in contrast to some other models with social concerns, like Fershtman *et al.* (1996), Fershtman and Weiss (1993), Gottfries and McCormick (1995), McCormick (1990).

regardless of his talent and chosen effort. There may be several reasons for not making the benefit depend on effort:

- Effort is unobservable. This is quite natural if effort for example corresponds to the number of firms to which an individual applies for a job, or to studiousness in doing homework.
- Discrimination according to effort choices may be deemed an illegitimate basis for government action, or it may be optimal and feasible for a government to commit not to use such information. See Brito *et al.* (1991) for a discussion about related issues. As a possible illustration, assume that effort corresponds to education. In Sweden, information about individuals' school background is public information. Yet, unemployment benefits are not conditioned on individuals' educational degrees.

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

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

where the strictly increasing income valuation function $u: \mathbb{R}_+ \rightarrow \mathbb{R}$ is normalised so that $u(0) = 0$, $u(1) = 1$. If u is sufficiently concave a utilitarian policy maker (for example) would choose $\beta \in (0, 1)$, and we will concentrate on this case. K is a positive constant large enough to guarantee that $x = 1$ will never be chosen in equilibrium. The interpretation is that no individual finds it worthwhile spending all his time or effort reducing the risk of being unemployed. (An explicit restriction on K is specified in equation (5) in section 2.4). Hence, an individual of talent t motivated only by material payoff would choose effort equal to

$$\frac{1}{K}[1 - u(\beta)]t. \quad (2)$$

1.3. *Social Norms, Social Respect, Social Payoffs*

Individuals care about the social respect that they bestow on one another. One important reason why social respect matters may be the presence of externalities in decision making. Coleman (1990, p. 275) claims that for social rewards to be effective '[t]he existence of externalities is a necessary condition.'³ In our framework, this means that an individual's social respect depends on his effort. With $\beta > 0$, an individual's effort has positive external effects, since effort reduces the probability of becoming unemployed and social insurance benefits are paid through public funds. Therefore, with $\beta > 0$, others may be

³ This is consistent with a view expressed by Arrow (1970, p. 20): 'I suggest as one possible interpretation that [norms of social behaviour, including ethical and moral codes] are reactions of society to compensate for market failures.'

inclined not to respect a lazy individual and to hold a diligent individual in high esteem.⁴

Where can the border between laziness and diligence be drawn? We assume that this is judged with reference to how much effort an individual would exert if he did not care about social respect at all. An individual perceived to have talent $\tau \in T$ is awarded zero social respect if he makes an effort choice of $\frac{1}{K}[1 - u(\beta)]\tau$ —the effort choice that would maximise his material payoff if this talent were τ (cf. (2)). It is assumed that an individual's choice of effort is observed by some neighbours, so the inference τ should be thought of as the inference of these neighbours. We refer to $\frac{1}{K}[1 - u(\beta)]\tau$ as the *social norm* for an individual perceived to have talent τ . Note how crucially this norm depends on the inference τ . If the individual makes some other choice then he is awarded respect in proportion to the difference between his actual choice of effort and the effort prescribed by the norm.⁵ Hence, using (2), an individual who chooses effort x and is perceived to have talent $\tau \in T$ is accorded a *social respect* of $x - \frac{1}{K}[1 - u(\beta)]\tau$. Note that with this specification the social respect of an individual measures how hard he is perceived to exert effort relative to the norm, not the actual externality associated with his behaviour. Finally, he gets a *social payoff* of

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

where $\sigma > 0$ is a number (common to all individuals) which measures sensitivity to social respect.

The inference τ made about the talent of an individual who makes a particular choice is determined endogenously in equilibrium. The details of this are explained in Section 2.

1.4. 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 τ has total payoffs as given by the utility function $U: X \times T^2 \rightarrow \mathbb{R}$ defined by

⁴ We follow the suggestion of a referee and talk about social *respect* rather than social *status* (which we used in earlier versions of this paper). This terminology seems most apt, given the following view expressed by the referee: 'I respect Mexican-Americans for being generally hard-working, honest, and committed to their families, even though their social status is low. And I hold in some contempt 'poor rich kids' who squander their wealth and talent, even though they have high social status. The point is semantic. At least in North American usage, social status tends to be almost an objective concept – like caste, though not as extreme. Social respect is more subjective. Senior politicians have considerable social status, but often command little respect.'

⁵ A similar approach to modelling social respect is used by Kandel and Lazear (1992). We invite reflection on our use of the word 'norm' by citing one of our referees: 'A norm is a social moral expectation, a definition of which acts people in society will judge as right or wrong. In the paper, each person is expected to exert so-and-so much effort or be stigmatised. This fits. Too many authors use "norm" just to mean "conformity in behavior".'

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

We assume that

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

which, as explained above, guarantees that $x = 1$ is never chosen in equilibrium.

2. Equilibrium

We next solve the model for equilibrium behaviour. The critical point is that information about individual talent is private. To determine an individual's social respect conclusions about his level of talent must be *inferred* from the chosen actions in society.⁶ We assume that the inferences can be represented by a function $\tau: X \rightarrow T$ such that $\tau(x) \in T$ is the perceived talent of an individual who chooses effort $x \in X$.

We require that individuals of all talents maximise their total payoffs given the equilibrium inference function. Furthermore, the inference function must be consistent 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.

Second, we impose a restriction on inferences concerning effort choices that do not occur in equilibrium. Restrictions of this sort are standard in the literature on signalling games (see, e.g., Cho and Kreps (1987); cf. Bernheim (1994)), because otherwise it is hard to get clearcut results and a host of equilibria with unintuitive interpretations may result. In the current context, we could (for many values of $x > 0$) sustain a 'strange' equilibrium such as the following: each individual chooses the same effort $x > 0$. The inference function τ satisfies that $\tau(x)$ equals the expected talent of an individual drawn at random from the pool, while $\tau(x') = 1$ for any $x' \neq x$. With σ high enough, this pattern of inferences sustains pooling at x as an equilibrium. The inferences concerning choices that do not occur in the equilibrium conform with Bayes' rule, but are arguably nevertheless unintuitive or unreasonable. For any $x' < x$, the lower is an individual's level of talent the stronger is his incentive to deviate. The inference $\tau(x') = 1$ effectively means that it is the individual with the *weakest* incentive to deviate who is deemed to have done so. By contrast, we shall demand that the inference concerning any choice that

⁶ This aspect resembles the conformity model in Bernheim (1994), except that his counterpart to our social norm is constant over types. The models share the feature that individuals care about others' inferences. The resulting signalling games are non-standard; usually the inferring parties take *actions* about which others are concerned. However, as noted by Bernheim (1994, footnote 6), this has no formal significance and the analysis could be recast in more traditional ways.

does not occur in the equilibrium must be on the talent level of the individual with the *strongest* incentive to deviate under that particular inference.⁷

To introduce the equilibrium concept that we shall use, let $x^*: T \rightarrow X$ be a function which describes which effort levels are chosen in equilibrium. We refer to the image of x^* as ‘equilibrium efforts’ and to complementary effort levels as ‘out-of-equilibrium efforts’. Let $\tau^*: X \rightarrow T$ describe inferences in equilibrium.

DEFINITION. *An equilibrium is a pair of functions (x^*, τ^*) such that*

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

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

Condition (i) in the Definition requires optimality of chosen actions while conditions (ii) and (iii) correspond to the requirements on the inference function discussed above. We now solve the model under the assumption that $0 < \beta < 1, \sigma > 0$:

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

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

We present the proof in the appendix, and concentrate here on interpreting and giving intuition for the result.

First, we illustrate the equilibrium in Fig. 2 and Fig. 3. The graph x^* in Fig. 2 depicts equilibrium effort choices for each level of (actual) talent, but may inversely also be read to depict the inferred talent for any equilibrium choice of effort (cf. Fig. 3). The graphs I_t^* and I_1^* depict indifference curves (inferred talent/effort combinations) of the individuals with talent $t < 1$ and 1, respectively, to which the graph x^* is tangent at the points $(t, x^*(t))$ and $(1, x^*(1)) = (1, \frac{1}{K}[1 - u(\beta) + \sigma])$. It is clear that x^* depicts optimum choices for any $t \in T$: An individual with talent t increases his payoff only if he could choose a level of effort and achieve an inference about his talent corresponding to a point contained in the area circumscribed by the vertical axis and I_t^* . However, this is not feasible given the pattern of equilibrium inferences.

To underscore the intuition for why x^* depicts equilibrium choices, it is

⁷ We opt for this inference restriction because we find it natural, and perhaps more natural than some other restrictions that have been proposed (like Cho and Kreps’ D1 criterion). We think it would be worthwhile to explore the properties of our restriction in general signalling games, but that task falls outside the scope of the present paper.

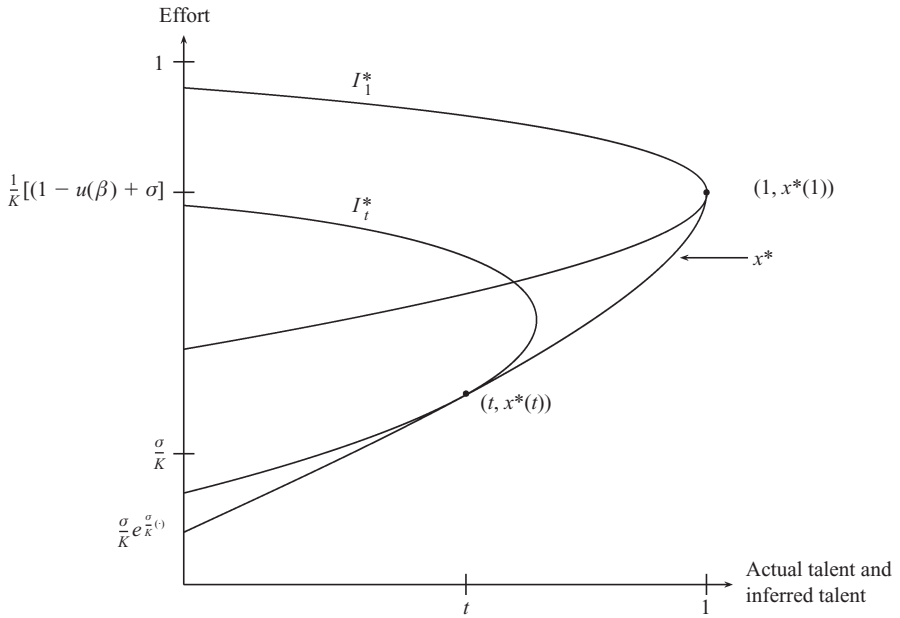


Fig. 2. *The Equilibrium*

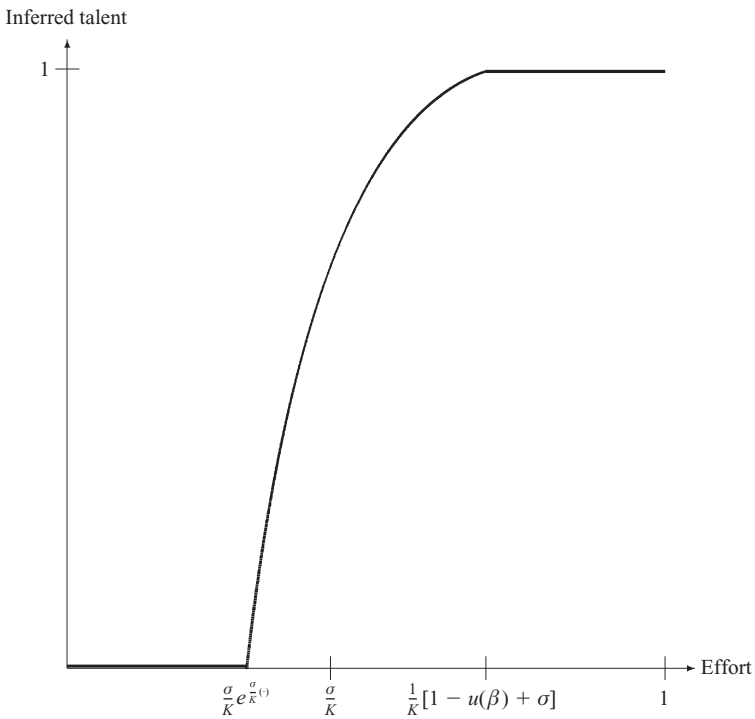


Fig. 3. *Equilibrium Inferences*

helpful to explain why in equilibrium individuals do not simply behave as if their level of talent were known. To see this, consider for the moment the effort choice that an individual with talent t would make if he took the view that the inference being made about his talent was t independently of his choice. He would choose x so as to maximise (4) with $\tau = t$. If everyone behaved in this manner, then the choice for different individuals could be described by a function x^c defined by $x^c(t) = 1/K\{[1 - u(\beta)]t + \sigma\}$. To facilitate comparison, the function x^c is plotted in Fig. 4(a). The super index c stands for ‘complete information’, suggesting that there is complete information about each individual’s level of talent. However, x^c cannot also describe equilibrium choices when talent is private information, because then (for any $t > 0$) reduced choices of effort would lead to reduced levels of inferred talent. This would create an additional incentive for reduced effort which was absent when x^c was derived. Given inferences consistent with x^c , with private information about individual talent, each individual with talent $t < 0$ would want to deviate from $x^c(t)$ to a lower level of effort.

The equilibrium choices described by x^* are just enough lower than choices described by x^c to make further ‘shading’ of effort, of the kind described in the previous paragraph, pointless. Note that in the equilibrium, the individual with talent 1 chooses as if his talent were known; i.e., $x^*(1) = x^c(1)$. The equilibrium is separating, so his level of talent is revealed and the highest conceivable social norm must apply to him. This could never be consistent with equilibrium behaviour unless he chose the effort level that would be optimal for him had his level of talent been known.

3. Discussion

In this section we discuss the impact of moral hazard, the effect of marginal policy changes, and welfare issues. We also discuss some possible alternative modelling approaches.

3.1. *The Impact of Moral Hazard*

The model we consider has moral hazard in the sense that the payment of unemployment benefits cannot be conditioned on any individual’s choice of effort. By the *impact of moral hazard* we refer to the difference in effort before and after a social insurance system, characterised by some $\beta > 0$, is introduced.

‘Before social insurance’ can be thought of as an autarkic situation before the introduction of a welfare state. Without social insurance, no unemployment benefits are paid from public funds, and in this sense individual behaviour carries no externalities. Since the presence of externalities is what we assume motivates the existence of social rewards, social respect does not matter if there is no social insurance. This regime may be viewed as a situation where $\beta = \sigma = 0$ and individuals choose x so as to maximise (4) with these values inserted; an individual with talent t maximises $tx - (\frac{K}{2})x^2$ with respect to x . The optimal effort choices for different individuals can be described by a

function x^a defined by $x^a(t) = \frac{t}{K}$, which is illustrated in Fig. 4(a). The super index a stands for ‘autarky’. We then define the impact of moral hazard regarding an individual of talent $t \in T$ as

$$x^a(t) - x^*(t), \tag{6}$$

which is illustrated in Fig. 4(b).

Before we discuss this measure it is natural to consider the impact of moral hazard had social rewards not been included in the model. In this baseline case, the individual’s effort can be described by a function x^b defined by

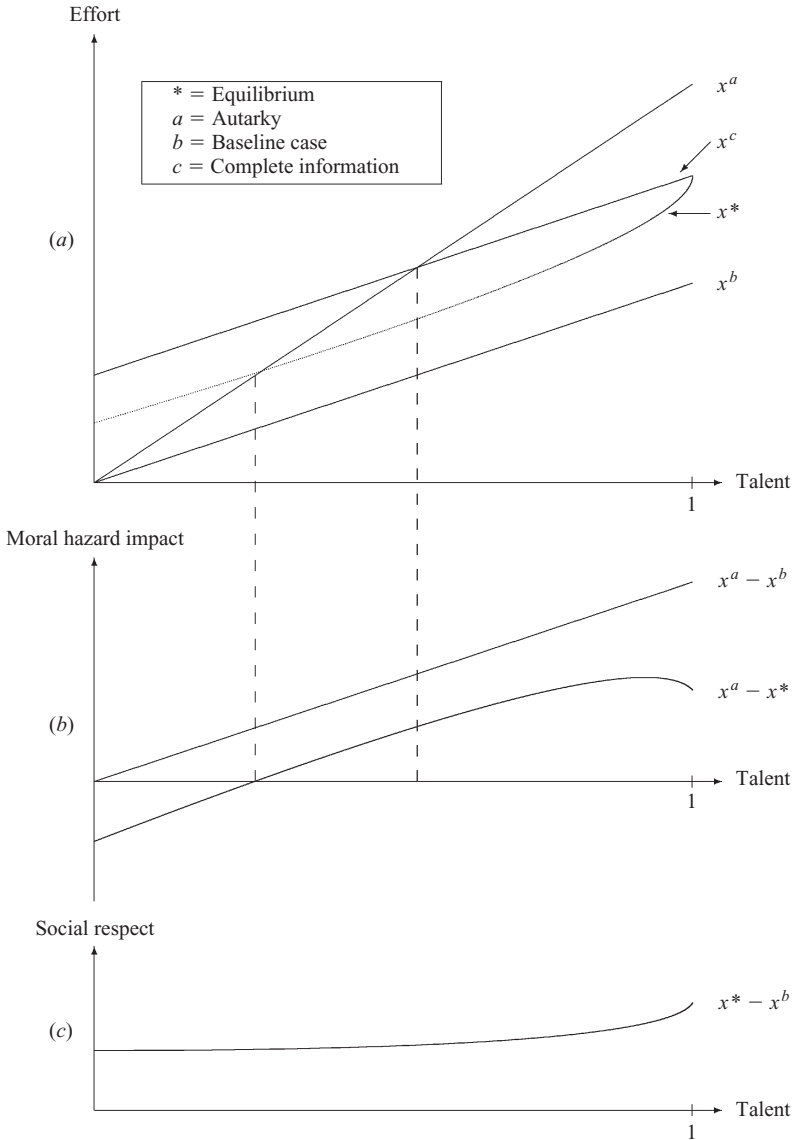


Fig. 4. *Effort choices (a), Moral hazard impact (b), Social respect (c)*

$x^b(t) = \frac{1}{K}[1 - u(\beta)]t$, where the right-hand side of the equality is given by (2). The super index b stands for 'baseline'. x^b is also illustrated in Fig. 4(a). The impact of moral hazard would be $x^a(t) - x^b(t) = \frac{t}{K}u(\beta)$; which is also illustrated in Fig. 4(b). The impact of moral hazard is positive (i.e., insurance always reduces effort) and increasing in talent.

The situation in the equilibrium we analyse is more complex. Effort is higher compared to a situation without social rewards and therefore the impact of moral hazard is lower. The impact of moral hazard is *negative* for low talents and positive only for high talents. The impact of moral hazard is first increasing in talent, reaches a peak, and finally decreases.

Summing up, the impact of moral hazard is unambiguously smaller with than without social rewards. This finding was not *a priori* obvious, because although effort *per se* tends to increase individuals' respect, there was the counter-effect that individuals might try to fake their level of talent by shading effort choices. Still, social norms may to some extent alleviate free-riding.⁸

3.2. Policy Changes

We now consider the effect of a change in the benefit level β holding constant the social preference parameter σ . First recall that in the absence of social rewards (with $\sigma = 0$) all individuals reduce their effort in proportion to their talent when the benefit level is increased, i.e., $\partial x^b(t)/\partial\beta = -u'(\beta)t/K < 0$. By contrast, in the equilibrium with $\sigma > 0$ one can show that there is a 'cutoff level of talent' somewhere in $(0, 1)$ such that all individuals with talents below this level *increase* effort when the benefit level is increased, while all individuals with higher talents decrease effort. This is illustrated in Fig. 5.

One can get the intuition for this result by verifying that individuals are influenced by two 'effects' that influence individuals of different talent differently. On the one hand, there is the (traditional) effect that a higher benefit makes unemployment appear less unattractive, which provides an incentive for lower effort. This effect is most important for individuals with high talent, who are most efficient in terms of translating effort into enhanced employment probability. The effect is irrelevant for an individual with talent zero.

The second effect is more subtle, and related to signalling aspects. To get a feel for it, recall that if the talent of an individual with talent $t < 1$ were known, then he would choose the effort $x^c(t)$. With private information about talent, however, he chooses $x^*(t) < x^c(t)$. The reason he does not choose a higher effort is the higher inference of his talent that would be made, and the more harsh social norm that would be applied to him. Now, consider what happens if β is raised. The function x^* become flatter relative to x^c , the economic

⁸ The conclusion that 'peer monitoring' may mitigate moral hazard is also obtained by Arnott and Stiglitz (1991) though their result has a different cause (e.g., non-market insurance within families). Our finding may furthermore be compared to the discussion in Goffman (1963). He argues that the stigmatised 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.'

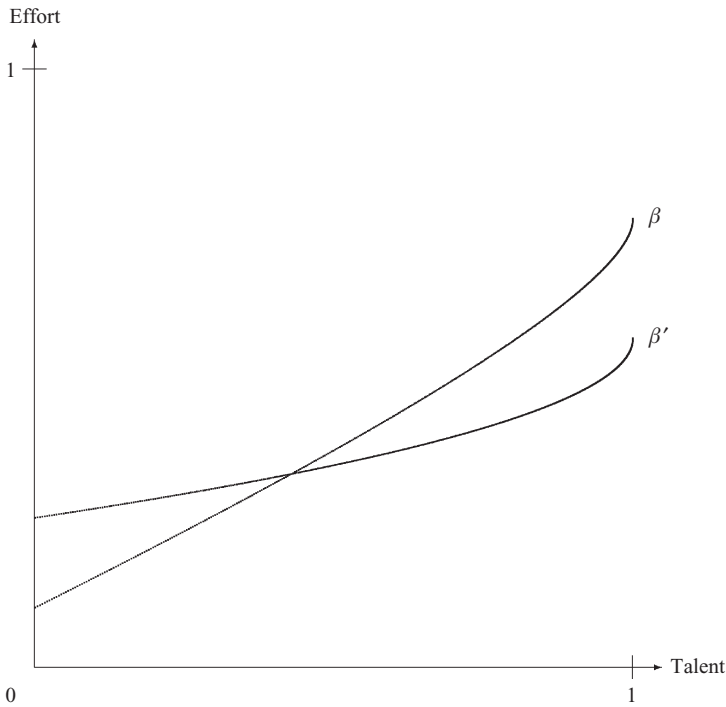


Fig. 5. Policy Change with Private Information; $\beta < \beta'$

interpretation being that mimicry of less talented individuals pays off less. The social norm for an individual perceived to have talent τ is $\frac{1}{K}[1 - u(\beta)]\tau$, so by manipulating τ an individual can manipulate the social norm that applies to him. However, the higher is β the less a change in τ influences the social norm. Since $x^*(1) = x^c(1)$, it follows that for all $t < 1$, $x^*(t)$ comes closer to $x^c(t)$. This is the second effect. Essentially, as β is raised, the individuals are less engaged in shading effort so as to manipulate the perception of their talent and they come closer to exerting the effort they would choose if their talent were known.

This second effect is most important for individuals with low levels of talent. For an individual with talent 1 it is completely unimportant; $x^*(1) = x^c(1)$ independently of β . However, for individuals with low talents the effect is strong enough to swamp the other (traditional) effect. This is most clearly visible in the case of the individual with a talent of 0. We know that the higher is β the less $x^*(0)$ drops below $x^c(0)$. However, $x^c(0)$ is *independent* of β , so this means that when β is raised so is $x^*(0)$. The individual can choose a higher effort without any fear that a harsher social norm will apply to him.⁹

⁹ David de Meza suggested an alternative way to think of this result: Consider the choice of the individual with $t = 0$ when β is increased, assuming that the effort of all others are *unchanged*. The social respect of those just above the least talented individual is now higher since their baseline effort is lower. This boosts the incentive of the least talented individual to increase effort. And this cascades up the scale to maintain separation.

The prediction that an increase in the unemployment insurance benefit will increase effort for the least talented should in principle be empirically testable. Would it be rejected? Some hints may be found in the fairly large empirical literature that attempts to measure the elasticity of unemployment with respect to unemployment benefits. As a general matter, the evidence seems to be mixed. Narendranathan *et al.* (1985, p. 307) report that ‘estimates may now be found anywhere from negative to four’, and they cite several studies. The bulk of this work is not directly concerned with the issue of talent, which is central to our model, but there are at least some weak pointers that reality does not clearly contradict our prediction. Narendranathan *et al.* find that ‘[b]enefits have no impact on the conditional probability of leaving unemployment for the long-term unemployed (over six months) except in the case of teenagers’ (p. 328). In fact, they report that the effect is positive, though insignificantly different from zero (see p. 322). If the long-term unemployed have low talent, this provides some support for our prediction.¹⁰

3.3. *Welfare*

In our model, the presence of social rewards mitigates the impact of moral hazard. At first glance, this results may seem to indicate that social rewards help welfare. However, there are at least two reasons why one should be cautious in drawing such a conclusion.

First, individuals with low talent exert lots of costly effort even though this does not significantly affect their chances of becoming employed, and even though the positive externality they create is small. Especially for populations where f has a fat bottom tail, so that a large share of the population has low talent, this seemingly wasteful behaviour may appear to be socially undesirable.

Second, social respect 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. Note also that effort increases with talent, so it is also true that more social respect attaches to those working hardest. The distribution of social respect is illustrated in Fig. 4(c). In equilibrium the social respect of an individual with talent t is equal to the difference between the effort actually chosen and the choice of effort which would maximise that individual’s material payoff. Hence, the individual’s social respect is measured by $x^*(t) - x^b(t)$. One might have expected that social respect would be the same for individuals of all talents; after all, the social norms are sensitive to individual levels of (perceived) talent. However, as illustrated by Fig. 4(c), this is not the case.

¹⁰ There is also some related work in sociology. Nordenmark (1999) investigates the psychosocial meaning of (un)employment. The finding that unemployment is best explained with reference to ‘non-financial employment motivation’ permeates all his work. He writes that his study ‘provides little support for the argument that it is primarily differences in the level of [financial] employment motivation that explain why some people find a paying job while others get stuck in unemployment’ (p. 13). This finding is at least not at odds with our result.

3.4. *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 $\sigma > 0$ only if $\beta > 0$, but do not consider more elaborate connections. However, given our assumption that social respect derives from perceived effort to reduce negative externalities, it may seem natural to postulate some monotone relationship between β and σ . The larger is β , the more important are externalities, so the stronger should be the sensitivity to social respect as measured by σ . Investigating such a connection could be particularly interesting if government is introduced as a player who *chooses* β , or if β is determined as an equilibrium in some political process. Moreover, there may be some inertia associated with the formation of σ , so that if β is suddenly changed σ adjusts only slowly.

The social preference parameter σ is moreover independent of the action profile chosen by the population in the model. As an alternative, σ could be a function of the effort choices in society.¹¹ Yet another possibility is that σ could be determined by forces of natural selection.¹²

In our model, the social norm that applies to an individual is endogenously determined in that it depends on the formation of equilibrium inferences. However, for a given level of perceived talent, the social norm itself is taken as given (see Section 2.3). One could consider alternative specifications where the norm is affected by environmental or policy changes, and ask if some particular form of social norm is an optimal response to the externality.¹³

We have assumed that individual effort choices are observable to other individuals but not explored other observability assumptions. We note that there are reasonable alternatives: individuals could for example observe whether an individual is unemployed or employed, or they could observe xt . The latter assumption seems natural if one adopts the interpretation that xt corresponds to an average unemployment spell. What assumption is most reasonable partly depends on the interpretation of effort. We think our assumption is reasonable (at least if effort is education) but it may nevertheless be interesting to explore other possibilities.

One could also imagine different observability assumptions concerning the government. One interesting possibility may be that government observes xt . This would be somewhat contrived given the interpretation we have given of our model (where xt is a probability), but for other interpretations (as in the previous paragraph) this may be more natural. In this case, the model would

¹¹ Cf. the models of Kandel and Lazear (1992) where the social norm depends on the average effort in society and Lindbeck *et al.* (1996) where the social payoff associated with some action is inversely related to the number of individuals choosing that action. Cf. also the model of Besley and Coate (1992) where stigma in equilibrium depends on the average type of all claimants of welfare and the average type of deserving (poor) claimants.

¹² Cf. Güth and Yaari (1992) who introduce the 'indirect evolutionary approach', in which individuals behave as rational agents for given preferences, but where preferences are selected by evolution. Fershtman and Weiss (1998) apply a related approach to analyse the evolutionary stability of certain social preferences.

¹³ Cf. the model of Bird (1999) where the social norm is determined by a 'median voter'. This specification enables Bird to show that in his model (which concerns lone parenthood) a norms system is eroded by increases in income.

have a structure that resembles the classical optimal income tax model (see Mirrlees (1971)), and presumably there may be a role for social rewards in that context. However, it seems these would have to be motivated differently than we have done here. It is not clear that the link effort-externality-social sanction, which is a key feature of our framework, can be immediately or naturally translated into an optimal taxation setting.

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 reasonable when one analyses problems of team production, profit sharing, work cooperatives, etc.

4. Concluding Remark

We initially motivated our paper with reference to a concern about deteriorating social values in public debate. Our results say something about the impact of social rewards of varying strength on economic behaviour, but the theory is silent on *why* an erosion of social rewards might occur. It seems appropriate at this point to offer a few speculative remarks on this topic.

Coleman (1990) connects social rewards to externalities in decision making. Of course, the smaller is a risk sharing community the larger is the externality that an individual's decision has on *any given other individual*. Hence, if the size of a risk-sharing community increases over time, and if social sanctions relate to 'per capita externalities', this could be a reason why social sanctions against opportunistic behaviour lose power over time.

This idea is roughly consistent with the development in Sweden over the last century. One major starting point of the Swedish welfare state was the encouragement of 'friendly societies', which organised sickness insurance and unemployment insurance for its members; see Lindquist (1990) and Berge (1995). In the beginning, these societies were fairly small, composed perhaps of all workers in some factory in a town or all members of a local labour union. The workers knew each other to a large extent, and the social ties between them were probably strong. Over time, the Swedish government added more and more centrally designed regulation of these societies, and many societies were merged into larger units. In the sickness insurance case, the development reached its peak in 1955, with a comprehensive nationally regulated system which completely replaced the friendly societies. A similar pattern applies to unemployment insurance, although such insurance is still voluntary and mainly organised by friendly societies related to (large) labour unions. The insurance conditions are nationally regulated and the societies are almost completely financed through government subsidies.

Stockholm University

Date of receipt of first submission: April 1998

Date of receipt of final typescript: October 2000

Appendix

The proof of the theorem relies on four lemmata. These show in turn that the equilibrium effort and inference functions x^* and τ^* are monotone, that the effort levels chosen in equilibrium form a connected set, and that there are no pools:

LEMMA 1. *If $t < t'$, then $x^*(t) \leq x^*(t')$.*

Proof of Lemma 1. Suppose $t < t'$. Define $x = x^*(t)$, $x' = x^*(t')$, $s = \tau^*(x)$ and $s' = \tau^*(x')$. We want to show that $x \leq x'$. Suppose to the contrary that $x' < x$. By Definition (i) (incentive compatibility) it holds that

$$U(x, t, s) \geq U(x', t, s') \quad \text{and} \quad U(x', t', s') \geq U(x, t', s) \quad (7)$$

which combined gives

$$Q = [U(x', t, s') - U(x, t, s)] - [U(x', t', s') - U(x, t', s)] \leq 0. \quad (8)$$

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

$$Q = [1 - u(\beta)](t - t')(x' - x). \quad (9)$$

Since $t < t'$ and $x' < x$ by assumption, it must be by (9) that $Q > 0$. This is a contradiction and therefore $x \leq x'$. \square

LEMMA 2. *If $x < x'$, then $\tau^*(x) \leq \tau^*(x')$.*

Proof of Lemma 2. It follows from Lemma 1 and Definition (ii) that τ^* is (strictly) increasing over equilibrium efforts. Hence it remains to prove τ^* is monotone also out-of-equilibrium. Note two things (a) and (b): (a) If $0 < t < t'$ and $x^*(t) < x^*(t')$ 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 $x^*(t)$). (b) Given τ^* , higher talents gain more (or lose less) than lesser talents by increasing effort from one given level to another given level, and lesser 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 (8) and (9) above, for different combinations of t, t', x, x' .) Combining (a) and (b) and Definition (ii) and (iii) one sees that $\tau^*(x) = 0$ for any $x < x^*(0)$, and that $\tau^*(x) = 1$ for any $x > x^*(1)$. It now only remains to prove monotonicity of τ^* when x is an out-of-equilibrium effort such that $x^*(0) < x < x^*(1)$. 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 by Definition (iii) we get $\tau^*(x) = \underline{t} = \bar{t}$. Combining this observation with the previous ones, one sees that τ^* is monotonic. \square

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

Proof of Lemma 3. Suppose Lemma 3 is not true. Then there exists an out-of-equilibrium effort x such that $x^*(0) < x < x^*(1)$. Let \underline{x} be the supremum of the equilibrium effort choices below x , let \bar{x} be the infimum of the equilibrium effort choice above x . By (b) in the proof of Lemma 2 and Definition (iii) one sees that $\tau^*(x)$ has a uniquely determined value. By continuity of total payoffs in talent, one infers that $u[\underline{x}, \tau^*(x), \tau^*(\underline{x})] = u[x, \tau^*(x), \tau^*(x)] = u[\bar{x}, \tau^*(x), \tau^*(\bar{x})]$. These equalities cannot hold unless $\tau^*(\underline{x}) \neq \tau^*(x) \neq \tau^*(\bar{x})$, since material payoffs are strictly concave in effort. Then, by Lemma 2, it holds that $\tau^*(\underline{x}) < \tau^*(x) < \tau^*(\bar{x})$. If \bar{x} is an equilibrium effort, this is impossible; any individual of talent $t \in T$ with $x^*(t) = x$ would gain by choosing $(\bar{x} - \varepsilon) \in X$ for ε 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 γ with $0 < \gamma < \varphi$ such that $\tau^*(\bar{x} + \gamma) \in T$ would gain by choosing \bar{x} . Hence no τ^* exist which sustains (x^*, τ^*) as an equilibrium, a contradiction. \square

LEMMA 4. $\nexists t, t' \in T$ s.t. $t \neq t', x^*(t) = x^*(t')$.

Proof of Lemma 4. Suppose Lemma 4 is not true. Then there exists $t, t' \in T$ such that $t \neq t'$ and $x^*(t) = x^*(t')$. By Lemma 1 we can find a connected set $\Theta \subseteq T$ of talents pooling at $x^*(t)$. Let \underline{t}, \bar{t} be the infimum and supremum of Θ . By Definition (ii), $\underline{t} < \tau^*[x^*(t)] < \bar{t}$. It is impossible that $x^*(t) > 0$ since then we can find $\varepsilon > 0$ small enough that individuals with talent $(\underline{t} + \varepsilon) \in \Theta$ would gain by choosing $x^*(t) - \varepsilon$ (the loss of material payoff is arbitrarily small and outweighed by a substantial social payoff gain). Hence it must be that $x^*(t) = 0$. 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 $\{[1 - u(b)] + \sigma\}/K$, 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 $t^*(\varepsilon)$ chooses ε , but he can in fact 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

Proof of the theorem: Combining Lemmata 1, 3, and 4 one sees that x^* must be strictly increasing and continuous. Moreover, x^* must be differentiable; in order to prescribe equilibrium choices for all talents x^* must for each t be tangent to an indifference curve as given by (6), and this would be impossible if x^* 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 τ^* must be differentiable since τ^* is the inverse of x^* , and x^* is differentiable. Hence we can proceed by maximising

$$tx + (1 - tx)u(\beta) - \frac{K}{2}x^2 + \sigma \left(x - \frac{1}{K} \{ [1 - u(\beta)]\tau(x) \} \right), \tag{10}$$

with respect to x to get the first order condition

$$\tau'(x) = K \frac{[1 - u(\beta)]\tau(x) + \sigma - Kx}{\sigma[1 - u(\beta)]}. \tag{11}$$

This is a linear first order differential equation which can be solved by standard methods. However, we need an initial condition. As explained in Section 2, in a separating equilibrium the individual with talent 1 must choose as if his talent were known. Hence he chooses $x^c(1) = \frac{1}{K}[1 - u(\beta) + \sigma]$ (as defined in Section 2), and the appropriate initial condition is $\tau\{\frac{1}{K}[1 - u(\beta) + \sigma]\} = 1$. The definite solution to (11) then is

$$\tau(x) = \frac{K}{1 - u(\beta)} x - \frac{\sigma}{1 - u(\beta)} e^{\frac{K}{\sigma}(x - \frac{1}{K}[1 - u(\beta) + \sigma])}. \tag{12}$$

Equation (12) describes the equilibrium inference function τ^* for equilibrium levels of effort. Equilibrium effort in the separating equilibrium is given by

$$x^*(t) = \frac{[1 - u(\beta)]t}{K} + \frac{\sigma e^{\frac{K}{\sigma}\{x^*(t) - \frac{1}{K}[1 - u(\beta) + \sigma]\}}}{K}. \quad (13)$$

Since equilibrium efforts increase with talent (Lemmata 1 and 4), effort levels above $x^*(1) = \frac{1}{K}[1 - u(\beta) + \sigma]$ are irrelevant in (12). Moreover, effort levels below $x^*(0)$ are irrelevant. By Lemma 2, $\tau^*(x) = 0$ for any $x \in [0, x^*(0))$ and $\tau^*(x) = 1$ for any $x \in (x^*(1), 1]$.

x^* is the inverse of τ^* for equilibrium effort levels. Hence substituting t for $\tau(x)$ and $x^*(t)$ for x in (12), and rearranging we get (13) which implicitly defines equilibrium efforts for all talents in T .

Letting $t = 1$ in (13) it is straightforward to verify that $x^*(1) = \frac{1}{K}[1 - u(\beta) + \sigma]$. Letting $t = 0$, and noticing that the exponential expression in (13) (in the following denoted $e^{K/\sigma(\cdot)}$) is less than unity (follows from Lemma 1 + separation), one sees that $x^*(0) = (\sigma/K)e^{K/\sigma(\cdot)} \in (0, \sigma/K)$. Finally, since $\partial^2 x^*/\partial t^2 = 1 - u(\beta)/K\sigma[1 - e^{K/\sigma(\cdot)}]^3 > 0$, x^* is strictly convex. \square

References

- Arnott, R. and Stiglitz, J. E. (1991). 'Moral hazard and nonmarket institutions: dysfunctional crowding out or peer monitoring?' *American Economic Review*, vol. 81, pp. 179–90.
- Arrow, K. J. (1963). 'Uncertainty and the welfare economics of medical care'. *American Economic Review*, vol. 53, pp. 941–71.
- Arrow, K. J. (1970). 'Political and economic evaluation of social effects and externalities.' In (J. Margolis, ed.), *The Analysis of Public Output*, pp. 1–23, New York. National Bureau of Economic Research.
- Berge, A. (1995). *Medborgarrätt och egenansvar. De sociala försäkringarna i Sverige 1901–1935*. Arkiv, Lund.
- Bernheim, B. D. (1994). 'A theory of conformity.' *Journal of Political Economy*, vol. 102, pp. 841–77.
- Besley, B. T. and Coate, S. (1992). 'Understanding welfare stigma: taxpayer resentment and statistical discrimination.' *Journal of Public Economics*, vol. 48, pp. 165–83.
- Bird, E. J. (1999). 'Can welfare policy make use of social norms?' *Rationality and Society*, vol. 11, pp. 343–68.
- Brito, D. L., Hamilton, J. H., Slutsky, S. M. and Stiglitz, J. E. (1991). 'Dynamic optimal income taxation with government commitment.' *Journal of Public Economics*, vol. 44, pp. 15–35.
- Bröms, J., Eriksson, I., Persson, I. and Schubert, G. (1994). *En social försäkring. Rapport till Expertgruppen för studier i offentlig ekonomi*. Ds 1994:81. Finansdepartementet, Stockholm.
- Cho, I.-K. and Kreps, D. M. (1987). 'Signaling games and stable equilibria.' *Quarterly Journal of Economics*, vol. 52, pp. 179–221.
- Coleman, J. S. (1990). *Foundations of Social Theory*. Cambridge, MA: and London: The Belknap Press of Harvard University Press.
- Dutta, P. and Radner, R. (1994). 'Moral hazard.' In (R. J. Aumann and S. Hart, eds), *Handbook of Game Theory*, volume 2, pp. 869–903, Amsterdam: Elsevier.
- Ehrlich, I. and Becker, G. (1972). 'Market insurance, self-insurance, and self-protection.' *Journal of Political Economy*, vol. 80, pp. 623–48.
- Fershtman, C., Murphy, K. M. and Weiss, Y. (1996). 'Social status, education, and growth.' *Journal of Political Economy*, vol. 104, pp. 108–32.
- Fershtman, C. and Weiss, Y. (1993). 'Social status, culture and economic performance.' *ECONOMIC JOURNAL*, vol. 103, pp. 946–59.
- Fershtman, C. and Weiss, Y. (1998). 'Social rewards, externalities and stable preferences.' *Journal of Public Economics*, vol. 70(1), pp. 53–73.
- Goffman, G. (1963). *Stigma: Notes on the Management of Spoiled Identity*. Englewood Cliffs, NJ: Prentice-Hall.
- Gottfries, N. and McCormick, B. (1995). 'Discrimination and open unemployment in a segmented labour market.' *European Economic Review*, vol. 39, pp. 1–15.
- Güth, W. and Yaari, M. E. (1992). 'Explaining reciprocal behavior in simple strategic games: an evolutionary approach.' In (U. Witt, ed.), *Explaining Process and Change—Approaches to Evolutionary Economics*, pp. 23–34. Ann Arbor, Michigan.
- Kandel, K. and Lazear, E. P. (1992). 'Peer pressure and partnership.' *Journal of Political Economy*, vol. 100, pp. 801–17.

- Lindbeck, A. (1995). 'Welfare state disincentives with endogenous habits and norms.' *Scandinavian Journal of Economics*, vol. 97, pp. 477-94.
- Lindbeck, A., Nyberg, S. and Weibull, J. W. (1996). 'Social norms and economic incentives in the welfare state.' *Quarterly Journal of Economics*, vol. 114(1), pp. 1-35.
- Lindquist, R. (1990). *Från folkrörelse till välfärdbyråkrati*. Arkiv, Lund.
- McCormick, B. (1990). 'A theory of signalling during job search, employment efficiency, and "stigmatised" jobs.' *Review of Economic Studies*, vol. 57, pp. 299-313.
- Mirrlees, J. A. (1971). 'An exploration in the theory of optimum income taxation.' *Review of Economic Studies*, vol. 38, pp. 135-208.
- Narendranathan, W., Nickell, S. and Stern, J. (1985). 'Unemployment benefits revisited.' *ECONOMIC JOURNAL*, vol. 95, pp. 307-29.
- Nordenmark, M. (1999). 'Non-financial employment motivation and well-being in different labour market situations: a longitudinal study.' In 'Unemployment, employment commitment, and well-being: the psychosocial meaning of unemployment among women and men', Doctoral Thesis No. 10. Department of Sociology, Umeå University.
- Pauly, M. (1974). 'Overinsurance and public provision of insurance: the roles of moral hazard and adverse selection.' *Quarterly Journal of Economics*, vol. 88, pp. 44-62.
- Shavell, S. (1979). 'On moral hazard and insurance.' *Quarterly Journal of Economics*, vol. 93, pp. 541-62.