Who bene…ts from workers with general skills? Countervailing incentives in labor contracts[¤]

Dorothea Kübler[y]
Humboldt University Berlin

January 10, 2002

Abstract

In a principal–agent model with adverse selection and moral hazard the impact of the agent's transferable human capital on incentives is analyzed. It is shown that under asymmetric information the employer (principal) prefers a worker (agent) with general skills to a similarly productive worker with …rm–speci…c skills although the reservation utility of a worker with general (i.e., marketable) skills is higher. The principal's information costs are lower when workers have general skills than in the case where workers possess only …rm–speci…c human capital because of countervailing incentives. The optimal contract for workers with general skills di¤ers from the standard screening contract in that it involves pooling.

[¤] Thanks to Pio Baake, András Lö - er, Jörg Oechssler, and Elmar Wolfstetter. Financial support from the Norwegian Research Council (NFR–Ruhrgas) and the Deutsche Forschungsgemeinschaft}, SFB 373 ("Quanti…kation und Simulation Ökonomischer Prozesse"), is gratefully acknowledged. Revision of the paper during a research fellowship at U.C. Berkeley has been made possible by a Post–Doc scholarship from the DAAD (Deutscher Akademischer Austauschdienst) funded by the "Gemeinsames Hochschulsonderprogramm III von Bund und Ländern".

[y] Institute for Economic Theory I, Spandauer Str. 1, 10178 Berlin, Germany, Fax +49 30 2093 5619, email: kuebler@wiwi.hu–berlin.de.

1 Introduction

From agency models of labor contracts with unobservable e¤ort and type of the agent, it is well–known that workers try to understate their true productivity (or overstate the di¢culty of the job) when the employer o¤ers a contract which is optimal under full information.[1] However, this seems to capture only half the truth. It is often the case that a worker tries to exaggerate her productivity. She thereby wants to increase her wage or press for more interesting tasks. In many cases, this will be successful if she is able to get job o¤ers from other employers. The more productive she is in the current job, the better these o¤ers are going to be. Hence, the current employer will have to increase her wage in response to the higher reservation wage, to prevent her from quitting the job. In other words, if a worker's ability is private information to her and outside opportunities depend on her ability, she has an incentive to overstate her productivity in order to receive a higher wage.[2]

In this setting, the optimal contract has the opposite properties of the standard screening contract: If there is an incentive to overstate one's productivity, an incentive compatible contract has to make lying expensive by demanding ine¢ciently high levels of output from all types. The worker earns an information rent which can be limited by output distortions. Only the least productive worker's output is not distorted because nobody will try to mimic her ('no distortion at the bottom'). The best type earns no rent because she cannot announce a better type.

It is puzzling at …rst sight why employers seem to take interest in general skills, because with …rm–speci…c training, a worker would not be able to transfer human capital to another job and it would be less costly to prevent turnover. If the worker loses some of her human capital by leaving the employer, her participation constraint is not as tight as that of an equally productive worker with general skills.

The starting point of the paper is to identify the incentive to exaggerate one's ability with the possession of general skills. General skills make a worker's outside

[1] See for example Gibbons (1987) and McAfee and McMillan (1987). The related problems of regulation and procurement when …rms have unknown costs and the e¤ort of the manager is unobservable are treated in La¤ont and Tirole (1986).

[2] Another way of creating an incentive to overstate one's ability are internal hierarchies where workers receive a higher wage and a better job in return for good performance.

option more sensitive to her ability. One reason is that general skills give workers a wide range of employment opportunities, allowing them to choose a job that makes full use of their various strengths, and thus widening the gap between those with high and low abilities.[3] Unskilled workers on the other hand usually receive a uniform (i.e., type independent) wage. I will use an incentive based argument to show that employers may prefer to employ workers with marketable skills instead of equally productive workers with firm–specific skills (coming from the internal labor market for example). If a worker has abilities which open up outside opportunities to her, she will be less inclined to underperform in order to mimic a less productive type.[4] On the other hand, exaggerating one's ability in order to increase the reservation utility is costly for a worker as it is accompanied by a higher disutility for producing higher output. Thus, if a worker has marketable skills and the employer cannot observe her productivity, incentives to misrepresent her type go into opposite directions. Depending on the slope of the disutility function and of the reservation utility, some types may find it more advantageous to overstate and some to understate their ability. It is possible that both incentives just cancel out for certain types.

In this paper, the theory of countervailing incentives, developed by Lewis and Sappington (1989a) in the context of regulation, is applied to the question of optimal skill specificity of workers.[5] It will be shown below that a positive level of general training introduces countervailing incentives and thus leads to an interval of pooling where optimal output and wage do not vary with the type of the worker. The simple optimal contract which emerges for a wide range of types offers an explanation for the pervasiveness of flat contracts, that is, the absence of screening in many real life situations.

The main purpose of the paper is to stress features germane to transferable skills in contrast to firm–specific skills.[6] Therefore, it is assumed that a worker's productivity

[3] Note that mobility of workers (which can be due to general skills among other things, such as proficiency in a foreign language) has the same effect on the reservation utility.

[4] The underlying mechanism that disciplines workers is similar to Fama's (1980) market for managers. In Fama (1980), market determined opportunity wages correctly reflect the manager's performance in the firm.

[5] In another paper Lewis and Sappington (1989b) sketch that employers can use mobility constraints to influence the reservation utility of their workers.

[6] Of course, these two types of skills are only the extreme points. Schlicht (1996) allows for a

is identical for both sorts of skills, and that general and speci…c skills are perfect substitutes. Productivity e¤ects and costs of acquiring skills are abstracted from as the model does not address issues of optimal productivity of workers. The question is rather whether the employer prefers to hire workers with general or with …rm-speci…c skills, holding productivity constant.

General skills can be obtained via university diplomas, trainee programs, or apprenticeships for example. In Germany, apprenticeships are characterized by a highly standardized system of schooling and exams.[7] And there is evidence that …rms incur net costs during training periods.[8] What is more, a large share of workers quits after an apprenticeship or a trainee program.[9] Some people stay, some accept job o¤ers by other employers, and some leave the employer who trained them after some years of working for him. The fact that employers o¤er apprenticeships although they are costly may be interpreted as evidence that employers take an interest in general skills of workers.

In the literature, the question of general versus …rm–speci…c skills has focused on why …rms train workers and on how they can prevent costly turnover after training them.[10] In the model presented below, the question is not why employers provide general training, but rather what are the speci…c advantages of employing a worker with

continuum of transferability and Stevens (1994) introduces "transferable" training as creating skills which are of some value to other …rms, but for which perfect competition is not assumed (as it typically is for general skills).

[7]Federal rules determine the duration, the minimum requirements for …nal exams, the organization of the training, and the occupational pro…le. There are more than 300 occupations for which such a nationally recognized certi…cate can be obtained.

[8]According to a study of the Bundesinstitut für Berufsbildung by von Bardeleben et al. (1994) for the year of 1991, the average net cost of one apprentice per year (total cost minus the bene…t from her work) amount to DM 18,051.

[9]Harho¤ and Kane (1997) estimate that 70% of apprentices leave the training …rm in the …rst 5 years. In contrast, 80% af all German workers report that they still have their …rst job.

[10]The most prominent explanation for the economic rationale of providing general skills is o¤ered by human capital theory. It is shown that other than in the case of …rm–speci…c training, the costs of providing general skills must be shifted to the worker (see Becker (1962)). To prevent turnover, employers can also make deferred payments or use the apprenticeship (with low or no pay) as an entrance fee to sort worker types (see Guasch and Weiss (1981)).

general skills as opposed to one with firm–specific skills.[11] In contrast to several articles in the existing literature, it is not assumed that turnover is limited by informational or cost asymmetries.[12] And fluctuation of workers does not make marketable skills less attractive to employers, but rather constitutes their beneficial incentive effect. Assuming asymmetric information between the current employer and the worker, the optimal mix of general and specific skills is shown to help solving the firm's incentive problem. In the model general skills are only useful to the employer when asymmetric information prevails, because it is always optimal to hire workers with firm–specific skills under full information.

In the next section the model is presented. In Section 3 the properties of the optimal contract are derived. In a series of lemmas (following the approach in Lewis and Sappington (1989a)), the structure of the contract is developed. Finally, a necessary and sufficient condition for hiring workers with a strictly positive level of general skills is derived. When workers are uniformly distributed, output distortions in the optimal contract are always smaller when workers are flexible and therefore welfare is larger (Section 4). Section 5 summarizes the results of the model.

2 The model

The employer offers a contract to the worker who can either accept or reject it. The worker produces an output y which is publicly observable. The employer attaches monetary value of $V(y)$ to output. This function is increasing and strictly concave in output, $V' > 0$ and $V'' < 0$ for all $y \geq 0$.

The worker's disutility of effort function $\phi(y;\theta)$ is common knowledge, but the

[11] However, the model also demonstrates that training costs can be relatively low if general skills are taught. This is due to the fact that apprentices or trainees work for the firm during their apprenticeship. The (anticipated) flexibility of a worker makes it worthwhile for her to perform well, already during the training period, because her performance influences her future wage directly through the effect on the participation constraint. Von Bardeleben et al. (1994) report that the costs of general training are lowered considerably by the apprentice's work for the firm.

[12] In a number of articles, general skill provision is profitable for firms, because the training firm gains superior information about the ability of a worker compared to firms that recruit skilled workers (see Katz and Ziderman (1990), Chiang and Chiang (1990), Bernhardt (1995), and Pischke (1998)).

employer can neither observe the worker's type µ, nor her e¤ort needed to produce some output y. Disutility is an increasing and convex function of output ($Á_y > 0$ and $Á_{yy}$ ¸ 0). Conversely, disutility is decreasing in the productivity of the worker with a decreasing rate, $Á_µ < 0$ and $Á_{µµ}$ · 0. The marginal disutility of output is lower for better types, $Á_{yµ} < 0$, which is the standard single–crossing–property.

The worker's type is drawn from a distribution on the closed interval $[\underline{µ};\overline{µ}]$ with distribution function F(µ) and density function f(µ). The distribution of types is common knowledge. The distribution function satis…es the monotonicity conditions which are commonly used to ensure that the optimal contract is fully separating:

$$\frac{d}{dµ}\left(\frac{F(µ)}{f(µ)}\right) ¸ 0 \quad \text{and} \quad \frac{d}{dµ}\left(\frac{1-F(µ)}{f(µ)}\right) · 0:$$

Both conditions are satis…ed by the uniform distribution for example. The employer's utility is the expected value of V(y)-w(y) where w denotes the wage paid to the worker for output y. The risk–neutral worker's net utility, that is, her rent in excess of the reservation level, is equal to w(y)-Á(y;µ) minus her reservation utility denoted by $\overline{U}$.

What distinguishes this model from the standard screening model is that the reservation utility is not constant, but depends on the worker's type and the level of general skills, $\overline{U}(µ;I)$.[13] It is assumed that it increases with the productivity parameter, $\overline{U}_µ > 0$ for all µ when I > 0: This means that the more productive a worker is, the higher her reservation utility as long as she has some general skills. This assumption can be justi…ed by the observation that workers of high productivity can get more attractive alternative jobs than workers with low productivity. Without any general skills, the reservation wage is normalized to zero, $\overline{U}(µ;0) = 0$, or in other words, without some basic general training, the utility from leaving the …rm is insensitive to a worker's type. Totally unskilled workers cannot make use of their abilities as their jobs will only involve highly standardized tasks, but for all others market opportunities increase, the higher their personal abilities.[14] Moreover, the reservation utility is strictly concave

[13] A similar assumption is made by Moore (1985) who considers optimal labor contracts when the reservation utility is private information to the worker.

[14] Note that other employers need not observe µ directly for a worker's outside option to increase as µ increases. In general, a worker can earn higher expected rents under any optimal incentive scheme with hidden information, the greater his ability. See Lewis and Sappington (1989a), note 7, for an analogous argument.

in μ, $\overline{U}_{\mu\mu} < 0$, that is, there are decreasing returns to ability.[15] The function $\overline{U}(\cdot)$ is common knowledge.

While firm–specific skills do not increase a worker's opportunities outside the firm, general skills are transferable and can thus be used to press for a higher wage. Thus, $\overline{U}_l > 0$. Furthermore, it is assumed that $\overline{U}_{\mu l} > 0$. Thus, a worker's reservation utility is more sensitive to type, the more general skills she possesses. This can be justified by the observation that the more different tasks a person is able to perform, the better she can choose a job which suits her own abilities and makes use of her specific strengths.[16] Notice that l is a parameter of the model. The effect of a variation in this parameter on the properties of the optimal contract will be studied. It is assumed that the employer can observe the general skills of a worker (for example because workers hold certificates or diplomas).

Note that general skills do not have a direct productivity effect nor any direct cost as the productivity level of a worker is held constant. The model deals with a different tradeoff: The only cost of general skills is the increase in reservation utility and its sole benefit to the employer is the reduction in information rent.

There are only two players to the game, the employer and the worker. The employer proposes a contract which specifies a set of wages and output levels contingent on the announced type. Using the revelation principle, the employer maximizes over all direct incentive compatible mechanisms denoted by the pair of functions $\{w(\mu); y(\mu)\}$. The worker reports her type μ truthfully, receives the wage $w(\mu)$, and produces an output of $y(\mu)$. The timing of the game is as follows:

stage 1: The employer offers a contract $\{w(\mu); y(\mu)\}$.

[15] Maggi and Rodríguez–Clare (1995) show that this assumption is crucial for pooling to emerge in the optimal contract. If the reservation utility is a convex function of μ, the contract is fully separating.

[16] Also, empirical evidence supports the relationship between ability and general education, expressed by $\overline{U}_{\mu l} > 0$. For example, it has been shown that a composite measure of aptitude is positively correlated with returns to postsecondary education in the U.S. See Altonji (1995) and Weiss (1995) for a discussion. Moreover, Arulampalam and Booth (1997) explain the finding that workers with a low level of general education receive relatively less work–related training than those with high levels of general education by the fact that "bright workers and workers with a sound educational background will learn faster than their less able colleagues" (p.199).

stage 2: The worker either rejects the contract or accepts it, and production takes place.

stage 3: The employer's payoff is realized and wages are paid.

The analysis proceeds as follows: First the optimal contract in stage 1 is characterized for all possible levels of l. Then, given the properties of the contract, the level of general skills which maximizes the employer's payoff is derived.

In the benchmark case of full information, the worker's type is common knowledge. The employer maximizes his expected payoff such that

$$V'(y) - \psi_y = 0 \quad \forall \theta \in [\underline{\theta}; \overline{\theta}]:$$

The first best output level is denoted by $y^*(\theta)$. The worker receives a wage which satisfies the participation constraint as an equality, $w(\theta) = \psi(y^*(\theta); \theta) + \overline{U}(\theta; l)$, because information rents are zero. Under full information, the employer never prefers to hire a worker with a positive level of general skills, because it would only increase his cost.

With asymmetric information, the contract has to ensure that the worker agrees to work for the employer and that she announces her type truthfully. To gain some intuition, consider the worker's incentive to misrepresent her type when she is offered the first-best contract. Given the contract which specifies the first-best wage and first-best output level for her announced type, a worker wants to maximize her information rent

$$U(w^*(\tilde{\theta}); y^*(\tilde{\theta}); \theta; l) = w^*(\tilde{\theta}) - \psi(y^*(\tilde{\theta}); \theta) - \overline{U}(\theta; l);$$

where $\tilde{\theta}$ denotes the worker's announced type. Maximizing this expression, using the envelope theorem, yields

$$\frac{\partial U}{\partial \tilde{\theta}} \Big|_{\tilde{\theta} = \theta} = \psi_\theta + \overline{U}_\theta:$$

As $\psi_\theta < 0$ and $\overline{U}_\theta > 0$ for $l > 0$, the right-hand side in general does not vanish, and it illustrates the countervailing incentives: overstating one's ability increases the reservation utility, but decreases the disutility from effort, whereas understating one's ability increases the disutility from effort, but decreases the reservation utility.

Now consider the optimal contract under asymmetric information. The employer's problem [P] in stage 1 is

$$\max_{y(\theta),w(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} [V(y(\theta)) - w(\theta)]\, dF(\theta)$$

subject to (for all $\theta, \tilde{\theta} \in [\underline{\theta}, \overline{\theta}]$)

$$\theta \in \arg\max_{\tilde{\theta}} \{ w(\tilde{\theta}) - \psi(y(\tilde{\theta}); \theta) \} \tag{1}$$

$$U(w(\theta), y(\theta), \theta, I) \equiv w(\theta) - \psi(y(\theta); \theta) - \overline{U}(\theta; I) \geq 0. \tag{2}$$

The participation constraint (2) is satisfied if the worker's net utility, defined as the difference between the reservation utility and the utility derived from the employment relation, is positive. For simplicity, assume that the employer wants to hire every type of worker, $y(\theta) > 0 \quad \forall \theta \in [\underline{\theta}, \overline{\theta}]$, in the solution to [P].[17]

3 The optimal contract

The optimal contract for a worker with a positive level of general skills is described by a series of lemmas. Lemmas 1 and 2 establish general properties of the rent and output function in the optimal contract while Lemmas 3 to 6 characterize the pooling interval and determine the optimal output distortions. The aim is to show under which condition the hidden information problem makes a higher level of general skills preferable for the employer than under symmetric information. The result is stated in Proposition 1. The employer's problem [P] can be solved using the theory of optimal control. However, the maximization problem can be simplified substantially by first computing the rent of the worker and substituting it into the objective function of [P]. The resulting indirect utility function is maximized pointwise for every θ.

As a first step, the participation and incentive compatibility constraints, (1) and (2), are used to derive properties of the optimal contract. Lemma 1 is a standard result of the optimal contract literature. The techniques used for the proof go back to Mirrlees

[17] It is implicit in this type of screening model that the supply of labor is limited or, alternatively, that search is costly. Hence, it is impossible or too expensive for a firm to search only for the best type of worker. The model captures this as a special case in which the firm offers only one contract from the menu which satisfies the participation constraint of the best type as an equality.

(1971) and Baron and Myerson (1982). Feasibility of a solution means that it satis…es the constraint sets of incentive compatibility (1) and participation (2).

Lemma 1 For any feasible solution of the program [P] it must hold that $U_\mu = -Á_\mu - \overline{U}_\mu$ and $y^0(\mu)$ ¸ 0 almost everywhere.

Incentive compatibility requires that output $y(\mu)$ be a nondecreasing function. Monotonicity of $y(\mu)$ and the condition $U_\mu = -Á_\mu - \overline{U}_\mu$ imply that truth–telling is a global optimum for every type μ.[18] Continuity of $y(\mu)$ in the optimal contract will be shown in the proof of Lemma 6, but it is not self–evident.

Lemma 2 In any continuous solution to the maximization program, the participation constraint holds as an equality for exactly one realization of the productivity parameter, μ_K 2 $[\underline{\mu}; \overline{\mu}]$, because the worker's rent is a strictly convex function, $U_{\mu\mu} = -Á_{\mu\mu} - Á_{y\mu} y^0 - \overline{U}_{\mu\mu} > 0$.

Proof Di¤erentiate the slope of the rent function $U_\mu = -Á_\mu - \overline{U}_\mu$ with respect to μ and notice that $\overline{U}_{\mu\mu} < 0$. 2

In the standard adverse selection model,[19] it is the least productive type who receives no rent, $\mu_K = \underline{\mu}$, because she cannot mimic a less productive type. Here, this is the case if the incentive to understate one's ability always dominates the incentive to overstate μ. However, the opposite is also possible. If the incentive to exaggerate μ is always stronger than the incentive to understate μ, it is the best type, $\overline{\mu}$, who receives no rent. The assumptions made above allow for both cases at the same time. Then the worker who earns no rent (i.e., who has no incentive to lie) is of an intermediate type. To see this note that if the incentive to overstate one's type is stronger than the incentive to understate one's type for low types, but it is dominated by the incentive to understate one's type for high types, then there must be an intermediate type μ_K for whom the participation constraint is binding. For this type the incentives to mimic a better or a worse type are equally strong and thus cancel out. The incentive constraints bind upwards for all $\mu < \mu_K$ and downwards for all $\mu > \mu_K$. This implies that a worker's rent decreases in μ for types below μ_K and it increases in μ for types above μ_K.

[18] For the proof see La¤ont and Tirole (1993), p. 121.

[19] See Baron and Myerson (1982) and La¤ont and Tirole (1993).

Now, if an intermediate type θ_k exists in the interior of the interval of types $[\underline{\theta};\bar{\theta}]$, it can be shown that the optimal contract involves pooling, that is, different types of workers $\theta \in [\theta_1;\theta_2]$ with $\underline{\theta} \leq \theta_1 \leq \theta_k \leq \theta_2 \leq \bar{\theta}$ receive a uniform wage denoted by $\bar{w}$ and are required to produce a uniform level of output $\bar{y}$.

Lemma 3 There is at most one pooling interval. If there exists a parameter θ_k in the interior of $[\underline{\theta};\bar{\theta}]$, pooling occurs in the interval $[\theta_1;\theta_2]$ at the first best level of output and wage for type θ_k, $\bar{y} = y^*(\theta_k)$ and $\bar{w} = w^*(\theta_k)$:

Proof see Appendix. 2

In short, the reason for pooling to emerge in the optimal contract is that incentive compatibility requires output to be nondecreasing in θ over the whole range of possible types.[20] If the slope of the output function $y(\theta)$ in the optimal contract is steeper than the first best $y^*(\theta)$ for small values of θ and flatter than $y^*(\theta)$ for high values of θ and if $y(\theta)$ is continuous, incentive compatibility can only be secured by a flat scheme with $y'(\theta) = 0$ for an intermediate range of types. This is illustrated by Figure 1. Note that the optimal ouput is determined by the optimality condition that output be larger than first best for types $\theta < \theta_k$ and smaller than first best for types $\theta > \theta_k$.

[Insert Figure 1 about here]

It has not been established yet whether the pooling interval $[\theta_1;\theta_2]$ extends all the way to $\underline{\theta}$ or $\bar{\theta}$. Lemma 4 shows that the existence of an interior type θ_k (as defined above) is necessary, but not sufficient for the boundaries of the pooling interval to be in the interior, too.

Lemma 4 The lowest type $\underline{\theta}$ and the highest type $\bar{\theta}$ are not pooled if the following condition holds:

$$\theta_1 > \underline{\theta} \wedge \theta_2 < \bar{\theta} \iff \bar{U}_\theta \in (-\phi_\theta(y^*(\underline{\theta});\underline{\theta}); -\phi_\theta(y^*(\bar{\theta});\bar{\theta})) \quad \forall\theta \in [\underline{\theta};\bar{\theta}]: \qquad (3)$$

[20] However, countervailing incentives are not sufficient for the optimal contract to entail pooling. It is necessary that the worker's utility be a quasiconvex function in the private parameter (see Maggi and Rodríguez–Clare (1995)), which is ensured by the assumptions made above.

Proof see Appendix. $\square$

Now, the optimal contract can be characterized regarding output distortions.

Lemma 5 If there is an interior μ_k, output levels of the least and the most productive type as well as of the intermediate type μ_k are not distorted,

$$y(\underline{\mu}) = y^*(\underline{\mu}); \quad y(\bar{\mu}) = y^*(\bar{\mu}); \quad y(\mu_k) = y^*(\mu_k):$$

Workers with low productivity are forced to produce more than the optimal output level,

$$y(\mu) > y^*(\mu) \quad \forall \mu \in (\underline{\mu}; \mu_1];$$

whereas workers with high productivity produce less than the optimal output level,

$$y(\mu) < y^*(\mu) \quad \forall \mu \in [\mu_2; \bar{\mu}):$$

Output levels for types to the left and to the right of the pooling interval are either distorted upwards or downwards. In particular, bad workers produce more than their optimum whereas good workers work less than the efficient amount. It remains to be shown how the pooling interval $[\mu_1; \mu_2]$ is defined and that pooling is a feature of the optimal contract when the employer hires a worker with general skills.

Lemma 6 In the optimal contract, $y(\mu)$ is a continuous function. The boundaries of the pooling interval are characterized by the following properties:

1. At μ_1, the profit of the employer with output level $y(\mu_1)$ is equal to his profit with output $\bar{y}$ (or greater if $\mu_1 = \underline{\mu}$).

2. At μ_2, the profit of the employer with output level $y(\mu_2)$ is equal to his profit with output $\bar{y}$ (or smaller if $\mu_2 = \bar{\mu}$).

3. When the pooling interval does not extend to $\underline{\mu}$ or $\bar{\mu}$, the parameter μ_k has the property that the expected marginal profit of the employer in the pooling interval is equal to the marginal increase in reservation utility for type μ_k.

Proof of Lemmas 5 and 6 see Appendix. 2

Lemma 6 is intuitive if one notices that the boundaries of the pooling interval are defined by the types for whom the employer is indifferent between separating and pooling. Monotonicity of $y(\theta)$ together with Lemma 5 implies that there is overproduction for all $\theta \in [\theta_1, \theta_k)$ and underproduction for all $\theta \in (\theta_k, \theta_2]$. These distortions are introduced to reduce the worker's rent. The employer would prefer to distort output even more in order to separate the different types of workers (the dotted lines in Figure 1). He would rather choose $y(\theta) > y^* \ \forall \theta \in [\theta_1, \theta_k)$ and $y(\theta) < y^* \ \forall \theta \in (\theta_k, \theta_2]$. But this violates incentive compatibility because the continuity of $y(\theta)$ would necessitate output to decrease in θ in the neighbourhood of θ_k.

Now the main result of the paper can be stated.

Proposition 1 The employer prefers a positive level of general skills, $l^* > 0$, if (at $l = 0$) the increase in reservation utility due to general skills is sufficiently stronger for good types than for bad types,

$$\int_{\underline{\theta}}^{\overline{\theta}} \overline{U}_{\theta l}(\theta;0)\frac{1-F(\theta)}{f(\theta)}dF(\theta) > \overline{U}_l(\theta_k;0). \tag{4}$$

Proof see Appendix. 2

The two opposing effects of general skills are a decrease in information rent (the left-hand side of expression (4)) and an increase in reservation utility (the right-hand side of expression (4)). The marginal increase in reservation utility is measured at θ_k, because at θ_k the participation constraint is binding. Intuitively Proposition 1 states that the employer prefers a strictly positive level of l if the cross derivative $\overline{U}_{\theta l}$ is sufficiently large, that is, if the effect of general skills on the reservation utility increases with the ability of the worker. Thus, it is advantageous for the principal to hire a worker with general skills if a good type, who has a strong incentive to understate her productivity, also has a strong opposite incentive. But a worker of low productivity who tends to overstate θ should not have a strong additional reason to do so because of a positive l. Then the higher cost due to a higher reservation utility is overcompensated by lower information rents.

[Insert Figure 2 about here]

Figure 2 shows the e¤ect of a positive level of general skills, l, on the rent paid to the worker. Countervailing incentives cut the interval of types into two halves. In each of them, one of the two incentives dominates the other. For type μ_k, both incentives are equally strong, $\overline{U}_\mu(\mu_k; l) = -Á_\mu(y(\mu_k); \mu_k)$, and the rent is zero. The condition of Proposition 1 assures that the slope of the left and right branch of the new rent function are not too steep, that is, that the area below the new rent function (weighted with $f(\mu)$) is not larger than the area below $U(\mu; 0)$. If general skills lower the incentive to understate μ for good types without creating too strong an incentive to overstate μ for bad types, the emplyoer prefers a worker with general skills.

Note that as soon as $l > 0$ (and l not too large), the contract involves pooling because the marginal rent of type $\underline{\mu}$ for understating her ability is zero and therefore $\overline{U}_\mu(\underline{\mu}; l) > -Á_\mu(y(\underline{\mu}); \underline{\mu}) = 0$. Thus, there is an interior μ_k with $\overline{U}_\mu(\mu_k; l) = -Á_\mu(y(\mu_k); \mu_k)$. Put somewhat loosely, it is easy (i.e., a small level of l su¢ces) to induce a bad type to overstate her productivity because she does not gain anything by understating it. And if there are types who want to overstate their ability (and if not all types do so), then pooling characterizes the optimal contract.

4 E¢ciency and welfare: An example

In this section, the optimal contract for workers with general skills is compared to the optimal contract when workers possess only …rm–speci…c skills. The main purpose of this section is to show for the simple case of uniformly distributed worker types that marketable skills of workers may improve allocative e¢ciency. With overall welfare de…ned as the sum of the employer's and the worker's utility, the distribution of rent is irrelevant. The crucial element of a welfare analysis is the e¤ect of asymmetric information on actual production. When the worker's ability is uniformly distributed over the interval of types, it can be shown that general skills reduce the output distortions of the optimal contract.

Proposition 2 *When the employer hires a worker with a positive level of general skills, output distortions relative to the …rst best are smaller than when a worker possesses only …rm–speci…c skills, assuming that worker types are uniformly distributed over $[\underline{\mu}; \overline{\mu}]$.*

Proof see Appendix. 2

Figure 1 provides some simple intuition for this result. Second-best ouput of workers with no general skills is given by y_L^0 whereas second-best output of workers with general skills is given by $y(\mu)$: In the proof it is shown that for all possible pooling intervals, the area between y_L^0 and $y^*(\mu)$ is always greater than the area between $y(\mu)$ and $y^*(\mu)$:

5 Discussion

The model demonstrates the value of general skills from an information economic perspective. It hints at a solution to the puzzle why employers offer general training instead of locking in workers by providing firm–specific training. If the hidden information problem is serious, for example because the interval of possible types is large, general skills are an effective means to lower information rents.

More specifically, it is shown that a worker with marketable skills may earn less information rents than a worker possessing mainly firm–specific skills. In addition, output can be closer to the efficient level when workers have general skills. The reason for these findings lies in the nature of countervailing incentives: When workers profit from overstating and from understating their ability, the overall incentive to misrepresent one's type is reduced. In the presence of countervailing incentives, the optimal contract differs from the standard screening contract. In particuar, the highest, the lowest, and an intermediate type produce the first-best ouput level whereas output of all other types is either distorted upwards or downwards relative to the first-best level. Furthermore, it is this intermediate type who earns no information rent, because her incentives to over- and understate ability are just equally strong. All other types earn rents, and rents are higher the more extreme a type is, that is, the closer her ability to the ability of the least or the most able type.

The model not only draws attention to the incentive effect of general skills, but also sheds some light on the pervasiveness of relatively simple contracts in practice. If the pooling interval is large, almost all types receive the same contract in the optimal mechanism. As a consequence, the losses of the employer from offering only one contract to all types can be relatively small in the presence of conflicting incentives.

Appendix

Proof of Lemma 1 The second part of the lemma is implied by incentive compatibility. Adding up the two incentive constraints which secure that any two types θ and θ' do not mimic each other respectively leads to, for all $\theta, \theta' \in [\underline{\theta}, \overline{\theta}]$,

$$\begin{aligned} & \phi(y(\theta);\theta') - \phi(y(\theta);\theta) \geq \phi(y(\theta');\theta') - \phi(y(\theta');\theta) \\ \Longleftrightarrow \; & \int_{\theta}^{\theta'} \int_{y(\theta')}^{y(\theta)} \phi_{xz}(x;z)\,dx\,dz \geq 0. \end{aligned}$$

If $\theta' > \theta$, then $y(\theta') \geq y(\theta)$ because $\phi_{y\theta}$ is negative. Thus, y is nondecreasing in θ. Therefore $y'(\theta) \geq 0$ almost everywhere.

The net rent of type θ who announces $\tilde{\theta}$ is

$$U(w(\tilde{\theta}); y(\tilde{\theta}); \theta; l) = w(\tilde{\theta}) - \phi(y(\tilde{\theta}); \theta) - \overline{U}(\theta; l). \tag{5}$$

The first–order condition for truth–telling is

$$w' - \phi_y y' = 0$$

where $\tilde{\theta} = \theta$ maximizes $U(w(\tilde{\theta}); y(\tilde{\theta}); \theta; l)$. The first part of Lemma 1 follows after plugging the first order condition into the first derivative of the rent function of the utility maximizing worker. (This procedure is equivalent to applying the envelope theorem directly to expression (5).) $\Box$

Proof of Lemma 3 First, define the functions y_L^0 and y_H^0 as output in the optimal contract when one of the incentives dominates for all types and the participation constraint binds as an equality either for type $\underline{\theta}$ or for type $\overline{\theta}$. Thus, y_L^0 is the optimal output scheme in the case that all types have an incentive to understate their ability (as in the standard model) and y_H^0 is the optimal output scheme when there is an incentive to overstate ability for all types. Now, in the solution to [P], the type θ_K who receives no rent is in the interior of $[\underline{\theta}, \overline{\theta}]$ if the following condition holds:

$$\theta_K \in (\underline{\theta}, \overline{\theta}) \Longleftrightarrow \overline{U}_\theta(\theta; l) \in (-\phi_\theta(y_L^0(\underline{\theta}); \underline{\theta}); -\phi_\theta(y_H^0(\overline{\theta}); \overline{\theta})) \quad \forall \theta \in [\underline{\theta}, \overline{\theta}]. \tag{6}$$

The right–hand side of expression (6) is satisfied if every type's incentive to overstate θ is strictly greater than type $\underline{\theta}$'s incentive to understate θ and strictly smaller than type $\overline{\theta}$'s incentive to understate θ when output is y_L^0 or y_H^0 respectively.[21]

[21] See Figure 1 for a graphical representation of y_H^0 and y_L^0.

To prove (6), first show that the second term follows from the first (sufficiency). Suppose $\underline{\theta} = \theta_k$. Because $\psi_{\theta\theta} \le 0$ and $\overline{U}_{\theta\theta} < 0$, it follows from $-\psi_\theta(\theta_k) = \overline{U}_\theta(\theta_k)$[22] that $\overline{U}_\theta \le -\psi_\theta$ for all θ. Then also $\overline{U}_\theta(\underline{\theta}) \le -\psi_\theta(y(\underline{\theta});\underline{\theta})$ which contradicts the right hand side of (6). Conversely, if $\theta_k = \overline{\theta}$ then $\overline{U}_\theta \ge -\psi_\theta$ and $\overline{U}_\theta(\overline{\theta}) \ge -\psi_\theta(y(\overline{\theta});\overline{\theta})$, which again contradicts the right hand side of (6).

The first expression in (6) is implied by the second (necessity). Suppose $\overline{U}_\theta(\underline{\theta}) \le -\psi_\theta(y(\underline{\theta});\underline{\theta})$. Then, the rent decreases in θ everywhere (see Lemmas 1 and 2) and $\underline{\theta} = \theta_k$. Similarly, suppose $\overline{U}_\theta(\overline{\theta}) \ge -\psi_\theta(y(\overline{\theta});\overline{\theta})$. The rent increases in θ and thus $\theta_k = \overline{\theta}$.

Using Lemma 1, integration by parts and substitution of the expected rent of the worker into the objective function of program [P] leads to

$$\max_{y(\theta);\theta_k} \int_{\underline{\theta}}^{\theta_k} V(y(\theta)) - \psi(y(\theta);\theta) - [\psi_\theta(y(\theta);\theta) + \overline{U}_\theta(\theta;l)]\frac{F(\theta)}{f(\theta)} - \overline{U}(\theta_k;l)\,dF(\theta)$$
$$+ \int_{\theta_k}^{\overline{\theta}} V(y(\theta)) - \psi(y(\theta);\theta) + [\psi_\theta(y(\theta);\theta) + \overline{U}_\theta(\theta;l)]\frac{1-F(\theta)}{f(\theta)} - \overline{U}(\theta_k;l)\,dF(\theta)$$

subject to the constraints

$$y(\theta) \le y(\theta_k) \quad \forall\theta \in [\underline{\theta};\theta_k) \tag{7}$$
$$y(\theta) \ge y(\theta_k) \quad \forall\theta \in (\theta_k;\overline{\theta}]: \tag{8}$$

These constraints are necesssary for $y'(\theta) \ge 0$ (Lemma 1), but not sufficient. However, it can be checked easily that the solution satisfies the stronger restriction of Lemma 1.

Pointwise maximization with respect to y yields[23]

$$V'(y) - \psi_y - \psi_{y\theta}\frac{F(\theta)}{f(\theta)} - \gamma = 0 \quad \forall\theta \in [\underline{\theta};\theta_k) \tag{9}$$

$$V'(y) - \psi_y + \psi_{y\theta}\frac{1-F(\theta)}{f(\theta)} + \delta = 0 \quad \forall\theta \in (\theta_k;\overline{\theta}] \tag{10}$$

[22] For reasons of clarity, the arguments $y(\theta)$ and l of the functions $\psi(y(\theta);\theta)$ and $\overline{U}(\theta;l)$ are suppressed in the proof.

[23] Throughout the paper, only local incentive compatibility constraints are checked. However, the corresponding global constraints are satisfied if $\psi_{\theta yy} = 0$ or relatively small and $\psi_{y\theta\theta} = 0$.

with γ and δ as the Lagrange multipliers. The first order conditions for the multiplier γ are

$$
\begin{aligned}
y(\theta_k)-y(\theta) &\geq 0 \\
\gamma[y(\theta_k)-y(\theta)] &= 0
\end{aligned}
$$

and for δ accordingly. With two distinct pooling intervals in $[\underline{\theta},\theta_k)$, it is necessary that $\gamma = 0$ in the first interval because output must be smaller than $y(\theta_k)$. By assumption, $\frac{F(\theta)}{f(\theta)}$ is increasing in θ. Pooling implies that $y'(\theta) = 0$ and we know that $\phi_{y\theta} < 0$ and $\phi_{y\theta\theta} = 0$. Thus, with pooling and $\gamma = 0$ expression (9) increases in θ, which leads to a contradiction. The same argument can be used to show that only one pooling interval exists in $(\theta_k,\overline{\theta}]$ because expression (10) increases in θ under a pooling contract and $\delta = 0$. Similarly, if there exist two different pooling intervals, one in $[\underline{\theta},\theta_k)$ and one in $(\theta_k,\overline{\theta}]$, at least one of the multipliers γ and δ is equal to zero, which contradicts (9) or (10) respectively.

Suppose there is a pooling interval with a level of output $\bar{y}$ greater than $y^*(\theta)$ for low values of θ and smaller than $y^*(\theta)$ for high values of θ. From (9) and (10) follows that $y(\theta) > y^*(\theta)$ for all θ with $\underline{\theta} < \theta < \theta_k$ and $y(\theta) < y^*(\theta)$ for all θ with $\overline{\theta} > \theta > \theta_k$. This can only hold if $\bar{y} = y^*(\theta_k)$.

Finally, if condition (6) holds and $y(\theta)$ is continuous, pooling must be optimal. Assume there is no pooling. Then it follows from (9) and (10) that $\lim_{\theta\to\theta_k^-} y(\theta) > \lim_{\theta\to\theta_k^+} y(\theta)$. If $y(\theta)$ is continuous, output must be decreasing in θ at some point, which violates the necessary condition for incentive compatibility that $y'(\theta) \geq 0$. $\square$

Proof of Lemmas 4, 5, and 6 Using Lemma 1 and integrating by parts, the worker's expected rent for all sections of the interval can be derived, for example

$$
\begin{aligned}
&\int_{\underline{\theta}}^{\theta_1}\Big[\int_{\theta}^{\theta_k} \phi_\theta + \overline{U}_\theta(\theta,l)d\theta \quad + \quad U(\theta_k)\Big]f(\theta)d\theta \\
= &\int_{\underline{\theta}}^{\theta_1} \phi_\theta + \overline{U}_\theta(\theta,l)\frac{F(\theta)}{f(\theta)}dF(\theta) \quad + \quad F(\theta_1)\int_{\theta_1}^{\theta_k} \phi_\theta + \overline{U}_\theta(\theta,l)d\theta \\
&\qquad\qquad\qquad\qquad\qquad\qquad + \quad U(\theta_k)F(\theta_1) \quad \forall\theta \in [\underline{\theta},\theta_1).
\end{aligned}
$$

Substitution of the expected rent payments into the objective function yields

$$
\int_{\underline{\theta}}^{\theta_1} V(y(\theta))-\phi(y(\theta),\theta) \quad - \quad [\phi_\theta(y(\theta),\theta) + \overline{U}_\theta(\theta,l)]\frac{F(\theta)}{f(\theta)} \qquad (11)
$$

$$
\begin{aligned}
&-\bar{U}(\theta_k;l)dF(\theta) \quad - \quad F(\theta_1)\int_{\theta_1}^{\theta_k} \psi_\theta(\bar{y};\theta) + \bar{U}_\theta(\theta;l)d\theta \\
&+\int_{\theta_1}^{\theta_k} V(\bar{y})-\psi(\bar{y};\theta) \quad - \quad [\psi_\theta(\bar{y};\theta) + \bar{U}_\theta(\theta;l)]\frac{F(\theta)-F(\theta_1)}{f(\theta)} \\
&\qquad - \quad \bar{U}(\theta_k;l)dF(\theta) \\
&+\int_{\theta_k}^{\theta_2} V(\bar{y})-\psi(\bar{y};\theta) \quad + \quad [\psi_\theta(\bar{y};\theta) + \bar{U}_\theta(\theta;l)]\frac{F(\theta_2)-F(\theta)}{f(\theta)} \\
&-\bar{U}(\theta_k;l)dF(\theta) \quad + \quad (1-F(\theta_2))\int_{\theta_k}^{\theta_2} \psi_\theta(\bar{y};\theta) + \bar{U}_\theta(\theta;l)d\theta \\
&+\int_{\theta_2}^{\bar{\theta}} V(y(\theta))-\psi(y(\theta);\theta) \quad + \quad [\psi_\theta(y(\theta);\theta) + \bar{U}_\theta(\theta;l)]\frac{1-F(\theta)}{f(\theta)} \\
&\qquad - \quad \bar{U}(\theta_k;l)dF(\theta).
\end{aligned}
$$

Note from Lemma 2 that $U(\theta_k;l) = \bar{U}(\theta_k;l)$.

Lemma 5 follows from the first order conditions for y:

$$
V'(y)-\psi_y-\psi_{\theta y}\frac{F(\theta)}{f(\theta)} = 0 \quad \forall\theta \in [\underline{\theta};\theta_1) \tag{12}
$$

and

$$
V'(y)-\psi_y + \psi_{\theta y}\frac{1-F(\theta)}{f(\theta)} = 0 \quad \forall\theta \in (\theta_2;\bar{\theta}]. \tag{13}
$$

For the highest and lowest type, $\bar{\theta}$ and $\underline{\theta}$, the optimality condition is reduced to $V_y-\psi_y = 0$. This results in first best output levels.

The cross derivative $\psi_{\theta y}$ is negative. Thus, output is greater than first best for all $\theta \in [\underline{\theta};\theta_1)$. Types with productivity $\theta \in (\theta_2;\bar{\theta}]$ produce less than the first best output. This proves Lemma 5.

Differentiating (11) with respect to θ_k and considering that $-\psi_\theta(\bar{y};\theta_k) = \bar{U}_\theta(\theta_k)$ leads to

$$
\begin{aligned}
&[\int_{\theta_1}^{\theta_k} V'(\bar{y})-\psi_{\bar{y}} \quad - \quad \psi_{\theta\bar{y}}\frac{F(\theta)}{f(\theta)}dF(\theta) \\
&+\int_{\theta_k}^{\theta_2} V'(\bar{y})-\psi_{\bar{y}} \quad + \quad \psi_{\theta\bar{y}}\frac{1-F(\theta)}{f(\theta)}dF(\theta)]\frac{\partial\bar{y}}{\partial\theta_k} \\
&\qquad - \quad \bar{U}_{\theta_k}(\theta_k;l) \quad \begin{matrix} < 0 \text{ for } \theta_1 = \underline{\theta}; \\ = 0 \text{ for } \theta_1 > \underline{\theta} \text{ and } \theta_2 < \bar{\theta}; \\ > 0 \text{ for } \theta_2 = \bar{\theta}. \end{matrix}
\end{aligned} \tag{14}
$$

This proves the third statement of Lemma 6.

Di¤erentiating (11) with respect to μ_1 leads to

$$[V(y(\mu_1)) - V(\bar{y}) \quad - \quad \acute{A}(y(\mu_1); \mu_1) + \acute{A}(\bar{y}; \mu_1)] f(\mu_1) \qquad (15)$$
$$+[-\acute{A}_\mu(y(\mu_1); \mu_1) \quad + \quad \acute{A}_\mu(\bar{y}; \mu_1)] F(\mu_1) \ \cdot \ 0:$$

If $\mu_1 > \underline{\mu}$, the FOC holds as an equality. Thus, for type μ_1 the pro…t of the employer must be equal independent of whether he chooses $\bar{y}$ or $y(\mu_1)$. If $\mu_1 = \underline{\mu}$, i.e. the pooling interval extends all the way to $\underline{\mu}$, a smaller μ_1 would be optimal, but it cannot be lowered further than $\underline{\mu}$.

An analogous condition can be derived for μ_2:

$$[V(\bar{y}) - V(y(\mu_2)) \quad - \quad \acute{A}(\bar{y}; \mu_2) + \acute{A}(y(\mu_2); \mu_2)] f(\mu_2) \qquad (16)$$
$$+[\acute{A}_\mu(\bar{y}; \mu_2) \quad - \quad \acute{A}_\mu(y(\mu_2); \mu_2)](1 - F(\mu_2)) \ \cdot \ 0:$$

Again, an interior solution with $\mu_2 < \overline{\mu}$ is characterized by the principal's indi¤erence between choosing $y(\mu_2)$ or $\bar{y}$. This proves the …rst two statements of Lemma 6.

Output $y(\mu)$ is a continuous function over $[\underline{\mu}; \mu_1)$ and $(\mu_2; \overline{\mu}]$ because the objective function is concave in y. The …rst order conditions (12) and (13) yield a unique output level for every type μ that maximizes the objective function. In addition, optimal output $y(\mu)$ is continuous over the pooling interval for types μ 2 $[\mu_1; \mu_2]$. The employer's pro…t function for type μ_1, $V(y(\mu_1)) - \acute{A}(y(\mu_1); \mu_1) - [\acute{A}_\mu(y(\mu_1); \mu_1) + \overline{U}_\mu(\mu; l)] \frac{F(\mu)}{f(\mu)} - \overline{U}(\mu_k; l)$ is concave in y. Thus, there exists a unique maximum $y(\mu_1)$. The same holds for μ_2 and for all other μ 2 $[\mu_1; \mu_2]$.

Now it is possible to show that Lemma 4 holds.

Suppose that $\overline{U}_\mu$ 2 $(-\acute{A}_\mu(y^{¤}(\underline{\mu}); \underline{\mu}); -\acute{A}_\mu(y^{¤}(\overline{\mu}); \overline{\mu}))$ for all μ 2 $[\underline{\mu}; \overline{\mu}]$ and $\mu_1 = \underline{\mu}$. Then it follows from (15) that $V(y(\mu_1)) - V(\bar{y}) - \acute{A}(y(\mu_1); \mu_1) + \acute{A}(\bar{y}; \mu_1) \ \cdot \ 0$. With $y(\underline{\mu}) = y^{¤}(\underline{\mu})$ it must hold for $\mu_1 = \underline{\mu}$ that $\bar{y} = y^{¤}(\underline{\mu})$ and $\acute{A}_\mu(\bar{y}) = \acute{A}_\mu(y^{¤}(\underline{\mu}))$. As $-\acute{A}_\mu(\bar{y}; \mu_k) = \overline{U}_\mu(\mu_k)$, it follows that $-\acute{A}_\mu(y^{¤}(\underline{\mu}); \underline{\mu}) = \overline{U}_\mu(\mu_k) > \overline{U}_\mu(\mu_2)$ which contradicts the second part of Lemma 4. An analogous argument can be made for μ_2. Therefore, if $\overline{U}_\mu$ 2 $(-\acute{A}_\mu(y^{¤}(\underline{\mu}); \underline{\mu}))$;-$\acute{A}_\mu(y^{¤}(\overline{\mu}); \overline{\mu}))$ for all μ 2 $[\underline{\mu}; \overline{\mu}]$, then the solution of the maximization problem has $\underline{\mu} < \mu_1$ and $\mu_2 < \overline{\mu}$.

Conversely, if $\overline{U}_\mu(\underline{\mu}) \ \cdot -\acute{A}_\mu(y^{¤}(\underline{\mu}); \underline{\mu})$ together with $\mu_1 > \underline{\mu}$, it must hold that $\overline{U}_\mu(\mu_k) =$-

$\varphi_\theta(\tilde{y};\theta_k) < \overline{U}_\theta(\underline{\theta}) \leq -\varphi_\theta(y^*(\underline{\theta});\underline{\theta})$. Because of $-\varphi_\theta(y^*(\underline{\theta});\underline{\theta}) = -\varphi_\theta(y(\underline{\theta});\underline{\theta})$, y is not increasing in θ everywhere if $-\varphi_\theta(y(\underline{\theta});\underline{\theta}) > -\varphi_\theta(y(\theta_k);\theta_k)$. Thus, it is not a feasible solution. Also, $\overline{U}_\theta(\overline{\theta}) \geq -\varphi_\theta(y^*(\overline{\theta});\overline{\theta})$ and $\theta_2 < \overline{\theta}$ do not characterize a feasible solution. Hence, $\theta_1 > \underline{\theta}$ and $\theta_2 < \overline{\theta}$ are necessary for
$\overline{U}_\theta \in (-\varphi_\theta(y^*(\underline{\theta});\underline{\theta}), -\varphi_\theta(y^*(\overline{\theta});\overline{\theta}))$ for all $\theta \in [\underline{\theta};\overline{\theta}]$. □

Proof of Proposition 1 Differentiating the objective function (11) with respect to l for the corresponding parameters θ_1, θ_k, and θ_2 (using the envelope theorem) yields the following expression:[24]

$$
\begin{aligned}
&\int_{\underline{\theta}}^{\theta_1} -\overline{U}_{\theta l}(\theta;l)\frac{F(\theta)}{f(\theta)} - \overline{U}_l(\theta_k;l)dF(\theta) \\
&+\int_{\theta_1}^{\theta_k} [V'(\tilde{y})-\varphi_{\tilde{y}}-\varphi_{\theta\tilde{y}}\frac{F(\theta)}{f(\theta)}]\frac{\partial\tilde{y}}{\partial l} - \overline{U}_{\theta l}(\theta;l)\frac{F(\theta)}{f(\theta)}-\overline{U}_l(\theta_k;l)dF(\theta) \\
&+\int_{\theta_k}^{\theta_2} [V'(\tilde{y})-\varphi_{\tilde{y}}+\varphi_{\theta\tilde{y}}\frac{1-F(\theta)}{f(\theta)}]\frac{\partial\tilde{y}}{\partial l} + \overline{U}_{\theta l}(\theta;l)\frac{1-F(\theta)}{f(\theta)}-\overline{U}_l(\theta_k;l)dF(\theta) \\
&+\int_{\theta_2}^{\overline{\theta}} \overline{U}_{\theta l}(\theta;l)\frac{1-F(\theta)}{f(\theta)} - \overline{U}_l(\theta_k;l)dF(\theta).
\end{aligned}
\tag{17}
$$

Without general skills, $l = 0$, the pooling interval $[\theta_1;\theta_2]$ collapses and $\underline{\theta} = \theta_1 = \theta_k = \theta_2$. Expression (17) becomes

$$\int_{\underline{\theta}}^{\overline{\theta}} \overline{U}_{\theta l}(\theta;0)\frac{1-F(\theta)}{f(\theta)}-\overline{U}_l(\underline{\theta};0)dF(\theta).$$

A sufficient condition for the optimal value of l to be greater than zero is

$$\int_{\underline{\theta}}^{\overline{\theta}} \overline{U}_{\theta l}(\theta;0)\frac{1-F(\theta)}{f(\theta)}dF(\theta) > \overline{U}_l(\underline{\theta};0).$$

□

Proof of Proposition 2 The standard adverse selection contract is reproduced by the model if $l = 0$. This leads to output distortions according to the function $y_L^0(\theta)$, which is defined by equation (13). Output $y(\theta)$ in the optimal contract with $l > 0$ is less distorted with respect to the efficient level than output $y_L^0(\theta)$ with $l = 0$ iff

$$\int_{\underline{\theta}}^{\overline{\theta}} y^*(\theta)-y_L^0(\theta)dF(\theta) > \int_{\underline{\theta}}^{\theta_k} y(\theta)-y^*(\theta)dF(\theta) + \int_{\theta_k}^{\overline{\theta}} y^*(\theta)-y(\theta)dF(\theta).$$

[24] The maximum value function is concave in l, which can be shown using standard methods.

As $y_L^0(\theta) = y(\theta)$ for all $\theta \in (\theta_2; \overline{\theta}]$, it is sufficient to show that the following three inequalities are satisfied:

$$\int_{\underline{\theta}}^{\theta_1} y(\theta) - y^*(\theta) dF(\theta) < \int_{\underline{\theta}}^{\theta_1} y^*(\theta) - y_L^0(\theta) dF(\theta); \quad (18)$$

$$\int_{\theta_1}^{\theta_k} \bar{y} - y^*(\theta) dF(\theta) < \int_{\theta_1}^{\theta_k} y^*(\theta) - y_L^0(\theta) dF(\theta); \quad (19)$$

$$\int_{\theta_k}^{\theta_2} y^*(\theta) - \bar{y} dF(\theta) < \int_{\theta_k}^{\theta_2} y^*(\theta) - y_L^0(\theta) dF(\theta): \quad (20)$$

As $\bar{y} = y^*(\theta_k) = y_L^0(\theta_2)$, condition (20) can be shown to hold by differentiating[25] the first order condition (13) for $y_L^0(\theta)$ with respect to θ:

$$y_L^{0'}(\theta) = \frac{\phi_{y\theta}}{V''(y) - \phi_{yy}}[2 + \frac{(1-F(\theta)) f'(\theta)}{(f(\theta))^2}] \quad \forall \theta \in [\underline{\theta}; \overline{\theta}]: \quad (21)$$

For the uniform distribution, the fraction in brackets on the right hand side disappears. Thus, $y_L^0(\theta)$ is increasing in θ everywhere and $\bar{y} > y_L^0(\theta)$ $\forall \theta \in [\theta_k; \theta_2)$. If the pooling interval extends to $\underline{\theta}$, then with uniformly distributed types $\theta_2 = \frac{1}{2}(\underline{\theta} + \overline{\theta})$.[26] Pooling cannot be optimal at a lower level than the bad type's efficient output. The pooling output $\bar{y}$ is closer to $y^*(\theta)$ than $y_L^0(\theta)$ for all $\theta \in [\underline{\theta}; \theta_2)$ because output does not decrease beyond $y_L^0(\theta_2) = \bar{y} = y^*(\underline{\theta})$.

Now take the first two conditions (18) and (19). Note that for the uniform distribution the slope of $y_L^0(\theta)$ is just twice as steep as the slope of $y^*(\theta)$ given by

$$y^{*'}(\theta) = \frac{\phi_{y\theta}}{V''(y) - \phi_{yy}} > 0 \quad \forall \theta \in [\underline{\theta}; \overline{\theta}] \quad (22)$$

The optimal output with $l > 0$ for all $\theta \in [\underline{\theta}; \theta_1)$, defined by equation (12), has the first derivative

$$y_H^{0'}(\theta) = \frac{\phi_{y\theta}}{V''(y) - \phi_{yy}}(2 - \frac{F(\theta) f'(\theta)}{(f(\theta))^2}) \quad \forall \theta \in [\underline{\theta}; \theta_1): \quad (23)$$

Again, for the uniform distribution the output function is twice as steep as the first–best output scheme. It is sufficient for condition (18) that the difference between $y(\theta)$ and $y^*(\theta)$ is smaller or at most equal to the difference between $y^*(\theta)$ and $y_L^0(\theta)$ for all

[25] Assume for simplicity that $V'''(y) = \phi_{yyy} = 0$ and $V''(y) \neq \phi_{yy}$. This ensures that the expressions in (21), (22), and (23) are well defined and, together with the assumptions of footnote 23, that all output schemes are linear for the uniform distribution.

[26] This can be shown with elementary geometry.

μ 2 $[\underline{\mu};\mu_1]$. Note that μ_1 cannot be to the right of $\frac{1}{2}(\underline{\mu}+\overline{\mu})$ because for $\mu_1 = \frac{1}{2}(\underline{\mu}+\overline{\mu})$ it must be that $y^0_H(\mu_1) = \bar{y} = y^{¤}(\overline{\mu})$. Only at $\mu_1 = \frac{1}{2}(\underline{\mu}+\overline{\mu})$, output distortions under both contracts are equal, $y(\mu_1)$-$y^{¤}(\mu_1) = y^{¤}(\mu_1)$-$y^0_L(\mu_1)$, but for all $\mu < \mu_1$ the strict inequality holds. Thus, condition (18) is satis…ed. From this follows immediately that condition (19) is satis…ed, too. If $\mu_1 < \frac{1}{2}(\underline{\mu}+\overline{\mu})$, moving from μ_1 to μ_k reduces the distance between $y(\mu)$ and $y^{¤}(\mu)$ to zero whereas the distance between $y^{¤}(\mu)$ and $y^0_L(\mu)$ decreases at a lower rate in this interval. If the pooling interval extends all the way to $\overline{\mu}$, output distortions are lower in the pooling interval than under a contract with $l = 0$. In particular if $\mu_1 = \frac{1}{2}(\underline{\mu}+\overline{\mu})$, the di¤erence between $\bar{y}$ and $y^{¤}(\mu)$ is smaller than the di¤erence between $y^0_H(\mu)$ and $y^{¤}(\mu)$ for all μ 2 $(\mu_1;\overline{\mu}]$ because output does not increase beyond $y^0_H(\mu_1)$.

2

References

[1] Altonji, J.G. (1995): "The e¤ects of high school curriculum on education and labor market outcomes", Journal of Human Resources, 30, pp. 409-483.

[2] Arulampalam, W., Booth, A.L. (1997): "Who gets over the training hurdle? A study of the training experiences of young men and women in Britain", Journal of Population Economics, 10, pp. 197-217.

[3] Baron, D., Myerson, R. (1982): " Regulating a monopolist with unknown costs", Econometrica, 50, pp. 911-930.

[4] Becker, G.S. (1962): Human Capital, Columbia University Press, New York.

[5] Bernhardt, D. (1995): "Strategic promotion and compensation", Review of Economic Studies, 62, pp. 315-339.

[6] Chiang, S.-H., Chiang, S.-C. (1990): "General human capital as a shared investment under asymmetric information", Canadian Journal of Economics, 23, pp. 175-188.

[7] Fama, E.F. (1980): "Agency problems and the theory of the …rm", Journal of Political Economy, 88, pp. 288-307.

[8] Gibbons, R. (1987): "Piece-rate incentive schemes", Journal of Labor Economics, 5, pp. 413-429.

[9] Guasch, J.L., Weiss, A. (1981): "Self-selection in the labor market", American Economic Review, 71, pp. 275-284.

[10] Harho¤, D., Kane, T.J. (1997): "Is the German apprenticeship system a panacea for the U.S. labor market?", Journal of Population Economics, 10, pp.171-196.

[11] Katz, E., Ziderman, A. (1990): "Investment in general training: The role of information and labour mobility", The Economic Journal, 100, pp. 1147-1158.

[12] La¤ont, J.-J., Tirole, J. (1986): "Using cost observation to regulate …rms", Journal of Political Economy, 94, pp. 614-641.

[13] Laffont, J.-J., Tirole, J. (1993): A Theory of Incentives in Procurement and Regulation, MIT Press, Cambridge, Massachusetts.

[14] Lewis, T.R., Sappington, D.E.M. (1989a): "Countervailing incentives in agency problems", Journal of Economic Theory, 49, pp. 294-313.

[15] Lewis, T.R., Sappington, D.E.M. (1989b): "Inflexible rules in incentive problems", American Economic Review, 79, pp. 69-84.

[16] Maggi, G., Rodríguez-Clare, A. (1995): "On countervailing incentives", Journal of Economic Theory, 66, pp. 238-263.

[17] McAfee, R.P., McMillan, J. (1987): "Competition for agency contracts", RAND Journal of Economics, 18, pp. 296-307.

[18] Mirrlees, J.A. (1971): "An exploration in the theory of optimum income taxation", Review of Economic Studies, 38, pp. 175-208.

[19] Moore, J. (1985): "Optimal labour contracts when workers have a variety of privately observed reservation wages", Review of Economic Studies, 52, pp. 37-67.

[20] Pischke, J.-S. (1998): " Why do firms train? Theory and evidence", Quarterly Journal of Economics, 113, pp. 79-119.

[21] Schlicht, E. (1996): "Endogenous on-the-job training with moral hazard", Labour Economics, 3, pp. 81-92.

[22] Stevens, M. (1994): "Labour contracts and efficiency in on-the-job-training", The Economic Journal, 104, pp. 408-419.

[23] v. Bardeleben, R., Beicht, U., Fehér, K. (1994): Kosten und Nutzen der betrieblichen Berufsausbildung. Zwischenergebnisse aus dem BIBB-Forschungsbericht, Bundesinstitut für Berufsbildung.

[24] Weiss, A. (1995): "Human capital vs. signalling explanations of wages", Journal of Economic Perspectives, 9, pp. 133-154.