General Training by Firms, Apprentice Contracts, and Public Policy*

James M. Malcomson
University of Oxford

James W. Maw
University of Swansea Wales

Barry McCormick
University of Southampton

First version: November 1997
This version: October 23, 2000

Abstract

Workers will not pay for general on-the-job training if contracts are not enforceable. Firms may if there are mobility frictions. Private information about worker productivities, however, prevents workers who quit receiving their marginal products elsewhere. Their new employers then receive external benefits from their training. Training firms increase profits by offering apprenticeships committing them to high wages for trainees retained on completion. At those wages, only good workers are retained, which signals their productivity and reduces the external benefits if they subsequently quit. Regulation of apprenticeship length (a historically important feature) can enhance efficiency, as can appropriate subsidies.

1 Introduction

Certain transferable skills are most efficiently acquired in the workplace and many policy makers view this general training as potentially crucial for enhancing labour productivity. However, there is little consensus about whether market provision of such general training is efficient.[1] Becker (1964) showed how, in a competitive economy without distortions, workers invest in, and firms supply, efficient levels of general training. His analysis provides an intellectual underpinning for the absence in the US of a national system of accreditation for post-high school training and was used to justify dismantling the UK system of trade apprenticeships.[2] Subsequent analysis of policy intervention has focused on the consequences of imperfect capital markets in which workers do not have access to funds to pay for general training. If, in a frictionless market, workers do not purchase transferable skills, firms will not supply them because trained workers can earn a wage equal to their trained marginal product at other firms and will quit unless this wage is matched by the training firm. The training firm will, as a result, be unable to recoup any return on the training provided. Consistent with this, Lynch (1992) finds that most on-the-job training in the US outside formal apprenticeships is firm-specific. Arguments of this type have prompted calls to facilitate workers' access to funds to pay for general training.

There are, however, reasons to think that workers' lack of access to capital to buy training is not the only reason for inefficient general training. The present paper explores other reasons, reasons that arise because training is a complex commodity provided by heterogeneous firms for heterogeneous trainees, and argues that these reasons are more consistent with the long history over several centuries of regulation of labour training provisions and the entitlements of employer and employee—see Earle (1989, ch.3). The complexity of training means that specifying the training to be provided in a way that is enforceable using a contract is problematic. We show that, as a result, trainees may not pay directly for general training even if they have the funds (or loans) to do so. Firms may pay for some general training if there are mobility frictions in skilled labour markets, as an extensive recent literature (discussed below) has shown. However, this leaves two crucial questions: what is the appropriate economic relationship between the training firm and its' trainees? What is the role of policy given that there is a potential for market failure in the provision of both training levels and the number of workers trained that a policy of providing loans will not overcome.

We analyse two forms of economic relationship between a general training firm and its trainees. In the first, a firm trains a worker and then set a wage schedule that just persuades that worker not to quit given other firms' imperfect information about the worker's productivity.

---

[1]See, for example, U.S. Department of Labor (1989) and Finegold and Soskice (1988).

[2]Thus, for example, Lees and Chiplin (1970) base their criticism of the UK grant-levy system to support training in the Industrial Training Act (1964) on the Becker human capital model. The system was subsequently abolished.

In the second, the firm again provides general training but offers a contract in which a worker accepts a low "apprentice" wage for a specified duration in return for a guaranteed high wage if retained after the end of the apprenticeship that ensures that the worker's productivity is revealed to the market. The apprentice contract delivers higher profits to training firms and more trained workers.

Apprenticeship is a contractual arrangement for general training that has been common in many countries since at least medieval times, see Pirenne (1936), and is still widely used in Germany. Common features of many apprenticeships are that they last for a duration specified at the start, not just until the apprentice has demonstrated satisfactory acquisition of the appropriate skills, and that the apprentice receives a substantial pay increase at the end. Apprenticeships have, moreover, been subject to extensive regulation by guilds and by governments. Particularly important historically has been regulation of their length. Regulation is regarded by some as improving the provision of training compared to what an unregulated market would provide, though policy differs considerably between countries—see, for example, Soskice (1994) and other chapters in Lynch (1994). This paper not only shows how apprenticeship contracts of this type can help resolve the incentive problems that non-contractibility of training introduces. It also shows that regulation to increase the apprenticeship length, coupled with a subsidy for each completed apprenticeship if the deadweight loss from raising taxes is not too high, can reduce the inefficiency of training provision. Importantly, these conclusions about apprenticeships apply even if training under an apprenticeship is no more effective at preventing trainees quitting than training without one. Thus, the difficulty of preventing premature quits by apprentices that historians such as Elbaum (1989) discuss in the context of the decline of apprenticeships in the US from the end of the 18th century is not sufficient reason to abandon their use even though, in our model, premature quitting will reduce both training levels and the numbers of apprentices trained.

The essential reasons for these conclusions are as follows. Because of the problems of specifying the training to be provided in a contract, workers will pay upfront only the expected value to them of the training that it is in firms' own interests to provide. As long as there are firms for which the cost of providing training is too high to be worthwhile, workers will not pay upfront for the full benefit they receive from the training by firms that actually train. With a large enough number of high cost training firms, they will not pay anything upfront.

While labour market frictions may result in firms providing some general training even if workers do not pay directly for it, there are two reasons recognised in the literature why the amount of such training may be inefficient. The first is that, when workers are free to quit without penalty, market frictions typically cannot prevent trainees using that freedom to capture some of the returns to training in the form of a higher wage even if they do not actually quit. Those returns are social returns to training that do not accrue to the training firm and so, as in the analysis of hold-up discussed extensively by Williamson (1985), firms invest too little

in training. To the extent that use of formal apprenticeship contracts reduces the ability of trainees to capture returns from training, it reduces the extent of hold-up. However, this role for apprenticeship contracts disappears if, as discussed by Elbaum (1989) for the US from the late 18th century on, apprenticeships cease to be effective in imposing additional penalties for premature quits.

The second reason why firm-supplied general training may be inefficient, recognised in the literature by Chang and Wang (1996), arises from asymmetric information about workers' productivities. When workers quit for jobs at other firms, that asymmetric information results in their wages being, on average, less than their marginal products. Thus part of the return to their general training goes to the new employers, an external benefit of training that is captured by neither trained workers nor training firms. Apprenticeship contracts have a role in mitigating this external effect. A contract commits at least one party to do something at some time in the future that it might not otherwise do when that time arrives. A practice well-documented for Germany, see Soskice (1994) and Acemoglu and Pischke (1998), is that training firms retain some trainees as skilled workers at the end of their apprenticeships. By committing in advance to a high wage for these retained trainees, a training firm ensures that it retains only the better workers with marginal product at least as great as that wage. Retention at the end of the apprenticeship then reveals information about the productivity of those retained and thus increases the wage of those who subsequently quit. That both reduces the externality to the workers' new employers and makes trainees willing to work for a lower wage during the training period. As a result, training firms make higher profits from training than they would if they made no commitment to future wages but merely determined them on a period by period basis. The shorter the contract length, the sooner a training firm reveals its information about a trainee's productivity and thus the sooner quitting workers capture the returns to their training that accrue after they quit, rather than those returns being an external benefit to their new employers. However, the shorter the contract length, the sooner retained workers receive a wage equal to their full trained marginal product and thus the less time the training firm has in which to recoup the costs of training. The profit maximizing contract length trades off these two effects. Although modelled here in the context of asymmetric information about workers' types, a similar phenomenon may arise with other market frictions that result in workers' current wages influencing (through, for example, bargaining as in Acemoglu (1997)) the wages they obtain when quitting. This role for apprenticeship contracts, unlike that of reducing hold-up, exists even when apprenticeships do not reduce the ability of trainees to capture returns from training.

Apprenticeship contracts cannot, however, in general prevent some of the return to general training being captured by trainees and, if they quit, by their new employers. As a result, even with apprenticeship contracts, trainees receive less training than is efficient. Moreover, when firms vary in the cost of providing training places, too few workers are trained. These two inefficiencies provide a natural role for regulation. We show that, by increasing the length of

the training contract, regulation can increase the amount of training towards the efficient level. That is consistent with the historical tradition of regulation of the length of apprenticeships. Of course, regulating the length of a contract reduces the profits from training and thus the number of firms that train. Even so, it is socially worthwhile. The essential reason is that, at the profit maximizing level of training, a small change in that level has only a second order effect on profit and, hence, only a second order effect on the number of workers trained. But, it has a first order effect on the amount of training received by each trainee and, because of the externalities, this has a first order effect on social welfare.

The adverse effect of regulation on the number of workers trained can be mitigated by a subsidy to firms for each completed apprenticeship. We show that, while it is worth providing some subsidy to increase the number of trainees if the deadweight loss of raising tax revenue is not too high, there are limits to what such subsidies can achieve. The difficulty of enforcing training by contract may, if the subsidy is large enough, result in firms using subsidized "trainees" simply as cheap labour without in fact providing training, an issue of serious concern with UK training policy, see Lee *et al.* (1990). That constrains the use of subsidies to enhance training, a constraint that can be important for policy.

A number of recent papers have studied mobility frictions that result in firms investing in some general training for their employees when workers do not pay for it. These papers fall into four broad categories. Katz and Ziderman (1990), Chang and Wang (1996), and Acemoglu and Pischke (1998) analyse asymmetric information between training firms and other potential employers. Stevens (1994a, 1996), Acemoglu (1996), Acemoglu (1997), and Booth and Chatterji (1997) consider imperfect competition in skilled labour markets. Burdett and Smith (1996) and Loewenstein and Spletzer (1998) study matching frictions. Booth and Chatterji (1995) and Acemoglu and Pischke (1997) discuss general training that is a joint product with specific training. Consistent with such frictions, Loewenstein and Spletzer (1998) conclude that firms in the US do indeed extract some of the returns to general training. The present paper differs from this literature in a number of important respects. First, it does not start from an *assumption* that workers do not pay directly for general training but derives that as a conclusion. Second, it shows how contracts, and in particular traditional apprenticeship contracts, are more profitable for training firms than simply determining wages optimally at each date in response to the quit behaviour of trainees. And third, it shows the role for regulation, and in particular the historically important regulation of apprenticeship length, in improving on the market provision of general training.

Of recent papers on training, only Cantor (1990) and Hermalin (1990) share our focus on contract length. Contract length in Cantor reflects a trade-off between the firm's preference for a long pay-back period and the mitigation of a moral hazard problem—workers' unobserved effort influences the effectiveness of training but has a value to the worker that diminishes with training contract length. Cantor's analysis is directed at firm-specific training, not general

training as here. Hermalin's concern with contract length is as a selection device in the face of worker adverse selection and the possible non-existence of equilibrium, again a very different issue from any discussed here. Neither paper analyses the role for policy.

The next section of the paper sets out the model of general training. Section 3 analyses the training provided in an unregulated market when contracts are not used. Section 4 shows that firms make greater profits by using an apprenticeship type arrangement. Regulation is discussed in Section 5. In Section 6 we discuss historical evidence for the approach adopted here.

2   The model

Firms can each employ one worker. All untrained workers have productivity w^0 per period of time. A worker can be trained on the job if the firm sets up a training programme at a fixed cost $k \in [0, \bar{k}]$. With training, productivity increases to $w^0 + \hat{\gamma} g(c)$, where $c \geq 0$ is the variable cost incurred in training, $g(c)$ is strictly concave with $g(0) = 0$ and $g'(0) = \infty$, and $\hat{\gamma} \in \{\gamma, \bar{\gamma}\}$ with $0 < \gamma < 1 < \bar{\gamma}$ is the realisation of a random variable determining worker type that is independently distributed across workers with mean 1. This realisation is unknown to firms and workers at the time of initial hiring and observed by only the training firm during training. The amount of training c is observed by trainees and by other firms once it has been provided but training is assumed to be too complex to specify its level in advance in an enforceable contract.

Firms differ only in the fixed cost k. Each knows its own fixed cost but this is unknown to anybody else and we assume that it is not possible for firms to establish a reputation for the value of k. The conditions on $g(c)$ ensure that some training is always worthwhile conditional on the fixed cost being incurred and so allow us to ignore the non-negativity constraint on c. There are $F(k)$ firms with fixed cost less than or equal to k, so the total number of *potential* jobs (and training places) is $F(\bar{k})$. We normalize the number of workers to 1 and assume both that $F(k)$ is twice continuously differentiable and that $F(\bar{k})$ is large (formally, $F(k) \to \infty$ as $k \to \bar{k}$). Thus workers can always get a job involving no training at wage w^0.

For simplicity, we assume firms and workers are risk neutral and discount the future at the same rate r. Workers enter the labour force at age 0 and have working lifetime T. Since job turnover is important in practice, we generate that in a relatively simple way by workers of both types deciding at the rate ρ that they wish to leave their current employer for a new one for exogenous reasons independent of pay. This can be viewed as a simple way of capturing that some turnover is efficient. The effective discount rate that a firm applies to the future profits from a particular worker is thus $r + \rho$. Firms are infinitely lived. Since we are interested in training contracts that last longer than required for the training itself, we lose little by assuming that training takes place instantaneously at the start of employment.

For what follows, it is convenient to treat time as continuous and to define

$$\beta(t) \equiv \frac{1 - e^{-rt}}{r} \tag{1}$$

$$\delta(t) \equiv \frac{1 - e^{-(r+\rho)t}}{r + \rho}. \tag{2}$$

The function $\beta(t)$ is the capitalisation factor for turning a constant flow over a period of length t into a present discounted value, $\delta(t)$ the effective capitalisation factor for firms whose workers quit at the rate ρ. For notational simplicity, let $\beta \equiv \beta(T)$ and $\delta \equiv \delta(T)$, the capitalisation factors for the whole working lifetime T.

Training is general and thus valuable even if a worker quits, so the expected lifetime value of the output of a worker trained to the level c is $\beta[w^0 + g(c)]$, that of an untrained worker βw^0. Conditional on a worker being trained, the efficient level of training maximizes the difference between the expected lifetime gain in output, $\beta g(c)$, and the cost. It is the solution to

$$\max_{c \geq 0} \beta g(c) - c, \tag{3}$$

that is, c^* uniquely defined by the first order condition

$$\beta g'(c^*) = 1. \tag{4}$$

It is, however, efficient to train only sufficient workers for the gains from training $\beta g(c^*) - c^*$ to exceed the fixed cost k. Define k_1^* by

$$k_1^* = \beta g(c^*) - c^*. \tag{5}$$

It is never efficient to train more than $F(k_1^*)$ workers. That may, however, be more workers than there are to train. Define k_2^* by $F(k_2^*) = 1$, the highest k required to train all workers. Then the efficient number of workers to train is such that the fixed cost of the highest cost firms that train is k^* defined by

$$k^* = \min\{k_1^*, k_2^*\}. \tag{6}$$

There are, however, three reasons in the model why achieving efficient training may be problematic. One is that a contract to provide a specified level of training is unenforceable. The second is that, if low wages during training or upfront payments (bonds) are used to get workers to pay for training, firms with high values of k may take on workers in order to get cheap labour or the upfront payment without having any intention of actually providing the training. Indeed, since such firms make zero profits from employing an untrained worker, they would always do better by agreeing to train a worker, taking the upfront payment, and then going out of business. With a large number of high training cost firms that would cheat on

any upfront payment by workers, workers are never prepared to offer such payments. While this conclusion follows straightforwardly from the assumptions about non-contractibility and heterogeneity of firms, we emphasise it here because of its important implications for general training.[3]

The third reason why achieving efficient training may be problematic is that workers, even when apprenticed, can quit at any time without having to compensate the training firm for doing so.[4] That limits the extent to which a firm can recoup costs of training by employing trained workers at a wage below their marginal product once training has been completed. To see the implications of this, consider the payoff to a trainee who contemplates quitting for an alternative firm. The alternative firm can observe how much training c a worker has actually received once it has taken place—the lack of enforceability arises because of the problem of describing the appropriate training in a contract beforehand. It does not, however, know the trainee's type. By the usual adverse selection argument (see, for example, Greenwald (1986)), the training firm, which knows the trainee's type, matches any outside offer no greater than the trainee's productivity. Thus the expected profit to an alternative firm from any offer above the productivity of a less productive worker (with $\hat{\gamma} = \gamma$) is negative unless that worker is quitting for exogenous reasons—the *winner's curse*. With continuous hiring, however, the flow of workers quitting for exogenous reasons is negligible relative to the stock of less productive trainees. (Formally, the flow is of order dt relative to the stock.) Thus, if there is even a small cost (in addition to the wage cost) of hiring a new worker, it is never profitable for an alternative firm to offer a wage higher than the training firm would pay to retain a less productive trainee. It is, however, worth offering a wage equal to the productivity of a less good trainee, less the flow value of the hiring cost, because that may attract workers quitting for exogenous reasons, some of whom are the more productive type. Competition between alternative employers ensures that they offer exactly that wage. The formal result is in Proposition 6 in Appendix A. Given the practical importance of hiring costs, we treat the model without hiring costs studied here as the limiting case in which the hiring cost goes to zero. Then the wage available to a trainee who quits approaches the productivity of a less productive type $w^0 + \gamma g(c)$ and the present

---

[3]If workers know each firm's fixed costs of training (for example, when fixed costs are the same for all firms as in Acemoglu and Pischke (1998)), they know which firms will actually train and so, even with the amount of training non-contractible, are prepared to pay upfront for the full benefit they receive from training by those firms. In that case, an appropriately chosen contract can overcome the hold-up problem. If fixed costs of training are unknown to workers but there are only a finite number of high training cost firms, workers are prepared to pay upfront an amount that depends inversely on the probability that a randomly selected firm will fail to train. Thus this variation of the model is consistent with payment of initial fees for apprenticeships that were historically common in England, see Hamilton (1996). Incorporating this variation into the analysis complicates the exposition because the maximum upfront fee is endogenous but does not alter the result below that an unregulated market provides less than the efficient amount of training, see footnote 6 below. Thus, the basic conclusions of the paper apply even where limited upfront fees are paid.

[4]Hamilton (1995) argues that relatively few apprentices in her sample from Montreal actually deserted. That is, of course, entirely consistent with training firms designing contracts to induce them not to do so. Her sample provides substantial evidence of contracts being drawn up with precisely this purpose in mind.

discounted value of the remaining lifetime earnings of a trainee who quits at $t \leq T$ becomes

$$\beta (T - t) \left[w^0 + \gamma g (c) \right], \text{ for } 0 \leq t \leq T. \tag{7}$$

(It applies from $t = 0$ because we have assumed training is instantaneous.) A training firm must thus ensure that trainees have expected future utility at least as great as that in (7) if they are not to quit.

3   Training without contracts

Suppose training firms do not offer training contracts but merely hire trainees, train them, and pay them a wage high enough to induce them not to quit. Let $w(t)$ denote the wage paid by the firm at t. Then the present discounted value of expected future wages received from the firm is $W(t, T)$ given by

$$W (t, T) = \int_t^T w (\theta) e^{-(r + \rho)(\theta - t)} d\theta. \tag{8}$$

In addition to this, the worker receives the wage $w^0 + \gamma g (c)$ at any date subsequent to quitting for exogenous reasons. By the standard calculation, the probability a worker employed at t has quit for exogenous reasons by θ is $1 - e^{-\rho(\theta - t)}$. Thus the expected future utility of a worker who does not quit at t is

$$W (t, T) + \left[w^0 + \gamma g (c) \right] \int_t^T e^{-r(\theta - t)} \left[1 - e^{-\rho(\theta - t)} \right] d\theta$$
$$= W (t, T) + \left[w^0 + \gamma g (c) \right] \left[\beta (T - t) - \delta (T - t) \right]. \tag{9}$$

For the worker not to quit for other than exogenous reasons, this expected future utility must be at least as great as that in (7) for all t, which implies

$$W (t, T) \geq \delta (T - t) \left[w^0 + \gamma g (c) \right], \text{ for } 0 \leq t \leq T. \tag{10}$$

The expected profit $\Pi (k, c, W, T, S)$ of a firm with fixed training cost k from training a worker to the level c when the worker quits only for exogenous reasons is its expected revenue $\delta \left[w^0 + g (c) \right]$, less its training costs $k + c$ and its expected wage costs denoted by W, plus the expected present value S of any subsidy received for training paid at date T. Such subsidies play a role in our discussion of regulation in Section 5.[5] That is,

$$\Pi (k, c, W, T, S) = \delta \left[w^0 + g (c) \right] - (k + c) - W + S. \tag{11}$$

[5]The government could offer a subsidy to take on a trainee that is paid at the start of training but such a subsidy provides no advantage and exacerbates the gains to firms from promising to train and taking the subsidy but not actually training.

We consider here the case in which it is not efficient to train all employees entering the market, that is $k_1^* < k_2^*$. Stevens (1994b, pp. 561-2) discusses the evidence for this being the relevant case for UK engineering during 1966-88. Other cases are discussed in Appendix C. When $k_1^* < k_2^*$, an unregulated market will never train all workers entering the workforce. Thus, there will always be more employees entering the market than will be trained. Moreover, because of the finite lifetimes of workers, it is never efficient to start training workers who are not new entrants to the market because they have a shorter remaining working lifetime to generate returns on that training. New entrants not starting training straightaway, therefore, remain untrained for ever and earn market wage w^0 for their remaining working lifetime. Untrained workers who quit for exogenous reasons can always get another job at the wage w^0. Thus the present discounted utility of a worker not successful in obtaining a training place on entering the market is βw^0 and the lifetime value of wages that equates the demand for workers with the supply of workers, denoted $\bar{W}$, is also βw^0. Thus to hire a trainee, a training firm must ensure that the utility from taking a training place, given by (9) for $t = 0$, is no less than βw^0, which implies

$$W\,(0,\,T) \geq \delta\left[w^0 + \gamma\,g\,(c)\right] - \beta\gamma\,g\,(c)\,. \tag{12}$$

But to prevent a trainee quitting immediately after being trained, the present discounted value of wages offered by the firm must satisfy the higher value given by (10) for $t = 0$. Thus (10) for $t = 0$ is a binding constraint on the lifetime wages that firms must pay to retain trainees. A wage of $w^0 + \gamma\,g\,(c)$ at each t satisfies this constraint with equality both for $t = 0$ and for all subsequent t so, with this wage, trained employees quit for an alternative employer only for exogenous reasons. Substitution of this wage into (11) gives expected profit from training without a contract when there is no subsidy

$$\begin{aligned}\Pi^0\,(k,\,c) \;&=\; \delta\left[w^0 + g\,(c)\right] - (k + c) - \delta\left[w^0 + \gamma\,g\,(c)\right] &\tag{13}\\ &=\; (1 - \gamma)\,\delta g\,(c) - (k + c)\,. &\tag{14}\end{aligned}$$

Proposition 1 *(i) The level of training c^0 that maximises the expected profit of a training firm in the absence of a contract is given by*

$$\delta g'(c^0) = \frac{1}{1 - \gamma}\,. \tag{15}$$

Moreover, $c^0 < c^*$, the efficient level of training, and decreases with ρ, the exogenous quit rate, and with γ. (ii) Training is carried out by only those firms with $k \leq k^0$ defined by

$$k^0 = (1 - \gamma)\,\delta g\left(c^0\right) - c^0\,. \tag{16}$$

Moreover, $k^0 < k_1^*$ defined in (5), so fewer workers are trained than is efficient.

Proof. (i) It follows from (14) that, because $g(c)$ is differentiable and strictly concave, $\Pi^0(k, c)$ is a differentiable and strictly concave function of c. Thus the level of training that maximises expected profit is given uniquely by the first-order condition that corresponds to (15). Given $g(c)$ strictly concave, that $c^0 < c^*$ follows directly from comparison of (15) with (4) and $0 < \gamma < 1$, that c^0 is decreasing in ρ follows from (15) and the definition of δ in (2), and that c^0 is decreasing in γ follows from (15). (ii) It follows from (14) that, k^0 defined in (16) is the highest fixed cost of training for which $\Pi^0(k, c^0) \geq 0$, so only firms with $k \leq k^0$ train. Note that c^* maximises $\beta g(c) - c$, whereas c^0 maximises $(1 - \gamma) \delta g(c) - c$, and that $\beta > (1 - \gamma) \delta$. It thus follows directly from comparison of (16) with (5) that $k^0 < k_1^*$. ∎

The intuition for these results is as follows. The level of training c^0 is less than c^*, the efficient level of training defined in (4) both because $\gamma > 0$ and because $\delta < \beta$. As long as $\gamma > 0$, the wage the worker can get by quitting increases with the amount of training the firm provides. Thus, the worker captures part of the return on the firm's investment, a form of Williamson's hold-up, and the firm under invests in training. In the extreme case as γ approaches 1, the firm is unable to earn any return on its investment in training and thus the amount of training approaches zero. At the other extreme, as γ approaches 0, the amount of training does not, however, approach the efficient level c^* because $\delta < \beta$. The reason for $\delta < \beta$, as is clear from (1) and (2), is that $\rho > 0$, that is, some employees quit for exogenous reasons. Because the training is general, it is valuable even when workers quit and so still generates a social return. However, the training firm does not capture any of the return arising after a worker quits for exogenous reasons. That return is received by the worker's new employer, who acquires an employee with expected productivity $w^0 + g(c)$ at a wage $w^0 + \gamma g(c)$. Only if both γ and ρ are zero do training firms invest efficiently. Moreover, the number of workers trained is then also efficient because the highest fixed cost that firms will incur in order to train is k^* defined in (6).[6]

---

[6]In the text, we have treated the case in which no worker makes an upfront payment because of the high probability that a randomly chosen firm will default on training by going out of business. (Formally, that probability approaches 1 because we have assumed $F(k) \to \infty$ as $k \to \bar{k}$.) If, however, $F(\bar{k})$ is finite, the probability that a randomly chosen firm trains is $F(k)/F(\bar{k})$ when k is the highest fixed cost at which it is profitable to train. In the absence of upfront payments, a worker receives a gain from joining a firm that trains of $\beta \gamma g(c)$, the difference between the expression in (7) at $t = 0$ and βw^0. The maximum upfront payment that would be made by a worker not knowing a firm's type, and so whether it will actually train, is thus $B = \beta \gamma g(c) F(k)/F(\bar{k})$. To a training firm, this payment is a lump sum independent of the training it actually chooses. (This is where the analysis differs crucially from the case where the level of training is contractible.) Thus training firms choose the same amount of training c^0 as without an upfront payment, which we already know to be below the efficient level c^*. The upfront payment, however, increases the highest fixed cost for which it is worth training to k^B given by

$$k^B = (1 - \gamma) \delta g\left(c^0\right) - c^0 + \frac{F\left(k^B\right)}{F(\bar{k})} \beta \gamma g\left(c^0\right),$$

which is greater than k^0. However, $k^B < k_1^* \equiv \beta g(c^*) - c^*$ because, by definition, c^* maximizes $\beta g(c) - c$, so $\beta g(c^*) - c^* > \beta g(c^0) - c^0$, and $\beta > (1 - \gamma) \delta + \gamma \beta F(k^B)/F(\bar{k})$. Thus the number of workers trained is also still below the efficient level.

10

4 Training contracts

In the analysis of the preceding section, training firms do not offer training contracts. They merely hire trainees, train them, and pay them a wage high enough to induce them not to quit. Alternatively, training firms may offer an apprenticeship contract with the following characteristics even if (by, for example, reducing the payoff to quitting below that in (7)) the contract does not itself reduce the ability of trainees to capture returns from training. Trainees undertake to work for a wage lower than if they were hired without a contract for a specified period in return for an assurance of a higher (skilled) wage if the firm continues to employ them after that period. The higher wage is set so that the training firm offers to continue the employment of more able trainees but not of less able ones. The information contained in that offer ensures that more able trained workers can earn their marginal product with any other employer and hence, by competition, with their training firm. A training firm's incentives to train are then unaffected if a trained worker quits after this stage because that worker's wage equals the marginal product. We shall show that this contract is more profitable, and that as a result more workers are trained.

In analysing the implications of such a contract, we again consider the case in which it is not efficient to train all workers ($k_1^* < k_2^*$). Formally, the training contract specifies a period of length τ (which we call *contract length*) after which the training firm is committed, if it subsequently employs the trained worker, to pay a specified wage w strictly between the productivity of a worker of type γ trained to whatever level c the firm wishes to commit itself to provide, namely $w^0 + \gamma g(c)$, and the productivity of a worker of type $\bar{\gamma}$ trained to that level, namely $w^0 + \bar{\gamma} g(c)$. The firm is not *required* to employ the worker after τ—it merely commits itself to the wage w if it does so. The contract also specifies $w(t)$, for $0 \leq t \leq \tau$, the wage paid at time $t \leq \tau$ after its start.[7]

The effect of this contract on the wage of workers trained to a given level c is illustrated in Figure 1. With no contract as in the previous section, the winner's curse ensures all workers earn wage $w^0 + \gamma g(c)$ throughout their lifetime T, just enough to induce them not to quit. With a contract of length τ, more able workers are retained at the end of the contract and earn their marginal product $w^0 + \bar{\gamma} g(c)$ from τ on. Less able workers are not retained and obtain another job with wage equal to their marginal product $w^0 + \gamma g(c)$ from τ on. Because workers do not know their own type until it is revealed by the firm, their expected wage from τ on if they stay until the end of their contract is greater than $w^0 + \gamma g(c)$, so the wage $w(t)$ during the period up to τ can be reduced below $w^0 + \gamma g(c)$ without inducing them to quit. Thus, by using a contract that makes it credible to reveal workers' types at the end of the contract period,

---

[7]This contract is similar to the "up-or-out" contracts analysed in Kahn and Huberman (1988), Waldman (1990) and Prendergast (1993) for specific investments. Other important differences are that here the contract induces firms, rather than employees, to invest in training and the date at which the "up or out" decision is made (τ) is agreed as part of the contract, not given exogenously, which enables us to study the role of contract length.
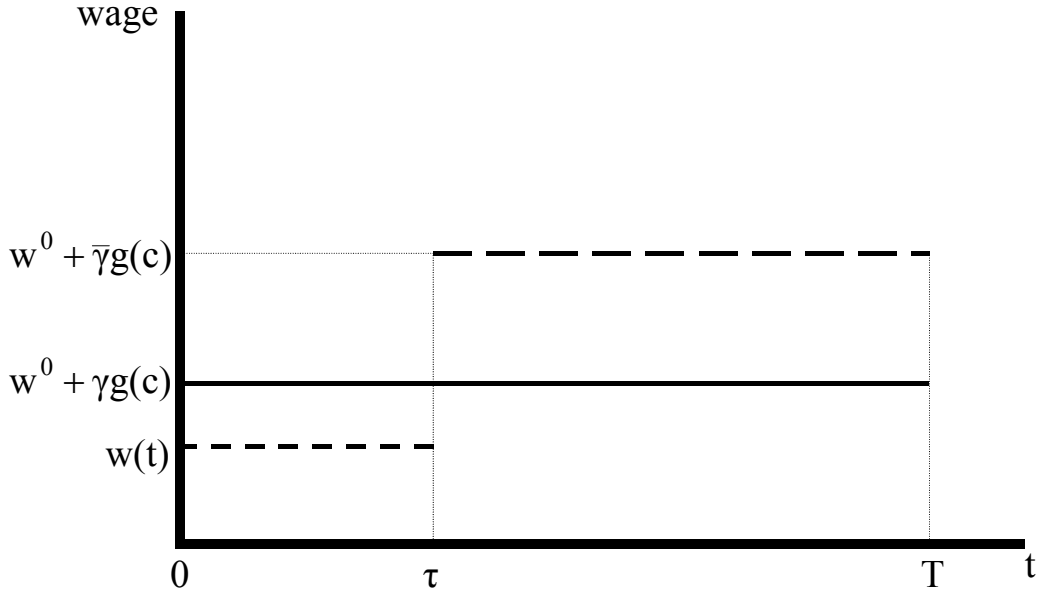
Figure 1: Effect of training contract on wages

a training firm induces trainees to accept lower wages during the contract period.

4.1 Incentive compatibility for training firms

Revelation will, of course, occur only if the amount of training actually provided by the firm makes it profitable to retain more able trained workers and unprofitable to retain less able ones. There are three possibilities to consider. The first is that the training c actually provided is such that $w^0 + \gamma g(c) \geq w$. In this case, the wage w is no higher than the wage $w^0 + \gamma g(c)$ a trained worker can obtain elsewhere, so the firm has to pay $w^0 + \gamma g(c)$ to retain the worker. This reduces to the case of the previous section for which we already know it is optimal for the firm to set $c = c^0$ and receive profit $\Pi^0(k, c^0)$. As long as the firm can make greater profits by setting w at some level such that it actually trains to $c \neq c^0$, it is optimal for the firm to offer a training contract.

The second possibility is that the training c actually provided by the firm is such that $w^0 + \gamma g(c) < w < w^0 + \bar{\gamma} g(c)$. We show in Appendix B that the training wage can be selected so that the firm wishes to retain even a less good trainee (of type γ) during the period up to date τ. Thus, during that period the training firm receives the expected output of the trained worker, $\delta(\tau)[w^0 + g(c)]$. On completion of the contract, a trained worker is either retained by the training firm at wage w or seeks a job in a firm with a vacancy. It is never in the firm's interest to retain a less good worker (type γ) since such a worker has productivity less than w. It is always in the training firm's interest to retain a better worker (type $\bar{\gamma}$) since such a worker has

12

productivity greater than w.[8] By retaining a worker, therefore, the firm signals that the worker is a good type. Outside firms will then be prepared to offer such a worker a wage of $w^0 + \bar{\gamma} g(c)$ and the firm will have to match that offer to retain the worker. The firm's profits from training are zero from τ on and are given by (11) with δ replaced by $\delta(\tau)$[9]

$$\Pi(k, c, W, \tau, S) \equiv \delta(\tau)\left[w^0 + g(c)\right] - (k + c) - W + S, \tag{17}$$

where W is the total expected wage payment up to τ and S is the present value of any subsidy to training, now assumed to be paid at the end of the contract period.

The third possibility is that the training c actually provided by the firm is such that $w^0 + \bar{\gamma} g(c) \leq w$. In this case it is not profitable to retain a less productive trainee (type γ) and, even if the firm retains a more productive trainee (type $\bar{\gamma}$), it cannot make positive profits from doing this, so its profits from training are the same as in (17). Thus the analysis for the previous case applies to this case too.[10]

Lemma 1 *With an apprentice contract that specifies τ, w and W, the optimal training level c for a firm that trains is given by*

$$\delta(\tau) g'(c) = 1. \tag{18}$$

The optimal apprentice training level is increasing in τ.

Proof. Given a training contract that specifies τ, w and W, the firm chooses the amount of training c to maximize $\Pi(k, c, W, \tau, S)$ given by (17).[11] It follows from (17) that, because $g(c)$ is differentiable and strictly concave, $\Pi(k, c, W, \tau, S)$ is a differentiable and strictly concave function of c. Thus the level of training that maximises expected profit is given uniquely by the first-order condition that corresponds to (18). Moreover, from (2), $\delta(\tau)$ is increasing in τ. Thus, given $g(c)$ strictly concave, it follows from (18) that training c is an increasing function of contract length τ. ∎

---

[8]Many trainees are in fact retained. For Germany, Acemoglu and Pischke (1998) report that 84% of apprentices stay on, at least initially, with their training firm. Soskice (1994, p. 56) argues that "the postapprentices which a nontraining company will be likely to hire are those whom their training companies have chosen not to keep."

[9]A training firm receives profits from retaining a good worker for the period until outside firms make that worker an offer of $w^0 + \bar{\gamma} g(\tilde{c})$. Formally, we consider the limit as this period goes to zero. If the period is longer, the optimal contract length is reduced and the training firm will, as Soskice (1994) argues for Germany, continue to make profits from retaining the better trained workers after the end of their training contract. Competition between firms in initial hiring will then ensure that firms do not recoup all the training costs during the training period, again as Soskice (1994) argues for Germany. We do not pursue this issue here.

[10]Because the wage available from other firms is $w^0 + \gamma g(c)$, there is potential for a profitable renegotiation between the training firm and a more productive trainee to reduce the wage to a level at which both gain by having the firm retain the employee. The trainee can, however, ensure that the renegotiated wage is strictly above $w^0 + \gamma g(c)$ because, for $w^0 + \gamma g(c)$, a threat to quit is credible. Moreover, once the firm agrees to a wage strictly above $w^0 + \gamma g(c)$, it reveals the worker to be of the more productive type so that, by the argument for the previous case, outside firms bid the wage up to $w^0 + \bar{\gamma} g(c)$. Thus the firm still makes no profits from retaining the trainee.

[11]Strictly, this maximisation is subject to the inequality constraint that c is not so large that, given W, the trainee quits. The choice of optimal contract, however, ensures that W is always such that this constraint does not bind.

The intuition for this result is simply that the longer τ, the longer is the period for which the firm receives returns to training and thus the more training it provides. The contract cannot, however, be longer than $\tau = T$, so the highest level of training that can be induced in this way is $\bar{c}$ defined by

$$\delta g'(\bar{c}) = 1. \tag{19}$$

The firm can therefore be induced to provide any amount of training $c \in [0, \bar{c}]$ by a contract of length $\tau(c)$ that satisfies (18) for that c. The definition of $\delta(\tau)$ in (2) can be used to write

$$\tau(c) \equiv -\frac{1}{r+\rho} \ln\left[1 - \frac{r+\rho}{g'(c)}\right], \quad \text{for } c \in [0, \bar{c}]. \tag{20}$$

4.2 Incentive compatibility for trainees

We next determine the constraints imposed by workers' incentives to join, and remain in, an apprenticeship. For this, we derive the implied relationship between the amount of training, c, and the total wages paid during the contract period, W. The expected remaining lifetime utility at $t < \tau$ from a trainee staying with the training firm can be derived as follows. Let $W(t, \tau)$ denote the expected present discounted value of wages over the remainder of the contract. Given training to level c, a worker who quits for exogenous reasons before the end of the contract receives wage $w^0 + \gamma g(c)$ from the date of quitting. A worker who has not quit for exogenous reasons before the end of the contract receives the wage $w^0 + \bar{\gamma} g(c)$ from τ on if of type $\bar{\gamma}$ and $w^0 + \gamma g(c)$ if of type γ, with expected value $w^0 + g(c)$. As in the case without a contract, the probability a worker employed at t has quit for exogenous reasons by θ is $1 - e^{-\rho(\theta-t)}$. The expected value of all these components for $\tau \le T$ is

$$W(t, \tau) + \left[w^0 + \gamma g(c)\right] \int_t^\tau e^{-r(\theta-t)} \left[1 - e^{-\rho(\theta-t)}\right] d\theta$$

$$+ \left\{\left[w^0 + \gamma g(c)\right]\left[1 - e^{-\rho(\tau-t)}\right] + \left[w^0 + g(c)\right] e^{-\rho(\tau-t)}\right\} \int_\tau^T e^{-r(\theta-t)} d\theta$$

$$= W(t, \tau) + \left[w^0 + \gamma g(c)\right]\left[\beta(\tau-t) - \delta(\tau-t)\right]$$

$$+ \left\{\left[w^0 + \gamma g(c)\right] + (1-\gamma) g(c) e^{-\rho(\tau-t)}\right\}\left[\beta(T-t) - \beta(\tau-t)\right] \tag{21}$$

$$= W(t, \tau) + \left[w^0 + \gamma g(c)\right]\left[\beta(T-t) - \delta(\tau-t)\right]$$

$$+ (1-\gamma) g(c) e^{-\rho(\tau-t)}\left[\beta(T-t) - \beta(\tau-t)\right]. \tag{22}$$

To induce a trainee to complete the contract, therefore, the expected value of wages paid during the contract period must ensure that this expression is at least as great as that in (7) for all $0 \le t < \tau$, which gives the following incentive compatibility condition for a worker to complete

14

the contract:

$$W(t, \tau) \geq \delta(\tau - t)\left[w^0 + \gamma g(c)\right]$$
$$- (1 - \gamma) g(c) e^{-\rho(\tau - t)}\left[\beta(T - t) - \beta(\tau - t)\right], \text{ for all } 0 \leq t < \tau. \quad (23)$$

Fortunately (23) has an intuitive interpretation: the apprentice expects to be paid a sum equal to the discounted value of potential outside earnings up to the end of the apprenticeship less the present value of the increase in earnings from completing the apprenticeship. When it is not efficient to train all workers, (23) is a binding constraint at $t = 0$ on wages during the contract because the expected utility from quitting immediately after being trained, given by (7) for $t = 0$, is strictly greater than βw^0. The lowest present discounted value of wages W for which this constraint is satisfied at $t = 0$ for a contract of length $\tau(c)$ is

$$W = \delta[\tau(c)]\left[w^0 + \gamma g(c)\right] - (1 - \gamma) g(c) e^{-\rho\tau(c)}\{\beta - \beta[\tau(c)]\}. \quad (24)$$

Of course, because potential trainees do not know a firm's fixed cost of training, they cannot rule out the possibility that a firm with a high fixed cost of training will agree to a training contract without intending actually to train if trainee wages are low enough for it to be profitable to do so. Because, however, workers can observe whether they have in fact been trained, such behaviour can be prevented by an initial training wage greater than w^0. The reason is as follows. An untrained worker can always quit to earn w^0 and will do so, thus ensuring the firm makes a loss from offering the training contract, unless the remainder of the training contract provides utility higher than provided by the wage w^0. But if the remainder of the training contract provides that higher utility, a firm that promises to train but does not do so will make a loss from a training contract if the initial wage is greater than w^0.[12] We show in Appendix B that for any W satisfying (24), there exists a wage path $w(t)$ with $w(0) > w^0$ that ensures (23) is satisfied for all $0 \leq t < \tau(c)$ without inducing the firm to dismiss a less good trainee before $\tau(c)$.

---

[12]If trainees do not observe whether they have actually been trained, they can be sure that a firm will not agree to a training contract unless it intends to train only if expected wages over the whole training contract are no less than the untrained wage, that is, $W \geq \delta[\tau(c)]w^0$. The reason is as follows. Untrained workers will not be kept on at the end of the training contract and will be paid only w^0 in the market thereafter. Because there are lots of high cost training firms which will never train (formally, $F(\bar{k}) \to \infty$), the probability of being trained in a randomly chosen firm is negligible unless there is self selection of firms. Thus, one of two conditions needs to be satisfied for potential trainees to agree to a contract. One of these is that trainees receive higher utility even if the firm does not train than they would by not taking on a training contract. The other is that a firm makes less profit by signing a training contract if it does not in fact intend to train than if it simply paid the wage w^0. Both of these are equivalent to the condition specified. Substitution for W from (24) allows that condition to be written as the constraint $\gamma \delta[\tau(c)] - (1 - \gamma) e^{-\rho\tau(c)}\left[\beta - \beta(\tau(c))\right] \geq 0$. In this case, it turns out that the additional constraint is always binding with the optimum contract, which implies an optimal contract length that is independent of the function $g(c)$. The length of the apprenticeship is then the same for different trades for which γ, ρ and r are the same even if the function $g(c)$ is different.

4.3 Optimal training contracts

We next determine the optimal training contract. Substitution from (24) into (17) gives the expected profit from choosing τ such that training is to level c

$$\tilde{\Pi}(k, c, S) = (1 - \gamma) g(c) \left\{ \delta[\tau(c)] + e^{-\rho\tau(c)} [\beta - \beta(\tau(c))] \right\} - (k + c) + S. \quad (25)$$

To compare the outcomes with and without training contracts, it is convenient to define

$$\tilde{\delta}(\tau) \equiv \delta(\tau) + e^{-\rho\tau} [\beta - \beta(\tau)]. \quad (26)$$

This corresponds to an adjusted discount factor that can be interpreted in the following way. The actual discount rate of both firm and worker is r, corresponding to the discount factor $\beta(\tau)$. Trainees quit at the rate ρ. If a trainee quits during the contract period, the firm receives no more return on its investment, so the effective discount rate during the contract period becomes $r + \rho$, corresponding to the discount factor $\delta(\tau)$. Once the contract period has ended, however, workers receive all the returns to training whether or not they quit, so the quit rate ρ no longer influences the effective discount rate after the end of the contract. Moreover, these returns are passed back to firms in the form of lower wages during the contract period. The present discounted value of an increase in the post-contract wage that gets passed back to the firm is $\beta - \beta(\tau)$, which increases with the length of the post-contract period. But this accrues only if the trainee does not quit before the end of the contract period, so it is multiplied by the probability $e^{-\rho\tau}$. The discount factor $\tilde{\delta}(\tau)$ adjusts $\delta(\tau)$ for this. From the definitions of $\beta(t)$ and $\delta(t)$ in (1) and (2), note that $\tilde{\delta}(0) = \beta$ and $\tilde{\delta}(T) = \delta$. Thus, if the contract lasts the whole lifetime T so there is no post contract period, the effective discount factor is δ whereas, if the contract has zero length so the post contract period consistent of the whole working lifetime, the effective discount rate becomes β. Moreover,

$$\begin{aligned}
\tilde{\delta}'(\tau) &= \delta'(\tau) - \rho e^{-\rho\tau} [\beta - \beta(\tau)] - e^{-\rho\tau} \beta'(\tau) \\
&= -\rho e^{-\rho\tau} [\beta - \beta(\tau)] \quad (27) \\
&< 0, \text{ for } \tau < T.
\end{aligned}$$

We thus have

$$\begin{aligned}
\tilde{\delta}(0) &= \beta \\
\beta > \tilde{\delta}(\tau) &> \delta, \text{ for } 0 < \tau < T \quad (28) \\
\tilde{\delta}(T) &= \delta.
\end{aligned}$$

Profit $\tilde{\Pi}(k, c, S)$ defined in (25) can then be written in terms of $\tilde{\delta}[\tau(c)]$ as

$$\tilde{\Pi}(k, c, S) = (1 - \gamma)\,\tilde{\delta}[\tau(c)]\,g(c) - (k + c) + S. \tag{29}$$

Proposition 2 *Use of an apprentice contract increases the profit a firm receives from training. It is profitable for more firms to train, and more workers are trained, with an apprentice contract than without.*

Proof. Consider the optimal training c^0 when no contract is used given by (15). It follows from comparison of (15) with (18) in Lemma 1 that to induce training of c^0 with an apprentice contract would require a contract of length τ^0 satisfying $\delta(\tau^0) = (1 - \gamma)\delta$, which implies $\tau^0 < T$ since, by definition, $\delta(T) = \delta$. It follows from (28) that $\tilde{\delta}(\tau^0) > \delta$. It then follows from comparison of (14) with (29) that, when there is no subsidy, profit is greater when c^0 is achieved with an apprentice contract than with no contract. Since c^0 maximizes expected profit when there is no contract, it follows that a training firm always makes greater expected profit with a contract than without. The highest fixed cost of training for which profits are non-negative is

$$k(c, S) \equiv (1 - \gamma)\,\tilde{\delta}[\tau(c)]\,g(c) - c + S. \tag{30}$$

Because profits for any given k are higher with an apprentice contract than without, more firms train and the number of workers trained is higher. ∎

Proposition 3 *The level of training and the number of trainees under apprentice contracts are lower than is efficient. The length of the apprentice contract is strictly less than the whole working lifetime T.*

Proof. The optimal training $\tilde{c}$ with a training contract ensures that the derivative of expected profit with respect to c, namely

$$\frac{\partial \tilde{\Pi}(k, \tilde{c}, S)}{\partial c} = (1 - \gamma)\left\{\tilde{\delta}[\tau(\tilde{c})]\,g'(\tilde{c}) + \tilde{\delta}'[\tau(\tilde{c})]\,\tau'(\tilde{c})\,g(\tilde{c})\right\} - 1, \tag{31}$$

is either zero, or positive but with $\tau(\tilde{c}) = T$, the longest possible contract length. (Note that $\tau'(c) > 0$.) However, from (27) and (28), $\tilde{\delta}(T) = \delta$ and $\tilde{\delta}'(T) = 0$. It thus follows from (18) in Lemma 1 that the derivative in (31) is negative for $\tau(c) = T$, which implies that $\tau(\tilde{c}) < T$. Thus the optimal training contract certainly lasts less than the trainee's whole working lifetime T. It also follows from (18) that $\tilde{c} < \bar{c}$ defined in (19). Moreover, since $\bar{c}$ is less than the efficient level c^* whenever $\rho > 0$, training is below the efficient level. It thus follows from (30) that $k(\tilde{c}, 0) < k_1^*$ defined in (5), so the number of workers trained is also below the efficient level in the absence of a subsidy. ∎

The intuition for these results is as follows. As explained in the previous section, one reason for training being inefficiently low when there is no training contract is that the training

17

firm does not capture any of the return to training arising after the trainee quits for exogenous reasons. That return is received by the worker's new employer, who acquires an employee with expected productivity $w^0 + g(c)$ at a wage $w^0 + \gamma g(c)$. A training contract provides a mechanism for the training firm to commit to ensuring its trainee receives a wage equal to marginal product after the end of the training contract, the period from τ to T in Figure 1. Thus, when an exogenous quit occurs after the end of the training contract, the worker's new employer does not capture any of the return to training. That return goes to the worker who, in turn, passes it to the training firm in the form of lower wages during the training contract. So the training firm captures more of the return to training. To see this more formally, consider the limiting case of no exogenous quits, $\rho = 0$. Then, from (1) and (2), $\delta(t) = \beta(t)$ and thus, from (26), $\tilde{\delta}(t) = \beta$. In that case, expected profits from any given level of training with a contract, $\tilde{\Pi}(k, c, 0)$ in (29), are identical to those without a contract, $\Pi^0(k, c)$ in (13), for any given k and c, so the optimal training level must also be the same.

By reducing the length of the contract, the training firm reduces the amount of the return to training that goes as an external benefit to the new employers of quitting workers because that increases the length of the post-contract period (τ to T in Figure 1). However, it also reduces the length of time during which the training firm receives returns to training, the period from 0 to τ in Figure 1. That, via (18), reduces the amount of training the firm provides. The optimal length of contract trades off these two effects. The optimal length depends in general on the form of the function $g(c)$ about which we have little information. There is, however, one case, the limiting one with no exogenous quits ($\rho = 0$), in which the optimal length of the contract is independent of the form of the function $g(c)$ and this case may be a useful guide for cases in which efficient turnover is low. For $\rho = 0$, recall that $\tilde{\delta}(t) = \beta$. Then $\tilde{\delta}'(t) = 0$ and, from (31), the first order condition derived from setting $\partial \tilde{\Pi}(k, \tilde{c}, S) / \partial c = 0$ can be written

$$\beta g'(\tilde{c}) = \frac{1}{1 - \gamma}. \tag{32}$$

Use of this in (20) with $\rho = 0$ gives

$$\tau(\tilde{c}) \equiv -\frac{1}{r} \ln\left[1 - r\beta(1 - \gamma)\right]. \tag{33}$$

Thus values of $\tau(\tilde{c})$ can be calculated directly for given values of r, γ and T. (β is determined by r and T.)

Some sample values are given in Table 1. For the calculations, we have used working life-times of 25 and 40 years, which span what would seem reasonable historically given mortality rates and the fact that people may change occupations, thus reducing the useful life of skills. We have used real discount rates of 5%, 10% and 20% per year, since these need to allow for depreciation of skills that we have not incorporated explicitly into the model. There is little evidence on what would be reasonable values for γ, the proportion of the average value of

18

	$T = 25$		$T = 40$	
r/γ	**0.5**	**0.9**	**0.5**	**0.9**
5% pa	8.8	1.5	11.3	1.8
10% pa	6.1	1.0	6.8	1.0
20% pa	3.4	0.5	3.5	0.5

Table 1: Market apprenticeship length (years)

acquired skills a trainee can earn by quitting before the end of the contract. Acemoglu and Pischke (1998, Table III) give point (but not very precise) estimates for Germany that completing an apprenticeship, but not staying with the training firm afterwards, adds between 0.024 and 0.041 to log wages. These can be interpreted as estimates of $\gamma\, g\,(c)\,/w^0$. Staying with the training firm for the initial period after training adds another 0.012 to log wages, which gives a point estimate for $\bar{\gamma}\, g\,(c)\,/w^0$ of between 0.036 and 0.053. These figures suggest point estimates for $\gamma\,/\bar{\gamma}$ of between 2/3 and 3/4. Since $\bar{\gamma}\,>\,1$, $\gamma\,/\bar{\gamma}$ is a lower bound for γ. Because these estimates are not very precise, we use values of 0.5 and 0.9 for γ in the table. It is clear from the table that the calculations are not very sensitive to the working lifetime, particularly with the higher discount rates, but differ a lot between the two values of γ. It is, however, reassuring that the contract lengths given in Table 1 encompass the lengths of apprenticeships typically experienced historically.[13]

5 Regulation of training contracts

The previous section has shown that, even with a training contract, the amount of training $\tilde{c}$ and the number of workers trained in an unregulated market are less than the efficient levels whenever ρ and/or $\gamma\,>\,0$. This section discusses appropriate policies of regulation and subsidy of training contracts under these circumstances. An obvious role for regulation is to monitor training by, for example, regulating the curriculum of training programmes and setting tests of competence. If this were 100% effective, it would make the amount of training verifiable and the issues discussed here would go away. Since, however, there are in practice many dimensions to quantity and quality of training, the cost of full monitoring may well be prohibitively high. The issues discussed here then continue to apply to any aspect of training that is not fully monitored by the regulator.

---

[13]There are a number of ways in which apprenticed workers traditionally suffered greater penalties from quitting than other workers, see Hamilton (1995). To the extent that these penalties reduce trainees' ability to capture returns from training, they can be incorporated into the model by a lower value of γ for apprenticed than for other workers. In that case, firms make higher profits with an apprenticeship than without one even with $\rho = 0$. There is then an additional reason for the use of apprenticeships which reinforces that discussed here. Since the value of γ in the period after the apprenticeship ends affects only wage differentials, not the amount of training or the length of the apprenticeship, the optimal apprenticeship length is then given by the length in the table that corresponds to the value of γ *during* the apprenticeship.

One role for regulation is in trying to ensure that trainees lose out by quitting early, which is analytically equivalent to reducing γ. The strict rules imposed by the medieval guilds against working as a journeyman before completion of an apprenticeship would obviously have served this purpose by preventing competitors competing away apprentices. For regulation of this type, apprenticeships have the advantage that the rules need to be imposed only for the length of the apprenticeship τ, not for the worker's whole working life as in the case without a training contract. But even if γ can be reduced to zero, this will still not achieve the efficient level of training as long as $\rho > 0$.

Other policies that might be considered by government are regulation of the length of the apprentice contract and a subsidy for some observable measure of training. There are two margins of concern to a regulator. The first is the amount of training for each employee trained. The second is the total number of employees trained. The regulator can affect the latter by offering a lump sum subsidy per trainee that increases the profits from training (29) and so, as can be seen from (30), induce firms with higher fixed costs of training to take on trainees. However, a lump sum subsidy does not affect the training received by each trainee because it does not affect the first order condition derived from differentiating profits (29) with respect to c. The regulator cannot observe training but can indirectly influence the training each trainee receives by regulating the length of training contract because there is a direct relationship between the length of the contract and the amount of training via (18).[14] However, the longest possible length of contract is the whole working lifetime T and thus the highest amount of training consistent with (18) is $\bar{c}$ defined in (19). So $\bar{c}$ is an upper bound on the amount of training that can be attained with these regulatory instruments.

Regulation of contract length is ineffective without at least some regulation of the wage. Without wage regulation, the firm and the trainee could effectively evade regulation of contract length because the firm could make the privately optimal contract nominally satisfy the regulation by extending its length by the required amount, but with a wage in this extension equal to the expected trained marginal product, and guaranteeing not to dismiss less good trainees. Such evasion can be prevented by regulating the total wage payment over the whole contract to the value of W given by (24) for the given regulated length and the value of c that satisfies (18).[15] The formal regulation would then be that any firm wishing to pay more than the untrained wage w^0 to a worker who has not had a previous training contract must offer the regulated training contract.[16]

There are also limitations on the amount of subsidy that can be paid. If the subsidy is large

---

[14]Alternatively, it can do so by offering a subsidy that increases with the length of training. It turns out that this has no advantages over regulation of the contract length, so we do not pursue it here.

[15]In Germany, unions regulate pay of apprentices which can serve equally well. We are grateful to Daron Acemoglu for pointing this out.

[16]If the regulation were limited to those offering training contracts, firms could choose to train without a training contract if that were more profitable. To avoid that would require another constraint on regulation that we do not explore here.

enough, it will be worth firms with high training costs offering a training contract in order to attract the subsidy even though they do not intend to train. To prevent that requires the amount of the subsidy to be less than the difference between expected wages under a training contract and expected wages to an untrained worker over the length of the training contract. That gives a constraint on the maximum subsidy of

$$S \leq W - \delta\left[t\left(c\right)\right]w^0 \text{ for } S > 0. \tag{34}$$

(Recall that S is defined as the expected present discounted value at the start of a contract of the subsidy paid at the end.)[17] Substitution for W from (24) and use of the definition of $\tilde{\delta}\left(\tau\right)$ in (26) allows that constraint to be written

$$S \leq g\left(c\right)\left\{\delta\left[\tau\left(c\right)\right] - \left(1 - \gamma\right)\tilde{\delta}\left[\tau\left(c\right)\right]\right\} \text{ for } S > 0. \tag{35}$$

With W fixed, the firm chooses c to maximise its expected revenue less the training costs over the regulated length of contract. That is the same c that maximises expected profits given in (25) for given τ and W and thus the relationship between the regulated length τ and the amount of training c is simply the first order condition (18). Given this, we can think in terms of the regulator choosing c rather than τ and then use (18) to determine the contract length that must be regulated in order to achieve that c. For any $c \in [0, \bar{c}]$, there is always a τ that implements that c.

We assume the regulator wishes to maximise the social gains from training but, in the conventional way, allowing for a proportional cost α of subsidies to account for distortions arising from having to raise revenue from taxation. These gains can be represented in the following way. The highest fixed cost of training that it is profitable for firms to incur when training to level c with subsidy S is $k\left(c, S\right)$ defined in (30). The number of firms with fixed cost below this, and hence the number of workers trained, is $F\left[k\left(c, S\right)\right]$. The total social benefit from each worker trained consists of the additional lifetime value of output per trainee $\beta g\left(c\right)$, less the training cost $c + k$ and the deadweight loss from the subsidy αS. Thus, the optimal regulatory policy is $(\hat{c}, \hat{S})$ given by the solution to

$$\max_{S, c \in [0, \bar{c}]} F\left[k\left(c, S\right)\right]\left[\beta g\left(c\right) - c - \alpha S\right] - \int_0^{k(c,S)} k \, dF\left(k\right) \text{ subject to (35),} \tag{36}$$

where the integral term is the total fixed cost of training $F\left[k\left(c, S\right)\right]$ workers.

For the moment, suppose the constraint (35) is not binding and note from (30) that $\partial k\left(c, S\right)/\partial S =$

---

[17]Regulating a wage W above that determined by the market would relax this constraint but not in a way that is helpful to the regulator. The only reason to offer a subsidy is to induce firms with higher fixed costs of training to train. Raising the wage has the opposite effect. It is the difference between S and W that determines the highest fixed cost for which it is profitable to train. Thus raising W does not relax the constraint in a way that enables the regulator to increase the number of workers trained.

1. The first order conditions for an interior solution to this problem, with $k_1(c, S)$ denoting the derivative of $k(c, S)$ with respect c, are then

$$-\alpha F[k(\hat{c}, \hat{S})] + dF[k(\hat{c}, \hat{S})]\left[\beta g(\hat{c}) - \hat{c} - \alpha \hat{S} - k(\hat{c}, \hat{S})\right] = 0 \quad (37)$$

$$\left[\beta g'(\hat{c}) - 1\right] F[k(\hat{c}, \hat{S})] + k_1(\hat{c}, \hat{S})dF[k(\hat{c}, \hat{S})]\left[\beta g(\hat{c}) - \hat{c} - \alpha \hat{S} - k(\hat{c}, \hat{S})\right] = 0. \quad (38)$$

Sufficient (but not necessary) conditions for the second order conditions for a maximum to be satisfied are that $k_1(\hat{c}, \hat{S}) < 0$ and $d^2 F\left[k(\hat{c}, \hat{S})\right] \leq 0$.

The intuition behind (37) is as follows. The marginal cost to a subsidy for given training $\hat{c}$ is the welfare loss α from the additional subsidy for each of the $F[k(\hat{c}, \hat{S})]$ workers trained. The marginal benefit is that $dF[k(\hat{c}, \hat{S})]$ more firms find it profitable to train, which increases welfare by the difference between the returns to training, $\beta g(\hat{c})$, and the cost of the training both in resources $\hat{c} + k(\hat{c}, \hat{S})$ and in the deadweight loss from the subsidy, $\alpha \hat{S}$. The optimal subsidy balances these two. The marginal benefit of the training received by each trainee for given $\hat{S}$ is the difference between the marginal returns to expenditure on training $\beta g'(\hat{c})$ and its cost of 1, multiplied by the number of workers trained $F[k(\hat{c}, \hat{S})]$. That is the first term in (38). The marginal cost is that increasing training above the profit maximizing level reduces the profits from training, so $dF[k(\hat{c}, \hat{S})]k_1(\hat{c}, \hat{S})$ fewer firms train and there is a welfare loss of the difference between the returns to training, $\beta g(\hat{c})$, and the cost of the training both in resources $\hat{c} + k(\hat{c}, \hat{S})$ and in the deadweight loss from the subsidy, $\alpha \hat{S}$. The optimal amount of training balances these two effects. Our first result on regulation follows directly from (38).

Proposition 4 *For any given subsidy S (including $S = 0$) that with unregulated contracts satisfies the constraint (35), social welfare is increased by regulating a longer contract, and thereby achieving a higher training per worker, than would be set without such regulation.*

Proof. Suppose contract length is not regulated. Then firms choose a contract length that results in training c that maximises $\tilde{\Pi}(k, c, S)$ in (29) for any given S. It follows from (29) and (30) that $k_1(c, S) = \partial \tilde{\Pi}(k, c, S)/\partial c$. Thus, in the absence of regulation of contract length, training is at a level $\tilde{c}$ for which $k_1(\tilde{c}, S) = 0$. (We know from Proposition 3 that the contract is for less than the whole lifetime so that the optimum is interior with $\tilde{c} < \bar{c}$. Note that $\tilde{c}$ is actually independent of S.) At any such $\tilde{c}$, therefore, the second term in (38) is zero. However, the first term would be strictly positive—it would be zero only for the first best level of training c^* defined in (4) and we showed in Proposition 3 that training without regulation $\tilde{c}$ is less than c^*. Thus, no such $\tilde{c}$ can satisfy the first-order condition (38). Moreover, the left hand side of (38) is the effect on welfare of an increase in c. That this left hand side is strictly positive implies that social welfare is increased by increasing c from any $\tilde{c}$ that results in the absence of regulation.

Given $\tilde{c} < \bar{c}$, such an increase in c is feasible as long as it does not violate the maximum subsidy constraint (35). For $S = 0$, that constraint does not apply. For $S > 0$, that constraint

can be satisfied at the unregulated contract length only if the right hand side is strictly positive at $\tilde{c}$. When, however, the right hand side of (35) is positive, its derivative with respect to c is also positive. Thus the constraint cannot become binding as regulated contract length is increased above the unregulated contract length. ∎

The intuition for this result is as follows. The number of workers trained is determined by the expected profit from training. At the profit maximizing amount of training, a small change in that amount has only a second order effect on expected profit and, hence, only a second order effect on the number trained. But it has a first order effect on the training received by each trainee and, because of the externalities, this has a first order effect on social welfare.[18]

This result has an interesting implication. In practice, there are administrative costs in setting up a subsidy system that may have been too high to be worthwhile to, for example, medieval and Tudor government authorities. A government with no mechanism for paying subsidies, or a regulator who has no tax raising powers, will nevertheless wish to regulate the length of training. The model is, therefore, consistent with the historically widespread regulation of apprenticeships even when training is not subsidised.[19]

When there is a subsidy to training, (37) and (38) together imply

$$\beta g'(\hat{c}) - 1 + \alpha k_1(\hat{c}, \hat{S}) = 0, \tag{39}$$

which, since $k_1(c, S)$ is actually independent of S, determines $\hat{c}$, though not necessarily uniquely if $k_1(c, S)$ is not monotone decreasing in c. The corresponding subsidy $\hat{S}$ can then be calculated for any distribution $F(k)$ from (37) which, with substitution for $k(\hat{c}, \hat{S})$ from (30), can be written in the form

$$\frac{dF\left[k(\hat{c}, \hat{S})\right]}{F\left[k(\hat{c}, \hat{S})\right]} \left\{ g\left(\hat{c}\right) \left[\beta - (1 - \gamma)\tilde{\delta}\left(\tau\left(\hat{c}\right)\right)\right] - (1 + \alpha)\hat{S} \right\} = \alpha. \tag{40}$$

For given $\hat{c}$ and any distribution of k for which $dF(k)/F(k)$ is monotone decreasing (a hazard rate assumption that is standard in contract theory, see Laffont and Tirole (1993, p. 66), and is satisfied by most of the usual distributions), there is a unique value of $\hat{S}$ that satisfies (40). In other cases, if there is more than one value of $\hat{S}$ that satisfies (40) or $\hat{c}$ that satisfies (39), a

---

[18]Although social welfare is increased by increasing contract length from the unregulated level, we have not been able to show that the global regulated optimum always has a longer contract than an unregulated market for the general form of the function $g(c)$ used here. However, it certainly does if $g(c)$ takes a constant elasticity form or if ρ is sufficiently small. We have not found examples in which it does not.

[19]The same conclusion applies even if the regulator is controlled by firms and thus wishes to maximize the total expected profits of all firms. The reason is that, because of the externality derived by non-training firms when trained workers quit for exogenous reasons, an unregulated market does not maximize total expected profits of all firms. The model is thus also consistent with medieval guilds regulating the length of apprenticeships even before government authorities did so.

further check must be used to establish which is a global maximum.[20]

When the deadweight cost of taxation gets large ($\alpha \to \infty$), (40) cannot be satisfied with equality and we have the corner solution with $\hat{S} = 0$. The maximum subsidy constraint (35) does not apply when $S = 0$, so this solution is optimal. The other extreme is no deadweight cost to raising funds via taxation ($\alpha = 0$). To satisfy (39) would then require $\hat{c} = c^*$, the efficient level defined by (4). But that is not possible whenever there are exogenous quits ($\rho > 0$) because the maximum training the regulator can induce (corresponding to $\tau = T$) is $\bar{c} < c^*$. Thus we have the corner solution with $\hat{c} = \bar{c}$, corresponding to a lifetime training contract. In this case, however, the maximum subsidy condition (35) always binds. This can be seen as follows. From (28) we know that $\tilde{\delta}(T) = \delta(T) \equiv \delta$. Thus, for $\hat{c} = \bar{c}$ and hence $\tau = T$, (35) is satisfied only if $S \leq \delta \gamma g(\bar{c})$. But then, with $\alpha = 0$, (40) implies $\hat{S} = g(\bar{c})[\beta - (1 - \gamma)\delta]$ which, since $\beta > \delta$ when $\rho > 0$, does not satisfy the maximum subsidy constraint. Because, however, that constraint cannot be relaxed by reducing c and because c cannot be increased above $\bar{c}$, it is nevertheless optimal to set $\hat{c} = \bar{c}$ and $\hat{S} = \delta \gamma g(\bar{c})$. From (30) we then have $k(\hat{c}, \hat{S}) = \delta g(\bar{c}) - \bar{c}$. This is less than the efficient level k_1^* defined in (5). Thus, despite there being no deadweight cost to taxation, both the number of workers trained and the amount of training received by each trainee are below the efficient levels.

This raises the obvious question of whether the regulator could do better with some other kind of regulation of training contracts. It is not, however, obvious what regulation would do better. It is clear from (25) that a wage bill W that is decreasing in c for given τ would induce more investment for a given contract length. However, because c is non-contractible, it is not clear how such an arrangement could be implemented.

In the limiting case with no exogenous quits ($\rho = 0$), the optimal regulatory solution has a particularly simple form that may be a useful guide for cases in which efficient turnover is low. For $\rho = 0$, (39) reduces to

$$\beta g'(\hat{c}) = \frac{1}{1 - \frac{\alpha}{1+\alpha}\gamma}, \tag{41}$$

which determines $\hat{c}$ uniquely. The regulated length of apprenticeship $\hat{\tau}$ that corresponds to $\hat{c}$ is given by use of (41) in (20),

$$\hat{\tau} = -\frac{1}{r}\ln\left[1 - r\beta\left(1 - \frac{\alpha}{1+\alpha}\gamma\right)\right]. \tag{42}$$

Moreover, in this case the maximum subsidy constraint (35) is never binding at an optimum.

---

[20]This argument has ignored the maximum subsidy constraint (35). As shown in the proof of Proposition 4, however, when the right hand side of (35) is positive (a necessary condition for the constraint to be satisfied), its derivative with respect to c is also positive. Thus one cannot relax the constraint by reducing c. The implication is that the optimal level $\hat{c}$ is at least as large as that implied by (39).

	$\alpha = 0.5$		$\alpha = 1$		$\alpha = 2$	
$r/^\gamma$	**0.5**	**0.9**	**0.5**	**0.9**	**0.5**	**0.9**
5% pa	25.5	18.6	20.9	12.9	17.2	8.5
10% pa	17.0	11.6	13.3	7.8	10.6	5.0
20% pa	9.0	6.0	6.9	4.0	5.5	2.6

Table 2: Optimal regulated apprenticeship length (years), working lifetime 40 years

To see this, note that, from (40),

$$\hat{S} \leq \frac{\gamma \beta g\left(\hat{c}\right)}{1 + \alpha}.$$ (43)

Thus (35) is satisfied as long as

$$\frac{\beta\left[\tau\left(\hat{c}\right)\right]}{\beta} - (1 - \gamma) \geq \frac{\gamma}{1 + \alpha}.$$ (44)

But (18) and (39) imply that this condition holds with equality and hence the optimal solution always satisfies (35).

In this case, the optimal regulated length of training $\hat{\tau}$ in (42) is independent of the form of the training function $g\left(.\right)$ and the distribution function $F\left(.\right)$. Some sample values of $\bar{\tau}$ for $T = 40$ years, r and γ as in Table 1, and values for α of 0.5, 1, and 2 are given in Table 2. For the low values of all the parameters in the table, the contract lengths look on the high side but for higher values are of orders of magnitude observed historically. The Elizabethan Statute of Artificers of 1563 regulated apprenticeships to 7 years in England, see Bindoff (1950, p. 201). In modern day Germany apprenticeships for a wide variety of occupations last 3-4 years, see Soskice (1994).

The general conclusion from this section is that regulation of contract length is always optimal even if subsidising training has too high a deadweight cost, or is otherwise too difficult, to be worthwhile. An unregulated market provides too little training because both trainees and the new employers of trainees who quit capture some part of the returns to the training firm's investment in training. This part of the returns is still a social benefit from training even though not a private benefit to the training firm. Regulation takes account of that.

6   Alternative views and historical evidence

In the model in this paper, general training by firms is provided within an apprenticeship type system in order to mitigate the market failure that arises when the amount of training cannot be specified contractually in advance. Instead, the length of the apprenticeship is specified contractually in advance, rather than being determined by each apprentice's ability to perform the specific tasks required of a skilled worker. The model also provides a welfare-enhancing

role for policy towards apprenticeships. Historically, it has certainly been the case that much general training undertaken in the workplace has occurred within the context of a formal apprentice system with its duration specified contractually in advance—for example, the métier in France, arte in Italy, Amt, Innung, Zunft or Handwerk in Germany, and craft guild in England. Contemporary training of lawyers, doctors and accountants also typically involves a specified duration of training on the job that is not reduced just because a trainee is quicker than average at acquiring the appropriate skills. Moreover, there is a long history of regulation of apprenticeships, certainly from medieval times onward, see Pirenne (1936). In England, this was put on a nationwide statutory basis by the Elizabethan Statute of Artificers in 1563, see Bindoff (1959, p. 201).

Reasons other than those modelled here have been suggested for these characteristics. The most notable are that (a) apprenticeships are a device to create monopoly rents by controlling the supply of skilled workers, and that (b) they are a way of overcoming worker's limited access to capital to pay for general training within Becker's (1964) human capital framework. There are, however, a number of historical features of apprenticeships that are not easily reconciled with these hypotheses.

Consider first the view that formal apprenticeships are a device to create monopoly rents by controlling the supply of skilled workers. Smith (1887, Book 1, Ch. 8) discussed extensively the view that apprenticeships acted as a way for craftsmen to control access to their trade in order to increase their earnings. He argued that a long period of apprenticeship unnecessarily raises the cost of becoming skilled, thereby artificially reducing the skilled labour supply. Certainly, regulation reducing the numbers entering skilled trades would have controlled the supply of skilled workers. Two things, however, suggest that this was not the only reason for apprenticeships. First, that purpose would have been served simply by regulating the length of training and who could provide it. In fact, there was also extensive regulation of the adequacy of training. The craft guilds of the middle ages had supervisory functions that included the right of search to ensure that good materials and appropriate processes of manufacture were employed, and that masters "took measure to secure that workmen should be properly trained by serving a regular apprenticeship, and they made rules affecting the hours of labour and well-being of those employed" (Cunningham and McArthur, 1920, p.61). Moreover, in modern Germany a range of institutions funded collectively by firms oversee the working of the apprentice system (Harhoff and Kane, 1993; Steedman, 1993; Soskice, 1994); and in the UK the Industrial Training Boards monitor industrial training. Second, one would expect apprenticeships to decline in importance once craft guilds consisting of skilled workers lost control of the numbers admitted to apprenticeships. While apprenticeship declined in Britain after the repeal of the Statute of Artificers in 1814, the diminished institution in fact continued into the mid-twentieth century, long after an apprenticeship was required to sell skilled labour services. Indeed, not only did apprenticeship survive, but it expanded into the new twentieth century industries that had not

existed in 1814 and were not, apparently at least, controlled by their skilled workers.

Consider now limited access to capital by workers. In Becker's (1964) analysis of human capital, it is efficient for workers to pay for general training and apprenticeships at a low wage are an obvious way for firms to recoup the cost of training from employees who may not have ready access to capital markets to pay for their own training in advance. Elbaum (1989) argues on the basis of statements from contemporaries, estimates of training costs, and the "extraordinarily large pay increases" on completion of apprenticeship, that the later stages of apprenticeship in early twentieth century Britain were indeed playing the role of enabling firms to recoup costs of training. If, however, apprenticeships were only about ensuring that workers pay for general training, there would be no obvious role for regulation apart from enforcing the agreements reached privately between workers and training firms. In fact regulation went far beyond this. As already noted, apprenticeship length was regulated in the middle ages and given a statutory basis in Elizabethan England by the Statute of Artificers. Moreover, there was extensive regulation of the adequacy of training. More recently, subsidies have been used to encourage training, for example the Youth Training Programmes in the UK, see Dolton (1993). Within the Becker framework, government financial inducements to undertake training may be appropriate if workers have higher discount rates than firms, with the result that the enforced low consumption during training at low wages implies a significant utility loss. However, the natural policy under these circumstances would be to make loans available to employees for on-the-job training, and that has not been the typical form of financial support used in practice, see Dolton (1993) for the UK and Lynch (1993) for the US. Moreover, apprenticeships have been used not only for artisan professions but also in the medical, legal and accounting professions for the younger offspring of middle class and wealthy families who were often wealthy enough to pay for general training directly, see Earle (1989, ch. 3). For all these reasons, it seems unlikely that trainees' limited access to capital was the primary reason for the use of apprenticeships.

In contrast, regulation of both length and adequacy of training contracts makes sense within the model developed here. Elbaum (1989, p. 344) in his discussion of apprenticeships in Britain is explicit about the problem of ensuring that firms deliver on training promises: "youths who sacrificed current wages in return for the promise of training were generally vulnerable to exploitation by employers who failed to live up to their training commitments." The extensive use of institutions to set and monitor standards of training is indicative that contractual enforcement of standards by individual trainees appealing to courts is not straightforward. That non-contracted considerations were important is indicated by the extensive evidence that potential apprentices and their parents placed central emphasis on the moral character of the master, resorting where possible to a relative, see (Earle, 1989). The more that efficient standards can be enforced either in this way or by the direct regulation of the curriculum as in modern Germany, the less the welfare loss in the model from underprovision of training by the market.

Moreover, within the model, regulation of length of apprenticeships is a welfare enhancing way to move market provision of training closer to the efficient level. Thus both these dimensions of regulation have an obvious role within the model developed here.

As well as being consistent with the institutional structures accompanying training, the model analysed here predicts that wages of trainees staying with their training firm after the end of their apprenticeship (a) increase to the value of their marginal product from a level below marginal product during the final stage of their apprenticeship, and (b) are, ceteris paribus, higher than wages of those who change employer. There is considerable evidence supporting point (a), for example, the extraordinarily large pay increases on completion of apprenticeship documented by Elbaum (1989). Careful evidence on point (b) is rare and made problematic by the heterogeneity in types of skills and observable characteristics of those completing apprenticeships. However, in a regression study of German apprentices, Acemoglu and Pischke (1998) find that, within occupations, retained workers earn more than those leaving for another non-military job.

7 Concluding remarks

When the amount of training is not contractible and firms have costs of training that are unknown to workers, workers will not pay directly for the full cost of general training even if they have access to funds to do so. We have shown that, when in addition workers are free to quit without penalty and some do so, apprenticeships are an institution that enables firms to capture more of the returns to the general training they provide. This applies even if training with an apprenticeship is no more effective at preventing employees quitting than training without one. Thus there remains a case for using apprenticeships even where they confer no advantage in preventing rival firms bidding away trainees. The model developed here captures many of the salient characteristics of traditional on-the-job general training. Moreover, it has implications for the analysis of policy. We have shown that it is in general efficient to regulate the length of apprenticeships and, if the deadweight loss from raising funds through taxation is sufficiently low, to subsidise their completion.

The model has been kept simple for expositional reasons but the basic insights do not depend on that. One obvious simplification concerns the timing of training during the apprenticeship period. In the model developed here, all training takes place straightaway at the start of the training contract. The rest of the training contract is to enable the firm to recoup the cost of training. There are, however, two reasons why firms may not wish to provide all training straightaway. The first follows from the standard assumption in human capital theory that a given amount of training is more effective if spread out over a longer period of time. The second follows from the result in the model that the binding constraint for trainee quitting is at the start of the training period. Thus firms could provide some training later in the training period

without making this crucial constraint tighter. Delaying training for either reason, of course, reduces the period over which the firm can recoup the cost of training and the optimal timing would have to balance the two effects, but one would expect at least some training to occur after the start of the training period.

Other obvious simplifications are that we have assumed the environment is stationary with no uncertainty except about the types of trainees, and that firms do not build reputations for providing levels of training beyond their immediate short term interest. These are not, however, crucial to the underlying insights. If labour productivity grew at a constant proportional rate for exogenous reasons unrelated to training, that would enter the model like a reduction in the discount rate. Moreover, if there were uncertainty about the return to training for a given type of trainee, one could simply interpret the return to training as the expected return at the time the training was given. Finally, even if information flows are sufficient to enable firms to develop reputations for providing good training, such reputations are worth maintaining only for those with fixed costs of training sufficiently low that their profits from training in the future are high enough to keep them honest in the short term. In particular, reputations will never keep marginally profitable firms honest, so the market will still provide too little training and the role for government policy remains.

It thus seems that the underlying insights in the model presented here are reasonably robust to these types of generalization of the model. Those insights are also consistent with much of the historical evidence on general training provided by firms. They therefore seem a promising basis for further analysis.

Appendices

A   The payoff to trainees who quit

This appendix sets out a game for the hiring process with hiring at discrete time intervals that is consistent with the employment model in the main text. It shows that the expression in (7) for the remaining lifetime earnings of a trainee who quits is a perfect Bayesian equilibrium payoff of a trainee who quits as the time interval between hiring dates goes to zero.

For simplicity, consider 1 worker and m firms, each with one vacancy, plus the training firm that hires and trains the worker at date 0. Divide the worker's working lifetime $[0, T]$ into a large number N of equally spaced intermediate hiring dates at which job offers can be made, the interval between them being $dt = T / (N + 1)$. The first hiring date at which a non-training firm can make an offer is thus $t = dt$, the training firm being the incumbent employer at this date. The worker works continuously between hiring dates at the wage agreed with the current employer at the most recent hiring date. She quits at the next hiring date for exogenous reasons with a probability that corresponds to the rate ρ during a period of employment. Thus, at each

	Player to move	**Moves**	**Information Set**
Stage *n*.1	Nature	1. Determines whether worker quits for exogenous reasons 2. Informs employing firm of worker's type	
Stage *n*.2	Firms with vacancies (simultaneously)	Offer wage to allocated worker	Worker's training level & wage history
Stage *n*.3	Firm with worker	Counter-offer wage to incumbent worker	1. Worker's training level & wage history 2. Worker's type 3. Offers made to worker at stage *n*.2

Stage *n*.4	Worker	Accept no offer (& quit market) or offer of 1 firm from:
		
		    1. outside offers, if quitting for exogenous reasons
		
		    2. all offers, otherwise
			1. Own training level & wage history
			
			2. Offers made at stage *n*.2 and *n*.3

Table 3: Stage game at stage n

hiring date, the probability the worker quits no matter what offer the current employer makes is $1 - e^{-\rho dt}$. Whatever the reason for a quit, the new employer incurs a once-off hiring cost $\varepsilon > 0$.

At each date ndt at which job offers can be made, the following stage game is played. First, at stage $n.1$, Nature determines whether the worker decides to quit for exogenous reasons (an event that has probability $1 - e^{-\rho dt}$) and informs the current employing firm of the worker's type. At stage $n.2$, the m firms with vacancies each make a simultaneous wage offer to the worker. Each firm knows what wage the worker has received at each previous stage, that is, the worker's wage history. With w_t denoting the wage at t, the wage history at date $t = ndt$ for some n is denoted by the $n - 1$ dimensional vector of real numbers $w^t = (w_0, w_{dt}, \ldots, w_{t-dt})$. It does not know the worker's type.[21] At stage $n.3$, the currently employing firm makes a counter-offer to the worker. The firm knows the worker's wage history, the worker's type, and the offers made at stage $n.2$. Finally, at stage $n.4$, the worker either accepts the wage offer of one firm or accepts no offer. If accepting no offer, the worker leaves the labour market permanently and receives a payoff at the rate $u < w^0 + \gamma g(c)$ for subsequent stages. If the worker is quitting for exogenous reasons, the offer accepted must be from an outside firm. Otherwise it can be from any firm. At this stage, the worker knows her own wage history and the offers made to her at stages $n.3$ and $n.4$. The stage game is summarised in Table 3.

For describing payoffs and strategies in the game, we use the following notation. Since all firms know the worker's level of training c, we can simplify notation by denoting the productivity of a better worker (type $\bar{\gamma}$) by $\bar{z}\ (= w^0 + \bar{\gamma} g(c))$, that of a less good worker (type $\underline{\gamma}$) by $\underline{z}\ (= w^0 + \underline{\gamma} g(c))$. The flow payoff to a firm from employing a worker of type $z \in \{\underline{z}, \bar{z}\}$ at wage w_t from t to $t + dt$ is $(z - w_t)\, dt$, that from not employing a worker is zero. The flow payoff to the worker is $w_t dt$. The structure of the game, and the payoffs conditional on worker

---

[21]This formulation requires that a firm forgets the worker's type when the worker quits its employment. This simplifies the formal model by ensuring that no more than one firm knows the worker's type at any hiring date. We discuss implications of relaxing this requirement later.

type, are common knowledge.

Strategies in the game can be described as follows. The strategy component at t for each firm $f \in \{1, \ldots, m+1\}$ conditional on having a vacancy at t is a mapping from the wage history w^t to a real number that corresponds to a wage offer, denoted $w_t^f(w^t)$. That for each firm $i \in \{1, \ldots, m+1\}$ conditional on employing the worker at $t - dt$ is a mapping from the worker's type z, the wage history w^t, and the wage offers by firms with vacancies at t which (with some abuse of notation) we denote $\{w_t^f, f \neq i\}$ to a real number that again corresponds to a wage offer, denoted $\hat{w}_t^i(z, w^t, \{w_t^f, f \neq i\})$. The strategy for the worker at t employed by firm i at $t - dt$ is a mapping from the wage history w^t, the offers by firms with vacancies at t $\{w_t^f, f \neq i\}$, and the counter-offer by firm i, denoted (again with some abuse of notation) $\hat{w}_t^i$, to the set of integers $\{0, 1, \ldots, m+1\}$ if not quitting for exogenous reasons and to the set of integers $\{0, 1, \ldots, i-1, i+1, \ldots, m+1\}$ if quitting for exogenous reasons. The integer chosen identifies the firm whose offer is accepted, with 0 denoting that no offer is accepted and the worker is leaving the labour market. We denote the worker's strategy component at t when currently employed by firm i by $a_t^i(w^t, \{w_t^f, f \neq i\}, \hat{w}_t^i), i \in \{1, \ldots, m+1\}$. The complete strategy space for all players is denoted $\mathcal{S}$, a profile of strategies (one for each player) by $S \in \mathcal{S}$.

To demonstrate the result in (7), we show that the wage path $w_t = \underline{z}$ for all $t \in (0, T]$ is a perfect Bayesian equilibrium for the continuous version of this game (as $dt \to 0$) as the number of firms with vacancies $m \to \infty$ and the hiring cost $\varepsilon \to 0$. For this demonstration, we use the following notation. Let $w_t^+ = (w_{t+dt}, w_{t+2dt}, \ldots, w_{T-dt})$ be a wage path after t. Let $\Pi_t^*(z, w_t, w_t^+)$ denote the expected profits a firm receives from t on from having in post a worker of type $z \in \{\underline{z}, \bar{z}\}$ when the wage path (w_t, w_t^+) is followed from t on and the worker quits only for exogenous reasons. Also let $\bar{\Pi}_t(w_t, w_t^+)$ denote the expected present discounted value of profits for a firm that starts period t with a vacancy in an equilibrium with future wage path (w_t, w_t^+) in which the worker quits only for exogenous reasons and the firm's belief is that the worker is type $\bar{z}$ with probability p. Note from the definition of w_t^+ that $w_t^+ = (w_{t+dt}, w_{t+2dt}^+)$, so we can use $\bar{\Pi}_{t+dt}(w_t^+)$ and $\bar{\Pi}_{t+dt}(w_{t+dt}, w_{t+dt}^+)$ interchangeably. Then $\bar{\Pi}_t(w_t, w_t^+)$ can be written

$$
\bar{\Pi}_t(w_t, w_t^+) = \max \left\{ e^{-rdt} \bar{\Pi}_{t+dt}(w_t^+), \frac{1 - e^{-\rho dt}}{m} \left[p \Pi_t^*(\bar{z}, w_t, w_t^+) + (1-p) \Pi_t^*(\underline{z}, w_t, w_t^+) - \varepsilon \right] \right.
$$
$$
\left. + \left(1 - \frac{1 - e^{-\rho dt}}{m}\right) e^{-rdt} \bar{\Pi}_{t+dt}(w_t^+) \right\}, \text{ for } t \in \{dt, 2dt, \ldots, T-dt\}, \tag{45}
$$

$$
= e^{-rdt} \bar{\Pi}_{t+dt}(w_t^+) + \frac{1 - e^{-\rho dt}}{m} \max \left\{ 0, \left[p \Pi_t^*(\bar{z}, w_t, w_t^+) + (1-p) \Pi_t^*(\underline{z}, w_t, w_t^+) \right. \right.
$$
$$
\left. \left. -\varepsilon \right] - e^{-rdt} \bar{\Pi}_{t+dt}(w_t^+) \right\}, \text{ for } t \in \{dt, 2dt, \ldots, T-dt\}, \tag{46}
$$

31

with $\bar{\Pi}_T \left(w_{T-dt}^+ \right) = 0$. The first term in braces in (45) is the expected profits from having a vacancy at $t + dt$, namely $\bar{\Pi}_{t+dt} \left(w_t^+ \right)$, discounted back to t. A firm with a vacancy cannot do worse than this because it can always offer a wage that will never be accepted, earn zero profit in period t and then retain the vacancy until $t + 1$. The second term in braces in (45) is constructed as follows. With probability $1 - e^{-\rho dt}$ the worker will quit at t for exogenous reasons. With m outside firms all bidding the same highest wage w_t, each has probability $1/m$ of acquiring the worker if the worker quits for exogenous reasons. With probability p the worker has high productivity $\bar{z}$ yielding expected future profit $\Pi_t^* \left(\bar{z}, w_t, w_t^+ \right)$, with probability $1 - p$ low productivity $\underline{z}$ yielding expected future profit $\Pi_t^* \left(\underline{z}, w_t, w_t^+ \right)$. In either case, the successful hiring firm incurs the hiring cost ε. The remaining term consists of the probability of not successfully hiring the worker multiplied by the expected profits from having a vacancy at $t + dt$ constructed as before. Since the game ends at T, it is necessarily the case that $\bar{\Pi}_T \left(w_{T-dt}^+ \right) = 0$. Note that, in evaluating $\bar{\Pi}_t \left(w_t, w_t^+ \right)$ and $\Pi_t^* \left(z, w_t, w_t^+ \right)$, we do not need to keep track of the number of firms with vacancies that actually bid w_t. Consider $\bar{\Pi}_t \left(w_t, w_t^+ \right)$. If $\bar{\Pi}_t \left(w_t, w_t^+ \right) > e^{-rdt} \bar{\Pi}_{t+dt} \left(w_t^+ \right)$, it is necessarily strictly optimal for all m to do so. If, on the other hand, $\bar{\Pi}_t \left(w_t, w_t^+ \right) = e^{-rdt} \bar{\Pi}_{t+dt} \left(w_t^+ \right)$, it makes no difference to the profit how many do so. Now consider $\Pi_t^* \left(z, w_t, w_t^+ \right)$. We can write $\Pi_t^* \left(z, w_t, w_t^+ \right)$ as

$$
\Pi_t^* \left(z, w_t, w_t^+ \right) = \sum_{k=1}^{(T-t)/dt} e^{-(r+\rho)(T-t-kdt)} \Bigg\{ \left(z - w_{T-kdt} \right)
$$
$$
+ e^{\rho dt} e^{-rdt} \left(1 - e^{-\rho dt} \right) \bar{\Pi}_{T-(k-1)dt} \left(w_{T-(k-1)dt}, w_{T-(k-1)dt}^+ \right) \Bigg\},
$$
$$
\text{for } t \in \{ dt, 2dt, \ldots, T - dt \}. \tag{47}
$$

This is constructed as follows. With probability $e^{-\rho(T-t-kdt)}$, the worker has not quit by $T - kdt$ so the firm receives profit $z - w_{T-kdt}$ for that period, discounted by the discount factor $e^{-r(T-t-kdt)}$. With probability $e^{-\rho(T-t-(k+1)dt)} \left(1 - e^{-\rho dt} \right)$, the worker has not quit by $T - (k+1) dt$ but quits at $T - kdt$. In that event, the firm starts period $T - (k-1) dt$ without a worker with expected payoff $\bar{\Pi}_{T-(k-1)dt} \left(w_{T-(k-1)dt}, w_{T-(k-1)dt}^+ \right)$ that is discounted by the factor $e^{-r(T-t-(k-1)dt)}$. Thus, since we do not need to keep track of the number of bidding firms in $\bar{\Pi}_t \left(w_t, w_t^+ \right)$, we do not need to do so in $\Pi_t^* \left(z, w_t, w_t^+ \right)$ either.

Next, define $\underline{w} \left(w_t^+ \right)$ and $\bar{w} \left(w_t^+ \right)$ by

$$
\underline{w} \left(w_t^+ \right) = \sup w \text{ such that } \Pi_t^* \left(\underline{z}, w, w_t^+ \right) \geq e^{-rdt} \bar{\Pi}_{t+dt} \left(w_t^+ \right), \text{ for } t \in \{ dt, 2dt, \ldots, T - dt \},
$$
$$
\tag{48}
$$
$$
\bar{w} \left(w_t^+ \right) = \sup w \text{ such that } \Pi_t^* \left(\bar{z}, w, w_t^+ \right) \geq e^{-rdt} \bar{\Pi}_{t+dt} \left(w_t^+ \right), \text{ for } t \in \{ dt, 2dt, \ldots, T - dt \}.
$$
$$
\tag{49}
$$

These are the highest wages a firm would be prepared to pay to retain a worker of type $\underline{z}$ and

$\bar{z}$ respectively at t given the future wage path w_t^+. Since $\bar{z} > \underline{z}$, it follows from (47) that $\bar{w}(w_t^+) > \underline{w}(w_t^+)$. Also, let $w^0(w_t^+)$ denote the lowest wage the worker will accept rather than leave the market given the future wage path w_t^+. Finally on notation, let $p^f(w^t)$ denote the probability assessment of firm f with a vacancy that a worker with wage history w^t is a better worker, that is, has productivity $\bar{z}$ rather than $\underline{z}$.

To demonstrate that the specified wage path is a perfect Bayesian equilibrium, we use strategies that consist of the following responses.

Definition 1 *The response of a firm with a vacancy at t is* Bertrand *if it offers the worker the highest wage at which the expected present value of its future profits is non-negative.*

Definition 2 *Suppose w_t^+ is an equilibrium wage path from $t + dt$ on. An employing firm's response at t is* matching *if, at stage n.3 for $n = t$, the firm offers a worker of type z: (1) a wage w that matches the higher of $w^0(w_t^+)$ and the best outside offer if $\Pi_t^*(z, w, w_t^+) \geq e^{-rdt}\bar{\Pi}_{t+dt}(w_t^+)$; (2) any wage strictly less than w otherwise.*

Definition 3 *Suppose w_t^+ is an equilibrium wage path from $t + dt$ on. The worker's response at t is* myopic *if, at stage n.4 for $n = t$, the worker accepts the highest wage w from among those available if $w \geq w^0(w_t^+)$ and no offer if $w < w^0(w_t^+)$.*

Definition 4 *The worker's response at t is* non-discriminating *if the following conditions hold. (1) If not quitting for exogenous reasons at t, the worker stays with the current employer if the current employer has matched the best outside offer. (2) When quitting at t, the worker chooses a new employer by randomising with equal probabilities between the best equal outside offers.*

This last response is non-discriminating in the sense that the worker treats equally all firms with vacancies that make equally good offers.

We next show that strategies consisting of the Bertrand, matching and myopic/non-discriminating responses for, respectively, firms with vacancies, the employing firm, and the worker form a perfect Bayesian equilibrium for off-equilibrium-path beliefs that are reasonable in the following sense.

Definition 5 *Suppose the worker has wage path $(w_0, \ldots, w_{T-dt})$ in a perfect Bayesian equilibrium. The beliefs of firm f with a vacancy in period $t + 1$ are* reasonable *if (1) for a wage at t of $w' < w_t$, $p^f(w^t, w') \leq p^f(w^t, w_t)$, and (2) for a wage at t of $w' > w_t$, $p^f(w^t, w') \geq p^f(w^t, w_t)$.*

For stating the result, it is useful to define the function $G(w, w_t^+)$ by

$$G_t(w, w_t^+) = p\Pi_t^*(\bar{z}, w_t, w_t^+) + (1 - p)\Pi_t^*(\underline{z}, w_t, w_t^+) - \varepsilon - e^{-rdt}\bar{\Pi}_{t+dt}(w_t^+),$$

$$\text{for } t \in \{dt, 2dt, \ldots, T - dt\}. \quad (50)$$

This is the gain in expected future profits of a firm with a vacancy that acquires, at date t and wage w, a worker who is type $\bar{z}$ with probability p over the expected future profits it would make if it did not acquire the worker at t.

Proposition 5 *Strategies consisting of the Bertrand, matching and myopic/non-discriminating responses for, respectively, firms with vacancies, the employing firm, and the worker form a perfect Bayesian equilibrium with reasonable beliefs for dt sufficiently small. For the wage path w_t in any such equilibrium, the following properties hold for all $t \in \{dt, 2dt, \ldots, T - dt\}$:*

1. *$p^f(w^t) = p$ for all firms f with vacancies at t.*

2. *If there exists a $w \in \left[w^0(w_t^+), \underline{w}(w_t^+)\right]$ such that $G_t(w, w_t^+) \geq 0$:*

$$w_t = \max w \in \left[w^0(w_t^+), \underline{w}(w_t^+)\right] \text{ such that } G_t(w, w_t^+) \geq 0. \tag{51}$$

3. *If there exists no $w \in \left[w^0(w_t^+), \underline{w}(w_t^+)\right]$ such that $G_t(w, w_t^+) \geq 0$:*

$$w_t = w^0(w_t^+). \tag{52}$$

Proof. Consider the last hiring period $n = T - dt$. With no future to be concerned about, the unique best response of the worker at stage $n.4$ is clearly the myopic response of accepting the offer with the highest wage w for period $T - dt$ from among those available, or no offer if $w < u$. It cannot do better than by being non-discriminatory. Thus the myopic/non-discriminatory response is a best response at $T - dt$. Also with no future to be concerned about, $\bar{\Pi}_T(w_{T-dt}^+) = 0$. Thus, a best response at stage $n.3$ for the incumbent employer, given that it knows the worker's productivity z, is the matching response of offering w that matches the higher of the worker's reservation wage u and the best outside offer if $\Pi_{T-dt}^*(z, w, w_{T-dt}^+) \geq 0$, or any wage offer less than that w if $\Pi_{T-dt}^*(z, w, w_{T-dt}^+) < 0$.

Consider now the best responses of firms with vacancies making offers to the worker at stage $n.2$ for $n = T - dt$. The only offers consistent with the claimed equilibrium are offers $w \leq \underline{w}(w_{T-dt}^+)$ because otherwise, given the best responses of the incumbent employer and the worker, the worker will certainly receive a wage $w_{T-dt} > \underline{w}(w_{T-dt}^+)$. Suppose firm f with a vacancy, and belief $p^f(w^{T-dt})$ that the worker is type $\underline{z}$, deviates by offering $w > \underline{w}(w_{T-dt}^+)$. Whatever its beliefs $p^f(w^{T-dt})$ about the worker's type, $w \geq \bar{w}(w_{T-dt}^+)$ is clearly not a best response because $\Pi_{T-dt}^*(z, w, w_{T-dt}^+) \leq 0$ for both $\bar{z}$ and $\underline{z}$ for any $w \geq \bar{w}(w_{T-dt}^+)$ so, given the hiring cost ε, it would certainly achieve a higher payoff by bidding $w < u$ and being sure not to hire the worker. For an offer w such that $\underline{w}(w_{T-dt}^+) < w < \bar{w}(w_{T-dt}^+)$, it follows from the definitions of $\underline{w}(w_t^+)$ and $\bar{w}(w_t^+)$ that the incumbent employer matches w for a $\bar{z}$ type worker but not for a $\underline{z}$ type worker. Thus firm f acquires a type $\bar{z}$ worker if quitting for exogenous reasons (that is, with probability $(1 - e^{-\rho dt}) p^f(w^{T-dt})$) and a type $\underline{z}$

worker whether or not quitting for exogenous reasons (that is, with probability $1 - p^f\left(w^{T-dt}\right)$). The probability it acquires a worker of either sort (and so incurs the hiring cost ε) is thus $\left(1 - e^{-\rho dt}\right) p^f\left(w^{T-dt}\right) + \left(1 - p^f\left(w^{T-dt}\right)\right) = 1 - e^{-\rho dt} p^f\left(w^{T-dt}\right)$. With the remaining probability it does not acquire the worker and receives zero flow payoff from $T - dt$ to T, so its payoff from the deviation is $\hat{\Pi}^f_{T-dt}\left(w^{T-dt}, w, w^+_{T-dt}\right)$ given by

$$\hat{\Pi}^f_{T-dt}\left(w^{T-dt}, w, w^+_{T-dt}\right) = \left[\left(1 - e^{-\rho dt}\right) p^f\left(w^{T-dt}\right) \Pi^*_{T-dt}\left(\bar{z}, w, w^+_{T-dt}\right)\right.$$
$$\left. + \left(1 - p^f\left(w^{T-dt}\right)\right) \Pi^*_{T-dt}\left(\underline{z}, w, w^+_{T-dt}\right) - \left(1 - e^{-\rho dt} p^f\left(w^{T-dt}\right)\right) \varepsilon\right]. \quad (53)$$

Now, it follows from (48) and $\bar{\Pi}_T\left(w^+_{T-dt}\right) = 0$ that $\Pi^*_{T-dt}\left(\underline{z}, w, w^+_{T-dt}\right) < 0$ for any $w > \underline{w}\left(w^+_{T-dt}\right)$. Moreover, as $dt \to 0$, $\left(1 - e^{-\rho dt}\right) \to 0$, whereas $\left(1 - e^{-\rho dt} p^f\left(w^{T-dt}\right)\right) \to \left(1 - p^f\left(w^{T-dt}\right)\right) > 0$ for $p^f\left(w^{T-dt}\right) < 1$. It thus follows that

$$\lim_{dt \to 0} \hat{\Pi}^f_{T-dt}\left(w^{T-dt}, w, w^+_{T-dt}\right) = \left(1 - p^f\left(w^{T-dt}\right)\right) \lim_{dt \to 0} \left[\Pi^*_{T-dt}\left(\underline{z}, w, w^+_{T-dt}\right) - \varepsilon\right]$$
$$< 0, \quad (54)$$

the strict inequality holding even if $\Pi^*_{T-dt}\left(\underline{z}, w, w^+_{T-dt}\right) \to 0$ as $w \to \underline{w}\left(w^+_{T-dt}\right)$. Thus, there exists $\delta > 0$ such that, for any $dt < \delta$, $\hat{\Pi}^f_{T-dt}\left(w^{T-dt}, w, w^+_{T-dt}\right) < 0$ for any deviation $w > \underline{w}\left(w^+_{T-dt}\right)$. Any such deviation is not, therefore, a best response for dt sufficiently small.

For the claimed equilibrium $p^f\left(w^{T-dt}\right) = p$ for all f. In that case, playing the Bertrand response is a best response by the standard argument for Bertrand games. That involves offering the highest w such that $G_{T-dt}\left(w, w^+_{T-dt}\right) \geq 0$ if that results in $u\left[= w^0\left(w^+_{T-dt}\right)\right] \leq w \leq \underline{w}\left(w^+_{T-dt}\right)$ and offering some $w < u$ otherwise. Given the matching and myopic responses of the incumbent employer and the worker, the wage w_{T-dt} actually paid then satisfies Properties 2 and 3 of the proposition. Note that, if any play before $T - dt$ were to result in $p^f\left(w^{T-dt}\right) \neq p$, the wage offer by firm f would be non-decreasing in $p^f\left(w^{T-dt}\right)$.

Now consider $n = T - 2dt$. Given the equilibrium plays at $T - dt$, the only reason the worker might deviate from the myopic response by turning down the highest wage offer for period $T - 2dt$ is if it would alter the beliefs of outside firms at $T - dt$ in such a way as to increase the wage w_{T-dt}. The possible deviations are for the worker to accept a lower offer or no offer ($w_{T-2dt} = 0$). But, given that the wage offer at $T - dt$ by outside firm f is non-decreasing in $p^f\left(w^{T-dt}\right)$, for reasonable beliefs as defined such a deviation is worse for the present and no better for the future, so it is not a best response. The worker still cannot do better than by being non-discriminatory. Thus the myopic/non-discriminatory response is a best response at $T - 2dt$. Similarly, the current employer's best response continues to be the matching response. If the higher of the worker's reservation wage and the best outside offer is w, if $\Pi^*_{T-2dt}\left(z, w, w^+_{T-2dt}\right) \geq e^{-rdt} \bar{\Pi}_{T-dt}\left(w^+_{T-2dt}\right)$, and if the current employer offers less than w, it loses the worker and so reduces its payoff, thus doing worse than by

matching. If, on the other hand, it offers more than w, that reduces profits in the current period. Moreover, given reasonable beliefs, it does not reduce the wages outside firms will offer for that worker at $T - dt$. Thus the incumbent employer offers no more than w. If $\Pi^*_{T-2dt}\left(z, w, w^+_{T-2dt}\right) < e^{-rdt}\bar{\Pi}_{T-dt}\left(w^+_{T-2dt}\right)$, only an offer less than w is a best response. But any such offer is a matching response as defined.

Consider now the best responses of firms with vacancies making offers to the worker at stage $n.2$ for $n = T - 2dt$. The only offers consistent with the claimed equilibrium are offers $w \leq \underline{w}\left(w^+_{T-2dt}\right)$ because otherwise, given the best responses of the incumbent employer and the worker, the worker will certainly receive a wage $w_{T-2dt} > \underline{w}\left(w^+_{T-2dt}\right)$. Suppose firm f with a vacancy, and belief $p^f\left(w^{T-2dt}\right)$ that the worker is type $\underline{z}$, deviates by offering $w > \underline{w}\left(w^+_{T-2dt}\right)$. Whatever its beliefs $p^f\left(w^{T-2dt}\right)$ about the worker's type, $w \geq \bar{w}\left(w^+_{T-2dt}\right)$ is clearly not a best response because $\Pi^*_{T-2dt}\left(z, w, w^+_{T-2dt}\right) \leq e^{-rdt}\bar{\Pi}_{T-dt}\left(w^+_{T-2dt}\right)$ for both $\bar{z}$ and $\underline{z}$ for any $w \geq \bar{w}\left(w^+_{T-2dt}\right)$ so, given the hiring cost ε, it would certainly achieve a higher payoff by bidding $w < w^0\left(w^+_{T-2dt}\right)$ and being sure not to hire the worker. For an offer w such that $\underline{w}\left(w^+_{T-2dt}\right) < w < \bar{w}\left(w^+_{T-2dt}\right)$, it follows from the definitions of $\underline{w}\left(w^+_t\right)$ and $\bar{w}\left(w^+_t\right)$ that the incumbent employer matches w for a $\bar{z}$ type worker but not for a $\underline{z}$ type worker. Thus firm f acquires a type $\bar{z}$ worker if quitting for exogenous reasons (that is, with probability $\left(1 - e^{-\rho dt}\right)p^f\left(w^{T-2dt}\right)$) and a type $\underline{z}$ worker whether or not quitting for exogenous reasons (that is, with probability $1 - p^f\left(w^{T-2dt}\right)$). The probability it acquires a worker of either sort (and so incurs the hiring cost ε) is thus $\left(1 - e^{-\rho dt}\right)p^f\left(w^{T-2dt}\right) + \left(1 - p^f\left(w^{T-2dt}\right)\right) = 1 - e^{-\rho dt}p^f\left(w^{T-2dt}\right)$. With the remaining probability it does not acquire the worker and receives zero payoff, so its payoff from the deviation is $\hat{\Pi}^f_{T-2dt}\left(w^{T-2dt}, w, w^+_{T-2dt}\right)$ given by

$$
\begin{aligned}
\hat{\Pi}^f_{T-2dt}\left(w^{T-2dt}, w, w^+_{T-2dt}\right) = & \left[\left(1 - e^{-\rho dt}\right)p^f\left(w^{T-2dt}\right)\Pi^*_{T-2dt}\left(\bar{z}, w, w^+_{T-2dt}\right)\right. \\
& + \left(1 - p^f\left(w^{T-2dt}\right)\right)\Pi^*_{T-2dt}\left(\underline{z}, w, w^+_{T-2dt}\right) \\
& \left. - \left(1 - e^{-\rho dt}p^f\left(w^{T-2dt}\right)\right)\varepsilon\right] + e^{-\rho dt}p^f\left(w^{T-2dt}\right)e^{-rdt}\bar{\Pi}_{T-dt}\left(w^+_{T-2dt}\right). \quad (55)
\end{aligned}
$$

Now, it follows from (48) that $\Pi^*_{T-2dt}\left(\underline{z}, w, w^+_{T-2dt}\right) < e^{-rdt}\bar{\Pi}_{T-dt}\left(w^+_{T-2dt}\right)$ for any $w > \underline{w}\left(w^+_{T-2dt}\right)$. Again, as $dt \to 0$, $\left(1 - e^{-\rho dt}\right) \to 0$, whereas $\left(1 - e^{-\rho dt}p^f\left(w^{T-2dt}\right)\right) \to \left(1 - p^f\left(w^{T-2dt}\right)\right) > 0$ for $p^f\left(w^{T-2dt}\right) < 1$. It thus follows that

$$
\begin{aligned}
\lim_{dt \to 0}\hat{\Pi}^f_{T-2dt}\left(w^{T-2dt}, w, w^+_{T-2dt}\right) = & \lim_{dt \to 0}\left\{e^{-rdt}\bar{\Pi}_{T-dt}\left(w^+_{T-2dt}\right)\right. \\
& + \left(1 - p^f\left(w^{T-2dt}\right)\right)\left[\Pi^*_{T-2dt}\left(\underline{z}, w, w^+_{T-2dt}\right) - \varepsilon\right. \\
& \left.\left. - e^{-rdt}\bar{\Pi}_{T-dt}\left(w^+_{T-2dt}\right)\right]\right\} \quad (56) \\
& < \lim_{dt \to 0}e^{-rdt}\bar{\Pi}_{T-dt}\left(w^+_{T-2dt}\right), \quad (57)
\end{aligned}
$$

the strict inequality holding even if $\Pi^*_{T-2dt}\left(\underline{z}, w, w^+_{T-2dt}\right) \to e^{-rdt}\bar{\Pi}_{T-dt}\left(w^+_{T-2dt}\right)$ as $w \to \underline{w}\left(w^+_{T-2dt}\right)$. Thus, there exists $\delta > 0$ such that, for any $dt < \delta$, $\hat{\Pi}^f_{T-2dt}\left(w^{T-2dt}, w, w^+_{T-2dt}\right) < e^{-rdt}\bar{\Pi}_{T-dt}\left(w^+_{T-2dt}\right)$ for any deviation $w > \underline{w}\left(w^+_{T-2dt}\right)$. Thus any such deviation gives a lower payoff than making an offer that the worker will never accept, and is therefore not a best response, for dt sufficiently small.

For the claimed equilibrium $p^f\left(w^{T-2dt}\right) = p$ for all f. In that case, playing the Bertrand response is a best response by the standard argument for Bertrand games. That involves offering the highest w such that $G_{T-2dt}\left(w, w^+_{T-2dt}\right) \geq 0$ if that results in $w^0\left(w^+_{T-2dt}\right) \leq w \leq \underline{w}\left(w^+_{T-2dt}\right)$ and offering some $w < w^0\left(w^+_{T-2dt}\right)$ otherwise. Given the matching and myopic responses of the incumbent employer and the worker, the wage w_{T-2dt} actually paid then satisfies Properties 2 and 3 of the proposition. Note that, if any play before $T - 2dt$ were to result in $p^f\left(w^{T-2dt}\right) \neq p$, the wage offer by firm f would be non-decreasing in $p^f\left(w^{T-2dt}\right)$.

Backwards iteration of the above argument for earlier dates establishes that the myopic and non-discriminating response by the worker, the matching response by the current employer, and the Bertrand response for firms with vacancies are all best responses for each date $t \in \{dt, 2dt, \ldots, T - dt\}$, conditional on $p^f\left(w^t\right) = p$, and result in a wage path that satisfies Properties 2 and 3 of the proposition.

It is common knowledge at $t = 0$ that the worker is type $\bar{z}$ with probability $p \in (0, 1)$. It follows that, in the first hiring date $t = dt$, $p^f\left(w^{dt}\right) = p \in (0, 1)$ for all firms f with a vacancy. Along the equilibrium path, the wage set at this date is $w_{dt} \leq \underline{w}\left(w^+_{dt}\right)$ by Properties 2 and 3 of the proposition and the worker quits only for exogenous reasons. Since the probability of exogenous quits is the same for both types, whether or not the worker quits provides no information about type. Thus, Bayesian updating implies $p^f\left(w^{2dt}\right) = p \in (0, 1)$ for all firms f with a vacancy. A similar argument applies at each hiring date t. Thus, along an equilibrium path $p^f\left(w^t\right) = p$ for all firms f with a vacancy and Property 1 of the proposition holds for all $t \in \{dt, 2dt, \ldots, T - dt\}$. ∎

This result has an important consequence. It is not just that strategies consisting of the Bertrand, matching and myopic/non-discriminating responses for, respectively, firms with vacancies, the incumbent employer, and the worker form a perfect Bayesian equilibrium. It is also that, in any such equilibrium, a firm with a vacancy at t never bids more for the worker than $\underline{w}\left(w^+_t\right)$, the wage that would be matched by the current employer for a worker of the less good type, if the time interval between offers is sufficiently short. The intuition is as follows. When the employing firm plays the matching response and the worker the myopic response, a bidding firm acquires the worker only at a wage for which the worker's future profitability is negative—unless, that is, the worker is quitting for exogenous reasons. For an offer above $\underline{w}\left(w^+_t\right)$, it acquires the worker only if of type $\underline{z}$ or if quitting for exogenous reasons. In the former case, the acquiring firm makes a loss in period t and no additional profit subsequently. But as the time interval dt goes to zero, the probability that a worker quits for exogenous reasons

becomes negligible. Thus, as long as there is a hiring cost, the acquiring firm necessarily loses by offering more than $\underline{w}\left(w_t^+\right)$.

Property 2 in Proposition 5 is the standard condition for an interior solution given the restriction to wage offers at t no more than $\underline{w}\left(w_t^+\right)$. Bertrand competition between firms with vacancies at t pushes the wage offers up to the highest level at which a firm that acquires the worker makes at least as much expected profit as from the alternative of making an offer that is refused for sure and then starting the next period dt later with a vacancy. Since, however, it is not an equilibrium to offer $w > \underline{w}\left(w_t^+\right)$ at t, firms with vacancies may make strictly positive expected profits from acquiring the worker at the equilibrium wage $\underline{w}\left(w_t^+\right)$ because the wage cannot be pushed higher to remove those profits. Formally, there is a discontinuity in the profit function at $\underline{w}\left(w_t^+\right)$ because, for any higher wage, the worker always accepts if of type $\underline{z}$, not just if quitting for exogenous reasons. Property 3 in Proposition 5 recognises that, because of the hiring cost ε, it may not be profitable for firms with vacancies to bid even the reservation wage $w^0\left(w_t^+\right)$ for the worker. Then the incumbent employer simply offers $w_t = w^0\left(w_t^+\right)$. (That $\underline{z} > u$ rules out the possibility that there is no wage at which employment can continue in the currently employing firm.)

It can in fact be shown (though at considerably greater length) that any wage path in a perfect Bayesian equilibrium with reasonable beliefs and a non-discriminating response by the worker satisfies Properties 1–3 of Proposition 5 for dt sufficiently small. Even if the worker's response is not non-discriminating, any perfect Bayesian equilibrium with reasonable beliefs can, for dt sufficiently small, be described by Properties 1–3 of Proposition 5 with the function $G\left(.\right)$ appropriately modified.[22]

The expression in (7) in the main text for the remaining lifetime earnings of a trainee who quits is for the limiting case as the interval dt between wage offers and the hiring cost ε go to zero, while the number of firms with vacancies m goes to infinity. In taking these limits, we specify that the limit first be taken as $dt \to 0$ in order to ensure that dt is always sufficiently small for the results of Proposition 5 to apply.

---

[22]In the game to which Proposition 5 applies, a firm quit by the worker at a previous hiring date forgets the worker's type. What happens if that is not the case? If two firms both know the worker's type at $t + dt$, Bertrand competition results in the worker receiving a wage equal to her true marginal product from $t + dt$ on. If the incumbent employer adopts the matching response at t, it is always worthwhile for the worker to quit at t because a second firm will then learn the worker's type, which results in a higher wage from $t + dt$ on, without the worker incurring a lower wage from t to $t + dt$. With a positive hiring cost, however, firms with vacancies offer a wage at t strictly less than $\underline{w}(w_t^+)$ for at least part of the working lifetime, and for all of it if that lifetime is sufficiently short. The best response of the incumbent employer may then be to more than match the best outside offer so that the worker incurs a short-term loss from quitting. As long as a wage at $t - dt$ no higher than $\underline{w}(w_{t-dt}^+)$ is enough to stop the worker quitting, a wage path satisfying Properties 1–3 of Proposition 5 remains a perfect Bayesian equilibrium. That will not be the case, however, as both the interval between hiring dates and the hiring cost go to zero because then both the difference between $\underline{w}(w_{t-dt}^+)$ and the best outside offer, and the period for which this difference can be paid, also go to zero. We do not explore this possibility further here because it requires moving away from the limiting case as dt and ε go to 0 that is a very convenient simplification for deriving the results in the main text.

Definition 6 *A limiting equilibrium wage path* is the limit as $dt \to 0$, $m \to \infty$ and $\varepsilon \to 0$ *(with limits taken in that order) of a wage path that satisfies Properties 1–3 of Proposition 5.*

Proposition 6 *There is a unique limiting equilibrium wage path. It satisfies* $w_t = \underline{w}\left(w_t^+\right) = \underline{z}$ *for all $t \in (0, T]$.*

Proof. It follows from (46) and (47) that

$$\lim_{m \to \infty} \bar{\Pi}_t \left(w_t, w_t^+\right) = 0, \text{ for all } t \in (0, T], \tag{58}$$

$$\lim_{m \to \infty} \lim_{dt \to 0} \Pi_t^* \left(\underline{z}, w_t, w_t^+\right) = \int_t^T e^{-(r+\rho)(\theta-t)} \left(\underline{z} - w_\theta\right) d\theta, \text{ for all } t \in (0, T]. \tag{59}$$

Thus, from (50),

$$\lim_{m \to \infty} \lim_{dt \to 0} G_t \left(w_t, w_t^+\right) = \int_t^T e^{-(r+\rho)(\theta-t)} \left[\underline{z} + p\left(\bar{z} - \underline{z}\right) - w_\theta\right] d\theta - \varepsilon, \text{ for all } t \in (0, T]. \tag{60}$$

As $\varepsilon \to 0$, therefore, the highest wage at t for which $\lim_{m \to \infty} \lim_{dt \to 0} G_t \left(w_t, w_t^+\right) \geq 0$ for all t approaches $\hat{w}$ defined by

$$\hat{w} = \underline{z} + p\left(\bar{z} - \underline{z}\right). \tag{61}$$

By Property 2 of Proposition 5, this is an equilibrium wage if $\hat{w} \leq \underline{w}\left(w_t^+\right)$. From (48) and (58), we have that $\underline{w}\left(w_t^+\right)$ is set such that

$$\lim_{m \to \infty} \lim_{dt \to 0} \Pi_t^* \left(\underline{z}, \underline{w}\left(w_t^+\right), w_t^+\right) = 0, \text{ for all } t \in (0, T]. \tag{62}$$

Because this must hold for all t, the derivative of the left hand side with respect to t must also be zero for all t. That and (59) imply

$$\underline{z} - \underline{w}\left(w_t^+\right) + (r + \rho) \int_t^T e^{-(r+\rho)(\theta-t)} \left(\underline{z} - w_\theta\right) d\theta = 0, \text{ for all } t \in (0, T]. \tag{63}$$

It follows that $\underline{w}\left(w_T^+\right) = \underline{z}$. But $\bar{z} > \underline{z}$ and we know from Proposition 5 that the wage cannot exceed $\underline{w}\left(w_t^+\right)$ for any t. Hence, it must be that $w_T = \underline{z}$. It follows directly from repeating this argument for earlier t that $w_t = \underline{w}\left(w_t^+\right) = \underline{z}$ for all t. ∎

The intuition behind this result is as follows. As the number of firms with vacancies becomes large and the hiring cost becomes small, firms with vacancies would be prepared to pay up to the average productivity of the two types, $\underline{z} + p\left(\bar{z} - \underline{z}\right)$, to hire the worker if quitting for exogenous reasons. But the incumbent employer is willing to pay only up to $\underline{z}$ to retain a worker of the less productive type. So, if firms with vacancies offer more than $\underline{z}$, a less productive worker will always accept, not just if quitting for exogenous reasons, and whichever

39

outside firm acquires the worker makes a loss on that type. That swamps the higher productivity of the more productive type quitting for exogenous reasons. Thus, outside offers never rise above $\underline{z}$ even though any outside firm that acquires the worker at that wage makes a strictly positive profit from doing so.

The expression in (7) for the remaining lifetime earnings of a trainee who quits follows directly from Proposition 6 in view of the definition of $\underline{z} = w^0 + \gamma g(c)$.

B The incentive constraints for $t > 0$

This appendix shows that, given any $W(0, \tau)$ that satisfies (23) for $t = 0$, it is always possible to find a wage path $w(t)$ that satisfies (23) for all $0 < t \leq \tau$ with $w(0) > w^0$ and with the training firm wishing to retain less good trainees up to τ.

First note that, since $w(t)$ is a flow, its value at the isolated point $t = 0$ is negligible relative to its present discounted value over the interval τ. Thus, we can set $w(0) = w^0 + \varepsilon$ for $\varepsilon > 0$ without affecting subsequent values of $w(t)$ for given $W(0, \tau)$. Let the wage over the remainder of the contract period be constant at

$$w = \frac{W(0, \tau)}{\delta(\tau)} \geq \left[w^0 + \gamma g(c)\right] - (1 - \gamma) g(c) e^{-\rho \tau} \frac{\beta(T) - \beta(\tau)}{\delta(\tau)}, \tag{64}$$

where the inequality follows from (23) being satisfied at $t = 0$. Then, from (8),

$$W(t, \tau) = \delta(\tau - t) w \tag{65}$$
$$\geq \delta(\tau - t) \left[w^0 + \gamma g(c)\right] - (1 - \gamma) g(c) e^{-\rho \tau} \frac{\delta(\tau - t)}{\delta(\tau)} \left[\beta(T) - \beta(\tau)\right] \tag{66}$$

Since, by construction, this satisfies (23) for $t = 0$, it is clear by inspection that it will also satisfy (23) for $0 < t \leq \tau$ if

$$e^{-\rho \tau} \frac{\delta(\tau - t)}{\delta(\tau)} \left[\beta(T) - \beta(\tau)\right] \leq e^{-\rho(\tau - t)} \left[\beta(T - t) - \beta(\tau - t)\right], \text{ for } 0 < t \leq \tau, \tag{67}$$

or

$$\frac{\delta(\tau - t)}{\delta(\tau)} \leq e^{\rho t} \frac{\beta(T - t) - \beta(\tau - t)}{\beta(T) - \beta(\tau)}, \text{ for } 0 < t \leq \tau, \tag{68}$$

or, using the definitions of $\beta(t)$ and $\delta(t)$ in (1) and (2),

$$\frac{1 - e^{-(r+\rho)(\tau - t)}}{1 - e^{-(r+\rho)\tau}} \leq e^{(r+\rho)t}, \text{ for } 0 < t \leq \tau,$$

or

$$e^{(r+\rho)t} \geq 1, \text{ for } 0 < t \leq \tau.$$

This is certainly true given r and $\rho \geq 0$.

We next show that, at the wage in (64), the training firm wishes to retain even less good trainees up to τ. To see this, note that the training firm would wish to continue the contract after $0 < t \leq \tau$ even for type γ trainees provided

$$\delta(\tau - t)\left[w^0 + \gamma g(c)\right] - W(t, \tau) + e^{(r+\rho)t}S \geq 0 \text{ for } 0 < t \leq \tau. \tag{69}$$

Now consider (69) for the total wage bill given by (24) so that (23) holds with equality at $t = 0$ and $W(t, \tau)$ is given by equality in (66). Then (69) can be written

$$e^{(r+\rho)t}S \geq -(1-\gamma)g(c)e^{-\rho\tau}\frac{\delta(\tau - t)}{\delta(\tau)}\left[\beta(T) - \beta(\tau)\right], \text{ for } 0 < t \leq \tau, \tag{70}$$

which, since $\gamma < 1$ and $\beta(T) \geq \beta(\tau)$, is clearly true for any $S \geq 0$. Finally, suppose that (23) holds with strict inequality, in which case any optimal apprenticeship has $W(0, \tau) = \bar{W}$. Then the addition over the present discounted value of wages making (23) hold with equality that is required to make $W(0, \tau) = \bar{W}$ can all be paid to the worker on agreeing to the contract at $t = 0$ and wages for $0 < t \leq \tau$ paid as above. Then clearly (70) is unaffected and continues to hold.

C Training when it is efficient to train all workers

This appendix considers training outcomes when it is efficient to train all workers, that is $k_1^* \geq k_2^*$. Then $\bar{W}$, the lifetime value of wages that equates demand and supply for workers at initial hiring, can be greater than βw^0 because the supply of workers may be less than the demand for workers by firms wanting to train at the lifetime value of wages βw^0. A market equilibrium can be defined formally as follows. Let $\bar{k}(\bar{W})$ denote the highest fixed cost of training at which it is profitable to train with a profit maximizing training arrangement when the expected lifetime utility of workers hired in the market at age 0 is $\bar{W}$. Then $\bar{W}$ is a market equilibrium if it satisfies

$$\bar{k}(\bar{W}) \leq k_2^* \tag{71}$$
$$\bar{W} \geq \beta w^0 \tag{72}$$
$$\text{either } \bar{W} = \beta w^0 \quad \text{ or } \bar{k}(\bar{W}) = k_2^*. \tag{73}$$

(k_2^* was defined above as the highest capital cost that needs to be incurred for all workers to be trained.)

C.1 Training without contracts

Suppose $\bar{W}$ is sufficiently high that (10) is not a binding constraint at $t = 0$ for any level of c chosen by a training firm. What then limits how low the wage can go is that the expected utility of joining a training firm given by (9) with $t = 0$ is at least as great as $\bar{W}$. That is, expected wages from the training firm W must satisfy

$$W \geq \bar{W} - (\beta - \delta)\left[w^0 + \gamma g\left(c\right)\right]. \tag{74}$$

Substitution of this into the expression for a training firm's expected profits (11) allows those expected profits to be written

$$\Pi^1\left(k, c, \bar{W}\right) = \left[(1 - \gamma)\delta + \gamma\beta\right]g\left(c\right) + \beta w^0 - \bar{W} - (k + c) \tag{75}$$

for which the first order condition for the optimal level of training c^1 is

$$\left[\delta + \gamma\left(\beta - \delta\right)\right]g'\left(c^1\right) = 1. \tag{76}$$

It follows from (15) and strict concavity of $g\left(.\right)$ that $c^1 > c^0$, though still $c^1 < c^*$ defined in (4). The intuition is that, although a firm's profits from training are lower when wages are higher, the worker captures less of the *marginal* returns to additional training. Indeed, since $\beta > \delta$, a higher value of γ actually *increases* the amount of training, which approaches the efficient level c^* as $\gamma \to 1$. The reason is that more training enables those workers who quit for exogenous reasons to obtain a higher wage in another firm and, because (10) is not binding, the training firm is able to lower its wage as a result. As γ approaches 1, those quitting for exogenous reasons capture all the return on training after they quit, so their new employers receive no external benefit from the training.

If (10) at $t = 0$ is not binding at c^0 but becomes binding at some $c < c^1$, then c is the profit maximizing level of training for training firms to provide. Moreover, for (10) not to be a binding constraint for some training c, it must be that the expected lifetime value of wages firms must pay to hire trainees, $\bar{W}$, exceeds what trained workers can obtain by quitting, given by (7) for $t = 0$. That is, $\bar{W} \geq \beta\left[w^0 + \gamma g\left(c\right)\right]$. For $\bar{W} > \beta w^0$ to be an equilibrium, it must be that all workers are trained at that wage. (This is expressed formally by (73).) Since k_2^* is defined as the fixed cost that must be incurred by the marginal training firm when all workers are trained, it follows from (75) that all workers will actually be trained at this wage only if

$$k_2^* \leq \left[(1 - \gamma)\delta + \gamma\beta\right]g\left(c\right) + \beta w^0 - \bar{W} - c, \tag{77}$$

or, with $\bar{W}$ at the lowest level $\beta\left[w^0 + \gamma g(c)\right]$,

$$k_2^* \le (1 - \gamma)\,\delta g(c) - c. \tag{78}$$

Thus, if (78) is satisfied for $c = c^0$, the market equilibrium level of training is the higher level that satisfies (78) with equality provided that level is less than c^1, and c^1 otherwise. The analysis in the main text applies whenever (78) is not satisfied at $c = c^0$. In view of the definition of k_1^* in (5), the condition for that can be written

$$k_1^* < k_2^* + \left[\beta g(c^*) - c^*\right] - (1 - \gamma)\,\delta g\left(c^0\right) + c^0. \tag{79}$$

For higher values of k_1^*, all workers are trained and the amount of training received by each trainee is the lower of c^1 and the amount that gives equality in (78).

C.2   Training contracts

Suppose $\bar{W}$ is sufficiently high that (23) is not a binding constraint at $t = 0$ for any level of c chosen by a training firm. That does not alter the analysis of the optimal training for any given contract, so (18) still applies. What, however, now limits how low the wage can go is that the expected utility of joining a training firm given by (22) with $t = 0$ is at least as great as $\bar{W}$. That is, expected wages W from the contract must satisfy

$$W \ge \bar{W} - \left[w^0 + \gamma g(c)\right]\left[\beta - \delta(\tau)\right] - (1 - \gamma)\,g(c)\,e^{-\rho\tau}\left[\beta - \beta(\tau)\right]. \tag{80}$$

Substitution of this into the expression for a training firm's expected profits (17), when there is no subsidy ($S = 0$) and training length $\tau(c)$ is chosen to induce training c, allows those expected profits to be written

$$\tilde{\Pi}^1\left(k, c, \bar{W}\right) = \left[(1 - \gamma)\,\tilde{\delta}(\tau(c)) + \gamma\beta\right]g(c) + \beta w^0 - \bar{W} - (k + c), \tag{81}$$

where $\tilde{\delta}(\tau)$ is defined in (26). Since, from (28), $\tilde{\delta}(\tau) > \delta$ for $\tau < T$, expected profits from offering a training contract given by (81) are greater than those from not offering a training contract given by (75) for $c < \bar{c}$ defined by (19), and are the same for $c = \bar{c}$, exactly as in the case of $k_1^* < k_2^*$ discussed in the main text. The difference in the case $k_1^* \ge k_2^*$ is that the level of training that is optimal with no contract if (78) is satisfied with strict inequality for $c = \bar{c}$ is not feasible with a training contract, for which the highest feasible level of training is $\bar{c}$ that is attained when $\tau = T$. Moreover, that optimal level yields higher expected profit than $c = \bar{c}$ when no contract is used and, since the expected profit with $c = \bar{c}$ is the same with a contract as without, it also yields higher profit than $c = \bar{c}$ with a contract. Thus the argument in the main text does not guarantee that firms make greater expected profit with a contract than without in

43

this case.

The argument in the main text does, however, apply as long as the level of training chosen without a contract is less than $\bar{c}$. From (78), that will be the case whenever

$$k_2^* > (1 - \gamma)\,\delta g\,(\bar{c}) - \bar{c}, \tag{82}$$

or, equivalently,

$$k_1^* < k_2^* + \left[\beta g\,(c^*) - c^*\right] - (1 - \gamma)\,\delta g\,(\bar{c}) + \bar{c}.$$

When this condition is satisfied, it is certainly more profitable for firms to use training contracts even though it is efficient to train all workers.

References

[1] Acemoglu, Daron. "Credit Constraints, Investment Externalities and Growth." In Booth and Snower (1996).

[2] Acemoglu, Daron. "Training and Innovation in an Imperfect Labour Market." *Review of Economic Studies,* 64 (July 1997), 445-464.

[3] Acemoglu, Daron and Pischke, Jörn-Steffen. *The Structure of Wages and Investment in General Training.* Massachusetts Institute of Technology, Department of Economics (1997).

[4] Acemoglu, Daron and Pischke, Jörn-Steffen. "Why Do Firms Train? Theory and Evidence." *Quarterly Journal of Economics,* 113 (February 1998), 79–119.

[5] Becker, Gary. *Human Capital.* New York: Columbia University Press (1964).

[6] Bindoff, S. T. *Tudor England.* Harmondsworth: Penguin (1950).

[7] Booth, Alison L. and Chatterji, Monojit. *Training and Contracts.* Working Paper 95-15, Research Centre on Micro-Social Change, University of Essex (1995).

[8] Booth, Alison L. and Chatterji, Monojit. *Training and Unions.* CEPR Discussion Paper 1573 (February 1997).

[9] Booth, Alison L. and Snower, Dennis J. *Acquiring Skills.* Cambridge: Cambridge University Press (1996).

[10] Burdett, Kenneth and Smith, Eric. "Education and Matching Externalities." In Booth and Snower (1996).

[11] Cantor, Richard. "Firm Specific Training and Contract Length." *Economica,* 57 (February 1990), 1–14.

[12] Chang, Chun and Wang, Yijiang. "Human Capital Investment under Asymmetric Information: The Pigovian Conjecture Revisited." *Journal of Labor Economics,* 14 (July 1996), 505–519.

[13] Cunningham, W. and McArthur, E. A. *Outlines of English Industrial History.* Cambridge: Cambridge University Press (1920).

[14] Dolton, Peter J. "The Economics of Youth Training in Britain." *Economic Journal,* 103 (September 1993), 1261–1278.

[15] Earle, P. *The Making of the English Middle Class, 1660-1730.* London: Methuen (1989).

[16] Elbaum, Bernard. "Why Apprenticeship Persisted in Britain But Not in the United States." *Journal of Economic History,* 49 (June 1989), 337–349.

[17] Finegold, D. and Soskice, D. "The Failure of British Training: Analysis and Prescription." *Oxford Review of Economic Policy,* 4(3) (Autumn 1988), 21–53.

[18] Greenwald, Bruce C. "Adverse Selection in the Labour market." *Review of Economic Studies,* 53(3) (July 1986), 325-347.

[19] Hamilton, Gillian. "Enforcement in Apprenticeship Contracts: Were Runaways a Serious Problem? Evidence from Montreal." *Journal of Economic History*, 55(3) (September 1995), 551-574.

[20] Hamilton, Gillian. "The Market for Montreal Apprentices: Contract Length and Information." *Explorations in Economic History*, 33 (1996), 496-523.

[21] Harhoff, Dietmar and Kane Thomas J. *Financing Apprenticeship Training: Evidence from Germany*, NBER Discussion Paper no.4557 (1993).

[22] Hermalin, Benjamin. "Adverse Selection, Short-Term Contracting, and the Underprovision of On-the-Job Training." Working Paper No. 90-139, Department of Economics, University of California at Berkeley (February 1990).

[23] Kahn, Charles and Huberman, Gur. "Two-Sided Uncertainty and "Up-or-Out" Contracts." *Journal of Labor Economics,* 6 (October 1988), 423-444.

[24] Katz, Eliakim and Ziderman, Adrian. "Investment in General Training: The Role of Information and Labour Mobility." *Economic Journal,* 100 (December 1990), 1147–1158.

[25] Laffont, Jean-Jacques and Tirole, Jean. *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press (1993).

[26] Lee, D., Marsden, Dennis, Rickman, Penny and Duncombe, Jean. *Scheming for Youth: A Study of YTS in the Enterprise Culture.* Buckingham: Open University Press (1990).

[27] Lees, Dennis and Chiplin, Brian "The Economics of Industrial Training." *Lloyds Bank Review,* 96 (1970), 29–41.

[28] Loewenstein, Mark A. and Spletzer, James R. "Dividing the Costs and Returns to General Training." *Journal of Labor Economics,* 16 (January 1998), 142–171.

[29] Lynch, Lisa M. "Private-Sector Training and the Earnings of Young Workers." *American Economic Review,* 82 (March 1992), 299–312.

[30] Lynch, Lisa M. "The Economics of Youth Training in the U.S." *Economic Journal,* 103 (September 1993), 1292-1302.

[31] Lynch, Lisa M. (ed.). *Training and the Private Sector: International Comparisons*, Chicago: University of Chicago Press (1994).

[32] Pirenne, Henri. *Economic and Social History of Medieval Europe.* London: Kegan Paul, Trench, Trubner (1936).

[33] Prendergast, Canice "The Role of Promotion in Inducing Specific Human Capital Acquisition." *Quarterly Journal of Economics,* 108 (May 1993), 523-534.

[34] Schlicht, Ekkehart. "Endogenous On-the-Job Training with Moral Hazard." *Labour Economics,* 3 (1996), 81–92.

[35] Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations.* London: Bell (1887).

[36] Soskice, David. "Reconciling Markets and Institutions: The German Apprenticeship System." In Lynch (1994), pp. 25–60.

[37] Steedman, Hilary. "The Economics of Youth Training in Germany." *Economic Journal,* 103 (September 1993), 1279–1291.

[38] Stevens, Margaret. "Labour Contracts and Efficiency in On-the-Job Training." *Economic Journal,* 104 (March 1994a), 408–419.

[39] Stevens, Margaret. "An Investment Model for the Supply of Training by Employers." *Economic Journal,* 104 (May 1994b), 556-570.

[40] Stevens, Margaret. "Transferable Training and Poaching Externalities." In Booth and Snower (1996).

[41] U.S. Department of Labor. *Report of the Commission on Workplace Quality and Labor Market Efficiency.* Washington: US GPO (1989).

[42] Waldman, Michael. ""Up-or-Out" Contracts: A Signaling Perspective." *Journal of Labor Economics,* 8 (April 1990), 230-250.

[43] Williamson, Oliver E. *The Economic Institutions of Capitalism.* New York: The Free Press (1985).