Rostocker Zentrum zur Erforschung des Demografischen Wandels
Rostock Center for the Study of Demographic Change
Konrad-Zuse-Strasse 1 · D-18057 Rostock · Germany
Tel.: + 49 (0) 381 2081 – 0 · Fax: +49 (0) 381 2081 – 202
www.rostockerzentrum.de

ROSTOCKER ZENTRUM – DISKUSSIONSPAPIER
ROSTOCK CENTER – DISCUSSION PAPER

No. 14

Team Composition and Knowledge Transfer within an Ageing Workforce

Michael Kuhn
Pascal Hetze

July 2007

Team Composition and Knowledge Transfer within an Ageing Workforce

Michael Kuhn
Max Planck Institute for Demographic Research and University of Rostock
kuhn@demogr.mpg.de

Pascal Hetze
Bundespraesidialamt
pascal.hetze@bpra.bund.de

ROSTOCKER ZENTRUM – DISKUSSIONSPAPIER
ROSTOCK CENTER – DISCUSSION PAPER

No. 14

Juli 2007

Rostocker Zentrum zur Erforschung des Demografischen Wandels
Rostock Center for the Study of Demographic Change
Konrad-Zuse-Strasse 1 · D-18057 Rostock · Germany
Tel.: + 49 (0) 381 2081 – 0 · Fax: +49 (0) 381 2081 – 202
www.rostockerzentrum.de

Accepted by the ‘editorial board’*

Team Composition and Knowledge Transfer within an Ageing Workforce*

Michael Kuhn† and Pascal Hetze‡

July 17, 2007

Abstract

This paper examines the transfer of know-how from old/experienced workers to their junior co-workers and how it is affected by the ageing of the workforce. We consider an OLG framework, where agents from different age groups form partnerships/teams to produce some output. Where teams are composed of young workers and experienced old workers there is scope for a costly transfer of knowledge. We derive the team structure and training rates for the social optimum and for a decentral setting, where matching and training rates are determined by the interplay of supply and demand. We show under which conditions population ageing leads to a reduction in training and establish the decentral outcomes for the cases where fees are determined competitively and by bargaining, respectively. We assess the efficiency of the decentral outcomes and discuss how it depends on the age-structure. Our model lends itself to the analysis of knowledge transfer within professional partnerships.

Key-words: age-structure, knowledge transfer, matching, overlapping-generations, partnerships

JEL-classification: J24, J41

*We are grateful for helpful comments to Jim Vaupel, to the participants of the Rostock Demographic Colloquium and research seminars at the Universities of Rostock and Essen, Vienna Institute of Demography and McMaster University. All remaining errors are ours.

†Corresponding author: Max-Planck-Institute for Demographic Research and University of Rostock, Germany; e-mail: kuhn@demogr.mpg.de

‡Bundespraesidialamt.

1 Introduction

In the presence of population ageing there are strong concerns about (re-)integrating older workers into the labour market. While these concerns frequently relate to the sustainability of social security and the assurance of a sufficient labour supply within a shrinking workforce, another argument raised in the debate is that insufficient use is being made of the experience and know-how older workers can contribute to the running of firms and organisations. This paper explores this latter argument by analysing the formation of young and old (possibly experienced) workers to teams and the incentives that govern the transfer of knowledge from experienced old workers to young workers within these teams.

We consider an over-lapping generations (OLG) framework, where agents from different (or possibly the same) age groups form partnerships/teams in order to produce some output. Whenever there is a match between young workers and experienced old workers there is also scope for the provision of training, albeit at a cost. We establish the socially efficient age-composition of partnerships and the efficient supply of training depending on the production and training technology, the discount rate and the age-structure of the profession (workforce).

We compare the social optimum with the allocation within a decentralised economy, where experienced old agents have to be given incentives to form teams with young partners and train them. Training is undertaken if and only if there exists scope for a transfer from young to old agents. Specifically this requires that the maximum willingness to pay for training of a young worker exceeds the minimum price charged by old workers. We show that the willingness to pay increases in the expectation of young workers to be able to sell training in their second period of life. This expectation depends on the demography of the workforce.

For a shrinking workforce there is a period-by-period over-supply of old workers – and thus a potential over-supply of training. While this may suggest that there should be sufficient scope for the transfer of knowedge, such an intuition is misleading. Indeed, the converse is tue. When there is an excess supply of (skilled) old workers not everyone can expect to find a trainee. While young workers benefit from this when purchasing training, they will anticipate that within an ageing workforce they will belong to the 'long-side' of the market once they have acquired skills. The expected future returns to training are adjusted downwards and, accordingly, the willingness to pay for training at present. Hence, the scope for training is curtailed within a shrinking work force.

The dependency of training on age-structure is particularly pertinent when either the fee for training is determined as the outcome of bargaining between representatives of the two age groups or when it is optimal to form mixed teams with skilled and unskilled partners even in the absence of training. In the latter case, two 'markets' interact, one for a place in a mixed team without training and one for training. We derive the decentral allocation depending on the discount rate, the production technology, the transfer structure, the age distribution and the process at which the fees are determined and show under which circumstances the rate of training is efficient. We find that if the training fee and where relevant the entry fee into mixed teams (without training) are subject to competition, an efficient training structure emerges. Inefficiencies arise whenever a uniform fee is set in the training sector as it gives rise to private returns to training. This pushes towards the over-provision of training. Under-provision may arise in the presence of bargaining if bargaining power is concentrated with either the young or the old. In such a case, the disadvantaged side will generally supply or demand only limited amounts of training so that rationing arises. In either case, a shrinking workforce increases the likelihood and extent of under-provision or otherwise reduce the extent of over-provision.

It is not always clear in the political debate as to what sort of know-how is being referred to and why older workers should have a particular advantage in providing it (as opposed to general education or learning-by-doing). Generally, one may think of problem solving-skills or social competencies. We believe that the handing-down of expertise and know-how plays a particularly important role in the context of professional partnerships including the free-professions (physicians, lawyers and/or consultants), craftsmen's workshops and scientific teams. Within these partnerships know-how and skills can frequently only be acquired through the collaboration with experienced partners *and/or* under their supervision (see e.g. Morrison and Wilhelm 2004).[1] For instance, medical know-how includes certain diagnostic or surgical skills related to rare conditions and/or complications. Almost by definition, the knowledge how to diagnose such conditions and/or how to deal with complications cannot be easily taught as part of the formal medical education but

[1] Such knowledge, which is difficult to codify is sometimes referred to as tacit knowledge (Polanyi, 1966).

requires the sustained collaboration between a junior and a senior partner.[2] Other examples relate to lawyers, where junior partners benefit from their senior collaborator's know-how of building a good case, of cross-examining witnesses and of making stick a final speech. In consultancy, senior partners may pass on important know-how about the acquisition of new clients or about how to competently deal with demanding clients. In science, junior researchers benefit from the experimental and paper writing skills of their senior counterparts. Thus, while we cast our analysis in the more general terms of 'training' and 'teams' and while we believe that our insights are of a general nature, our model lends itself most naturally to an economy composed of partnerships, in which trade-specific know-how is passed on by way of extensive interaction between a junior and a senior partner.

The issues of team formation and compensation within teams that are central to our model also feature in Legros and Newman (2003), who focus on the role of nontransferability of wealth between team-mates. This work is applied by Gall et al. (2006) to the timing of education (i.e. investments in productivity). While similar issues are touched upon, this work differs in substance. In Gall et al. (2006) education is attained outside the team and their timing relates to whether education is acquired before or after matching has occured. This is obviously different from our context where by definition training can only occur within a team. Furthermore, the model by Gall et al. (2006) is essentially static, in particular there is no OLG structure and the role of ageing is not addressed.[3] One work close in spirit to ours is Tykvova (2006), who studies the incentives for experienced partners to take on board and, implicitly, train junior partners. In her model, the team is a syndicate in the private equity and venture capital market. Young investors can benefit from the expertise of their established counterparts both by working in a more productive team and by being able to learn from them. In contrast to the static (one period) case, where experienced partners have no incentive to involve junior partners, heterogenous syndicates with skilled and unskilled investors can be found in the dynamic (two period) setting. Inexperienced investors initially accept comparably worse conditions with

[2] Of course, one could argue that to some extent this know-how can be acquired through learning-by-doing. However, the example drawn from medicine makes it very clear that the cost to learning-by-doing to the patient and physician may be very high.

[3] Besley and Ghatak (2006) study the effect of incentive contracts for teachers when teachers and students are matched to form 'educational teams'. Here, too, ages-structure plays no role.

respect to their payoff. Nonetheless, her model deviates substantially in that she, too, does not consider the dynamic incentives within an OLG economy consisting of many periods. Furthermore, she does not study the socially optimal allocation and cannot therefore address the efficiency of training.

Morrison and Wilhelm (2004) consider the incentives to pass on the non-contractible knowledge within partnerships. Thus, they have in mind a similar set-up as we do. However, they focus on an analysis how the structure of partnerships overcomes the moral hazard problem with regard to the provision of training. The argument is that only agents who have actually received training are willing to invest into the commonly illiquid shares of a partnership. If the partnership were taken over by untrained agents, it would lose reputation thus eroding the returns on the shares. In such a case senior partners have an incentive to provide effective training so as to ensure that they are able to profitably sell on their own illiquid shares. Morisson and Wilhelm consider a single infinitely lived firm and a stable population. Thus, they cannot address the role of age-structure and population shrinking. Furthermore, they do not consider the efficiency of training allocations. By identifying the mechanism by which training becomes enforceable within partnerships, however, Morrison and Wilhelm (2004) provide an important basis for our model, where we assume for simplicity the contractibility of training.[4]

The remainer of the paper is structured as follows. The next section presents the model. Section 3 examines the socially efficient team structure and derives the optimal rate of training. Section 4 derives the matching and training allocation in a decentralised economy, compares it with the social optimum and examines the impact of ageing. Section 5 concludes. Some of the more extensive proofs are contained in an appendix.

2 The model

We consider a profession consisting of over-lapping generations, where agents produce (and live) for two periods. Furthermore, agents may receive training

[4]In considering the role of demographic developments for the evolution of a profession over time our model also bears some resemblance to the literature on the demography of organisations (e.g. Keyfitz 1973, Vaupel 1981, Feichtinger et al. 2007). This literature is mostly concerned with the scope for promotion but does not consider the feedback on incentives such as those for training activities.

in the first period of their career. Thus, there are three types of agents who may form partnerships in any given period: young agents (y), old agents with skills (o^s), and old agents without skills (o^u). An agent can acquire skills only through receiving training while young. At the same time training can be provided only by old agents with skills. A partnership consists of two agents jointly producing output. Each partner can either have low (l) or high (h) productivity. Only skilled old agents, o^s, can be of type h, while the young, y, and unskilled old agents, o^u, are of type l. Partnerships composed of different types of agent can thus produce the following output levels

$$\begin{aligned}\theta_{l,l} &:= \theta(y,y) = \theta(o^u,o^u) = \theta(y,o^u),\\ \theta_{l,h} &:= \theta(o^s,o^u) = \theta(y,o^s),\\ \theta_{h,h} &:= \theta(o^s,o^s).\end{aligned}$$

We assume the natural ranking $0 \leq \theta_{ll} \leq \theta_{lh} \leq \theta_{hh}$. Furthermore, let us define

$$\widehat{\theta} = \theta_{h,h} + \theta_{ll} - 2\theta_{l,h} \lesseqgtr 0$$

as the net surplus of skill segretation. In analogy to the condition for assortative matching (Becker, 1973), it is then true that segregated teams are privately and socially preferrable to teams composed of skilled and unskilled workers (mixed teams) if and only if $\widehat{\theta} > 0$. This is the case if and only if the combined output of a team composed of high productivity workers only and a team composed of low productivity workers only exceeds the output of a couple of mixed teams.

Training will only take place between skilled old agents and young agents, thus only within (y, o^s) matches. Let δ denote the total costs associated with training a particular young worker. Young workers are heterogeneous with respect to their 'trainability', the latter reflecting their ability and motivation. Without loss of generality, we assume that δ is uniformly distributed on $[0, 1]$, where the most motivated/able can be trained at zero cost. The output in a team in which training takes place can thus be written as $\theta_{lh} - \delta$.

The total population of workers at time t is given by $N(t) = \sigma_o(t) + \sigma_y(t)$, where $\sigma_y(t)$ and $\sigma_o(t) = \sigma_{o^s}(t) + \sigma_{o^u}(t)$ denote the numbers of young and old workers, respectively. The population evolves as follows:

$$\sigma_o(t) = \sigma_y(t-1), \tag{1}$$

$$\sigma_y(t) = (1+n)\,\sigma_y(t-1). \tag{2}$$

Thus, all young workers survive into their second period and reproduce at a rate $1+n$, where n gives the net growth rate of the population. Dividing (2) through (1) gives

$$\frac{\sigma_y(t)}{\sigma_o(t)} = 1 + n =: \lambda$$

as a measure of the age-structure. Note that $\lambda < 1$ in a shrinking population, $\lambda = 1$ in a stable population and $\lambda > 1$ in a growing population. Using $\sigma_y(t) = \lambda\sigma_o(t)$ we can write $N(t) = (1+\lambda)\sigma_o(t) = \frac{1+\lambda}{\lambda}\sigma_y(t)$. We then obtain for the evolution of the population

$$\frac{N(t)}{N(t-1)} = \frac{(1+\lambda)\sigma_o(t)}{\frac{1+\lambda}{\lambda}\sigma_y(t-1)} = \lambda,$$

where the last equality follows from (1). Furthermore, it is easily verified that $N(t) = \lambda^t N_0$ and

$$\begin{aligned}\sigma_y(t) &= \frac{\lambda^{t+1}}{1+\lambda}N_0 \\ \sigma_o(t) &= \frac{\lambda^t}{1+\lambda}N_0,\end{aligned}$$

where $N(0) := N_0$ is the size of the initial population. In the following, we normalise $N_0 \equiv 1$.

Define $T(t) \in [0,1]$ as the share of young workers who receive training in period t. We then find $T(t)\sigma_y(t)$ as the incidence of training in period t so that

$$\begin{aligned}\sigma_{o^s}(t) &= T(t-1)\sigma_y(t-1) = T(t-1)\sigma_o(t), \\ \sigma_{o^u}(t) &= [1 - T(t-1)]\sigma_o(t), \\ \frac{\sigma_{o^s}(t)}{\sigma_o(t)} &= T(t-1).\end{aligned}$$

It follows that for a steady state with $T(t-1) = T(t) = T$ the share of skilled workers in the population is constant with $\frac{\sigma_{o^s}(t)}{\sigma_o(t)} = T$. Furthermore, minimising the cost of training $T(t)\sigma_y(t)$ workers implies that those young workers and only those should be trained for whom $\delta \leq T(t)$. Hence, we can write aggregate training costs as

$$\int_0^{T(t)} \delta d\delta = \frac{T(t)^2}{2}.$$

and marginal training costs as $T(t)$.

In the following section we use this framework to develop the training structure that maximises the present value of the flow of net production within the profession. Given that labour supply is exogenous in our model and no other inputs (or fixed quantities thereof) are used in production and assuming that output can be transferred freely between the agents, the maximisation of output over time corresponds to the maximisation of the flow of consumption and/or utility. For ease of reference, we call this allocation the 'social optimum' and use it as a benchmark for assessing potential inefficiencies within a decentralised economy.

3 Social optimum

In the following we focus on a weakly shrinking population. Thus, we consider $\lambda \in [0,1]$. Suppose there are $1+\bar{t}$, $\bar{t} \in [1,\infty)$ periods for which the economy exists. For the initial state in period $t=0$, we assume without loss of generality $N(0) \equiv 1$, $\sigma_y(0) = \frac{\lambda}{1+\lambda}$ and $\sigma_o(0) = \frac{1}{1+\lambda}$. Furthermore, we assume that $\sigma_{o^s}(0) = \sigma_o(0)$, i.e. that all old workers are skilled initially.[5] In the last period, $\bar{t}$, it is always optimal to set $T(\bar{t}) = 0$ in order to save the cost of training. We will then find the steady state value T^S that maximises the flow of net production over the interval $[0, \bar{t}-1]$ and show that it is, indeed, optimal for the planner to choose $T(t) = T^S$ in any given period $t \in [0, \bar{t}-1]$.

It is immediately clear that in the absence of training workers should be matched to teams according to the rule of assortative matching. Thus, segregated teams are preferred if and only if $\widehat{\theta} = \theta_{h,h} + \theta_{ll} - 2\theta_{l,h} > 0$. In the following we consider this case first.

[5] For $\sigma_{o^s}(0) < \sigma_o(0)$, a potential lack of skilled workers may lead to the incidence of training being bounded from above for an initial number of periods. For $\lambda < 1$ and $\bar{t}$ sufficiently large, the shrinking population will imply eventually that the supply of training outstrips the demand and the optimal rate of training will be realised as an interior solution.

3.1 Optimal training with assortative matching ($\widehat{\theta} > 0$)

Denoting $\rho := (1+r)^{-1}$ as the discount factor corresponding to an interest rate r, we can write the present value of the flow of net production as[6]

$$\Pi\left(T, \lambda, \bar{t}\right) = \left\{ \begin{array}{c} \frac{\theta_{h,h}-\theta_{l,l}}{2} + \frac{1-(\rho\lambda)^{\bar{t}+1}}{1-\rho\lambda}\left(1+\lambda\right)\frac{\theta_{l,l}}{2} \\ +\frac{1-(\rho\lambda)^{\bar{t}}}{1-\rho\lambda}\lambda T\left[\rho\frac{\theta_{h,h}-\theta_{l,l}}{2} - \frac{\widehat{\theta}+T}{2}\right] \end{array} \right\} \left(1+\lambda\right)^{-1}. \tag{3}$$

Differentiating with respect to T gives the first order condition

$$\begin{aligned} \frac{d\Pi\left(T, \lambda, \bar{t}\right)}{dT} &= \frac{1-(\rho\lambda)^{\bar{t}}}{1-\rho\lambda}\frac{\lambda}{1+\lambda}\left[\rho\frac{\theta_{h,h}-\theta_{l,l}}{2} - \frac{\widehat{\theta}+2T}{2}\right] = 0 \\ &\Leftrightarrow \; T = T^S = \rho\frac{\theta_{h,h}-\theta_{l,l}}{2} - \frac{\widehat{\theta}}{2}. \end{aligned} \tag{4}$$

Here, $\rho \in \left[\frac{\widehat{\theta}}{\theta_{h,h}-\theta_{l,l}}, \frac{\widehat{\theta}+1}{\theta_{h,h}-\theta_{l,l}}\right]$ guarantees an interior solution $T^S \in [0,1]$. The socially optimal rate of training increases with the discounted return $\rho\frac{\theta_{h,h}-\theta_{l,l}}{2}$ (per worker) from having a highly productive team instead of a skilled mixed team and decreases with the opportunity cost, amounting to the loss of output $\frac{\widehat{\theta}}{2}$ (per worker) as production takes place in mixed teams instead of the more productive segregated teams. Naturally, the optimal rate of training falls with the interest rate, but interestingly, it does not depend on the age structure of the population.

Finally, note that this result obtains irrespective of the number of periods, $\bar{t}$. But this implies that $T(t) = T^S$ constitutes not only the equilibrium over the full interval $\left[0, \bar{t}-1\right]$, but also for any sub-interval $\left[\widehat{t}, \bar{t}-1\right]$, with $\widehat{t} \in \left[0, \bar{t}-1\right]$. Thus, provided that $T(t) = T^S$ has been selected for $t \in \left[0, \widehat{t}-1\right]$, then $T(t) = T^S$ continues to be the optimum choice for the remaining time. But this implies that T^S is, indeed, a steady-state optimum over the full time path $\left[0, \bar{t}-1\right]$.[7]

[6] The derivation of equation (3) is provided in the Appendix.

[7] Alternatively, (4) can be derived by using recursively the Bellman-equation $\Pi\left(T(t), T(t-1), \lambda, t\right) + \rho\Pi\left(T(t+1), T(t), \lambda, t+1\right)$, with $\Pi\left(T(t), T(t-1), \lambda, t\right)$ denoting the net value of productino in period t. Solving the problem gives the time path for $T(t)$ for all $t \in \left[0, \bar{t}-1\right]$, where $T\left(\bar{t}\right) = 0$. It is easily checked that $T(t) = T^S$ for all $t \in \left[0, \bar{t}-1\right]$.

3.2 Optimal training with mixed teams ($\widehat{\theta} < 0$)

When $\widehat{\theta} = \theta_{h,h} + \theta_{ll} - 2\theta_{l,h} < 0$ it is optimal to form mixed teams even when training does not take place. As a preliminary, recall that for a steady-state value, T, we have in each period $t \in \left[0, \bar{t}\right]$ a number of $\sigma_{o^s}(t) = T\sigma_o(t)$ skilled agents facing $\sigma_{o^u}(t) + \sigma_y(t) = (1 - T + \lambda)\,\sigma_o(t)$ unskilled agents. As is readily checked, we then have

$$\sigma_{o^s}(t) \geq \sigma_{o^u}(t) + \sigma_y(t) \Leftrightarrow T \geq \frac{1+\lambda}{2}. \tag{5}$$

For $T < \frac{1+\lambda}{2}$ the number of unskilled agents (young or old) in each period exceeds the number of skilled agents. Thus, there is an insufficient number of skilled workers to form mixed teams and a number of unskilled-only teams are formed. In contrast, for $T > \frac{1+\lambda}{2}$ a surplus of skilled old workers is grouped into skilled-only teams within each period. For $T = \frac{1+\lambda}{2}$ (and for this value only) all agents are grouped into mixed teams (with or without training). As it turns out, the flow of net production depends on whether there is an excess or shortfall of skilled agents.

Defining $\underline{\Pi}\left(T, \lambda, \bar{t}\right) := \Pi\left(T, \lambda, \bar{t}\right) \left| T \in \left[0, \frac{1+\lambda}{2}\right]\right.$ and

$\overline{\Pi}\left(T, \lambda, \bar{t}\right) := \Pi\left(T, \lambda, \bar{t}\right) \left| T \in \left[\frac{1+\lambda}{2}, 1\right]\right.$, we can write the net present value of output as[8]

$$\underline{\Pi}\left(T, \lambda, \bar{t}\right) = \left\{ \begin{array}{c} (1-\lambda)\frac{\theta_{h,h}}{2} + \lambda\theta_{l,h} + \frac{1-(\rho\lambda)^{\bar{t}}}{1-\rho\lambda}\rho\lambda(1+\lambda)\frac{\theta_{l,l}}{2} \\ +\frac{1-(\rho\lambda)^{\bar{t}}}{1-\rho\lambda}\lambda T\left[\rho\left(\theta_{l,h} - \theta_{l,l}\right) - \frac{T}{2}\right] \end{array} \right\} (1+\lambda)^{-1} \tag{6}$$

and

$$\overline{\Pi}\left(T, \lambda, \bar{t}\right) = \left\{ \begin{array}{c} (1-\lambda)\frac{\theta_{h,h}}{2} + \lambda(\rho\lambda)^{\bar{t}}\theta_{l,h} \\ +\frac{1-(\rho\lambda)^{\bar{t}}}{1-\rho\lambda}\lambda\left[(1+\rho)\theta_{l,h} - \rho(1+\lambda)\frac{\theta_{h,h}}{2}\right] \\ \frac{1-(\rho\lambda)^{\bar{t}}}{1-\rho\lambda}\lambda T\left[\rho\left(\theta_{h,h} - \theta_{l,h}\right) - \frac{T}{2}\right] \end{array} \right\} (1+\lambda)^{-1} \tag{7}$$

Note that $\underline{\Pi}\left(T, \lambda, \bar{t}\right) = \overline{\Pi}\left(T, \lambda, \bar{t}\right)$ if and only if $.T = \frac{1+\lambda}{2}$ Thus the net production value is continuous at $T = \frac{1+\lambda}{2}$. Differentiating each branch with

[8] The derivation of equations (6) and (7) is provided in the Appendix.

respect to T, we obtain

$$\begin{aligned}\frac{d\underline{\Pi}\left(T,\lambda,\bar{t}\right)}{dT} &= \frac{\left(1-(\rho\lambda)^{\bar{t}}\right)\lambda}{(1+\lambda)(1-\rho\lambda)}\left[\rho\left(\theta_{l,h}-\theta_{l,l}\right)-T\right]=0 \\ &\Leftrightarrow\ T=T^{S}=\rho\left(\theta_{l,h}-\theta_{l,l}\right)=:\underline{T}^{S}\end{aligned}$$

and

$$\begin{aligned}\frac{d\overline{\Pi}\left(T,\lambda,\bar{t}\right)}{dT} &= \frac{\left(1-(\rho\lambda)^{\bar{t}}\right)\lambda}{(1+\lambda)(1-\rho\lambda)}\left[\rho\left(\theta_{h,h}-\theta_{l,h}\right)-T\right]=0 \\ &\Leftrightarrow\ T=T^{S}=\rho\left(\theta_{h,h}-\theta_{l,h}\right)=:\overline{T}^{S}.\end{aligned}$$

It is readily verified that $\widehat{\theta}<0$ implies $\overline{T}^{S}<\underline{T}^{S}$. Furthermore, the optimum values must satisfy $\overline{T}^{S}\geq\frac{1+\lambda}{2}$ and $\underline{T}^{S}\leq\frac{1+\lambda}{2}$, respectively. Assuming $\rho\leq\left(\theta_{h,h}-\theta_{l,h}\right)^{-1}$ so as to guarantee $\overline{T}^{S}\leq 1$ as an interior equilibrium, it is then easy to verify the following equilibrium structure

$$T^{S}=\begin{cases}\underline{T}^{S}\Leftrightarrow\rho\leq\frac{1+\lambda}{2\left(\theta_{l,h}-\theta_{l,l}\right)} \\ \frac{1+\lambda}{2}\Leftrightarrow\rho\in\left[\frac{1+\lambda}{2\left(\theta_{l,h}-\theta_{l,l}\right)},\frac{1+\lambda}{2\left(\theta_{h,h}-\theta_{l,h}\right)}\right] \\ \overline{T}^{S}\Leftrightarrow\rho\geq\frac{1+\lambda}{2\left(\theta_{h,h}-\theta_{l,h}\right)}\end{cases}.$$

By an argument similar to the one applied for the case with $\widehat{\theta}>0$ it can be demonstrated that $T(t)=T^{S}$ is a steady state for all $t\in\left[0,\bar{t}-1\right]$. The structure we find for the optimal rate of training is intuitive. If and only if the discount factor is sufficiently low (implying a high interest rate r), then it is socially optimal to train only a small share of young agents $\underline{T}^{S}\leq\frac{1+\lambda}{2}$ and rather save the high upfront training costs. This is the case even if it leads to an excess supply of untrained agents. For high values of the discount factor, it is optimal to train at the rate $\overline{T}^{S}\geq\frac{1+\lambda}{2}$ in order to realise future productivity gains amounting to $\rho\left(\theta_{h,h}-\theta_{l,h}\right)$, even if this implies an excess supply of skilled agents. Finally, for intermediate values of the discount factor, it is optimal to train at the rate $\frac{1+\lambda}{2}$ so that in any period the number of skilled agents just balances the number of unskilled agents. In this last case only does the rate of training depend on the age structure. As expected, the rate of training increases with the rate of reproduction, λ. This is because a greater number of future young/unskilled agents would have to be matched with skilled agents. In this intermediate case, the rate of training is thus independent of the discount factor. In contrast, the rate of training increases with the discount factor both at the upper and lower end.

4 The decentralised economy

We now turn to the allocation of training when the agents organise themselves into partnerships and decide decentrally on whether or not to engage in training activities. In order to focus on the incidence of training, we assume that the matching process is frictionless. Once a team has formed and production takes place the output is shared between the partners. In homogeneous teams (composed of either skilled or unskilled agents only) output is split equally. For mixed teams composed of a skilled and an unskilled worker, a sharing rule determines a net transfer to the skilled agents as an incentive to take on board unskilled partners and, possibly, to provide training. We continue to assume a weakly shrinking population, i.e. $\lambda \leq 1$. Note that this implies that a skilled old agent cannot expect with certainty to be matched with a young agent. Similar to the case of the social optimum we determine the steady state rate of training, as the share of young workers receiving training within each period.

4.1 Decentral training with assortative matching ($\widehat{\theta} > 0$)

Again, we commence with the case of assortative matching. Recall that for $\widehat{\theta} = \theta_{h,h} + \theta_{l,l}$-$2\theta_{l,h} > 0$ segregated teams dominate (in terms of total output) mixed teams. It is then readily shown in analogy to Becker (1973) that mixed teams will never be formed unless for the purpose of training. Assume as a benchmark that the partners split the net output of the team and then define a as the net transfer from the unskilled to the skilled partner. Furthermore, define

$$\underline{a} := \frac{\theta_{h,h} - \theta_{l,h}}{2}, \qquad \overline{a} := \frac{\theta_{l,h} - \theta_{l,l}}{2},$$

with $\underline{a} < \overline{a}$ if and only if $\widehat{\theta} < 0$.

Lemma 1 *(i) Mixed teams form if and only if* $a \in [\underline{a}, \overline{a}]$. *(ii) Mixed teams do not form if* $\widehat{\theta} > 0$.

Proof (i) Skilled old agents form mixed teams (without training) only if $\theta_{l,h}/2 + a \geq \theta_{h,h}/2$ or $a \geq \frac{\theta_{h,h}-\theta_{l,h}}{2} =: \underline{a}$. In any one period unskilled agents are willing to pay for entering a mixed team (without training) only if $\theta_{l,h}/2 - a \geq \theta_{l,l}/2$ or $a \leq \frac{\theta_{l,h}-\theta_{l,l}}{2} =: \overline{a}$. Hence, mixed teams form if and only if $a \in [\underline{a}, \overline{a}]$. (ii) Existence of the interval $[\underline{a}, \overline{a}]$ contradicts $\widehat{\theta} > 0$.■

Similar to Becker (1973) efficient matching is attained within a decentralised economy. For the case $\widehat{\theta} > 0$ this implies that mixed teams are not formed unless this is for the purpose of training.

We can now turn to the provision of training. Denote with b the net transfer from young to old workers in teams where training takes place.[9] Skilled agents prefer training a younger colleague at cost δ over forming a segregated (o^s, o^s)-team if and only if $(\theta_{l,h} - \delta)/2 + b \geq \theta_{h,h}/2$ or

$$\pi_T^o(\delta, b) := \frac{\theta_{l,h} - \theta_{h,h}}{2} - \frac{\delta}{2} + b \geq 0. \tag{8}$$

The discounted expected net surplus (over two periods) of a young worker who engages in training can be written as

$$\begin{aligned} \pi_T^y\left(\delta, b, \widehat{\delta}, \widehat{b}, p\right) &= \left\{\frac{\theta_{l,h} - \delta}{2} - b + \rho\left[p\left(\frac{\theta_{l,h} - \widehat{\delta}}{2} + \widehat{b}\right) + (1-p)\frac{\theta_{h,h}}{2}\right]\right\} \\ &\quad - (1+\rho)\frac{\theta_{l,l}}{2} \\ &= \frac{\theta_{l,h} - \theta_{l,l}}{2} - \frac{\delta}{2} - b + \rho\left[\frac{\theta_{h,h} - \theta_{l,l}}{2} + p\pi_T^o\left(\widehat{\delta}, \widehat{b}\right)\right]. \end{aligned} \tag{9}$$

Here, $p \in [0,1]$ is the conditional probability of a trained worker to enter a mixed team and train in the second period, $\pi_T^o\left(\widehat{\delta}, \widehat{b}\right)$ is the expected profit from training a future worker, where $\widehat{\delta}$ is the expected training cost and $\widehat{b} = b\left(\widehat{\delta}\right)$ the corresponding transfer. The term in bracelets in the first

[9] Along the lines of Morrison and Wilhelm (2004) such a transfer may be implicit in the value of the shares sold on by a senior partner to an (junior) associate.

line is thus the expected flow of surplus (over two periods) when the young worker takes out training. The surplus $\frac{\theta_{l,h}-\delta}{2} - b$ during the training period is amended by the present value of the surplus expected as an old worker. $\rho\left[\frac{\theta_{h,h}}{2} + p\pi_T^o\left(\widehat{\delta}, b\left(\widehat{\delta}\right)\right)\right]$. As the supply condition (8) has to hold, it follows that the expected future surplus increases (weakly) in the probability of being matched with a future young worker, p. The expected flow of surplus from training is balanced against the flow of surplus (over two periods) when remaining untrained. In the following we will make the following assumption.

Assumption A *Consider any two values $\delta, \delta' \in [0,1]$, where $\delta < \delta'$. Then any restrictions, such as uniformity constraints or limited liability constraints that govern the transfer $b(\delta)$ will also govern the transfer $b(\delta')$.*

The assumption implies that for a given rate of training, T, all workers with $\delta \in [0, T]$, and only those, will train. It is easy to see why there are no gaps in training. Using (8) and (9) we can write the total return to training as $\pi_T^y(\cdot) + \pi_T^o(\cdot) = T^S - \delta + \rho p \pi_T^o\left(\widehat{\delta}, b\left(\widehat{\delta}\right)\right)$. This expression is strictly decreasing with δ. Thus, if it is profitable to the team to train a young agent at cost δ' then the return to the team is even larger for an agent with $\delta < \delta'$. Whether or not training is implemented then depends only on the restrictions on the transfers $b(\delta)$ and $b(\delta')$. Our assumption implies that if these restrictions rule out the training of δ they rule out a fortiori the training of δ'. Gaps in training cannot therefore arise.[10]

We can thus determine T and $\widehat{T}$, respectively, as the current and expected rate of training. Assuming without much loss of generality that $\widehat{T} \leq T$ we can write $p = \frac{\widehat{T}\sigma_y}{T\sigma_o} = \frac{\widehat{T}\lambda}{T} \leq 1$, as determined by the current and expected rate of training as well as by the age-structure, λ. We can now proceed to determine the steady state allocation of training.

[10] To see why assumption A is not trivial, consider a set-up where a young agent's outside utility decreases in δ (perhaps because agents with a high training cost are less attractive prospects for outside employers). In this case, the upper bound $\overline{b}(\delta)$ on the transfer that is implied by the outside utility increases with δ. As $\overline{b}(\delta') \geq \overline{b}(\delta)$ this implies a more severe restriction on the transfer for the low cost type. Assumption A rules out such cases. Note that the set-up of our model does not include elements in contradiction with Assumption A.

4.1.1 Steady-state training intensity and incidence of training

Substituting from (8) into (9) we can write

$$\begin{aligned}\pi_T^y(\cdot) &= \frac{\theta_{l,h}-\theta_{l,l}}{2}-\delta+\frac{\theta_{l,h}-\theta_{h,h}}{2}-\pi_T^o(\delta,b)+\rho\left[\frac{\theta_{h,h}-\theta_{l,l}}{2}+p\pi_T^o\left(\widehat{\delta},b\left(\widehat{\delta}\right)\right)\right]\\ &= T^S-\delta-\pi_T^o(\delta,b(\delta))+\rho p\pi_T^o\left(\widehat{\delta},b\left(\widehat{\delta}\right)\right),\end{aligned}$$

where the second line follows when observing $T^S=\frac{\theta_{l,h}-\theta_{l,l}}{2}+\frac{\theta_{l,h}-\theta_{h,h}}{2}+\rho\frac{\theta_{h,h}-\theta_{l,l}}{2}$ as derived in (4). Training is demanded by young agents as long as $\pi_T^y(\cdot)\geq 0$ or, equivalently, as long as

$$\delta\leq T^S-\pi_T^o(\delta,b(\delta))+\rho p\pi_T^o\left(\widehat{\delta},b\left(\widehat{\delta}\right)\right)=:T^D.$$

Here, T^D denotes the marginal worker as well as the equilibrium rate of training in a steady-state. The expression allows us to interpet the incentives that govern the demand for training. The social value of training, T^S, forms one part of the private incentive. However, the full private return to training contains additional rent-shifting elements which do not contribute to the social value: On the one hand, the private return (to a young agent) is reduced to the extent that an amount $\pi_T^o(\delta,b)\geq 0$ of the return to training falls to the old agent (as trainer). On the other hand, the private return is increased to the extent that training can be sold at a profit to the next generation of young workers. Note that the discounted value of the expected return $\rho p\pi_T^o\left(\widehat{\delta},b\left(\widehat{\delta}\right)\right)$ increases in the prbability p that a trainee can be found. Which, if any, of these rent-shifting elements dominates the total incentive depends on the pricing process and the expectations about the future as determined, inter alia, by the age-structure of the workforce.

Consider now a steady state, where

$$\widehat{T}=T=T^D=T^S-\pi_T^o\left(T^D,b\left(T^D\right)\right)+\rho\lambda\pi_T^o\left(\frac{T^D}{2},b\left(\frac{T^D}{2}\right)\right)\qquad(10)$$

and where $\widehat{\delta}=\frac{\widehat{T}}{2}=\frac{T^D}{2}$ for a uniform distribution.[11] The socially optimal rate of training is then realised only if $\pi_T^o\left(T^D,b\right)=\rho\lambda\pi_T^o\left(\frac{T^D}{2},b\left(\frac{T^D}{2}\right)\right)$. This is the case either if the current profit from training the marginal worker

[11] Obviously, equation (10) defines T^D only implicitly.

cancels out the expected profit from training - an unlikely event -, or if both the current profit from training the marginal worker and the expected profit from training are zero.

Define $\Theta := \frac{T^D \sigma_y}{\sigma_y + \sigma_o} = T^D \frac{\lambda}{1+\lambda}$ as the per-capita incidence of training in a steady state, i.e. as the share of the total population that receives training. The effect of age structure on the per-capita incidence of training is then given by[12]

$$\frac{d\Theta}{d\lambda} = \frac{1}{(1+\lambda)^2} T^D + \frac{\lambda}{1+\lambda} \frac{dT^D}{d\lambda}$$

$$\text{with } \frac{dT^D}{d\lambda} = \frac{\rho \pi_T^o \left(\frac{T^D}{2}, b\left(\frac{T^D}{2} \right) \right)}{1 + \left(\frac{d\pi_T^o (T^D, b(T^D))}{dT^D} - \frac{\rho\lambda}{2} \frac{d\pi_T^o \left(\frac{T^D}{2}, b\left(\frac{T^D}{2} \right) \right)}{dT^D} \right)} \geq 0.$$

The impact of population ageing (as measured by a decrease in λ below 1) on the incidence of training can thus be decomposed into two distinct effects, one demographic, one economic. Both effects are negative. The incidence of training is reduced due to a lower share of (trainable) young workers in the population (the demographic effect) and by a reduction in the steady-state rate of training (the economic effect, reflecting the reduced return to training). We can thus make a first summary.

Proposition 1 *(i) The steady state rate of training implemented in a decentral economy is below (above) the social optimum if and only if the old agents attain a profit on the marginal trainee that exceeds (falls short of) the expected profit from selling training to future generations. (ii) The negative demographic impact of population ageing (i.e. a lower λ) on the per-capita incidence of training in a steady state is compounded by an economic effect driven by the negative impact of ageing on training incentives.*

4.1.2 Training in the presence of price competition

So far we have not made any assumptions about the mechanism by which the training fee b is determined. Naturally, one would expect the details of

[12] $\frac{dT^D}{d\lambda} > 0$ follows as the denominator of the RHS has to be positive for a stable training rate.

this mechanism to have a bearing on the rate of training. This is because it determines the current and future distribution of rents and thus the incentives to train. In the following, we examine in greater detail the role of the price mechanism, distinguishing two cases: (i) prices are determined together with the matches in a process of price competition and (ii) prices are set before matching takes place as the outcome of a bargaining process between representatives of the young and the old.

We consider (i) first, assuming that the training cost, δ, is perfectly observable and contractible and that young and old partners compete a la Bertrand seeking to attract favourable partners. For a shrinking population there is always an excess supply of skilled agents. But then, Bertrand competition for trainees implies that $\pi_T^o(\delta, b) = 0$. The skilled agent's surplus from training is fully exctracted. Naturally, this holds for the future as well so that $\pi_T^o\left(\widehat{\delta}, \widehat{b}\right) = 0$. From (10) it then follows immediately that $T^D = T^S$.

Proposition 2 (i) *In the case of price competition, the optimal rate of training is attained in a decentral setting. (ii) Age-structure has no effect on the rate of training T^D.*

Note that ex-post price competition for trainees implies partnership specific fees. These are bargained down to the minimum at which a skilled worker is willing to supply training. The permanent excess supply of training in a steady state with a shrinking population gives rise to the rational expectation that the surplus of skilled workers will be fully extracted in future periods, too. But then, the private return to becoming trained is independent of whether or not one is able to train in the future and is thus independent of the age structure. Indeed, the private return to training equals the (shared) excess surplus generated by a highly skilled team as compared to an unskilled team, which corresponds to the social return.

4.1.3 Training when the fee is determined by bargaining

In this section we consider a setting, where a uniform fee b is determined by way of negotiations between old workers and young workers before matching takes place. A uniform fee of training obviously conforms with Assumption A and implies that $\pi_T^o(\delta, b)$ strictly falls in δ. The supply of training is then governed by the condition

$$\pi_T^o(T, b) = \frac{\theta_{l,h} - \theta_{h,h}}{2} - \frac{T}{2} + b \geq 0. \tag{11}$$

Furthermore, observing $\pi_T^o\left(\frac{T}{2},b\right)=\pi_T^o\left(T,b\right)+\frac{T}{4}$, we can express the expected surplus for a skilled agent as $\frac{\theta_{h,h}}{2}+\frac{\lambda T}{T}\left[\pi_T^o\left(T,b\right)+\frac{T}{4}\right]$.

The net surplus of a young agent of type δ who engages in training at fee b is given by

$$\begin{aligned}\pi_T^y\left(\delta,b,T,\widehat{b},\widehat{T},\right) \quad :&= \frac{\theta_{l,h}-\theta_{l,l}}{2}-\frac{\delta}{2}-b+\rho\left[\frac{\theta_{h,h}-\theta_{l,l}}{2}+\frac{\lambda\widehat{T}}{T}\pi_T^o\left(\frac{\widehat{T}}{2},\widehat{b}\right)\right]\\ &= \frac{\theta_{l,h}-\theta_{l,l}}{2}-\frac{\delta}{2}-b+\rho\left\{\frac{\theta_{h,h}-\theta_{l,l}}{2}+\frac{\lambda\widehat{T}}{T}\left[\pi_T^o\left(\widehat{T},\widehat{b}\right)+\frac{\widehat{T}}{4}\right]\right\}\end{aligned}$$

where $\frac{\lambda\widehat{T}}{T}\leq 1$ is the conditional 2nd period probability for a trainee to be matched with a (future) young worker, and where $\frac{\widehat{T}}{2}$ is the expected training cost in the 2nd period.[13]. The uniform pricing rule implies that young agents will demand training for all $\delta\leq T$ so that the demand for training is governed by the condition

$$\pi_T^y\left(T,b,T,\widehat{b},\widehat{T}\right)\geq 0. \tag{12}$$

Using the conditions in (12) and (11) one can show that there exists a value $\widetilde{b}\geq\frac{\theta_{h,h}-\theta_{l,h}}{2}$ such that

$$\pi_T^o\left(T,b\right) \;=\; 0\leq\pi_T^y\left(T,b,T,\widehat{b},\widehat{T}\right)\Leftrightarrow b\leq\widetilde{b} \tag{13}$$

$$\pi_T^o\left(T,b\right) \;>\; 0=\pi_T^y\left(T,b,T,\widehat{b},\widehat{T}\right)\Leftrightarrow b>\widetilde{b}. \tag{14}$$

Solving the respective equalities with respect to T we find the equilibirum rate of training that is realised for a fee set at a value b

$$T\left(b,\lambda\right)=\begin{cases}2b+\left(\theta_{l,h}-\theta_{h,h}\right) & \text{if } \; b\leq\widetilde{b}\\ A-b+\sqrt{\left(A-b\right)^2+2B} & \text{if } \; b>\widetilde{b}\end{cases} \tag{15}$$

$$A \;:= \frac{\theta_{l,h}-\theta_{l,l}}{2}+\rho\frac{\theta_{h,h}-\theta_{l,l}}{2}, \tag{16}$$

$$B \;:= \rho\lambda\widehat{T}\left[\pi_T^o\left(\widehat{T},\widehat{b}\right)+\frac{\widehat{T}}{4}\right], \tag{17}$$

[13] Again we assume without much loss of generality that $\widehat{T}\leq T$.

where

$$T_b(b,\lambda) = \begin{cases} 2 & \text{if } \; b \le \widetilde{b} \\ -\frac{2T(b,\lambda)^2}{T(b,\lambda)^2+2B} & \text{if } \; b > \widetilde{b} \end{cases}. \tag{18}$$

Hence, the rate of training is first increasing and then decreasing in the fee. This reflects the fact that the rate of training is driven by the short-side of the market. For low levels of b there is an excess demand for training, which is rationed by the level of supply $T(b) := T\,|\pi_T^o(T,b) = 0$. Naturally, as more training is supplied for higher b supply and thus the rate of training increase. However, if the fee exceeds the boundary $\widetilde{b}$, there is an excess supply of training, which is now rationed by the level of demand $T(b,\lambda) := T\left|\pi_T^y\left(T,b,T,\widehat{b},\widehat{T}\right) = 0\right.$. Naturally, on this branch the rate of training falls in b.

We posit that the bargaining parties maximise a generalised Nash product. The outside utilities are determined by the outcome in the absence of any training. In this case, the agents produce in fully segregated teams, implying a total per capita surplus $\frac{\theta_{l,l}}{2} + \overline{T}\frac{\theta_{h,h}-\theta_{l,l}}{2}$ for the old and $(1+\rho)\frac{\theta_{l,l}}{2}$ for the young, where $\overline{T}$ denotes the current share of skilled workers. It can be shown that the Nash product is then given by[14]

$$\begin{aligned} &\max_b \lambda\sigma_o T\left[\pi_T^o(T,b) + \frac{T}{4}\right]^\beta \left[\pi_T^y\left(T,b,T,\widehat{b},\widehat{T},\right) + \frac{T}{4}\right]^{1-\beta} \quad (19) \\ \text{subject to } T &= T(b,\lambda) \end{aligned}$$

with $\beta \in [0,1]$ being a measure of the bargaining power of the old. Note that $\pi_T^o(T,b) + \frac{T}{4}$ and $\pi_T^y\left(T,b,T,\widehat{b},\widehat{T},\right) + \frac{T}{4}$ correspond to the average (net) return to training for an old and young worker, respectively, conditional on training taking place. Thus, the bargaining partners maximise the weighted product of the average returns to training multiplied, in turn, by the incidence of training $\lambda\sigma_o T$. In the following, we consider a steady state, where $\widehat{T} = T$ and $\widehat{b} = b$. Let $b^B(\beta,\lambda)$ denote the steady-state outcome of the bargaining, so that $T^B(\beta,\lambda) = T\left(b^B(\beta,\lambda),\lambda\right)$. Furthermore, define

$$\underline{\beta} := \frac{2}{4+\rho\lambda}, \qquad \overline{\beta} := \frac{2+\rho\lambda}{4+\rho\lambda}.$$

[14]The derivation of equation (19) is provided in the Appendix.

The following can then be shown (contained in the proof of Lemma 2)

$$\begin{aligned}\pi_T^y\left(T^B(\beta,\lambda),b^B(\beta,\lambda)\right) &\geq \pi_T^o\left(T^B(\beta,\lambda),b^B(\beta,\lambda)\right)=0\Rightarrow\beta<\overline{\beta}\\ \pi_T^o\left(T^B(\beta,\lambda),b^B(\beta,\lambda)\right) &\geq \pi_T^y\left(T^B(\beta,\lambda),b^B(\beta,\lambda)\right)=0\Rightarrow\beta\geq\underline{\beta}.\end{aligned}$$

Hence, a bargaining outcome at which the surplus of the marginal old (young) agent is fully extracted can arise only if the bargaining power of the old, β, is sufficiently small (large). Since, $\underline{\beta}<\overline{\beta}$, however, mutliple outcomes are feasible for intermediate levels $\beta\in\left[\underline{\beta},\overline{\beta}\right]$. We can then establish the following result.

Lemma 2 *Consider a steady state with $\widehat{T}=T=T^B(\beta,\lambda)$ and $\widehat{b}=b=b^B(\beta,\lambda)$. (i) If $\beta\leq\underline{\beta}$, the rate of training is given by*

$$T^B=T_o^B(\beta,\lambda):=\frac{2(1+\beta)}{3-\beta\lambda\rho}\left(\rho\frac{\theta_{h,h}-\theta_{l,l}}{2}-\frac{\widehat{\theta}}{2}\right)=\frac{2(1+\beta)}{3-\beta\lambda\rho}T^S$$

(ii) If $\beta\geq\overline{\beta}$, the rate of training is given by

$$T^B=T_y^B(\beta,\lambda):=\frac{2[2-\beta(1+\lambda\rho)]}{3(1-\beta\lambda\rho)}\left(\rho\frac{\theta_{h,h}-\theta_{l,l}}{2}-\frac{\widehat{\theta}}{2}\right)=\frac{2[2-\beta(1+\lambda\rho)]}{3(1-\beta\lambda\rho)}T^S.$$

(iii) If $\beta\in\left[\underline{\beta},\overline{\beta}\right]$, the rate of training is given by $T^B\in\left\{T_y^B(\beta,\lambda),T_o^B(\beta,\lambda)\right\}$.

Proof See Appendix.

The following properties are readily verified. They can be used to represent the bargaining outcome as depicted in figure 1.

Corollary 1 *(i) $\frac{\partial T_o^B(\beta,\lambda)}{\partial\beta}>0;\frac{\partial T_o^B(\beta,\lambda)}{\partial\lambda}>0$ and $\frac{\partial T_y^B(\beta,\lambda)}{\partial\beta}<0;\frac{\partial T_y^B(\beta,\lambda)}{\partial\lambda}>0$;*
(ii) $T_o^B(0,\lambda)=T_y^B(1,\lambda)=\frac{2}{3}T^S$;
(iii) $T_y^B\left(\frac{1}{2},\lambda\right)=\frac{2(3-\lambda\rho)}{3(2-\lambda\rho)}T^S>\frac{6}{6-\lambda\rho}T^S=T_o^B\left(\frac{1}{2},\lambda\right)$ for all $\lambda\rho>0$.
(iv) $T_y^B\left(\underline{\beta},\lambda\right)=\frac{4}{4-\lambda\rho}T^S>\frac{2(6+\lambda\rho)}{12+\lambda\rho}T^S=T_o^B\left(\underline{\beta},\lambda\right)>T^S$ for all $\lambda\rho>0$.
(v) $T_o^B\left(\overline{\beta},\lambda\right)=\frac{4}{4-\lambda\rho}T^S>\frac{2[6-\lambda\rho(1+\lambda\rho)]}{3[4-\lambda\rho(1+\lambda\rho)]}T^S=T_y^B\left(\overline{\beta},\lambda\right)>T^S$ for all $\lambda\rho>0$.

[Insert figure 1]

The equilibrium rate of training depends on the distribution of bargaining power and on the age-structure. Consider a situation in which the bargaining power of the old is low, i.e. $\beta \in \left[0, \underline{\beta}\right]$. Here, the rate of training increases in the bargaining power of the old, reflecting their greater willingness to supply training as the fee increases. Similarly, for $\beta \in \left[\overline{\beta}, 0\right]$ the rate of training decreases when bargaining power is shifted towards the old. The corresponding increase in the fee drives some of the potential trainees out of the market. The result is ambiguous for intermediate distributions of bargaining power, i.e. for $\beta \in \left[\underline{\beta}, \overline{\beta}\right]$. In this case, two outcomes can be supported: An outcome with a high fee, $b_y(\beta, \lambda)$, leading to $T^B = T_y^B(\beta, \lambda)$ and an outcome with a low fee, $b_o(\beta, \lambda)$, leading to $T^B = T_o^B(\beta, \lambda)$.[15] These outcomes cannot be Pareto-ranked.[16] As one would expect the old prefer the high fee outcome and the young prefer the low fee outcome for any level of $\beta \in \left[\underline{\beta}, \overline{\beta}\right]$. Intuitively, one would expect the choice of outcome to depend on the expectations about the future bargaining outcome. If the current bargaining partners expect future bargaining to lead to a low fee , then this restricts the expected return to training and thus the current willingness to engage in training. In this case, only a low fee will be implemented. The converse applies to the expectation of high fees in the future. Note that the ranking of the training intensities $T_y^B(\beta, \lambda)$ and $T_o^B(\beta, \lambda)$ for high and low fees, respectively, depends on β. More specifically, $T_y^B(\beta, \lambda) < T_o^B(\beta, \lambda)$ applies if and only if β is sufficiently high. Maximal rates of training are reached at $\underline{\beta}$ and $\overline{\beta}$ for the low and high fee outcomes $T_o^B\left(\underline{\beta}, \lambda\right) = T_y^B\left(\overline{\beta}, \lambda\right)$, respectively. Note that in these two cases both the marginal young and old agent make a zero profit. This also illustrates why for increases in β above $\overline{\beta}$, the low fee outcome is no longer supported. This is because the marginal young agent would now make a negative profit. Similarly, a low fee outcome is not supported for $\beta < \underline{\beta}$ as it would imply a negative profit for the marginal old agent. Finally, note the asymmetry where for an equal distribution of bargaining power, $\beta = \frac{1}{2}$, a higher rate of training is implemented in the

[15] Define $b_y(\beta, \lambda) := b\left|\pi_T^y\left(T_y^B(\beta, \lambda), b\right) = 0\right.$ and $b_o(\beta, \lambda) := b\left|\pi_T^o\left(T_o^B(\beta, \lambda), b\right) = 0\right.$. It can then be shown that $b_o(\beta, \lambda) \leq b_o\left(\overline{\beta}, \lambda\right) = b_y\left(\underline{\beta}, \lambda\right) \leq b_y(\beta, \lambda)$ is true for all $\beta \in \left[\underline{\beta}, \overline{\beta}\right]$.

[16] We provide a sketch of the proof in the Appendix.

high fee outcome. This is because the joint return from selling training to future agents $\frac{\lambda\widehat{T}}{T}\left[\pi_T^o\left(\widehat{T},\widehat{b}\right)+\frac{\widehat{T}}{4}\right]$ is greater when a high fee is supported in the steady-state. But this raises the willingness to train on the part of the current young even when having to pay a higher fee now.

In contrast to the case with price competition, we now find that the rate of training depends on the age structure. This is because for a uniform fee rents are always accruing to old agents who are supplying training within the margin. Recall that the expected return to providing training (conditional on matching with a young worker) is given by $\pi_T^o(T,b)+\frac{T}{4}\geq\frac{T}{4}>0$. In a steady-state this return is expected for the future, which increases the willingness to undertake training on the part of the current young. However, attaining this return depends on whether or not a trainee can be found in the future. The (conditional) probability of finding a trainee is given by λ, implying that the expected return $\lambda\left[\pi_T^o\left(\widehat{T},\widehat{b}\right)+\frac{\widehat{T}}{4}\right]$ and thus the rate of training increases with the reproduction rate. Turning this argument around, ageing of the workforce will lead to a reduction in training rates because it leads to an erosion of the expected returns to training.

Under bargaining, the training rate generally deviates from the rate that would be socially optimal. The situation is depicted in figure 2 and can be summarised as follows.

[Insert figure 2]

Proposition 3 *There exists a pair of boundary values β_l and β_h with $0<\beta_l<\underline{\beta}<\overline{\beta}<\beta_h<1$ such that (i) training is socially optimal if and only if $\beta\in\{\beta_l,\beta_h\}$; (ii) training is over-provided if and only if $\beta\in(\beta_l,\beta_h)$; (iii) training is under-provided if $\beta<\beta_l$ or $\beta>\beta_h$. (iv) For any given distribution of bargaining power β the tendency towards under (over-)provision falls (increases) in the reproduction rate λ.*

Proof Parts (i) to (iii) follow immediately from (iv) and (v) of Corollary 1 together with the observation that $\frac{\partial T_o^B(\beta,\lambda)}{\partial\beta}>0$ *and* $\frac{\partial T_y^B(\beta,\lambda)}{\partial\beta}<0$ as by part (i) of Corollary 1. Part (iv) follows immediately from the observation that $\frac{\partial T_o^B(\beta,\lambda)}{\partial\lambda}>0$ *and* $\frac{\partial T_y^B(\beta,\lambda)}{\partial\lambda}>0$ as by part (i) of Corollary 1.■

The findings are intuitive. A concentration of bargaining power on either side of the market will lead to rationing of training activities to levels that fall short from the social optimum. This is because the dominant party extracts surplus from the bargaining partner by way of charging excessive fees (if dominance lies with the old) or depressing fees (if dominance lies with the young). The degree to which training is under-provided is limited, however, as even the dominant bargaining party has an interest in not stifling the 'market' too much. For an equal distribution of bargaining power, not too much surplus can be extracted from the other bargaining party. In this case bargaining parties seek to jointly extract surplus $\lambda\left[\pi_T^o\left(\widehat{T},\widehat{b}\right)+\frac{\widehat{T}}{4}\right]$ from the future generation by way of 'generating' future trainers. Socially excessive training results. Note that this applies irrespective of whether a high or low fee is realised in the case of multiple outcomes. It is thus immaterial, in a sense, whether it is the young or the old who benefit more. The important mechanism is that the generations currently active in the profession seek to extract rent from the next generation. The degree of training increases in the reproduction rate: A larger proportion of young agents drives up the expected rents to be captured. In figure 2 we have plotted the training rates for three possible levels of λ. Evidently, the excess supply of training is most pronounced (and most likely) in a stationary population, where $\lambda = 1$. Every young agent receiving training today will then be able to sell it on in the future. In the opposite extreme of a workforce not reproducing at all, i.e. for $\lambda = 0$, training will always be under-supplied. Naturally, the bargaining generations can at best receive the social return from training but there are no rents to be gained. In this case, the only way for bargaining parties to acquire surplus is to extract it from their present-day oponnents This leads to rationing and an under-supply of training.

Finally, we note that the generation of rents which leads to possible over-production is not so much due to the presence of bargaining parties but rather due to the fact that only a uniform fee is charged. It is the uniform fee that rules out the downward competition of the fee by skilled agent to the level at which no surplus accrues to the skilled partner in any of the teams. To see this, suppose there is a market for training operating under the requirement that a single price is implemented. In this case, market clearing implies a level of the fee $b^* = b\left|T^B(b,\lambda) = T_o^B(b,\lambda) = T_y^B(b,\lambda)\right.$. It is readily verified that this implies the maximimal supportable training level $T^B(b^*,\lambda) = \frac{4}{4-\lambda\rho}T^S$. Hence, we see that a market with a uniform price (rather than a bargaining procedure) would always lead to over-supply, at indeed, the most excessive level. To this end, bargaining arrangements do, in fact, mitigate the scope for over-provision.

4.2 Decentral training with mixed teams ($\widehat{\theta} < 0$)

Recall from Lemma 1, part (i) that for $\widehat{\theta} < 0$ there is scope for mixed teams (without training) if (and only if) the transfer a is chosen within the interval $[\underline{a}, \overline{a}]$. In this case, both the skilled and unskilled partners to a mixed team benefit from the arrangement and, thus, we may assume that mixed teams will always be formed for $\widehat{\theta} < 0$ whenever this is possible. Recalling (5), it must then be true then that a decentral economy features (a) no segregated teams if $T = \frac{1+\lambda}{2}$, (b) segregated teams with unskilled workers exclusively if $T < \frac{1+\lambda}{2}$, and (c) segregated teams with skilled workers exclusively if $T > \frac{1+\lambda}{2}$.

We can now turn to the provision of training. Consider a mixed team consisting of a skilled agent and a young agent. Skilled agents prefer to train their younger colleague at cost δ if and only if $(\theta_{l,h} - \delta)/2 + b \geq \theta_{l,h}/2 + a$ or

$$b \geq \frac{\delta}{2} + a = \underline{b}(\delta, a) \tag{20}$$

The distinguishing feature of the setting in which mixed teams form is that there are now two 'markets': one for the formation of mixed partnerships (without training) and one for training. The relationship in (20) establishes an arbitrage condition between the two markets from the perspective of the old agent.

Consider now the young partner who is matched with a skilled agent. When engaging in training (rather than just working as an untrained junior partner) the young worker expects a discounted net surplus (over two periods) amounting to

$$\begin{aligned}
\pi_T^y\left(\delta, b, a, \widehat{\delta}, \widehat{b}, \widehat{a}\right) &= \left\{\frac{\theta_{l,h}-\delta}{2} - b + \rho\left[\begin{array}{c} p_T\left(\frac{\theta_{l,h}-\widehat{\delta}}{2}+\widehat{b}\right) \\ +p_M\left(\frac{\theta_{l,h}}{2}+\widehat{a}\right)+p_S\frac{\theta_{h,h}}{2}\end{array}\right]\right\} \\
&\quad -\left\{\frac{\theta_{l,h}}{2} - a + \rho\left[p_{NT}\left(\frac{\theta_{l,h}}{2}-\widehat{a}\right)+\left(1-p_{NT}\right)\frac{\theta_{l,l}}{2}\right]\right\} \\
&= \left\{-\frac{\delta}{2} - b + a + \rho\left[\begin{array}{c} \frac{\theta_{h,h}-\theta_{l,l}}{2} + \left(p_T+p_M\right)\left(\widehat{a}-\underline{a}\right) \\ +p_T\left(\widehat{b}-\underline{b}\left(\widehat{\delta},\widehat{a}\right)\right) \\ -p_{NT}\left(\overline{a}-\widehat{a}\right)\end{array}\right]\right\},
\end{aligned}$$

where $\left(\widehat{\delta}, \widehat{b}, \widehat{a}\right)$ refer to the expected values in the future period. Moreover, p_T, p_M and p_S denote the conditional probabilities that a skilled worker enters a mixed team with training, a mixed team without training and a segregated team, respectively. Similarly, p_{NT} is the conditional probability that an unskilled worker is offered a place in a mixed team (without training). Obviously, $p_T + p_M + p_S = 1$, which has been used together with the definitions $\underline{a} = \frac{\theta_{h,h}-\theta_{l,h}}{2}$ and $\overline{a} = \frac{\theta_{l,h}-\theta_{l,l}}{2}$ in transforming the expression. The second expression is easily interpreted. In the first period, the young agent receives a net surplus from training (as opposed to working in a mixed team without training) equal to $-\frac{\delta}{2} - b + a$, where $b - a$ is the 'surcharge' for training. The net return to training is composed of four parts: First, a trained agent is able to obtain at least the extra return from working in a high skilled team as opposed to an unskilled team. Second, with probability $p_T + p_M$ the agent is able to become the skilled partner in a mixed team and attain an extra surplus amounting to $\widehat{a} - \underline{a}$. Third, with probability p_T the agent is able to obtain the return from training herself. These returns are offset against the surplus the agent attains when entering a mixed team as an unskilled partner with probablity $p_T + p_M$. In order to obtain more explicit expressions for the probabilities we need to distinguish the following two cases.

Case 1: $T \geq \frac{1+\lambda}{2}$. There is a surplus of skilled workers and we obtain

$$\begin{aligned}
p_T &= \frac{\widehat{T}\sigma_y}{T\sigma_o} = \frac{\widehat{T}\lambda}{T} \\
p_M &= \frac{\overbrace{(1-T)\,\sigma_o + \sigma_y}^{\text{unskilled}} - \overbrace{\widehat{T}\sigma_y}^{\text{unskilled receiving training}}}{T\sigma_o} = \frac{1-T+\left(1-\widehat{T}\right)\lambda}{T}, \\
p_T + p_M &= \frac{1-T+\lambda}{T}, \\
p_{NT} &= 1,
\end{aligned}$$

where we assume again that $\widehat{T} \leq T$. Here, unskilled agents are always able to find a skilled partner, while the same is not true for skilled agents. Consequently,

$$\begin{aligned}
\overline{\pi}_T^y\left(\delta, b, a, \widehat{\delta}, \widehat{b}, \widehat{a}\right) &= \pi_T^y \Big| T \geq \frac{1+\lambda}{2} \\
&= -\frac{\delta}{2} - b + a + \rho\left[\underline{a} + \widehat{a} + \frac{\widehat{T}\lambda}{T}\left(\widehat{b} - \underline{b}\left(\widehat{\delta}, \widehat{a}\right)\right) + \frac{1-T+\lambda}{T}\left(\widehat{a} - \underline{a}\right)\right]
\end{aligned}$$

Case 2: $T \leq \frac{1+\lambda}{2}$. There is a surplus of unskilled workers and we obtain

$$\begin{aligned}
p_T &= \frac{\widehat{T}\lambda}{T}, \qquad p_M = 1 - \frac{\widehat{T}\lambda}{T}, \qquad p_T + p_M = 1, \\
p_{NT} &= \frac{\overbrace{T\sigma_o - \widehat{T}\sigma_y}^{\text{skilled who do not train}}}{\underbrace{(1-T)\,\sigma_o + \left(1-\widehat{T}\right)\sigma_y}_{\text{unskilled}}} = \frac{T - \widehat{T}\lambda}{(1-T) + \left(1-\widehat{T}\right)\lambda}.
\end{aligned}$$

In this case, skilled agents are always able to find an unskilled partner, while the same is not true for unskilled agents. Note, however, that due to the shortage of young workers not all skilled agents are able to train. Consequently,

$$\begin{aligned}\underline{\pi}_T^y\left(\delta, b, a, \widehat{\delta}, \widehat{b}, \widehat{a}\right) &= \pi_T^y \Big| T \leq \frac{1+\lambda}{2} \\ &= \left\{ -\frac{\delta}{2} - b + a + \rho \left[\begin{array}{c} \overline{a} + \widehat{a} + \frac{\widehat{T}\lambda}{T}\left(\widehat{b} - \underline{b}\left(\widehat{\delta}, \widehat{a}\right)\right) \\ -\frac{T-\widehat{T}\lambda}{(1-T)+\left(1-\widehat{T}\right)\lambda}\left(\overline{a} - \widehat{a}\right) \end{array} \right] \right\}.\end{aligned}$$

We can now proceed to determine the steady state allocation of training. In order to facilitate the analysis we focus on the case where the training fee is determined competitively.

4.2.1 Training in the presence of price competition

In a steady state, it must be true that $T = \widehat{T}$, $a = \widehat{a}$ and $b = \widehat{b}$. Price competition may arise in either or both of the markets for mixed partnerships and for training. We begin by considering the allocation of training when there is price competition with respect to b, while taking as given for the moment the price a. We will subsequently study the impact of (further) price competition with respect to a. As in the previous case with segregated teams, a shrinking population implies an excess supply of potential trainers in a steady state, where for $T = \widehat{T}$ the conditional probability of finding a trainee is given by $p_T = \lambda < 1$. But then, Bertrand competition for trainees implies that $b = \underline{b}(\delta, a)$. The surplus from the supply of training is exctracted to the point that skilled agents are indifferent between supplying training or taking on a junior partner in a mixed team. Naturally, this holds for the future as well so that $\widehat{b} = \underline{b}\left(\widehat{\delta}, \widehat{a}\right)$. But then we obtain

$$\overline{\pi}_T^y\left(\overset{-}{\delta}, \overset{+}{a}, \overset{-}{T}, \overset{+}{\lambda}\right) = -\delta + \rho\left[\underline{a} + a + \frac{(1-T+\lambda)}{T}(a - \underline{a})\right], \tag{21}$$

$$\underline{\pi}_T^y\left(\overset{-}{\delta}, \overset{+}{a}, \overset{-}{T}, \overset{+}{\lambda}\right) = -\delta + \rho\left[\overline{a} + a - \frac{(1-\lambda)T}{(1-T)(1+\lambda)}(\overline{a} - a)\right], \tag{22}$$

where $\underline{a} = \frac{\theta_{h,h} - \theta_{l,h}}{2}$ and $\overline{a} = \frac{\theta_{l,h} - \theta_{l,l}}{2}$. As one would expect the net surplus from becoming trained increases in a irrespective of whether or not the equilibrium rate of training generates a surplus or shortfall of skilled workers. Furthermore, in both cases the net surplus from receiving training falls in the rate of training and increase in the reproduction rate. Again this is intuitive. A higher rate of training implies more competition between (future) skilled agents, thus lowering the probability of becoming the skilled partner in a mixed team and acquiring the associated surplus. Likewise a higher rate of reproduction increases the number of unskilled partners and, thus, the probability of a (future) skilled agent to lead a mixed team.

Evoking again Assumption A, training will only be demanded by young agents whose individual training cost δ is low enough to guarantee $\overline{\pi}_T^y(\cdot) \geq 0$ and $\underline{\pi}_T^y(\cdot) \geq 0$, respectively. Setting to zero the expressions in (21) and (22), respectively, and solving for δ gives

$$\delta = \overline{T}^D(a,\lambda) = \rho\underline{a} + \sqrt{(\rho\underline{a})^2 + \rho(1+\lambda)(a-\underline{a})}$$

if $T \geq \frac{1+\lambda}{2}$ and

$$\begin{aligned} \delta &= \underline{T}^D(a,\lambda) = C - \sqrt{C^2 - \rho(\overline{a}+a)} \\ C &:= \frac{1}{2} + \frac{\rho(\overline{a}+\lambda a)}{(1+\lambda)} \end{aligned}$$

if $T \leq \frac{1+\lambda}{2}$. It can then be verified that the rate of training is given by

$$T^D(a,\lambda) = \begin{cases} \underline{T}^D\left(\overset{+}{a},\overset{+}{\lambda}\right) & \text{for} \quad a \leq \frac{1+\lambda}{4\rho} \\ \overline{T}^D\left(\overset{+}{a},\overset{+}{\lambda}\right) & \text{for} \quad a \geq \frac{1+\lambda}{4\rho} \end{cases}.$$

It can be checked that the rate of training increases both in the transfer a and in the rate of reproduction λ, where an increase in either implies a higher (expected) future return to training. Thus, in this case, too, a shrinking population implies a lower rate of training. It remains to be checked whether the boundary value $\frac{1+\lambda}{4\rho}$ falls within the interval $[\underline{a},\overline{a}]$. It can be shown that

$$T^D(a,\lambda) = \begin{cases} \underline{T}^D(a,\lambda) \quad \forall a \in [\underline{a},\overline{a}] \quad \text{if} \quad \rho \leq \frac{1+\lambda}{2(\theta_{l,h}-\theta_{l,l})} \\ \begin{cases} \underline{T}^D(a,\lambda) & \text{for} \quad a \in \left[\underline{a}, \frac{1+\lambda}{4\rho}\right] \\ \overline{T}^D(a,\lambda) & \text{for} \quad a \in \left[\frac{1+\lambda}{4\rho}, \overline{a}\right] \end{cases} \quad \text{if} \quad \rho \in \left[\frac{1+\lambda}{2(\theta_{l,h}-\theta_{l,l})}, \frac{1+\lambda}{2(\theta_{h,h}-\theta_{l,h})}\right] \\ \overline{T}^D(a,\lambda) \quad \forall a \in [\underline{a},\overline{a}] \quad \text{if} \quad \rho \geq \frac{1+\lambda}{2(\theta_{h,h}-\theta_{l,h})} \end{cases}.$$

We can now compare this with the social optimum, which is given by

$$T^S = \begin{cases} \underline{T}^S \Leftrightarrow \rho \leq \frac{1+\lambda}{2(\theta_{l,h}-\theta_{l,l})} \\ \frac{1+\lambda}{2} \Leftrightarrow \rho \in \left[\frac{1+\lambda}{2(\theta_{l,h}-\theta_{l,l})}, \frac{1+\lambda}{2(\theta_{h,h}-\theta_{l,h})}\right] \\ \overline{T}^S \Leftrightarrow \rho \geq \frac{1+\lambda}{2(\theta_{h,h}-\theta_{l,h})} \end{cases}.$$

with $\underline{T}^S = \rho(\theta_{l,h} - \theta_{l,l})$ and $\overline{T}^S = \rho(\theta_{h,h} - \theta_{l,h})$. The following can be shown.

Lemma 3 *(i) For* $\rho \leq \frac{1+\lambda}{2(\theta_{l,h}-\theta_{l,l})}$ *it is true that* $T^D(a,\lambda) = \underline{T}^D(a,\lambda) \leq \underline{T}^D(\overline{a},\lambda) = \underline{T}^S = T^S$.

(ii) For $\rho \in \left[\frac{1+\lambda}{2(\theta_{l,h}-\theta_{l,l})}, \frac{1+\lambda}{2(\theta_{h,h}-\theta_{l,h})}\right]$ *it is true that* $T^D(a,\lambda) = \underline{T}^D(a,\lambda) \leq \frac{1+\lambda}{2} = T^S$ *for all* $a \in \left[\underline{a}, \frac{1+\lambda}{4\rho}\right]$ *and* $T^D(a,\lambda) = \overline{T}^D(a,\lambda) \geq \frac{1+\lambda}{2} = T^S$ *for all* $a \in \left[\frac{1+\lambda}{4\rho}, \overline{a}\right]$.

(iii) For $\rho \geq \frac{1+\lambda}{2(\theta_{h,h}-\theta_{l,h})}$ *it is true that* $T^D(a,\lambda) = \overline{T}^D(a,\lambda) \geq \overline{T}^D(\underline{a},\lambda) = \overline{T}^S = T^S$.

Thus, there is scope for over-provision (under-provision) of training, when the discount factor is low (high). For intermediate levels of the discount rate, there is scope for under-provision (over-provision) when a is low (high). Therefore, the question how the fee for mixed teams without training is determined turns out to be crucial for the efficiency of training. We can envisage several scenarios. For instance, the presence of a minimum wage for unskilled workers, which implies an upper-bound on a would curb a potential over-provision of training in case (iii) yet exarcerbate under-provision in case (i).

As an alternative, consider now that the fee a is determined competitively. Specifically, we would expect a to increase (decrease) whenever there is an excess demand for (supply of) skilled workers, i.e. whenever $T < \frac{1+\lambda}{2}$ ($T > \frac{1+\lambda}{2}$). The following then applies

$$a = \begin{cases} \overline{a} \Leftrightarrow \rho \leq \frac{1+\lambda}{2(\theta_{l,h}-\theta_{l,l})} \\ \frac{1+\lambda}{4\rho} \Leftrightarrow \rho \in \left[\frac{1+\lambda}{2(\theta_{l,h}-\theta_{l,l})}, \frac{1+\lambda}{2(\theta_{h,h}-\theta_{l,h})}\right] \\ \underline{a} \Leftrightarrow \rho \geq \frac{1+\lambda}{2(\theta_{h,h}-\theta_{l,h})} \end{cases}.$$

Hence, in situations, where a low (high) discount factor leads to a perpetual under-supply (over-supply) of trained partners, the entry fee into mixed partnerships increases up to the maximum level (decreases down to the minimum level). For intermediate levels of the discount factor, the fee adjusts so as to balance the number of skilled and unskilled workers.Drawing on the previous Lemma, we can now establish the following result.

Proposition 4 *If both the training fee and the entry fee into mixed partnerships adjust competitively, the resulting allocation of training is efficient.*

Similar to the case with $\widehat{\theta} > 0$ the competition for trainees eliminates all current and future profit from the provision of training. The same does not apply with respect to the entry fee into a mixed team as long as there is not a surplus of skilled workers. Hence, for $T \leq \frac{1+\lambda}{2}$ a skilled agent is able to make a profit by charging an entry fee from unskilled workers. However, competition implies that this profit exactly matches up the private with the social return to training. In contrast, for $T \geq \frac{1+\lambda}{2}$, unskilled agents are able to make a profit when entering a mixed team. This profit reduces the private net return to training exactly to the level of the social return.

5 Conclusions

We have studied private and social incentives to provide training within partnerships under the condition of a shrinking workforce. We show that population ageing reduces the total incidence of training (per capita) due to a demographic effect (a lower number of potential trainees per capita) and an economic effect (a lower rate of training). The reason is that the competition for young and/or unskilled partners becomes tougher within an ageing workforce. This tends to erode the returns to training whenever they include the profit from selling on training to the next cohort and/or from becoming the senior partner in a mixed partnership. In such a case, the current demand for training drops even beyond what would be predicted on the basis of demographics alone.

If the fees for training and, where relevant, the entry fee for unskilled workers into mixed teams are set competitively, then training activities tend to be efficient even within a shrinking population. This is because any expected returns to training are eroded in the competition of skilled-old workers for the lower number of young trainees. Therefore, there are no private rents associated with the acquisition of training and the optimum is attained even in a decentralised setting. If production takes place in mixed teams, a further condition for efficiency is that the entry fee into such teams adjusts competitively to inbalances between the number of skilled or unskilled workers. If a uniform training fee is the outcome of a bargaining process that takes place before partners are matched then an efficient allocation is generally not attained. Both over- or under-supply of training may occur, the former being more likely in the presence of equal bargaining power and under circumstances of a stable population or only moderate decline. Thus, within an ageing workforce the under-provision of training becomes somewhat more likely.

While we have cast our model in terms of training, the idea extends to other forms of investments that are undertaken to attain/maintain a high value of the partnership business. The value of the partnership business may be based on an advanced technology and/or on a reputation for high quality. Thus, the cost δ may reflect (regular) investments in order to replace obsolete technology technology with a view to remaining efficient and/or maintaining quality. Upon taking over an established (i.e. valuable) business the junior partner will have to compensate the leaving partner for the investments undertaken. However, he is the less inclined to purchase the established business (rather than open up a less valuable business on his own) the lower the probability that he is able to sell on the established business in the future. Within an ageing profession the scope for selling on such businesses declines over time, thus, eroding current demand and the rate of investment.

The model lends itself to a number of extensions. Apart from an analysis of the bargaining problem for the case in which mixed teams are formed, further analysis may take into account the implications of wealth constraints faced by young partners. Note that the purchase of training may well imply a fee that exceeds their share of output, i.e. $b > \frac{\theta_{l,h}}{2} - \delta$. In this case, wealth constraints imply either a maximum bound on the fee or a reduction in the demand for training below the level attained in the absence of wealth constraints. In either case, under-supply of training may occur. For this case, there is an obvious policy issue. One way of overcoming wealth constraints may lie in a mechanism, where training is initially provided at a low fee but young workers repay the (then retired) old workers out of their future earnings. While this would necessitate a three-generation model (two active - one retired) such an analysis may be insightful.

Another issue amenable to extensions arises from our assumption that productivities do not change over time. In particular, this implies that in the absence of training the economy actually loses productivity. Obviously, this runs counter to most empirical evidence. The issue may be resolved by assuming that there is a positive trend on the productivity of young workers, which would capture the progress of education. Our analysis remains valid in the sense that even then the provision of training by skilled old workers would further enhance the capabilities of young workers, generating thus a further productivity gain. We would thus expect the inclusion of a growth trend of productivity into the model not to severly alter our results.

Finally, the matching process in our model is frictionless a la Becker (1973). There is naturally scope for extending the model towards frictitious matching, a la Shimer and Smith (2000) who extend Becker's (1973) analysis to allow for time-intensive partner search.

6 Appendix

Derivation of equation (3): Consider $1+\bar{t}$, $\bar{t} \in [1,\infty)$ periods. For an initial population $N(0) \equiv 1$, the population in each period t is described by

$$\sigma_y(t) = \frac{\lambda^{t+1}}{1+\lambda}, \qquad \sigma_o(t) = \frac{\lambda^t}{1+\lambda}.$$

Furthermore, assuming a steady state training rate of $T(t) = T\ \forall t \in [0,\bar{t}-1]$ and $T(\bar{t}) = 0$ we obtain the following skill distribution among old agents for each period $t \in [t,\bar{t}]$

$$\sigma_{o^s}(t) = T\frac{\lambda^t}{1+\lambda}, \qquad \sigma_{o^u}(t) = (1-T)\frac{\lambda^t}{1+\lambda}.$$

Furthermore, we assume $\sigma_{o^s}(0) = \frac{1}{1+\lambda}$ and, corrsespondingly, $\sigma_{o^u}(0) = 0$. The incidence of training for each period $t \in [0,\bar{t}-1]$ is given by $T\frac{\lambda^{t+1}}{1+\lambda}$. On the basis of this we can calculate the corresponding team/output structure for each period. The following table 1 provides an illustration for an exemplary number of periods.

	$\theta_{h,h}$	$\theta_{l,l}$	$\theta_{l,h} - \frac{T}{2}$
0	$\frac{1-T\lambda}{2(1+\lambda)}$	$\frac{(1-T)\lambda}{2(1+\lambda)}$	$\frac{T\lambda}{1+\lambda}$
1	$\frac{T(1-\lambda)\lambda}{2(1+\lambda)}$	$\frac{(1-T)\lambda}{2}$	$\frac{T\lambda^2}{1+\lambda}$
$t \in [2,\bar{t}-1]$	$\frac{T(1-\lambda)\lambda^t}{2(1+\lambda)}$	$\frac{(1-T)\lambda^t}{2}$	$\frac{T\lambda^{t+1}}{1+\lambda}$
$\bar{t}$	$\frac{T\lambda^{\bar{t}}}{2(1+\lambda)}$	$\frac{(1+\lambda-T)\lambda^{\bar{t}}}{2(1+\lambda)}$	0

Table 1: Output structure for $\widehat{\theta} > 0$.

When calculating the number of segregated teams ($\theta_{h,h}$ or $\theta_{l,l}$) we need to divide by two the total population from which these teams are drawn. This is not neccessary for the mixed teams as their number is represented by the incidence of training. We evaluate the output of mixed teams with training at its average $\theta_{l,h} - \frac{T}{2}$. Adding up the output levels from table 1 across teams and periods $t \in [0, \bar{t}]$ while applying a discount factor ρ we obtain

$$\Pi(T, \lambda, \bar{t}) = \left\{ \begin{array}{c} (1 - T\lambda)\frac{\theta_{h,h}}{2} + (1 - T)\lambda\frac{\theta_{l,l}}{2} \\ + \left(1 + \rho\lambda + ... + (\rho\lambda)^{\bar{t}-1}\right)\lambda T\left(\theta_{l,h} - \frac{T}{2}\right) \\ + \left(1 + \rho\lambda + ... + (\rho\lambda)^{\bar{t}-1}\right)\left[\begin{array}{c} T(1 - \lambda)\frac{\theta_{h,h}}{2} \\ + (1 - T)(1 + \lambda)\frac{\theta_{l,l}}{2} \end{array} \right] \\ - \left[T(1 - \lambda)\frac{\theta_{h,h}}{2} + (1 - T)(1 + \lambda)\frac{\theta_{l,l}}{2}\right] \\ + (\rho\lambda)^{\bar{t}}\left[T\frac{\theta_{h,h}}{2} + (1 + \lambda - T)\frac{\theta_{l,l}}{2}\right] \end{array} \right\}(1 + \lambda)^{-1}.$$

Collecting the terms including T, applying the appropriate operations for series, and collecting terms again, we obtain from this the expression in (3).■

Derivation of equations (6) and (7): On the basis of the population structure provided in the previous proof we can calculate the corresponding team/output structure for each period. The following tables 2 and 3 provide an illustration for an exemplary number of periods for the cases $T \in \left[0, \frac{1+\lambda}{2}\right]$ and $T \in \left[\frac{1+\lambda}{2}, 1\right]$, respectively.

	$\theta_{h,h}$	$\theta_{l,l}$	$\theta_{l,h}$	$\theta_{l,h} - \frac{T}{2}$
0	$\frac{1-\lambda}{2(1+\lambda)}$	0	$\frac{(1-T)\lambda}{1+\lambda}$	$\frac{T\lambda}{1+\lambda}$
1	0	$\frac{(1+\lambda-2T)\lambda}{2(1+\lambda)}$	$\frac{T(1-\lambda)\lambda}{1+\lambda}$	$\frac{T\lambda^2}{1+\lambda}$
$t \in [2, \bar{t} - 1]$	0	$\frac{(1+\lambda-2T)\lambda^t}{2(1+\lambda)}$	$\frac{T(1-\lambda)\lambda^t}{1+\lambda}$	$\frac{T\lambda^{t+1}}{1+\lambda}$
$\bar{t}$	0	$\frac{(1+\lambda-2T)\lambda^{\bar{t}}}{2(1+\lambda)}$	$\frac{T\lambda^{\bar{t}}}{1+\lambda}$	0

Table 2: Output structure for $\widehat{\theta} < 0$ and $T \in \left[0, \frac{1+\lambda}{2}\right]$.

	$\theta_{h,h}$	$\theta_{l,l}$	$\theta_{l,h}$	$\theta_{l,h}-\frac{T}{2}$
0	$\frac{1-\lambda}{2(1+\lambda)}$	0	$\frac{(1-T)\lambda}{1+\lambda}$	$\frac{T\lambda}{1+\lambda}$
1	$\frac{(2T-1-\lambda)\lambda}{2(1+\lambda)}$	0	$(1-T)\lambda$	$\frac{T\lambda^2}{1+\lambda}$
$t\in\left[2,\bar{t}-1\right]$	$\frac{(2T-1-\lambda)\lambda^t}{2(1+\lambda)}$	0	$(1-T)\lambda^t$	$\frac{T\lambda^{t+1}}{1+\lambda}$
$\bar{t}$	$\frac{(2T-1-\lambda)\lambda^{\bar{t}}}{2(1+\lambda)}$	0	$\frac{(1+\lambda-T)\lambda^{\bar{t}}}{1+\lambda}$	0

Table 3: Output structure for $\widehat{\theta}<0$ and $T\in\left[\frac{1+\lambda}{2},1\right]$.

Adding up the output levels from tables 2 and 3, respectively, across teams and periods $t\in\left[0,\bar{t}\right]$ while using a discount factor ρ, and then following the same steps as in the derivation of (3) we obtain the expressions in (6) and (7), respectively.■

Derivation of equation (19): From an ex-ante perspective old workers cannot be sure whether or not they are able to attract a young worker to train, neither do they know the training cost. Using (8) we can write the expected surplus for a skilled agent from providing training at fee b as

$$\frac{\theta_{h,h}}{2}+\frac{\lambda T}{\overline{T}}\pi_T^o\left(\frac{T}{2},b\right),$$

where we assume without further loss of generality that the current share of skilled workers, $\overline{T}$, does not exceed the training rate, T implying the conditional probability $\frac{\lambda T}{\overline{T}}\leq 1$ for a skilled agent to find a trainee. As before, $\frac{T}{2}$ is the expected training cost, when training takes place at rate T. Furthermore, observing $\pi_T^o\left(\frac{T}{2},b\right)=\pi_T^o\left(T,b\right)+\frac{T}{4}$, we can express the expected surplus for a skilled agent as $\frac{\theta_{h,h}}{2}+\frac{\lambda T}{\overline{T}}\left[\pi_T^o\left(T,b\right)+\frac{T}{4}\right]$. The total (expected) surplus of old agents (skilled and unskilled) is then given by

$$\begin{aligned}\pi^o\quad:&=\left\langle\overline{T}\left\{\frac{\theta_{h,h}}{2}+\frac{\lambda T}{\overline{T}}\left[\pi_T^o\left(T,b\right)+\frac{T}{4}\right]\right\}+\left(1-\overline{T}\right)\frac{\theta_{l,l}}{2}\right\rangle\sigma_o\\&=\left\langle\frac{\theta_{l,l}}{2}+\overline{T}\frac{\theta_{h,h}-\theta_{l,l}}{2}+\lambda T\left[\pi_T^o\left(T,b\right)+\frac{T}{4}\right]\right\rangle\sigma_o,\end{aligned}\tag{23}$$

where $\overline{T}\sigma_o$ and $\left(1-\overline{T}\right)\sigma_o$ denotes the number of skilled and unskilled old agents respectively. Consider now the expected surplus of young workers.

Aggregating over the types δ we obtain the total surplus of young workers as

$$
\begin{aligned}
\pi^{y} &= \left[\int_{0}^{T} \pi_T^{y}\left(\delta, b, T, \widehat{b}, \widehat{T},\right) + (1+\rho)\frac{\theta_{l,l}}{2}\right]\sigma_y \\
&= \left[T\pi_T^{y}\left(\frac{T}{2}, b, T, \widehat{b}, \widehat{T},\right) + (1+\rho)\frac{\theta_{l,l}}{2}\right]\sigma_y \\
&= \left[T\left[\pi_T^{y}\left(T, b, T, \widehat{b}, \widehat{T},\right) + \frac{T}{4}\right] + (1+\rho)\frac{\theta_{l,l}}{2}\right]\sigma_y. \qquad (24)
\end{aligned}
$$

Using the outside utilities $\frac{\theta_{l,l}}{2} + \overline{T}\frac{\theta_{h,h}-\theta_{l,l}}{2}$ for the old and $(1+\rho)\frac{\theta_{l,l}}{2}$ for the young, respectively, together with (23) and (24) we can write the Nash product as

$$
\begin{aligned}
&\left[\pi^{o} - \left(\frac{\theta_{l,l}}{2} + \overline{T}\frac{\theta_{h,h}-\theta_{l,l}}{2}\right)\right]^{\beta}\left[\pi^{y} - (1+\rho)\frac{\theta_{l,l}}{2}\right]^{1-\beta} \\
= \; &\lambda\sigma_o T\left[\pi_T^{o}(T,b) + \frac{T}{4}\right]^{\beta}\left[\pi_T^{y}\left(T, b, T, \widehat{b}, \widehat{T},\right) + \frac{T}{4}\right]^{1-\beta}
\end{aligned}
$$

where $T = T(b,\lambda)$ as given in 15.■

Proof of Lemma 2: The first-order condition to the problem in (19) can be expessed as

$$
Z(b,\beta) = \left\langle \begin{array}{c} T\left\{\beta\left(\pi_T^{y} + \frac{T}{4}\right) - (1-\beta)\left(\pi_T^{o} + \frac{T}{4}\right) + \frac{T_b}{4}\left[\beta\pi_T^{o} + (1-\beta)\pi_T^{y}\right]\right\} \\ -(1-\beta)\left(\pi_T^{o} + \frac{T}{4}\right)\frac{B}{T}T_b \end{array}\right\rangle = 0. \qquad (25)
$$

The second-order condition can be shown to be satisfied at the relevant equilibria. Recall from (13) and (14) that either $\pi_T^{y} \geq \pi_T^{o} = 0$ or $\pi_T^{o} \geq \pi_T^{y} = 0$ will obtain for any (b,β). We consider these cases in turn.

Case 1: $\pi_T^{y} \geq \pi_T^{o} = 0$: Using this together with $T_b = 2$, as from (18), in (25) we obtain

$$
Z(b,\beta) = \frac{T}{4}\left\{-(1-2\beta)T + 2(1+\beta)\pi_T^{y}\right\} - (1-\beta)\frac{B}{2} = 0. \qquad (26)
$$

Setting $\pi_T^{o} = 0$ in (11) and solving for b we obtain $b = \frac{T}{2} + \frac{\theta_{h,h}-\theta_{l,h}}{2}$. Substituting for b in the expression for $\pi_T^{y}\left(T, b, \widehat{T}, \widehat{b}\right)$ we then obtain $\pi_T^{y} =$

$A-\frac{\theta_{h,h}-\theta_{l,h}}{2}-T+\frac{B}{T}$, with A and B as defined in (16) and (17), or $\pi_T^y=T^S-T+\frac{B}{T}$, where we recall from (4) that $T^S=A-\frac{\theta_{h,h}-\theta_{l,h}}{2}$. Substituting into (26) and rearranging we thus obtain

$$Z(b,\beta)=\frac{T}{4}\left\{2(1+\beta)T^S-3T\right\}+\beta B=0.$$

In a steady-state, it must be true that $\pi_T^o\left(\widehat{T},\widehat{b}\right)=\pi_T^o(T,b)=0$ and $\widehat{T}=T$. Using this in (17) we obtain $B=\rho\lambda\frac{T^2}{4}$ and, thus, $Z(b,\beta)=\frac{T}{4}\left\{2(1+\beta)T^S-3T\right\}+\beta\rho\lambda\frac{T^2}{4}=0$. Solving for T gives us $T^B=T_o^B(\beta,\lambda)=\frac{2(1+\beta)}{3-\beta\lambda\rho}T^S$ as reported in part (i) of the Lemma. We need to verify now that $\pi_T^y\geq 0$ holds for this outcome. Writing $\pi_T^y=T^S-T_o^B(\beta,\lambda)+\rho\lambda\frac{T_o^B(\beta,\lambda)}{4}$ it is readily checked that the RHS of this expression is non-negative if and only if $\beta\leq\frac{2+\lambda\rho}{4+\lambda\rho}=\overline{\beta}$.

Case 2: $\pi_T^o\geq\pi_T^y=0$: Using this together with $T_b=-\frac{2T^2}{T^2+2B}$, as from (18), in (25) we obtain after some rearrangements

$$Z(b,\beta)=\frac{-T^2}{4(T^2+2B)}\left\{(1-2\beta)T^2-2\beta B+2(2-\beta)T\pi_T^o\right\}=0,\quad(27)$$

which holds if and only if the equation in bracelets adds up to zero. Setting $\pi_T^y=0$ and solving for b we obtain $b=A-\frac{T}{2}+\frac{B}{T}$,with A and B as defined in(16)and(17). Substituting for b in the expression for $\pi_T^o(T,b)$ we then obtain $\pi_T^o=T^S-T+\frac{B}{T}$, where we use again $T^S=A-\frac{\theta_{h,h}-\theta_{l,h}}{2}$. Substituting into the equation in bracelets in (27) and rearranging we thus obtain

$$Z(b,\beta)=0\Longleftrightarrow -3T^2+4(1-\beta)B+2(2-\beta)TT^S=0.$$

In a steady-state, it must be true that $\pi_T^o\left(\widehat{T},\widehat{b}\right)=\pi_T^o(T,b)=T^S-T+\frac{B}{T}$ and $\widehat{T}=T$. Using this in (17) we obtain $B=\rho\lambda T\left(T^S-\frac{3T}{4}+\frac{B}{T}\right)$ or $B=\frac{\rho\lambda T}{1-\rho\lambda}\left(T^S-\frac{3T}{4}\right)$. Thus,

$$Z(b,\beta)=0\Longleftrightarrow(1-\rho\lambda)^{-1}T\left\{-3(1-\beta\rho\lambda)T+2\left[2-\beta(1+\rho\lambda)\right]T^S\right\}=0$$

. Solving for T gives us $T^B=T_y^B(\beta,\lambda)=\frac{2[2-\beta(1+\lambda\rho)]}{3(1-\beta\lambda\rho)}T^S$ as reported in part (ii) of the Lemma. We need to verify now that $\pi_T^o\geq 0$ holds for this outcome.

Writing $\pi_T^o = T^S - T_y^B(\beta,\lambda) + \frac{\rho\lambda}{1-\rho\lambda}\left(T^S - \frac{3T_y^B(\beta,\lambda)}{4}\right)$ it is readily checked that the RHS of this expression is non-negative if and only if $\beta \geq \frac{2}{4+\lambda\rho} = \underline{\beta}$. Noting that $\underline{\beta} < \overline{\beta}$ the structure of possible outcomes follows as reported in the Lemma.This completes the proof.■

Proof of claim that $T_y^B(\beta,\lambda)$ and $T_o^B(\beta,\lambda)$ cannot be Pareto-ranked for $\beta \in \left[\underline{\beta},\overline{\beta}\right]$ (sketch): Consider the expected surplus from training $\widetilde{\pi}_T^o(T,b) := \lambda\sigma_o T\left[\pi_T^o(T,b) + \frac{T}{4}\right]$ falling to the old and the expected surplus from training $\widetilde{\pi}_T^y(T,b) := \lambda\sigma_o T\left[\pi_T^y\left(T,b,T,\widehat{b},\widehat{T},\right) + \frac{T}{4}\right]$ falling to the young, respectively. These can be expressed as functions of β, where Table 4 provides the relationships for the high fee equilibrium with $T = T_y^B(\beta,\lambda)$ and $b = b_y(\beta,\lambda)$ and low fee equilibrium with $T = T_o^B(\beta,\lambda)$ and $b = b_o(\beta,\lambda)$, respectively.[17]

	$\widetilde{\pi}_T^o(T,b)$	$\widetilde{\pi}_T^y(T,b)$
$\left\{T_y^B, b_y\right\}$	$\widetilde{\pi}_y^o(\beta) = \frac{T_y^B}{1-\lambda\rho}\left(T^S - \frac{3T_y^B}{4}\right)$ $\frac{d\widetilde{\pi}_y^o}{d\beta} = \frac{1}{1-\lambda\rho}\left(T^S - \frac{3T_y^B}{2}\right)\frac{dT_y^B}{d\beta} > 0$	$\widetilde{\pi}_y^y(\beta) = \frac{\left(T_y^B\right)^2}{4}$ $\frac{d\widetilde{\pi}_y^y}{d\beta} = \frac{T_y^B}{2}\frac{dT_y^B}{d\beta} < 0$
$\left\{T_o^B, b_o\right\}$	$\widetilde{\pi}_o^o(\beta) = \frac{\left(T_o^B\right)^2}{4}$ $\frac{d\widetilde{\pi}_o^o}{d\beta} = \frac{T_o^B}{2}\frac{dT_o^B}{d\beta} > 0$	$\widetilde{\pi}_o^y(\beta) = T_o^B\left[T^S - \frac{T_o^B}{4}(3-\lambda\rho)\right]$ $\frac{d\widetilde{\pi}_o^y}{d\beta} = \left[T^S - \frac{T_o^B}{2}(3-\lambda\rho)\right]\frac{dT_o^B}{d\beta} < 0$

Table 4.

The signs of the derivatives can be readily verified when using the relationships reported in Lemma 2 and Corollary 1 and observing $\beta \in \left[\underline{\beta},\overline{\beta}\right]$. The following can then be established

$$\widetilde{\pi}_y^o(\beta) \geq \widetilde{\pi}_y^o\left(\underline{\beta}\right) = \frac{T_y^B\left(\underline{\beta},\lambda\right)^2}{4} = \frac{T_o^B\left(\overline{\beta},\lambda\right)^2}{4} = \widetilde{\pi}_o^o\left(\overline{\beta}\right) \geq \widetilde{\pi}_o^o(\beta)$$

and

$$\widetilde{\pi}_y^y(\beta) \leq \widetilde{\pi}_y^y\left(\underline{\beta}\right) = \frac{T_y^B\left(\underline{\beta},\lambda\right)^2}{4} = \frac{T_o^B\left(\overline{\beta},\lambda\right)^2}{4} = \widetilde{\pi}_o^y\left(\overline{\beta}\right) \leq \widetilde{\pi}_o^y(\beta).$$

Hence, the old (young) always prefer the high fee equilibrium $\left\{T_y^B, b_y\right\}$ (low fee equilibrium $\left\{T_o^B, b_o\right\}$) which therefore cannot be Pareto-ranked.■

[17] In the table, we supress the arguments of the functions.

References

[1] Becker, G. (1973): A theory of marriage: Part I, Journal of Political Economy, 81, 813 - 846.

[2] Besley, T. and M. Ghatak (2006), Sorting with motivated agents: implications for school competition and teacher incentives, Journal of the European Economic Association, 4 (2-3), 404 - 414.

[3] Feichtinger, G., M. Winkler-Dworak, I. Freund and A. Prskawetz (2007): On the age dynamics of learned societies: taking the example of the Austrian Academy of Sciences, VID Working Paper 02/2007.

[4] Gall, T., P. Legros, and A. Newman (2006): The timing of education, Journal of the European Economic Association, 4 (2-3), 427 - 435.

[5] Keyfitz, N. (1973): Individual mobility in a stationary population, Population Studies 27, 335-352.

[6] Legros, P. and A.F. Newman (2003): Beauty is a beast, frog is a prince: assortative matching with nontransferabilitites, EconomicsWorking Paper #30, School for Social Science, Institute for Advanced Study, Princeton University.

[7] Morrison, A.D. and W.J. Wilhelm, Jr. (2004): Partnership firms, reputation, and human capital, American Economic Review 94, 1682-1692.

[8] Polanyi, M (1966): The tacit dimension, GardenCity: Doubleday.

[9] Shimer, R. and L. Smith (2000): Assortative matching and search, Econometrica, 68(2), 343 - 369.

[10] Tykvova, T. (2006): Who chooses whom? Syndication, skills and reputation, Review of Financial Economics, in press.

[11] Vaupel, J.W. (1981), Over-tenured universities: the mathematics of reduction, Management Science 27, 904-913.

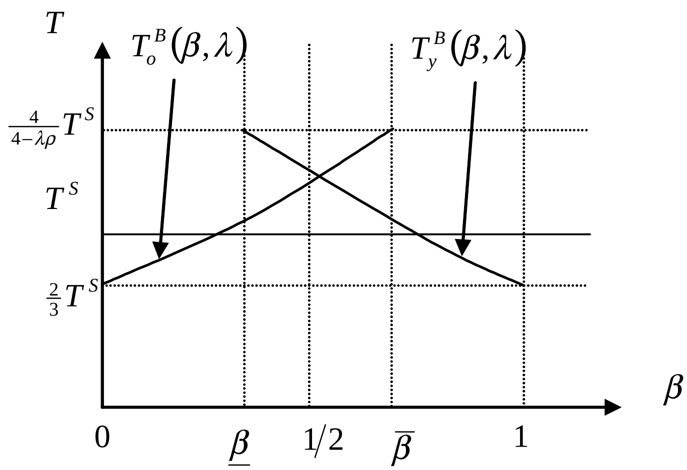

Figure 1

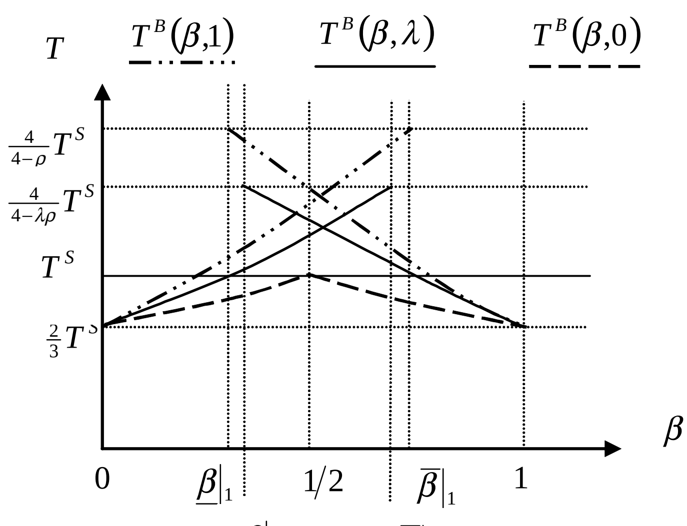

Figure 2