Deferred Training*

Michael Waldman†   Zhenda Yin‡

July 2023

Abstract

In standard human-capital investment models, firms train workers early during employment relationships. In real-world settings, however, training is frequently provided after workers have been employed for substantial time. Why is this the case given later training means a firm receives training benefits for fewer periods? We explore the dynamic allocation of training investments in general human capital, and identify three main factors that can result in deferred training: i) increasing returns to training with worker age; ii) human capital deterioration or obsolescence; and iii) poaching deterrence. In each case, we discuss the relevance of the factor for real-world labor markets.

Keywords: deferred training; general human capital; firm sponsored training

JEL Code: J2; J3; M5

*For valuable conversations, comments, and suggestions, we thank Dan Bernhardt, Peter Kuhn, Gabrielle Pepin, Michael Powell, James Spletzer, Pengfei Wang and seminar and conference participants at PKU, SUFE, China Economics Annual Conference (SYSU), Contemporary Labor Economics Symposium (Jinan), MEA-SOLE (MSU), and SOLE (virtual). All remaining errors are our own.

†Cornell University. Email: mw46@cornell.edu.

‡Peking University. Email: zdyin@pku.edu.cn.

1 Introduction

Economists have been interested in training investments since at least the publication of Becker's seminal work on human capital acquisition. Basically, there are two important views concerning the benefits of training. First, training investments lead to human capital accumulation, and in some cases a firm can reap future benefits from the increased productivity of its workforce. Focusing on this perspective, much of Becker's (1962; 1964) original focus was on how the nature of the human capital investments, whether general or firm specific, determines who finances the training—workers in the case of general human capital and shared financing when the human capital is firm specific. Second, training investments can lead to the revelation of worker attributes, such as worker abilities, which can lead to various returns. For instance, Autor (2001) finds that employers sometimes use the services of temporary help supply firms, because the temporary help firms are knowledgeable about their workers' abilities, and thus contracting with these firms improves selection and the matching of workers to job assignments.

One aspect of the mainstream literature on this topic is that little attention has been paid to the *timing* of firm sponsored training investments. Specifically, because training investments lead to subsequent returns, in standard models investments are made as early as possible to maximize those subsequent returns. However, substantial evidence indicates that training is not solely provided at the beginning of careers. Examples include firm sponsored career development programs, subsidized part-time MBA education, and sabbatical leaves that are frequently provided to experienced workers (Lazear and Oyer, 2013). Moreover, a number of empirical studies find systematic evidence supporting the idea that many workers receive training spanning their whole careers, as we discuss in Section 2.

As a complement to existing models of firm sponsored training, which mainly focus on training investments that occur early in careers, we explore the dynamic allocation of training investments; in particular, circumstances in which firms offer deferred training. Extending the classic analysis of Acemoglu and Pischke (1999b), our focus is investments in general training. We begin with a two-period baseline model in which firms can invest in general training and capture a return on their investments in each of the two periods. In particular, we consider two cases of the baseline model. In the first firms can commit to training levels while contracting, while in the second such commitment is not possible.

We find in our baseline model that, in both commitment and no-commitment cases, firms do not provide deferred training, i.e., when training investments are positive, all of the investment occurs in the first period. The logic is that, holding the total investment across the two periods fixed, it is always optimal to move investment dollars from period 2 to period 1, because the return would occur over a longer time period. In other words, the baseline model illustrates the standard argument in which firm sponsored training occurs early in careers, because that maximizes the number of periods in which the firm can reap returns from the investment.

We then enrich the baseline model to explore factors that can cause the logic from the baseline model to break down, and thus lead to firms deferring investments in general training. We identify three main factors: i) increasing returns to training with worker age; ii) human capital deterioration or obsolescence; and iii) poaching deterrence. In each case, in addition to formally showing that deferred training can be an equilibrium outcome, we discuss aspects of real-world labor markets that naturally cause the factor to be important, and thus lead to the emergence of deferred training.

As indicated, in our first enrichment, we allow the effect of training on worker output to increase with worker age. In the baseline model, training has the same effect on worker output in the two periods. Here we show that, if the effect of additional training on worker output is higher in period 2 than in period 1, then deferred training can arise in equilibrium. The logic is that the return to training at the beginning of any period depends to an extent on the productivity increase in the subsequent period. So, if the productivity increase associated with training is substantially higher in period 2, then there will be deferred training in equilibrium. In addition to formally showing this result, we discuss various factors that can cause the productivity increase associated with training to effectively rise with worker age such as learning by doing, promotions to higher level jobs, and employer learning of worker abilities.

In our second enrichment, we incorporate human capital deterioration or obsolescence into the model. By human capital deterioration or obsolescence, we mean that a dollar invested in training in period 1 has a bigger impact on the stock of human capital in period 1 than on the stock in period 2. This enrichment builds on insights in the well-known study of Ben-Porath (1967) which focuses on the time profile of human capital investments made by workers during their careers. Ben-Porath shows that investments are concentrated at early ages, but given human capital deterioration or obsolescence there are also investments later in careers. We show that Ben-Porath's basic insights also apply to the case of firm sponsored training. That is, if the rate of deterioration or obsolescence is sufficiently high, some of the investment will be deferred. The basic logic is that human capital deterioration or obsolescence decreases the stock of human capital in later periods, and thus increases the rate of return to investing in later periods. Besides extending Ben-Porath's basic insights to firm sponsored training, we also relate the results to various factors in real-world labor markets such as job changes within firms and technological advances which effectively engender human capital deterioration or obsolescence.

The factors identified in the previous two enrichments can be considered as *natural* reasons for deferred training, i.e., factors that arise as part of common human-resource management practices or due to an efficiency rationale. We next consider *strategic* reasons for delaying training, where firms defer training investments to limit the size of employee poaching from outside firms. In our third enrichment, we consider asymmetric employer learning, where incumbent firms obtain more precise information about worker ability than outsiders. In this case, training in period 1 signals worker ability to outsiders whenever incumbent firms invest more in high ability workers than in low ability workers. As a result, to compress outside wages and, in turn, capture informational rents in period 2, incumbent firms have incentives to hide information from outsiders by not varying training in period 1 across employees of heterogeneous ability levels. This type of signal-jamming strategy means that firms will invest at a low level in period 1 and defer some investment to period 2.

In our final enrichment, we modify the baseline model such that workers can quit after receiving training in period 2. Specifically, workers draw some random utility in period 1, where a bad draw can induce worker turnover in period 2. In this case, the baseline analysis suggests no deferred training. Yet, by requiring workers to partially repay training investments when they leave after receiving training in period 2, the repayment effectively limits the size of employee poaching from outside firms. In equilibrium, we show that firms can offer deferred training in period 2.

The outline for the paper is as follows. Section 2 discusses the relevant literature. Section 3 presents the baseline model, which we then analyze in Section 4. In Section 5, we consider enrichments of the baseline model, where the focus is identifying factors that can lead to deferred training. Section 6 concludes. Technical details can be found in the Appendix.

2 Related Literature

2.1 Theoretical Literature

The modern theory of human capital acquisition within firms begins with the seminal contribution of Becker (1962, 1964). There are two main arguments in Becker's analysis. First, the financing of firm specific human capital investments will be shared between workers and firms. Second, firms will not finance or sponsor investments in general training because of an inability to collect returns from such investments. In other words, classic human capital theory suggests that workers themselves will finance investments in general training.

Contrary to the classic theory of human capital, numerous empirical studies find that firms frequently sponsor general training for their workers (Lynch and Black, 1998; Loewenstein and Spletzer, 1998, 1999; Acemoglu and Pischke, 1998; Autor, 2001; Cappelli, 2004). A number of papers provide theoretical explanations for this phenomenon.[1] For instance, several papers focus on the informational advantage of incumbent employers. Autor (2001) argues that temporary help firms have an informational advantage concerning the abilities of temporary workers. As

[1] There are also theoretical studies concerning firms sponsoring investments in firm specific training such as Hashimoto (1981), Kahn and Huberman (1988), and Scoones and Bernhardt (1998).

a result, these firms earn information rents by efficiently matching temporary workers to job assignments. Moreover, this informational advantage becomes even more valuable when the provision of free training allows temporary help firms to attract higher ability workers. Acemoglu and Pischke (1998), Chang and Wang (1996), and Katz and Ziderman (1990) argue that if current employers are better informed about a worker's abilities or skills than prospective employers, they may finance general training investments to capture information rents in the future. This is because information asymmetry turns general skills into firm specific skills in the sense that poaching wage offers made to trained workers do not reflect the full marginal products of the workers.[2]

Our paper is closest to Acemoglu and Pischke (1999b). That paper generalizes results found in several other papers (Acemoglu, 1997; Acemoglu and Pischke, 1998, 1999a, 2003) by identifying a single unifying idea for why firms invest in general training. Specifically, firms sponsor general training investments when wages in the labor market are compressed relative to workers' marginal products in the sense that general training raises productivity at the incumbent employer more than it increases outside wage offers. They further establish that this type of wage compression can arise from various plausible mechanisms, e.g., search costs, asymmetric information, complementarity between general human capital accumulation and firm specific human capital, efficiency wages, and unions. Note that in Acemoglu and Pischke's analysis firms can only invest in training at the beginning of workers' careers. Our paper generalizes their analysis by investigating the dynamic allocation of training investments.

Our paper is also related to Garicano and Rayo (2017) concerning apprenticeships. That analysis focuses on optimal contracting when an important aspect of the relationship is the speed with which the employer transfers knowledge to the apprentice. They show that, given a credit constrained apprentice, an optimal contract can entail a gradual transfer of knowledge, rather than all the employer's knowledge being transferred to the apprentice at the beginning of the relationship. This arises in equilibrium because gradual knowledge transfer stops the apprentice from exiting the relationship prior to the employer being fully compensated for the transfer. Note that this is a type of deferred training, but only in a limited way—the training does not all occur at the very beginning of careers, but it is concentrated over the first part of careers. In contrast, our focus is deferred training which can arise even later in careers, which is inconsistent with apprenticeships, but the empirical evidence discussed next indicates that it is an important aspect of real-world labor markets.

Other related papers include Balmaceda (2005), Kessler and Lülfesmann (2006), and Carter (2021). The first two papers explore the implications of complementarity between firm spe-

[2] Waldman (1990) makes a related argument concerning the use of up-or-out contracts. Also, see Bar-Isaac and Leaver (2022) for a related analysis in which asymmetric information endogenously arises with the result that firms provide general training in equilibrium.

cific and general training, while Carter's paper focuses on deferred training that results because of uncertainty concerning the quality of worker-firm matching. For example, in Balmaceda's model, training that increases a worker's firm specific human capital increases the marginal returns to investing in general training. When workers and firms first invest in firm specific training and firms subsequently choose general training levels, both parties favor high investments in firm specific training because of the subsequent positive effect on investments in general training. Note, however, that in Balmaceda's two-period model, deferred training is the result of the assumed sequence of moves—workers and firms first choose investments in firm specific training, followed by firms choosing general training levels. In Carter's two-period model, firms acquire information about the firm-specific matches of untrained workers in period 1 and use this information to decide which workers will receive (delayed) training in period 2. In contrast, our model allows firms to make training investments in any period and whether or not deferred training is provided is an equilibrium outcome rather than an assumption of the model.

2.2 Empirical Literature

Inspired by classic human capital theory, early empirical research mostly focused on training near the beginning of employment relationships. For example, Barron, Black, and Loewenstein (1987, 1989) and Brown (1989) assess the impact of training received by workers soon after being hired.

Subsequent studies have focused more on training received by workers later in employment relationships. Using the Current Population Survey and the National Longitudinal Survey of Youth, Loewenstein and Spletzer (1997) document that in the US setting formal training frequently does not start at the beginning of an employment relationship, but rather after the worker has spent significant time at the firm. Similarly, Pischke (2001) uses data from a survey focused on the German apprenticeship training system to explore the practice of continuous training. He documents that 28% of the workers in the survey took at least one employer sponsored training course during the three years prior to the survey, and many of them were likely longer tenured workers. Melero (2010) also finds results concerning training and promotion rates in the US consistent with training being provided to long tenured workers.

There are two studies that directly estimate the provision of formal employer sponsored training, where attention is paid to the timing of training during the employment relationship. Using data on professional and technical employees at a large US manufacturing company, Bartel (1995) finds that training declines with tenure but is still significant for workers with high levels of tenure. More precisely, she finds that 72% of workers at the firm with one to two years of tenure receive training, while that figure falls to a still substantial 47% for workers with ten or more years of tenure at the firm. Using national survey data from the Australian Bureau of

Statistics, Waddoups (2012) generalizes Bartel's findings and reports that training varies little with tenure among longer-tenured workers.

3 The Baseline Model

This section presents a two-period baseline model, whereby firms can make training investments and earn returns to training in both periods. In particular, we consider a competitive labor market consisting of $B > 2$ identical firms, and a continuum of workers with mass one. All players are rational, risk neutral, and share a common discount factor, $\delta \in (0, 1]$. Each worker lives for two periods. A firm decides in each period how much to invest in general-skills training for each worker the firm employs. In terms of productivity, investments are cumulative. That is, given $g_t \geq 0$ as the skill investment in period $t \in \{1, 2\}$, a worker produces $r(g_1)$ in period 1 and $r(g_1 + g_2)$ in period 2.[3,4]

We assume an exogenous turnover process. In particular, between periods 1 and 2 there is a probability $s \in (0, 1)$ that a worker receives an adverse productivity shock associated with working at the period 1 employer which forces the worker to move to a new firm in period 2.[5] A worker who switches firms at the beginning of period 2 earns $v(g_1 + g_2)$, where in this case g_2 is the training investment chosen by the new employer. Since human capital is fully general, when the worker does move at the beginning of period 2, the new employer chooses the same second period training investment as the initial employer chooses in period 2 when the worker does not move. We assume that $v(\cdot)$ is such that in period 2 employers have an incentive to retain workers for whom there was not an adverse productivity shock.[6] This means only workers who experience adverse productivity shocks move to a new employer.

We consider short-term or one-period contracts, where we investigate both no-commitment and commitment cases. In the no-commitment case, training investments are non-contractible and/or non-enforceable, which means that a contract offer specifies a wage but not a training

[3]In our model, in order to simplify the exposition, we assume that the total investment is *linear* in the aggregate accumulation of human capital, i.e., $c(g_1) = g_1$ and $c(g_1 + g_2) = g_1 + g_2$. Alternatively, to reflect that training a more skilled worker is more costly, one can assume a strictly *convex* total investment, i.e., $c(g_1)$ and $c(g_1 + g_2)$ with $c'(\cdot) > 0$ and $c''(\cdot) > 0$, but our results should remain qualitatively unchanged. However, if one assumes a strictly convex investment in each period, i.e., $c(g_1)$ in period 1 and $c(g_2)$ in period 2 with $c'(\cdot) > 0$ and $c''(\cdot) > 0$, deferred training can arise to smooth training costs across time periods.

[4]Following Acemoglu and Pischke (1999b) and much of the traditional human capital literature, we assume that human capital is one dimensional rather than multi-dimensional as suggested by papers such as Gibbons and Waldman (2004) and Lazear (2009). That is, we do not distinguish between training of different types, e.g., new employee training versus managerial training. Instead, our focus is the timing of training in a setting in which the human capital is one dimensional. This approach, although somewhat restrictive, is tractable and as we show can be used to identify a number of factors that can result in deferred training.

[5]In Section 5.3, we endogenize this job separation rate, which is negatively related to training investments.

[6]Sufficient conditions that guarantee this are $r(g) \geq v(g)$ for all $g \geq 0$, and a worker does not switch employers when the worker is indifferent between moving and staying.

investment. In the commitment case, in contrast, a contract specifies both a wage and a training investment for that period.[7,8]

The sequence of moves for the no-commitment case is as follows. In period 1, firms make wage offers at the beginning of the period and workers choose which offers to accept. Each firm that hires a worker then decides on the worker's training level, and then production takes place. In period 2, firms again offer wages and workers again decide which offers to accept, where workers who receive an adverse productivity shock between periods 1 and 2 move to new employers. Firms then decide on second period training levels, which is followed by second period production. As for the commitment case, the only change in the sequence of moves is that contract offers in each of periods 1 and 2 include a training level for that period.

Our focus is Subgame Perfect Nash equilibrium. Throughout our analysis, we assume $r(\cdot)$ and $v(\cdot)$ are non-negative, strictly increasing, strictly concave, twice continuously differentiable, and satisfy the Inada conditions. One can think of our baseline model as a generalization of the two-period analysis in Acemoglu and Pischke (1999b), whereby firms invest in training in period 1 and earn a return to training in period 2, i.e., the return to training investments is delayed by one period. In our model, we assume instead an immediate return to training investments, and firms can invest and capture a return in both periods, so we can analyze the possibility of deferred training which Acemoglu and Pischke did not consider.

4 Analysis of the Baseline Model

In this section, we show the baseline model generalizes the classical analyses in Becker (1962, 1964) and Acemoglu and Pischke (1999b), and yields a basic principle concerning the dynamic allocation of human-capital investments.

4.1 The No-Commitment Case

Denote $g_{t,N}$ as the equilibrium choice of training investment in period $t \in \{1, 2\}$. When a firm cannot commit to a training level in contract offers, the wage payment just matches outside wages. In this case, taking the current period's wage as given, a firm will choose training investment levels that maximize the profit. That is, after the occurrence of job separations, the

[7] The analysis would be unchanged if we allowed long-term or dynamic contracts in which—in addition to the training level and wage of period 1—firms can commit in period 1 to training levels and wages for period 2. See the Appendix for details.

[8] We assume the outside wage function $v(\cdot)$ does not vary between the commitment and no commitment cases. Most of our results do not depend on this assumption. See Rendahl (2013) for an analysis in which outside wages are derived endogenously as a function of the contractual setting.

period 2 investment maximizes the profit of period 2,

$$g_{2,N}(g_1) = \arg\max_{g_2} r(g_1 + g_2) - g_2. \tag{1}$$

in which $g_{2,N}(g_1)$ is a function of the period 1 training level; taking the period 2 wage $w_{2,N}(g_1, g_2) = v(g_1 + g_2)$ as given, the period 1 investment maximizes the profit over periods 1 and 2,

$$g_{1,N} = \arg\max_{g_1} r(g_1) - g_1 + \delta(1-s)\left[r\left(g_1 + g_{2,N}(g_1)\right) - g_{2,N}(g_1) - v\left(g_1 + g_{2,N}(g_1)\right)\right]. \tag{2}$$

Nesting classical analyses

When the firm can only invest in period 1 and capture a delayed return in period 2, the above model is similar to the "constrained regime" analysis in Acemoglu and Pischke (1999b). That is, by setting $r(g_{1,N}) \equiv 0$ and $g_{2,N} \equiv 0$, (2) can be rewritten as

$$g_{1,N} = \arg\max_{g_1} \delta(1-s)\left[r(g_1) - v(g_1)\right] - g_1,$$

and the corresponding first order condition is given by

$$\delta(1-s)\left[r'(g_{1,N}) - v'(g_{1,N})\right] = 1. \tag{3}$$

In Becker's seminal analysis, assuming human capital is general means that outside wages equal productivity at the current employer, i.e., $r(g) = v(g)$ for all $g \geq 0$. In this case, (3) yields $g_{1,N} = 0$. In other words, in the no-commitment case, assuming the classic approach to general human capital due to Becker yields no firm sponsored training.

The main contribution of Acemoglu and Pischke (1999b) is to show that the nature of outside wages plays a crucial role in determining when firms invest in general training. For example, in the no-commitment case, (3) tells us that firms will sponsor investments in general training in period 1 if outside wages are compressed relative to productivity, i.e., $0 < v'(g) < r'(g)$ for all $g > 0$, and the degree of compression is substantial. To be precise, a sufficient condition for firms to sponsor investments in general training is $\lim_{g \searrow 0} \delta(1-s)[r'(g) - v'(g)] > 1$. We refer to this condition as an adequate level of wage compression.

No deferred training

We now consider what happens when the firm can invest and capture an immediate return to training investments in periods 1 and 2.

Our first result is that the main insight of Acemoglu and Pischke (1999b) holds in our model. That is, compressed outside wages can result in a positive aggregate investment in training, i.e.,

$g_{1,N} + g_{2,N} > 0$. For example, consider the special case in which $v(g) = kr(g)$ for all $g \geq 0$, where $k \in (0, 1]$ captures the degree of wage compression. If outside wages are compressed, i.e., $k \in (0, 1)$, then the aggregate training investment is positive even if k is very close to (but below) one. This follows given $r(\cdot)$ satisfies the Inada conditions. On the other hand, if outside wages are not compressed, i.e., $k = 1$, then the aggregate investment in training equals zero.

Given that compressed outside wages lead to a positive level for the aggregate training investment, the question is, how is the aggregate investment allocated across the two periods? To understand the answer to this question, consider (1) above. If the second period investment is positive, i.e., $g_{2,N}(g_{1,N}) > 0$, then the first order condition for $g_{2,N}(g_{1,N})$ is given by $r'(g_{1,N} + g_{2,N}(g_{1,N})) = 1$. In other words, the aggregate investment level is independent of the period 1 investment level, which indicates that a one unit increase in the period 1 investment decreases the period 2 investment by a unit, i.e., $\frac{\partial g_{1,N}+g_{2,N}(g_{1,N})}{\partial g_{1,N}} = 0 \Rightarrow g'_{2,N}(g_{1,N}) = -1$.

The above reasoning leads to our second result. That is, in the absence of commitment, when the aggregate investment is positive, all of the investment is made in the first period while none of the investment is deferred, i.e., $g_{2,N} = 0$. This result is driven by the basic logic discussed in the Introduction. The return to investing in general training is higher at the beginning of careers because there are more periods in which to reap the returns. As a result, investments in general training occur at the beginning of careers and there is no deferred training.

Proposition 1. *Given an adequate level of wage compression and no commitment to training, $g_{1,N} > 0$ and $g_{2,N} = 0$ where $r'(g_{1,N}) + \delta(1-s)[r'(g_{1,N}) - v'(g_{1,N})] = 1$.*

4.2 The Commitment Case

We now consider the case in which firms can commit to a training level in the contracts offered to workers. In this case, the wage payment becomes an internal transfer between the worker and the firm which does not affect training incentives, so a firm will choose training investment levels that maximize the joint surplus between the contracting parties. Let $g_{t,C}$ denote the equilibrium choice of training level in period $t \in \{1, 2\}$. After the occurrence of job separations, the period 2 investment maximizes the joint surplus of period 2,

$$g_{2,C}(g_1) = \arg\max_{g_2} r(g_1 + g_2) - g_2, \tag{4}$$

in which $g_{2,C}(g_1)$ is a function of the period 1 training level; taking the period 2 wage $w_{2,C}(g_1, g_2) = v(g_1 + g_2)$ as given, the period 1 investment maximizes the joint surplus over periods 1 and 2,

$$g_{1,C} = \arg\max_{g_1} r(g_1) - g_1 + \delta \left\{(1-s)\left[r\left(g_1 + g_{2,C}(g_1)\right) - g_{2,C}(g_1)\right] + sv\left(g_1 + g_{2,C}(g_1)\right)\right\}. \tag{5}$$

Nesting classical analyses

Assuming no opportunity for investment in period 2 and delayed returns yield an analysis similar to the "full-competition regime" analysis in Acemoglu and Pischke (1999b). That is, by setting $r(g_{1,C}) \equiv 0$ and $g_{2,C} \equiv 0$, (5) can be rewritten as

$$g_{1,C} = \arg\max_{g_1} \delta \left[(1-s)r(g_1) + sv(g_1)\right] - g_1,$$

and the corresponding first order condition is given by $\delta \left[(1-s)r'(g_{1,C}) + sv'(g_{1,C})\right] = 1$.

As in our case of commitment, in Becker's world a worker and a firm can sign an enforceable contract that specifies a wage and a training investment level. Then, the period 1 wage is given by

$$w_{1,C} = \delta(1-s)\left[r(g_{1,C}) - v(g_{1,C})\right] - g_{1,C},$$

which yields zero expected firm profit. Like in the no-commitment case, imposing $r(g) = v(g)$ for all $g \geq 0$ yields the period 1 wage $w_{1,C} = -g_{1,C}$. This exactly captures the classic Beckerian story: i) without credit constraints (meaning $w_{1,C}$ can be negative), general training is provided at the efficient level with firms bearing none of the training costs and workers financing training by taking a wage cut in the first period of employment; and ii) with credit constraints (meaning $w_{1,C}$ cannot be negative), general training is below the efficient level.

No deferred training

We now consider what happens when the firm can invest and capture an immediate return in periods 1 and 2. As in the no-commitment case, if $g_{2,C}(g_{1,C}) > 0$, then the first order condition for $g_{2,C}(g_{1,C})$, as defined in (4) above, is given by $r'\left(g_{1,C} + g_{2,C}(g_{1,C})\right) = 1$. So, the aggregate investment level is independent of the period 1 investment level, which suggests that a one unit increase in the period 1 investment decreases the period 2 investment by a unit, i.e., $\frac{\partial g_{1,C}+g_{2,C}(g_{1,C})}{\partial g_{1,C}} = 0 \Rightarrow g'_{2,C}(g_{1,C}) = -1$. Thus, all of the investment is made in period one, where the logic is basically the same as in the no-commitment case.

Proposition 2. *Given commitment to training,* $g_{1,C} > 0$ *and* $g_{2,C} = 0$ *where* $r'(g_{1,C}) + \delta[(1-s)r'(g_{1,C}) + sv'(g_{1,C})] = 1$.

The basic logic for (no) deferred training. The main message of the baseline model is that, independent of whether firms can commit to training, no training is deferred in equilibrium. And the rationale is basically the same across the two cases. That is, starting from an investment profile in which there is deferred training, increasing training in period 1 by a dollar decreases the period 2 investment by a dollar. In turn, since investing at the beginning of period 1 has a higher return because there are more periods following the investment, a firm has an incentive to increase the period 1 investment level until the period 2 investment level equals zero. In the next section, we explore changes to the baseline model that cause this basic logic to break down, and as a result can explain the existence of deferred training.

5 Why Deferred Training?

As discussed in Section 2, there is substantial evidence indicating that deferred training is a common feature of real-world labor markets. However, as was shown in the previous section, our baseline model which extends the well-known two-period Acemoglu and Pischke (1999b) model so that investing in both periods is possible does not exhibit deferred training. In this section, we consider three enrichments of the baseline model that can result in deferred training in equilibrium: i) increasing returns to training with worker age; ii) human capital deterioration or obsolescence; and iii) poaching deterrence. In each case, in addition to formally showing that the enrichment can lead to deferred training, we discuss the relevance of the enrichment for real-world labor markets.

The goal here is to highlight major ideas rather than fully analyze each mechanism in detail, so we keep our exposition as simple as possible. For length reasons, we focus on the no-commitment case below, while similar results could be obtained for the commitment case.

5.1 Increasing Returns to Training with Worker Age

In this subsection, we investigate how equilibrium behavior changes when the returns to training rise in later periods. As we discuss in more detail at the end of this subsection, there are a number of factors that can naturally lead to increasing returns to training with age.

In the baseline model, production as a function of the aggregate training investment is the same in the two periods. In this subsection, we allow the production function in period 2 to be different from the production function in period 1. In particular, our focus is the special case in which the production function is $r(\cdot)$ in period 1 and $\alpha r(\cdot)$ in period 2, with $\alpha > 0$.

For the no-commitment case, if the second period investment is positive, i.e., $g_2(g_1) > 0$, the corresponding first order condition for $g_2(g_1)$ is given by $\alpha r'(g_1 + g_2(g_1)) = 1$. As was true in the baseline analysis, a one unit increase in the period 1 investment decreases the period 2 investment by a unit, i.e., $g_2'(g_1) = -1$. However, the basic logic from the baseline analysis can break down here when a dollar invested in period 2 has a larger marginal effect on future returns than a dollar invested in period 1. If that marginal effect rises significantly with worker age, i.e., α is sufficiently larger than 1, then the return to sponsoring general training in period 2 can be higher than in period 1 even though there are fewer subsequent periods, which means that deferred training can arise in equilibrium.

Proposition 3. *Given returns to training that can vary with worker age, $g_1 > 0$, and there exists a threshold value $\hat{\alpha} > 1$ such that $g_2 > 0$ if and only if $\alpha > \hat{\alpha}$.*

Now we discuss various reasons that the return to training can rise as workers age.[9]

5.1.1 Learning by doing

We first consider learning by doing as a class of explanations for deferred training. That is, even without formal training, workers can gain skills via on-the-job learning which improves productivity. By deferring training towards more experienced (senior) workers, the rate of return effectively increases over time. In particular, firms will have incentives to defer training investments in more experienced employees when the effect of learning by doing on improving productivity or $\alpha > 1$ in the extended model described above is sufficiently large.[10]

5.1.2 Job changes

We then consider a standard job ladder model à la Gibbons and Waldman (1999, 2006).[11] In that model, a firm has multiple jobs a worker can be assigned to, where workers accumulate general human capital as they age and the productivity return to human capital is higher at jobs higher on the job ladder. In equilibrium, as workers age, they move up the ladder into jobs which value human capital more highly. The analysis in the current subsection captures what would happen in a Gibbons and Waldman type model if firms can sponsor general human capital and there is wage compression. The wage compression leads to positive investments as initially pointed out by Acemoglu and Pischke. And then, because as workers get older they move up the ladder into jobs where human capital has a bigger incremental impact on productivity, the result can be deferred investments rather than all the investments occurring early in worker careers.

Example 1. Suppose that a worker's productivity is $r(g)$ on job 1 and $\alpha r(g)$ on job 2, where $\alpha > 1$ captures the higher return to human capital investments on job 2 than on job 1. When α is sufficiently large, the worker assigned to job 1 in period 1 will receive deferred training if she gets promoted in period 2.

[9] We label this enrichment as "increasing returns to training with worker age," but our formal analysis is equally consistent with "increasing returns to training with labor market experience," since in our model age and labor market experience are perfectly correlated.

[10] An equivalent argument is that learning by doing reduces investment costs. Consequently, training costs decline with worker age, so firms can defer training to save investment costs.

[11] Earlier related analyses can be found, for example, in Lucas (1978) and Rosen (1982).

5.1.3 Employer learning

Another setting along the lines of this subsection's argument is a variant of an employer learning model with heterogeneous worker abilities.[12] Suppose workers vary in their ability levels, ability and general human capital are complementary inputs in the production function, and each worker's ability is gradually revealed to firms after labor market entry, possibly through publicly observed output realizations. In this setting, there can be deferred training, although when it arises it will typically be received only by the workers revealed to be high ability. The basic idea is that, because of the complementarity, the amount of training is positively related to worker ability. So, when a worker is revealed to be high ability, the return to additional training is high for that worker which can result in an equilibrium where everyone receives training in the first period, but only those revealed to be high ability receive deferred training. This result is consistent with various real-world settings, such as large law firms, where workers revealed to be high ability such as those promoted to partner positions receive further training after their high abilities are revealed.

Example 2. Suppose a worker with ability $\theta \in \{l, h\}$ and training $g \geq 0$ produces output $r(g;\theta) = \theta r(g)$. To ease the exposition, we normalize abilities to $0 = l < h = 1$, so an l type should never receive any training from an efficiency perspective. If information is not belated, i.e., employers can immediately observe employees' abilities at the beginning of period 1, h types receive training only. In particular, they receive training in period 1 but not in period 2 since the rate of return to training equals $r'(\cdot)$ in periods 1 and 2. If information is, instead, belated, i.e., employees' abilities are unknown in period 1 but revealed at the beginning of period 2, then the rate of return to training is $qr'(\cdot)$ in period 1 and $r'(\cdot)$ in period 2, where $q = \Pr(\theta = h)$ denotes the prior probability that a worker is h type. In this case, both types will receive the same amount of training in period 1 because employers cannot identify h types, and employers will select h types only for training in period 2 when $\alpha = 1/q$ is sufficiently larger than 1.

In the setting just described, one can think of deferred training as a cost-saving strategy. If firms know which workers are h types, then it would be optimal to not defer any training, but rather vary the initial training level across workers such that h types receive more training than l types in period 1. However, without knowing who are h types when workers are first trained, such variation is not possible. In particular, training in period 1 is tailored to the majority of the workforce. That is, when q is small or the proportion of l types is high, training in period 1 is at a positive but low level, so delaying training in this case serves to reduce costs because training in period 2 is only provided to h types for whom training is efficient.[13]

[12] There is a large literature on employer learning that starts with the seminal analyses of Jovanovic (1979), Fama (1980), and Holmström (1982). Note that this argument is related to the argument in Carter (2021), mentioned in Section 2, where training is delayed until firms learn whether the worker-firm match is high quality.

[13] One can combine the above two ideas, i.e., there are multiple jobs, workers vary in terms of where their

5.2 Human Capital Deterioration or Obsolescence

In our second enrichment, we build on the insights of Ben-Porath (1967) concerning the role of human capital deterioration or obsolescence in the timing of human capital investments. Ben-Porath's focus was a setting in which workers choose how much to invest in human capital at different ages, where the human capital created by an investment in period t gradually deteriorates or becomes obsolete in subsequent periods. The result is that investment levels are higher at earlier ages, but there are positive investments at later ages. Here we show that Ben-Porath's basic insight also applies to firm sponsored investments in general training.

In the baseline model, in terms of the stock of human capital, a dollar spent on training in period 1 has the same effect on period 2's stock as on period 1's stock. But because of human capital deterioration or obsolescence, it might be more realistic to assume that a dollar spent on training in period 1 has a larger impact on period 1's stock than on period 2's stock. To be precise, in the baseline model the stock of human capital in period 2 is $g_1 + g_2$; incorporating human capital deterioration or obsolescence into our model means that the stock of human capital in period 2 equals $\lambda g_1 + g_2$, with $\lambda \in (0, 1)$.

For the no-commitment case, if the second period investment is positive, i.e., $g_2(g_1) > 0$, the corresponding first order condition for $g_2(g_1)$ is given by $r'(\lambda g_1 + g_2(g_1)) = 1$. Like in the baseline analysis, the aggregate investment is independent of the period 1 investment, which indicates that a one unit increase in the period 1 investment decreases the period 2 investment by λ units, i.e., $\frac{\partial \lambda g_1 + g_2(g_1)}{\partial g_1} = 0 \Rightarrow g_2'(g_1) = -\lambda$. This stands in contrast to the baseline analysis where a one unit increase in the period 1 investment decreases the period 2 investment by a unit. In other words, the basic logic from the baseline analysis breaks down here since the tradeoff between the investments in periods 1 and 2 is no longer one for one. So, in contrast to what was true in the baseline analysis, it is not always optimal to shift investment dollars from period 2 to period 1. In particular, if λ is sufficiently small, it is now possible that the firm will choose a positive training investment in both periods. The reason is that, even if firms sponsor high levels of training early in careers, in the absence of deferred training the stock of human capital later in careers would be limited. So deferred training would be the efficient way to ensure that older workers are endowed with an adequate stock of human capital.

Proposition 4. *Given human capital deterioration/obsolescence, $g_1 > 0$, and there exists a*

comparative advantage lies, and there is initial uncertainty concerning a worker's productivity type. A model along these lines was considered in MacDonald (1982). The basic idea is that when a worker enters the labor market there is uncertainty concerning which job the worker would be most efficiently assigned to, and as the worker gains labor market experience firms learn the worker's efficient job assignment. If human capital is more valuable when a worker is assigned optimally, then the basic logic of this subsection's analysis applies. There will be a positive investment in sponsored training given wage compression. Further, some training will be deferred if the initial level of uncertainty is sufficiently high, because the returns to human capital will be higher in later periods given that job assignments will be more efficient, on average, when workers are older.

threshold value $\hat{\lambda} \in (0, 1)$ *such that* $g_2 > 0$ *if and only if* $\lambda < \hat{\lambda}$.

The question is, how realistic is it to assume that firm sponsored human capital accumulation is characterized by substantial rates of deterioration or obsolescence? We believe that it is quite realistic. In particular, there are a number of factors that can naturally lead to substantial human capital deterioration. For example, memory constraints can lead to skills acquired through training early in a career to fade as the memory of the training fades.[14]

Another reason that human capital deterioration or obsolescence can arise is job movements, either lateral moves or promotions, where different jobs employ different skills. This idea is related to Gibbons and Waldman (2004, 2006) notion of task specific human capital, and also Lazear (2009, 2012) arguments concerning the skill-weights approach and the human capital needed for corporate leadership positions. The basic logic is that when a worker moves to a new job that employs different skills, the human capital the worker previously accumulated becomes less valuable, and there is a high return to investing in human capital that builds skills important for the new job. Gathmann and Schönberg (2010), Frederiksen and Kato (2018) and Jin and Waldman (2019) present evidence consistent with this view of the human capital accumulation process.

Examples above demonstrate that job changes such as job transfers and promotions can cause deferred training. In many contexts, however, actual job changes are not essential; deferred training can arise with virtual job changes due to technological changes in which workers do not experience actual job changes. Specifically, human capital obsolescence is clearly more important in industries experiencing rapid technological changes which affect the nature of work. For example, deferred training is more common in industries in which automation has been growing, such as for many manufacturing jobs, rather than in industries such as teaching where automation has played a less significant role. In other words, skills that significantly improve productivity given the current technology can become much less useful when technology advances and different skills are needed to stay highly productive. Note that the recent increase in remote teaching may lead to an increase in deferred training among teachers going forward.

5.3 Poaching Deterrence

One can view mechanisms identified in the previous two subsections as *natural* reasons for delaying investments in training that arise as part of common human-resource management practices or due to an efficiency rationale. In this subsection, we show that training can be *strategically* deferred in order to limit the size of employee poaching from outside firms. In other

[14] See Mullainathan (2002) and Wilson (2014) for analyses focused on limited memory, and Chari and Hopenhayn (1991) for an analysis related to human capital obsolescence.

words, factors identified in this subsection distort the optimal allocation of training investments, whereby some investments are deferred to period 2.

5.3.1 Signal jamming

When workers differ in ability, the preceding subsection establishes that belated information can engender deferred training via an increasing rate of return over time. This subsection studies another type of information friction: asymmetric employer learning (Waldman, 1984), whereby incumbent employers obtain more accurate information about worker abilities than outsiders. In this case, training in period 1 sponsored by incumbent employers signals worker ability to outsiders whenever training investments increase with worker ability. To compress outside wages and, in turn, capture informational rents in period 2, we show that incumbent employers can hide information from outsiders by making training in period 1 invariant to worker abilities. The ensuing investment level is relatively low in period 1, indicating that deferred training is likely to occur in period 2.

Suppose the productivity of a worker of type $\theta \in \{l, h\}$ is $r(g; \theta) = \theta r(g)$, with $0 = l < h = 1$, which suggests that it is efficient for h types to receive more training than l types. Employer learning is asymmetric across employers: immediately after hiring workers at the beginning of period 1, incumbent employers can observe worker abilities, whereas outsiders cannot. Outsiders, however, know that a worker is h type with probability $q \in (0, 1)$. Then, training in period 1 serves as a signal of worker abilities to outsiders whenever incumbent employers invest more in h types than in l types. Focusing on pure strategy Perfect Bayesian equilibrium, there are two possible forms equilibrium can take: i) a separating equilibrium (SE) where in period 1 incumbent employers invest more in h types than in l types; and ii) a pooling equilibrium (PE) where in period 1 incumbent employers make the same amount of investments in both types.

In an SE, outsiders can perfectly distinguish worker abilities based on training in period 1, and as a result outside wages are not compressed. In contrast, in a PE, outsiders cannot distinguish worker abilities so outside wages are compressed. A full analysis of this setting is beyond the scope of the paper. Instead our focus is the equilibrium which maximizes expected worker utility.[15] Below let g_h^* be the first best efficient first period training investment for type h workers.

Proposition 5. *If $r(0) > [1 - \delta(1 - s)(1 - q)] (r(g_h^*) - g_h^*)$, there is deferred training, i.e., neither type receives training in period 1 and only h types receive training in period 2; otherwise, there is no deferred training, i.e., only h types receive training in period 1.*

From an efficiency perspective, incumbent employers should make more training investments

[15] To be precise, expected worker utility is q times the discounted expected lifetime utility of being an h ability worker plus $1 - q$ times the discounted expected lifetime utility of being an l ability worker.

in h types than in l types in period 1. However, given asymmetric learning, training in period 1 signals worker abilities to outsiders and drives up outside wages to h types. In the setting described above, this means non-compressed outside wages. In equilibrium, to limit the size of employee poaching from outside firms, incumbent employers can adopt a signal-jamming strategy by making training in period 1 invariant to worker abilities. This results in compressed outside wages and, in turn, informational rents in period 2, which is associated with a higher period 2 firm profit.[16] The proposition shows that, if $r(0) > [1 - \delta(1-s)(1-q)]\,(r(g_h^*) - g_h^*)$, then the incentive for firms to practice this type of poaching deterrence means this behavior is exhibited in equilibrium even if the focus is the equilibrium which maximizes expected worker utility.

5.3.2 Endogenous turnover

In our last enrichment of the baseline model, we consider endogenous worker turnover. That is, in contrast to the baseline model where the separation probability $s \in (0,1)$ at the beginning of the second period is exogenous, we now endogenize worker turnover. In particular, we explore separation processes in which the turnover rate drops with an increased level of training investments. We show that firms can defer training given breakup fees that employees (partially) repay training costs if they quit jobs before a specified time.[17]

So far we maintain the assumption that incumbent employers make training investments in period 2 after the occurrence of worker turnover. In other words, only job stayers can ever receive training in period 2. We now modify the baseline model such that workers can quit after receiving training in period 2. Specifically, we make two changes to the baseline model. First, each worker receives training at the beginning of periods 1 and 2, and draws some random utility η from a zero-mean log-concave distribution $G(\cdot)$ during period 1, where a bad draw can induce a worker to change employers in period 2 (after receiving period 2 training) to find a better draw. Second, incumbent employers can request a worker to repay $R(g_2) \geq 0$ if she quits after receiving period 2 training $g_2 > 0$. So, a worker receives an effective outside wage $v(g_1 + g_2) - R(g_2)$ if she quits in period 2, i.e., the repayment $R(g_2)$ limits the size of employee poaching from outside firms. Furthermore, the repayment schedule $\{R(g_2) | g_2 \geq 0\}$ is specified

[16] Note that this signal-jamming strategy compresses outside wages and gives rise to deferred training in a different manner than in the two-period framework in Acemoglu and Pischke (1998), where firms can observe worker ability at the beginning of period 2 but make training investments in period 1 only. In their analysis, though employer learning is asymmetric, training in period 1 does not signal worker abilities. In turn, outside wages are compressed because outsiders cannot distinguish worker types. In equilibrium, incumbent employers capture informational rents in period 2 and thus have incentives to pay for general training in period 1.

[17] See Mukherjee and Vasconcelos (2018) for a recent analysis of breakup fees and human capital.

before a worker receives any training at the beginning of period 1.[18,19] To ease the exposition, we focus on a linear repayment schedule $R(g_2) = \rho g_2$ with $\rho \in [0, 1]$.

Given no repayment, i.e., $\rho = 0$, a worker with random utility η separates if $v(g_1 + g_2) > w_2 + \eta$. So, the separation rate is $G(v(g_1 + g_2) - w_2)$ and period 2 training solves

$$\begin{aligned} g_2(g_1) &= \arg\max_{w_2, g_2} \{[1 - G(v(g_1+g_2) - w_2)][r(g_1+g_2) - w_2] - g_2\} \\ &= \arg\max_{g_2} \left\{ \underbrace{\max_{w_2} [1 - G(v(g_1+g_2) - w_2)][r(g_1+g_2) - w_2]}_{=\phi(r(g_1+g_2) - v(g_1+g_2))} - g_2 \right\}, \end{aligned}$$

where $\phi(\cdot)$, with $\phi'(\cdot) \in (0, 1)$, denotes the expected return to training. In turn, period 1 training solves

$$g_1 = \arg\max_{g_1} \{r(g_1) - g_1 + \delta[\phi(r(g_1 + g_2(g_1)) - v(g_1 + g_2(g_1))) - g_2(g_1)]\},$$

where the return rate decreases over time since $\phi'(\cdot) < 1$. In equilibrium, as suggested by the baseline analysis, no training is deferred.

Given repayment, i.e., $\rho \in (0, 1]$, a worker with random utility η separates if $v(g_1 + g_2) - R(g_2) > w_2 + \eta$. Then, the separation rate is $G(v(g_1 + g_2) - R(g_2) - w_2)$ and period 2 training solves

$$\begin{aligned} g_2(g_1) &= \arg\max_{w_2, g_2} \left\{ \begin{array}{c} [1 - G(v(g_1+g_2) - R(g_2) - w_2)][r(g_1+g_2) - w_2] \\ +G(v(g_1+g_2) - R(g_2) - w_2) R(g_2) - g_2 \end{array} \right\} \\ &= \arg\max_{g_2} \left\{ \underbrace{\max_{w_2} [1 - G(v(g_1+g_2) - R(g_2) - w_2)][r(g_1+g_2) - R(g_2) - w_2]}_{=\phi(r(g_1+g_2) - v(g_1+g_2))} - \underbrace{[g_2 - R(g_2)]}_{=(1-\rho)g_2} \right\}, \end{aligned}$$

where the investment cost effectively reduces to $(1 - \rho)g_2$ in period 2. In turn, period 1 training solves

$$g_1 = \arg\max_{g_1} \{r(g_1) - g_1 + \delta[\phi(r(g_1 + g_2(g_1)) - v(g_1 + g_2(g_1))) - (1 - \rho)g_2(g_1)]\}.$$

In equilibrium, training is deferred given a sufficiently large repayment schedule $R(g_2) = \rho g_2$.

[18] We assume that firms cannot commit to training (i.e., they can renege on promised training investments), yet they can make the repayment schedule contingent on training investments.

[19] This repay-if-quit practice is consistent with evidence in the real world that many part-time MBA students, who are sponsored by employers, are required to repay tuition fees (at least partially) if they leave before a specified time.

Proposition 6. *When a worker repays a sufficiently large fraction of training investments if she quits after receiving training in period 2, i.e.,* $\rho \in (0, 1]$ *is sufficiently close to* 1*, firms defer training.*

Note here that we abstract away from the optimal setting of the repayment schedule, i.e., $\rho \in (0, 1]$ is taken as given. Furthermore, we assume no credit constraints meaning that ρ can be as close to 1 as possible. If firms can, instead, set the repayment schedule, i.e., $\rho \in (0, 1]$ is a choice variable, a Beckerian type of argument yields that firms set $\rho = 1$, i.e., credit unconstrained workers pay for deferred training.

This result is related to the classical hold-up problem, where relationship-specific investments create payoffs that can be bargained over after investments have been made and anticipation of bargaining can create disincentives for firms to invest in firm-specific training. In other words, relationship-specific investments made by firms can be held hostage, so firms will not make efficient investments.

In our model, training in period 2, however, improves a worker's general skills. Given a bad draw of utility, a worker can separate and capture some returns to deferred training when there is no repayment, so firms will not make training investments in period 2 from an efficiency perspective. When there is repayment, the firm and the worker can, however, co-share investment costs in period 2. As a consequence, firms have incentives to defer training investments, which improves the value of employment relationships in period 2 and, in turn, results in a higher profit, a higher wage, and a lower separation rate.[20,21]

6 Conclusion

Evidence from real-world labor markets clearly shows that firms frequently sponsor investments in deferred training. Standard theoretical models of firm sponsored training, however, typically have all training provided early in workers' careers because doing so maximizes the number of periods the firm can reap the returns from the training investments. In this paper we extend the classic two-period model of Acemoglu and Pischke (1999b) concerning firm sponsored general training to investigate the factors that can lead to deferred investments in general training.

We first study a baseline model which illustrates both Acemoglu and Pischke's main results concerning the importance of wage compression for understanding the provision of firm sponsored general training, and also illustrates the standard theoretical result that all training is

[20]Consistent with our results, Benson, Finegold, and Mohrman (2004) and Manchester (2010, 2012) find evidence that training is positively related to more employee retention.

[21]It is also possible that deferred training enhances job satisfaction, i.e., workers derive a higher value of utility in period 2 than in period 1, which reinforces the effects of deferred training on improving the value of employment relationships.

provided early in workers' careers. We then consider enrichments of the model which identify three main factors that can lead to firm sponsored investments in general training: i) increasing productivity returns to human capital with worker age; ii) human capital deterioration or obsolescence; and iii) poaching deterrence. We also discuss underlying theoretical mechanisms that can result in each factor being important, and related empirical studies.

We think a fruitful avenue for future research is to investigate the testable predictions of our analysis. As we discussed in Section 2, existing empirical evidence indicates that deferred training is important in real-world labor markets. But there is little focus on factors which lead to higher investments in deferred training. Our analysis points to various factors that lead to deferred training, and we believe investigating the possible interactions across the various factors we have identified and testable predictions of our analysis is a worthwhile topic for future research.

A Appendix

Proof of Proposition 1. As $\lim_{g \searrow 0} \delta(1-s)[r'(g) - v'(g)] > 1$, $g_{1,N} > 0$. If $g_{2,N} > 0$, $g'_{2,N}(g_{1,N}) = -1$. Then, the first order condition for $g_{1,N} > 0$, as defined in (2), is $r'(g_{1,N}) + \delta(1-s) = 1$, which contradicts the first order condition for $g_{2,N} > 0$ as defined in (1), i.e., $r'(g_{1,N} + g_{2,N}) = 1$. Given $g_{2,N} = 0$, $g_{1,N} > 0$ solves $r'(g_{1,N}) + \delta(1-s)\left[r'\left(g_{1,N}\right) - v'\left(g_{1,N}\right)\right] = 1$. □

Proof of Proposition 2. As $\lim_{g \searrow 0} r'(g) = +\infty$, $g_{1,C} > 0$. If $g_{2,C} > 0$, $g'_{2,C}(g_{1,C}) = -1$. Then, the first order condition for $g_{1,C} > 0$, as defined in (5), is $r'(g_{1,C}) + \delta(1-s) = 1$, which contradicts the first order condition for $g_{2,C} > 0$ as defined in (4), i.e., $r'(g_{1,C} + g_{2,C}) = 1$. Given $g_{2,C} = 0$, $g_{1,C} > 0$ solves $r'(g_{1,C}) + \delta\left[(1-s)r'\left(g_{1,C}\right) + sv'\left(g_{1,C}\right)\right] = 1$. □

Long-term or dynamic contracts. In period 1, firms can now commit to a wage-training combination over periods 1 and 2, i.e., $\{(w_{1,D}, g_{1,D}); (w_{2,D}, g_{2,D})\}$, which maximizes a worker's lifetime utility

$$\max_{\{(w_{1,D},g_{1,D});(w_{2,D},g_{2,D})\}} w_{1,D} + \delta\left[(1-s)w_{2,D} + sv(g_{1,D} + g_{2,D})\right]$$
$$\text{s.t.} \quad r(g_{1,D}) - w_{1,D} - g_{1,D} + \delta(1-s)\left[r(g_{1,D} + g_{2,D}) - w_{2,D} - g_{2,D}\right] \geq 0,$$

where the constraint means a non-negative firm profit. If the constraint binds, the above problem reduces to

$$\max_{(g_{1,D},g_{2,D})} r(g_{1,D}) - g_{1,D} + \delta\left\{(1-s)\left[r(g_{1,D} + g_{2,D}) - g_{2,D}\right] + sv(g_{1,D} + g_{2,D})\right\},$$

which is a standard constraint optimization problem whose Kuhn-Tucker conditions are

$$[g_{1,D}]: \quad r'(g_{1,D}) + \delta\left[(1-s)r'(g_{1,D} + g_{2,D}) + sv'(g_{1,D} + g_{2,D})\right] = 1 - \eta, \quad \eta g_{1,D} = 0, \quad g_{1,D} \geq 0;$$
$$[g_{2,D}]: \quad \delta(1-s)r'(g_{1,D} + g_{2,D}) = \delta(1-s) - \mu, \quad \mu g_{2,D} = 0, \quad g_{2,D} \geq 0.$$

As $\lim_{g \searrow 0} r'(g) = +\infty$, $g_{1,D} > 0$, so $\eta = 0$. If $g_{2,D} > 0$, $\mu = 0$. In turn, there is a contradiction, i.e., $r'(g_{1,D} + g_{2,D}) = 1 > 1 - \delta(1-s) = r'(g_{1,D}) + \delta sv'(g_{1,D} + g_{2,D})$. Thus, as in the commitment case, no training is deferred, whether or not the outside wage is compressed.

Proof of Proposition 3 (increasing returns). With increasing returns to training, $g_2(g_1)$ maximizes the profit of period 2

$$g_2(g_1) = \arg\max_{g_2} \alpha r(g_1 + g_2) - g_2,$$

while g_1 maximizes the profit over periods 1 and 2,

$$g_1 = \arg\max_{g_1} r(g_1) - g_1 + \delta(1-s)\left[\alpha r\left(g_1 + g_2(g_1)\right) - g_2(g_1) - \alpha v\left(g_1 + g_2(g_1)\right)\right].$$

Invoking the first order conditions, as in the baseline analyses, $g_1 > 0$; if $g_2 > 0$, $g_2'(g_1) = -1$ suggests $r'(g_1) + \delta(1-s) = 1 = \alpha r'\left(g_1 + g_2\right)$, which yields $g_2 > 0$ if and only if $\alpha > \hat{\alpha} = \frac{1}{1-\delta(1-s)}$. □

Details of Example 2 (employer learning). With belated information, $g_2(g_1)$ maximizes the profit of period 2

$$g_2(g_1) = \arg\max_{g_2} r(g_1 + g_2) - g_2,$$

while g_1 maximizes the profit over periods 1 and 2,

$$g_1 = \arg\max_{g_1} qr(g_1) - g_1 + \delta(1-s)q\left[r\left(g_1 + g_2(g_1)\right) - g_2(g_1) - v\left(g_1 + g_2(g_1)\right)\right].$$

Invoking the first order conditions, $q\left[r'(g_1) + \delta(1-s)\right] = 1 = r'\left(g_1 + g_2\right)$ suggests $g_2 > 0$ if and only if $q < \hat{q} = \frac{1}{1+\delta(1-s)}$.

Proof of Proposition 4 (human capital deterioration). With human capital deterioration, $g_2(g_1)$ maximizes the profit of period 2

$$g_2(g_1) = \arg\max_{g_2} r(\lambda g_1 + g_2) - g_2,$$

while g_1 maximizes the profit over periods 1 and 2,

$$g_1 = \arg\max_{g_1} r(g_1) - g_1 + \delta(1-s)\left[r\left(\lambda g_1 + g_2(g_1)\right) - g_2(g_1) - v\left(\lambda g_1 + g_2(g_1)\right)\right].$$

Invoking the first order conditions, as in the baseline analyses, $g_1 > 0$; if $g_2 > 0$, $g_2'(g_1) = -\lambda$ suggests $r'(g_1) + \delta(1-s)\lambda = 1 = r'\left(\lambda g_1 + g_2\right)$, where g_1 increases in λ, g_2 decreases in λ, and $\lambda g_1 + g_2$ is invariant to λ. Thus, there exists a unique threshold $\hat{\lambda} \in (0,1)$ for which $g_2 > 0$ if and only if $\lambda < \hat{\lambda}$. □

Proof of Proposition 5 (signal jamming). In an SE, outsiders can perfectly distinguish worker abilities based on training in period 1, whereby incumbent firms invest $g_1^{SE} > 0$ in h types only. In this case, outside wages are non-compressed in period 2, i.e., $\max_{g_2} r\left(g_1^{SE} + g_2\right) - g_2$ to h types and zero to l types. Consequently, the period 2 profit equals zero and, in turn, training in period 1 is $g_1^{SE} = \arg\max_{g_1} r(g_1) - g_1$, where $g_1^{SE} = g_h^* > 0$ solves $r'(g_h^*) = 1$. As the rate of return to training is $r'(\cdot)$ in both periods, incumbent employers will not defer any training, i.e., $g_2^{SE} = 0$. In this case, a worker's expected lifetime utility is given by

$$\underbrace{q\left(r(g_h^*) - g_h^*\right)}_{\text{period 1 wage}} + \delta q \underbrace{\left(r(g_h^*) - g_h^*\right)}_{h\text{'s period 2 wage}} = q(1+\delta)\left(r(g_h^*) - g_h^*\right).$$

In a PE, outsiders cannot distinguish worker abilities. As incumbent employers hiring an l type invests zero in period 1, incumbent employers hiring an h type can pretend to hire an l type by also investing zero in period 1. In this case, training in period 1 is $g_1^{PE} = 0$ for both types. Consequently, outside wages are compressed in period 2, i.e., $\max_{g_2} q\left(r(g_2) - g_2\right)$ to both types, where training $g_2^{SE} = g_h^*$ solves $r'(g_h^*) = 1$. In turn, incumbent employers can capture informational rents $(1-q)\left(r(g_h^*) - g_h^*\right)$ in period 2. In this case, a worker's expected lifetime utility is given by

$$\underbrace{q\left[r(0) + \delta(1-s)(1-q)\left(r(g_h^*) - g_h^*\right)\right]}_{\text{period 1 wage}} + \delta \underbrace{q\left(r(g_h^*) - g_h^*\right)}_{\text{period 2 wage}} = q\left\{r(0) + \delta\left[1 + (1-s)(1-q)\right]\left(r(g_h^*) - g_h^*\right)\right\}.$$

Comparing a worker's expected lifetime utility given an SE and given a PE, we find that firms defer training whenever

$$\begin{aligned} &q\left\{r(0) + \delta\left[1 + (1-s)(1-q)\right]\left(r(g_h^*) - g_h^*\right)\right\} - q(1+\delta)\left(r(g_h^*) - g_h^*\right) \\ &\qquad = q\left\{r(0) - \left[1 - \delta(1-s)(1-q)\right]\left(r(g_h^*) - g_h^*\right)\right\} > 0, \end{aligned}$$

which is true if $r(0) > \left[1 - \delta(1-s)(1-q)\right]\left(r(g_h^*) - g_h^*\right)$. □

Proof of Proposition 6 (endogenous turnover). If $g_2 > 0$, invoking the first order conditions for the no-repayment case yields $\phi'\left(r\left(g_1+g_2\right) - v\left(g_1+g_2\right)\right)\left[r'\left(g_1+g_2\right) - v'\left(g_1+g_2\right)\right] = 1$ which suggests $g_2'(g_1) = -1$. In turn, $r'(g_1) + \delta = 1$, which contradicts

$$r'\left(g_1+g_2\right) = \frac{1}{\phi'\left(r\left(g_1+g_2\right) - v\left(g_1+g_2\right)\right)} + v'\left(g_1+g_2\right) > 1,$$

where $v'\left(g_1+g_2\right) > 0$ and $\phi'(\cdot) \in (0,1)$. Likewise, if $g_2 > 0$, invoking the first order conditions for the repayment case yields $\phi'\left(r\left(g_1+g_2\right) - v\left(g_1+g_2\right)\right)\left[r'\left(g_1+g_2\right) - v'\left(g_1+g_2\right)\right] = 1-\rho$ which again suggests $g_2'(g_1) = -1$. In turn, $r'(g_1) + \delta(1-\rho) = 1$, while

$$r'\left(g_1+g_2\right) = \frac{1-\rho}{\phi'\left(r\left(g_1+g_2\right) - v\left(g_1+g_2\right)\right)} + v'\left(g_1+g_2\right) < 1 - \delta(1-\rho)$$

if ρ is sufficiently close to 1. □

References

Acemoglu, Daron. 1997. "Training and innovation in an imperfect labour market." *Review of Economic Studies* 64 (3):445–464.

Acemoglu, Daron and Jörn-Steffen Pischke. 1998. "Why do firms train: theory and evidence." *Quarterly Journal of Economics* 113:79–119.

———. 1999a. "Beyond Becker: Training in imperfect labour markets." *Economic Journal* 109 (453):112–142.

———. 1999b. "The structure of wages and investment in general training." *Journal of Political Economy* 107:539–572.

———. 2003. "Minimum wages and on-the-job training." *Research in Labor Economics* 22 (2003):159–202.

Autor, David. 2001. "Why do temporary help firms provide free general skills training?" *Quarterly Journal of Economics* 116:1409–1448.

Balmaceda, Felipe. 2005. "Firm-sponsored general training." *Journal of Labor Economics* 23 (1):115–133.

Bar-Isaac, Heski and Clare Leaver. 2022. "Training, recruitment, and outplacement as endogenous asymmetric information." *Economica* 89:849–861.

Barron, John, Dan Black, and Mark Loewenstein. 1987. "Employer size: The implications for search, training, capital investment, starting wages, and wage growth." *Journal of Labor Economics* 5 (1):76–89.

———. 1989. "Job matching and on-the-job training." *Journal of Labor Economics* 7 (1):1–19.

Bartel, Ann. 1995. "Wage growth, and job performance: evidence from a company database." *Journal of Labor Economics* 13:401–425.

Becker, Gary. 1962. "Investment in human capital: a theoretical analysis." *Journal of Political Economy* 70:9–49.

———. 1964. *Human Capital*. New York: Columbia University Press.

Ben-Porath, Yoram. 1967. "The production of human capital and the life cycle of earnings." *Journal of Political Economy* 75 (4, Part 1):352–365.

Benson, George S, David Finegold, and Susan Albers Mohrman. 2004. "You paid for the skills, now keep them: Tuition reimbursement and voluntary turnover." *Academy of Management Journal* 47 (3):315–331.

Brown, James. 1989. "Why do wages increase with tenure? On-the-job training and life-cycle wage growth observed within firms." *American Economic Review* 79:971–991.

Cappelli, Peter. 2004. "Why do employers pay for college?" *Journal of Econometrics* 121 (1-2):213–241.

Carter, Thomas. 2021. "Delayed firm-paid general training." *Southern Economic Journal* 87:937–951.

Chang, Chun and Yijiang Wang. 1996. "Human capital investment under asymmetric information: the Pigovian conjecture revisited." *Journal of Labor Economics* 14 (3):505–519.

Chari, V. V. and Hugo Hopenhayn. 1991. "Vintage human capital, growth, and the diffusion of new technology." *Journal of Political Economy* 99:1142–1165.

Fama, Eugene. 1980. "Agency problems and the theory of the firm." *Journal of Political Economy* 88:288–307.

Frederiksen, Anders and Takao Kato. 2018. "Human capital and career success: Evidence from linked employer-employee data." *Economic Journal* 128:1952–1982.

Garicano, Luis and Luis Rayo. 2017. "Relational knowledge transfers." *American Economic Review* 107 (9):2695–2730.

Gathmann, Christina and Uta Schönberg. 2010. "How general is human capital? A task-based approach." *Journal of Labor Economics* 28:1–49.

Gibbons, Robert and Michael Waldman. 1999. “A theory of wage and promotion dynamics inside firms.” *Quarterly Journal of Economics* 114:1321–1358.

———. 2004. “Task-specific human capital.” *American Economic Review* 94 (2):203–207.

———. 2006. “Enriching a theory of wage and promotion dynamics inside firms.” *Journal of Labor Economic* 24:59–107.

Hashimoto, Masanori. 1981. “Firm-specific human capital as a shared investment.” *American Economic Review* 71 (3):475–482.

Holmström, Bengt. 1982. “Managerial incentive problems: A dynamic perspective.” In *Essays in Economics and Management in Honor of Lars Wahlbeck*. Helsinki, Finland: Swenska Handelshogkolan; reprinted 1999 in *Review of Economic Studies* 66:169–82.

Jin, Xin and Michael Waldman. 2019. “Lateral moves, promotions, and task-specific human capital: Theory and evidence.” *Journal of Law, Economics, and Organization* 36:1–46.

Jovanovic, Boyan. 1979. “Job matching and the theory of turnover.” *Journal of Political Economy* 87:972–990.

Kahn, Charles and Gur Huberman. 1988. “Two-sided uncertainty and "up-or-out" contracts.” *Journal of Labor Economics* 6 (4):423–444.

Katz, Eliakim and Adrian Ziderman. 1990. “Investment in general training: the role of information and labour mobility.” *Economic Journal* 100:1147–1158.

Kessler, Anke S and Christoph Lülfesmann. 2006. “The theory of human capital revisited: on the interaction of general and specific investments.” *Economic Journal* 116 (514):903–923.

Lazear, Edward. 2009. “Firm-specific human capital: A skill-weights approach.” *Journal of Political Economy* 117 (5):914–940.

———. 2012. “Leadership: A personnel economics approach.” *Labour Economics* 19:92–101.

Lazear, Edward and Paul Oyer. 2013. *Personnel Economics*, chap. 12. Handbook of Organizational Economics, 479–519.

Loewenstein, Mark and James Spletzer. 1997. “Delayed formal on-the-job training.” *Industrial and Labor Relations Review* 51:82–99.

———. 1998. “Dividing the costs and returns to general training.” *Journal of labor Economics* 16 (1):142–171.

———. 1999. “General and Specific Training: Evidence and Implications.” *Journal of Human Resources* 4:710–733.

Lucas, Robert. 1978. “On the size distribution of business firms.” *Bell Journal of Economics* 9:508–523.

Lynch, Lisa M and Sandra E Black. 1998. “Beyond the incidence of employer-provided training.” *ILR Review* 52 (1):64–81.

MacDonald, Glenn. 1982. “A market equilibrium theory of job assignment and sequential accumulation of information.” *American Economic Review* 72:1038–1055.

Manchester, Colleen Flaherty. 2010. “Investment in general human capital and turnover intention.” *American Economic Review* 100 (2):209–13.

———. 2012. “General human capital and employee mobility: How tuition reimbursement increases retention through sorting and participation.” *ILR Review* 65 (4):951–974.

Melero, Eduardo. 2010. “Training and promotion: allocation of skills or incentives?” *Industrial Relations* 49 (4):640–667.

Mukherjee, Arijit and Luís Vasconcelos. 2018. “Assignment and turnover: the role of breakup fees.” *Journal of Law, Economics, and Organization* 34:230–271.

Mullainathan, Sendhil. 2002. “A memory-based model of bounded rationality.” *Quarterly Journal of Economics* 117:735–774.

Pischke, Jörn-Steffen. 2001. “Continuous training in Germany.” *Journal of Population Economics* 14 (3):523–548.

Rendahl, Pontus. 2013. “Equilibrium contracts and firm sponsored training.” Mimeo, University of Cambridge.

Rosen, Sherwin. 1982. “Authority, control, and the distribution of earnings.” *Bell Journal of Economics* 13:311–323.

Scoones, David and Dan Bernhardt. 1998. “Promotion, turnover, and discretionary human capital acquisition.” *Journal of Labor Economics* 16 (1):122–141.

Waddoups, Jeffrey. 2012. “Employer-Sponsored Training and Longer-Tenured Workers: Evidence from Australia.” *Industrial Relations* 51 (4):966–986.

Waldman, Michael. 1984. “Job assignments, signalling, and efficiency.” *RAND Journal of Economics* 15:255–267.

———. 1990. “Up-or-out contracts: a signaling perspective.” *Journal of Labor Economics* 8:230–250.

Wilson, Andrea. 2014. “Bounded memory and biases in information processing.” *Econometrica* 82:2257–2294.