General Training in Labor Markets:
Common Value Auctions with Unobservable Investment

Neel Rao[*]

University at Buffalo, SUNY

July 10, 2015

Abstract

This paper studies the puzzle of employer financing for the general training of workers. A parsimonious theory is developed based on asymmetric information between employers about the quantity of training. The labor market is modeled as a common value auction with an informed and an uninformed bidder. The novel feature of the game is that one of the bidders can make an unobservable investment that increases the value of the item before the auction. By randomizing the amount of training provided, an employer can create an endogenous adverse selection problem, enabling it to compress the wage structure and capture some returns from its training investment. The model generates continuous equilibrium wage and training distributions, and identical employees can receive different wage offers and training levels. A parametric example is used to illustrate how the shape of the wage distribution depends on the elasticity of production with respect to human capital.

JEL Classification: D44, D82, J24, J31
Keywords: common value auctions, asymmetric information, investment behavior, general human capital, wage dispersion, labor markets

---

[*]E-mail: neelrao@buffalo.edu. Tel: 716-645-8674. Address: 423 Fronczak Hall, Buffalo, NY 14260. I thank Robert Gibbons, Oliver Hart, two anonymous referees, and the editor for useful comments.

1 Introduction

Why do firms pay for the general training of employees? This paper constructs a model of common value auctions with unobservable investment in order to address this question, which has interested several generations of economists. In *Wealth and Welfare*, Pigou (1912) observed that "under a free economy...socially profitable expenditure by employers in the training of their workpeople...does not carry a corresponding private profit." A trained worker might quit his or her current position for a higher paying job, or an employer might need to pay a higher wage in order to retain a trained worker. Hence, a firm may not capture all of the returns to an investment in general training. Some of the gains may accrue to the worker or even other firms in the labor market. Because of this sort of poaching externality, firms might underinvest in training, resulting in inefficiently low levels of human capital and labor productivity.

In *Human Capital*, Becker (1964) presented an influential analysis of training in perfectly competitive labor markets. A key assumption of this study, which is relaxed in the current paper, is the observability and contractibility of training. According to Becker (1964), training can be either specific or general. Both a worker and a firm can share the costs of and returns to specific training, which is useful only at the firm where it is received. By contrast, general training is widely applicable, augmenting the productivity of a worker in numerous firms. If the labor market is competitive, then a worker is paid a wage equal to his or her marginal product. In this case, employers cannot recover any of the returns to general training and are unwilling to fund the general training of employees. Ideally, a worker can finance the efficient level of training by incurring a tuition fee or accepting a wage cut.

Some evidence supports the prediction that workers must bear a large portion of training costs. For example, Minns and Wallis (2013) describe the premiums that apprentices in preindustrial England paid to masters before receiving instruction. Nonetheless, as Bishop (1996) notes when reviewing empirical work on the subject, employers often cover some of the expenses for general training. For example, Barron, Berger, and Black (1999) find only a small impact of training on the starting wage of an employee, and Loewenstein and Spletzer (1999) find little difference between the wage gains from general and specific training. In addition, Acemoglu and Pischke (1998) argue that the monetary cost of apprenticeships in Germany is largely borne by employers, and Autor (2001) discusses the free provision of general training by temporary help firms in the United States. Picchio and van Ours (2011) observe that labor market frictions raise training investments by employers.

A considerable theoretical literature has developed in order to account for outlays by

firms on the general training of employees. Several authors have suggested that adverse selection might reduce the mobility of workers between employers, thereby enabling firms to recover some returns to investments in general human capital. Chiang and Chiang (1990) analyze the role of asymmetric information between employers about the teachability of a worker. Katz and Ziderman (1990) propose that an incumbent employer might have better knowledge than outside firms about the value of training provided, and Chang and Wang (1996) attempt to formalize this idea. Acemoglu and Pischke (1998) construct a model in which the innate ability of a worker is observable to the current employer of the worker but not to outside firms. Autor (2001) argues that more able workers self-select into jobs that offer training, which helps employers screen workers for their ability.

Informational asymmetries are not the only sort of market imperfection that can stimulate general training by employers. Acemoglu (1997) demonstrates how search frictions might facilitate training investments by enabling firms to collect some of the surplus from an employment relationship. A linkage between general and specific human capital may also be relevant. Stevens (1994) studies skill acquisition in situations where training is a mixture of specific and general components. Lazear (2009) argues that each firm in the labor market uses a specific combination of general skills; so that, training is only partially transferable across employers. Finally, Acemoglu and Pischke (1999a) catalogue a variety of labor market institutions like labor unions, efficiency wages, minimum wages, and unemployment insurance that can enhance training investments. Overall, as Acemoglu and Pischke (1999b) demonstrate, firms may have an incentive to fund general training whenever the wage structure is compressed, in which case training investments raise the marginal product of a worker by more than they increase the wage rate.

The current paper endogenizes the process through which the wage structure is compressed. It models general training in labor markets by introducing pre-bidding investment into a first-price, sealed-bid auction with asymmetric information. The framework retains the essential features of a perfectly competitive labor market, except for the observability and contractibility of training. The key insight is that the opportunity for employers to train workers might in itself be a justification for employers to finance training. In particular, if the amount of training received from an incumbent employer cannot be accurately ascertained by the outside market, then firms may have an incentive to randomize the quantity of training provided, so as to endogenously generate uncertainty about an employee's productivity. By rationing training to workers in this way, an employer can create a winner's curse problem that deters other firms from offering high wages to its employees. This mechanism

may lower the wage below the marginal product of a worker, enabling an employer to earn a return on its training outlays.

The basic intuition behind the model is straightforward. First, suppose that an employer spends the same positive amount on training each worker. Outside firms would rationally anticipate that each worker has received this particular amount of training and so would be willing to offer each worker a wage equal to his or her marginal product at this training level. Hence, an employer could retain a worker only by paying the worker a wage no less than his or her marginal product. This arrangement cannot be supported as an equilibrium because an employer invests in training but obtains no return on its investment.

Next, suppose that an employer never trains a worker. Outside firms would believe that each worker has no training and so would be willing to offer each worker a wage no greater than the marginal product of untrained labor. Hence, an employer could retain a worker by paying the worker a wage equal to his or her marginal product without training. This situation may not be an equilibrium outcome because an employer might have an incentive to deviate by secretly training a worker and paying the worker a wage equal to his or her marginal product without training. This deviation could enable an employer to obtain a return on its training investment equal to the difference between the marginal products of a trained and untrained worker.

Finally, suppose that an employer uses a mixed strategy such that the amount of training provided varies across workers. Outside firms would be uncertain about the training received by each worker and so could not offer each worker a wage equal to his or her marginal product. As a result, an employer could potentially employ a trained worker at a wage below his or her marginal product, thereby reaping a return on its training investment. The model in the current paper has a unique Nash equilibrium, in which an employer implements this sort of investment policy.

Although the game is simple, it generates a complex pattern of behavior. Ex ante identical employees will receive unequal amounts of training. The equilibrium distribution of training levels will have a positive density on the interval between a zero training level and the socially optimal quantity. Furthermore, an incumbent employer and the outside market will both offer the same atomless distribution of wages. An incumbent employer offers different wages to employees with different training levels and the same wage to employees with the same training level. By contrast, wage offers from the outside market do not depend on the amount of training, which is unobservable except to the incumbent employer.

The remainder of this paper is structured as follows. Section 2 explains the relationship

and contribution of the paper to the literatures on training, auctions, search, and holdup. Section 3 presents a model of common value auctions with unobservable investment and discusses some of the assumptions behind the model as applied to training and wages. Section 4 characterizes the equilibria of the model. It considers the benchmark cases in which training is contractible or observable before analyzing the outcome with unobservable investment. Section 5 motivates and interprets the prediction that training varies among workers at the same firm. Section 6 provides a simple example with a Cobb-Douglas production function in which closed-form solutions for the distributions of wage offers and training levels can be obtained. Section 7 performs some modeling extensions to assess the robustness of the findings. Section 8 concludes. The proofs of all the theoretical results are given in the appendix.

2   Relation to Existing Literature

This paper lies at the intersection of four lines of research in economics. The first is the literature on training in labor markets. The current paper models training as an unobservable investment. This idea has some precedent in existing work. Katz and Ziderman (1990) discuss the notion that employers are not well informed about the training supplied by other firms, but they do not formally solve for an equilibrium with asymmetric information between employers about the training level. Chang and Wang (1996) also analyze a setting in which the amount of training provided by an employer is unobservable to the outside market. However, their model relies on match-specific human capital and results in a single level of training. In equilibrium, all workers receive the same amount of training; so that, there is effectively no asymmetric information between employers about the quantity of training. By contrast, the model in the current paper does not assume match-specific shocks to productivity and generates a continuous equilibrium distribution of training investments.

Other relevant papers are Chiang and Chiang (1990) and Acemoglu and Pischke (1998). In these studies, all workers receive the same amount of training. However, workers differ in their learning capacity according to Chiang and Chiang (1990) and in their innate ability according to Acemoglu and Pischke (1998). Such differences are known to an incumbent firm but not to prospective employers. These papers show that employers make positive investments in general training due to adverse selection. However, this result requires workers to be risk averse in Chiang and Chiang (1990) and relies on complementarities between ability and training in Acemoglu and Pischke (1998). Moreover, exogenous differences in worker characteristics are invoked in order to produce asymmetric information. The theory

in the current paper is more parsimonious. The model does not depend on risk aversion or complementary skills. The asymmetric information is endogenous, resulting from the use of a mixed strategy.

Second, this paper builds on and adds to the literature on auctions. The model of wage setting is derived from existing work on first-price, sealed-bid auctions with asymmetric information. In particular, wage determination is based on the framework in Engelbrecht-Wiggans, Milgrom, and Weber (1983), who characterize equilibrium strategies in an auction where one bidder has access to some private information that is unavailable to other bidders. This setting was originally studied by Wilson (1969) and Weverbergh (1979). It was also mentioned in Milgrom and Weber (1982a). Milgrom and Weber (1982b) analyze the incentives for bidders in this environment to acquire information about the object for sale. Hendricks and Porter (1988) apply the theory to empirically investigate auctions for oil fields. The current paper contributes by incorporating unobservable investment that affects the value of the item being auctioned. It endogenizes the private information.

In relation to labor markets, Li (2013) adapts the setup from Engelbrecht-Wiggans, Milgrom, and Weber (1983) to study job mobility and wage inequality when employers have asymmetric information. Virág (2007) considers a repeated version of this auction, deriving a number of results about bidding strategies and wage dynamics. Another sequential variant of this auction is presented in Hörner and Jamison (2008). However, these papers do not examine the investment or training decisions of bidders or firms. In this context, the contribution of the current paper is to show that informational asymmetries can arise endogenously as a result of the incentive for employers to capture some of the returns from general training.

Third, another pertinent literature comes from search theory. The role of search frictions in explaining price dispersion has been studied by several authors including Butters (1977), Burdett and Judd (1983), Lang (1991), and Burdett and Mortensen (1998). Some papers also examine how wage dispersion due to matching frictions can generate variation in investment decisions across firms. Acemoglu and Shimer (2000) demonstrate that search costs may create differences among producers in the stock of physical capital. Mortensen (1998) and Quercioli (2005) analyze firm-specific human capital in a frictional labor market. A search model with investments in general human capital can be found in Fu (2011).

In the last paper, general training is assumed to be observable and contractible, whereas the current paper investigates unobservable investment. Existing research shows that wage dispersion due to search costs causes investment differences. The direction of causation

is the opposite in the current paper. Training dispersion due to unobservable investment causes asymmetric information and wage variation. If training is assumed to be observable or contractible, then the model in the current paper would predict that all workers should receive the same wage.

Fourth, this paper shares common themes with research on the hold-up problem. In general, hold-up refers to a situation in which one party makes an investment specific to a relationship but must share the returns to the investment with another party. As a result, agents underinvest in a joint project. Gul (2001) constructs a bargaining model in which the hold-up problem between a buyer and a seller is mitigated if investment is unobservable. Lau (2008) characterizes the optimal degree of informational asymmetry in this framework.

Recent papers on similar topics are Hermalin (2013) and Kawai (2014). These authors consider a hold-up problem where a seller makes an unobservable investment that increases the value of an object to a buyer. The model in Kawai (2014) is static, whereas that in Hermalin (2013) is dynamic. Another related paper is González (2004), which analyzes a principle-agent problem in which the agent undertakes a cost-reducing investment that is unobservable to the principal. In addition, Goldlücke and Schmitz (2014) present a signaling model in which a seller follows a mixed strategy when making an observable but noncontractible investment. The existing literature on hold-up studies bargaining games between a buyer and a seller and does not examine general training by employers. By contrast, the current paper investigates the training investments of firms in the context of an auction between buyers. It extends the literature by introducing unobservable investment into an environment where trade occurs through auctions instead of bargaining.

3 Model of Training and Wages

This section presents a simple model of general training and wage setting. The framework is based on a common value auction with asymmetric bidders and unobservable investment. The game comprises two stages. First, an insider firm invests in training a worker, and the size of the investment is its private information. Second, the insider and an outsider firm simultaneously offer wages to the worker, and the worker accepts the higher offer. The details are as follows.

The labor market comprises one worker and two firms I and O. Firms I and O represent an incumbent and a competing employer, respectively. At time $t = 0$, firm I invests an amount $h \in \mathbb{R}_+$ in general training for the worker. The choice of the training level h is observable to firm I but not to firm O. Given the specification of the bidding game between

firms and the action space of the worker, it is irrelevant whether or not the worker knows the quantity h of training received.

At time $t = 1$, firms I and O make simultaneous wage offers to the worker. Let $b_I \in \mathbb{R}_+$ and $b_O \in \mathbb{R}_+$ denote the respective offers of firms I and O. The worker accepts employment from the firm offering the higher wage. If both offers are the same, then the worker accepts employment from firm I with probability α and from firm O with probability $1 - \alpha$, where $\alpha \in [0, 1]$. Firm I can use its knowledge of the training level h when formulating a wage offer b_I, whereas the wage b_O offered by firm O cannot depend on h. The worker produces an output worth $g(h)$ for the firm that hires him or her at time $t = 1$.

The function g mapping training into output is assumed to satisfy the following regularity conditions.

Assumption 1 *The production function $g : \mathbb{R}_+ \to \mathbb{R}_+$ has the following properties:*

a) $g'(h) > 0$ and $g''(h) < 0$ for all $h \in \mathbb{R}_+$

b) $\lim_{h \to 0} g'(h) = \infty$ and $\lim_{h \to \infty} g'(h) = 0$

The first part of the assumption states that the output of the worker is a concave, increasing, and twice differentiable function of training. The Inada conditions in the second part of the assumption help ensure that the efficient level of training is positive.

The firms are assumed to be risk neutral, and there is no discounting between periods. Hence, the payoffs to firms I and O are defined as follows. If firm I chooses training level h and wage offer w_I, then its payoff is $[g(h) - w_I] - h$ when the worker accepts its offer, and its payoff is $-h$ when the worker rejects its offer. If firm O offers the wage w_O, then its payoff is $[g(h) - w_O]$ when the worker accepts its offer, and its payoff is 0 when the worker rejects its offer.

Note that firms are restricted to offering flat-wage contracts. In other words, output is assumed to be unverifiable by courts and thus noncontractible. This restriction is standard in the literature on adverse selection in labor markets. For example, see the models of asymmetric information in Greenwald (1986) and Gibbons and Katz (1991). Moreover, if output were instead assumed to be verifiable, then training would effectively be contractible since output is a function of training. However, this situation would be inconsistent with the aim of the current paper, which is to understand investment behavior when training is unobservable and noncontractible.

In practice, there are several reasons why employers may avoid output-contingent contracts. First, performance pay might significantly raise an employer's compensation costs if a worker is risk averse and the output measure is noisy. See Prendergast (2002) for a discussion of the literature on economic uncertainty and performance pay. Second, Baker, Jensen, and Murphy (1987) observe that employees may substitute quantity for quality if they are paid on a piece rate. Third, Kanemoto and MacLeod (1992) analyze a dynamic model in which a piece rate induces a ratchet effect that may inefficiently lower productivity. Fourth, an employee's individual contribution to production may be difficult to determine if agents work in teams with only group output being observed. See Alchian and Demsetz (1972) for a classic description of the metering problem associated with team production.

The central assumption of the model is that training investments are not publicly observable. The literature on general training provides substantial support for this premise. Katz and Ziderman (1990) argue that "potential recruiters do not possess much information on the extent and type of workers' on-the-job training." Likewise, Acemoglu and Pischke (1999a) note that "training practices inside the firm are hard to observe by outsiders" and "important parts of the training programme are intangible, involving mentoring, advice, and practice." The hypothesis that training is unobservable has some precedent in the literature, but the current paper is unique in deriving an equilibrium in which there is asymmetric information between employers about the quantity of training as well as a nondegenerate distribution of training levels.

Although a worker might have knowledge of training received, he or she may be unable to credibly signal this information to prospective employers. Acemoglu and Pischke (1999b) mention that employer-sponsored training often provides "uncredentialed skills." That is, workers normally do not receive a certificate of completion or competency that they can present to outsiders as evidence of on-the-job learning. It may be infeasible to design a standardized test or interview question to assess the quality of instruction. Even if possible, formal verification of training may not generate an efficiency improvement. As Katz and Ziderman (1990) observe, "certification might lead to less rather than more general training" because of a reduction in the returns that accrue to employers.

Some training activities like mentoring and advising may be spontaneous and idiosyncratic. As Bishop (1996) notes, a substantial component of "training is informal and hard to measure and its effects on productivity are even more difficult to quantify." Similarly, Brown (1990) argues that "much of the conceptual difficulty of measuring employer-provided training is due to the fact that an important part of such training occurs informally, on the job."

These problems of measurement are less characteristic of investment in physical capital such as machinery or equipment. Becker (1962) describes physical capital as "tangible assets," whereas human capital comprises "intangible resources." Accordingly, human capital may be subject to greater uncertainty than physical capital, which is typically concrete, material, and observable. Furthermore, human capital is "inalienable" as emphasized by Hart and Moore (1994). An employee is generally free to leave a job at will, in which case the employer loses access to the skills embodied in the worker. This special property of human capital makes it difficult for a firm to obtain a return on its investment.

Another important assumption of the model is that wage offers are made simultaneously. That is, each firm selects a wage offer without observing the wage chosen by the other firm. Other papers that explore the implications of this specification are Virág (2007) and Li (2013). Those authors provide a number of justifications for modeling offers as simultaneous. First, a wage offer may be easy for a worker to falsify because a worker is not compelled to accept an offer of employment. Hence, workers might be unable to credibly reveal a wage offer from one firm to another employer. Second, compensation may be partly in the form of nonmonetary perks, which could benefit different workers unequally. Hence, a firm might have difficulty appraising the value of an employment offer from another firm. Third, an employer may enforce a policy of ignoring wage offers from other firms. Otherwise, if an employer matched offers from competitors, then workers might try to solicit offers from other firms in order to start a bidding war among employers that raises wages and lowers profits.

4 Analysis of Investment Behavior

This section examines training decisions under different assumptions about the contracting environment. It begins with two benchmark scenarios. Section 4.1 considers the situation where training is contractible; so that, an efficient level of investment can be implemented. Section 4.2 documents the outcome when training is observable but not contractible, in which case no investment occurs on the path of play. It then proceeds to analyze the setting of interest, where investment is unobservable and so noncontractible. Section 4.3 derives an equilibrium with randomized training and wage dispersion.

4.1 Contractible Investment

In this section, training is assumed to be contractible. Hence, the worker can pay firm I a fee in exchange for the efficient amount of training. The efficient investment level is defined as follows.

Definition 1 *The training level $\hat{h} \in \mathbb{R}_+$ is efficient if $g(\hat{h}) - \hat{h} \geq g(h) - h$ for all $h \in \mathbb{R}_+$.*

In other words, the efficient investment level maximizes revenue minus training costs. The result below characterizes the socially optimal quantity of training.

Proposition 1 *The efficient training level is given by $h^e = g'^{-1}(1)$.*

That is, the marginal revenue of $g'(h)$ from an additional unit of human capital is equated with the marginal cost of 1 for a unit of human capital. An agreement between the worker and firm I might have the following form. First, the worker would pay firm I a fee f. Depending on the bargaining powers of the two parties, the fee can vary from h^e to $g(h^e) - g(0)$, where $f = h^e$ and $f = g(h^e) - g(0)$ correspond to the respective cases where all the bargaining power is vested in the worker and the firm. Second, firm I would provide the efficient amount h^e of training to the worker. Third, Bertrand competition between firms I and O would result in the worker being paid a wage equal to his or her marginal product $g(h^e)$.

4.2 Observable Investment

In this section, training is assumed to be observable but not contractible. The setup is as described in section 3, except that firms I and O both observe the amount of training received by the worker. A formal definition of strategies for this game is provided below.

Definition 2 *A mixed strategy for firm I consists of a distribution function D on $\mathbb{R}_+$ as well as a collection $\{B_I^h\}$ of distribution functions on $\mathbb{R}_+$ indexed by $h \in \mathbb{R}_+$. A mixed strategy for firm O is a collection $\{B_O^h\}$ of distribution functions on $\mathbb{R}_+$ indexed by $h \in \mathbb{R}_+$.*

The preceding expressions can be interpreted as follows. $D(h)$ is the probability that firm I chooses a training level no greater than h, and $B_I^h(w_I)$ is the probability that firm I chooses a wage offer no greater than w_I when the training level is h. Given these terms, let $B_I(h, w_I) = D(h) \cdot B_I^h(w_I)$ be the probability that firm I chooses a training level no greater than h and a wage offer no greater than w_I. Finally, $B_O^h(w_O)$ is the probability that firm O offers a wage no greater than w_O when the training level is h.

To paraphrase, firm I chooses a training level h that is publicly observable. Thereafter, firms I and O simultaneously and independently place bids for the labor services of the worker. The wage offers w_I and w_O can vary with the amount h invested. The solution concept for this model is subgame perfect Nash equilibrium, which is described below. The notation $B_I^h(w^-) = \lim_{v\uparrow w} B_I^h(v)$ and $B_O^h(w^-) = \lim_{v\uparrow w} B_O^h(v)$ is used when $w > 0$. In addition, let $B_I^h(0^-) = 0$ and $B_O^h(0^-) = 0$.

Definition 3 *A subgame perfect Nash equilibrium consists of mixed strategies $(D, \{B_I^h\})$ and $\{B_O^h\}$ for firms I and O such that:*

a) *for any $h \in \mathbb{R}_+$, there is probability one that $\hat{w}_I$ drawn from distribution B_I^h satisfies $\{B_O^h(\hat{w}_I^-) + \alpha[B_O^h(\hat{w}_I) - B_O^h(\hat{w}_I^-)]\}[g(h) - \hat{w}_I] \geq \{B_O^h(w_I^-) + \alpha[B_O^h(w_I) - B_O^h(w_I^-)]\}[g(h) - w_I]$ for all $w_I \in \mathbb{R}_+$;*

b) *for any $h \in \mathbb{R}_+$, there is probability one that $\hat{w}_O$ drawn from distribution B_O^h satisfies $\{B_I^h(\hat{w}_O^-) + (1 - \alpha)[B_I^h(\hat{w}_O) - B_I^h(\hat{w}_O^-)]\}[g(h) - \hat{w}_O] \geq \{B_I^h(w_O^-) + (1 - \alpha)[B_I^h(w_O) - B_I^h(w_O^-)]\}[g(h) - w_O]$ for all $w_O \in \mathbb{R}_+$;*

c) *there is probability one that $(\hat{h}, \hat{w}_I)$ drawn from distribution B_I satisfies $\{B_O^{\hat{h}}(\hat{w}_I^-) + \alpha[B_O^{\hat{h}}(\hat{w}_I) - B_O^{\hat{h}}(\hat{w}_I^-)]\}[g(\hat{h}) - \hat{w}_I] - \hat{h} \geq \{B_O^h(w_I^-) + \alpha[B_O^h(w_I) - B_O^h(w_I^-)]\}[g(h) - w_I] - h$ for all $(h, w_I) \in \mathbb{R}_+^2$.*

The equilibrium conditions can be stated as follows. Conditional on the amount h invested, each firm chooses a wage offer so as to maximize its profits given the distribution of wages offered by the other firm. Moreover, firm I selects a training level and wage offer so as to maximize its profits given the distribution of wages offered by firm O upon observing the choice of investment. The result below summarizes the equilibrium of the game.

Proposition 2 *The unique subgame perfect Nash equilibrium outcome is as follows. With probability one, firm I chooses training level $h = 0$ and wage offer $w_I = g(0)$. With probability one, firm O chooses wage offer $w_O = g(0)$. The resulting payoffs are 0 to firm I, 0 to firm O, and $g(0)$ to the worker.*

In other words, no investment occurs in equilibrium. Consequently, each firm receives zero profits, and the worker earns a wage equal to the marginal product of untrained labor. The reasoning behind this finding is straightforward. Due to Bertrand competition between firms I and O, the worker is always paid his or her marginal product. Hence, firm I would obtain no return from training but would incur the cost of investment. That is, it is not profitable for firm I to invest in training the worker.

4.3 Unobservable Investment

As in section 3, training is assumed to be unobservable and so noncontractible. Below is the specification of strategies in this case.

Definition 4 *A mixed strategy for firm I is a distribution function G_I on $\mathbb{R}_+^2$. A mixed strategy for firm O is a distribution function F_O on $\mathbb{R}_+$.*

The terms in the preceding statement have the following meanings. $G_I(h, w_I)$ is the probability that firm I chooses a training level no greater than h and a wage offer no greater than w_I. $F_O(w_O)$ is the probability that firm O offers a wage no greater than w_O. That is, firm I selects an amount h to invest, after which both firms I and O make simultaneous wage offers w_I and w_O to the worker. A critical aspect of the model is that firm I but not firm O knows the training level h when placing a bid.

This is a game of complete but imperfect information. Any informational asymmetry between firms I and O will arise not from a random move of Nature but from the endogenous behavior of firm I. Hence, the appropriate solution concept is just a Nash equilibrium, which is defined below for the current setting. The notation $F_I(w^-) = \lim_{v \uparrow w} F_I(v)$ and $F_O(w^-) = \lim_{v \uparrow w} F_O(v)$ is used when $w > 0$. In addition, let $F_I(0^-) = 0$ and $F_O(0^-) = 0$.

Definition 5 *A Nash equilibrium consists of mixed strategies G_I and F_O for firms I and O such that:*

a) *there is probability one that $(\hat{h}, \hat{w}_I)$ drawn from distribution G_I satisfies $\{F_O(\hat{w}_I^-) + \alpha[F_O(\hat{w}_I) - F_O(\hat{w}_I^-)]\}[g(\hat{h}) - \hat{w}_I] - \hat{h} \geq \{F_O(w_I^-) + \alpha[F_O(w_I) - F_O(w_I^-)]\}[g(h) - w_I] - h$ for all $(h, w_I) \in \mathbb{R}_+^2$;*

b) *there is probability one that $\hat{w}_O$ drawn from distribution F_O satisfies $\mathbb{E}\{[(1-\alpha)\mathbb{1}(\hat{w}_O \geq w_I) + \alpha\mathbb{1}(\hat{w}_O > w_I)][g(h) - \hat{w}_O]\} \geq \mathbb{E}\{[(1-\alpha)\mathbb{1}(w_O \geq w_I) + \alpha\mathbb{1}(w_O > w_I)][g(h) - w_O]\}$ for all $w_O \in \mathbb{R}_+$, where the expectation $\mathbb{E}\{\cdot\}$ is taken over (h, w_I) drawn from distribution G_I, and $\mathbb{1}(\cdot)$ is an indicator function equal to 1 if the statement in parentheses is true and equal to 0 otherwise.*

The requirements for an equilibrium are simple. Firm I chooses a wage offer and training level so as to maximize its profits given the distribution of wages offered by firm O. Firm O chooses a wage offer so as to maximize its profits given the joint distribution of wage offers and training levels selected by firm I.

The process of solving the model is subdivided into several parts. The basic properties of an equilibrium are first derived. The unique strategy profile satisfying these conditions

is then identified. The result below specifies the equilibrium relationship between the wage offered and the training provided by firm I. It is convenient to denote $P_O(w_I) = F_O(w_I^-) + \alpha[F_O(w_I) - F_O(w_I^-)]$ for $w_I \in \mathbb{R}_+$. That is, $P_O(w_I)$ represents the probability that firm I retains the worker when offering the wage w_I given that firm O chooses a bid according to the distribution F_O.

Lemma 1 *Suppose that the mixed strategies G_I and F_O for firms I and O constitute a Nash equilibrium. There is probability one that $(\hat{h}, \hat{w}_I)$ drawn from distribution G_I satisfies $\hat{h} = g'^{-1}[1/P_O(\hat{w}_I)]$.*

In other words, the wage offer $\hat{w}_I$ and training level $\hat{h}$ chosen by firm I almost surely fulfill the equation $\hat{h} = g'^{-1}[1/P_O(\hat{w}_I)]$. This property maps each wage offered by the incumbent employer to a unique amount of training. As explained below, it helps simplify the specification of an equilibrium strategy for firm I.

In particular, consider a Nash equilibrium in which G_I and F_O are the respective strategies of firms I and O. Denote $x(w_I) = g'^{-1}[1/P_O(w_I)]$ for all $w_I \in \mathbb{R}_+$. The strategy G_I for firm I can be represented as a pair (F_I, x), where F_I is a distribution function on $\mathbb{R}_+$, and the function x is as just described. In this formulation, $F_I(w_I)$ is the probability that firm I offers a wage no greater than w_I, and $x(w_I)$ can be interpreted as the training level chosen by firm I when it selects the wage offer w_I. Note that G_I and (F_I, x) are related through the equation $F_I(w) = G_I[x(w), w]$ for all $w \in \mathbb{R}_+$.

The preceding lemma reflects a simple intuition. Given the wage offered by firm I, there is a particular probability that firm I has of retaining the worker. In turn, firm I should optimally provide a certain amount of training to the worker based on the likelihood of retention, which affects the fraction of the return on an investment that firm I captures. According to the relation derived above, the training level is nondecreasing in the wage offer. When firm I offers the worker a higher wage, the worker may be more likely to remain with firm I, enabling firm I to recover more of the gains from training. Consequently, firm I may find it profitable to invest a larger amount in the human capital of the worker.

The basic structure of an equilibrium is deduced by a lengthy process of elimination. As detailed in the appendix, the results below are mostly proven by contradiction. An equilibrium of a particular form is hypothesized, but a profitable deviation is shown to exist. The next lemma relates to the infima and suprema of the supports for the distributions of wages offered by firms I and O. For $Y \in \{I, O\}$, let w_Y^l be the least $w \in \mathbb{R}_+$ such that $F_Y(w) > 0$ for all $w > w_Y^l$. That is, w_I^l and w_O^l respectively denote the infimum of the support of F_I and F_O. For $Y \in \{I, O\}$, let w_Y^u be the greatest $w \in \mathbb{R}_+$ such that $F_Y(w) < 1$

for all $w < w_Y^u$. That is, w_I^u and w_O^u respectively denote the supremum of the support of F_I and F_O.

Lemma 2 *Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium. It must be that $w_I^l = w_O^l$ and $w_I^u = w_O^u$.*

In other words, the supports of the wage distributions offered by the incumbent and outside employers have the same infimum and supremum. Henceforth, the notation $w^l = w_I^l = w_O^l$ and $w^u = w_I^u = w_O^u$ is used given equilibrium strategies (F_I, x) and F_O for firms I and O.

The following is a sketch of the reasoning behind the result above. Let (F_I, x) and F_O be the respective strategies of firms I and O in a Nash equilibrium. Consider first the suprema of the supports of F_I and F_O. Suppose that $w_I^u < w_O^u$. In this case, firm O can get a higher expected payoff by offering a wage slightly below w_O^u than by offering the wage w_O^u. Hence, firm O has an incentive to deviate from choosing the bid w_O^u. Intuitively, this implies that there cannot exist an equilibrium in which $w_I^u < w_O^u$. A similar argument holds under the supposition that $w_I^u > w_O^u$. Therefore, it must be that $w_I^u = w_O^u$ in any equilibrium.

Consider next the infima of the supports of F_I and F_O. Suppose that $w_I^l > w_O^l$. If so, firm I has a positive probability of retaining the worker by offering a wage no less than w_I^l. Hence, it is optimal for firm I to provide a positive amount of training when it chooses any bid greater than or equal to w_I^l. Given that firm I must get a nonnegative expected payoff in equilibrium, it can be argued that firm O will get a positive expected payoff by offering a wage slightly above w_I^l. Intuitively, firm O may reap some of the gains from the training received by the worker, even though it does not pay for this investment. However, firm O gets zero profits by offering the worker a wage below w_I^l. Hence, firm O has an incentive to deviate from choosing a bid lower than w_I^l. Thus, there cannot exist an equilibrium in which $w_I^l > w_O^l$.

Suppose instead that $w_I^l < w_O^l$. When firm I offers any wage below w_O^l, the worker accepts the offer of firm I with probability zero. Hence, it is not optimal for firm I to provide any training when it chooses any bid less than w_O^l. Assume for simplicity that F_I does not have an atom at w_O^l. If $w_O^l > g(0)$, then firm O gets a negative expected payoff from offering the wage w_O^l, whereas it can get zero profits by bidding $g(0)$. If $w_O^l = g(0)$, then firm O gets zero profits from offering the wage w_O^l, whereas it can obtain a positive expected payoff by bidding slightly below $g(0)$. If $w_O^l < g(0)$, then firm I receives zero profits from choosing a wage offer lower than w_O^l and a training level of 0, although it can get a positive expected payoff by choosing a wage offer slightly higher than w_O^l along with a training level

of 0. In each case, firm I or O appears to have a profitable deviation. Thus, there cannot exist an equilibrium in which $w_I^l < w_O^l$. Overall, it must be that $w_I^l = w_O^l$ in any equilibrium.

The lemma below states additional properties regarding the infimum of the supports for the distributions of wages offered by firms I and O. In particular, the supports of the equilibrium bid distributions F_I and F_O have an infimum w^l equal to $g(0)$, which is the output produced by a worker with no training. Furthermore, the equilibrium distribution F_O of wages offered by the outside employer does not have a mass point at the infimum w^l of its support.

Lemma 3 *Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium. It must be that $w^l = g(0)$. Moreover, F_O does not have an atom at w^l.*

It will also follow from later steps of the solution that there is no mass point at the infimum w^l of the support for the equilibrium bid distribution F_I of the incumbent employer. The logic underlying the lemma above can be summarized as follows. Consider a Nash equilibrium in which (F_I, x) and F_O are the respective strategies of firms I and O. Assume provisionally that $P_O(w^l) = 0$, which implies that $x(w^l) = 0$. That is, firm I is assumed to have zero probability of retaining the worker by offering the wage w^l, in which case it is not profitable for firm I to provide any training to the worker.

Suppose first that $w^l > g(0)$. If F_I has an atom at w^l, then firm O receives a negative expected payoff by bidding w^l, whereas it can get zero profits by offering a wage less than w^l. If F_I does not have an atom at w^l, then firm O receives zero profits by offering the wage w^l, but it can be argued that firm O can obtain a positive expected payoff by bidding slightly higher than w^l. The intuition is that firm O may collect some returns from the training of the employee without incurring any costs for the training. In both cases, firm O has an incentive to deviate from offering the wage w^l. Hence, there does not seem to exist an equilibrium in which $w^l > g(0)$ under the assumption that $P_O(w^l) = 0$.

Suppose next that $w^l < g(0)$. Firm I receives zero profits by choosing wage offer w^l and training level $x(w^l)$, whereas it can get a positive expected payoff by selecting a bid slightly greater than w^l along with the training level $x(w^l)$. This means that it is profitable for firm I to deviate from offering the wage w^l and providing the training $x(w^l)$. Thus, an equilibrium with $w^l < g(0)$ does not appear to exist under the assumption that $P_O(w^l) = 0$. In sum, it must be that $w^l = g(0)$ whenever $P_O(w^l) = 0$.

It now remains to argue that $P_O(w^l) = 0$. Suppose to the contrary that $P_O(w^l) > 0$, which implies that $x(w) > 0$ for $w \geq w^l$. That is, if firm I offers the wage w^l, then firm I is presumed to retain the employee with positive probability. This would make it optimal

16

for firm I to provide a positive amount of training whenever it offers a wage no less than w^l. Noting that the expected payoff to firm I must be nonnegative in equilibrium, firm O can increase its profits by bidding slightly higher than w^l instead of offering the wage w^l. Intuitively, firm O improves its chance of hiring the worker by bidding above w^l, and firm O gains from hiring the worker because it captures some returns to training without paying for the investment. Hence, it is profitable for firm O to deviate from offering the wage w^l. This implies that there cannot be an equilibrium with $P_O(w^l) > 0$. Consequently, $w^l = g(0)$.

The final task is to explain why F_O does not have an atom at w^l. Suppose instead that F_O has an atom at w^l. If so, firm I receives zero profits by choosing wage offer $w^l = g(0)$ and training level $x(w^l) = 0$, whereas firm I can obtain a positive expected payoff by offering a wage slightly higher than w^l and providing somewhat greater training than $x(w^l)$. That is, firm I has an incentive to deviate from selecting the bid w^l along with the training level $x(w^l)$. Thus, there does not appear to be an equilibrium where F_O has an atom at w^l.

The result below reports the average profits of firms I and O in equilibrium. The expected payoffs are defined as follows. For any $(h, w) \in \mathbb{R}_+^2$, let $\pi_I(h, w) = P_O(w)[g(h) - w] - h$ denote the expected payoff to firm I if firm I chooses training level h and wage offer w. Choose any $w \in \mathbb{R}_+$. Let $\mathbb{E}\{g[x(w_I)] - w | w_I \leq w\}$ be the conditional expectation of $g[x(w_I)] - w$ given that $w_I \leq w$, where w_I is drawn from F_I. If there is zero probability that w_I drawn from F_I satisfies $w_I \leq w$, then $\mathbb{E}\{g[x(w_I)] - w | w_I \leq w\}$ is allowed to be any real number. Let $\pi_O(w) = F_I(w)\mathbb{E}\{g[x(w_I)] - w | w_I \leq w\} - \alpha[F_I(w) - F_I(w^-)]\{g[x(w)] - w\}$ denote the expected payoff to firm O if firm O offers wage w.

Lemma 4 *Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium. There is probability one that $\hat{w}_I$ drawn from F_I satisfies $\pi_I[x(\hat{w}_I), \hat{w}_I] = 0$. There is probability one that $\hat{w}_O$ drawn from F_O satisfies $\pi_O(\hat{w}_O) = 0$.*

To paraphrase, firm I almost surely chooses a wage offer and training level for which its expected payoff is zero, and firm O almost surely offers a wage such that its expected payoff is zero. The following is the main idea behind the preceding lemma. Assume that (F_I, x) and F_O are the respective strategies of firms I and O in a Nash equilibrium. Firm O gets zero profits by offering the wage w^l because it has zero probability of hiring a trained worker if it bids w^l. Intuitively, firm O should be indifferent among the wages that it offers in equilibrium because any such wage should maximize its expected payoff. Hence, firm O would generally obtain an expected payoff of zero in equilibrium.

A similar argument can be applied to the expected payoff of firm I. Note that the actual proof of the lemma in the appendix is more complicated than the logic presented above. The

17

reason is as follows. Given any $\epsilon > 0$, there is positive probability that firm O chooses a bid in the interval $[w^l, w^l + \epsilon]$. However, firm O does not offer the wage w^l with positive probability in equilibrium. Hence, it is not entirely obvious that firm O maximizes its expected payoff by bidding exactly w^l. A related issue makes the proofs of several other results in the appendix more elaborate than the intuition described in the body of the paper.

The next result pertains to the supremum of the supports for the distributions of wages offered by firms I and O. Specifically, it shows that the supports of the equilibrium bid distributions F_I and F_O have a supremum w^u equal to $g[g'^{-1}(1)] - g'^{-1}(1)$. This value represents the output of a worker with the efficient level of training minus the cost of such a training investment.

Lemma 5 *Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium. It must be that $w^u = g[g'^{-1}(1)] - g'^{-1}(1)$.*

The following is an outline of the proof for the lemma above. Consider a Nash equilibrium in which firms I and O follow the respective strategies (F_I, x) and F_O. If $w^u < g[g'^{-1}(1)] - g'^{-1}(1)$, then firm I gets zero profits by choosing wage offer w^l and training level $x(w^l)$, but firm I would obtain a positive expected payoff by offering a wage slightly greater than w^u and providing the training level $g'^{-1}(1)$. If $w^u > g[g'^{-1}(1)] - g'^{-1}(1)$, then firm I can be shown to obtain a negative expected payoff by choosing wage offer w^u and training level $x(w^u)$, whereas it would get zero profits by choosing wage offer w^l and training level $x(w^l)$. Hence, firm I seems to have an incentive to deviate in both cases, contradicting the existence of an equilibrium in which $w^u \neq g[g'^{-1}(1)] - g'^{-1}(1)$.

The lemma below establishes some important characteristics for the distributions of wages offered by firms I and O. It demonstrates that the equilibrium wage offer distributions F_I and F_O do not have a mass point above the infimum w^l of their supports. In addition, the distribution functions F_I and F_O are shown to be strictly increasing on the interval $[w^l, w^u]$, which forms the support of F_I and F_O. In other words, the equilibrium bid distribution for each firm is gapless and atomless above the infimum of its support.

Lemma 6 *Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium. Then neither F_I nor F_O has an atom at any $w > w^l$. Moreover, for any w^a and w^b such that $w^l \leq w^a < w^b \leq w^u$, it must be that $F_I(w^a) < F_I(w^b)$ and $F_O(w^a) < F_O(w^b)$.*

A synopsis of the argument behind the statement above would be as follows. Let (F_I, x) and F_O be the respective strategies of firms I and O in a Nash equilibrium. Suppose first that

F_I has an atom at some wage $\hat{w} > w^l$. Assume for concreteness that $\alpha > 0$. That is, there is positive probability of the worker accepting employment with firm I when both firms offer the same wage.

It can be shown that there does not exist an equilibrium in which firm O offers a wage equal to or slightly less than $\hat{w}$ with positive probability. To understand this claim, suppose to the contrary that such an equilibrium exists. Given that firm I receives an expected payoff of zero in equilibrium, firm O can increase its expected payoff by bidding slightly above $\hat{w}$ instead of offering a wage equal to or slightly less than $\hat{w}$. Intuitively, firm O substantially improves its likelihood of recruiting a trained worker by incrementally raising its bid to just over $\hat{w}$, and firm O profits from employing such a worker because it captures some of the gains from training but does not finance the cost of training. Hence, firm O would have an incentive to deviate under such a scenario.

Since there is zero probability of firm O choosing a bid equal to or slightly below $\hat{w}$ in equilibrium, there exists a profitable deviation for firm I. In particular, firm I can increase its expected payoff by choosing a wage offer slightly below $\hat{w}$ along with a training level of $x(\hat{w})$ instead of offering the wage $\hat{w}$ and providing the training $x(\hat{w})$. By doing so, firm I reduces the wage paid when it retains the worker without lowering its chance of hiring the worker. Hence, F_I cannot have an atom at any wage $\hat{w} > w^l$. A similar style of reasoning suggests that F_O does not have a mass point at any wage above the infimum of its support.

Suppose now that there exists $w^a \geq w^l$ such that one can find $w^b \leq w^u$ for which $w^b > w^a$ but $F_I(w^b) = F_I(w^a)$. That is, firm I offers a wage greater than w^a but less than or equal to w^b with probability zero. Noting that F_I is continuous on the interval $[w^l, w^u]$, let w^c be the largest number $w^c \leq w^u$ with $w^c > w^a$ but $F_I(w^c) = F_I(w^a)$. It can be shown that no equilibrium exists in which firm O chooses a bid higher than w^a but no greater than w^c with positive probability. To see this, suppose otherwise that such an equilibrium exists. Then firm O can increase its expected payoff by offering the wage w^a instead of bidding over w^a but less than or equal to w^c. Thus, firm O would have an incentive to deviate in this setting.

Since there is zero probability of firm O offering a wage above w^a but below or the same as w^c, there appears to be a profitable deviation for firm I. Specifically, firm I can increase its expected payoff by choosing a bid slightly under w^c along with the investment level $x(w^c)$ instead of selecting the bid w^c and investing the amount $x(w^c)$. By doing so, firm I does not change its probability of retaining the worker but decreases the wage paid in the event that the worker accepts its offer. Hence, there cannot exist w^a and w^b with $w^l \leq w^a < w^b \leq w^u$ for which $F_I(w^a) = F_I(w^b)$. A related justification can be given for why F_O is strictly increasing

between w^l and w^h.

The discussion so far has concentrated on the distributions of wages offered by firms I and O. However, the preceding results also have implications for the equilibrium distribution of human capital. The lemma below lists some key features of the investment distribution. Given a Nash equilibrium strategy profile (F_I, x) and F_O, let the distribution function K on $\mathbb{R}_+$ denote the resulting equilibrium distribution of training levels. That is, $K(h)$ is the probability that firm I chooses a training level no greater than h. In addition, let h^l and h^u respectively denote the infimum and supremum of the support of K. That is, h^l is the least $h \in \mathbb{R}_+$ such that $K(h) > 0$ for all $h > h^l$, and h^u is the greatest $h \in \mathbb{R}_+$ such that $K(h) < 1$ for all $h < h^u$.

Lemma 7 *Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium. Then $h^l = 0$ and $h^u = g'^{-1}(1)$. Furthermore, K does not have an atom at any $h > h^l$. Lastly, for any h^a and h^b such that $h^l \le h^a < h^b \le h^u$, it must be that $K(h^a) < K(h^b)$.*

According to the lemma above, the support of the equilibrium training distribution has an infimum h^l of 0 and a supremum h^u of $g'^{-1}(1)$. Recall that 0 is the amount invested when training is observable but not contractible as in section 4.2 and that $g'^{-1}(1)$ is the amount invested when training is observable and contractible as in section 4.1. Noting that $g'^{-1}(1)$ is the efficient quantity of investment, training is generally below the socially optimal level in equilibrium.

The basic logic behind this part of the lemma is straightforward. Consider equilibrium strategies (F_I, x) and F_O. First, it is optimal for firm I to provide the training $h^l = 0$ if it offers the wage $w^l = g(0)$, which is the infimum of the support for the wage offer distribution F_I. In this case, firm I hires the worker with probability zero and so receives none of the returns to investment. Second, it is optimal for firm I to invest the amount $h^u = g'^{-1}(1)$ if it chooses the bid $w^u = g[g'^{-1}(1)] - g'^{-1}(1)$, which is the supremum of the support for the equilibrium bid distribution F_I. In this case, firm I retains the employee with probability one and so obtains all of the returns to training.

The preceding lemma also shows that the distribution function K is continuous and strictly increasing on the interval $[h^l, h^u]$. In other words, the distribution of human capital is gapless and atomless above the infimum of its support. Intuitively, the investment distribution inherits these properties from the bid distribution. Let (F_I, x) and F_O be equilibrium strategies. Then the wage offer distribution F_I of firm I is continuous and strictly increasing on the interval $[w^l, w^u]$. Furthermore, the wage offer $\hat{w}_I$ and training level $\hat{h}$ chosen by firm

I satisfy the condition $\hat{h} = g'^{-1}[1/P_O(\hat{w}_I)]$ with probability one, where P_O is continuous and strictly increasing on the interval $[w^l, w^u]$. Consequently, K should have the attributes specified above.

The next theorem is the main result of the paper. It demonstrates the existence of a unique Nash equilibrium for the model. It also characterizes the equilibrium bid distributions of firms I and O. Remarkably, the distributions F_I and F_O of wages offered by the two employers are identical to each other. Because the relation $\hat{h} = x(\hat{w}_I)$ holds almost surely for the bid selected $\hat{w}_I$ and amount invested $\hat{h}$ by firm I in equilibrium, the joint distribution G_I of training levels and wage offers chosen by firm I is known given the equilibrium distributions of wages offered by firms I and O. Recall that $x(w_I) = g'^{-1}[1/P_O(w_I)]$ for all $w_I \in \mathbb{R}_+$, where $P_O(w_I) = F_O(w_I)$ for all $w_I \in \mathbb{R}_+$ because the distribution F_O has no mass point.

Theorem 1 *There exists a unique Nash equilibrium profile of strategies (F_I, x) and F_O for firms I and O. For $Y \in \{I, O\}$, the distribution function F_Y is defined by:*

$$F_Y(w) = \begin{cases} 0 & \text{if } w \leq g(0) \\ x(w)/\{g[x(w)] - w\} & \text{if } g(0) < w < g[g'^{-1}(1)] - g'^{-1}(1) \\ 1 & \text{if } w \geq g[g'^{-1}(1)] - g'^{-1}(1) \end{cases} .$$

Suppose that the strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium. The following is an explanation for why F_O must be as described in the theorem. It was shown before that the bid $\hat{w}_I$ chosen by firm I fulfills the indifference condition $\pi_I[x(\hat{w}_I), \hat{w}_I] = 0$ with probability one. Given the properties of F_I and F_O, this condition translates into a requirement that $F_O(w)\{g[x(w)] - w\} - x(w) = 0$ holds on the interval $[w^l, w^u]$. Rearrangement of this expression gives $F_O(w) = x(w)/\{g[x(w)] - w\}$ for $w \in (w^l, w^u]$, where $F_O(w^l) = 0$ because F_O must be continuous. It can be further argued that there actually exists a unique distribution function satisfying the description in the theorem.

The next topic for discussion is why F_I has the same formula as F_O. As stated earlier, firm O almost surely offers a wage $\hat{w}_O$ that complies with the indifference condition $\pi_O(\hat{w}_O) = 0$. Given the characteristics of F_I and F_O, this condition can be rewritten as $\int_{w^l}^w g[x(v)] - w \, dF_I(v) + F_I(w^l)\{g[x(w^l)] - w\} = 0$ for $w \in [w^l, w^u]$. Differentiation of this equation yields $F_I'(w)\{g[x(w)] - w\} = F_I(w)$ on the interval (w^l, w^u), where the Leibniz integral rule is applied. Now recall the condition $F_O(w)\{g[x(w)] - w\} - x(w) = 0$ for $w \in [w^l, w^u]$. Differentiation of this identity with the help of the envelope theorem gives $F_O'(w)\{g[x(w)] - w\} = F_O(w)$ on the interval (w^l, w^u). That is, F_I and F_O satisfy the same differential

equation with the same boundary condition $F_I(w^u) = F_O(w^u) = 1$. This implies that the distribution functions F_I and F_O should be the same as each other due to the existence and uniqueness of the solution to a first-order linear differential equation.

Intuitively, the distribution functions F_I and F_O parallel each other because of an underlying symmetry between the payoff functions of firms I and O. When firm I contemplates a marginal increase in its wage offer w_I, it weighs the expected gain $F_O'(w_I)\{g[x(w_I)] - w_I\}$ in profits from a higher retention rate against the expected rise $F_O(w_I)$ in wage payments. Note that firm I chooses the training level $x(w_I)$ to maximize its expected payoff given its anticipated wage offer w_I. Hence, the envelope theorem implies that changes in $x(w_I)$ can be suppressed when calculating the effect of a marginal change in w_I on the expected payoff of firm I.

When firm O considers raising its wage offer w_O infinitesimally, it faces a tradeoff between the expected gain $F_I'(w_O)\{g[x(w_O)] - w_O\}$ in profits from a higher recruitment rate and the expected rise $F_I(w_O)$ in wage payments. Note that each wage w_I offered by firm I can be mapped to a unique training level $x(w_I)$. Hence, firm O can perfectly predict the productivity of the marginal worker potentially being recruited when it incrementally raises its wage offer. It can be seen that the problems faced by firms I and O are identical to each other on the margin. This suggests that both employers would offer the same distribution of wages in equilibrium, observing also that the supports for the bid distributions of the two firms should have the same supremum.

The following corollary provides an explicit solution for the equilibrium distribution K of training levels selected by firm I. In brief, the result follows from two attributes of an equilibrium. First, the wage $\hat{w}_I$ offered and training $\hat{h}$ provided by firm I satisfy the condition $\hat{h} = g'^{-1}[1/F_O(\hat{w}_I)]$ with probability one, where F_O is continuous and strictly increasing from 0 to 1 on the interval $[w^l, w^h]$. Second, the distributions F_I and F_O of wages offered by firms I and O are identical.

Corollary 1 *Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium. The distribution function K is given by:*

$$K(h) = 1/g'(h), \ \ 0 \leq h \leq g'^{-1}(1).$$

The properties of the production function determine the shape of the equilibrium distribution of human capital. Differentiating the cumulative distribution function in the corollary above, the probability density function of investment is $k(h) = -g''(h)/[g'(h)]^2$ for $h \in (h^l, h^u)$. Remember that $g'(h) > 0$ and $g''(h) < 0$ for all $h \in \mathbb{R}_+$ by assumption. Hence, if the second

derivative $g''(h)$ of the production function is nonincreasing in human capital h, then the density $k(h)$ of the training distribution will be increasing. If so, the distribution function $K(h)$ will be convex. In the case where $g''(h)$ is increasing in h, it is possible for $k(h)$ to be decreasing and for $K(h)$ to be concave.

It now remains to characterize the expected payoffs of the bidders and the seller. If (F_I, x) and F_O are equilibrium strategies for firms I and O, then $\hat{w}_I$ drawn from F_I satisfies $\pi_I[x(\hat{w}_I), \hat{w}_I] = 0$ with probability one, and $\hat{w}_O$ drawn from F_O satisfies $\pi_O(\hat{w}_O) = 0$ with probability one. That is, firms I and O each receive an expected payoff of zero in equilibrium. An expression for the expected payoff to the worker is stated in the corollary below. Let $\phi_S = \int_{w^l}^{w^u} \int_{w^l}^{w^u} \max(u, v) \, dF_I(u) dF_O(v)$ denote the expected payoff to the worker.

Corollary 2 *Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium. Then ϕ_S is given by:*

$$\phi_S = g[g'^{-1}(1)] - 2g'^{-1}(1) + \int_0^{g'^{-1}(1)} \frac{1}{g'(h)} \, dh.$$

The preceding formula is derived in two steps. First, the expected payoff ϕ_S to the worker is written as the difference between the expected value of output produced and the expected amount invested in training. This equality holds because the expected payoffs to firms I and O are zero in equilibrium. Second, the solution for K presented above is used to substitute for the cumulative distribution function of investment, and the resulting expression is simplified using integration by parts.

This section closes with some remarks about the possible robustness of the solution to changes in the setup of the model. It is straightforward to generalize the framework to markets with a larger number of participants. Obviously, if the market contains several independent workers, then each play of the game can be regarded as the outcome for a given worker. Furthermore, the solution does not differ materially when there is an arbitrary number $N \geq 2$ of bidders. If the worker receives wage offers from the incumbent employer as well as from more than one outside firm, then it can easily be seen that the model has an equilibrium of the following form. The incumbent employer bids according to the same distribution F_I as in the theorem above. Moreover, the wage offer $\hat{w}_I$ and training level $\hat{h}$ chosen by the incumbent employer continue to satisfy $\hat{h} = x(\hat{w}_I)$ with probability one. The distribution of the highest wage offered by an outside employer is F_O, which has the same formula as F_I in the theorem above.

Another possibility is that an individual firm cannot provide its employees with differential access to training. A specific firm may be constrained to choose a single training level for all of its workers. In this situation, the solution presented above may still hold if there is a continuum of agents with random matching of workers to firms. For example, suppose that there are continua of incumbent and outside firms. Every incumbent firm trains a certain number of workers, after which the incumbent firm is paired with a randomly selected outside firm, and both firms make simultaneous wage offers to the workers trained by the incumbent firm. Even if each employer is restricted to playing pure strategies, the equilibrium distributions of wage offers across incumbent and outside employers would be identical to those in the theorem above. Likewise, the mapping from wage offers to training levels and the resulting distribution of human capital would be as previously specified. That is, randomization would occur at the market scale instead of within a firm.

Nonetheless, some extensions of the model would alter the structure of the solution, although they may not eliminate the incentive for firms to provide general training by randomizing their actions. Perhaps, an outside employer can observe the skills of a worker by conducting an interview or administering a test before placing a bid. If so, the incumbent firm might not provide training as in the case where investment is observable but not contractible. However, such screening may be costly for employers to perform, in which case a time-inconsistency problem would prevent the existence of an equilibrium without employer-financed training. To see this, suppose otherwise that the incumbent employer never trains a worker in equilibrium. Then an outside firm would surmise any employee of the incumbent firm to be untrained and would not incur the cost for checking the skills of a worker. Consequently, the incumbent employer might have an incentive to deviate by secretly training a worker.

The solution would also change if the bidding game were extended beyond one period. In a model with one period of bidding as above, the wage offer of the incumbent employer would perfectly reveal the training level of a worker if the bid were observed after the game. However, if there are several periods of bidding, then an incumbent employer might have an incentive to offer the same wage to workers with different training levels, so as to conceal information about the productivity of a worker from the outside market. In addition, an employer might provide additional random increments of training between rounds of bidding, thereby ensuring that outside firms remain uncertain about the productivity of a worker. The solution derived in this section can be regarded as modeling the special case where all of the training provided by an employer depreciates after each period of bidding. Under such

an interpretation, each play of the game in this section would represent a particular round of investment and bidding.

5   Randomness in Training Outcomes

The model suggests that training should vary across workers at the same firm. A relevant question is whether this prediction is plausible given the empirical evidence. If so, how might such differences in skill provision arise in the workplace? Some economists have documented variability in training outcomes within a given establishment. For example, Lynch and Black (1998) examine the proportion of workers that receive training on the job. They observe that "while most firms provide training, relatively few workers appear to be getting it." Likewise, Liu and Batt (2007) find that different managers at the same company provide unequal levels of training to staff members. They also demonstrate that the effectiveness of training is related to the seniority of the trainer.

A mechanism that might generate disparities in training is the process of mentoring, whereby an experienced worker guides, advises, and sponsors a junior colleague. As noted by Laband and Lentz (1995), mentor-protégé relationships are common in some professions such as law. However, a number of studies have uncovered substantial dispersion in the quality of such interactions. Ragins, Cotton, and Miller (2000) note that "mentoring relationships fall along a continuum" with some situations being productive, others being detrimental, and many being ineffectual but not harmful. Scandura (1998) examines cases in which mentoring arrangements become dysfunctional despite their potential to provide both vocational and emotional support.

Furthermore, mentoring can confer benefits and impose costs not only on protégés but also on mentors. The benefits from being a mentor comprise personal satisfaction, professional recognition, and loyal support from trainees. The costs include time and energy spent on mentoring instead of other tasks, the risk of being replaced by a trainee, and reputational damage from an underperforming protégé. Ragins and Scandura (1999) conduct a survey illustrating how the perceived costs and benefits of being a mentor as well as overall preferences for becoming a mentor differ across corporate executives.

Indeed, the framework in this paper can be interpreted in terms of a story about mentoring. One version of the story is as follows. Suppose that mentors differ in their coaching and tutoring abilities. Let $s \in \mathbb{R}_+$ represent the teaching skills of a mentor. If a worker is assigned to a mentor with skill level s, then the firm incurs the cost s, and the output of the worker is $g(s)$, where the function g has the properties listed in assumption 1. In this

case, the human capital distribution K in corollary 1 would arise if a worker is assigned to a mentor whose skill level s is distributed according to K.

Note that the output of a worker is higher if he or she is assigned to a mentor with stronger teaching skills. Furthermore, assigning a worker to a mentor with stronger teaching skills is more costly for the firm. This assumption is reasonable if better instructors also have greater productivity in other roles. For example, more senior employees may be more effective at mentoring. However, they have a greater opportunity cost for serving as mentors instead of participating in alternative activities like meetings, speeches, and strategic planning.

The following is another variant of the story about mentoring. Suppose that mentors differ in their preferences for assisting and educating younger coworkers. These preferences might derive from the tradeoff between the costs and benefits of mentoring described above. Let $t \in \mathbb{R}_+$ represent the taste of a mentor for teaching. In particular, t is the amount of training that a mentor prefers to provide. If a worker is assigned to a mentor with taste level t, then the firm incurs the cost t, and the output of the worker is $g(t)$, where the function g has the properties listed in assumption 1. The cost of training might reflect the opportunity cost of time that a mentor could be spending on other productive tasks. The human capital distribution K in corollary 1 would arise if a worker is assigned to a mentor whose taste level t is distributed according to K.

6   Example with Constant Output Elasticity

This section presents a tractable parametric example to illustrate the main features of the solution to the model with unobservable investment. The production function is assumed to have the Cobb-Douglas form $g(h) = Ah^\theta$ for all $h \in \mathbb{R}_+$, where $\theta \in (0,1)$ is the elasticity of output with respect to human capital, and $A > 0$ represents other factors affecting worker productivity. Both the total amount of output and the marginal return to training at a given level of investment h are increasing in the efficiency A. Note that this specification is consistent with assumption 1. The equilibrium of the game is given below in closed form. The derivations are explained in the appendix.

Example 1 *Consider the game with an unobservable training level. Assume that the production function is given by $g(h) = Ah^\theta$ for all $h \in \mathbb{R}_+$, where $\theta \in (0,1)$ and $A > 0$. There exists a unique Nash equilibrium profile of strategies (F_I, x) and F_O for firms I and O. The*

cdf of the wage offer distributions for firms I and O is:

$$F_I(w) = F_O(w) = \left(\frac{w}{A^{\frac{1}{1-\theta}}(\theta^{\frac{\theta}{1-\theta}} - \theta^{\frac{1}{1-\theta}})}\right)^{\frac{1-\theta}{\theta}}, \ 0 \leq w \leq A^{\frac{1}{1-\theta}}(\theta^{\frac{\theta}{1-\theta}} - \theta^{\frac{1}{1-\theta}}),$$

and the function relating the wage offered to the training provided by firm I is:

$$x(w) = \left(\frac{w}{A(1-\theta)}\right)^{\frac{1}{\theta}}.$$

In equilibrium, the cdf of the human capital distribution is:

$$K(h) = \frac{h^{1-\theta}}{A\theta}, \ 0 \leq h \leq (A\theta)^{\frac{1}{1-\theta}},$$

and the expected payoff to the worker is:

$$\phi_S = \frac{2(1-\theta)^2(A\theta^\theta)^{\frac{1}{1-\theta}}}{2-\theta}.$$

The shape of the equilibrium bid distributions F_I and F_O depends critically on the elasticity θ. In particular, the density functions F_I' and F_O' of the wage offer distributions are increasing if $\theta < 1/2$, decreasing if $\theta > 1/2$, and uniform if $\theta = 1/2$. That is, the distribution of wages offered by each firm is left-skewed if $\theta < 1/2$, right-skewed if $\theta > 1/2$, and symmetric if $\theta = 1/2$. Note that the equilibrium distribution of wages received by the worker is given by $F_I \cdot F_O$ because the worker accepts the higher of the offers from firms I and O. Consequently, the density of wages received by the worker is increasing if $\theta < 2/3$, decreasing if $\theta > 2/3$, and uniform if $\theta = 2/3$. Overall, the distributions of amounts offered and offers accepted shift from being negatively to positively skewed as the elasticity of output with respect to training increases.

The parameter θ also influences the equilibrium training distribution K. However, the distribution function K is always concave, provided that the production technology has the above functional form. That is, the density function K' of the human capital distribution is decreasing, and so the distribution of human capital is skewed to the right. More generally, if the restriction on the functional form is removed, then the distribution function K may be convex as explained in the discussion following corollary 1.

The above expressions for the distributions of wage offers and human capital can be related to some common probability distributions in statistics and economics. Let the dis-

tribution functions F_I, F_O, and K be as specified in the example above. First, consider the random variables $\hat{w}/w^u$ and $w^u/\hat{w}$, where $\hat{w}$ is drawn from either distribution F_I or F_O, and w^u is the supremum of the support for both distributions F_I and F_O. It can easily be shown that $\hat{w}/w^u$ has a beta distribution with left parameter $(1-\theta)/\theta$ and right parameter 1 and that $w^u/\hat{w}$ has a Pareto distribution with scale parameter 1 and shape parameter $(1-\theta)/\theta$. Next, consider the random variables $\hat{h}/h^u$ and $h^u/\hat{h}$, where $\hat{h}$ is drawn from distribution K, and h^u is the supremum of the support for distribution K. It can similarly be shown that $\hat{h}/h^u$ has a beta distribution with left parameter $1-\theta$ and right parameter 1 and that $h^u/\hat{h}$ has a Pareto distribution with scale parameter 1 and shape parameter $1-\theta$.

Recall from the discussion preceding corollary 2 that the expected payoff to the worker is positive, whereas the expected payoffs to firms I and O are zero. In the example above, it is straightforward to show from the formula that the expected payoff ϕ_S to the worker has the following properties. If $A \leq 1$, then the expected payoff ϕ_S is decreasing in the elasticity θ, and ϕ_S approaches zero in the limit as θ goes to one. If $A > 1$, then the expected payoff ϕ_S is a non-monotonic function of the elasticity θ, and ϕ_S becomes infinite in the limit as θ tends to one. In this case, the relationship between ϕ_S and θ is U-shaped, with ϕ_S being first decreasing and then increasing in θ.

The model in this paper is consistent with some important stylized facts about training, turnover, and wages. Lynch (1992) and Parent (1999) find that training has a positive impact on wages. In addition, Parent (1999) infers that "firms tend to keep their trained workers longer as compared with other workers because they are more productive." Similarly, Loewenstein and Spletzer (1997) as well as Royalty (1996) identify a negative association between training and turnover. In the current framework, the wage offered by the incumbent employer is increasing in the training level. Consequently, the training level is positively related to the probability of the worker remaining with the incumbent employer.

The model can also match some empirical regularities concerning the shape of the wage and training distributions. According to Neal and Rosen (2000), one of the main properties of the wage distribution is that it is positively skewed with a long right tail. The example above possesses this feature provided that the elasticity parameter is sufficiently high. In addition, Frazis and Loewenstein (2005) observe that the "training distribution is quite skewed to the right," and Frazis and Spletzer (2005) document a similar pattern. Likewise, the example above generates a human capital distribution with a decreasing density.

7 Modeling Extensions and Robustness Checks

This section discusses some variations of the main model. Section 7.1 examines the case in which the training decision is dichotomous. Section 7.2 addresses the possibility of commitment by a worker to remain with an employer. Section 7.3 identifies a strategy that might enable the incumbent employer to obtain positive expected profits from training.

7.1 Binary Training

The main model assumes that the training level is a continuous instead of discrete choice. However, this feature is not critical for supporting an equilibrium with randomness in training. The framework can be modified as follows to permit only a binary choice of training level by the incumbent employer. Restrict firm I to choose between training levels $h = 0$ and $h = \kappa$. That is, the worker is either trained or untrained. The constant $\kappa > 0$ should satisfy $\kappa < g(\kappa) - g(0)$, in which case the training cost is less than the gain in productivity from training. All the other assumptions are the same as in the original version of the model with unobservable investment.

The proposition below demonstrates the existence of an equilibrium when the training decision is dichotomous. The associated distribution of human capital is nondegenerate. The uniqueness of the equilibrium can also be shown for the model with binary training. However, the result is omitted because the proof is lengthy and similar in style to the analysis of the main model.

Proposition 3 *The model with unobservable investment and binary training has a Nash equilibrium in which the incumbent employer trains the worker with probability $p = 1 - \kappa/[g(\kappa) - g(0)]$.*

7.2 Worker Commitment

In some situations, a worker might commit to remain with an employer in return for the opportunity to receive training. For example, Cappelli (2004) describes programs in which an employer covers the cost of attending college as long as a worker does not leave the firm too early, and Benson, Finegold, and Mohrman (2004) analyze how such tuition reimbursement policies affect the turnover of employees. There is a straightforward way to incorporate this possibility into the above model with binary training. At time $t = 0$, firm I chooses between training levels $h = 0$ and $h = \lambda$, where $\lambda > 0$. At time $t = 1/2$, the worker is constrained to

stay at firm I, and firm I receives an output worth $g(h)$ from the worker. For concreteness, assume that firm I pays the worker the wage $g(0)$ at time $t = 1/2$.

The events at time $t = 1$ are the same as in the original model with unobservable investment. That is, firms I and O make simultaneous wage offers to the worker, and the worker is free to accept employment from the firm offering the higher wage. The worker produces an output worth $g(h)$ for his or her employer at time $t = 1$. There is no discounting over time. This extension is equivalent to the model with only two levels of training and just one round of production. The commitment by the worker to remain with the incumbent employer simply serves as a reduction in the training cost. If firm I trains the worker, then it enjoys a profit of $g(\lambda) - g(0)$ at time $t = 1/2$, and so the cost of training is effectively decreased by this amount.

The solution to the extended model is as follows. For $\lambda \leq g(\lambda) - g(0)$, the commitment scheme fully covers the training cost, and so the incumbent employer trains the worker with probability one. For $\lambda \geq 2[g(\lambda) - g(0)]$, the cost of training is sufficiently large that there is no efficiency gain from training, and the incumbent employer trains the worker with probability zero. If $g(\lambda) - g(0) < \lambda < 2[g(\lambda) - g(0)]$, then the provision of training is random. In particular, there exists an equilibrium in which the incumbent employer trains the worker with probability $1 - \{\lambda - [g(\lambda) - g(0)]\}/[g(\lambda) - g(0)]$. This case matches the basic framework with binary training when the output of an untrained worker is worth $g(0)$, the output of a trained worker is worth $g(\lambda)$, and the cost of training is $\lambda - [g(\lambda) - g(0)]$.

Nevertheless, there are important legal and technical limitations on policies that bind a worker to an employer. First, common law normally allows a worker to leave a job at will. Because of protections against involuntary servitude, an employer cannot coerce an individual to work, but it may legally require a worker to repay the cost of training if he or she does not return to work. Thus, an employee may be unable to commit to staying with a firm unless credit constraints prevent him or her from compensating the employer for the cost of training. Second, the requirement for a worker to reimburse an employer for the cost of training may be enforceable only if the receipt of training is observable and contractible.

By contrast, this paper focuses on activities like mentoring and advising that are more likely to be unobservable and noncontractible. If training is unverifiable, then a worker can leave an employer and claim that training was not supplied. The courts may be reluctant to compel the employee to repay the firm because the provision of training by the firm cannot be confirmed. As noted by Rubin and Shedd (1981), an employee may nonetheless sign a restrictive covenant, which limits his or her employment options after leaving a job. For

example, a person might agree not to work for a direct competitor for a certain length of time after leaving an employer. However, the enforcement of such noncompetition clauses varies substantially across jurisdictions with some states prohibiting them almost entirely.

7.3 Positive Profits

A counterintuitive property of the equilibrium with randomized training is that the ability to train the worker does not confer any profits upon a firm. Although the incumbent employer derives an informational rent by training the worker, the rent is exactly offset by the cost of training required to generate the rent. Thus, the incumbent employer obtains an expected payoff of zero in equilibrium. This result is a consequence of the behavioral restrictions imposed by a Nash equilibrium. Consider the model with a binary choice of training level. In order for the incumbent employer to randomize the investment decision, it must be indifferent between supplying and withholding training. Since its expected profits are zero when the worker is untrained, its expected profits must also be zero when the worker is trained.

If this indifference condition is relaxed, then the incumbent employer can earn a positive expected payoff. The resulting strategies do not constitute a Nash equilibrium because the incumbent employer has an incentive to deviate by training the worker with probability one instead of randomizing the training decision. Nevertheless, the outcome might be reasonable if a firm can credibly commit to a particular training policy. The model with binary training is adjusted as follows. At time $t = 0$, the worker is trained with exogenous probability $q \in [0, 1]$. That is, firm I provides training level $h = 0$ with probability $1 - q$ and training level $h = \kappa$ with probability q, where $0 < \kappa < g(\kappa) - g(0)$. At time $t = 1/2$, firm I observes whether the worker is trained or untrained, but firm O does not learn the training status of the worker.

The events at time $t = 1$ are the same as in the original setup. Firms I and O place bids for the services of the worker, and his or her employer receives an output worth $g(h)$. Assume no discounting. The result below identifies the value of the training probability q that maximizes the expected profits of the incumbent employer. Since the distribution of values is exogenous in this version of the model, the bidding game at time $t = 1$ is analogous to that in Engelbrecht-Wiggans, Milgrom, and Weber (1983), who examine a first-price, sealed-bid auction with asymmetrically informed bidders. The proposition simply solves for the distribution of values that makes the bidding game most profitable to the better informed bidder given the cost of investment. The training probability is less than in the Nash equilibrium characterized above, and the incumbent employer obtains positive expected

profits.

Proposition 4 *Suppose that a Bayesian Nash equilibrium is played in the bidding game between firms I and O in the model with binary training. The exogenous training probability that maximizes the expected payoff to firm I is $q = \frac{1}{2}\{1 - \kappa/[g(\kappa) - g(0)]\}$. The maximum expected payoff to firm I is positive, and the expected payoff to firm O is zero.*

8   Conclusion

A model of common value auctions with unobservable investment has been presented to explain firm sponsorship for the general training of workers. The setup preserves the basic assumptions of perfect competition, except for the premise that training is observable and contractible. In equilibrium, firms play mixed strategies, randomizing the amount invested and the wage offered. These actions create an endogenous adverse selection problem that allows employers to recover some returns to training investments by compressing the wage structure. Unlike existing research on general training in labor markets with informational imperfections, the framework in this paper does not rely on risk aversion among workers, complementarities between training and ability, exogenous sources of asymmetric information, or match-specific heterogeneity across worker-firm pairs.

Despite its simplicity, the game produces a sophisticated pattern of equilibrium behavior. The training distribution has a positive density on the interval between zero investment and the efficient level. The incumbent employer and the outside market offer the same continuous distribution of wages, although only the bids of the incumbent employer can vary with the quantity invested. As a result, ex ante identical workers can have different outcomes in terms of human capital and labor income. Overall, firms receive zero expected profits in equilibrium, whereas the expected payoff to the worker is positive. A parametric example indicates that the wage distribution is positively skewed if the elasticity of output with respect to training is sufficiently high.

The following are some extensions of the model for future study. One possibility is to introduce variation among workers in characteristics like schooling, aptitude, and talent. If firms have symmetric information about these variables, then the analysis would not change substantially because such heterogeneity can simply be modeled as differences across workers in the applicable production function. However, asymmetric information among employers about worker attributes may be significantly more complicated to incorporate. Other potential variants are to include several rounds of bidding between firms and to enable

firms to screen workers for training. The firm specificity of human capital and interactions between innate and acquired skills may also be factors to explore.

References

ACEMOGLU, D. (1997): "Training and Innovation in an Imperfectly Competitive Market," *Review of Economic Studies*, 64(3), 445–464.

ACEMOGLU, D., AND J.-S. PISCHKE (1998): "Why Do Firms Train? Theory and Evidence," *Quarterly Journal of Economics*, 113(1), 79–119.

——— (1999a): "Beyond Becker: Training in Imperfect Labour Markets," *Economic Journal*, 109(453), F112–F142.

——— (1999b): "The Structure of Wages and Investment in General Training," *Journal of Political Economy*, 107(3), 539–572.

ACEMOGLU, D., AND R. SHIMER (2000): "Wage and Technology Dispersion," *Review of Economic Studies*, 67(4), 585–607.

ALCHIAN, A. A., AND H. DEMSETZ (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62(5), 777–795.

AUTOR, D. H. (2001): "Why Do Temporary Help Firms Provide Free General Skills Training?," *Quarterly Journal of Economics*, 116(4), 1409–1448.

BAKER, G. P., M. C. JENSEN, AND K. J. MURPHY (1987): "Compensation and Incentives: Practice vs. Theory," *Journal of Finance*, 43(3), 593–616.

BARRON, J. M., M. C. BERGER, AND D. A. BLACK (1999): "Do Workers Pay for On-the-Job Training?," *Journal of Human Resources*, 34(2), 235–252.

BECKER, G. (1964): *Human Capital*. University of Chicago Press, Chicago.

BECKER, G. S. (1962): "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy*, 70(5), 9–49.

BENSON, G. S., D. FINEGOLD, AND S. A. MOHRMAN (2004): "You Paid for the Skills, Now Keep Them: Tuition Reimbursement and Voluntary Turnover," *Academy of Management Journal*, 47(3), 315–331.

BISHOP, J. H. (1996): "What We Know About Employer-Provided Training: A Review of Literature," Mimeo, Cornell University, School of Industrial and Labor Relations.

BROWN, C. (1990): "Empirical Evidence on Private Training," in *Research in Labor Economics*, ed. by R. Ehrenberg, vol. 11, pp. 97–113. JAI Press.

BURDETT, K., AND K. L. JUDD (1983): "Equilibrium Price Dispersion," *Econometrica*, 51(4), 955–969.

BURDETT, K., AND D. T. MORTENSEN (1998): "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39(2), 257–273.

BUTTERS, G. R. (1977): "Equilibrium Distributions of Sales and Advertising Prices," *Review of Economic Studies*, 44(3), 465–491.

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

CHANG, C., AND Y. WANG (1996): "Human Capital Investment under Asymmetric Information: The Pigovian Conjecture Revisited," *Journal of Labor Economics*, 14(3), 505–519.

CHIANG, S.-H., AND S.-C. CHIANG (1990): "General Human Capital as a Shared Investment under Asymmetric Information," *Canadian Journal of Economics*, 23(1), 175–188.

ENGELBRECHT-WIGGANS, R., P. R. MILGROM, AND R. J. WEBER (1983): "Competitive Bidding and Proprietary Information," *Journal of Mathematical Economics*, 11(2), 161–169.

FRAZIS, H. J., AND M. A. LOEWENSTEIN (2005): "Reexamining the Returns to Training: Functional Form, Magnitude, and Interpretation," *Journal of Human Resources*, 40(2), 453–476.

FRAZIS, H. J., AND J. R. SPLETZER (2005): "Worker Training: What We've Learned from the NLSY79," *Monthly Labor Review*, 128(2), 48–58.

FU, C. (2011): "Training, Search and Wage Dispersion," *Review of Economic Dynamics*, 14(4), 650–666.

GIBBONS, R., AND L. F. KATZ (1991): "Layoffs and Lemons," *Journal of Labor Economics*, 9(4), 351–380.

GOLDLÜCKE, S., AND P. W. SCHMITZ (2014): "Investments as Signals of Outside Options," *Journal of Economic Theory*, 150, 683–708.

GONZÁLEZ, P. (2004): "Investment and Screening under Asymmetric Endogenous Information," *RAND Journal of Economics*, 35(3), 502–519.

GREENWALD, B. C. (1986): "Adverse Selection in the Labour Market," *Review of Economic Studies*, 53(3), 325–347.

GUL, F. (2001): "Unobservable Investment and the Hold-Up Problem," *Econometrica*, 69(2), 343–376.

HART, O., AND J. MOORE (1994): "A Theory of Debt Based on the Inalienability of Human Capital," *Quarterly Journal of Economics*, 109(4), 841–879.

HENDRICKS, K., AND R. H. PORTER (1988): "An Empirical Study of an Auction with Asymmetric Information," *American Economic Review*, 78(5), 865–883.

HERMALIN, B. E. (2013): "Unobserved Investment, Endogenous Quality, and Trade," *RAND Journal of Economics*, 44(1), 33–55.

HÖRNER, J., AND J. JAMISON (2008): "Sequential Common-Value Auctions with Asymmetrically Informed Bidders," *Review of Economic Studies*, 75(2), 475–498.

KANEMOTO, Y., AND W. B. MACLEOD (1992): "The Ratchet Effect and the Market for Secondhand Workers," *Journal of Labor Economics*, 10(1), 85–98.

KATZ, E., AND A. ZIDERMAN (1990): "Investment in General Training: The Role of Information and Labour Mobility," *Economic Journal*, 100(403), 1147–1158.

KAWAI, K. (2014): "Dynamic Market for Lemons with Endogenous Quality Choice by the Seller," *Games and Economic Behavior*, 84(1), 152–162.

LABAND, D. N., AND B. F. LENTZ (1995): "Workplace Mentoring in the Legal Profession," *Southern Economic Journal*, 61(3), 783–802.

LANG, K. (1991): "Persistent Wage Dispersion and Involuntary Unemployment," *Quarterly Journal of Economics*, 106(1), 181–202.

LAU, S. (2008): "Information and Bargaining in the Hold-Up Problem," *RAND Journal of Economics*, 39(1), 266–282.

LAZEAR, E. P. (2009): "Firm-Specific Human Capital: A Skill-Weights Approach," *Journal of Political Economy*, 117(5), 914–940.

LI, J. (2013): "Job Mobility, Wage Dispersion, and Technological Change: An Asymmetric Information Perspective," *European Economic Review*, 60(C), 105–126.

LIU, X., AND R. BATT (2007): "The Economic Pay-Offs to Informal Training: Evidence from Routine Service Work," *Industrial and Labor Relations Review*, 61(1), 75–89.

LOEWENSTEIN, M. A., AND J. R. SPLETZER (1997): "Delayed Formal On-the-Job Training," *Industrial and Labor Relations Review*, 51(1), 82–99.

——— (1999): "General and Specific Training: Evidence and Implications," *Journal of Human Resources*, 34(4), 710–733.

LYNCH, L. M. (1992): "Private-Sector Training and the Earnings of Young Workers," *American Economic Review*, 82(1), 299–312.

LYNCH, L. M., AND S. E. BLACK (1998): "Beyond the Incidence of Employer-Provided Training," *Industrial and Labor Relations Review*, 52(1), 64–81.

MILGROM, P., AND R. J. WEBER (1982a): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50(5), 1089–1122.

——— (1982b): "The Value of Information in a Sealed-Bid Auction," *Journal of Mathematical Economics*, 10(1), 105–114.

MINNS, C., AND P. WALLIS (2013): "The Price of Human Capital in a Pre-Industrial Economy: Premiums and Apprenticeship Contracts in 18th Century England," *Explorations in Economic History*, 50(3), 335–350.

MORTENSEN, D. (1998): "Equilibrium Unemployment with Wage Posting: Burdett-Mortensen Meet Pissarides," Mimeo, Northwestern University.

NEAL, D., AND S. ROSEN (2000): "Theories of the Distribution of Earnings," in *Handbook of Income Distribution*, ed. by A. B. Atkinson, and F. Bourguignon, vol. 1, pp. 379–427. North Holland.

PARENT, D. (1999): "Wages and Mobility: The Impact of Employer-Provided Training," *Journal of Labor Economics*, 17(2), 298–317.

PICCHIO, M., AND J. C. VAN OURS (2011): "Market Imperfections and Firm-Sponsored Training," *Labour Economics*, 18(5), 712–722.

PIGOU, A. (1912): *Wealth and Welfare*. Macmillan, London.

PRENDERGAST, C. (2002): "The Tenuous Trade-off between Risk and Incentives," *Journal of Political Economy*, 110(5), 1071–1102.

QUERCIOLI, E. (2005): "Training, Turnover, and Search," *International Economic Review*, 46(1), 133–143.

RAGINS, B. R., J. L. COTTON, AND J. S. MILLER (2000): "Marginal Mentoring: The Effects of Type of Mentor, Quality of Relationship, and Program Design on Work and Career Attitudes," *Academy of Management Journal*, 43(6), 1177–1194.

RAGINS, B. R., AND T. A. SCANDURA (1999): "Burden or Blessing? Expected Costs and Benefits of Being a Mentor," *Journal of Organizational Behavior*, 20(4), 493–509.

ROYALTY, A. B. (1996): "The Effects of Job Turnover on the Training of Men and Women," *Industrial and Labor Relations Review*, 49(3), 506–521.

RUBIN, P. H., AND P. SHEDD (1981): "Human Capital and Covenants Not to Compete," *Journal of Legal Studies*, 10(1), 93–110.

SCANDURA, T. A. (1998): "Dysfunctional Mentoring Relationships and Outcomes," *Journal of Management*, 24(3), 449–467.

STEVENS, M. (1994): "A Theoretical Model of On-the-Job Training with Imperfect Competition," *Oxford Economic Papers*, 46(4), 537–562.

VIRÁG, G. (2007): "Repeated Common Value Auctions with Asymmetric Bidders," *Games and Economic Behavior*, 61(1), 156–177.

WEVERBERGH, M. (1979): "Competitive Bidding with Asymmetric Information Reanalyzed," *Management Science*, 25(3), 291–294.

WILSON, R. B. (1969): "Competitive Bidding with Disparate Information," *Management Science*, 15(7), 446–448.

A Appendix

A.1 Proof of Proposition 1

The efficient training level h^e solves $\max_{h \in \mathbb{R}_+} g(h) - h$. The first-order condition is $g'(h^e) - 1 = 0$. Recall that $\lim_{h \downarrow 0} g'(h) = \infty$, $\lim_{h \uparrow \infty} g'(h) = 0$, and $g''(h) < 0$ for all $h \in \mathbb{R}_+$. Hence, the global maximizer is $h^e = g'^{-1}(1)$. ∎

A.2 Proof of Proposition 2

Suppose that the training level is $h \in \mathbb{R}_+$. The corresponding subgame involves Bertrand competition between firms I and O for a worker whose productivity is $g(h)$ at each firm. It is widely known and easily shown that the subgame has a unique Nash equilibrium, in which firms I and O both offer wage $w_O = w_I = g(h)$ with probability one. Hence, if firm I chooses training level $h \in \mathbb{R}_+$, then the expected payoff of firm I is $-h$, which is maximized when $h = 0$. It follows that the unique subgame perfect Nash equilibrium outcome is as described in the statement of the proposition. ∎

A.3 Proof of Lemma 1

Suppose that firm O plays mixed strategy F_O. If firm I chooses training level $h \in \mathbb{R}_+$ and wage offer $w_I \in \mathbb{R}_+$, then the expected payoff of firm I is $\pi_I(h, w_I) = P_O(w_I)[g(h) - w_I] - h$. A necessary condition for $(\hat{h}, \hat{w}_I) \in \mathbb{R}_+^2$ to maximize $\pi_I(h, w_I)$ is $\partial \pi_I(\hat{h}, \hat{w}_I)/\partial h = P_O(\hat{w}_I)g'(\hat{h}) - 1 = 0$, which gives $\hat{h} = g'^{-1}[1/P_O(\hat{w}_I)]$. Hence, a Nash equilibrium must have the property described in the statement of the lemma. ∎

A.4 Proof of Lemma 2

The lemma follows from the claims below. Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium.

Claim 1 *It must be that $w_I^u = w_O^u$.*

Proof Suppose that $w_I^u > w_O^u$. Choose any $\hat{w} \in (w_O^u, w_I^u]$. If firm I chooses training level $x(\hat{w})$ and wage offer $\hat{w}$, then the payoff to firm I is $\pi_I[x(\hat{w}), \hat{w}] = \{g[x(\hat{w})] - \hat{w}\} - x(\hat{w})$ with probability one. Choose any $\tilde{w} \in (w_O^u, \hat{w})$. If firm I chooses training level $x(\hat{w})$ and wage offer $\tilde{w}$, then the payoff to firm I is $\pi_I[x(\hat{w}), \tilde{w}] = \{g[x(\hat{w})] - \tilde{w}\} - x(\hat{w})$ with probability one. Because $\pi_I[x(\hat{w}), \tilde{w}] > \pi_I[x(\hat{w}), \hat{w}]$, there cannot be a Nash equilibrium in which $\hat{w}$ drawn

from distribution F_I satisfies $w_O^u < \hat{w} \le w_I^u$ with positive probability. This contradicts the fact that w_I^u is the supremum of the support of F_I. Hence, it must be that $w_I^u \le w_O^u$.

Suppose that $w_I^u < w_O^u$. Let $\mathbb{E}\{\cdot\}$ denote the expectation taken over w_I drawn from distribution F_I. Choose any $\hat{w} \in (w_I^u, w_O^u]$. If firm O offers the wage $\hat{w}$, then the expected payoff to firm O is $\pi_O(\hat{w}) = \mathbb{E}\{g[x(w_I)]\} - \hat{w}$. Choose any $\tilde{w} \in (w_I^u, \hat{w})$. If firm O offers the wage $\tilde{w}$, then the expected payoff to firm O is $\pi_O(\tilde{w}) = \mathbb{E}\{g[x(w_I)]\} - \tilde{w}$. Because $\pi_O(\tilde{w}) > \pi_O(\hat{w})$, there cannot be a Nash equilibrium in which $\hat{w}$ drawn from distribution F_O satisfies $w_I^u < \hat{w} \le w_O^u$. This contradicts the fact that w_O^u is the supremum of the support of F_O. Hence, it must be that $w_I^u \ge w_O^u$. ∎

Claim 2 *It must be that $w_I^l \le w_O^l$.*

Proof Suppose that $w_I^l > w_O^l$. Consider any wage offer $\hat{w} \in \mathbb{R}_+$ such that $\{g[x(\hat{w})] - \hat{w}\} - x(\hat{w}) < 0$. Firm I would obtain a negative expected payoff from choosing training level $x(\hat{w})$ and wage offer $\hat{w}$. Firm I could obtain a zero expected payoff from choosing training level 0 and wage offer $g(0)$. Hence, $\tilde{w}$ drawn from F_I must satisfy $\{g[x(\tilde{w})] - \tilde{w}\} - x(\tilde{w}) \ge 0$ with probability one. Because $w_I^l > w_O^l$, it must be that $P_O(w) > 0$ and $x(w) > 0$ for all $w \ge w_I^l$. Moreover, note that $P_O(w)$ and $x(w)$ are nondecreasing for $w \ge w_I^l$. Thus, $\tilde{w}$ drawn from F_I satisfies $g[x(\tilde{w})] - [\tilde{w} + x(w_I^l)] \ge 0$ with probability one. It follows that for any w such that $w_I^l \le w < w_I^l + x(w_I^l)$, $\tilde{w}$ drawn from F_I satisfies $g[x(\tilde{w})] - w > 0$ with probability one.

Choose any w^* satisfying $w_I^l < w^* < w_I^l + x(w_I^l)$ such that F_I does not have an atom at w^*. Let $\mathbb{E}\{g[x(w_I)] - w^* | w_I \le w^*\}$ denote the conditional expectation of $g[x(w_I)] - w^*$ given that $w_I \le w^*$, where w_I is drawn from F_I. If firm O offers the wage w^*, then the expected payoff to firm O is $\pi_O(w^*) = F_I(w^*)\mathbb{E}\{g[x(w_I)] - w^* | w_I \le w^*\}$. Because w_I drawn from F_I satisfies $g[x(w_I)] - w^* > 0$ with probability one, it must be that $\pi_O(w^*) > 0$. If firm O offers a wage $w^{**} \in [w_O^l, w_I^l)$, then the expected payoff to firm O is $\pi_O(w^{**}) = 0$. Because $\pi_O(w^*) > \pi_O(w^{**})$, there cannot be a Nash equilibrium in which w_O drawn from distribution F_O satisfies $w_O^l \le w_O < w_I^l$ with positive probability. This contradicts the fact that w_O^l is the infimum of the support of F_O. Hence, it must be that $w_I^l \le w_O^l$. ∎

Claim 3 *It must be that $w_I^l \ge w_O^l$.*

Proof Suppose that $w_I^l < w_O^l$. It must be that $P_O(w) = 0$ and so $x(w) = 0$ for all $w \in [w_I^l, w_O^l)$. If F_O does not have an atom at w_O^l, then $P_O(w_O^l) = 0$ and so $x(w_O^l) = 0$. If F_O and F_I each have an atom at w_O^l, then I argue in the next paragraph that $P_O(w_O^l) = 0$ and so $x(w_O^l) = 0$.

Suppose that F_O and F_I each have an atom at w_O^l. Assume that $P_O(w_O^l) > 0$ and so $x(w_O^l) > 0$. For any $w \in \mathbb{R}_+$, the expected payoff to firm I is $\pi_I[x(w), w] = \{F_O(w^-) + \alpha[F_O(w) - F_O(w^-)]\}\{g[x(w)] - w\} - x(w)$ if firm I chooses training level $x(w)$ and wage offer w. For any $w > w_I^l$, let $\mathbb{E}\{g[x(w_I)] - w|w_I \leq w\}$ denote the conditional expectation of $g[x(w_I)] - w$ given that $w_I \leq w$, where w_I is drawn from F_I. For any $w > w_I^l$, the expected payoff to firm O is $\pi_O(w) = F_I(w)\mathbb{E}\{g[x(w_I)] - w|w_I \leq w\} - \alpha[F_I(w) - F_I(w^-)]\{g[x(w)] - w\}$ if firm O offers wage w. Noting that F_I and F_O are right continuous, $\lim_{w \downarrow w_O^l} \pi_I[x(w), w] = F_O(w_O^l)\{g[x(w_O^l)] - w_O^l\} - x(w_O^l)$ and $\lim_{w \downarrow w_O^l} \pi_O(w) = F_I(w_O^l)\mathbb{E}\{g[x(w_I)] - w|w_I \leq w_O^l\}$. Because firm I can obtain a payoff of zero by choosing training level 0 and wage offer $g(0)$, it must be that $\pi_I[x(w_O^l), w_O^l] \geq 0$ in any Nash equilibrium. Since $P(w_O^l) > 0$ and $x(w_O^l) > 0$, it must be that $g[x(w_O^l)] - w_O^l > 0$. Hence, $\lim_{w \downarrow w_O^l} \pi_I[x(w), w] > \pi_I[x(w_O^l), w_O^l]$ if $\alpha \in [0, 1)$ and $\lim_{w \downarrow w_O^l} \pi_O(w) > \pi_O(w_O^l)$ if $\alpha \in (0, 1]$. This implies that there exists $\epsilon > 0$ such that firm I obtains a higher expected payoff by choosing training level $x(w_O^l + \epsilon)$ and wage offer $w_O^l + \epsilon$ than by choosing training level $x(w_O^l)$ and wage offer w_O^l or such that firm O obtains a higher expected payoff by offering wage $w_O^l + \epsilon$ than by offering wage w_O^l. It follows that there cannot be a Nash equilibrium in which $P_O(w_O^l) > 0$ and $x(w_O^l) > 0$ if F_O and F_I each have an atom at w_O^l.

I now argue that the assumption $w_I^l < w_O^l$ leads to a contradiction. First, consider the case where $w_O^l > g(0)$. For any $w > w_I^l$, the expected payoff to firm O is $\pi_O(w) = F_I(w)\mathbb{E}\{g[x(w_I)] - w|w_I \leq w\} - \alpha[F_I(w) - F_I(w^-)]\{g[x(w)] - w\}$ if firm O offers wage w. Note that $\pi_O(w_O^l) = \{F_I(w_O^{l-}) + (1 - \alpha)[F_I(w_O^l) - F_I(w_O^{l-})]\}[g(0) - w_O^l] < 0$ and that $\lim_{w \downarrow w_O^l} \pi_O(w) = F_I(w_O^l)[g(0) - w_O^l] < 0$. Hence, there exists $\epsilon > 0$ such that firm O obtains a negative expected payoff by choosing a wage $w \in [w_O^l, w_O^l + \epsilon)$. However, firm O can obtain a zero payoff with probability one by offering a wage $\tilde{w} = g(0)$. It follows that there cannot be a Nash equilibrium in which w_O drawn from distribution F_O satisfies $w_O^l \leq w_O < w_O^l + \epsilon$ with positive probability. This contradicts the fact that w_O^l is the infimum of the support of F_O. Hence, there cannot be a Nash equilibrium in which $w_I^l < w_O^l$ and $w_O^l > g(0)$.

Second, consider the case where $w_O^l = g(0)$. For any $w > w_I^l$, the expected payoff to firm O is $\pi_O(w) = F_I(w)\mathbb{E}\{g[x(w_I)] - w|w_I \leq w\} - \alpha[F_I(w) - F_I(w^-)]\{g[x(w)] - w\}$ if firm O offers wage w. Choose any $\tilde{w} \in (w_I^l, w_O^l)$. Note that $\pi_O(w_O^l) = \{F_I(w_O^{l-}) + (1 - \alpha)[F_I(w_O^l) - F_I(w_O^{l-})]\}[g(0) - w_O^l] = 0$, $\lim_{w \downarrow w_O^l} \pi_O(w) = F_I(w_O^l)[g(0) - w_O^l] = 0$, and $\pi_O(\tilde{w}) = \{F_I(\tilde{w}^-) + (1 - \alpha)[F_I(\tilde{w}) - F_I(\tilde{w}^-)]\}[g(0) - \tilde{w}] > 0$. Hence, there exists $\epsilon > 0$ such that $\pi_O(\tilde{w}) > \pi_O(w)$ for all $w \in [w_O^l, w_O^l + \epsilon)$. Because firm O obtains a higher expected payoff by offering wage $\tilde{w}$ than by offering any wage $w \in [w_O^l, w_O^l + \epsilon)$, there cannot be a

Nash equilibrium in which w_O drawn from distribution F_O satisfies $w_O^l \le w_O < w_O^l + \epsilon$ with positive probability. This contradicts the fact that w_O^l is the infimum of the support of F_O. Hence, there cannot be a Nash equilibrium in which $w_I^l < w_O^l$ and $w_O^l = g(0)$.

Third, consider the case where $w_O^l < g(0)$. If firm I chooses training level $x(\hat{w}) = 0$ and wage offer $\hat{w}$ satisfying $w_I^l \le \hat{w} < w_O^l$, then the payoff to firm I is $\pi_I[x(\hat{w}), \hat{w}] = 0$ with probability one. If firm I chooses training level 0 and wage offer $\tilde{w}$ satisfying $w_O^l < \tilde{w} < g(0)$, then the expected payoff to firm I is $\pi_I(0, \tilde{w}) = P_O(\tilde{w})[g(0) - \tilde{w}] > 0$. Because $\pi_I(0, \tilde{w}) > \pi_I[x(\hat{w}), \hat{w}]$, there cannot be a Nash equilibrium in which w_I drawn from distribution F_I satisfies $w_I^l \le w_I < w_O^l$ with positive probability. This contradicts the fact that w_I^l is the infimum of the support of F_I. Hence, there cannot be a Nash equilibrium in which $w_I^l < w_O^l$ and $w_O^l < g(0)$. ∎

A.5   Proof of Lemma 3

The lemma follows from the claims below. Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium.

Claim 4 *Assume that $P_O(w^l) = 0$. Then it must be that $w^l \le g(0)$.*

Proof Suppose that $w^l > g(0)$. Because $P_O(w^l) = 0$, it must be that $x(w^l) = 0$. I first argue that F_I does not have an atom at w^l. Suppose to the contrary that F_I has an atom at w^l. For any $w \ge w^l$, let $\mathbb{E}\{g[x(w_I)] - w | w_I \le w\}$ denote the conditional expectation of $g[x(w_I)] - w$ given that $w_I \le w$, where w_I is drawn from F_I. For any $w \ge w^l$, the expected payoff to firm O is $\pi_O(w) = F_I(w)\mathbb{E}\{g[x(w_I)] - w | w_I \le w\} - \alpha[F_I(w) - F_I(w^-)]\{g[x(w)] - w\}$ if firm O offers wage w. Note that $\pi_O(w^l) = (1 - \alpha)F_I(w^l)[g(0) - w^l] < 0$ and $\lim_{w \downarrow w^l} \pi_O(w) = F_I(w^l)[g(0) - w^l] < 0$. Hence, there exists $\epsilon > 0$ such that $\pi_O(w) < 0$ for all $w \in [w^l, w^l + \epsilon)$. If firm O offers wage $g(0)$, then firm O obtains a zero payoff with probability one. Because firm O obtains a higher payoff by offering wage $g(0)$ than by offering a wage $w \in [w^l, w^l + \epsilon)$, there cannot be a Nash equilibrium in which w_O drawn from F_O satisfies $w^l \le w_O < w^l + \epsilon$ with positive probability. This contradicts the fact that w^l is the infimum of the support of F_O. Hence, assuming that $P_O(w^l) = 0$, there cannot be a Nash equilibrium in which $w^l > g(0)$ and F_I has an atom at w^l.

I next argue that F_O has an atom at w^l. Suppose to the contrary that F_O does not have an atom at w^l. For any $w \in \mathbb{R}_+$, the expected payoff to firm I is $\pi_I[x(w), w] = P_O(w)\{g[x(w)] - w\} - x(w)$ if firm I chooses training level $x(w)$ and wage offer w. Note that $\lim_{w \downarrow w^l} P_O(w) = 0$, $\lim_{w \downarrow w^l} x(w) = 0$, $\lim_{w \downarrow w^l} g[x(w)] = g(0)$, and $\lim_{w \downarrow w^l}\{g[x(w)] - w\} = g(0) - w^l < 0$. Hence,

41

there exists $\epsilon > 0$ such that $g(0) - w^l < 0$ for all $w \in (w^l, w^l + \epsilon)$. Because $P_O(w) > 0$ and $x(w) > 0$ for all $w > w^l$, it must be that $P_O(w)\{g[x(w)] - w\} - x(w) < 0$ for $w \in (w^l, w^l + \epsilon)$. If firm I chooses training level 0 and wage offer $g(0)$, then firm I obtains a zero payoff with probability one. Because firm I obtains a higher payoff by choosing training level 0 and wage offer $g(0)$ than by choosing training level $x(w)$ and wage offer w satisfying $w \in (w^l, w^l + \epsilon)$, there cannot be a Nash equilibrium in which w_I drawn from F_I satisfies $w^l < w_I < w^l + \epsilon$ with positive probability. Moreover, since F_I does not have an atom at w^l, there cannot be a Nash equilibrium in which w_I drawn from F_I satisfies $w_I = w^l$ with positive probability. These results contradict the fact that w^l is the infimum of the support of F_I. Hence, assuming that $P_O(w^l) = 0$, there cannot be a Nash equilibrium in which $w^l > g(0)$ and F_O does not have an atom at w^l.

I now show that there is a contradiction if F_I does not have an atom at w^l and F_O has an atom at w^l. If firm I chooses training level 0 and wage offer $g(0)$, then firm I obtains a zero payoff with probability one. Hence, $\tilde{w}$ drawn from F_I must satisfy $\{g[x(\tilde{w})] - \tilde{w}\} - x(\tilde{w}) \geq 0$ with probability one. Noting that F_O has an atom at w^l, it must be that $P_O(w) \geq \lim_{w \downarrow w^l} P_O(w) > 0$ and $x(w) \geq \lim_{w \downarrow w^l} x(w) > 0$ for $w > w^l$. Noting that F_I does not have an atom at w^l, it must be that $\tilde{w}$ drawn from F_I satisfies $g[x(\tilde{w})] - [\tilde{w} + \lim_{w \downarrow w^l} x(w)] \geq 0$ with probability one. Hence, for any w satisfying $w^l \leq w < w^l + \lim_{w \downarrow w^l} x(w)$, it must be that $\tilde{w}$ drawn from F_I satisfies $g[x(\tilde{w})] - w > 0$ with probability one.

Choose any $\hat{w}$ satisfying $w^l < \hat{w} < w^l + \lim_{w \downarrow w^l} x(w)$ such that F_I does not have an atom at $\hat{w}$. Let $\mathbb{E}\{g[x(w_I)] - \hat{w}|w_I \leq \hat{w}\}$ denote the conditional expectation of $g[x(w_I)] - \hat{w}$ given that $w_I \leq \hat{w}$, where w_I is drawn from F_I. If firm O offers the wage $\hat{w}$, then the expected payoff to firm O is $\pi_O(\hat{w}) = F_I(\hat{w})\mathbb{E}\{g[x(w_I)] - \hat{w}|w_I \leq \hat{w}\}$. Because w_I drawn from F_I satisfies $g[x(w_I)] - \hat{w} > 0$ with probability one, it must be that $\pi_O(\hat{w}) > 0$. Noting that F_I does not have an atom at w^l, the payoff to firm O is $\pi_O(w^l) = 0$ with probability one if firm O offers the wage w^l. Because $\pi_O(\hat{w}) > \pi_O(w^l)$, there cannot be a Nash equilibrium in which F_O has an atom at w^l. This contradicts the result that F_O has an atom at w^l. Hence, assuming that $P_O(w^l) = 0$, it must be that $w_l \leq g(0)$.　∎

Claim 5 *Assume that $P_O(w^l) = 0$. Then it must be that $w^l \geq g(0)$.*

Proof Suppose that $w^l < g(0)$. Because $P_O(w^l) = 0$, it must be that $x(w^l) = 0$. If $P_O(w^l) = 0$, then either F_O does not have an atom at w^l or F_O has an atom at w^l but $\alpha = 0$. I show that there is a contradiction in each of these cases. It will then follow that $w^l \geq g(0)$ if $P_O(w^l) = 0$.

Consider the case where F_O does not have an atom at w^l. Choose any $\hat{w}$ satisfying $w^l < \hat{w} < g(0)$. If firm I chooses training level 0 and wage offer $\hat{w}$, then the expected payoff to firm I is $\pi_I(0, \hat{w}) = P_O(\hat{w})[g(0) - \hat{w}] > 0$. For any $w \in \mathbb{R}_+$, the expected payoff to firm I is $\pi_I[x(w), w] = P_O(w)\{g[x(w)] - w\} - x(w)$ if firm I chooses training level $x(w)$ and wage offer w. Note that $\pi_I[x(w^l), w^l] = 0$ and $\lim_{w \downarrow w^l} \pi_I[x(w), w] = 0$. Hence, there exists $\epsilon > 0$ such that $\pi_I[x(w), w] < \pi_I(0, \hat{w})$ for all $w \in [w^l, w^l + \epsilon)$. Because firm I obtains a higher expected payoff by choosing training level 0 and wage offer $\hat{w}$ than by choosing training level $x(w)$ and wage offer w satisfying $w^l \le w < w^l + \epsilon$, there cannot be a Nash equilibrium in which w_I drawn from F_I satisfies $w^l \le w_I < w^l + \epsilon$ with positive probability. This contradicts the fact that w^l is the infimum of the support of F_I. Hence, assuming that $P_O(w^l) = 0$, there cannot be a Nash equilibrium in which $w^l < g(0)$ and F_O does not have an atom at w^l.

Consider the case where F_O has an atom at w^l but $\alpha = 0$. I begin by arguing that F_I does not have an atom at w^l. Suppose to the contrary that F_I has an atom at w^l. If firm I chooses training level $x(w^l) = 0$ and wage offer w^l, then the payoff to firm I is $\pi_I[x(w^l), w^l] = 0$ with probability one. For any w satisfying $w^l < w < g(0)$, the expected payoff to firm I is $\pi_I(0, w) = P_O(w)[g(0) - w] > 0$ if firm I chooses training level 0 and wage offer w. Because $\pi_I(0, w) > \pi_I[x(w^l), w^l]$, there cannot be a Nash equilibrium in which F_I has an atom at w^l.

I now show that there is a contradiction if F_O has an atom at w^l and F_I does not have an atom at w^l. If firm O offers wage w^l, then the payoff to firm O is $\pi_O(w^l) = 0$ with probability one. Choose any $\tilde{w}$ satisfying $w^l < \tilde{w} < g(0)$ such that F_I does not have an atom at $\tilde{w}$. Let $\mathbb{E}\{g[x(w_I)] - w | w_I \le \tilde{w}\}$ denote the conditional expectation of $g[x(w_I)] - \tilde{w}$ given that $w_I \le \tilde{w}$, where w_I is drawn from F_I. If firm O offers wage $\tilde{w}$, then the expected payoff to firm O is $\pi_O(\tilde{w}) = \mathbb{E}\{g[x(w_I)] - \tilde{w} | w_I \le \tilde{w}\} \ge F(\tilde{w})[g(0) - \tilde{w}] > 0$. Because firm O obtains a higher expected payoff by offering wage $\tilde{w}$ than by offering wage w^l, there cannot be a Nash equilibrium in which w_O drawn from F_O satisfies $w_O = w^l$ with positive probability. This contradicts the fact that F_O has an atom at w^l. Hence, assuming that $P_O(w^l) = 0$, there cannot be a Nash equilibrium in which $w^l < g(0)$ and F_O has an atom at w^l but $\alpha = 0$. ■

Claim 6 *It must be that $P_O(w^l) = 0$.*

Proof Suppose that $P_O(w^l) > 0$. In this case, F_O must have an atom at w^l and $\alpha > 0$. If firm I chooses training level 0 and wage offer $g(0)$, then firm I obtains a zero payoff with probability one. Hence, $\tilde{w}$ drawn from F_I must satisfy $\{g[x(\tilde{w})] - \tilde{w}\} - x(\tilde{w}) \ge 0$ with probability one. Note that $P_O(w) > 0$ and $x(w) > 0$ for all $w \ge w^l$ and that $P_O(w)$ and $x(w)$ are nondecreasing for $w \ge w^l$. Thus, $\tilde{w}$ drawn from F_I satisfies $g[x(\tilde{w})] - [\tilde{w} + x(w^l)] \ge 0$

43

with probability one. It follows that for any w such that $w^l \leq w < w^l + x(w^l)$, $\tilde{w}$ drawn from F_I satisfies $g[x(\tilde{w})] - w > 0$ with probability one.

For any $w > w^l$, let $\mathbb{E}\{g[x(w_I)] - w | w_I \leq w\}$ denote the conditional expectation of $g[x(w_I)] - w$ given that $w_I \leq w$, where w_I is drawn from F_I. For any $w > w^l$, the expected payoff to firm O is $\pi_O(w) = F_I(w)\mathbb{E}\{g[x(w_I)] - w | w_I \leq w\} - \alpha[F_I(w) - F_I(w^-)]\{g[x(w)] - w\}$ if firm O offers wage w. There are two cases, one in which F_I does not have an atom at w^l, and one in which F_I has an atom at w^l. I argue that there is a contradiction in each case. It will then follow that $P_O(w^l) = 0$.

Consider the case where F_I does not have an atom at w^l. If firm O offers wage w^l, then firm O obtains a zero payoff with probability one. Choose any $\hat{w}$ satisfying $w^l < \hat{w} < w^l + x(w^l)$ such that F_I does not have an atom at $\hat{w}$. Noting that w_I drawn from F_I satisfies $g[x(w_I)] - \hat{w} > 0$ with probability one, the expected payoff to firm O is $\pi_O(\hat{w}) = F_I(\hat{w})\mathbb{E}\{g[x(w_I)] - \hat{w} | w_I \leq \hat{w}\} > 0$ if firm O offers wage $\hat{w}$. Because firm O obtains a higher payoff by offering wage $\hat{w}$ than by offering wage w^l, there cannot be a Nash equilibrium in which firm O offers wage w^l with positive probability. This contradicts the fact that F_O has an atom at w^l. Hence, there cannot be a Nash equilibrium in which $P_O(w^l) > 0$ and F_I does not have an atom at w^l.

Consider the case where F_I has an atom at w^l. If firm O offers wage w^l, then the expected payoff to firm O is $\pi_O(w^l) = (1 - \alpha)[F_I(w^l) - F_I(w^{l-})]\{g[x(w^l)] - w^l\}$. Note that $\lim_{w \downarrow w^l} \pi_O(w) = [F_I(w^l) - F_I(w^{l-})]\{g[x(w^l)] - w^l\}$. It follows from $g[x(w^l)] - w^l > 0$, $F_I(w^l) - F_I(w^{l-}) > 0$, and $\alpha > 0$ that $\lim_{w \downarrow w^l} \pi_O(w) > \pi_O(w^l)$. This implies that there exists $\epsilon > 0$ such that firm O gets a higher expected payoff by offering wage $w^l + \epsilon$ than by offering wage w^l. It follows that there cannot be a Nash equilibrium in which firm O offers wage w^l with positive probability. This contradicts the fact that F_O has an atom at w^l. Hence, there cannot be a Nash equilibrium in which $P_O(w^l) > 0$ and F_I has an atom at w^l. ∎

Claim 7 *It must be that F_O does not have an atom at w^l.*

Proof Suppose to the contrary that F_O has an atom at w^l. There are two cases, one in which F_I does not have an atom at w^l, and one in which F_I has an atom at w^l. I argue that there is a contradiction in each case. It will then follow that F_O does not have an atom at w^l.

Consider the case where F_I does not have an atom at w^l. If firm I chooses training level 0 and wage offer $g(0)$, then firm I obtains a zero payoff with probability one. Hence, $\tilde{w}$ drawn from F_I must satisfy $\{g[x(\tilde{w})] - \tilde{w}\} - x(\tilde{w}) \geq 0$ with probability one. Note that

$P_O(w) > 0$ and $x(w) > 0$ for all $w > w^l$ and that $P_O(w)$ and $x(w)$ are nondecreasing for $w > w^l$. Thus, $\tilde{w}$ drawn from F_I satisfies $g[x(\tilde{w})] - [\tilde{w} + \lim_{w \downarrow w^l} x(w)] \geq 0$ with probability one. It follows that for any w such that $w^l \leq w < w^l + \lim_{w \downarrow w^l} x(w)$, $\tilde{w}$ drawn from F_I satisfies $g[x(\tilde{w})] - w > 0$ with probability one.

For any $w > w^l$, let $\mathbb{E}\{g[x(w_I)] - w | w_I \leq w\}$ denote the conditional expectation of $g[x(w_I)] - w$ given that $w_I \leq w$, where w_I is drawn from F_I. For any $w > w^l$, the expected payoff to firm O is $\pi_O(w) = F_I(w)\mathbb{E}\{g[x(w_I)] - w | w_I \leq w\} - \alpha[F_I(w) - F_I(w^-)]\{g[x(w)] - w\}$ if firm O offers wage w. If firm O offers wage w^l, then firm O obtains a zero payoff with probability one.

Choose any $\hat{w}$ satisfying $w^l < \hat{w} < w^l + \lim_{w \downarrow w^l} x(w)$ such that F_I does not have an atom at $\hat{w}$. Noting that w_I drawn from F_I satisfies $g[x(w_I)] - \hat{w} > 0$ with probability one, the expected payoff to firm O is $\pi_O(\hat{w}) = F_I(\hat{w})\mathbb{E}\{g[x(w_I)] - \hat{w} | w_I \leq \hat{w}\} > 0$ if firm O offers wage $\hat{w}$. Because firm O obtains a higher payoff by offering wage $\hat{w}$ than by offering wage w^l, there cannot be a Nash equilibrium in which firm O offers wage w^l with positive probability. This contradicts the fact that F_O has an atom at w^l. Hence, there cannot be a Nash equilibrium in which F_O but not F_I has an atom at w^l.

Consider the case where F_I has an atom at w^l. It follows from $P_O(w^l) = 0$ that $x(w^l) = 0$. For any $(h, w) \in \mathbb{R}_+^2$, the expected payoff to firm I is $\pi_I(h, w) = P_O(w)[g(h) - w] - h$ if firm I chooses training level h and wage offer w. If firm I chooses training level $x(w^l) = 0$ and wage offer $w^l = g(0)$, then the payoff to firm I is zero with probability one. Note that $\lim_{w \downarrow w^l} \pi_I(h, w) = F_O(w^l)[g(h) - g(0)] - h$, where $F_O(w^l) > 0$ because F_O is assumed to have an atom at w^l. The statement $\lim_{w \downarrow w^l} \pi_I(h, w) > 0$ is equivalent to $g(h) - g(0) > h/F_O(w^l)$. Because g is a differentiable concave function, it must be that $hg'(h) < g(h) - g(0)$ for all $h > 0$. Hence, if $h > 0$ is such that $hg'(h) > h/F_O(w^l)$, then $\lim_{w \downarrow w^l} \pi_I(h, w) > 0$. It follows that $\lim_{w \downarrow w^l} \pi_I(h, w) > 0$ whenever $0 < h < g'^{-1}[1/F_O(w^l)]$, where $g'^{-1}[1/F_O(w^l)] > 0$ is well defined because $\lim_{h \downarrow 0} g'(h) = \infty$, $\lim_{h \uparrow \infty} g'(h) = 0$, and $g''(h) < 0$ for all $h \in \mathbb{R}_+$.

Choose any $\hat{h}$ satisfying $0 < \hat{h} < g'^{-1}[1/F_O(w^l)]$. Since $\lim_{w \downarrow w^l} \pi_I(\hat{h}, w) > 0$, there exists $\epsilon > 0$ such that $\pi_I(\hat{h}, w^l + \epsilon) > 0$. Because firm I obtains a higher expected payoff by choosing training level $\hat{h}$ and wage offer $w^l + \epsilon$ than by choosing training level $x(w^l)$ and wage offer w^l, there cannot be a Nash equilibrium in which firm I offers wage w^l with positive probability. This contradicts the fact that F_I has an atom at w^l. Hence, there cannot be a Nash equilibrium in which F_O and F_I each have an atom at w^l. ∎

A.6 Proof of Lemma 4

The lemma follows from the claims below. Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium.

Claim 8 *There is probability one that $\hat{w}_I$ drawn from F_I satisfies $\pi_I[x(\hat{w}_I), \hat{w}_I] = 0$.*

Proof If firm I chooses training level 0 and wage offer $g(0)$, then firm I obtains a zero payoff with probability one. Hence, $\hat{w}_I$ drawn from F_I must satisfy $\pi_I[x(\hat{w}_I), \hat{w}_I] \geq 0$ with probability one in any Nash equilibrium. Suppose now that $\hat{w}_I$ drawn from F_I satisfies $\pi_I[x(\hat{w}_I), \hat{w}_I] > 0$ with positive probability. Choose any $\tilde{w} \in \mathbb{R}_+$ such that $\pi_I[x(\tilde{w}), \tilde{w}] > 0$.

Consider first the case where F_I has an atom at w^l. Recall that $P_O(w^l) = 0$. If firm I chooses training level $x(w^l) = 0$ and wage offer $w^l = g(0)$, then the expected payoff to firm I is $\pi_I[x(w^l), w^l] = 0$. However, the expected payoff to firm I is $\pi_I[x(\tilde{w}), \tilde{w}] > 0$ if firm I chooses training level $x(\tilde{w})$ and wage offer $\tilde{w}$. Consequently, there cannot be a Nash equilibrium in which $\hat{w}_I$ drawn from F_I satisfies $\hat{w}_I = w^l$ with positive probability. This contradicts the fact that F_I has an atom at w^l. Thus, $\hat{w}_I$ drawn from F_I must satisfy $\pi_I[x(\hat{w}_I), \hat{w}_I] = 0$ with probability one in any Nash equilibrium such that F_I has an atom at w^l.

Consider next the case where F_I does not have an atom at w^l. Recall that F_O does not have an atom at w^l. Hence, $\lim_{w \downarrow w^l} P_O(w) = P_O(w^l) = 0$ and $\lim_{w \downarrow w^l} \pi_I[x(w), w] = \pi_I[x(w^l), w^l] = 0$, noting that F_O is a right continuous function. This implies that there exists $\epsilon > 0$ such that $\pi_I[x(w), w] < \pi_I[x(\tilde{w}), \tilde{w}]$ for all $w \in [w^l, w^l + \epsilon)$. Because firm I obtains a higher expected payoff by choosing training level $x(\tilde{w})$ and wage offer $\tilde{w}$ than by choosing any training level $x(w)$ and wage offer w such that $w \in [w^l, w^l + \epsilon)$, there cannot be a Nash equilibrium in which $\hat{w}_I$ drawn from F_I satisfies $\hat{w}_I \in [w^l, w^l + \epsilon)$ with positive probability. This contradicts the fact that w^l is the infimum of the support of F_I. Thus, $\hat{w}_I$ drawn from F_I must satisfy $\pi_I[x(\hat{w}_I), \hat{w}_I] = 0$ with probability one in any Nash equilibrium such that F_I does not have an atom at w^l. ∎

Claim 9 *There is probability one that $\hat{w}_O$ drawn from F_O satisfies $\pi_O(\hat{w}_O) = 0$.*

Proof Recall that $P_O(w^l) = 0$ and so $x(w^l) = 0$. If firm O offers the wage $w^l = g(0)$, then firm O obtains a zero payoff with probability one. Hence, $\hat{w}_O$ drawn from F_O must satisfy $\pi_O(\hat{w}_O) \geq 0$ with probability one in any Nash equilibrium. Suppose now that $\hat{w}_O$ drawn from F_O satisfies $\pi_O(\hat{w}_O) > 0$ with positive probability. Choose any $\tilde{w} \in \mathbb{R}_+$ such that $\pi_O(\tilde{w}) > 0$.

Recall that F_O does not have an atom at w^l. Hence, $\lim_{w \downarrow w^l} P_O(w) = P_O(w^l) = 0$, $\lim_{w \downarrow w^l} x(w) = x(w^l) = 0$, and $\lim_{w \downarrow w^l} \pi_O(w) = \pi_O(w^l) = 0$, noting that F_O is a right continuous function. This implies that there exists $\epsilon > 0$ such that $\pi_O(w) < \pi_O(\tilde{w})$ for all $w \in [w^l, w^l + \epsilon)$. Because firm O obtains a higher expected payoff by offering the wage $\tilde{w}$ than by offering any wage $w \in [w^l, w^l + \epsilon)$, there cannot be a Nash equilibrium in which $\hat{w}_O$ drawn from F_O satisfies $\hat{w}_O \in [w^l, w^l + \epsilon)$ with positive probability. This contradicts the fact that w^l is the infimum of the support of F_O. Thus, $\hat{w}_O$ drawn from F_O must satisfy $\pi_O(\hat{w}_O) = 0$ with probability one in any Nash equilibrium. ∎

A.7 Proof of Lemma 5

The lemma follows from the claims below. Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium.

Claim 10 *It must be that $w^u \le g[g'^{-1}(1)] - g'^{-1}(1)$.*

Proof Suppose to the contrary that $w^u > g[g'^{-1}(1)] - g'^{-1}(1)$. Choose any wage $\hat{w} > g[g'^{-1}(1)] - g'^{-1}(1)$. Note that $P_O(\hat{w}) > 0$ because $\hat{w} > g[g'^{-1}(1)] - g'^{-1}(1) > g(0) = w^l$. If firm I chooses training level $x(\hat{w})$ and wage offer $\hat{w}$, then the expected payoff to firm I is $\pi_I[x(\hat{w}), \hat{w}] = P_O(\hat{w})\{g[x(\hat{w})] - \hat{w}\} - x(\hat{w})$, where $x(\hat{w}) = g'^{-1}[1/P_O(\hat{w})]$ by definition. Observe that $\pi_I[x(\hat{w}), \hat{w}]$ satisfies:

$$\pi_I[x(\hat{w}), \hat{w}] = P_O(\hat{w}) \cdot g\{g'^{-1}[1/P_O(\hat{w})]\} - P_O(\hat{w}) \cdot \hat{w} - g'^{-1}[1/P_O(\hat{w})]$$
$$< P_O(\hat{w}) \cdot g\{g'^{-1}[1/P_O(\hat{w})]\} - P_O(\hat{w}) \cdot \left(g\{g'^{-1}[1/P_O(\hat{w})]\} - g'^{-1}[1/P_O(\hat{w})]\right) - g'^{-1}[1/P_O(\hat{w})],$$
$$= P_O(\hat{w}) \cdot g'^{-1}[1/P_O(\hat{w})] - g'^{-1}[1/P_O(\hat{w})] \le 0$$

where the strict inequality holds because $\hat{w} > g[g'^{-1}(1)] - g'^{-1}(1)$ and $g\{g'^{-1}[1/P_O(\hat{w})]\} - g'^{-1}[1/P_O(\hat{w})] \le g[g'^{-1}(1)] - g'^{-1}(1)$.

Hence, the expected payoff to firm I is negative if firm I chooses training level $x(\hat{w})$ and wage offer $\hat{w}$, where $\hat{w} > g[g'^{-1}(1)] - g'^{-1}(1)$. However, if firm I chooses training level 0 and wage offer $g(0)$, then firm I obtains a zero payoff with probability one. Consequently, there cannot be a Nash equilibrium in which $\tilde{w}$ drawn from F_I satisfies $\tilde{w} > g[g'^{-1}(1)] - g'^{-1}(1)$ with positive probability. This contradicts the fact that $w^u > g[g'^{-1}(1)] - g'^{-1}(1)$, where w^u is the supremum of the support of F_I. Thus, it must be that $w^u \le g[g'^{-1}(1)] - g'^{-1}(1)$ in any Nash equilibrium. ∎

Claim 11 *It must be that $w^u \ge g[g'^{-1}(1)] - g'^{-1}(1)$.*

Proof Suppose to the contrary that $w^u < g[g'^{-1}(1)] - g'^{-1}(1)$. Choose any wage $\hat{w}$ satisfying $w^u < \hat{w} < g[g'^{-1}(1)] - g'^{-1}(1)$. Note that $P_O(\hat{w}) = 1$ because $\hat{w} > w^u$. Hence, $x(\hat{w}) = g'^{-1}[1/P_O(\hat{w})] = g'^{-1}(1)$. If firm I chooses training level $x(\hat{w}) = g'^{-1}(1)$ and wage offer $\hat{w}$, then the expected payoff to firm I is given by:

$$\pi_I[x(\hat{w}), \hat{w}] = P_O(\hat{w})\{g[x(\hat{w})] - \hat{w}\} - x(\hat{w}) = \{g[g'^{-1}(1)] - \hat{w}\} - g'^{-1}(1) = \{g[g'^{-1}(1)] - g'^{-1}(1)\} - \hat{w}.$$

Observe that $\pi_I[x(\hat{w}), \hat{w}] > 0$ because $\hat{w} < g[g'^{-1}(1)] - g'^{-1}(1)$.

As shown above, there is probability one in any Nash equilibrium that $\tilde{w}$ drawn from F_I satisfies $\pi_I[x(\tilde{w}), \tilde{w}] = 0$. By definition, there is probability one in any Nash equilibrium that $\tilde{w}$ drawn from F_I satisfies $\pi_I[x(\tilde{w}), \tilde{w}] \geq \pi_I[x(w), w]$ for all $w \in \mathbb{R}_+$. It follows that $\pi_I[x(w), w] \leq 0$ for all $w \in \mathbb{R}_+$. However, this contradicts the fact that $\pi_I[x(\hat{w}), \hat{w}] > 0$. Thus, it must be that $w^u \geq g[g'^{-1}(1)] - g'^{-1}(1)$ in any Nash equilibrium. ∎

A.8 Proof of Lemma 6

The lemma follows from the claims below. Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium.

Claim 12 *It must be that F_I does not have an atom at any $w > w^l$.*

Proof Suppose to the contrary that there exists $\hat{w} > w^l$ such that F_I has an atom at $\hat{w}$. I begin by arguing that for any $\epsilon > 0$, there is positive probability that w drawn from F_O satisfies $w \in [\hat{w} - \epsilon, \hat{w}]$.

Suppose that there exists $\epsilon > 0$ such that w drawn from F_O satisfies $w < \hat{w} - \epsilon$ or $w > \hat{w}$ with probability one. If firm I chooses training level $x(\hat{w})$ and wage offer $\hat{w}$, then the expected payoff to firm I is given by:

$$\pi_I[x(\hat{w}), \hat{w}] = P_O(\hat{w})\{g[x(\hat{w})] - \hat{w}\} - x(\hat{w}).$$

Because $\hat{w} > w^l$, it must be that $P_O(\hat{w}) > 0$. Choose any $\tilde{w} \in (\hat{w} - \epsilon, \hat{w})$. If firm I chooses training level $x(\tilde{w})$ and wage offer $\tilde{w}$, then the expected payoff to firm I is given by:

$$\begin{aligned}
\pi_I[x(\tilde{w}), \tilde{w}] &= P_O(\tilde{w})\{g[x(\tilde{w})] - \tilde{w}\} - x(\tilde{w}) = P_O(\hat{w})\{g[x(\hat{w})] - \tilde{w}\} - x(\hat{w}) \\
&> P_O(\hat{w})\{g[x(\hat{w})] - \hat{w}\} - x(\hat{w}) = \pi_I[x(\hat{w}), \hat{w}],
\end{aligned}$$

where the second equality follows because $P_O(\tilde{w}) = P_O(\hat{w})$ and so $x(\tilde{w}) = x(\hat{w})$. Because firm I obtains a higher expected payoff by choosing training level $x(\tilde{w})$ and wage offer $\tilde{w}$

than by choosing training level $x(\hat{w})$ and wage offer $\hat{w}$, there cannot be a Nash equilibrium in which w drawn from F_I satisfies $w = \hat{w}$ with positive probability. This contradicts the fact that F_I has an atom at $\hat{w}$. It follows that for any $\epsilon > 0$, there is positive probability that w drawn from F_O satisfies $w \in [\hat{w} - \epsilon, \hat{w}]$.

Consider first the case where F_O has an atom at $\hat{w}$ and $\alpha > 0$. If firm O offers the wage $\hat{w}$, then the expected payoff to firm O is given by:

$$\pi_O(\hat{w}) = F_I(\hat{w})\mathbb{E}\{g[x(w_I)] - \hat{w}|w_I \le \hat{w}\} - \alpha[F_I(\hat{w}) - F_I(\hat{w}^-)]\{g[x(\hat{w})] - \hat{w}\}.$$

Because F_I has an atom at $\hat{w}$, it must be that $F_I(\hat{w}) - F_I(\hat{w}^-) > 0$. In addition, $g[x(\hat{w})] - \hat{w} > 0$, noting that $P_O(\hat{w}) > 0$, $x(\hat{w}) > 0$, and $\pi_I(\hat{w}) \ge 0$ in a Nash equilibrium. Observe that:

$$\lim_{w \downarrow \hat{w}} \pi_O(w) = F_I(\hat{w})\mathbb{E}\{g[x(w_I)] - \hat{w}|w_I \le \hat{w}\}$$
$$> F_I(\hat{w})\mathbb{E}\{g[x(w_I)] - \hat{w}|w_I \le \hat{w}\} - \alpha[F_I(\hat{w}) - F_I(\hat{w}^-)]\{g[x(\hat{w})] - \hat{w}\} = \pi_O(\hat{w}).$$

Hence, there exists $\eta > 0$ such that firm O obtains a higher expected payoff by offering the wage $\hat{w} + \eta$ than by offering the wage $\hat{w}$. It follows that there cannot be a Nash equilibrium in which w drawn from F_O satisfies $w = \hat{w}$ with positive probability. This contradicts the fact that F_O has an atom at $\hat{w}$. Thus, if $\alpha > 0$, then there cannot be a Nash equilibrium in which F_I and F_O both have an atom at $\hat{w}$.

Consider next the case where F_O has an atom at $\hat{w}$ and $\alpha = 0$. If firm I chooses training level $x(\hat{w})$ and wage offer $\hat{w}$, then the expected payoff to firm I is given by:

$$\pi_I[x(\hat{w}), \hat{w}] = F_O(\hat{w}^-)\{g[x(\hat{w})] - \hat{w}\} - x(\hat{w}).$$

Because F_O has an atom at $\hat{w}$, it must be that $F_O(\hat{w}^-) < F_O(\hat{w})$. In addition, $g[x(\hat{w})] - \hat{w} > 0$, noting that $P_O(\hat{w}) > 0$, $x(\hat{w}) > 0$, and $\pi_I(\hat{w}) \ge 0$ in a Nash equilibrium. Observe that:

$$\lim_{w \downarrow \hat{w}} \pi_I[x(\hat{w}), w] = F_O(\hat{w})\{g[x(\hat{w})] - \hat{w}\} - x(\hat{w})$$
$$> F_O(\hat{w}^-)\{g[x(\hat{w})] - \hat{w}\} - x(\hat{w}) = \pi_I[x(\hat{w}), \hat{w}].$$

Hence, there exists $\eta > 0$ such that firm I obtains a higher expected payoff by choosing training level $x(\hat{w})$ and wage offer $\hat{w} + \eta$ than by choosing training level $x(\hat{w})$ and wage offer $\hat{w}$. It follows that there cannot be a Nash equilibrium in which w drawn from F_I satisfies $w = \hat{w}$ with positive probability. This contradicts the fact that F_I has an atom at $\hat{w}$. Thus,

if $\alpha = 0$, then there cannot be a Nash equilibrium in which F_I and F_O both have an atom at $\hat{w}$.

Consider finally the case where F_O does not have an atom at $\hat{w}$. Observe that:

$$\lim_{w \uparrow \hat{w}} \pi_O(w) = F_I(\hat{w})\mathbb{E}\{g[x(w_I)] - \hat{w}|w_I \leq \hat{w}\} - [F_I(\hat{w}) - F_I(\hat{w}^-)]\{g[x(\hat{w})] - \hat{w}\}$$

and that:

$$\lim_{w \downarrow \hat{w}} \pi_O(w) = F_I(\hat{w})\mathbb{E}\{g[x(w_I)] - \hat{w}|w_I \leq \hat{w}\}.$$

Because F_I has an atom at $\hat{w}$, it must be that $F_I(\hat{w}) - F_I(\hat{w}^-) > 0$. In addition, $g[x(\hat{w})] - \hat{w} > 0$, noting that $P_O(\hat{w}) > 0$, $x(\hat{w}) > 0$, and $\pi_I(\hat{w}) \geq 0$ in a Nash equilibrium. It follows that $\lim_{w \uparrow \hat{w}} \pi_O(w) < \lim_{w \downarrow \hat{w}} \pi_O(w)$. Hence, there exist $\eta > 0$ and $\xi > 0$ such that firm O obtains a higher expected payoff by offering the wage $\hat{w} + \eta$ than by offering any wage $w \in [\hat{w} - \xi, \hat{w})$. Noting also that F_O does not have an atom at $\hat{w}$, it follows that there cannot be a Nash equilibrium in which w drawn from F_O satisfies $w \in [\hat{w} - \xi, \hat{w}]$. This contradicts the fact that for any $\epsilon > 0$, w drawn from F_O satisfies $w \in [\hat{w} - \epsilon, \hat{w}]$ with positive probability. Thus, there cannot be a Nash equilibrium in which F_I but not F_O has an atom at $\hat{w}$. ∎

Claim 13 *It must be that F_O does not have an atom at any $w > w^l$.*

Proof Suppose to the contrary that there exists $\hat{w} > w^l$ such that F_O has an atom at $\hat{w}$. I begin by arguing that for any $\epsilon > 0$, there is positive probability that w drawn from F_I satisfies $w \in [\hat{w} - \epsilon, \hat{w}]$.

Suppose that there exists $\epsilon > 0$ such that w drawn from F_I satisfies $w < \hat{w} - \epsilon$ or $w > \hat{w}$ with probability one. Recall that F_I does not have an atom at any $w > w^l$. If firm O offers the wage $\hat{w}$, then the expected payoff to firm O is given by:

$$\pi_O(\hat{w}) = F_I(\hat{w})\mathbb{E}\{g[x(w_I)] - \hat{w}|w_I \leq \hat{w}\}.$$

Because $\hat{w} > w^l$, it must be that $F_I(\hat{w}) > 0$. Choose any $\tilde{w} \in (\hat{w} - \epsilon, \hat{w})$. If firm O offers the wage $\tilde{w}$, then the expected payoff to firm O is given by:

$$\begin{aligned}\pi_O(\tilde{w}) &= F_I(\tilde{w})\mathbb{E}\{g[x(w_I)] - \tilde{w}|w_I \leq \tilde{w}\} = F_I(\hat{w})\mathbb{E}\{g[x(w_I)] - \tilde{w}|w_I \leq \hat{w}\} \\ &> F_I(\hat{w})\mathbb{E}\{g[x(w_I)] - \hat{w}|w_I \leq \hat{w}\} = \pi_O(\hat{w}),\end{aligned}$$

where the second equality follows because $F_I(\tilde{w}) = F_I(\hat{w})$ and so $\mathbb{E}\{g[x(w_I)]|w_I \leq \tilde{w}\} = \mathbb{E}\{g[x(w_I)]|w_I \leq \hat{w}\}$. Because firm O obtains a higher expected payoff by offering the wage

$\tilde{w}$ than by offering the wage $\hat{w}$, there cannot be a Nash equilibrium in which w drawn from F_O satisfies $w = \hat{w}$ with positive probability. This contradicts the fact that F_O has an atom at $\hat{w}$. It follows that for any $\epsilon > 0$, there is positive probability that w drawn from F_I satisfies $w \in [\hat{w} - \epsilon, \hat{w}]$.

Recall that F_I does not have an atom at $\hat{w}$. Observe that:

$$\lim_{w \downarrow \hat{w}} \pi_I[x(\hat{w}), w] = F_O(\hat{w})\{g[x(\hat{w})] - \hat{w}\} - x(\hat{w})$$

and that:

$$\lim_{w \uparrow \hat{w}} \pi_I[x(w), w] \geq \lim_{w \uparrow \hat{w}} \pi_I[x(\hat{w}), w] = F_O(\hat{w}^-)\{g[x(\hat{w})] - \hat{w}\} - x(\hat{w}),$$

where the inequality follows because $x(w) = \arg\max_{h \in \mathbb{R}_+} \pi_I(h, w)$. Because F_O has an atom at $\hat{w}$, it must be that $F_O(\hat{w}) > F_O(\hat{w}^-)$. In addition, $\mathbb{E}\{g[x(w_I)] - \hat{w}|w_I \leq \hat{w}\} \geq 0$, noting that $F_I(\hat{w}) > 0$ because $\hat{w} > w^l$ and that $\pi_O(\hat{w}) \geq 0$ in a Nash equilibrium. Since $g[x(\hat{w})] > \mathbb{E}\{g[x(w_I)]|w_I \leq \hat{w}\}$, it must be that $g[x(\hat{w})] - \hat{w} > 0$. It follows that $\lim_{w \uparrow \hat{w}} \pi_I[x(w), w] < \lim_{w \downarrow \hat{w}} \pi_I[x(\hat{w}), w]$. Hence, there exist $\eta > 0$ and $\xi > 0$ such that firm I obtains a higher expected payoff by choosing training level $x(\hat{w})$ and wage offer $\hat{w} + \eta$ than by choosing any training level $x(w)$ and wage offer w such that $w \in [\hat{w} - \xi, \hat{w})$. Noting also that F_I does not have an atom at $\hat{w}$, it follows that there cannot be a Nash equilibrium in which w drawn from F_I satisfies $w \in [\hat{w} - \xi, \hat{w}]$ with positive probability. This contradicts the fact that for any $\epsilon > 0$, w drawn from F_I satisfies $w \in [\hat{w} - \epsilon, \hat{w}]$ with positive probability. Thus, there cannot be a Nash equilibrium in which F_O has an atom at $\hat{w}$. ∎

Claim 14 *For any w^a and w^b such that $w^l \leq w^a < w^b \leq w^u$, it must be that $F_I(w^a) < F_I(w^b)$.*

Proof Assume to the contrary that there exist w^a and w^b with $w^l \leq w^a < w^b \leq w^u$ such that $F_I(w^a) = F_I(w^b)$. Let w^c be the least number such that $F_I(w) > F_I(w^a)$ for all $w > w^c$. Because $F_I(w^a) = F_I(w^b)$ where $w^a < w^b$, it must be that $w^c > w^a$. Recall that F_I does not have an atom at any $w > w^l$. This implies that $F_I(w^c) = F_I(w^a)$. Hence, it must be that $w^c < w^u$. Otherwise, if $w^c = w^u$, then it would have to be that $F_I(w^a) = F_I(w^c) = 1$ where $w^a < w^c$, which would lead to the contradiction that $w^a < w^u$ is no less than the supremum of the support of F_I.

First, suppose that $F_O(w) > F_O(w^c)$ for all $w > w^c$. Choose any $\tilde{w} \in (w^a, w^c)$. Recall

that F_I does not have an atom at any $w > w^l$. Noting that $F_I(w^c) = F_I(\tilde{w}) > 0$, one obtains:

$$\pi_O(w^c) = \lim_{w \downarrow w^c} \pi_O(w) = F_I(w^c)\mathbb{E}\{g[x(w_I)] - w^c | w_I \le w^c\} = F_I(\tilde{w})\mathbb{E}\{g[x(w_I)] - w^c | w_I \le \tilde{w}\}$$

$$< F_I(\tilde{w})\mathbb{E}\{g[x(w_I)] - \tilde{w} | w_I \le \tilde{w}\} = \pi_O(\tilde{w}).$$

Thus, there exists $\epsilon > 0$ such that firm O obtains a higher expected payoff by offering the wage $\tilde{w}$ than by offering any wage $w \in [w^c, w^c + \epsilon)$. It follows that there cannot be a Nash equilibrium in which w drawn from F_O satisfies $w \in [w^c, w^c + \epsilon)$ with positive probability. This contradicts the fact that $F_O(w) > F_O(w^c)$ for all $w > w^c$.

Hence, there must exist $\epsilon > 0$ such that $F_O(w) = F_O(w^c)$ for all $w \in (w^c, w^c + \epsilon)$. Recall that F_O does not have an atom. For any $w \in (w^c, w^c + \epsilon)$, note that $F_O(w^c) = F_O(w) > 0$ and so $x(w^c) = x(w) > 0$. The following holds for all $w \in (w^c, w^c + \epsilon)$:

$$\pi_I[x(w^c), w^c] = F_O(w^c)\{g[x(w^c)] - w^c\} - x(w^c) = F_O(w)\{g[x(w)] - w^c\} - x(w)$$

$$> F_O(w)\{g[x(w)] - w\} - x(w) = \pi_I[x(w), w].$$

Thus, firm I obtains a higher expected payoff by choosing training level $x(w^c)$ and wage offer w^c than by choosing training level $x(w)$ and wage offer w, where $w \in (w^c, w^c + \epsilon)$. Noting also that F_I does not have an atom at w^c, it follows that there cannot be a Nash equilibrium in which w drawn from F_I satisfies $w \in [w^c, w^c + \epsilon)$ with positive probability. This contradicts the fact that w^c is the least number such that $F_I(w) > F_I(w^a)$ for all $w > w^c$. Thus, there cannot exist w^a and w^b with $w^l \le w^a < w^b \le w^u$ such that $F_I(w^a) = F_I(w^b)$. ■

Claim 15 *For any w^a and w^b such that $w^l \le w^a < w^b \le w^u$, it must be that $F_O(w^a) < F_O(w^b)$.*

Proof Assume to the contrary that there exist w^a and w^b with $w^l \le w^a < w^b \le w^u$ such that $F_O(w^a) = F_O(w^b)$. Let w^c be the least number such that $F_O(w) > F_O(w^a)$ for all $w > w^c$. Because $F_O(w^a) = F_O(w^b)$ where $w^a < w^b$, it must be that $w^c > w^a$. Recall that F_O does not have an atom. This implies that $F_O(w^c) = F_O(w^a)$. Hence, it must be that $w^c < w^u$. Otherwise, if $w^c = w^u$, then it would have to be that $F_O(w^a) = F_O(w^c) = 1$ where $w^a < w^c$, which would lead to the contradiction that $w^a < w^u$ is no less than the supremum of the support of F_O.

Recall that $F_I(w)$ is increasing for $w \in [w^l, w^u]$. Hence, $F_I(w) > F_I(w^c)$ for all $w > w^c$. Choose any $\tilde{w} \in (w^a, w^c)$. Recall that F_O does not have an atom. Noting that $F_O(w^c) =$

$F_O(\tilde{w}) > 0$ and so $x(w^c) = x(\tilde{w}) > 0$, one obtains:

$$\pi_I[x(w^c), w^c] = \lim_{w \downarrow w^c} \pi_I[x(w), w] = F_O(w^c)\{g[x(w^c)] - w^c\} - x(w^c)$$

$$= F_O(\tilde{w})\{g[x(\tilde{w})] - w^c\} - x(\tilde{w}) < F_O(\tilde{w})\{g[x(\tilde{w})] - \tilde{w}\} - x(\tilde{w}) = \pi_I[x(\tilde{w}), \tilde{w}].$$

Thus, there exists $\epsilon > 0$ such that firm I obtains a higher expected payoff by choosing training level $x(\tilde{w})$ and wage offer $\tilde{w}$ than by choosing any training level $x(w)$ and wage offer w such that $w \in [w^c, w^c + \epsilon)$. It follows that there cannot be a Nash equilibrium in which w drawn from F_I satisfies $w \in [w^c, w^c + \epsilon)$ with positive probability. This contradicts the fact that $F_I(w) > F_I(w^c)$ for all $w > w^c$. Hence, there cannot exist w^a and w^b with $w^l \leq w^a < w^b \leq w^u$ such that $F_O(w^a) = F_O(w^b)$. $\blacksquare$

A.9 Proof of Lemma 7

The lemma follows from the claims below. Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium.

Claim 16 *It must be that $h^l = 0$ and $h^u = g'^{-1}(1)$.*

Proof Recall that the training level $\hat{h}$ and wage offer $\hat{w}_I$ chosen by firm I satisfy $\hat{h} = x(\hat{w}_I)$ with probability one, where $x(\hat{w}_I) = g'^{-1}[1/F_O(\hat{w}_I)]$. Let $\Pr(\cdot)$ represent the probability that w_I drawn from F_I is such that the enclosed statement is true. Recall that F_O does not have an atom at any $w \in \mathbb{R}_+$ and is increasing for $w \in [w^l, w^u]$. Note that F_O is invertible over the set consisting of all $w \in \mathbb{R}_+$ such that $F_O(w) \in (0, 1)$. The distribution function K is given by the following for all $h \in \mathbb{R}_+$:

$$K(h) = \Pr[x(w_I) \leq h] = \Pr\{g'^{-1}[1/F_O(w_I)] \leq h\} = \Pr[F_O(w_I) \leq 1/g'(h)]$$
$$= \Pr\{w_I \leq F_O^{-1}[1/g'(h)]\} = F_I\{F_O^{-1}[1/g'(h)]\},$$

where the function $F_O^{-1}(z)$ is the inverse of F_O for $z \in (0, 1)$ and is defined as w^l for $z = 0$ and as w^u for $z \geq 1$. Note that $h^l = 0$ is the smallest number r such that $K(h) > 0$ for all $h > r$ and that $h^u = g'^{-1}(1)$ is the greatest number r such that $K(h) < 1$ for all $h < r$. $\blacksquare$

Claim 17 *It must be that K does not have an atom at any $h > h^l$.*

Proof Recall that the training level $\hat{h}$ and wage offer $\hat{w}_I$ chosen by firm I satisfy $\hat{h} = x(\hat{w}_I)$ with probability one, where $x(\hat{w}_I) = g'^{-1}[1/F_O(\hat{w}_I)]$. Let $\Pr(\cdot)$ represent the probability that

w_I drawn from F_I is such that the enclosed statement is true. Recall that F_O does not have an atom at any $w \in \mathbb{R}_+$ and is increasing for $w \in [w^l, w^u]$. Note that F_O is invertible over the set consisting of all $w \in \mathbb{R}_+$ such that $F_O(w) \in (0, 1)$. The distribution function K is given by the following for all $h \in \mathbb{R}_+$:

$$K(h) = \Pr[x(w_I) \leq h] = \Pr\{g'^{-1}[1/F_O(w_I)] \leq h\} = \Pr[F_O(w_I) \leq 1/g'(h)]$$
$$= \Pr\{w_I \leq F_O^{-1}[1/g'(h)]\} = F_I\{F_O^{-1}[1/g'(h)]\},$$

where the function $F_O^{-1}(z)$ is the inverse of F_O for $z \in (0, 1)$ and is defined as w^l for $z = 0$ and as w^u for $z \geq 1$. Recall that g' and F_O^{-1} are continuous on the domain $\mathbb{R}_{++}$. Moveover, F_I is continuous on the domain $[w^l, \infty)$. Hence, $K(h)$ is continuous for all $h \in \mathbb{R}_{++}$. It follows that K does not have an atom at any $h > h^l$. ∎

Claim 18 *For any h^a and h^b such that $h^l \leq h^a < h^b \leq h^u$, it must be that $K(h^a) < K(h^b)$.*

Proof Recall that the training level $\hat{h}$ and wage offer $\hat{w}_I$ chosen by firm I satisfy $\hat{h} = x(\hat{w}_I)$ with probability one, where $x(\hat{w}_I) = g'^{-1}[1/F_O(\hat{w}_I)]$. Let $\Pr(\cdot)$ represent the probability that w_I drawn from F_I is such that the enclosed statement is true. Recall that F_O does not have an atom at any $w \in \mathbb{R}_+$ and is increasing for $w \in [w^l, w^u]$. Note that F_O is invertible over the set consisting of all $w \in \mathbb{R}_+$ such that $F_O(w) \in (0, 1)$. The distribution function K is given by the following for all $h \in \mathbb{R}_+$:

$$K(h) = \Pr[x(w_I) \leq h] = \Pr\{g'^{-1}[1/F_O(w_I)] \leq h\} = \Pr[F_O(w_I) \leq 1/g'(h)]$$
$$= \Pr\{w_I \leq F_O^{-1}[1/g'(h)]\} = F_I\{F_O^{-1}[1/g'(h)]\},$$

where the function $F_O^{-1}(z)$ is the inverse of F_O for $z \in (0, 1)$ and is defined as w^l for $z = 0$ and as w^u for $z \geq 1$. Recall that g' is decreasing on $\mathbb{R}_+$, F_O^{-1} is increasing on $[0, 1]$, and F_I is increasing on $[w^l, w^u]$. Hence, K is increasing on $[h^l, h^u]$ as desired. ∎

A.10   Proof of Theorem 1

The theorem follows from the claims below. Suppose that the mixed strategies (F_I, x) and F_O for firms I and O constitute a Nash equilibrium.

Claim 19 *The distribution function F_O is uniquely defined by:*

$$F_O(w) = \begin{cases} 0 & \text{if } w \leq g(0) \\ x(w)/\{g[x(w)] - w\} & \text{if } g(0) < w < g[g'^{-1}(1)] - g'^{-1}(1) \\ 1 & \text{if } w \geq g[g'^{-1}(1)] - g'^{-1}(1) \end{cases}.$$

Proof Recall that $\hat{w}_I$ drawn from F_I satisfies $\pi_I[x(\hat{w}_I), \hat{w}_I] = 0$ with probability one. Noting that F_O does not have an atom at any $w \in \mathbb{R}_+$, this implies that $\hat{w}_I$ drawn from F_I satisfies $F_O(\hat{w}_I)\{g[x(\hat{w}_I)] - \hat{w}_I\} - x(\hat{w}_I) = 0$ with probability one. Because F_I is strictly increasing on the interval $[w^l, w^u]$, it follows that $F_O(w)\{g[x(w)] - w\} - x(w) = 0$ for all $w \in [w^l, w^u]$. Hence, $F_O(w) = x(w)/\{g[x(w)] - w\}$, where $w^l < w \leq w^u$. In addition, it must be that $F_O(w^l) = 0$ because F_O does not have an atom at any $w \in \mathbb{R}_+$.

I next argue that the expression in the statement of the claim defines a unique strategy F_O for firm O. By definition, $x(w) = g'^{-1}[1/F_O(w)]$ for all $w \in [w^l, w^u]$. Noting that F_O is strictly increasing on the interval $[w^l, w^u]$, let F_O^{-1} be the inverse of F_O when F_O is restricted to the interval $[w^l, w^u]$. It follows that $w = F_O^{-1}\{1/g'[x(w)]\}$, where $w^l \leq w \leq w^u$. Thus, the condition $F_O(w)\{g[x(w)] - w\} - x(w) = 0$ for all w such that $w^l \leq w \leq w^u$ is equivalent to the condition $\{1/g'[x(w)]\}\big(g[x(w)] - F_O^{-1}\{1/g'[x(w)]\}\big) - x(w) = 0$ for all $x(w)$ such that $h^l \leq x(w) \leq h^u$. For $h^l < x(w) \leq h^u$, the preceding equality can be written as $1/g'[x(w)] = F_O\{g[x(w)] - x(w)g'[x(w)]\}$.

The left-hand side of this equation is continuous and increasing in $x(w)$, approaching 0 in the limit as $x(w)$ goes to h^l, and equal to 1 for $x(w) = h^u$. The term inside braces on the right-hand side of this equation is continuous and increasing in $x(w)$, approaching w^l in the limit as $x(w)$ goes to h^l, and equal to w^u for $x(w) = h^u$. Hence, this equation is satisfied by a unique distribution function F_O, which is continuous and increasing on the interval $[w^l, w^u]$ and satisfies $F_O(w^l) = 0$ and $F_O(w^u) = 1$. ∎

Claim 20 *The distribution function F_I is uniquely defined by:*

$$F_I(w) = \begin{cases} 0 & \text{if } w \leq g(0) \\ x(w)/\{g[x(w)] - w\} & \text{if } g(0) < w < g[g'^{-1}(1)] - g'^{-1}(1) \\ 1 & \text{if } w \geq g[g'^{-1}(1)] - g'^{-1}(1) \end{cases}.$$

Proof Recall that $\hat{w}_O$ drawn from F_O satisfies $\pi_O(\hat{w}_O) = 0$ with probability one. Noting that F_O does not have an atom at any $w \geq w^l$ and that F_I does not have an atom at any

55

$w > w^l$, this implies that $\hat{w}_O$ drawn from F_O satisfies $F_I(\hat{w}_O)\mathbb{E}\{g[x(w_I)] - \hat{w}_O | w_I \leq \hat{w}_O\} = 0$ with probability one. Because F_O is strictly increasing on the interval $[w^l, w^u]$, it follows that $F_I(w)\mathbb{E}\{g[x(w_I)] - w | w_I \leq w\} = 0$ for all $w \in [w^l, w^u]$. This condition for F_I can be written as follows for $w \in [w^l, w^u]$:

$$\int_{w^l}^{w} g[x(v)] - w \, dF_I(v) + F_I(w^l)\{g[x(w^l)] - w\} = 0.$$

The preceding expression yields the differential equation $F_I'(w)\{g[x(w)] - w\} = F_I(w)$ on the interval $(w^l, w^u]$, where the Leibniz integral rule is used to help compute the derivative.

Recall that F_O is defined by the following equation for $w \in [w^l, w^u]$:

$$F_O(w)\{g[x(w)] - w\} - x(w) = 0.$$

The preceding expression yields the differential equation $F_O'(w)\{g[x(w)] - w\} = F_O(w)$ on the interval $(w^l, w^u]$, where the envelope theorem is applied when calculating the derivative. Hence, F_I and F_O are required to satisfy the same differential equation with the same boundary condition $F_I(w^u) = F_O(w^u) = 1$. By the existence and uniqueness theorem for first-order linear differential equations, it must be that $F_I(w) = F_O(w)$ for $w \in (w^l, w^u]$, where $F_O(w)$ is defined as above. Moreover, it must be that $F_I(w^l) = 0$ because F_I is a nonnegative and nondecreasing function with $\lim_{w \downarrow w^l} F_I(w) = 0$. ∎

It is straightforward to confirm that the strategies (F_I, x) and F_O fulfill all of the requirements for a Nash equilibrium when F_I and F_O are as defined in the two preceding claims and x is as specified in the main text.

A.11 Proof of Corollary 1

Recall that the training level $\hat{h}$ and wage offer $\hat{w}_I$ chosen by firm I satisfy $\hat{h} = x(\hat{w}_I)$ with probability one, where $x(\hat{w}_I) = g'^{-1}[1/F_O(\hat{w}_I)]$. Let $\Pr(\cdot)$ represent the probability that w_I drawn from F_I is such that the enclosed statement is true. Recall that F_O does not have an atom at any $w \in \mathbb{R}_+$ and is increasing for $w \in [w^l, w^u]$. Note that F_O is invertible over the set consisting of all $w \in \mathbb{R}_+$ such that $F_O(w) \in (0, 1)$. As shown above, $F_I(w) = F_O(w)$ for all $w \in \mathbb{R}_+$. The distribution function K is given by the following for all $h \in [h^l, h^u]$:

$$K(h) = \Pr[x(w_I) \leq h] = \Pr\{g'^{-1}[1/F_O(w_I)] \leq h\} = \Pr[F_O(w_I) \leq 1/g'(h)]$$
$$= \Pr\{w_I \leq F_O^{-1}[1/g'(h)]\} = F_I\{F_O^{-1}[1/g'(h)]\} = 1/g'(h),$$

where the function $F_O^{-1}(z)$ is the inverse of F_O for $z \in (0,1)$ and is defined as w^l for $z = 0$ and as w^u for $z = 1$. ∎

A.12 Proof of Corollary 2

The expected payoffs to the firms and the worker must satisfy:

$$\int_{w^l}^{w^u} \pi_I[x(v), v] \, dF_I(v) + \int_{w^l}^{w^u} \pi_O(v) \, dF_O(v) + \phi_S = \int_{h^l}^{h^u} g(h) \, dK(h) - \int_{h^l}^{h^u} h \, dK(h).$$

On the left-hand side, the first, second, and third terms are the expected payoffs to firm I, firm O, and the worker, respectively. On the right-hand side, the first term is the expected output produced by the worker, and the second term is the expected cost of training the worker. As shown above, w_I drawn from F_I satisfies $\pi_I[x(w_I), w_I] = 0$ with probability one, and w_O drawn from F_O satisfies $\pi_O(w_O) = 0$ with probability one. It follows that $\int_{w^l}^{w^u} \pi_I[x(v), v] \, dF_I(v) = 0$ and $\int_{w^l}^{w^u} \pi_O(v) \, dF_O(v) = 0$. Hence, the preceding expression reduces to:

$$\phi_S = \int_{h^l}^{h^u} g(h) - h \, dK(h).$$

Substituting for $K(h)$ using the formula derived above, one obtains:

$$\phi_S = \int_{h^l}^{h^u} [g(h) - h] \frac{-g''(h)}{[g'(h)]^2} \, dh = \frac{g(h) - h}{g'(h)} \Big|_{h^l}^{h^u} - \int_{h^l}^{h^u} 1 - \frac{1}{g'(h)} \, dh = g(h^u) - 2h^u + \int_{h^l}^{h^u} \frac{1}{g'(h)} \, dh,$$

where integration by parts is used. ∎

A.13 Derivations for Example 1

Given the production function $g(h) = Ah^\theta$ for all $h \in \mathbb{R}_+$, the first derivative is $g'(h) = A\theta/h^{1-\theta}$, and the second derivative is $g''(h) = -A\theta(1 - \theta)/h^{2-\theta}$. The inverse of the first derivative of the production function is $g'^{-1}(r) = (A\theta/r)^{1/(1-\theta)}$, where $r > 0$. Plugging the above functional forms for g and g'^{-1} into the expressions from lemmata 3 and 5, the infimum w^l and supremum w^u of the supports for the equilibrium wage offer distributions F_I and F_O are the following:

$$w^l = g(0) = 0,$$
$$w^u = g[g'^{-1}(1)] - g'^{-1}(1) = A^{\frac{1}{1-\theta}}(\theta^{\frac{\theta}{1-\theta}} - \theta^{\frac{1}{1-\theta}}).$$

Recall from lemmata 3 and 6 that F_O is atomless. Substituting for g'^{-1} in the definition of x following lemma 1, the function mapping the wage offered to the training provided by firm I is given by:

$$x(w) = g'^{-1}[1/P_O(w)] = [A\theta F_O(w)]^{\frac{1}{1-\theta}}.$$

From theorem 1, the cdf of the bid distribution is simply $F_Y(w) = 0$ for $w \leq 0$ and $F_Y(w) = 1$ for $w \geq A^{1/(1-\theta)}(\theta^{\theta/(1-\theta)} - \theta^{1/(1-\theta)})$, where $Y \in \{I, O\}$. If $0 < w < A^{1/(1-\theta)}(\theta^{\theta/(1-\theta)} - \theta^{1/(1-\theta)})$, then the bid distribution satisfies the following condition for $Y \in \{I, O\}$:

$$F_Y(w) = \frac{x(w)}{g[x(w)] - w} = \frac{[A\theta F_Y(w)]^{\frac{1}{1-\theta}}}{A[A\theta F_Y(w)]^{\frac{\theta}{1-\theta}} - w},$$

where the functional form for g, the above expression for x, and the fact that $F_I = F_O$ are used. Solving the preceding equation for F_Y yields:

$$F_Y(w) = \left(\frac{w}{A^{\frac{1}{1-\theta}}(\theta^{\frac{\theta}{1-\theta}} - \theta^{\frac{1}{1-\theta}})} \right)^{\frac{1-\theta}{\theta}}.$$

Using this result to substitute for F_O in the above expression for x, one has:

$$x(w) = A^{\frac{1}{1-\theta}}\theta^{\frac{1}{1-\theta}} \left(\frac{w}{A^{\frac{1}{1-\theta}}(\theta^{\frac{\theta}{1-\theta}} - \theta^{\frac{1}{1-\theta}})} \right)^{\frac{1}{\theta}} = \left(\frac{w}{A(1-\theta)} \right)^{\frac{1}{\theta}}.$$

Plugging the above functional form for g'^{-1} into the expressions from lemma 7, the infimum h^l and supremum h^u of the support for the equilibrium human capital distribution K are the following:

$$h^l = 0,$$
$$h^u = g'^{-1}(1) = (A\theta)^{\frac{1}{1-\theta}}.$$

From corollary 1, the cdf of the training distribution is as follows for $0 \leq h \leq (A\theta)^{1/(1-\theta)}$:

$$K(h) = \frac{1}{g'(h)} = \frac{h^{1-\theta}}{A\theta},$$

where the functional form for g' is used. Based on corollary 2, the expected payoff to the worker is equal to:

$$\phi_S = g[g'^{-1}(1)] - 2g'^{-1}(1) + \int_0^{g'^{-1}(1)} \frac{1}{g'(h)}\, dh = A(A\theta)^{\frac{\theta}{1-\theta}} - 2(A\theta)^{\frac{1}{1-\theta}} + \int_0^{(A\theta)^{\frac{1}{1-\theta}}} \frac{h^{1-\theta}}{A\theta}\, dh,$$

where the above expressions are substituted for g, g', and g'^{-1}. Evaluating the integral and simplifying the result, one has:

$$\phi_S = \frac{2(1-\theta)^2(A\theta^\theta)^{\frac{1}{1-\theta}}}{2-\theta}.$$

∎

A.14 Proof of Proposition 3

Consider the following strategy profile. As in the statement of the proposition, firm I chooses training level $h = \kappa$ with probability $p = 1 - \kappa/[g(\kappa) - g(0)]$. The cumulative distribution function Ω_I^κ of wages offered by firm I to a worker with training level $h = \kappa$ satisfies $\Omega_I^\kappa(w) = 0$ for $w < g(0)$, $\Omega_I^\kappa(w) = 1$ for $w > g(\kappa) - \kappa$, and is defined as $\Omega_I^\kappa(w) = \{[w - g(0)](1 - p)\}/\{[g(\kappa) - w]p\}$ for $g(0) \leq w \leq g(\kappa) - \kappa$. The cumulative distribution function Ω_I^0 of wages offered by firm I to a worker with training level $h = 0$ is $\Omega_I^0(w) = 0$ if $w < g(0)$ and $\Omega_I^0(w) = 1$ if $w \geq g(0)$. The cumulative distribution function Ω_O of wages offered by firm O satisfies $\Omega_O(w) = 0$ for $w < g(0)$, $\Omega_O(w) = 1$ for $w > g(\kappa) - \kappa$, and is defined as $\Omega_O(w) = \kappa/[g(\kappa) - w]$ for $g(0) \leq w \leq g(\kappa) - \kappa$. I confirm that these strategies constitute a Nash equilibrium by showing that neither firm I nor firm O has an incentive to deviate.

Assume that firm O follows the strategy Ω_O described above. Suppose first that firm I chooses training level 0 and wage offer $w \in \mathbb{R}_+$. If $w = g(0)$, then firm I receives the payoff $\alpha\Omega_O[g(0)][g(0) - g(0)] = 0$ with probability one. Otherwise, the expected payoff to firm I is zero for $w < g(0)$ and negative for $w > g(0)$. Suppose next that firm I chooses training level κ and wage offer $w \in \mathbb{R}_+$. If $g(0) < w \leq g(\kappa) - \kappa$, then firm I receives the expected payoff $\Omega_O(w)[g(\kappa) - w] - \kappa = \{\kappa/[g(\kappa) - w]\}[g(\kappa) - w] - \kappa = 0$. Otherwise, the expected payoff to firm I is negative for $w < g(0)$, nonpositive for $w = g(0)$, and negative for $w > g(\kappa) - \kappa$. Hence, firm I has no incentive to deviate from the strategy $(p, \Omega_I^\kappa, \Omega_I^0)$ described above because this strategy yields an expected payoff of zero, which is the maximum expected payoff attainable by firm I given that firm O follows the strategy Ω_O.

Assume that firm I follows the strategy $(p, \Omega_I^\kappa, \Omega_I^0)$ described above. If firm O offers a wage w with $g(0) < w \le g(\kappa) - \kappa$, then firm O receives the expected payoff $p\Omega_I^\kappa(w)[g(\kappa) - w] + (1 - p)\Omega_I^0(w)[g(0) - w] = [g(\kappa) - w]p\{[w - g(0)](1 - p)\}/\{[g(\kappa) - w]p\} + [g(0) - w](1 - p) = 0$. If firm O offers the wage $g(0)$, then the payoff to firm O is $p\Omega_I^\kappa[g(0)][g(\kappa) - g(0)] + (1 - p)(1 - \alpha)\Omega_I^0[g(0)][g(0) - g(0)] = 0$ with probability one. If firm O offers a wage $w < g(0)$, then the payoff to firm O is zero with probability one. If firm O offers a wage $w > g(\kappa) - \kappa$, then firm O receives a negative expected payoff. Hence, firm O has no incentive to deviate from the strategy Ω_O described above because this strategy yields an expected payoff of zero, which is the maximum expected payoff attainable by firm O given that firm I follows the strategy $(p, \Omega_I^\kappa, \Omega_I^0)$. ∎

A.15 Proof of Proposition 4

The analysis of Engelbrecht-Wiggans, Milgrom, and Weber (1983) shows that the following are the equilibrium strategies in the bidding game. The cumulative distribution function Γ_I^κ of wages offered by firm I to a worker with training level $h = \kappa$ satisfies $\Gamma_I^\kappa(w) = 0$ for $w < g(0)$, $\Gamma_I^\kappa(w) = 1$ for $w \ge (1 - q)g(0) + qg(\kappa)$, and is defined as $\Gamma_I^\kappa(w) = \{[w - g(0)](1 - q)\}/\{[g(\kappa) - w]q\}$ for $g(0) \le w < (1 - q)g(0) + qg(\kappa)$. The cumulative distribution function Γ_I^0 of wages offered by firm I to a worker with training level $h = 0$ is $\Gamma_I^0(w) = 0$ if $w < g(0)$ and $\Gamma_I^0(w) = 1$ if $w \ge g(0)$. The cumulative distribution function Γ_O of wages offered by firm O satisfies $\Gamma_O(w) = 0$ for $w < g(0)$, $\Gamma_O(w) = 1$ for $w \ge (1 - q)g(0) + qg(\kappa)$, and is defined as $\Gamma_O(w) = [g(\kappa) - g(0)](1 - q)/[g(\kappa) - w]$ for $g(0) \le w < (1 - q)g(0) + qg(\kappa)$.

Assume that firms I and O follow the strategies in the preceding paragraph. If firm O offers the wage $g(0)$, then it receives the payoff $q\Gamma_I^\kappa[g(0)][g(\kappa) - g(0)] + (1 - q)(1 - \alpha)\Gamma_I^0[g(0)][g(0) - g(0)] = 0$ with probability one. The expected payoff to firm O from offering a wage w with $g(0) < w \le (1 - q)g(0) + qg(\kappa)$ is $q\Gamma_I^\kappa(w)[g(\kappa) - w] + (1 - q)\Gamma_I^0(w)[g(0) - w] = 0$. If $h = 0$, then the expected payoff to firm I from offering the wage $g(0)$ is $\psi_I^0 = \alpha\Gamma_O[g(0)][g(0) - g(0)] = 0$. If $h = \kappa$, then the expected payoff to firm I from offering a wage w with $g(0) < w \le (1 - q)g(0) + qg(\kappa)$ is $\psi_I^\kappa = \Gamma_O(w)[g(\kappa) - w] - \kappa = [g(\kappa) - g(0)](1 - q) - \kappa$. Since q is the training probability, the expected payoff to firm I is $(1 - q)\psi_I^0 + q\psi_I^\kappa$ or, equivalently, $\psi_I(q) = [g(\kappa) - g(0)]q(1 - q) - q\kappa$.

The value of the training probability q that maximizes the expected payoff to firm I solves the first-order condition $\psi_I'(q) = [g(\kappa) - g(0)](1 - 2q) - \kappa = 0$, which gives $q = \frac{1}{2}\{1 - \kappa/[g(\kappa) - g(0)]\}$. Note that the second-order condition $\psi_I''(q) = -2[g(\kappa) - g(0)] < 0$ is satisfied. The expected payoff to firm I at this value of q is $\{[g(\kappa) - g(0)] - \kappa\}^2/\{4[g(\kappa) - g(0)]\} > 0$. ∎