AN EXPLORATION OF THE PROMOTION SIGNALING DISTORTION

by

Michael Waldman
Johnson Graduate School of Management
Cornell University
Sage Hall
Ithaca, NY 14853
(607) 255-8631, mw46@cornell.edu

and

Ori Zax
Department of Economics
Tel-Hai College
Upper Galilee
ISRAEL
orizax@gmail.com

September 2013

* We thank Jed DeVaro, Suman Ghosh, and David Wettstein for conversations that were helpful in the initial formulation of the paper.

ABSTRACT

Beginning with Waldman (1984a), it is well understood that in a world characterized by asymmetric learning promotions can serve as a signal of worker ability which can, in turn, lead to an inefficiently small number of promotions. In this paper we explore two related issues. First, how robust is the finding of a promotion signaling distortion to different ways of modeling the promotion process? Second, what are the various forms that the promotion signaling distortion can take? Our first conclusion is that a promotion signaling distortion exists across a wide range of settings, including some for which earlier work suggests no distortion. Our second conclusion is that, even if there is no inefficiency concerning the number of promotions, there can still be a promotion distortion that takes the form of inefficiencies concerning who is promoted.

I. INTRODUCTION

Most workers provide a detailed history of the jobs they previously held on their resumes which suggests that the history of job assignments and promotions provides valuable information to prospective employers. In the economics literature this phenomena is captured by the idea that job assignments and promotions serve as signals as initially modeled in Waldman (1984a) and extended in numerous subsequent papers. One of the main results in this literature is that the signaling role of promotions leads to promotion distortions, i.e., the promotion process is not fully efficient. In this paper we focus on the nature and robustness of this promotion distortion.

In Waldman's (1984a) model there are two periods, two jobs, and workers vary in terms of ability. In the first period all workers are assigned to a low level job where ability is not valuable, and workers accumulate both general and firm specific human capital. At the end of the period a worker's first period employer privately learns the worker's ability and then at the beginning of the second period high ability workers are promoted to a high level job where ability is more valuable. Further, because of the asymmetric nature of the learning process, promotions serve as signals of high worker ability to prospective employers which results in these firms bidding more for promoted workers. The end result is that promotions are associated with large wage increases which are paid in order to stop workers from being bid away.

In addition to showing that promotions can have a signaling role, Waldman also shows that signaling can lead to a distortion. The argument is straightforward. Because a promotion serves as a signal of high ability, prospective employers are willing to bid more for workers who receive a promotion so incumbent employers give promoted workers large wage increases in order to stop promoted workers from being bid away. The result is that, if a worker is only a little more productive at the high level job than the low level job, the worker is not promoted because this allows the firm to increase profits by avoiding the large wage increase associated with promotion. In other words, from an efficiency standpoint too few workers are promoted.

This analysis has been extended in various ways both theoretically and empirically. For example, Bernhardt (1995) allows workers to be in the labor market for more periods and shows

that for many workers the distortion takes the form of inefficiently delayed promotion rather than no promotion. Bernhardt also shows that the degree of distortion falls with the worker's education level, where the logic is that the promotion signal is smaller for more highly educated workers and thus there is less incentive to distort the promotion decision.[1] DeVaro and Waldman (2012) extend Bernhardt's schooling analysis and then show evidence consistent with the resulting testable implications using the well known Baker, Gibbs, and Holmstrom (1994a,b) dataset that is based on the personnel records of a financial services firm.[2] And a number of other recent papers also find evidence consistent with the promotion-as-signal argument (see Belzil and Bognanno (2010), Okamura (2011), Bognanno and Melero (2012), and Cassidy, DeVaro, and Kauhanen (2012)).

But it has also been argued that the promotion signaling distortion is not a robust theoretical result. Specifically, some authors argue that, even if one assumes that learning in the labor market is asymmetric, a promotion distortion will not arise under many realistic ways of modeling the promotion process.[3] The first author to make an argument along these lines is Limor Golan in her 2005 paper. She reconsiders the model originally considered in Waldman (1984a) but allows for counteroffers in the wage determination process. She finds that introducing counteroffers on the part of initial employers eliminates the promotion distortion associated with signaling and, in addition, she incorporates training into the analysis and shows that training decisions are also efficient.

---

[1] There is an extensive theoretical literature on this topic. Some of the other papers in this literature include Ricart i Costa (1988), Waldman (1990), Zabojnik and Bernhardt (2001), Owan (2004), Zabojnik (2012), and Zax (2012).

[2] To be precise, DeVaro and Waldman (2012) develop predictions concerning probability of promotion and predictions concerning wage inceases that accompany promotion. They find that the promotion probability predictions hold for all education groups, while the predictions concerning the size of promotion wage increases hold only for bachelors and masters degree holders.

[3] There is substantial empirical evidence pointing to asymmetric learning in labor markets as initially modeled in Greenwald (1979,1986) and Waldman (1984a). The first paper to provide evidence consistent with asymmetric learning was Gibbons and Katz (1991). In addition to DeVaro and Waldman (2012) mentioned above and other recent papers that provide evidence for the promotion-as-signal hypothesis, other recent papers that provide evidence consistent with asymmetric learning in labor markets include Pinkston (2009) and Kahn (2013). Schoenberg (2007) argues that she finds weak evidence for asymmetric learning, but as argued in Waldman (2012), it is unclear that the test for which she claims to find no evidence for asymmetric learning is, in fact, a valid test of the asymmetric learning hypothesis.

No promotion distortion has also been found in analyses such as DeVaro and Kauhanen (2013) and Waldman (2013) in which promotions serve as signals but there is a slot constraint, i.e., each firm has a single managerial position. As a result, in those models firms do not have a choice concerning how many workers to promote and there is no promotion distortion. This is not fully surprising because the promotion distortion identified in papers like Waldman (1984a) and Bernhardt (1995) concerns the probability of promotion which is not a choice variable in these more recent papers characterized by slot constraints.

In this paper we investigate two related issues. First, we consider whether Golan's argument is correct both in terms of an enriched version of the specific setting she considered and more generally. In other words, we consider whether the promotion distortion due to signaling is a fragile result that disappears when modeling is made more realistic or a more robust result? Second, we consider the nature of the promotion distortion. Earlier papers focus on promotion distortions that concern the frequency or timing of promotions. In a setting with slot constraints, frequency or timing may not be a choice variable for the firm. We thus ask whether there can, nevertheless, be important promotion inefficiencies in such a setting.

In our first analysis we focus on a setting similar to the one considered by Golan (2005). Following Waldman (1984a), Golan assumes that productivity on the low level job is independent of worker ability. Our model is similar but we more realistically allow productivity on the low level job to vary positively with worker ability. We show that Golan is correct that, if the wage determination process allows for counteroffers by current employers, then there is no promotion distortion when output on the low level job is independent of worker ability. However, when we more realistically assume that productivity on the low level job is a strictly positive function of worker ability, then there is a promotion distortion even when counteroffers are possible. In other words, in what is arguably the most realistic case, i.e., counteroffers and worker ability affecting productivity on each job level, too few workers are promoted as in

Waldman's original analysis.[4] We also show how Golan's results concerning training change when productivity on the low level job is a strictly positive function of ability.

In our second analysis we increase what prospective employers can observe concerning the promotion process. Previous models of promotion signaling typically assume either a single worker at the low level job or multiple workers but the market cannot observe the number of workers promoted. We assume multiple low level workers and allow the market to observe the number of promoted workers, where our focus is on whether allowing the market to observe the number of promotions eliminates or at least reduces the promotion signaling distortion.

One might conjecture that by promoting a higher number of workers a firm can signal that some promoted workers are of lower ability. In turn, this should reduce the wage offers prospective employers make to promoted workers which should reduce the incentive for the initial employer to distort the promotion decision. We find some, but limited, support for this argument. Specifically, we find that allowing the market to observe the number promoted improves the efficiency of the promotion process somewhat, where the improvement follows from the ability of a firm to use the number of promoted workers as a type of signal along the lines just discussed. But it is still the case that the promotion process is never fully efficient.

The last model we consider is a model with a single managerial job as in DeVaro and Kauhanen (2013) and Waldman (2013), but we allow for the possibility of multiple schooling groups. We first show that, if there is a single schooling group, then the promotion process is fully efficient as in the earlier analyses characterized by slot constraints. We then show, however, that with multiple schooling groups the promotion process is not fully efficient because the wrong worker is sometimes promoted. Specifically, highly educated workers are inefficiently favored in promotion decisions because the signal and thus the extra wage associated with promotion is smaller for highly educated workers.

---

[4] See Barron, Berger, and Black (2006) for an empirical analysis that shows that counteroffers are frequently employed in real world labor markets.

So our overall conclusion is that the promotion signaling distortion is a robust result found across multiple settings characterized by asymmetric learning. As in the initial analysis of Waldman (1984a), in many cases the distortion takes the form of too few promotions. But in other settings where the frequency of promotion is fixed as can be the case in a setting with slot constraints, the distortion takes the form of the wrong worker sometimes being promoted. The bottom line is, thus, given the evidence in the literature that supports the asymmetric learning hypothesis (see footnote 3), theory suggests that promotion decisions should frequently be characterized by distortions of various types.

The outline for the paper is as follows. Section II considers a two-period model where each firm hires a single young worker and focuses on the role of the counteroffer assumption. Section III changes Section II's model so that each firm can hire many low level workers and the market can observe the number of promoted workers. Section IV considers the efficiency of the promotion process when there are slot constraints for the high level position so inefficiency in the frequency of promotion is not a possibility. Section V provides an overview of our analyses and results. Section VI presents concluding remarks.

II. AN INVESTIGATION OF THE COUNTEROFFER ASSUMPTION

In this section we explore how the introduction of counteroffers affects the promotion signaling distortion. Further, like Golan (2005) who also considers this issue, we additionally consider the efficiency of human capital investments in this type of setting. Our main result is that, if productivity on the low level job is a strictly positive function of worker ability, then incorporating counteroffers frequently results in inefficiencies.[5]

A) The Model

In this section's model there are two periods, all firms are identical, and there is free entry

[5] The analysis in this section is related to analyses in Ghosh and Waldman (2010) and DeVaro and Waldman (2012).

into production. Labor is the only input and each worker is in the labor market both periods. Further, each firm employs either zero or one worker in each period, where this assumption is relaxed in the next section's analysis.

Worker i's ability is denoted θ_i, where θ_i is a random draw from a probability density function $f(\theta)$. We further assume $f(\theta)>0$ for all $\theta_L \leq \theta \leq \theta_H$ and $f(\theta)=0$ for all θ outside of this interval. In our initial specification the worker acquires no human capital during his or her career, but later we consider what happens when general human capital and firm specific human capital are introduced.

Each firm can assign a worker to either of two jobs, denoted 1 and 2, where assigning a worker to job 2 who was previously in job 1 is referred to as a promotion. If worker i is assigned to job 1 in period t, the worker's output is given by

$$y_{i,t}=c_1+d_1\theta_i, \tag{1}$$

while if the worker is assigned to job 2 the worker's output is given by

$$y_{i,t}=c_2+d_2\theta_i. \tag{2}$$

We assume $c_1>c_2$, $0\leq d_1<d_2$, and θ' is such that $c_1+d_1\theta'=c_2+d_2\theta'$. In other words, if $\theta_i<(>)\theta'$, then it is efficient to assign worker i to job 1(2). Job 1 is thus the low level job and job 2 the high level job, where as in Rosen (1982) and Waldman (1984b) there is a larger return to ability in the high level job. Let $E(\theta)$ be the expected ability level of workers in the population. We assume that $c_1+d_1E(\theta)>c_2+d_2E(\theta)$, i.e., a worker of average ability is efficiently assigned to job 1 rather than job 2. And we further assume that $\theta_L<\theta'<\theta_H$. That is, low ability workers are more efficiently assigned to job 1 and high ability workers to job 2.

At the beginning of period 1 each worker's ability level is unknown but all firms know that each worker's ability is drawn from the probability density function f(.), so each worker has an expected ability at the beginning of period 1 equal to $E(\theta)$. At the end of the period a worker's first period employer privately observes the worker's ability level, while as described in more detail below other firms infer information about the worker's ability by observing the second period job the first period employer assigns the worker to.

The wage determination process in the second period allows for counteroffers. To be specific, at the beginning of period 2 a worker's first period employer assigns the worker to a job. The other firms, in turn, observe this job assignment and make wage offers. The first period employer then observes the market wage offers and makes a wage counteroffer, where we assume that the worker stays if the first period employer matches the market wage offer and that the first period employer matches if it is indifferent between matching and not matching.[6]

The timing of the full game is as follows. At the beginning of period 1 firms simultaneously make wage offers and each worker chooses a firm to work for.[7] Each firm with a worker then assigns the worker to a job, production takes place and workers are paid, and then at the end of the period each operating firm privately observes the ability level of its period 1 worker (if $d_1>0$, this can be because the firm privately observes the worker's output).

At the beginning of period 2 each firm that employed a worker in period 1 offers the worker a job assignment. The other firms in the market then observe these job assignments and make wage offers to the worker. The period 1 employers then observe these market wage offers and make wage counteroffers. At the end of this wage bidding process, each worker chooses to work at the firm that offers the highest wage. Also, if multiple firms are tied in terms of the highest wage offer, the worker chooses randomly among these firms unless one was the first period employer in which case, as indicated earlier, the worker stays. Finally, after each worker chooses a firm to work at in period 2, firms assign workers to jobs, workers produce, and then get paid. Our focus is on pure strategy Perfect Bayesian equilibria where beliefs concerning off-the-equilibrium path actions are consistent with each such action being taken by the type with the

---

[6] Assuming that the first period employer matches is equivalent to assuming that workers accumulate an infinitesimal amount of firm specific human capital, while the assumption that the worker stays when the first period employer matches is equivalent to assuming an infinitesimally small moving cost. We also assume that the market wage offer is at least the minimum output of a worker who moves given beliefs about which workers move in equilibrium. This assumption is consistent with a trembling-hand type assumption where there is a small probability that the first period employer does not match even when doing so weakly increases profitability. We impose this assumption to rule out implausible equilibria in which an old worker stays with the first period employer and is paid less than the minimum output associated with moving.

[7] If two or more workers choose the same firm, then the firm chooses randomly among those who applied and the remaining workers are allowed to switch to a different employer.

smallest cost of choosing that action. This assumption concerning off-the-equilibrium path actions is similar to the notion of a Proper Equilibrium first discussed in Myerson (1978). Further, when there are multiple equilibria our focus is on equilibria that minimize inefficiencies. This last assumption raises the hurdle required to find a promotion distortion.

B) Analysis without Human Capital Investments

In this subsection we consider the model without human capital investments which is the initial analysis in Golan (2005). In the next subsection we introduce human capital investments. We begin with a benchmark analysis that concerns what happens when there is symmetric learning of worker ability which in our model means each worker's ability becomes public information at the end of the first period. Because of our assumption that $E(\theta)<\theta'$, in period 1 all workers are assigned to job 1 in this benchmark case and are paid $c_1+d_1E(\theta)$. Further, in period 2 worker i is assigned to job 1(2) if $\theta_i<(>)\theta'$ (if $\theta_i=\theta'$ the worker can be assigned to either job 1 or job 2), is paid $c_1+d_1\theta_i$ ($c_2+d_2\theta_i$) if $\theta_i<(>)\theta'$, and the worker remains with the first period employer.[8] In other words, job assignments are efficient, pay in period 1 equals expected productivity while pay in period 2 equals realized productivity, and there is no turnover.

Now consider asymmetric learning, i.e., only the first period employer directly observes a worker's ability at the end of the first period but other firms draw inferences about the worker's ability by observing the second period job assignment. We start with some preliminary results. First, as in the benchmark case, all workers are assigned to job 1 in period 1 and are paid the same wage. Second, in contrast to the symmetric learning benchmark, this wage exceeds expected productivity because it also reflects expected rents that a worker's first period employer earns in the following period. As is described in more detail below, these rents are due to a winner's curse problem that arises because of counteroffers and asymmetric information.

We now formally state what happens in this case. Below w_Y is the wage paid to young

[8] Related analyses of symmetric learning include Harris and Holmstrom (1982) and Gibbons and Waldman (1999).

workers in period 1, while $w_O(\theta_i)$ is the wage paid to an old worker in period 2 as a function of the worker's value for θ_i. All proofs are in the Appendix.

Proposition 1: If each worker's ability is privately observed at the end of period 1 by the worker's first period employer, then i) through iii) describe equilibrium behavior.

i) Each worker is assigned to job 1 in period 1 and is paid $w_Y > c_1 + d_1 E(\theta)$.

ii) If $d_1 = 0$, then in period 2 each worker i is assigned to job 1 (job 2) if $\theta_i < (>) \theta'$ (if $\theta_i = \theta'$ either assignment is possible), is paid $w_O(\theta_i) = c_1 + d_1\theta_L = c_1$ ($w_O(\theta_i) = c_2 + d_2\theta' = c_1$), and each worker remains with the first period employer.[9]

iii) If $d_1 > 0$, then each worker is assigned to job 1 in period 2, is paid $w_O(\theta_i) = c_1 + d_1\theta_L$, and each worker remains with the first period employer.

There are a number of results of interest in Proposition 1. First, similar to results found initially in Greenwald (1986) and Milgrom and Oster (1987), in period 2 a worker's wage is equal to the lowest productivity of any worker with the same labor market signal and this is due to the presence of the winner's curse. The idea here is that the first period employer knows the worker's ability while prospective employers can only observe the job assignment offered to the worker. Because of the counteroffer assumption, this yields that prospective employers will not bid above the lowest productivity of workers with the same job assignment. If they did, the first period employer would only match when the actual productivity was greater than or equal to the offered wage, so any actual hire by a prospective employer would result in losses for the hiring firm. As a result, the market wage offer equals the lowest possible productivity of the worker which is then matched by the first period employer, so this is the wage.

[9] ii) follows given our focus on equilibria that minimize inefficienices. See the proof of Proposition 1 in the Appendix for details. Golan (2005), on the other hand, argues that this equilibrium is more stable because it survives employers making assignment mistakes while the other equilibria do not.

Second, we find the main result in Golan (2005). When productivity at the low level job is independent of ability, i.e., $d_1=0$, then period 2 job assignments are efficient. The logic is that promotions serve as a signal of high ability as in Waldman (1984a), but in equilibrium the wage for promoted workers is the same as the wage for non-promoted workers. In turn, since there is no wage increase associated with promotion, firms do not have an incentive to distort the promotion decision so, in contrast to Waldman (1984a), period 2 job assignments are efficient.[10]

Third, we find that the result that assignments are efficient is fragile. For all $d_1>0$, including very small but strictly positive values for d_1, assignments are inefficient and, in fact, no one is assigned to job 2. The logic for why there is a promotion distortion here is similar to the logic in Waldman (1984a) for why the promotion decision is distorted in that analysis. That is, a promotion in this case results in a wage increase and, because firms have an incentive to avoid promotion wage increases, the probability of promotion is reduced below the efficient level.[11]

To see the argument more fully, suppose $d_1>0$ and there is a value θ^+ such that workers for whom $\theta_i<\theta^+$ are not promoted and those for whom $\theta_i>\theta^+$ are promoted. Given the winner's curse argument discussed above, the wage for the non-promoted workers would be $c_1+d_1\theta_L$ while the wage for the promoted workers would be $c_2+d_2\theta^+$ (the latter is true under the condition that prospective employers would find it efficient to assign such a worker to job 2 if the worker were to switch employers). So the extra wage the firm pays a promoted worker is $(c_2+d_2\theta^+)-(c_1+d_1\theta_L)$. The value for θ^+, in turn, is the value such that the extra productivity associated with promoting a worker with this value just equals the increase in the wage, i.e., $(c_2+d_2\theta^+)-(c_1+d_1\theta^+) =(c_2+d_2\theta^+)-(c_1+d_1\theta_L)$. But there is no value for θ^+ that satisfies this condition with the result that

---

[10] The above argument is incomplete in the sense that it takes wages as given and then shows that assignments are efficient given these wages. But in equilibrium the wages themselves are functions of the assignments. In the proof of Proposition 1 in the Appendix there is a more complete argument for why assignments are efficient when $d_1=0$.

[11] Golan (2005), page 382 and footnote 13, does indicate that her main efficiency result is due to no wage increase upon promotion in her model and this is "a consequence of production technology and distribution assumptions." Our analysis shows that this claim is correct and, further, that even if ability has just an infinitesimally small effect on productivity in the low level job, there will be a wage increase upon promotion and the result is a severe inefficiency in the promotion decision in the sense that no one is promoted.

no one is promoted.[12]

So overall, Proposition 1 shows that Golan (2005) is correct that Waldman's (1984a) result that promotions are inefficient disappears when counteroffers are added to the analysis. That is, in Waldman's initial model output in the low level job did not depend on worker ability and in that case introducing counteroffers eliminates the promotion distortion. But from another perspective Golan is incorrect in claiming that the promotion distortion is not robust to introducing counteroffers into the wage determination process. A more realistic assumption is that output on the low level job does depend on worker ability. And when we enrich the model to allow for this, even if we assume that the extra output on the low level job is vanishingly small, the result is a severe promotion distortion where, in fact, no one is promoted.

It is also worth noting that, as discussed above, when output on the low level job does not depend on ability, there is no promotion distortion but there is also a prediction of no wage increases upon promotion. In contrast, when output on the low level job does depend on ability, then there is a promotion distortion and a prediction that promotions will be associated with wage increases. Since various empirical studies such as Lazear (1992), Baker, Gibbs, and Holmstrom (1994a,b), and McCue (1996) show that promotions are typically associated with large wage increases, the case with a promotion distortion is more plausible both because the assumption that yields this case is more realistic and because the equilibrium in this case better matches the real world evidence (see footnote 16 for a related discussion).

C) Analysis with Human Capital Investments

The second main result in Golan (2005) is that the outcome remains efficient when human capital investments are introduced. In this subsection we consider investments in human capital when output on the low level job is allowed to increase with worker ability.[13]

---

[12] Setting $\theta^+=\theta_L$ seems to satisfy the condition which suggests that in equilibrium everyone rather than no one is promoted. But setting $\theta^+=\theta_L$ violates the condition that prospective employers would find it efficient to assign a worker who moves to job 2 rather than job 1. As we show in the proof of Proposition 1 in the Appendix, a full analysis that takes this into account yields that no one is promoted.

Following Golan (2005), we assume that after observing worker ability at the end of the first period, a firm can invest in the worker's human capital. Specifically, the firm makes a take-it or leave-it offer to the worker of a fee, $p(\theta)$, that the worker pays the firm, where $C(\theta)$ is the cost to the firm of providing the training. If the worker rejects the offer, then the worker's productivity in period 2 is as described earlier. If the worker accepts the offer, then output in the high level job at the first period employer increases to $c_2+d_2\theta_i+\Delta(\theta_i)$, while output in the low level job is unchanged. It is also assumed that $\Delta(\theta)-C(\theta)$ is increasing in θ and that $\Delta(\theta_H)-C(\theta_H)>0$ while $\Delta(\theta_L)-C(\theta_L)<0$. This last assumption tells us that it is efficient for high ability workers who are promoted to be trained but training is not efficient for workers who are low ability and/or not promoted. Further, at the beginning of period 2 prospective employers can observe whether a worker was trained or not and the fee when training takes place, but cannot observe either the training cost or the job 2 increase in productivity. For this analysis we also assume that a worker learns his or her own ability at the end of period 1.[14]

The efficient training rule in this model is independent of whether training is general or firm specific because efficiency is consistent with no turnover. Let θ^* be the value for θ such that $\Delta(\theta^*)-C(\theta^*)=0$. If $\theta^*>\theta'$, then it is efficient for workers to be promoted whenever $\theta_i>\theta'$ and promoted workers to be trained when $\theta_i>\theta^*$ (efficiency requires that non-promoted workers are never trained since training does not affect output in the low level job). If $\theta^*<\theta'$, then efficiency requires that all promoted workers are trained and workers are promoted whenever equation (3) is satisfied (when (3) holds as an equality then training with promotion and not training without promotion are both efficient).

$$(3) \qquad (c_2+d_2\theta_i)+[\Delta(\theta_i)-C(\theta_i)]\geq(c_1+d_1(\theta_i))$$

---

[13] Other papers that consider human capital investments in the case of asymmetric learning include Katz and Ziderman (1990), Waldman (1990), Chang and Wang (1996), and Acemoglu and Pischke (1998).

[14] Adding this assumption to the analysis in the previous subsection has no effect on equilibrium behavior. It is also worth noting that assuming the fee is publicly observable has no effect on the results. We make this assumption so that our analysis is consistent with Golan's.

Call θ'' the efficient cutoff ability level for promotion in this model, i.e., $\theta''=\theta'$ when $\theta^* \geq \theta'$ and from (3) we know $\theta''<\theta'$ when $\theta^*<\theta'$.

We now consider training decisions in equilibrium. We start with the case in which training is firm specific, i.e., the increase in productivity with training is given above when a worker does not switch employers at the beginning of period 2, but training has no effect on productivity at the new firm when a move occurs.

Proposition 2: If each worker's ability is observed at the end of period 1 by the worker and the worker's first period employer and this employer then makes firm specific human capital training decisions, then i) through iii) describe equilibrium behavior.

i) Each worker is assigned to job 1 in period 1 and is paid $w_Y > c_1 + d_1 E(\theta)$.

ii) If $d_1=0$, then in period 2 all job assignment and training decisions are efficient, all workers are paid c_1, training fees equal zero, and each worker remains with the first period employer.

iii) If $d_1>0$, then there exists a value θ^+, $\theta^+ \leq \theta_H$, such that in period 2 worker i is not promoted and not trained (is promoted and trained) if $\theta_i<\theta^+$ ($\theta_i>\theta^+$), is paid $c_1+d_1\theta_L$ ($\max\{c_1+d_1\theta^+, c_2+d_2\theta^+\}$), each promoted worker pays a training fee equal to $\max\{c_1+d_1\theta^+, c_2+d_2\theta^+\} - (c_1+d_1\theta_L)$, and each worker remains with the first period employer. Also, $\theta^+=\theta''$ when $\theta^* \leq \theta'$ while $\theta^+>\theta''$ when $\theta^*>\theta'$.

One result in Proposition 2 is the same finding as in Golan (2005). When $d_1=0$, both promotion decisions and training decisions are efficient. The logic extends the reasoning given earlier for ii) of Proposition 1. Because training is firm specific, the fact that there is training has no effect on second period equilibrium wages, i.e., a worker in period 2 is paid c_1 whether or not the worker is promoted and whether or not the worker is trained. In turn, since neither the promotion decision nor the training decision affects a worker's second period wage, firms have an incentive to choose both efficiently.

The second result of interest in Proposition 2 concerns what happens when $d_1>0$. Here we find that promotion and training decisions are not always fully efficient. Also, all promoted workers are trained. The logic behind these results is as follows. Suppose first that the lowest ability worker who is promoted is not trained. Then a logic similar to that presented earlier to explain why there are no promotions in Proposition 1 when $d_1>0$ yields there would also be no promotions here when $d_1>0$. Given this, suppose the lowest ability worker who is promoted is trained. Call this value for ability θ^+ as in the proposition. The value for θ^+ is determined by the value for θ such that the firm is indifferent between assigning a worker to jobs 1 and 2. Given the winner's curse result from earlier, this condition translates into (4).

$$(4) \qquad (c_2+d_2\theta^+)+[\Delta(\theta^+)-C(\theta^+)]-\max\{c_1+d_1\theta^+,c_2+d_2\theta^+\}+\max\{c_1+d_1\theta^+,c_2+d_2\theta^+\}-(c_1+d_1\theta_L)$$
$$=(c_1+d_1\theta^+)-(c_1+d_1\theta_L)$$

Equation (4) reduces to equation (5).

$$(5) \qquad (c_2+d_2\theta^+)+[\Delta(\theta^+)-C(\theta^+)]=c_1+d_1\theta^+$$

Comparing this equation to the definition of θ'' yields that the cutoff ability level for promotion in this equilibrium is efficient when $\theta^*\leq\theta'$ but not when $\theta^*>\theta'$

Intuitively, the training fee allows the firm to capture the increase in the wage associated with receiving training and being promoted. As a result, the signal need not distort the promotion decision even when $d_1>0$. In particular, when efficiency requires the lowest ability worker promoted to also be trained, then promotions are efficient because there is no cost to the firm associated with promoting the θ'' worker since the wage increase due to the signal is captured by the training fee. But when efficiency requires that the lowest ability worker promoted is not trained, then there is a promotion distortion because promoting and training in the efficient manner would not allow the firm to capture wage increases due to promotion signaling for the θ'' worker through the training fee.

We now consider equilibrium behavior when training is general, i.e., there is an increase in productivity with training whether or not a move occurs.

Proposition 3: If each worker's ability is observed at the end of period 1 by the worker and the worker's first period employer and this employer then makes general human capital training decisions, then i) through iii) describe equilibrium behavior.

i) Each worker is assigned to job 1 in period 1 and is paid $w_Y > c_1 + d_1 E(\theta)$.

ii) If $d_1 = 0$, then in period 2 all job assignment and training decisions are efficient, each promoted and trained worker i is paid $w_O(\theta_i) = c_2 + d_2\theta^* + \Delta(\theta^*)$ and pays a training fee given by $p(\theta_i) = (c_2 - c_1) + d_2\theta^* + \Delta(\theta^*)$, each non-promoted worker i and promoted but not trained worker i is paid $w_O(\theta_i) = c_1$, and each worker remains with the first period employer.

iii) If $d_1 > 0$, then there exists a value θ^+, $\theta^+ \leq \theta_H$, such that in period 2 worker i is not promoted and not trained (is promoted and trained) if $\theta_i < \theta^+$ ($\theta_i > \theta^+$), is paid $c_1 + d_1\theta_L$ ($\max\{c_1 + d_1\theta^+, c_2 + d_2\theta^+ + \Delta(\theta^+)\}$), each promoted worker pays a training fee equal to $\max\{c_1 + d_1\theta^+, c_2 + d_2\theta^+ + \Delta(\theta^+)\} - (c_1 + d_1\theta_L)$, and each worker remains with the first period employer. Also, $\theta^+ = \theta''$ when $\theta^* \leq \theta'$ while $\theta^+ > \theta''$ when $\theta^* > \theta'$.

Proposition 3 tells us that the model works similarly with general training as it did with firm specific training.[15] First, as found initially in Golan (2005), when $d_1 = 0$ the training decisions and job assignment decisions are both efficient and there is a wage increase upon promotion for those that are trained. But when $d_1 > 0$, as was also true in Proposition 2, training and promotion decisions are efficient when $\theta^* \leq \theta'$ but not when $\theta^* > \theta'$.

The logic for why promotion and training decisions are efficient when $d_1 = 0$ is basically the same as why promotion decisions are efficient in Proposition 1 in the case $d_1 = 0$. The net payment to workers is c_1 independent of whether they are assigned to job 1 or job 2 and whether

[15] As indicated earlier, following Golan (2005) we assume that prospective employers observe whether or not a worker is trained. It might be more realistic, however, to assume that training is not publicly observable. Under this alternative assumption it is still the case that, given both general and firm specific human capital training decisions, job assignment and training decision are efficient when $d_1 = 0$ and not fully efficient when $d_1 > 0$.

they are trained or not.[16] So, since at the point in time that promotion decisions and training decisions are made the firm making the decisions bears all the costs and benefits, firms have an incentive to make these decisions in an efficient fashion.

The logic for what happens when $d_1>0$ is basically the same as the logic for the case $d_1>0$ in Proposition 2. The training fee allows the firm to capture increases in the wage associated with a worker being trained and promoted. The result is that there is no promotion distortion when efficiency requires all promoted workers to be trained, but when this is not the case then promotion decisions are distorted.[17]

In summary, Golan (2005) shows that adding counteroffers to the model in Waldman (1984a) removes the finding of a promotion inefficiency. But Waldman's (1984a) model had the unrealistic feature that productivity on the low level job is independent of worker ability. In this section we explored what happens when counteroffers are allowed and we simultaneously relax the assumption that productivity on the low level job does not depend on worker ability. Our analysis reconfirms Golan's findings but we also show that these findings are not robust to allowing productivity on the low level job to be an increasing function of worker ability. Specifically, even if increases in worker ability have vanishingly small effects on productivity in the low level job, the results are: i) no promotions in the absence of training decisions; ii) inefficiencies in the promotion decision for some parameterizations when firm specific training is introduced; and iii) inefficiencies in the promotion decision for some parameterizations when general training is introduced. In other words, our analysis indicates that introducing

[16] By net payment we mean the payment after subtracting the training fee that the worker pays. Focusing on the gross rather than net payment associated with promotion in this case, there is a wage increase upon promotion which can be used to explain why numerous empirical studies find that promotions are associated with wage increases (see the discussion at the end of the previous subsection). But in this explanation the present discounted value of future promotion wage increases equals the training costs paid by current low level workers. And it seems unlikely that in reality such training fees are large enough to justify the large wage increases associated with promotion found in the empirical literature.

[17] Golan (2005) also considers what happens when workers can invest in general human capital before labor market entry, where the human capital increases productivity on the low level job. Another interesting case is when human capital acquisition during employment is not a choice variable but rather all workers acquire it via learning-by-doing. Although both are of interest, for length reasons we skip these cases.

counteroffers does not in general eliminate the promotion distortion in Waldman (1984a) type settings, although as shown by Golan this is true in the special case where productivity in the low level job is independent of worker ability.

III. AN ANALYSIS WHEN THE NUMBER OF PROMOTIONS IS OBSERVABLE

In Section II we assumed that each firm hires a single low level worker in period 1 and then decides whether or not to promote the worker in period 2. Previous models of promotion signaling typically assume either a single low level worker as we assumed in Section II or multiple low level workers but that the market has limited information about the initial employer's promotion decisions. That is, in models of the latter type it is typically assumed that the market only observes whether or not a worker is promoted and not the number or proportion of workers promoted by the worker's initial employer.[18] It is easy to show that these two approaches yield similar equilibria in terms of the cutoff ability level required for promotion.

But suppose instead that prospective employers could observe the number or proportion of workers who receive promotions. What is of interest is whether the promotion distortion is robust to this change in what is publicly observable. One might conjecture that allowing prospective employers to observe the proportion of workers promoted would have an important effect on the incentive for an initial employer to distort the promotion decision. That is what we investigate in this section.

A) The Model

Here we consider the initial model considered in the previous section, i.e., the model with no training decisions, except firms can now hire multiple workers in period 1. Specifically, in period 1 each firm can hire any number of workers between 0 and N, $N<\infty$, where there are

[18] An exception is promotion signaling models characterized by slot constraints for the managerial job. We consider a model of this type in the next section.

constant returns to scale up to hiring N workers and a worker's productivity on each job and in each period is the same as before.

At the beginning of period 2 each firm gives each of its period 1 employees a period 2 job assignment. Prospective employers then observe each worker's job assignment, the number of these employees that the firm assigns to job 1 at the beginning of period 2, and the number the firm assigns to job 2. Other than these extra pieces of information that prospective employers observe at the beginning of period 2 the timing of moves is the same as before.

B) Analysis

As discussed above, the basic question is whether having the ability to promote in a publicly observable way a higher number of workers reduces the incentive for a firm to distort the promotion decision. The argument for why it might is that by promoting a higher number of workers a firm can signal that the set of workers it promotes includes some workers of lower ability. Signaling in this way should lower the wage that prospective employers offer which, in turn, should reduce the subsequent counteroffers the initial employer needs to pay to stop promoted workers from being bid away.

We formally consider this issue in Proposition 4.

Proposition 4: Suppose each worker's ability is privately observed at the end of period 1 by the worker's first period employer and in period 2 the number of workers a firm assigns to job 1 and the number assigned to job 2 are both publicly observed. Holding all parameters other than θ_H fixed, there exists a value θ_H^* such that i) through v) describe equilibrium behavior.

i) Each worker is assigned to job 1 in period 1 and is paid $w_Y > c_1 + d_1 E(\theta)$.

ii) If $d_1 = 0$, then in period 2 each worker i is assigned to job 1 (job 2) if $\theta_i < (>) \theta'$ (if $\theta_i = \theta'$, then either assignment is possible), is paid $w_O(\theta_i) = c_1 + d_1\theta_L = c_1$ ($w_O(\theta_i) = c_2 + d_2\theta' = c_1$), and each worker remains with the first period employer.

iii) If $d_1 > 0$ and N=1, then in period 2 each worker is assigned to job 1, each worker i is paid

$w_O(\theta_i)=c_1+d_1\theta_L$, and each worker remains with the first period employer.

iv) If $d_1>0$, $N>1$, and $\theta_H<\theta_H^*$, then in period 2 each worker is assigned to job 1, each worker i is paid $w_O(\theta_i)=c_1+d_1\theta_L$, and each worker remains with the first period employer.

v) If $d_1>0$, $N>1$, and $\theta_H>\theta_H^*$, then equilibria are characterized by a strictly positive frequency of promotions, but promotion decisions are not fully efficient.[19]

Proposition 4 tells us that for many parameterizations having the ability to signal the number or proportion of workers promoted does not affect equilibrium behavior, but there is a range of parameterizations for which the frequency of promotions does rise. Consider first parameterizations in which $d_1=0$, i.e., ii) of Proposition 4. The finding that in this case allowing firms to signal the number or proportion of workers promoted results in no change in behavior is not surprising. In this case equilibrium was efficient when firms were constrained to hire just a single worker in period 1, so allowing firms to hire multiple young workers will not result in a change in the nature of equilibrium since no change can improve the efficiency of the second period promotion decision. iii) of Proposition 4 is also straightforward since that is just the case considered in iii) of Proposition 1.

To understand the logic for $d_1>0$ and $N>1$, i.e., iv) and v) of Proposition 4, consider what happens when $d_1>0$ and $N=2$. Specifically, consider a firm that hires 2 workers in period 1 and the wage offers received by these workers from prospective employers in period 2. This wage can depend on the number of workers the firm promotes, where consistent with a discussion above the wage potentially falls with the number promoted. Another property of this wage function is that it must satisfy the winner's curse constraint discussed in Section II. That is, the wage should be equal to the productivity at a prospective employer of the lowest ability worker who receives a promotion on the equilibrium path given the number of workers promoted.

[19] Due to space considerations, we do not provide a full characterization of equilibria in this case, although the following discussion provides some additional results concerning properties of these equilibria.

Given this, suppose the firm promotes a single worker in period 2 and let θ_1^+ be the lowest value for θ that results in this outcome on the equilibrium path. Then prospective employers would offer this worker $\max\{c_1+d_1\theta_1^+, c_2+d_2\theta_1^+\}$. We also know, however, that if the worker was not promoted, then the prospective employers would offer $c_1+d_1\theta_L$. For the firm to find it profitable to promote this lowest ability worker it must be the case that equation (6) is satisfied.

(6) $$(c_2+d_2\theta_1^+)-\max\{c_1+d_1\theta_1^+, c_2+d_2\theta_1^+)\geq(c_1+d_1\theta_1^+)-(c_1+d_1\theta_L)$$

But there is no value for θ_1^+ that satisfies equation (6). So there cannot be a strictly positive probability that in equilibrium the firm will promote a single worker. Another way to put this is that the logic for why there are no promotions when $d_1>0$ in Proposition 1 also tells us that on the equilibrium path a single worker cannot be promoted in this section's model.

Now suppose the firm promotes both workers and let θ_2^+ be the lowest value for θ consistent with this outcome on the equilibrium path and let θ_2^{++} be the ability level of the other worker in a realization of abilities where this lowest ability worker is promoted. Prospective employers would offer each worker $\max\{c_1+d_1\theta_2^+, c_2+d_2\theta_2^+\}$. We also know, however, that if neither worker was promoted, then prospective employers would offer each worker $c_1+d_1\theta_L$. When a pair of workers is promoted, it must be more profitable for the firm to promote the pair than to promote neither or then to promote one. Given that promoting a single worker is an off-the-equilibrium path event, the binding constraint concerns the option of promoting neither worker.[20] Given that the winner's curse means that when both workers are promoted neither leaves, we have that for the firm to promote both workers equation (7) must be satisfied.

(7) $$(c_2+d_2\theta_2^+)+(c_2+d_2\theta_2^{++})-2\max\{c_1+d_1\theta_2^+, c_2+d_2\theta_2^+\}\geq(c_1+d_1\theta_2^+)+(c_1+d_1\theta_2^{++})-2(c_1+d_1\theta_L)$$

If equation (7) is not satisfied, then the firm has an incentive to deviate from the proposed equilibrium behavior and promote no one.

[20] Because promoting a single worker is an off-the-equilibrium path event, we can choose a wage that prospective employers will pay when they observe a single worker promoted such that promoting a single worker is not an attractive option for the first period employer.

Equation (7) immediately tells us that there are always equilibria with no promotions. If there are no promotions, then observing both workers promoted is an off-the-equilibrium path event and θ_2^+ in the equation can be set at any value between θ_L and θ_H. If we set $\theta_2^+=\theta_H$, then inspection yields that (6) cannot be satisfied.[21] So no promotions is always an equilibrium. However, since we restrict attention to equilibria that minimize inefficiencies, this is the outcome only when there are no other equilibria that are more efficient.

To consider the statements in Proposition 4 concerning θ_H, first note that equation (7) can be rewritten as equation (8).

(8) $(c_2+d_2\theta_2^{++})-(c_1+d_1\theta_2^{++})\geq(c_1+d_1\theta_2^{+})-(c_2+d_2\theta_2^{+})+2\max\{c_1+d_1\theta_2^{+}, c_2+d_2\theta_2^{+}\}-2(c_1+d_1\theta_L)$

Note that for any fixed value of θ_2^+, if equation (8) is satisfied for some value for θ_2^{++}, $\theta_2^+\leq\theta_2^{++}\leq\theta_H$, then it is satisfied when $\theta_2^{++}=\theta_H$. Given this, hold all other parameters fixed and let $\theta_2^{++}=\theta_H$ and fix θ_2^+ at some constant value. Then increasing θ_H causes the left hand side of (8) to rise with no upper bound and the right hand side to remain unchanged. So for θ_H sufficiently large (8) is necessarily satisfied. This means that for θ_H sufficiently large there will be equilibria in which both workers are promoted with positive probability, so this is the outcome given our focus on equilibria that minimize inefficiencies.

On the other hand, suppose we hold all other parameters fixed and consider what happens as θ_H falls and approaches θ'. The upper bound on the left hand side of (8) falls and approaches zero while the right hand side of (8) has a lower bound that is strictly positive and bounded away from zero. So for θ_H sufficiently small (8) cannot be satisfied. This means that for θ_H sufficiently small there cannot be an equilibrium in which both workers are promoted with strictly positive probability.

The above analysis further tells us that, even for parameterizations for which there are equilibria with a strictly positive probability of both workers being promoted, there are no equilibria in which promotion decisions are fully efficient. There are two reasons for this. First,

[21] Notice that this belief is consistent with our assumption concerning beliefs associated with off-the-equilibrium path actions.

if one worker's ability is above θ' and the second worker's ability is below θ', then it is efficient to promote a single worker. But we know that promoting a single worker is not consistent with equilibrium behavior. So whenever a firm's workers have ability realizations where it is efficient to promote just one worker, then period 2 promotion decisions will be distorted.

Second, equation (8) tells us that equilibrium cannot be consistent with two workers being promoted if and only if it is efficient to promote both workers. To see this, suppose there was an equilibrium that had this property and there was a firm with realizations for worker ability equal to $\theta_1=\theta'+\varepsilon$ and $\theta_2=\theta'+2\varepsilon$. In this case the left hand side of equation (8) reduces to $3d_2\varepsilon$ while the right hand side reduces to $2d_1(\theta'-\theta_L)+3d_1\varepsilon$. So for ε sufficiently small equation (8) is not satisfied. In other words, there cannot be an equilibrium in which two workers are promoted if and only if this is efficient because, starting from such an outcome, a firm would have an incentive to deviate and promote no one if the two realizations for θ were both above but sufficiently close to θ'.

Note that the above discussion concerns the case N=2. But it is easy to generalize the above discussion to show that for any N the following must be true. First, a single worker cannot be promoted in equilibrium. So, if a firm's first period employees have realizations for θ such that it is efficient to promote just one worker, then for this firm promotion decisions will not be efficient. Second, for any n, $2\leq n\leq N$, promotion decisions where exactly n workers are promoted will not be fully efficient. Third, as captured in the proposition, the magnitude of θ_H (holding all other parameters fixed) determines whether or not there are equilibria with a strictly positive probability of promotions.

So, in summary, having the ability to signal the number or proportion of workers promoted can improve the efficiency of the promotion process. Specifically, in the case of no human capital investments, this change can move equilibrium from one in which inefficiency is severe because there are no promotions to one in which inefficiency is reduced in the sense that there is a strictly positive frequency of promotions. But this change never results in the promotion process being fully efficient.

A final point to consider in this section is the issue of commitment. In our analysis we assume that a firm cannot commit at the beginning of the game to the number of workers who will be promoted at the beginning of period 2. But possibly some limited type of commitment is posible. We have decided for length reasons not to include a formal analysis of this possibility. But it is worthwhile pointing out that giving firms some or even substantial commitment ability will not result in fully efficient promotion decisions in this model. For example, suppose each firm at the beginning of period 1 could commit to a minimum number of workers it will promote at the beginning of period 2. Since the efficient number of workers to promote will vary with the realizations of worker ability levels and there is always a strictly positive probability this efficient number will be zero, this type of commitment ability may improve the efficiency of the promotion process but cannot result in fully efficient promotion decisions.

IV. SLOT CONSTRAINTS

In Section III's model where a firm could hire multiple young workers in period 1 there were no slot constraints. Specifically, in period 2 the firm faced no constraint concerning how many workers it could promote. But in many real world firms managerial positions are characterized by slot constraints that limit the number of workers who can be assigned to managerial positions. In a setting where this type of slot constraint is important the standard promotion signaling distortion that concerns the number of promotions can disappear because the firm has no discretion concerning how many workers are promoted. Here we show that, even in this type of setting, there can still be a promotion distortion that concerns who is promoted.

A) The Model

In this section we consider a two-period model in which there are F firms, where each firm hires N or zero young workers in period 1 (there is no constraint on the number of old workers employed in period 2) and production functions are the same as in Section II. Workers can also produce in self-employment where output in self-employment is independent of ability

and equals z for workers in the first period of self-employment and z′, z′>z, for workers in their second period of self-employment, i.e., self-employment exhibits learning-by-doing.[22] We further assume that z and z′ are such that firms find it profitable to hire young workers in period 1 and between periods a worker never switches between self-employment and working at a firm.

There are also three other new assumptions that further distinguish this model from the earlier ones considered. First, there is a managerial slot constraint. That is, in each firm there is a single managerial job or level 2 position. Second, due to the importance of firm specific human capital for the managerial position, only an old worker who previously worked at a firm can staff the firm's managerial position. Note that this means that in period 1 the managerial position is left unfilled.

Third, there are S, S≥1, schooling levels where the ability of worker i with schooling level s_i equals $B(s_i)+\theta_i$. We assume B′>0, while θ_i is a random draw from a probability density function f(.) which does not vary across schooling groups and which has the same properties as previously. Given that the schooling level is not a choice variable, the schooling level does not serve as a traditional signal in our model. But, since B′>0, firms correctly believe that the ability distribution does vary with schooling and, in particular, both average ability and the range of abilities varies positively with the schooling level. We assume $B(1)+\theta_H>B(S)+\theta_L$. This means that the highest ability old worker at a firm can be a worker from any of the schooling groups.[23] Also, there are m(s) workers in schooling group s, where the total number of workers exceeds NF. Further, $c_2+d_2(B(1)+\theta_L)>c_1+d_1(B(1)+\theta_L)$ which ensures it is always profitable for a firm to staff the managerial position in period 2.

Note that the model is characterized by a number of simplifying assumptions that reduce the scope for a promotion distortion. In particular, the number of young workers a firm hires in

[22] Given our assumption that firms hire either N or zero young workers in each period, assuming free entry raises a number of complications. So instead we assume a fixed number of firms and a self-employment option which allows us to avoid these complications.

[23] Our assumption that the ability of worker i with schooling level s_i equals $B(s_i)+\theta_i$ is not essential for our results. The qualitative nature of the results would be unchanged as long as the minimum worker ability is a strictly positive function of the schooling level.

period 1 is fixed and the single managerial position must be staffed by promotion from within. These assumptions reduce the scope for a promotion distortion because with these assumptions the number of old workers promoted is fixed. Starting with Waldman (1984a) the standard promotion signaling distortion is that too few workers are promoted or promotion is delayed, but that type of distortion cannot arise here. Our main result in this section is that, even though this is the case, promotion decisions are not fully efficient when there are multiple schooling groups.

B) Analysis

We begin by considering how the model works in the case of symmetric learning. Given our assumption that a young worker cannot produce in a level 2 position, in period 1 NF young workers with the highest values for expected ability (highest schooling levels) are hired by firms and assigned to job 1. Further, young worker i with schooling level s_i hired by a firm is paid a wage $w_Y(s_i)$ that satisfies $w_Y(s_i)>c_1+d_1[B(s_i)+E(\theta_i)]$. The reason the wage exceeds expected first period productivity is that, as described below, for each worker there is a strictly positive probability the worker's productivity in period 2 will exceed the worker's wage. So, in anticipation of this, competition for young workers drives the wage above expected productivity.

For old workers, because learning is symmetric, wages depend on worker ability but not directly on the schooling level or the job assignment (the wage does depend indirectly on the schooling level since the schooling level helps determine worker ability). Specifically, since an old worker who switches employers must be assigned to job 1, the wage for each old worker i with schooling level s_i equals the worker's productivity at job 1, i.e., $w_O(\theta_i,s_i)=c_1+d_1[B(s_i)+\theta_i]$.

Further, since an old worker's pay is independent of the worker's job assignment, firms assign old workers to jobs efficiently. That is, given that the return to ability is higher in job 2 than in job 1, in period 2 a firm assigns the highest ability old worker to job 2 and keeps all the other old workers in job 1 (as in the previous models, there is no turnover in equilibrium). In other words, in this benchmark analysis the probability of promotion rises with a worker's schooling level, but this is solely because expected ability increases with schooling and not

because of any other advantage associated with schooling.[24]

We now turn our attention to what happens when learning is asymmetric rather than symmetric. With asymmetric learning the promotion decision will not be efficient because wages for old workers will depend on ability, schooling, and job assignment, as opposed to just ability. In order to illustrate the central role of schooling in inefficient promotion decisions, we start with the case of a single schooling group, i.e., S=1. This case is analyzed in Proposition 5. Note, below $w_Y(s)$ again denotes the young worker wage for workers in schooling group s, while $w_{O,j}(\theta_i,s_i)$ is now the old worker wage for worker i with schooling level s_i assigned to job j.

Proposition 5: If S=1, then i) through iii) describe equilibrium behavior.

i) Each firm hires N young workers in period 1 and the remaining young workers are self-employed and remain in self-employment when old. Also, $w_Y(1)$ is such that young workers are indifferent between self-employment and working at a firm, i.e., the expected payment over a worker's two-period lifetime of working at a firm equals $z+z'$.

ii) In period 2 each old worker i employed at a firm when young stays with the first period employer and is paid $w_{O,j}(\theta_i,1)=c_1+d_1[B(1)+\theta_L]$.

iii) In period 2 each firm assigns the old worker in its employ with the highest ability to job 2 and the remaining old workers are assigned to job 1.

Proposition 5 has a number of interesting results. First, wages are again determined by the winner's curse. Specifically, the wage for old workers employed at firms equals the productivity in job 1 of the lowest ability worker, where this is the case both for workers assigned

[24] One question concerning this model is whether firms will hire workers of the same schooling level or hire a more heterogeneous set of workers. Although we do not show it formally, the answer is that under both symmetric and asymmetric learning firms have incentives to hire heterogeneous workforces. The reason is that in period 1 a young worker with high expected ability, i.e., a high schooling level, is more valuable to a firm without other young workers with high expected ability because there is a higher probability the worker will be promoted into the managerial job in period 2. Also, a related argument yields that young workers with high expected ability will typically prefer to work at firms that employ few other young workers with high expected ability.

to job 1 and those assigned to job 2. Second, a consequence of this result is that, even if $d_1>0$, the wage for a promoted old worker is the same as the wage for an old worker assigned to job 1. This is in contrast to the Proposition 1 result concerning the wage for promoted workers given $d_1>0$ and no slot constraints. There a promoted worker received a higher wage because of the signal associated with promotion. But that does not arise here when S=1 because it is possible that a promoted worker has an ability level infinitesimally close to $B(1)+\theta_L$ and combining this with the winner's curse yields that promoted old workers and non-promoted old workers are paid the same wage. Third, both because firms have no discretion concerning the proportion of workers promoted and because promoted and non-promoted workers receive the same wage, there is no incentive to distort the promotion decision and the promotion rule is the efficient one.

We now consider equilibrium behavior given multiple schooling groups, i.e., S>1. The main result is that the promotion wage for old workers in lower schooling groups exceeds the non-promotion wage and, as a result, there is a promotion distortion.

Proposition 6: If S>1, then i) through vi) describe equilibrium behavior.

i) Each firm hires N young workers in period 1 and the remaining young workers are self-employed and remain in self-employment when old.

ii) There exists a schooling level s', $1\leq s'\leq S$, such that all young workers in each schooling group s^+, $s^+>s'$, are employed at firms, all workers in each schooling group s^-, $s^-<s'$, are self-employed, and some (maybe all) old workers in group s' are employed at firms.

iii) The expected payment over a worker's two-period lifetime of working at a firm for a worker in schooling group s' equals $z+z'$, while this expected payment is at least this amount for a worker in schooling group s, $s\geq s'$.

iv) In period 2 each old worker employed at a firm when young stays with the first period employer and each such old worker i with schooling level s_i assigned to job 1 is paid $w_{O,1}(\theta_i,s_i)=c_1+d_1[B(s_i)+\theta_L]$.

v) In period 2, if $d_1=0$, then old worker i with schooling level s_i assigned to job 2 is paid

$w_{O,2}(\theta_i,s_i)=w_{O,1}(\theta_i,s_i)=c_1+d_1[B(s_i)+\theta_L]=c_1$ and in each firm the old worker assigned to job 2 is the one with the highest ability, i.e., promotion decisions are efficient.

vi) In period 2, if $d_1>0$, then old worker i with schooling level s_i, $s_i<S$, assigned to job 2 is paid $w_{O,2}(\theta_i,s_i)>w_{O,1}(\theta_i,s_i)=c_1+d_1[B(s_i)+\theta_L]$, while $w_{O,2}(\theta_i,S)=w_{O,1}(\theta_i,S)=c_1+d_1[B(S)+\theta_L]$. Also, for some realizations of worker abilities the promoted worker has higher schooling but lower ability than a worker not promoted, i.e., promotion decisions are not fully efficient.[25]

Proposition 6 tells us that when there are multiple schooling groups there is a similarity between this model and the model analyzed in Section II. In particular, in both models promotion decisions are efficient when $d_1=0$ but there is a promotion distortion when $d_1>0$. We start by discussing the case $d_1=0$. The first important result in this case is that the wage for an old worker in each schooling group s is independent of whether the worker is assigned to job 1 or job 2. The logic is that promoting an old worker sends a positive signal about the worker's ability, but because in this model prospective employers cannot assign the worker to job 2 they are only willing to bid what the worker could produce in job 1. In turn, given $d_1=0$, the positive signal about ability has no effect on the market wage offer with the result that promoted and non-promoted old workers are paid the same wage.

The second important result concerning the case $d_1=0$ is that promotion decisions are efficient, i.e., the promoted old worker at each firm is the one with the highest ability. This result follows from the first result concerning wages. That is, since whether or not an old worker is promoted has no effect on the worker's wage, the total wage bill is not a factor in the promotion decision. This means the promotion decision is determined by the choice that maximizes current

[25] vi) is written to be consistent with a firm that has at least a single old worker of each schooling level s_i, $s_i \geq s'$, and where $s'<S$. If these conditions are not satisfied but not all old workers at the firm have the same schooling level, then it is still true that for some realizations of worker abilities promotion decisions are not efficient. See footnote 24 for a related discussion.

output and, since high ability is more valuable in the managerial or job 2 position, the firm promotes the old worker with the highest ability just like in the symmetric learning case.

We now consider the case $d_1>0$. The first main result in this case is that for each schooling group other than the top one, S, the wage for a promoted old worker is higher than the non-promotion wage. This is just the standard promotion signaling result. To see the logic here, note first that for an old worker of schooling level S the promotion and non-promotion wages are the same. This is because a worker of schooling level S with ability equal to to $B(S)+\theta_L$ will be promoted if the other workers have low enough ability levels (if the firm employs multiple workers of schooling level S the worker will have a strictly positive probability of promotion given an ability equal to $B(S)+\theta_L+\varepsilon$, for any $\varepsilon>0$). As a result, due to the winner's curse, both promoted and non-promoted old workers of schooling level S are paid $c_1+d_1[B(S)+\theta_L]$. But the same logic does not hold for old workers of schooling level s_i, $s_i<S$. That is, a worker of schooling level s_i, $s_i<S$, with ability close to $B(s_i)+\theta_L$ has a zero probability of promotion because a worker from schooling group S has higher ability for sure, so the firm would prefer to promote the group S worker (see footnote 25). So promotion of a group s_i worker means the worker has a value for θ_i strictly above θ_L which, in turn, via the winner's curse means the promotion wage strictly exceeds the non-promotion wage.

The second main result is that when $d_1>0$ promotion decisions are not fully efficient, where the result follows from what we know about wages just discussed. Suppose that the two highest ability old workers at a particular firm are from schooling group s_i, $s_i<S$, and S. As just discussed, if the firm promotes the worker with schooling level S there is no effect on the firm's total wage bill since the worker receives $c_1+d_1[B(S)+\theta_L]$ whether or not the worker is promoted. But if the firm promotes the worker with schooling level s_i, then the worker receives a higher wage meaning the firm's total wage bill increases. What this means is that, if the worker with the higher ability is the worker with schooling level s_i but abilities are similar, then the worker with schooling level S will be promoted even though he or she is of lower ability.

Notice that the nature of the promotion distortion here and the logic for why there is a distortion is related to the reason for promotion inefficiencies in earlier sections, but there are some differences. In earlier sections, as in Waldman(1984a), promotion serves as a signal of high worker ability which drives up the wage, so a firm only promotes a worker if the worker's productivity on the high level job is significantly more than productivity on the low level job. The result is that the frequency or probability of promotion is inefficiently low.

In this section's model the promotion distortion is not that the frequency of promotion is too low since the frequency of promotion is fixed at 1/N in this model. Rather, the distortion is in terms of who is promoted. The signaling effect of promotion on wages varies with the worker's schooling level. Specifically, as captured in the statement of the proposition and the above discussion, promotion signaling has no effect on wages for the highest schooling group but for lower levels of schooling a promotion causes an increase in the wage. As a result, the decision concerning who to promote is not fully efficient. In order to lower its costs, a firm will sometimes promote a worker of lower ability but a higher schooling level because the wage premium associated with promotion is lower for this worker.[26]

A final point concerning the argument in this section is that it can be generalized beyond the idea that varying education levels can result in inefficiencies concerning who is promoted. The more general point is that, if workers vary for almost any reason in the extent to which signaling causes wage increases upon promotion, firms will have an incentive to distort the promotion decision in favor of workers for whom this signaling effect is smaller. For example, consider two workers who vary in terms of a moving cost associated with switching employers where this difference between the workers is publicly observable. The positive signal associated

[26] The result is related to theoretical findings in Bernhardt (1995) and DeVaro and Waldman (2012). Those papers build on Waldman's (1984a) analysis by introducing workers of varying schooling levels. They show that higher levels of schooling reduce the wage premium due to promotion signaling with the result that the inefficient reduction in the probability of promotion is smaller for workers with high education levels. In our analysis the wage premium due to promotion signaling similarly varies with education. But instead of having an effect on the severity of the inefficiency concerning the probability of promotion, the result is that the wrong worker is sometimes promoted into the single managerial position.

with promotion would likely result in a higher wage increase for the worker with the lower moving cost. So, even if the worker with the lower moving cost had a very low probability of leaving after a promotion, the worker's employer would have an incentive to favor the worker with the high moving cost in the promotion decision because of the smaller increase in the firm's total wage bill. In other words, just like in our formal analysis above, promotion decisions would not be fully efficient because the firm would inefficiently favor the worker for whom the promotion wage increase due to signaling is smaller.

V. DISCUSSION

In Section II through IV we explored a series of models characterized by promotion signaling in order to understand the extent to which promotion signaling results in a distortion of the promotion decision. In each analysis we found that, depending on the parameterization, promotion decisions can be fully efficient or not. For example, along the lines of Golan (2005), in the models considered in Sections II through IV, the promotion decision was fully efficient when worker ability had no effect on productivity in the low level job. But when productivity on the low level job increased with worker ability, then in each of those models a promotion distortion could arise in equilibrium. We also found in Section IV that having multiple education groups can be important. That is, in that model there is no promotion distortion when all workers at a firm have the same education level, but when there are multiple education groups then a promotion distortion arises when productivity on the low level job increases with worker ability.

Based on the importance of whether or not ability affects productivity in the low level job for the existence of a promotion distortion in all of our models, one might be tempted to conclude that the existence of a promotion distortion requires ability to positively affect productivity on the low level job. But this is obviously incorrect. The promotion signaling distortion was first identified in Waldman (1984a) and in the models investigated in that paper ability had no effect on productivity in the low level job. So assuming that worker ability has no

effect on productivity in the low level job is not sufficient to guarantee there will not be a promotion distortion due to signaling.

But a comparison of the analysis in Waldman (1984a) with the analyses in Golan (2005) and in Section II can be used to identify the key feature required for a promotion signaling distortion. In the wage determination process for old workers in Waldman (1984a), a worker's current employer first offered a wage/job assignment pair, prospective employers then observed the job assignment and offered a wage, and then the worker chose a firm. In the resulting equilibrium, being promoted to the high level job served as a signal of high ability which led prospective employers to offer high wages to promoted workers. In turn, the current employer paired a high wage with promotions to stop workers from being bid away but, at the same time, reduced the probability of promotion below the efficient level in order to avoid paying the high promotion wage when the worker was not sufficiently more productive in the high level job.

Now consider the analysis in Golan (2005) and in Section II when $d_1=0$. The difference in terms of assumptions is the wage determination process for old workers. In particular, in those analyses the current employer announces a job assignment, each prospective employer observes the job assignment and offers a wage, the current employer makes a wage counteroffer, and then workers choose firms. One result is a winner's curse where the old workers do not move and a worker's wage equals the lowest possible productivity at a prospective employer of any worker with the same labor market history. Further, promotions serve as signals of high worker ability in the sense that prospective employers correctly believe that promoted workers are of higher expected ability than non-promoted workers. But in the case $d_1=0$ this does not translate into a promotion wage that exceeds the non-promotion wage. Rather, the promotion and non-promotion wages are the same and, because they are the same, a firm only considers productivity when deciding who to promote and thus promotion decisions are fully efficient.

This difference is the key feature that explains whether or not there is a promotion distortion in each of our analyses. That is, whenever promotion signaling results in promoted workers receiving a higher wage than non-promoted workers, then firms have an incentive to

distort the promotion decision to reduce the cost of this promotion wage premium. But when the signal does not result in a high promotion wage, then firms base their promotion decisions solely on productivity and promotion decisions are fully efficient. In Golan (2005) and in Sections II and III when $d_1=0$, promotion signaling did not result in a promotion wage increase and there was no distortion. But in Waldman (1984a) and in Sections II and III with $d_1>0$, promotion signaling did result in a a promotion wage increase and firms promoted an inefficiently small number of workers.[27,28] Similarly, in Section IV when $S=1$ and/or $d_1=0$, there was no promotion wage increase due to signaling and promotion decisions were fully efficient. But when $S>1$ and $d_1>0$, then signaling did result in a promotion wage increase and the promotion decision was biased towards workers with the highest level of education.

In summary, our analysis indicates that promotion distortions due to signaling can arise as long as the signal results in a higher promotion wage. In Sections II through IV, whether or not the models analyzed met this condition depended on specific properties of the specifications. But, in general, the more realistic specifications, i.e., ability positively affecting productivity on the low level job and multiple schooling groups, are the ones where signaling does have a positive effect on the promotion wage and thus the specifications characterized by a promotion distortions. It also seems from a real world perspective quite realistic that, if promotion signals

[27] The reason that whether or not d_1 is strictly positive is important for a higher promotion wage due to signaling was discussed in Section II. The basic idea is that, because of the counteroffer assumption and the winner's curse, the wage paid to each old worker equals the productivity at a prospective employer of the worst worker with the same labor market signal. The worst promoted worker is higher ability than the worst non-promoted worker. When $d_1>0$ this results in a higher wage for promoted workers since, independent of the job assignment at a prospective employer, the worst promoted worker is more productive than the worst non-promoted worker. But when $d_1=0$ the worst promoted worker has the same productivity in job 1 as the worst non-promoted worker. In turn, what happens in equilibrium is that the worst promoted worker is efficiently assigned to job 1 (to be precise, this worker's productivity is the same in the two jobs), so promoted workers earn the same amount as non-promoted workers.

[28] One exception is Section II's analysis with human capital training for parameterizations for which it was efficient for all promoted workers to be trained. In that case promotions did result in wage increases but there was no promotion distortion. The logic was that after subtracting off the training fee the worker's net compensation was independent of whether or not the worker was promoted. So the firm had no incentive to distort the promotion decision in order to avoid the higher wage associated with promotion.

higher ability, the result would be a higher wage. So we believe that our analysis shows that theory supports the importance of a promotion signaling distortion.

VI. CONCLUSION

Starting with Waldman (1984a), a large literature has investigated the signaling role of promotions. Many of the papers in this literature find that the signal is accompanied by promotion inefficiencies. But in some of the papers there is no promotion distortion and it is argued that inefficiencies are not a general result in promotion signaling models. In this paper we investigated the robustness of the promotion signaling distortion to different ways of modeling the promotion process.

We considered three different models in which asymmetric learning in the labor market leads to a promotion serving as a signal of high worker ability: i) a two-period model characterized by counteroffers; ii) an extension of our first framework in which firms hire multiple young workers and the number promoted is publicly observable; and iii) a two-period model characterized by slot constraints and multiple education groups. In each case we found that in the most realistic specifications promotions served as a signal, signaling led to higher promotion wages, and the higher promotion wages frequently led to distortions of the promotion decision. For example, in each of our first two analyses this was the result when worker ability had a positive effect on productivity in the low level job, while the distortion was not part of equilibrium behavior given the unrealistic assumption that this is not the case. Also, in our third model these properties describe equilibrium behavior given multiple education levels, but there was never a distortion under the unrealistic assumption of a single education level.

We also identified the key property necessary for a promotion signaling model to exhibit a promotion distortion. We found that when the signaling role of promotion results in a wage increase upon promotion, then there is typically a promotion distortion because in making promotion decisions firms have an incentive to avoid the wage increase due to signaling. In contrast, when signaling does not result in promotion wage increases, then firms focus solely on productivity in making promotion decisions with the result that promotion decisions are fully

efficient. This perspective explains when we see and do not see promotion distortions in all the analyses in this paper and also explains the existence or non-existence of promotion distortions in earlier papers such as Waldman (1984a) and Golan (2005).

In terms of future research, one topic that we feel deserves more attention is investigating competitive responses to promotion distortions. If the labor market is competitive, then contracts between workers and firms should minimize inefficiencies which from the standpoint of the analyses here means minimize promotion distortions due to signaling. A few papers have investigated this idea. For example, in Waldman's (1984a) original paper on promotion signaling he considered the extent to which commitments to future promotion and non-promotion wages can reduce promotion distortions, while more recently Mukherjee and Vasconelos (2013) investigate the extent to which break-up fees can be used to reduce the distortion. We believe, however, that these are just two possibilities among many and that more attention to competitive responses to promotion signaling distortions is warranted.

APPENDIX

Proof of Proposition 1: We start with period 2. Consider first wages. For actions on the equilibrium path, because the initial employer can make counteroffers, other firms are willing to offer a worker assigned to job j the worker's minimum possible output at one of these other firms which is based on when the initial employer assigns the worker to job j in equilibrium. In turn, given the tie-breaking rules assumed, the initial employer just matches these offers and then the worker stays with the initial employer.

Now consider period 2 job assignments. Since output rises faster with ability on job 2 than on job 1, there must be a value θ^+ such that a worker's initial employer assigns the worker to job 2 if $\theta_i>\theta^+$ and assigns the worker to job 1 if $\theta_i<\theta^+$ (if the worker is assigned to job 1 with probability one we will say $\theta^+=\theta_H$ while $\theta^+=\theta_L$ refers to the case where the worker is assigned to job 2 with probability one).

Suppose $\theta^+=\theta_L$. Consider the return to promoting the worker when $\theta_i=\theta_L+\gamma$, γ small. The extra productivity associated with such a promotion equals $[c_2+d_2(\theta_L+\gamma)]-[c_1+d_1(\theta_L+\gamma)]$ which is strictly negative for γ close to zero. Starting from a situation in which $\theta^+=\theta_L$, when the off-the-equilibrium path action of the worker not being promoted is observed by the market, the inference is that the worker's ability is θ_L (this follows from our assumption that beliefs concerning off-the-equilibrium path actions are consistent with each such action being taken by the type with the smallest cost of choosing that action). The extra cost of promoting the worker is therefore zero. Thus, since the extra cost of promoting the worker exceeds the extra productivity, the firm will not promote the worker so we have a contradiction. Hence, $\theta^+>\theta_L$.

Suppose $\theta_L<\theta^+<\theta_H$. Then θ^+ is the value for θ_i such that the firm is indifferent between assigning the worker to jobs 1 and 2. In this case θ^+ satisfies (A1).

(A1) $$[c_1+d_1\theta^+]-[c_1+d_1\theta_L]=[c_2+d_2\theta^+]-\max\{c_1+d_1\theta^+, c_2+d_2\theta^+\}$$

Suppose $d_1>0$. Then the left hand side of (A1) is strictly positive while the right hand side is weakly negative. So, if $d_1>0$, θ^+ is not in the interval (θ_L, θ_H). Given our earlier result, we have that in this case $\theta^+=\theta_H$, i.e., no one is promoted. This proves iii).

Suppose $d_1=0$. Then the left hand side of (A1) equals zero while the right hand side equals zero for any value for θ^+ that satisfies $\theta^+\geq\theta'$. So there are multiple equilibria where the equilibria differ in terms of the value for θ^+. Focusing on equilibria that minimize inefficiencies yields $\theta^+=\theta'$. The reason is that this outcome is characterized by no inefficiencies in job assignments. This proves i).

Now consider job assignments and wages in period 1. Given that from above we know that each employer in period 1 earns positive expected profits in period 2, competition means that w_Y must exceed expected productivity in period 1. We also know that, given $c_1+d_1E(\theta)>c_2+d_2E(\theta)$, in period 1 all workers are assigned to job 1. Combining this result with the previous result yields $w_Y>c_1+d_1E(\theta)$. This proves i).

Proof of Proposition 2: i) follows using the same logic as in the proof of Proposition 1. Given this, consider period 2 and the case $d_1=0$. As argued in the proof of Proposition 1, given our focus is on equilibria that minimize inefficiencies, if job assignments and training decisions being made optimally is an equilibrium, then all equilibria must satisfy the condition. Suppose each firm that hired workers in the first period promotes (does not promote) a worker when $\theta_i>(<)\theta''$, trains (does not train) the worker when $\theta_i>(<)\theta^*$, the training fee always equals zero, outside firms offer c_1 to every worker, the initial employer responds by offering c_1 to every worker, and the worker always stays.

Investigation of best responses yields this is an equilibrium. First, if the market offers c_1 and the initial employer responds by offering c_1, then by assumption the worker stays. Second, given the promotion and training rules specified, if the market offers c_1, it is always a best response for the initial employer to respond with c_1 since c_1 is the minimum wage offer required for the worker to stay and the worker's productivity at the initial employer is always at least equal to c_1. Third, given the promotion and training rules specified, the worst worker either not promoted, promoted and not trained, or promoted and trained has a productivity in an outside firm equal to c_1 so the market wage offer is always c_1. Fourth, given the market wage offer is always c_1, the initial employer always matches, and the worker never leaves, the firm will choose to promote and train efficiently and the training fee for workers who are trained must equal zero since training does not increase the wage a worker receives. Finally, one can also show using related arguments that there is no equilibrium in which promotion and training decisions are made efficiently that does not satisfy the specified strategies. This proves ii).

Now consider period 2 and the case $d_1>0$. The argument used in the proof of Proposition 1 to show that there are no promotions when $d_1>0$ now yields that there cannot be workers promoted who are not trained. We also know that since training is costly and does not increase productivity in job 1, no workers will be trained and then not promoted.

Let θ^+ now denote the lowest ability worker trained and promoted. Using arguments similar to the ones used to show that $\theta^+>\theta_L$ in Proposition 1, we can show that here $\theta_L<\theta^+<\theta_H$.

Given this, the winner's curse yields that the market offers $\max\{c_1+d_1\theta^+, c_2+d_2\theta^+\}$ to promoted and trained workers, the first period employer matches, and the worker stays. But anticipating this the firm sets the training fee equal to $\max\{c_1+d_1\theta^+, c_2+d_2\theta^+\}-(c_1+d_1\theta_L)$, so the net payment made by the firm equals $c_1+d_1\theta_L$. The firm's choice of θ^+ therefore satisfies (5). Comparing (3) and (5) yields that $\theta^+=\theta''$ when $\theta^*\leq\theta'$, but comparing (5) to the definition of θ'' yields $\theta^+>\theta''$ when $\theta^*>\theta'$. This proves iii).

Proof of Proposition 3: i) follows using the same logic as in the proof of Proposition 1. Now consider period 2 and the case $d_1=0$. The basic argument here is the same as in the proof of Proposition 2. Given our focus is on equilibria that minimize inefficiencies, any equilibrium must be fully efficient if there exists an equilibrium that is fully efficient. As in the proof of Proposition 2, an examination of best responses shows that such an equilibrium exists. What is different here is the training fee. If a worker is promoted and trained, then the wage equals $c_2+d_2\theta^*+\Delta(\theta^*)$ while a worker either not promoted or promoted and not trained is paid c_1. So in period 2 first period employers have an incentive to extract all potential period 2 surplus by offering a training fee equal to $(c_2-c_1)+d_2\theta^*+\Delta(\theta^*)$ to all workers for whom promotion and training is efficient. Finally, as is the case for ii) of Proposition 2, one can also show using related arguments that there is no equilibrium in which promotion and training decision are made efficiently that does not satisfy the specified strategies. This proves ii).

Now consider period 2 and the case $d_1>0$. As was true in the proof of Proposition 2, the argument used in the proof of Proposition 1 to show that there are no promotions when $d_1>0$ now yields that there cannot be workers promoted who are not trained. We also know that since training is costly and does not increase productivity in job 1, no workers will be trained and then not promoted.

Let θ^+ now denote the lowest ability worker trained and promoted. Using arguments similar to the one used to show that $\theta^+>\theta_L$ in the proof of Proposition 1 we can show here that $\theta_L<\theta^+<\theta_H$. Given this, the winner's curse yields that the market offers $\max\{c_1+d_1\theta^+, c_2+d_2\theta^+$

$+\Delta(\theta^+)\}$ to promoted and trained workers, the first period employer matches, and the worker stays. But anticipating this, the firm sets the training fee equal to $\max\{c_1+d_1\theta^+, c_2+d_2\theta^++\Delta(\theta^+)\}-(c_1+d_1\theta_L)$, so the net payment made by the firm equals $c_1+d_1\theta_L$. The firm's choice of θ^+ therefore satisfies (5). Comparing (3) and (5) yields that $\theta^+=\theta''$ when $\theta^*\leq\theta'$, but comparing (5) to the definition of θ'' yields $\theta^+>\theta''$ when $\theta^*>\theta'$. This proves iii).

Proof of Proposition 4: i), ii), and iii) follow from arguments in the proof of Proposition 1. Now suppose $d_1>0$ and $N>1$. A variant of the argument that shows that there are no promotions when $N=1$ yields that a single worker cannot be promoted when $N>1$. Given this, suppose that in equilibrium there is a strictly positive probability a firm promotes n workers, $1<n\leq N$. There must be a lowest ability level corresponding to the firm promoting n workers. Call this ability level θ_n^+. The winner's curse yields that when n workers are promoted the market will offer each worker $\max\{c_1+d_1\theta_n^+, c_2+d_2\theta_n^+\}$, the initial employer will match this wage, and the n workers will all stay.

For a firm to promote n workers in equilibrium it must earn higher profits from this action than promoting no one. If promoting no one is an on-the-equilibrium path action, then since the worst worker to promote is a θ_L worker it must be the case that some realizations of abilities in which no one is promoted has at least one worker be a θ_L worker. So the winner's curse means that when no one is promoted the market wage is $c_1+d_1\theta_L$, the initial employer matches, and all the workers stay. If promoting no one is an off-the-equilibrium path action, then our assumption about beliefs concerning off-the-equilibrium path actions yields that when no one is promoted the market wage is $c_1+d_1\theta_L$, the initial employer matches, and all the workers stay.

So, for a firm to find it profitable to promote n workers when θ_n^+ is the lowest ability worker promoted it must be the case that (A3) holds. Note, below θ_{n-1}^A is the average ability level of the n-1 other workers.

(A3) $$(c_2+d_2\theta_n^+)+(n-1)(c_2+d_2\theta_{n-1}^A)-n\max\{c_1+d_1\theta_n^+, c_2+d_2\theta_n^+\}$$
$$\geq(c_1+d_1\theta_n^+)+(n-1)(c_1+d_1\theta_{n-1}^A)-n(c_1+d_1\theta_L)$$

For any fixed value for θ_n^+, $\theta_n^+ \leq \theta_{n-1}^A$, if $\theta_{n-1}^A = \theta'$, then the left hand side of (A3) is strictly less than the right hand side. In turn, since θ_{n-1}^A cannot exceed θ_H, both sides of (A3) are continuously increasing in θ_{n-1}^A, and $d_2 > d_1$ so the left hand side increases faster with θ_{n-1}^A, there exists a smallest value for θ_H, holding all other parameters fixed, such that (A3) cannot be satisfied for any feasible $(\theta_n^+, \theta_{n-1}^A)$ pair given θ_H is strictly less than this value. There is such a value for every n, $2 \leq n \leq N$. Call the smallest of these values θ_H^*. We now have that, if $\theta_H < \theta_H^*$, then in equilibrium a firm cannot promote n, $2 \leq n \leq N$ workers, so the winner's curse means each worker is assigned to job 1, is offered $c_1 + d_1\theta_L$ by both the market and the initial employer, and the worker stays with the initial employer. This proves iv).

Now suppose $d_1 > 0$, $N > 1$, and $\theta_H > \theta_H^*$. We first show there are equilibria that satisfy the description in iv). Suppose each firm hires n young workers in period 1, while in period 2 no one is promoted, the market offers $c_1 + d_1\theta_L$ to all workers, initial employers always match, and all workers stay at their initial employers. Further, the off-the-equilibrium path action of a worker being promoted would be followed by the market offering $c_2 + d_2\theta_H$ which is consistent with our assumption concerning beliefs following off-the-equilibrium path actions. Based on these market wage offers a first period employer would never promote a worker in period 2, so the situation described in iv) is an equilibrium.

Finally, based on the definition of θ_H^* above, if $\theta_H > \theta_H^*$, there exist values for n and θ_n^+ such that (A3) is valid for certain realizations of ability so there will be equilibria characterized by a positive frequency of promotions. Since our focus is equilibria that minimize inefficiencies, we now have that the outcome is characterized by a positive frequency of promotions. But note that we know from earlier that a single worker cannot be promoted in one of these equilibria. Since it is possible that all workers at a firm but one are characterized by $\theta_i < \theta'$ while the single worker is characterized by $\theta_i > \theta'$, it is possible that efficiency requires the promotion of a single worker. Since this cannot happen in equilibrium, none of these equilibria are fully efficient. This proves v).

Proof of Proposition 5: The pair of assumptions that production is profitable and that a firm hires either N or zero young workers in period 1 yields that each firm hires N young workers in period 1. We also assumed that z and z′ are such that a worker who starts in self-employment remains in self-employment when he or she is old, while market clearing requires that $w_Y(1)$ is such that the expected compensation of working at firm over a worker's two-period lifetime equals z+z′. This proves i).

Because $d_2>d_1$ and a firm takes market wage offers as given, a firm will always have an incentive to promote the highest ability old worker. This proves iii).

Since in period 2 a firm assigns the N-1 old workers with the lowest abilities to job 1, the winner's curse yields that the market offers $c_1+d_1[B(1)+\theta_L]$ to the old workers assigned to job 1, the initial employer matches, and all the workers remain with the initial employer. Now consider the worker who is promoted. For any $\varepsilon>0$, there is a strictly positive probability that all the old workers have ability less than $\theta_L+\varepsilon$. So the winner's curse yields that the market also offers $c_1+d_1[B(1)+\theta_L]$ to the promoted worker, the initial employer matches, and the worker remains with the initial employer. This proves ii).

Proof of Proposition 6: The pair of assumptions that production is profitable and that a firm hires either N or zero young workers in period 1 yields that each firm hires N young workers in period 1. We also assumed z and z′ are such that a worker who starts in self-employment remains in self-employment when he or she is old. This yields i).

Our assumption that self-employment productivity does not depend on schooling and/or worker ability while expected productivity at a firm rises with a worker's schooling level means that the NF young workers with the highest schooling levels are employed at firms. There is thus a critical schooling level such that young workers at firms with less schooling are self-employed, those with more schooling work at firms, while some or all of the workers with this critical value for schooling work at firms. This yields ii).

Market clearing requires that a worker with schooling level s′ must be indifferent between working and not working. Given workers are risk neutral and no discounting, this requires that a young worker with schooling level s′ who works at a firm has an expected aggregate compensation over t and t+1 equal to z+z′. Further, expected lifetime compensation for workers with schooling level s, s>s′, must be at least this amount since otherwise the worker would prefer self-employment. This proves iii).

iv) follows using the same logic as in the proof of Proposition 1. That is, because of the winner's curse, a prospective employer will be unwilling to pay more than $c_1+d_1[B(s_i)+\theta_L]$ for any worker with schooling level s_i assigned to job 1 since if it did the period 1 employer would match if and only if the worker's productivity was at least equal to the wage offer. So prospective employers offer $c_1+d_1[B(s_i)+\theta_L]$, the first period employer matches, and the worker stays with the first period employer.

Now consider v). If $d_1=0$ and given that an old worker who switches employers can only be assigned to job 1, independent of the worker's schooling level, a prospective employer is only willing to offer c_1 to a worker assigned to job 2. In turn, by assumption the firm matches and the worker stays. Further, we now have that a firm's total wage bill in period 2 is independent of which old worker is assigned to job 2. Further, given $d_2>d_1=0$, a firm maximizes profits by assigning to job 2 the worker who maximizes output on job 2, i.e., the firm ignores the education level and simply promotes the worker with the highest ability. This proves v).

Now consider vi). Consider firm j that employed in period 1 at least one young worker from each schooling group s_i, $s_i \geq s'$ (see footnote 25). Suppose $w_{O,2}(\theta_i,s_i)=c_1+d_1[B(s_i)+\theta_L]$ for all s_i, $s_i \geq s'$. Then the firm's period 2 wage bill for old workers would be independent of its promotion decision, so it would promote the worker with the highest ability. But then a worker with schooling level s′ could only be promoted if the worker's ability was at least equal to $B(S)+\theta_L>B(s')+\theta_L$. But this is inconsistent with the promotion wage for a worker with schooling level s′ being equal to $c_1+d_1[B(s')+\theta_L]$. So it is not the case that $w_{O,2}(\theta_i,s_i)=c_1+d_1[B(s_i)+\theta_L]$ for all s_i, $s_i \geq s'$.

Suppose $w_{O.2}(\theta_i,s_i)>c_1+d_1[B(s_i)+\theta_L]$ for all s_i, $s_i \geq s'$. For any set of wages that satisfy this condition, there will be realizations of worker abilities for the young workers employed by firm j in period 1 such that each worker's productivity in job 1 is less than the wage. Since one of the workers would have to be promoted, this worker's wage would exceed the worker's productivity in job 1 in which case the wages proposed are inconsistent with how the winner's curse determines wages in this model. Combining this with the previous result and that the winner's curse means that $w_{O,2}(\theta_i,s_i)\geq c_1+d_1[B(s_i)+\theta_L]$, we have that $w_{O,2}(\theta_i,s_i)=c_1+d_1[B(s_i)+\theta_L]$ for some of the relevant s_i while $w_{O,2}(\theta_i,s_i)>c_1+d_1[B(s_i)+\theta_L]$ for other of the relevant s_i.

Suppose $w_{O,2}(\theta_i,S)=c_1+d_1[B(S)+\hat{\theta}]$, where $\hat{\theta}>\theta_L$. Suppose $\theta_i=\hat{\theta}-\varepsilon$ and the realizations for all the θs for the other old workers at the firm are close to θ_L. If ε is sufficiently small, then the firm would promote the worker with schooling level S. This contradicts how the winner's curse determines promotion wages, so $w_{O,2}(\theta_i,S)=c_1+d_1[B(S)+\theta_L]$. But given this and $w_{O,2}(\theta_i,s_i)\geq c_1+d_1[B(s_i)+\theta_L]$ for all relevant s_i, the firm would never promote a worker of schooling level s_i, $s_i<S$, if θ_i is sufficiently close to θ_L. So $w_{O,2}(\theta_i,s_i)>c_1+d_1[B(s_i)+\theta_L]$ for all relevant s_i, $s_i<S$.

Now suppose $c_2+d_2[B(s_i)+\theta_i]=c_2+d_2[B(S)+\theta_S]+\varepsilon$, where θ_i is the realization of θ for the s_i worker and θ_S is the realization of θ for the S worker (and that managerial ability for each of the other old workers in the firm in period 2 is below $B(S)+\theta_S$). Promoting the S worker does not increase the period 2 old worker wage bill since the promotion and non-promotion wages for schooling group S are the same, while promoting the s_i worker would increase the old worker wage bill given earlier results. So for ε sufficiently small the firm would have an incentive to promote the worker from schooling group S even though managerial productivity would be higher if the worker from schooling group s_i was promoted instead. This proves vi).

REFERENCES

Acemoglu, D. and J.S. Pischke (1998), "Why do Firms Train? Theory and Evidence," *Quarterly Journal of Economics*, 113, pp. 79-119.

Baker, G., M. Gibbs, and B. Holmstrom (1994a), "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics*, 109, pp. 881-919.

Baker, G., M. Gibbs, and B. Holmstrom (1994b), "The Wage Policy of a Firm," *Quarterly Journal of Economics*, 109, pp. 921-955.

Barron, J., M. Berger, and D. Black (2006), "Selective Counteroffers," *Journal of Labor Economics*, 24, pp. 385-409.

Belzil, C. and M.L. Bognanno (2010), "The Promotion Dynamics of American Executives," *Research in Labor Economics*, 30, pp. 189-231.

Bernhardt, D. (1995), "Strategic Promotion and Compensation," *Review of Economic Studies*, 62, pp. 315-339.

Bognanno, M. and E. Melero (2012), "Promotion Signals, Age and Education," Working Paper 12-05, Department of Economics, Temple University.

Chang, C. and Y. Wang (1996), "Human Capital Investment Under Asymmetric Information: The Pigovian Conjecture Revisited," *Journal of Labor Economics*, 14, pp. 505-519.

Cassidy, H., J. DeVaro, and A. Kauhanen (2012), "The Signaling Role of Promotions: New Evidence from European Firms," Mimeo, Department of Economics, California State University, East Bay.

DeVaro, J. and A. Kauhanen (2013), "An 'Opposing Responses' Test of Competing Promotion Models, Mimeo, Department of Economics, California State University, East Bay.

DeVaro, J. and M. Waldman (2012), "The Signaling Role of Promotions: Further Theory and Empirical Evidence," *Journal of Labor Economics*, 30, pp. 91-147.

Ghosh, S. and M. Waldman (2010), "Standard Promotion Practices Versus Up-Or-Out Contracts," *Rand Journal of Economics*, 41, pp. 301-325.

Gibbons, R. and L. Katz (1991), "Layoffs and Lemons," *Journal of Labor Economics*, 9, pp. 351-380.

Gibbons, R. and M. Waldman (1999), "A Theory of Wage and Promotion Dynamics Inside Firms," *Quarterly Journal of Economics*, 114, pp. 1321-1358.

Golan, L. (2005), "Counteroffers and Efficiency in Labor Markets with Asymmetric Information," *Journal of Labor Economics*, 23, pp. 373-393.

Greenwald, B. (1979), *Adverse Selection in the Labor Market*, Garland Press, New York.

Greenwald, B. (1986), "Adverse Selection in the Labour Market," *Review of Economic Studies*, 53, pp. 325-347.

Harris, M. and B. Holmstrom (1982), "A Theory of Wage Dynamics," *Review of Economic Studies*, 49, pp. 315-333.

Kahn, L. (2013), "Asymmetric Information Between Employers," IZA Discussion Paper #7147 (Forthcoming *American Economic Journal: Applied Economics).*

Katz, E. and A. Ziderman (1990), "Investment in General Training: The Role of Information and Labour Mobility," *Economic Journal*, 100, pp. 1147-1158.

Lazear, E. (1992), "The Job as a Conept," in W. Bruns, (ed.), *Performance Measurement, Evaluations, and Incentives* (Harvard University Press: Boston, MA), pp. 183-215.

Milgrom, P. and S. Oster (1987), "Job Discrimination, Market Forces, and the Invisibility Hypothesis," *Quarterly Journal of Economics*, 102, pp. 453-474.

McCue, K. (1996), "Promotions and Wage Growth," *Journal of Labor Economics*, 14, pp. 175-209.

Mukherjee, A. and L. Vasconcelos (2013), "On the Trade-Off Between Efficiency in Job Assignment and Turnover: The Role of Break-Up Fees," Mimeo, Department of Economics, Michigan State University.

Myerson, R.B. (1978), "Refinements of the Nash Equilibrium Concept," *International Journal of Game Theory*," 7, pp. 73-80.

Okamura, K. (2011), "The Signalling Role of Promotion in Japan," Discussion Paper No. 1112, Graduate School of Economics, Kobe University.

Owan, H. (2004), "Promotion, Turnover, Earnings, and Firm-Sponsored Training," *Journal of Labor Economics*, 22, pp. 955-978.

Pinkston, J. (2009), "A Model of Asymmetric Learning with Testable Implications," *Review of Economic Studies*, 76, pp. 367-394.

Ricart i Costa, J. (1988), "Managerial Task Assignment and Promotion," *Econometrica*, 56, pp. 449-466.

Rosen, S. (1982), "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, 13, pp. 311-323.

Schoenberg, U. (2007), "Testing for Asymmetric Employer Learning," *Journal of Labor Economics*, 25, pp. 651-692.

Waldman, M. (1984a), "Job Assignments, Signalling, and Efficiency," *Rand Journal of Economics*, 15, pp. 255-267.

Waldman (1984b), "Worker Allocation, Hierarchies, and the Wage Distribution," *Review of Economic Studies*, 51, pp. 95-109.

Waldman, M. (1990), "Up-or-Out Contracts: A Signaling Perspective," *Journal of Labor Economics*, 8, pp. 230-250.

Waldman, M. (2012), "Theory and Evidence in Internal Labor Markets," in R. Gibbons and J. Roberts (eds.), *The Handbook of Organizational Economics* (Princeton University Press, Princeton, NJ), pp. 520-574.

Waldman, M. (2013), "Classic Promotion Tournaments Versus Market-Based Tournaments," *International Journal of Industrial Organization*," 31, pp. 198-210.

Zabojnik, J. (2012), "Promotion Tournaments in Market Equilibrium," *Economic Theory*, 51, pp. 213-240.

Zabojnik, J. and D. Bernhardt (2001), "Corporate Tournaments, Human Capital Acquisition and the Firm Size-Wage Relation," *Review of Economic Studies*, 68, pp. 693-716.

Zax, O. (2012), "Promotion Policy and Firm Size," *Economics Bulletin*, 32, pp. 3347-3356.