NBER WORKING PAPER SERIES

THE PETER PRINCIPLE: PROMOTIONS AND DECLINING PRODUCTIVITY

Edward P. Lazear

Working Paper 8094
http://www.nber.org/papers/w8094

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2001

This research was supported in part by the National Science Foundation. The views expressed herein are those of the author and not necessarily those of the National Bureau of Economic Research.

The Peter Principle: Promotions and Declining Productivity
Edward P. Lazear
NBER Working Paper No. 8094
January 2001
JEL No. J00, J6

ABSTRACT

Many have observed that individuals perform worse after having received a promotion. The most famous statement of the idea is the Peter Principle, which states that people are promoted to their level of incompetence. There are a number of possible explanations. Two are explored. The most traditional is that the prospect of promotion provides incentives which vanish after the promotion has been granted; thus, tenured faculty slack off. Another is that output as a statistical matter is expected to fall. Being promoted is evidence that a standard has been met. Regression to the mean implies that future productivity will decline on average. Firms optimally account for the regression bias in making promotion decisions, but the effect is never eliminated. Both explanations are analyzed. The statistical point always holds; the slacking off story holds only under certain compensation structures.

Edward P. Lazear
Graduate School of Business
Stanford University
Stanford, CA 94305-5015
and NBER
lazear_edward@gsb.stanford.edu

"He wrote such good papers until we gave him tenure."

The quote above reflects a view often held by faculty members about their colleagues. Individuals who are striving for tenure often produce very good work, only to be followed by output far below the pre-tenure level after tenure is awarded. One possibility is that individuals game the system. Knowing that their senior colleagues are going to judge them on the basis of their pre-tenure work, junior academicians put forth extraordinary effort in order to convince their seniors that the long-term research prospects are favorable. After tenure is awarded, the value of effort declines and with it the amount of effort supplied.

Another possibility is that there is no strategic behavior at all. Instead, the decline in productivity may be the natural outcome of a statistical process which has in it regression to the mean. Workers are promoted on the basis of having met some standard. To the extent that output is the sum of both permanent and transitory components, those who meet the standard will have expected transitory components that are positive. The expectation of the post-promotion transitory component is zero, implying a reduction in expected output. Firms that understand the statistical process take this phenomenon into account, but the result remains: Expected output is lower after promotion than before.[1]

[1]It is also true that those who are denied promotion do better after they are turned down than they did before the decision was made, for the same reason.

Furthermore, the individuals who behave strategically do not necessarily put forth more effort than is efficient, or than they do after promotion. In fact, the reverse may be true. The nature of the action depends specifically on the compensation formula offered to the post-promotion workers. If workers are paid on the basis of their output after promotion, then somewhat surprisingly, some workers will strategically underwork before tenure. If, on the other hand, workers receive post-promotion wages that are independent of their post-promotion output, then pre-promotion overproduction is the rule.

The Peter Principle, which states that workers are promoted to their level of incompetence, is one version of this phenomenon. After workers are promoted, they do worse than they did before promotion.[2] More often, the Peter Principle is interpreted in a multi-factor context. Individuals who are good in one job are not necessarily good in the job into which they are promoted. As a result, individuals appear incompetent in the job in which they settle. No further promotions result. To obtain this result, it is merely necessary to make a slight modification in the regression to the mean structure. Here, general ability is combined with a job-specific ability to produce output. Regression to the mean results because positive readings on the job-specific component prior to promotion is

[2]Two recent papers [Fairburn and Malcolmson (2000) and Faria (2000)] on this topic use a very different approach from this paper and from each other. The Peter Principle is a by-product of using promotion to solve a moral hazard problem in Fairburn and Malcolmson. Rather than motivate through money, which induces influence activity, firms choose promotion because then managers must live with the consequences of their decisions. Too many workers are promoted under certain circumstances resulting in a Peter Principle effect. In Faria (2000), workers have two skills. Those who are good at one are necessarily less good at another when on the frontier. Faria argues that this is what is meant by the Peter Principle.

uncorrelated with the job-specific component after promotion.[3]

The fact that promoted individuals are less able than their apparent pre-promotion ability induces firms to adjust in two respects. First, firms select their promotion rule with the understanding that the pre-promotion ability is a biased estimate of true ability for those who exceed some standard. Second, as the variance in the transitory component of ability rises relative to the variance in the permanent component, the length of time over which a promotion decision is made increases. Noisier information results in longer optimal probationary periods.

The model presented below yields the following results:

1. Promoted individuals' performance falls, on average, relative to their pre-promotion performance.
2. Firms that take the decline into account adjust their promotion rule accordingly, but this does not negate the observation that ability declines after promotion.
3. Individuals who know that promotion is based on pre-promotion performance game the system. When their post-promotion decision compensation is based on absolute output, high-ability workers over-produce before the decision is made. More surprisingly, low-ability workers strategically under-produce before the decision is made. When compensation is tournament-like, so that all promoted individuals earn a fixed amount more than those not promoted, all workers over-produce.
4. The length of the pre-promotion period depends on the ratio of transitory variation

[3] The structure is a variant of the Jovanovic (1979a,b) model that was modified and used in a context closer to this structure in Lazear, (1986).

to permanent variation. As the transitory component becomes more important, firms lengthen the pre-promotion period.

A. Model

Let there be two periods. Each worker has a time invariant component of ability, denoted A~f(A), and a time-varying component of ability, denoted ϵ_1 for period 1 and ϵ_2 for period 2. Let the time-varying components be i.i.d. with density $g(\epsilon)$. The firm can observe $A+\epsilon_t$ in each period, but cannot disentangle the time-varying component of ability from the permanent component. There are a variety of interpretations that are consistent with this specification. One can think of the ϵ_t as being a true transitory component or just measurement error. Later, the interpretation of different jobs will be considered.

There are two jobs (two are sufficient), which we denote boss and worker. An individual's productivity in the worker job is given by

$$\alpha + \beta(A+\epsilon_t)$$

and in the boss job is given by

$$\gamma + \delta(A+\epsilon_t)$$

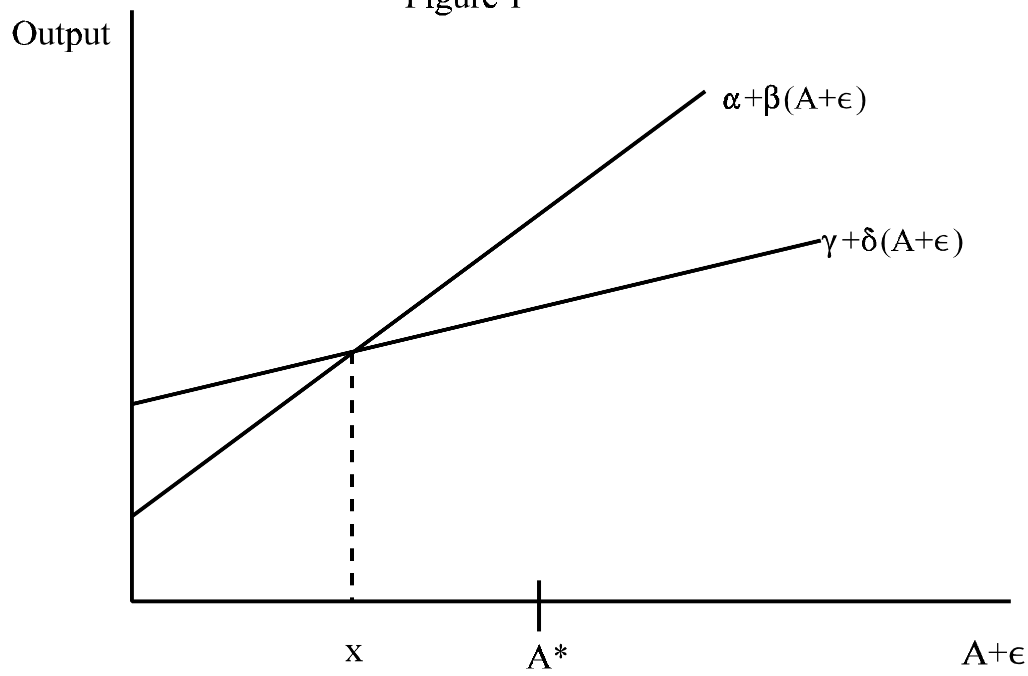
where $\alpha>\gamma$ and $\delta>\beta$. Thus, it pays to assign a worker to the boss's job if and only if

$$A+\epsilon_t > x$$

where

$$x \equiv (\alpha-\gamma) / (\delta-\beta)$$

Figure 1



The situation and the crossing point that correspond to x are shown in figure 1.[4]

[4]This production structure is similar to that used in a comprehensive analysis by Gibbons and Waldman (1999), who also allow for transitory and permanent components with regression. The focus of their paper is earnings and promotion. Neither optimal decision making by firms given the transitory component, nor strategic effort in response to promotion rules are central to their discussion.

Assume that individual ability f(A) is such that in the absence of information, it pays to assign everyone in period 1 to the worker job.[5] Intuitively, this assumption amounts to saying that most people are not boss material and that in the absence of countervailing information, individuals are assigned to the worker job.

After the first period, firms obtain an estimate of A, namely $\hat{A} = A + \epsilon_1$. Since ϵ_1 is the period one transitory component (either measurement error or transitory ability), it is A and not $\hat{A}$ on which a promotion decision should be made. But A is not observed, so firms are forced to base their decision on $\hat{A}$.

1. Workers perform worse after being promoted

Firms must select some criterion level, A*, such that if $\hat{A} > A^*$ the worker is promoted to the boss job. If $\hat{A}$ is less than A*, the worker remains in his current job. It is now shown that workers who are promoted have levels of ability in period 1 that are higher on average than their ability in period 2.

First, note that the expectation of ϵ_1 given that an individual is promoted, is

$$E(\epsilon_1 | A + \epsilon_1 > A^*) = \int_{-\infty}^{\infty} \int_{A^*-A}^{\infty} \frac{1}{1 - G(A^* - A)} \varepsilon\, g(\varepsilon) f(A) d\varepsilon\, dA$$

---

[5]This amounts to assuming that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\alpha + \beta(A+\varepsilon))dGdF > \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\gamma + \delta(A+\varepsilon))dGdF$.

$$= \int_{-\infty}^{\infty} E(\varepsilon \mid \varepsilon > A^* - A)\, f(A)\, dA$$

which is positive since f(A) is positive and the conditional expectation of ϵ given ϵ greater than any number is positive (because the unconditional expectation of ϵ is zero). Thus, the conditional expectation of ϵ_1 is positive among those who are promoted.

Now, in period 2, the expectation of the transitory component is

$$E(\epsilon_2 \mid A + \epsilon_1 > A^*) = 0$$

because ϵ_2 is independent of A and of ϵ_1 . As a result, for any promoted individual with ability A,

$$A + E(\epsilon_1 \mid A + \epsilon_1 > A^*) > A + E(\epsilon_2 \mid A + \epsilon_1 > A^*) .$$

Thus, expected ability falls for promoted individuals from period 1 to period 2.

Individuals who are promoted are promoted in part because they are likely to have high permanent ability,[6] but also because the transitory component of their ability is high. One of the reasons that academics tend to write better papers before they receive tenure is that they would not have received tenure had they not written the better-than-average papers. The point is obvious, but is made graphic by the following example. Suppose that a firm promotes all individuals who can obtain three heads on three consecutive coin tosses. Only one in eight will be promoted. But when the firm asks their promoted individuals to repeat the feat, only one in eight will measure up. Seven

[6]The notation ϵ_1 and A could be swapped in the above discussion to show that $E(A|A+\epsilon_1>A^*) > E(A)$.

out of eight will do worse than they did before being promoted.

2. The Promotion Rule

Firms know in advance that there will be some expected fall in productivity among those promoted and adjust their promotion standard accordingly. Below, the general optimization problem for the firm is presented. Then, the way in which the rule operates is demonstrated by an example.

The firm's problem is to maximize profits (or worker utility) by selecting the job for each candidate with the highest expected value. Recall that individuals who have period 2 ability greater than x, defined above, would be assigned to job 2 were second period ability known. The firm does not see A, but only $\hat{A}$ and must choose some criterion, A*, such that it promotes workers whose observed ability in period 1 is greater than A*. This is equivalent to promoting individuals when $A > A^* - \epsilon_1$. Thus, the firm wants to choose A* so as to maximize

$$(1) \quad \underset{A^*}{Max} \int_{-\infty}^{\infty} \int_{A^*-\varepsilon_1}^{\infty} \int_{-\infty}^{\infty} (\gamma + \delta(A+\varepsilon_2)) f(A) g(\varepsilon_1) g(\varepsilon_2)\, d\varepsilon_2\, dA\, d\varepsilon_1$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{A^*-\varepsilon_1} \int_{-\infty}^{\infty} (\alpha + \beta(A+\varepsilon_2)) f(A) g(\varepsilon_1) g(\varepsilon_2)\, d\varepsilon_2\, dA\, d\varepsilon_1$$

Because the expectation of ϵ_2 is zero, (1) can be written as

(2) $$\underset{A^*}{Max} \int_{-\infty}^{\infty} \int_{A^*-\varepsilon_1}^{\infty} (\gamma + \delta\, A) dFdG + \int_{-\infty}^{\infty} \int_{-\infty_1}^{A^*-\varepsilon_1} (\alpha + \beta\, A) dFdG$$

The choice of A* depends on the distribution. However, two examples reveal that A* does not equal x as a general rule. Instead, in typical cases, firms adjust A* upward. Knowing that worker ability in period 2 will differ from worker ability in period 1, firms set the bar higher than they would if all of ability observed in period 1 carried over directly to period 2.

Consider, for example, the case where A, ϵ_1, and ϵ_2 are all distributed normally, with mean zero and variance equal to 1. Let α=1, β=.5, γ=0, δ=1. Then x, the ability level at which jobs produce equal value, is 2 since

$$\alpha + \beta(A+\epsilon) = \gamma + \delta(A+\epsilon)$$

for A+ϵ = 2. However, A* is 4.01. The firm sets its promotion standard more than two standard deviations higher than the crossing point in figure 1 because it understands that the worker's ability in period 2 is likely to be lower than it was in period 1 for the promoted group. As a result, the firm insists on a very high level of observed ability in period 1 in order to warrant promotion. Statements like, "tenure requires that the faculty member be "the best in his field, having produced outstanding research" is a manifestation of the upward adjustment.[7]

Consider the same example with a twist. Let the distribution of A remain the same, namely, normal with mean 0 and standard deviation of 1, but let the standard deviation of ϵ_1 fall to 0.1.

[7]This is the stated criterion for tenure at Stanford.

Then, A* drops from 4.01 to 2.08. Although the firm still adjusts its promotion criterion upward from x, the adjustment is much smaller because the importance of the transitory component has been diminished. There is regression to the mean, but the regression that takes place is small relative to the amount in the prior example. When the standard deviation of ϵ is zero, the promotion standard is 2, which is exactly x as expected. Then, the distribution of ϵ_1 is degenerate, so that all observed ability in period 1 is permanent ability. The problem in (2) becomes

$$\underset{A^*}{Max} \int_{A^*}^{\infty} (\gamma + \delta A) f(A) dA \; + \; \int_{-\infty}^{A^*} (\alpha + \beta A) f(A) dA$$

which has first order condition

$$\frac{\partial}{\partial A^*} = (\gamma + \delta A^*) f(A^*) + (\alpha + \beta A^*) f(A^*) = 0 \quad .$$

The solution is

$$(\gamma + \delta A^*) = (\alpha + \beta A^*)$$

which is the crossing point, i.e., x, in figure 1. When there is no transitory component, the firm simply promotes those whose permanent ability places them better in the boss job than in the worker job.

B. Strategic Behavior by Workers

So far, worker effort has been assumed to be given. As such, there could be no strategic behavior by workers. In this section, we relax the assumption that effort is given in order to determine how workers may game the system to alter their promotion possibilities. As will be shown, it is not necessarily the case that the worker will over-produce during the pre-promotion period. The nature of strategic behavior depends on the compensation scheme.

In order to examine incentives, it is necessary to define three more terms: μ_1 which is effort in period 1, μ_2 which is effort in period 2 if the worker is not promoted, and μ_2*, which is effort in period 2 if the worker is promoted. Note that effort in period 1 is determined before the promotion decision is made so it makes no sense to define effort in period 1 as contingent on promotion. Of course, it is possible that effort in period 1 will be contingent on a worker's ability, since ability affects the probability of promotion.

Now, let the cost of effort be given by C(μ). For simplicity, this is assumed to be independent of ability and the same across periods, although this is unnecessary.

1. Piece Rates

The compensation scheme matters greatly, so begin by supposing that a worker is paid a piece rate. Then, in period 2, a worker who has not been promoted chooses effort, μ_2, so as to maximize

$$\underset{\mu_2}{Max} \quad \alpha + \beta\, E(A+\mu_2+\epsilon_2) - C(\mu_2)$$

or

(3) $$\underset{\mu_2}{Max}\ \alpha + \beta(A+\mu_2) - C(\mu_2).$$

The first order condition to (3) is

(4) $$C'(\mu_2) = \beta\ .$$

An analogous problem can be solved for those who are promoted. Their problem is

(5) $$\underset{\mu_2^*}{Max}\ \gamma + \delta(A_2+\mu_2^*) - C(\mu_2^*)$$

which has first-order condition

(6) $$C'(\mu_2^*) = \delta\ .$$

Eqq. (4) and (6) define μ_2 and μ_2*. Promoted workers put forth more effort in period 2 because the marginal return to effort is higher in the boss job than in the worker job, i.e., $\delta>\beta$. Given this, the worker solves a two-period problem in period 1, knowing that he will choose μ_2*and μ_2 , depending on whether or not he is promoted.

The worker who knows his own ability has a first period problem given by[8]

[8]The discount rate is assumed to be zero.

$$\underset{\mu_1}{Max}\ \alpha+\beta\ E(\mu_1+A+\epsilon_1)-C(\mu_1) + Prob(A+\mu_1+\epsilon_1>A^*)\{\gamma+\delta(A+\ \mu_2^*) - C(\mu_2^*))\}$$

$$+\ \ Prob(A+\epsilon_1+\mu_1\ \leq A^*)\ \{\alpha+\beta(A+\mu_2\ )- C(\mu_2)\}$$

or

$$(7)\quad \underset{\mu_1}{Max}\ \ \alpha+\beta(\ \mu_1+A) - C(\mu_1)\ \ + [1\text{-}G\ (A^* + \mu_1 - A)]\ [\gamma + \delta\ (A + \mu^*_2) - C\ (\mu^*_2)]$$

$$+ G\ (A^* - \mu_1 - A)\ [\alpha + \beta\ (A + \mu_2) - C\ (\mu_2)]\quad .$$

The first-order condition is

$$(8)\qquad \beta\ - C'(\mu_1)+g(A^*\text{-}\mu_1\ \text{-}A)\ \{[\gamma+\delta(A+\mu_2^*)\text{-}C(\mu_2^*)] - [\alpha+\beta(A+\mu_2\ )\text{-}C(\mu_2\ )]\} = 0\qquad .$$

Efficient effort is supplied when workers set $C'(\mu_1) = \beta$. According to the first-order condition in (8), this occurs only when the last term on the l.h.s. is equal to zero. In general, it will not be zero. In fact, the last term is positive, implying over-investment, when

$$[\gamma+\delta(A+\mu_2^*) - C(\mu_2^*)]\quad >\quad [\alpha+\beta(A+\mu_2) -\ C(\mu_2))\ ]\qquad .$$

Sufficiently high-ability workers prefer job 1 because they earn more in job 1. As a result, they overwork in period 1 to enhance the probability that they will be promoted. Because the firm cannot distinguish effort from ability, workers who want to be promoted have an incentive to work too hard in order to fool the firm into believing that their ability levels are higher than they actually are.

Less intuitive, the converse is also true. Low-ability workers, i.e., those for whom A is sufficiently low so that

$$[\gamma+\delta(A+\mu_2^*) - C(\mu_2^*)] \quad < \quad [\alpha+\beta(A+\mu_2) - C(\mu_2))] \qquad ,$$

underwork.[9] These workers underachieve because they do not want to take the chance of being promoted. From their point of view, a promotion is bad because they are likely to earn less in the boss job than in the worker job.

The intuition is quite clear. Because workers are paid a piece rate in the second period, they want to be in the job for which they are appropriately suited. To illustrate, suppose there were only two jobs in a university: secretary and professor. On average, professors are paid more than secretaries. But if a worker were told that as a professor, he would be paid only on the basis of the number of articles that he published in top journals, many workers would prefer to be secretaries. The lowest-ability workers would be particularly anxious to stay in the secretary job and would reduce their period 1 output accordingly so as to avoid being mistaken for professor material.

The workers who are most likely to overwork or underwork in period 1 are those for whom $g(A^* - \mu_1 - A)$ is high (see eq. (8)). Under standard assumptions about the distribution of ϵ, in particular that

$$\lim_{\varepsilon \to -\infty,\infty} g(\varepsilon) = 0,$$

[9]Since μ_2 is independent of A, there is always an A sufficiently low to make this condition hold.

very high- and very low-ability workers choose the efficient level of effort in period 1. They have little to fear in terms of incorrect promotion decisions. The extremely able are almost certain to be promoted, so that extra effort has very little effect on the probability of promotion. Conversely, the totally inept are almost certain to avoid promotion, so that reducing effort has almost no effect on lowering the probability of promotion. It is the workers who are closest to the margin on the promotion decision for whom effort distortion is most likely. For example, if ϵ were unimodal with a mode at zero, then it would be those whose expected output ($A+\mu_1$) just equaled the cutoff value (A^*) that would be most likely to distort effort.

2. Tournaments

The standard gaming intuition that most have about promotions inducing atypically high effort in period 1 comes from a tournament-like payment structure. When period 2 wages depend on the job rather than the output in the job, all workers, even low-ability ones, put forth more effort than they would in the absence of period 2 promotion concerns.

In a tournament, wages in period 2 are fixed in advance and depend only on whether or not a worker gets promoted. Even if workers receive no wage prior to promotion, they put forth effort in order to maximize

$$\underset{\mu_1}{Max}\ W_b \Pr(A+\mu_1+\varepsilon_1 > A^*) + W_w \Pr(A+\mu_1+\varepsilon_1 \leq A^*)$$

where W_b is the boss wage and W_w is the worker wage. This can be rewritten as

(9) $$\underset{\mu_1}{Max}\ W_b \Pr(A+\mu_1+\varepsilon_1 > A^*) + W_w[1-\Pr(A+\mu_1+\varepsilon_1 > A^*)]$$

The first-order condition is

$$(W_b - W_w)\frac{\partial \Pr(A+\mu_1+\varepsilon_1 > A^*)}{\partial \mu_1} = C'(\mu_1)$$

or

(10) $$(W_b - W_w)g(A^* - \mu_1 - A) = C'(\mu_1) \quad .$$

The firm can obtain any level of effort, μ_1, simply by setting the spread between the boss wage and worker wage appropriately. Then it is only necessary to set the expected wage sufficiently high to attract the marginal worker.[10]

It is impossible to obtain efficiency even in period 1 with a tournament structure. The efficiency condition is that $C'(\mu_1) = 1$ for all workers. This requires that the spread is set such that

$$(W_b - W_w) = 1/g(A^* - \mu_1 - A)$$

[10]Higher-ability workers earn rents.

for all workers. But since g( ) depends on A, satisfying the condition for one ability level A_0 will not, in general, satisfy it for all other ability levels $A \neq A_0$. As a general proposition, the tournament structure will induce more than the efficient level of effort for some workers and less than the efficient level of effort for others.[11]

What is clear, however, is that effort in period 1 exceeds that in period 2. The tournament structure induces individuals to work at some positive level in period 1, but to reduce effort in period 2. In this stylized model, since there is no contingent reward in period 2, effort falls to zero. But the general point is that the tournament against a standard creates incentives to perform better in the pre-promotion period than in the post-promotion period.

Firms understand that their compensation schemes induce strategic behavior by workers and set A* accordingly. Although this may mitigate the effects of the behavior, it in no way changes the results of this section. Since all derivations hold for any given A*, they hold for the A* chosen to take these effects into account.

C. Other Issues

1. Occupational Choice

The previous structure allows the worker to choose effort and the firm to assign a worker to the job. But it is possible to allow the worker to make the choice of job and the choice of effort. Indeed, if workers have information about their own ability that firms do not have and if output is

[11]The reason for using tournaments is not so much that it guarantees efficiency with heterogeneous workers, but rather that relative comparisons are easier to make than absolute assessments of output.

perfectly observable, then the optimal scheme allows workers to choose both job and effort. Firms simply allow the worker occupational choice and pay a straight piece rate.

When workers know their ability, they can always be induced to do make the right choice, both in terms of job and effort, simply by paying them on their basis of their output. Rather than having a probation period, the worker can be asked which job he prefers. Individuals for whom

$$(11)\qquad \int_{-\infty}^{\infty}(\gamma+\delta(A+\varepsilon))g(\varepsilon)d\varepsilon \;-\; C(\mu^*) \;>\; \int_{-\infty}^{\infty}(\alpha+\beta(A+\varepsilon))g(\varepsilon)d\varepsilon \;-\; C(\mu)$$

choose the boss job. Those for whom the condition in (11) does not hold choose the worker job. Effort levels in (11) are merely the optimal levels, given the job chosen, i.e., μ is the optimal level of effort, given that the worker job is chosen so μ is the solution to (4) and (6), respectively. Note that time subscripts are deleted because there is no longer any reason to split the worklife into pre- and post-promotion. The job chosen at the beginning of the career is appropriate throughout.[12]

Because the expectation of ϵ is zero, (11) can be rewritten as

$$(12)\qquad \alpha + \beta(A+\mu) - C(\mu) \quad < \quad \gamma + \delta(A+\mu^*) - C(\mu^*)\,.$$

If the condition in (12) holds, a worker prefers the boss job. If it does not, the worker job is

[12] Human capital accumulation and other reasons for changing jobs are ignored.

selected.[13]

In a tournament against a standard, workers cannot be allowed to choose the job because they would always prefer the boss job since it pays $W_b > W_w$. Thus, the issue is whether a tournament would ever be used instead of a piece rate.

The analysis implies that strategic underproduction cannot be a major factor. In order for this to be prevalent, it must be that low-ability workers who will be paid a piece rate post-promotion fear that they will be mis-classified into the boss job when the worker job is actually appropriate. If this is the fear, an easy solution is simply to allow the worker to choose his job. If output is observable and a piece rate can be paid, it has already been shown that workers choose jobs efficiently. There is no reason for the firm to make inappropriate job assignments when workers have the information and appropriate incentives can be constructed.

To the extent that strategic overproduction is an issue, it only arises when tournaments dominate. The usual argument has to do with the observability of output. If it is cheaper to obtain an unbiased estimate of rank than of individual output, then a competitive tournament may dominate paying a piece rate. Similarly, if it is cheaper to ascertain that output has exceeded a certain standard than it is to obtain an unbiased estimate of actual output, then a tournament against a standard may be used. To the extent that firms prefer to set basic hurdles that must be exceeded for promotion,

[13] In a competitive market with rising supply price for workers (because they are distinguished by ability), firms earn zero profit. The marginal worker is the one for whom ability A_0 is low enough that

$$\alpha + \beta(A_0 + \mu) - C(\mu) = 0 .$$

then strategic overproduction may occur.

2. The Peter Principle in Reverse

Just as those who are promoted have higher than average pre-promotion transitory error, ϵ_1, so do those who fail to be promoted have lower than expected transitory components. Other things equal, this implies that those who do not get a promotion should do better after being turned down than they did before. Thus, faculty who are denied tenure and move to other schools should do better on average at those other schools than they did when they were assistant professors at the first institution.

Observing this effect may be difficult for a number of reasons. For example, a worker's output might depend on the individuals with whom he works. In an up-or-out context,[14] those who fail to be promoted may find that the complementary factors in the new job are not as productivity-enhancing as those in the first job. Furthermore, motivation is an issue. To the extent that an individual believes that he is in the running for promotion, tournament effects are present, inducing effort. After the promotion has been denied, the incentives vanish, reducing effort and output.

3. Another Interpretation of the Peter Principle

One interpretation of being promoted to a level of incompetence is that a worker who is good in one job is not necessarily good in the job one level up. Fine professors do not necessarily make good deans (although not all would interpret moving to the dean's job as a promotion).[15] A slight

[14]See Kahn and Huberman, (1988).

[15]Anderson, Dubinsky, and Mehta (1999) claim that the data reveal this for sales managers because the skills needed by salespeople are generally distinctly different from those needed by sales managers.

modification of the definitions above and some of the formulas allow for this interpretation.

To see this, allow ϵ_1 to be defined as the job-specific component of ability associated with the worker job and ϵ_2 as the job-specific component of ability associated with the boss job. Individuals are assigned to the worker job in period 1 for the reason given before: Most are better at the worker job and in the absence of information, the worker job is the right assignment. After evaluation, $\hat{A}$ is observed and the worker is promoted or not. If he is not promoted, then his ability post-promotion is $A+\epsilon_1$. If he is promoted then his ability after promotion is $A+\epsilon_2$.

Under this interpretation, workers who are not promoted have output that remains constant over time and equal to

$$\alpha + \beta(A+\epsilon_1) \quad .$$

Those who are promoted have output equal to

$$\gamma + \delta(A+\epsilon_2) \quad .$$

The argument of the first section holds:

$$E(\epsilon_1 | A + \epsilon_1 > A^*) = \int_{-\infty}^{\infty} \int_{A^*-A}^{\infty} \frac{1}{1 - G(A^* - A)} \varepsilon\, g(\varepsilon) f(A) d\varepsilon\, dA$$

$$= \int_{-\infty}^{\infty} E(\varepsilon \,|\varepsilon > A^* - A) f(A) dA$$

which is positive since f(A) is positive and the conditional expection of ϵ given ϵ greater than any number is positive (because the unconditional expectation of ϵ is zero). But the expectation of ϵ_2 is zero for promoted workers because ϵ_1 and ϵ_2 are uncorrelated. As a result, expected ability is higher in pre-promotion than for promoted workers post-promotion.

This does not necessarily imply that output is lower after promotion because workers are in different jobs. On the contrary, if A* is chosen optimally, it must be the case that expected output for the promoted workers is higher in the boss job than in the worker job. If it were not, it would be better to raise A* until expected output were higher. If this could never be obtained, then setting A* equal to infinity, i.e., never promoting anyone, would be optimal. Rather the point is that after promotion, the average promoted worker is not as able in the boss job as he was in the worker job, i.e.,

$$E(A+\epsilon_2 \mid \text{promoted}) < E(A+\epsilon_1 \mid \text{promoted}) .$$

Also true is that within any job, those left behind and not promoted are of lower than average ability. If there were a series of promotion rounds, then at every level, those who were not promoted would have a job-specific error that is negative. This can be seen simply by examining the first round, which can be thought of as a “promotion” from being out of the firm to being hired as a worker. (Individuals must exceed some standard in order to be hired.) Since it has already been shown that

$$E(\epsilon_1 \mid A+\epsilon_1 > A^*) > 0$$

and since $E(\epsilon_1) = 0$, it must be true that

$$E(\epsilon_1 \mid A+\epsilon_1 \leq A^*) < 0 \quad .$$

They appear "incompetent" because within any given job, the actual ability of those who are not promoted out of the job is lower than the unconditional expectation of ability for that job. Those who are left behind and become the long-termers are worse than those who come into the job. They are incompetent relative to the entry pool because the competent workers are promoted out of the job. In a tournament with enough steps, each competent worker would continue to be promoted until he too is incompetent, i.e., until $E(\epsilon_t)$. This is the Peter Principle: Workers are promoted to their level of incompetence. Those who are "competent" are promoted again.

4. Length of Probationary Period and Relative Importance of the Transitory Component

The longer a firm waits to make a promotion decision, the better the information. One would expect that transitory components that bias a decision could be reduced or eliminated if the firm waited long enough to make a promotion decision. The cost of waiting, however, is that workers are in the wrong job for more of their lifetimes. For example, suppose that it were possible to get a perfect reading on *A* by waiting until the date of retirement. The information would have no value because the worker would have spent his entire working career in the worker job, even if he were better suited to the boss job. The tradeoff is modeled. The conclusion is that as the variance of ϵ_1 rises, it becomes more valuable to wait on a promotion decision.

To see this, let us add one period to the previous model (without effort). Now, $\epsilon_{1,}\ \epsilon_{2,}$ and ϵ_3 refer to the transitory component in periods 1, 2 and 3 and assume that they are distributed i.i.d. Suppose that by waiting two periods, an employer can obtain a perfect reading of *A*. Under those circumstances, the optimum is simply to promote all and only those for whom *A*>x. The cost is that when the firm delays its promotion decision to the end of period 2, all workers are in the worker job during period 2 even though it might be better to place some in the boss job in period 2. Expected output over the lifetime is then

(13)

$$
\begin{aligned}
\textit{Expected Output if Wait} &= 2(\alpha+\beta E(A+\varepsilon_1)) + \int_x^{\infty}(\gamma+\delta A)dF + \int_{-\infty}^{x}(\alpha+\beta A)dF \\
&= 2\alpha + \int_x^{\infty}(\gamma+\delta A)dF + \int_{-\infty}^{x}(\alpha+\beta A)dF
\end{aligned}
$$

The alternative is to make a decision after one period, using imperfect information and recognizing that sorting will be imperfect. To make things simple, assume that a firm that makes a promotion decision at the end of period 1 cannot reevaluate at the end of period 2. The gain is that workers are sorted early so that very able people can be put in the boss job more quickly. The cost is that more errors are made in assigning workers to jobs. Then, expected output over the three periods is (14):

$$
\begin{aligned}
\textit{Expected Output Early} &= (\alpha+\beta E(A+\varepsilon_1)) + 2\int_{-\infty}^{\infty}\int_{A^*-\varepsilon}^{\infty}(\gamma+\delta A)dFdG + 2\int_{-\infty}^{\infty}\int_{-\infty}^{A^*-\varepsilon}(\alpha+\beta A)dFdG \\
&= \alpha + 2\int_{-\infty}^{\infty}\int_{A^*-\varepsilon}^{\infty}(\gamma+\delta A)dFdG + 2\int_{-\infty}^{\infty}\int_{-\infty}^{A^*-\varepsilon}(\alpha+\beta A)dFdG
\end{aligned}
$$

In an extreme case, it is clear that it pays to decide early. If the distribution of ϵ is degenerate so that there is no error, then (14) becomes

(15) $$Expected\ Output\ Early = \alpha + 2\int_{x}^{\infty}(\gamma + \delta\ A)dFdG + 2\int_{-\infty}^{x}(\alpha + \beta\ A)dF$$

The r.h.s. of the expression in (15) must exceed the r.h.s. of (13) because

$$\gamma+\delta A > \alpha+\beta A \qquad \text{for} \quad A>x$$

since that is how x is defined. Thus, when the variance in ϵ shrinks to zero, it always pays to promote early.

The example used earlier shows that it sometimes pays to defer the promotion decision until the end of period 2. As before, let α=1, β=.5, γ=0 and δ=1, where the distributions of A and ϵ_t are normal with variance equal to one. As shown earlier, the optimal cut point is A*=4.01. Then, the r.h.s. of (13) equals 3.004. The r.h.s. of (14) is 3.000. Thus, deferring the promotion decision until the second period pays when the variance in ϵ is 1. Other numerical examples show that the advantage of deferring the promotion decision becomes larger for higher variances in ϵ.

The general point is that when the distribution of ϵ is sufficiently tight, it pays to make the promotion decision early. When it is sufficiently diffuse, it pays to make the promotion decision later. Later promotion decisions are more accurate, but result in workers' spending a longer proportion of their worklife in the wrong job.

Conclusion

Workers who are promoted (or hired in the first place) receive this treatment because they are observed to have exceeded some standard. Part of the observation is based on lasting ability, but part is based on transitory components that may reflect measurement difficulties, short term luck, or skills that are job specific. As a result, there is regression to the mean, creating a "Peter Principle." Workers who are promoted do not appear to be as able as they were before the promotion.

The statistical argument has been contrasted with incentive arguments. Whether workers over-produce because they are gaming the system depends on the payment structure. If, for example, workers were paid on the basis of output both before and after the promotion decision, those who were very able would over-produce, but those who were the least able would under-produce before the production decision was made. Each type of worker wants to be avoid being mis-classified into the wrong job and exaggerates the pre-promotion reading on ability to avoid being put in the wrong job. Underproduction is not likely to be a problem, because when a piece rate is used, the worker can simply be permitted to choose his job. Allowing the worker choice eliminates strategic under- or over-production. When a tournament structure is chosen because of inability to observe output, workers produce more before promotion than they do after promotion.

The Peter Principle can be interpreted as ex post unhappiness with a promotion decision, either because workers are not as good as perceived before promotion or because they were better in their prior job relative to their peers than they are in their current one. In a multi-level firm, virtually all workers who remain at a given level will be "incompetent" in that they are not as good as the average worker coming into the job, nor are they as good as they were in their previous job.

Finally, one way to offset the Peter Principle is to wait for a longer time before making a promotion decision. The advantage is that the job assignment is better than it would have been had the decision been made earlier. The disadvantage is that able workers remain in the wrong job for a longer period of time.

References

Anderson, Rolph E., Alan J. Dubinsky, and Rajiv Mehta. "Sales Managers: Marketing's Best Example of the Peter Principle?" *Business Horizons,* 4, 2, 1, 19 (1999), 19-26.

Fairburn, James A. and James M. Malcomson. "Performance, Promotion, and the Peter Principle," *Review of Economic Studies*, forthcoming, 2000.

Faria, Joao Ricardo. "An Economic Analysis of the Peter and Dilbert Principles," Unpublished manuscript, University of Technology, Sydney, Australia, (2000).

Gibbons, Robert and Michael Waldman. "A Theory of Wage and Promotion Dynamics," *Quarterly Journal of Economics,* 114: 4 (November 1999) 1321-58 .

Jovanovic, Boyan. "Job Matching and the Theory of Turnover," *Journal of Political Economy* 87 (October 1979): 972-90. (1979a)

Jovanovic, Boyan. "Firm-Specific Capital and Turnover," *Journal of Political Economy* 87 (December 1979): 1246-60. (1979b)

Kahn, Charles, and Huberman, Gur. "Two-Sided Uncertainty and `Up-or-Out' Contracts," *Journal of Labor Economics* 6 (October 1988): 423-44.

Lazear, Edward P. "Raids and Offer-Matching," in Ehrenberg, Ronald, ed. *Research in Labor Economics,* Vol. 8, 1986, part A: 141-65.

Peter, L.J. and R. Hull. *The Peter Principle: Why Things Always Go Wrong.* New York: Morrow (1969).