Journal of Economic Perspectives—Volume 3, Number 4—Fall 1989—Pages 65–83

Self-Enforcing Contracts, Shirking, and Life Cycle Incentives

H. Lorne Carmichael

The labor market is a rich and complicated place. When a worker takes a job he expects to earn a wage, but will also care about rates of wage growth, fringe benefits, levels of risk, retirement practices, pensions, promotion and layoff rules, seniority rights, and grievance procedures. In return the worker must give up some time, but he is also asked to upgrade his skills, train other workers, provide effort and ideas, and defer to authority in questions of how his time is spent.

Great changes are occurring in the way labor market institutions such as these are modelled. Central to the new approach is the concept of a self-enforcing implicit contract. This is the latest in a series of variations on the ideas of Azariadis (1975), Baily (1974) and Gordon (1974), and is one of the more exciting new developments in the theory of labor markets. This essay will outline this approach and evaluate some of its accomplishments.

The ideas will be presented in the context of an incentive model that is of some independent interest. The model was introduced by Becker and Stigler (1974), and was extended by Lazear (1979) to account for wage growth and mandatory retirement. In these papers the workers retire at a fixed age, lending the name "life cycle incentives" to the approach. More recently, Shapiro and Stiglitz (1984) have used a version without a fixed retirement date to account for unemployment. This version is often called the "shirking" model. These models have generated some controversy, but much of it can be resolved when the models are presented in the framework of self-enforcing contracts.

In fact, this simple incentive model can do even more. It can provide a rationale for "competitive hierarchies," where firms place their workers in competition with

■ *H. Lorne Carmichael is Associate Professor of Economics, Queen's University, Kingston, Ontario, Canada.*

each other to get them to provide effort. It makes some interesting and new predictions about how the use of performance bonuses will vary with conditions in the labor market. It can provide insights into retirement practices in Japan, which are quite different from those in the West. The approach also forces us to consider the role of firm reputations in a more careful way than was done in the past. There do remain some anomalies and puzzles, however, and these will also be discussed ahead.

The focus of this essay is on developments in the theory of self-enforcing contracts, and how these can be used to derive predictions about labor market institutions. There is not the space here to adequately review the empirical literature which has attempted to test these predictions. The reader is referred to the companion piece by Robert Hutchens in this issue for a more complete discussion of the empirical issues.

Self-Enforcing Contracts: The Shirking Model

It has long been recognized that the current wage is but one aspect of an employment agreement. Indeed, this is the basis of Adam Smith's (1776) claim that wages would be set to equalize the "net advantage" of different occupations. Nowadays net advantage is called "expected utility," but the principle remains.

Many nonpecuniary aspects of a job come in the form of expected future benefits. Pensions, chances of promotion, protection from layoffs, and expected future wage increases all fit in this category. Azariadis, Baily and Gordon coined the term "implicit contract" to describe the promises a firm would make to its workers about benefits such as these. The modifier "implicit" was used since explicit agreements of this kind are seldom observed outside of the union sector. Nonetheless, the agreements were modeled entirely as if they were explicit, legally enforceable contracts.

This brought up the so called "enforceability problem." If the firm's promises are not legally binding, and they specify behavior which turns out to be against the firm's best interest in some future period, why will the firm ever keep its promises? At the time it was argued that a concern about its reputation would keep the firm honest, but the formation of reputations was not formally modeled until much later (Holmstrom, 1981; Carmichael, 1984).

The theory of self-enforcing implicit contracts is specifically designed to address this enforceability problem. The goal of the theory is to determine precisely the restrictions which must be placed on labor market agreements such that they will in fact be upheld in future periods. Many of these restrictions are consistent with observed labor market institutions. In this sense, the theory can provide explanations for them.

The theory begins, as always, with some assumptions and definitions. If an agreement is to be legally enforceable, it must condition future behavior only on variables which are verifiable. These variables are observable by the firm, the workers, and an outside enforcement agency such as the courts. Examples might include wage rates, hours of work, and industrial accidents which require medical attention. Other

variables which are observable (perhaps with error) by the workers and the firm but which would be difficult to establish in a court are called nonverifiable. Examples here might include effort levels of workers, objectionable behavior by supervisors, and the like. Implicit agreements (that is, promises) about future behavior need not be based only on verifiable variables. A firm can promise to reward workers for their effort, for example, even if this effort is not verifiable.[1]

At this point it is perhaps best to introduce the basic incentive model we will be using throughout the paper. Consider the problem faced by a firm and a worker who form a relationship that may last for several periods into the future. (At this point we shall assume there is no fixed retirement date.) The production process requires that the worker provide effort, which other things equal he would prefer not to do. Given that effort is provided, the worker's productivity is constant over time. Verifiable quantities include the wage paid (as certified by cheque stubs), and the time that the worker has been on the premises (as certified by a punch clock). However, while the firm may have a good idea as to how much the worker has produced and thus how much effort he has provided, this will be hard to verify to any third party. We will therefore model the worker's effort, and thus his output, as nonverifiable.

The firm and the worker are able to sign an explicit agreement which specifies the wage to be paid and the times the worker must be in attendance.[2] The worker promises to provide effort and the firm promises to reward him if and only if this effort is forthcoming. The reward can in principle take on many forms. It could be a promotion, a monetary bonus, or simply a renewal of the explicit contract for another period under the same terms. In some contexts, the worker may further promise to quit if he has provided effort and the promised reward is not forthcoming. These promises, along with the explicit agreement covering wages and hours, will be called a self-enforcing implicit contract if and only if it is in each party's interest to keep its promises.[3]

The approach provides an explicit model of an implicit agreement. Self-enforcing contracts are collections of promises that, while they might not be legally binding, are nonetheless credible. Everyone can be confident that the promises will be kept. The major difficulty in trying to implement this approach is that we can no longer rely on the usual technique of profit maximization subject only to the constraint that workers want to participate. We must also worry about whether it will be in each party's interest to keep its promises.

The basic task is now to characterize the set of self-enforcing implicit contracts for our incentive model. Let us call such a contract nontrivial if it specifies behavior

[1]This sharp distinction between types of variables is of course an exaggeration. In practice, courts are often required to determine what has happened when there is little hard evidence either way. The point is simply that this method of settling disputes is very costly, and there are incentives to design agreements that minimize these costs.

[2]The agreement about wages may not actually be written down, but it is assumed to be legally enforceable. The terms "explicit" and "legally enforceable" are synonymous throughout the paper.

[3]In the jargon of game theory, the requirement is simply that the promises of the firm and the workers form a subgame perfect equilibrium.

which is not in the immediate short term interest of one or both of the parties. The fundamental result in this area is that nontrivial self-enforcing contracts will exist only if upholding the agreement will generate a surplus for the two parties (MacLeod and Malcolmson, 1987, 1989). One or both of the parties must be earning strictly more than in his best alternative, and neither can be earning less.

The logic behind this result is quite simple, and can be illustrated in the framework we have introduced. Consider the case where the worker's wage is fixed at the same level each period, and renewal of this (explicit) contract is the reward for effort. Since effort is costly, the worker always has the short run incentive to be lazy. If the threat of not having his contract renewed (that is, being fired) is to prevent this, then the worker must expect to earn more if he is retained than if he has to look elsewhere for work.

Assume that the worker would lose money by changing jobs, so that he does provide effort. The firm can nonetheless claim that the worker has cheated and fire him anyway. The worker has no redress through the courts if this happens since the courts cannot verify that the worker has provided effort. This "firm cheating" can be prevented only if the benefits the firm expects to get from a continuation of the contract are at least as great as what it can get by firing the worker. But we now have a situation where the worker must earn more than his alternative if the relationship continues while the firm can earn no less. Thus, there must be a surplus available from their match.

As a second, slightly more complicated example, consider the case where the firm promises a monetary bonus to workers if they provide effort. This promise will prevent cheating only if the worker believes it and only if the size of the bonus is large enough to outweigh the personal benefits of being lazy. Assume both of these conditions are true, so the worker does provide effort. The firm can again claim that the worker has cheated and refuse to pay the bonus. The most the worker can do in this case is quit. A quit will deprive the firm of any profits it expects to make from employing that worker in future periods. Thus the worker's threat of a quit will prevent the firm from cheating only if the present value of these lost profits exceeds the value of the bonus. Since the worker is sure to quit if he expects to earn less than his alternative in future periods and the firm must expect to earn more than its alternative, again there must be a surplus.

It is still necessary to check to see that the worker's threat to quit will be carried out if and only if the firm cheats. This can be guaranteed if the worker's full pay, given honesty on both sides, is equal to his alternative in another job. This way, even if the worker believes the firm will never cheat on him again, his threat to leave is believable. (Remember that even though the honest worker is earning the same wage as in his alternative, part of this wage is contingent on his good performance. Thus he still has the incentive to be honest.)

A surplus is needed only because both parties to the agreement are potential cheaters. If firms are always honest, then the bonus contract above can get workers to be honest without the need for a surplus. Workers will know that they will get their bonuses since they will know that the firm always keeps its promises. Similarly, if the

workers are always honest, there is no need for them to earn a surplus in the fixed wage contract. They can be paid their alternative wage and they will provide effort simply because they always keep their promises.

These examples highlight some important differences between the theory of self-enforcing contracts and the previous theory of explicit contracts. First, there is no presumption that there will be a unique self-enforcing contract. Our example has already produced two possibilities, and there are many more.[4] Other factors, some of which we shall be exploring shortly, will in general have to be introduced to choose among the possible agreements.

Second, and this may be more surprising, it is clear from the last example that it is not necessary that output be verifiable for wages to be contingent on performance. This result is contrary to the intuition some readers may have developed from earlier work on the form of explicit contracts under asymmetric information (Grossman and Hart, 1981; Green and Kahn, 1983; Cooper, 1983). In that context, where contracts are explicit and the firm is the only one to observe the value of output, wages can only be contingent on variables such as employment.

There are many potential sources for a surplus in employment relationships. Specific human capital is a natural one, as are the savings in direct mobility costs for each party if they do not have to search for new partners. Reputations can also generate a surplus. For example, if workers are not homogeneous then if one is fired other firms may decide he is inherently lazy and refuse to hire him. Similarly, if a firm has a reputation for firing a lot of workers, it may find that it has to pay workers more in order to attract them away from their alternative jobs.

Equilibrium unemployment can also do the trick. This is the basis of shirking style efficiency wage models (Stiglitz and Shapiro, 1984). In this context, a fired worker must spend some time unemployed before he can find another job. Even though all firms pay the same wage in equilibrium, fear of unemployment will keep employed workers honest.

A fifth possibility is for one or both of the two parties to explicitly contract with a third one to pay it some money if a verifiable breach should occur. Fines paid to the legal system are probably the best example of such payments, although we shall see later that promotion tournaments can introduce third parties to the relationship in a rather unexpected way.

The list does not stop here. Of historical interest is the practice of exchanging live hostages (Williamson, 1983). The person who sends the hostage will care greatly about whether he or she survives. The person receiving it has little to gain from murder, creating the necessary surplus. (Live hostages differ from monetary ones in this respect: a transfer of money creates immediate incentives for the receiving party to falsely claim that a breach has occurred). A final source, emphasized more by parents and clergymen than by economists, is the sense of guilt some people feel when they are unable to keep their promises.

[4] In fact, there is a whole continuum of self-enforcing contracts which lie between these two extremes. See MacLeod and Malcolmson (1989).

The way in which the surplus is divided between the two parties is important, because this is what determines the form of the contract (MacLeod and Malcolmson, 1989). If we can pin down the division of the surplus, we can pin down the contract. This allows us to make some predictions about how the form of the contract will change with conditions in the labor market.

Suppose that unemployment is the source of at least a part of the surplus. When the labor market is tight and there are many vacancies, firms will be able to offer bonuses to motivate their workers. Workers will trust the firm to pay these bonuses because they know that they will be hard to replace if they quit and that the costs to them of quitting are very low. The cost to the firm of "unemployed jobs" is high. When there is high unemployment of workers, replacements are easy to find and firms cannot be trusted to provide bonuses. Then firms will pay fixed wages and motivate workers by threatening to fire them.

Notice that it is not a worker's wage which determines his share of the surplus in this model, but the difference between what he is earning in the match and what he can get elsewhere. Workers may earn more money in tight labor markets, but relative to the unemployed they are not doing as well as they do in bad times. Thus the threat of a fire is not enough to motivate them. When there are more workers than jobs it is difficult to give firms the surplus because they can replace their workers too easily. Thus the employed workers must get the surplus and the unemployment (which is exogenous here) will be involuntary. Odd though it may seem, workers receive more of the surplus in a condition of high unemployment, while firms receive more of the surplus in a market with low unemployment.

We have seen that with the notion of a self-enforcing contract, labor theorists are finally beginning to capture in their models the ideas which caught their imaginations in the late 1970s. Implicit contracts are now truly implicit. However, there remains an indeterminacy which is a bit troublesome. There are many sources for a surplus and many ways of dividing the surplus, each leading to a slightly different agreement. One way to tie things down more tightly is to introduce a fixed retirement date for workers.

Life Cycle Incentives

The life cycle incentive models differ from the shirking model just discussed in one simple way. In the life cycle framework, both parties know that at some point in the future the relationship will end. The reasons for making this assumption are as follows. First, the length of time remaining in the relationship gets shorter as the years go by, so that the character of the self-enforcing contract may change from one period to the next. This allows us to consider questions of wage profiles and mandatory retirement. Second, the last period presents a special problem because neither party can hope to gain from extending the relationship to future periods. This puts some severe restrictions on the kinds of things which can generate a surplus, and also helps

to determine which party must get the surplus. The model with a fixed retirement date therefore makes much more precise predictions than the one without.

The discussion in this section will be a little bit more formal than in the last. Fortunately, it turns out that the implications of the model are quite robust to changes in the assumptions. This means the exposition can be kept simple by considering one specific case in detail and outlining afterwards the effects of changes in the assumptions.

Consider a model where workers are monitored at discrete intervals by the firm. Workers can work at one of only two effort levels. If the worker has been honest over the latest period the firm will find this out with certainty. If the worker has been dishonest, the firm will catch him with probability P, with P naturally lying between zero and one. If the worker stays his entire working lifetime with the firm he will be monitored T times, with the last time coming just before he retires.

When making decisions about current behavior, each agent will be concerned about what will happen in future periods. The assumption made here, and in all models of this kind, is that to form expectations about what will happen at some time in the future, people mentally place themselves in the position they and the others will be in at that time and determine how they will behave. Thus to solve this particular model and characterize the worker's wage profile we can first of all determine what will happen in the last period. Once this is worked out we can look at what will happen in the second last period assuming that the agents know what we do about what will happen in the last one. This process, known as "backwards induction," can then be carried all the way back to the first period to solve the entire model.

Reputations and the Division of the Surplus

The special character of the last period determines a great many things. In particular, it is clear that a fixed wage contract will not deter cheating. The worker knows he will be retired after this period whether he cheats or not, and he knows he will not be monitored until the last moment before he goes home. Thus the only way for the firm to prevent workers from cheating in their last period is to make some portion of their pay contingent upon their performance. The firm will have to hold back some of the worker's pay until the very last moment and give it to him as a pension, or lump sum bonus, after he has been monitored for the last time.

We know from the earlier analysis that when the contract is of the contingent wage variety, the firm must get a surplus. Otherwise it may falsely claim that cheating has occurred. But where is this surplus to arise? Both parties know the relationship will end regardless of the worker's performance, and (we assume) that the worker's enjoyment of his retirement will not be marred by a poor performance report in his last period. This eliminates mobility costs, specific capital, unemployment, and guilt as possible sources. If we are also willing to rule out live hostages, then we are left with third parties (to be discussed later) and reputations.

Since the firm is the one that needs the surplus, if reputations are to work they must be attached to the firm. Thus we have the result (Bull, 1987) that in the absence

of third parties, firm reputations are necessary for the existence of self-enforcing contracts of finite length.

Firm reputations arise when workers have imperfect information about the characteristics of a firm and thus cannot predict with certainty its behavior in future periods. They observe the present and past behavior of the firm to help form their expectations. If a firm starts to deny a larger proportion of older workers their bonuses, the other workers may begin to suspect that they will be treated in the same way in the future. The firm will therefore find it harder to hire new workers and harder to prevent its other workers from cheating. A reputation acts a bit like an asset that is a product of the current match but that lives on into future periods and can thus help to create a surplus.

If the firm is to care about its reputation it must be earning quasi-rents from being in business (Klein and Leffler, 1981; Shapiro, 1983). When there are no costs to a firm going out of business and moving its capital elsewhere, the opinions of the other workers cannot hurt it. The firm can cheat on all of its workers, leave the industry, and re-enter the same or another one with a new identity.

There are other important issues which arise in reputational models, but it would be quite impossible to do them justice here. Instead, since we know that firm reputations are necessary for the analysis, we will just assume that they exist and are sufficient to keep the firm honest. With this we can move directly to the model's implications for wage profiles and mandatory retirement.

Wage Profiles

We wish to determine the wages workers will be paid over their lifetime. First of all we need some more notation. The firm's explicit contract prespecifies the times the worker must be in attendance each period and the minimum wage he will be paid for showing up. Denote the minimum wage in period t by W_m^t. The worker has the alternative of working at another firm which pays him W_a in each period. The firm promises to pay the worker a bonus B^t in period t if he has been honest. The net benefit to the worker of providing minimal effort in any period is C. Recall that there are T periods and the probability of a cheater getting caught is P.

The worker will provide effort in the last period as long as the expected benefits from doing so exceed the benefits of cheating. If he is honest, he knows he will get $W_m^T + B^T$. If he cheats he will get the benefit C and at least the minimum wage. The overall expected benefits from cheating are therefore $W_m^T + C + (1 - P)B^T$. (If he cheats he might not get caught, in which case he will still get the bonus.) A simple manipulation then gives us the intuitive result that the bonus must equal or exceed the gains from cheating divided by the probability of getting caught; that is, $B > C/P$.

We must also ensure that the honest worker will not want to quit in the last period. This requires simply that his wages be as high as they are in his alternative; that is, $W_m^T + B^T > W_a$.

In general there will be some limits on the punishments the firm can impose on its workers. It is unlikely that the minimum wage could be negative, for example.

However, let us begin by considering the case where the guaranteed wage can be anything at all. If we assume that the firm will set its wage and bonus so as to minimize its labor cost subject to being able to attract workers and keep them from cheating, then the two expressions can be thought of as equalities and the period T wages are determinate. To keep the honest worker from quitting, he is paid his alternative wage. The minimum payment W_m^T is then set to ensure the worker stays honest. What keeps the worker from cheating is the threat that he will receive a very low (or negative) wage for time already spent on the job.

Now that the firm's last period policies are known, consider what happens the period before. The firm has another option in this period, which is to fire the worker if he cheats. This threat will affect the worker's effort decision only if he has something to lose by being fired. However, by the preceding logic, the worker expects to earn exactly his alternative in the next period, so he is indifferent to being fired, and the threat of a fire is ineffective. Nonetheless, the firm can again threaten to pay the worker his alternative if he is honest and the same minimum wage as in the last period if he is not, and this will again prevent cheating. This logic can be carried right back to the first period with the surprising result that with honest firms and no restrictions on the minimum wage paid by firms, the cost minimizing wage is the same in every period. The wage profile is flat!

Suppose now that there is a legal minimum wage which is the same every period, and, of course, is no higher than the worker's alternative wage.[5] To prevent cheating in the last period honest workers may now have to be paid an amount greater than their alternative. (The necessary condition for this to occur is just $W_m + C/P > W_a$.) In this more realistic case, the threat of being fired at the end of the second last period carries some weight. This means that wages in the penultimate period do not have to be so high.

Here's why. Assume for simplicity that the discount rate is zero. In the second to the last period the worker considers the effects of cheating on his income in both remaining periods. It is still true that the amount he expects to earn by being honest must exceed what he gets if he is caught cheating by at least C/P. If he is caught cheating and fired, then in his last period he will be in his alternative instead of being in the firm earning the high last period wage. Firing thus imposes a penalty of $W_m + C/P - W_a$. The firm will choose its current period bonus to bring the overall punishment up to C/P. This means simply that $W_m + C/P - W_a + B^{T-1} = C/P$. A little bit of algebra reveals that $B^{T-1} = W_a - W_m$, so that honest workers in their second to the last period are paid exactly their alternative wage, while dishonest ones are paid W_m and are fired.

The wages in previous periods can be derived in the same way through backwards induction, with the result that the firm's policies in periods 1 through $T - 1$ are

[5]This is not the only way to put limits on punishments, but it is the simplest. If the firm occasionally mistakes honest workers for cheaters, for example, then worker risk aversion will also put a limit on punishments.

the same. The intuition here is that with a zero discount rate, the prospect of losing the last period's bonus is equally distasteful no matter how long it is until the last period. Thus the current period bonus is independent of the number of periods remaining in the contract. The wage profile of an honest worker in this model is again flat, but has a bonus at the end.

Since the worker is paid a higher wage in his last period, the expected present value of joining the firm in period one exceeds that of the alternative by the amount $W_m + C/P - W_a$. (The discount rate of zero makes the present value simple to calculate.) Thus, the worker is earning a surplus as well as the firm. However the worker's surplus is a consequence of the limits we have placed on the ability of the firm to punish the worker. In the first case where there were no restrictions on the minimum wage, this surplus did not arise.

If the firm can charge entrance fees to new workers, it can reduce its labor costs even further. The worker pays a fee equal to the extra earnings he expects in the last period, and his overall expected earnings fall to that in his alternative. In equilibrium the firm will hire workers up to the point where the last one is paid his marginal product, and workers will be just as productive in this firm as they are elsewhere. (Remember, however, that there must be quasi-rents that the firm will lose if it goes out of business.)

One question that naturally arises is whether the firm will just claim that the new worker has cheated, fire him, and keep his entrance fee. However, we know already that the firm must be able to resist firing workers in their last period even though it would mean an immediate benefit of C/P. If a firm wanted to cheat a new worker it would let him work for one period and produce W_a, then it would fire him, pay him W_m and keep his entrance fee. The short term benefits from doing this work out to exactly C/P. Thus, if reputations can prevent the firm from cheating on its oldest workers (and for us to get far this must be true), then reputations can prevent them from cheating on their youngest ones.

Wage Profiles: Extensions and Comparisons

The model we have been discussing can be extended in several ways without affecting the basic result that wage profiles are flat with a bonus at the end and an entrance fee at the beginning. Further modifications will get wages to rise, but the predictions about entrance fees and final bonuses are quite robust.

For example, the firm can increase its monitoring intensity by shortening the period between "inspections." This will simply reduce the benefits the worker might receive from cheating since the length of time he gets to enjoy his leisure gets shorter. Thus, the only effect is that the size of the up front payment gets smaller. The wage rate in any period does not change at all.

When the firm's monitoring becomes less precise, so that the worker who shirks has a greater chance of not being caught, the penalty for being caught must increase. This means that the last period wage is higher as well as the entrance fee. Nothing else in the model changes. When discount rates are positive, workers must be paid somewhat more in all periods but the last since the gains from cheating are received in

the current period and the costs of being fired are in the form of lost wages in the future.[6] Again, the essential character of the wage profile does not change.

The introduction of imperfect capital markets will reduce the size of the entrance fee, but will not eliminate it entirely (Carmichael, 1985). This is because the worker dislikes entrance fees (they distort his consumption stream), and a smaller fee is all that is required to bring the expected utility of the offer down to that of the alternative. Again, however, the basic character of the results is unaffected.

It is possible to get rising wages out of the model. One way is to make the worker's alternative wage increase over time. This could be the result of on-the-job training, and in this sense the model is complementary with human capital models of wage growth. Another way is to assume that the benefits from cheating rise with experience and/or the probability of getting caught falls.[7] However, even in these cases there should be entrance fees and last period bonuses.

The wage profiles here are different from those derived by Lazear (1979, 1981). Lazear argues in a similar model that there are many possible wage profiles which can prevent worker cheating, and in particular that smoothly upward sloping profiles with no entrance fees can be optimal. However in his model, by assumption, the worker's alternative is strictly less than his marginal product until the time at which retirement occurs. Firms make zero profits, so that honest workers produce (and earn) more than they would in their alternative. The surplus created in this way may be more than is required to prevent workers from cheating. If so, the wage profiles will be indeterminate.

In the version of the model presented here, firms will lower wages whenever they can do so without causing workers to cheat or quit. The wage profile derived is the unique cost minimizing profile.[8] This precision is laudable, but it presents its own problems, as we shall see after we discuss the model's predictions about retirement.

[6]Consider the worker's problem in the second last period. With a positive discount rate r, the present value of the extra wages the honest worker gets in the last period is $(1/1 + r)(W_m + B^T - W_a)$. Thus the penultimate period's bonus must be increased by $(r/1 + r)(W_m + B^T - W_a)$ to ensure that the total losses from being caught cheating still add up to C/P. Again, the induction carries back to the first period.

[7]The value of C/P in the last period determines the final period bonus. Moving backwards, C/P falls so that the bonus in each period can be smaller. Observed wages thus rise with experience.

[8]The use of backwards induction to derive the wage profile ensures that the firm is minimizing costs at each point in the worker's career. This is essentially the same thing as allowing the contract to be renegotiated each period. An alternative assumption is that the explicit contract sets all wages at the time the worker is hired. In this case there are many profiles which ensure a worker's honesty and provide him with the same expected income over his lifetime. For example, if capital markets are perfect, the firm could pay the worker his entire earnings on the day he retires. After paying his entrance fee, the worker will be just indifferent about cheating, and as he gains seniority he will strictly prefer honesty. When we consider firm cheating, however, the old profile comes back again. The quasi-rents that the firm must be earning if it is to care enough about its reputation to prevent it from cheating depend on the potential short term gains to the firm from dishonesty. These gains are precisely the excess the worker will earn by being honest. Since the firm can choose when to fire the worker, to minimize the potential for firm cheating the wage profile should ensure that this excess is as small as possible at each point in the worker's career. The profile derived here does precisely that.

Mandatory Retirement

So far we have simply assumed that the match has a last period. Of course the worker will die eventually, but our assumption is stronger than this. We have assumed that both parties know the specific time at which the match will end. Can this assumption be justified within the model?

The argument, due to Lazear (1979), goes like this. In equilibrium the worker is paid his marginal product over his lifetime (else the firm would hire more (or fewer) workers). However, at the beginning he has had to pay an entrance fee and at the end he expects to get a bonus. Thus, once he is hired, he expects to earn more than his productivity if he stays with the firm. It is efficient for the match to end at a time when the worker is producing less on the job than he would elsewhere. However, the worker cannot be trusted to leave at this time because he is getting paid more than his marginal product. Thus the firm must make the decision. The firm doesn't know the worker's alternative with accuracy, but it does know that social security benefits start at age 65. The value of being at home takes a big jump at this age, and it is thus a good time for the match to end.

Retirement must be mandatory here because workers who quit lose their bonus as well as those who are fired. The only way to get a pension is to stay until mandatory retirement. If it is possible to give the bonus to genuine quitters but withhold it from cheaters, then retirement can be like any other quit. This does not change the implications of the model about wage profiles, but it does obviate the need for retirement to be mandatory. Indeed, with the current legal restrictions on mandatory retirement firms seem to be moving in this direction. Workers are given the option of leaving with full pension over a range of ages.

The question remains, of course, as to why retirement was ever mandatory in the United States and why it remains so for many firms in Canada, Britain, and elsewhere. Part of the answer may be that if retirement is voluntary, as above, there will be incentives for both parties to misrepresent the nature of the separation. If a worker has cheated and is about to be fired, he will claim in a court that he is just quitting. The firm will want to claim that all separations are fires, so as to deny quitting workers their pensions. The assumption that the court could easily resolve such disputes is tantamount to the assumption that it could determine whether or not the worker had cheated. But the whole basis of this approach is that cheating is not verifiable.

The firm's concern about its reputation will surely mitigate the problem somewhat. But it seems reasonable that reputations will work better when workers' inference problems are less severe. If the firm gives pensions to only some of the workers who claim they are quitting, then those watching must determine for themselves whether it is the firm or the workers who are cheating. It would take an explicit model of reputations to say much more, however, and this would require another paper.

Anomalies: Entrance Fees and Vested Pensions

The results of this model concerning wage growth correspond reasonably well with the data, largely because the model can be modified to be consistent with almost

any wage profile. There is even some recent evidence that wage profiles are quite flat, apart from very early in the worker's career and at the end, where he receives a pension (Altonji and Shakotko, 1987; Brown, 1985). These results are consistent with the simpler versions of the model. The data on wage profiles is still controversial, however, and the reader is again referred to the companion piece by Robert Hutchens for a fuller discussion.

The biggest problem is, of course, the prediction that firms should be charging workers entrance fees. I know of no labor markets anywhere in the world or in history where this practice has been widespread.[9] The problem arises in the shirking efficiency wage literature as well (Carmichael, 1985, 1990; Shapiro and Stiglitz, 1984; 1985). Here it is sometimes argued that the firm would steal such fees if they were offered. If the employed workers are getting all of the surplus and the firm is getting none, then this claim is correct. However, in the life cycle context we have seen that the firm must be disciplined by its reputation, and therefore must get a surplus. Thus this argument will not work.

Lazear in his original paper argued that the worker could pay the entrance fee by installment, by working for less than his marginal product while he is young. This kind of a wage profile cannot be derived in the simple model we have here unless we follow Lazear and abandon the assumption that firms choose wages to minimize costs. However, it has an intuitive appeal (especially in professions like academia), and an extension of the model may be consistent with it.

The problem is that with identical firms and no entrance fees, workers just after they are hired must be earning their alternative wage. (Their alternative is to go to a firm which offers the same wage profile.) However, at this stage in their careers there are many sources for a surplus to prevent them from cheating. Mobility costs and worker reputations may be sufficient, particularly if firms concentrate monitoring efforts on their new workers. Also, firms may be interested in selecting workers who work hard even when the economic incentives are weak (again academia comes to mind), and can do this by paying new workers the going wage and observing their performance closely. These are just ideas at this stage, however, and need to be worked out. Akerlof and Katz (1989) make a start.

The data present another problem for the life cycle approach. Pension reform in the United States has now made it quite clear that in most cases after several years of service the worker's pension must become vested with the worker. Even workers who are dismissed will receive their pensions. Yet we have argued that the firm must retain some contingent payment if it is to prevent cheating in the last period. What happens to the model in this case?

[9] I do know of one example. In Northern Ontario there are small companies which subcontract with the Ontario government to plant trees every spring. Workers (mostly university students) are hired some two months before planting starts, and are very hard to hire later since the planting sites are far out in the bush. New hires are asked to pay as much as $300 up front at the time the job offer is accepted, to ensure that they are not just taking this job until a better one shows up. This fee is nonrefundable even if the worker is fired later. Our model suggests that the workers must believe the firms expect to make excess profits of at least $300 from those workers who do show up.

One possibility is for the firm simply to allow its oldest workers to cheat. Old workers will earn more than their alternative because they have "tenure" and do not have to provide effort. The prospect of reaching this stage will be an inducement to younger workers not to cheat. If the firm is disciplined by reputation (as it must be in this finite horizon model), then it will not fire people dishonestly just before they get "tenure." Mandatory retirement is necessary because older workers are again getting paid more than they are worth, this time because their productivity is lower.

This contract has a certain intuitive appeal, but it is clear that there is an inefficiency, since older workers are not working. To prevent this outcome the firm could, for example, retain a further amount to be given to good workers upon retirement. It may be possible to interpret voluntary inflation protection in this way (and gold watches?), but it is not clear if the firms which provide this also threaten to discriminate among the eligible recipients.

The discussion of the merits (and demerits) of this model will be continued in the concluding section of the paper. Before this we will discuss briefly the potential role for third parties in agreements between workers and firms. This leads naturally to the theory of tournaments, and the special role they play in the theory of self-enforcing contracts.

Third Parties and Tournaments

The first models of tournaments (Lazear and Rosen, 1981) emphasized that by placing workers in competition with each other the firm could induce workers to provide effort in much the same way as could be done with piece rates. Tournaments might be preferred when relative output is easier to observe than individual output, or when there is uncertainty in the measurement of output which is correlated across workers. These reasons for tournaments all assume that output, or relative output, is verifiable so that the firm can advertise explicit contracts which make payments conditional on these quantities, and compete for workers based on the expected utility they offer.

In a self-enforcing contracts model, a tournament will play quite a different role. Here, the tournament can mimic the addition of an outside third party to the agreement between a firm and its workers, and can therefore "create" a surplus where there was none before. To see how this is done, consider first the role of an explicit third party. The best example is probably the legal enforcement system.

Suppose that in our life cycle incentive model the worker is paid exactly what he is worth in all periods and this is equal to his alternative. The worker has the opportunity to cheat and would clearly do so if it were not for a law which states that all workers who separate from their firm must pay a fine to the state. The fine is sufficient to deter cheating, and since it is paid to the legal system the firm has nothing to gain by claiming that cheating has occurred when it has not. All is not perfect since workers who genuinely want to quit will also have to pay the fine, but this is no worse than was the case in our earlier model.

Of course, this is a very unrealistic view of how the legal system could be expected to operate. It is also extremely unrealistic to assume that this system is

costless. However, if arrangements of this kind could be made at low cost there would be a great advantage in using them. Both the firm and the worker are earning their alternatives in this example, and yet a self-enforcing contract is feasible. The surplus needed for the self-enforcing contract to work is created by artificially lowering the outcomes for workers who cheat, and workers are still optimally allocated to jobs.

Given the costs of using the legal system, the question arises as to whether there could be a market for third parties. Serious problems arise here as well when it is recognized that these third parties will themselves be economic agents whose behavior must also be specified as a part of any self-enforcing contract. There must not only be no opportunities for any single agent to cheat, but as well there must be no opportunities for any pair of them to collude in order to cheat on the other. For example, the firm may offer to fire the worker and split the fine with the third party, or the worker may offer to cheat if the third party will split the fine with him.[10]

Tournaments among the workers introduce a third party to the implicit contract in a rather interesting way. In essence, the workers become third parties for each other! Papers which exploit this idea include Carmichael (1983a, 1983b), Malcolmson (1984, 1986), and Bhattacharya (1986). To see how it works we can consider a simple one period model.

Assume that the firm hires M workers. As before, there is a minimum wage W_m that the firm guarantees to pay the workers just for showing up. It also guarantees to pay the bonus B to a fixed number N of the workers. This bonus could be interpreted as a promotion, or it could be a prize of the sort that is sometimes used to motivate salespeople. The number of bonus winners, the amount of the bonus, and the minimum wage are all assumed to be verifiable, and are part of the firm's explicit contract. The firm promises to give first preference for the bonus to workers who have provided effort, and to use a random draw when otherwise necessary. The workers promise to work.

To understand the properties of this contract, note first of all that the firm's wage bill for all of its workers is completely predetermined. It pays M workers the guaranteed wage and N of these get a extra bonus. This is true regardless of how many workers cheat. Thus it is clear that the firm has nothing to gain by misrepresenting the identities of any cheaters.

If workers are to provide effort they must perceive that it will increase their rewards. If each worker believes that the others will provide effort, he will perceive a probability of N/M of getting the bonus if he also works. The benefits of cheating are again denoted by C and the probability of detection again given by P. As before, the

[10]One example of a "market" for third parties is the labor arbitration business. The potential for collusion here seems quite strong and yet the market seems to function quite well. One reason is the emphasis arbitrators place on maintaining their reputations. In choosing an arbitrator to settle a dispute, the two sides typically start with a list of potential candidates, and each side has the right to veto any ones they do not want. A particularly one sided decision would therefore lose an arbitrator much of his future business. For this system to work, arbitrators must expect to earn some rents in settling future disputes. As well, if there is a great deal of money involved in the dispute, it would be wise to choose an arbitrator who can command high fees in the future.

worker must expect to get an amount if he is honest which is C/P higher than what he gets if caught cheating. Thus to ensure effort, we must have $(N/M)B > C/P$. Note that this will also be sufficient to prevent cheating even if the worker assumes some or all of his opponents will cheat.

Where is the third party to this contract? When a worker cheats, he lowers his chances of getting the bonus. This makes him worse off from the purely monetary standpoint. However the firm becomes no better off, as it would if the bonus were a pension as in the last section. Instead, all the other workers find that their chances of getting the bonus are increased. The third party which benefits when a worker cheats is made up of all the other workers.

Extension of this idea to the life cycle framework becomes complicated since in general it is necessary to keep track of a worker's past successes and failures to determine the prizes he should face in each period. Rosen (1986) works out the case where the contest is a single elimination tournament, such as a tennis match. The results here are simpler because losers are eliminated at each round. Without going into the details, if a predetermined number of workers are fired each period in our life cycle model, the wage profile of successful ones is again flat with a bonus at the end and an entrance fee. The last period bonus serves to motivate workers in all previous periods, just as before.

The advantage of the tournament, or promotion contest contract is that the firm has nothing to gain by being dishonest. Under the first agreement, the firm must earn a surplus so that it can be relied upon to care about its reputation. This surplus is in the form of excess profits or quasi-rents due to fixed entry or exit costs for firms. In the tournament model, this kind of a surplus is not necessary. In the long run, therefore, tournament firms should drive out the firms which compete only on the basis of their reputations. Life cycle incentive models can therefore provide another rationale for the existence of hierarchies where workers compete for promotions.

Tournaments may involve extra costs of their own, and it is important that these be considered. Workers may try to bribe or otherwise influence those responsible for their promotions (Milgrom, 1988). They may collude to jointly reduce output, or they may waste time trying to "sabotage" the efforts of other workers (Lazear, 1989). There is also the problem that tournament schemes are generally riskier for the workers than schemes which do not compare them with their peers.

It is difficult to evaluate the importance of these problems on a purely theoretical basis. In practice they do not seem to be so serious as to eliminate the use of promotion contests altogether. Part of the reason may be that these contests also serve other purposes — including perhaps that of selecting the best people for promotion into new jobs.

The theory of tournaments may provide some insights into mandatory retirement practices in Japan. In the larger Japanese firms, retirement of most "permanent" workers occurs at an average age of about 58, while social security benefits do not start until age 65. Retired workers may get a lump sum pension and move on to the outside labor market, may be retained as a temporary worker at lower wages, or may be retained as a director at high wages for some years.

These institutions can be interpreted as a self-enforcing contract which keeps workers active right up to the end of their careers. The competition for directorship positions among those workers eligible for them is reported to be quite intense (Clark, 1979). Even the other workers will have the incentive to work right up to the very end of their careers since the firm has some control over where they end up afterwards. On its part, the firm has little to gain by misrepresenting those workers who have worked the hardest. Of course, the deeper question of why Japan has one set of institutions and the West another is left unanswered.

Conclusions

Self-enforcing implicit contracts are quickly taking over in the theory of labor markets. The main reason is that they seem to capture explicitly the idea that a firm's promises to its workers might be reliable even though they are not legally enforceable. This paper has illustrated some of the important ideas in the context of a very simple incentive model. In the process it has outlined some of the accomplishments and limitations of the life cycle incentives explanation for wage growth and mandatory retirement.

We can summarize with a brief taxonomy of the things the shirking and life cycle incentive models can and cannot do. The theories: a) are consistent with the most recent evidence on wage profiles; b) are consistent with mandatory retirement at an age where it is efficient for most workers to go home; c) can provide another rationale for the use of promotion contests and other reward schemes based on relative performance; d) can provide some predictions about how the form of labor contracts changes with the state of the labor market; and e) may provide an explanation of retirement practices in Japan.

However, the theories: a) do not help account for the existence of pensions themselves, given the current laws which make them the property of the workers; and, more disturbingly, b) make the seemingly robust prediction that entrance fees should be observed in the labor market.

All in all, this is probably not too bad. The agreements made by a firm and its workers must attempt to solve a great many problems, from the provision of incentives to the assignment of workers to firms and to the jobs within a firm, and the sharing of any inherent risks. It is unlikely that a model focusing on only one of these issues will be able to account for all of the observed institutions surrounding the exchange of labor. As well, the general framework of self-enforcing contracts is proving valuable in investigating all aspects the labor market, and greater things may be on the way.

Nonetheless, there is one final caveat which must be revealed. Self-enforcing contracts as modelled here contain both explicit and implicit promises. It is clear in all of the examples that the firm's explicit contract is based on variables which are easy to verify, so that such a contract could indeed be written down and enforced. In practice, however, it is still true that explicit contracts are seldom observed. Firms will "attach

wages to jobs" and "promote from within" as a matter of policy, but this undertaking is seldom explicit outside of the (shrinking) union sector.

Given this fact, readers might suspect that self-enforcing contract theory is just an excuse to avoid the much more difficult problem of developing explicit models of reputations. There is some truth to this but, less cynically, one can hope that self-enforcing contract theory may yet turn out to be an important stop on the road to such a model. Perhaps "verifiable" variables will turn out to be those which workers outside the firm can observe most easily. A change in a verifiable variable will then have a large effect on worker expectations, and thus the firm's reputation. In this case, many of the results from self-enforcing contract theory may turn out to be robust. We are still some years away from results such as this, however, and there is much to do in the meantime.

■ *The editors, Timothy Taylor, Joseph Stiglitz, Carl Shapiro, and Ed Lazear, have all provided superb comments. Ed Nosal pointed out a problem with an earlier draft. The material on self-enforcing contracts and the shirking model is covered more extensively in my forthcoming survey of efficiency wage models, Carmichael (1990).*

References

Akerlof, George and Lawrence Katz, "Workers' Trust Funds and the Logic of Wage Profiles," *Quarterly Journal of Economics*, August, 1989, *104*, 525–536.

Altonji, Joseph and Robert Shakotko, "Do Wages Rise with Job Seniority?," *Review of Economic Studies*, July 1987, *54*, 437–59.

Azariadis, Costas, "Implicit Contracts and Underemployment Equilibria," *Journal of Political Economy*, December 1975, *83*, 1183–202.

Baily, Martin, "Wages and Employment Under Uncertain Demand," *Review of Economic Studies*, January 1974, *41*, 37–50.

Becker, Gary, and George Stigler, "Law Enforcement, Malfeasance, and the Compensation of Enforcers," *Journal of Legal Studies*, January 1974, 1–18.

Bhattacharya, Sudipto, "Tournaments, Termination Schemes, and Forcing Contracts," Bell Communications Research, Working Paper, Dec. 1986.

Brown, Jim, "Are Those Paid More Really No More Productive?," mimeo, Princeton University, 1985.

Bull, Clive, "The Existence of Self-Enforcing Implicit Contracts," *Quarterly Journal of Economics*, February 1987, 147–59.

Carmichael, H. Lorne, "The Agent Agents Problem; Payment by Relative Output," *Journal of Labor Economics*, January 1983a, *1*, 37–50.

Carmichael, H. Lorne, "Firm Specific Human Capital and Promotion Ladders," *Bell Journal*, Spring 1983b, *14*, 251–58.

Carmichael, H. Lorne, "Reputations in the Labor Market," *American Economic Review*, September 1984, *74*, 713–25.

Carmichael, H. Lorne, "Can Efficiency Wage Unemployment be Involuntary?," *American Economic Review*, December 1985, *75*, 1213–14.

Carmichael, H. Lorne, "Efficiency Wage Models of Unemployment: A Survey," *Economic Inquiry*, forthcoming, 1990.

Clark, Robert, *The Japanese Company*. New Haven and London: Yale University Press, 1979.

Cooper, Russell, "A Note on Overemployment/Underemployment in Labor Contracts Under Asymmetric Information," *Economics Letters*, 1983, 81–87.

Gordon, Donald F., "A Neoclassical Theory of Keynesian Unemployment," *Economic Inquiry*, 1974, *12*, 431–59.

Green, Jerry, and Charles Kahn, "Wage Employment Contracts," *Quarterly Journal of Economics*, 1983, *98*, 173–87.

Grossman, Sanford, and Oliver Hart, "Implicit Contracts, Moral Hazard, and Unemployment," *American Economic Review*, Papers and Proceedings, May 1981, *61*, 301–07.

Holmstrom, Bengt, "Contractual Models of the Labor Market," *American Economic Review*, Papers and Proceedings, May 1981, *61*, 308–13.

Klein, Benjamin, and Keith Leffler, "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, June 1981, *89*, 615–32.

Lazear, Edward, "Why is There Mandatory Retirement?," *Journal of Political Economy*, December 1979, *87*, 1261–84.

Lazear, Edward, "Agency, Earnings Profiles, Productivity and Hours Restrictions", *American Economic Review*, September 1981, *71*, 606–20.

Lazear, Edward, "Pay Equality and Industrial Politics," *Journal of Political Economy*, June 1989, *97*, 561–580.

Lazear, Edward, and Sherwin Rosen, "Rank Order Tournaments as Optimal Salary Schemes," *Journal of Political Economy*, 1981, 841–64.

MacLeod, Bentley, and James Malcolmson, "Involuntary Unemployment in Dynamic Contract Equilibria," *European Economic Review*, 1987, *31*, 427–35.

MacLeod, Bentley, and James Malcolmson, "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment", *Econometrica*, March 1989, *57*, 312–22.

Malcolmson, James, "Work Incentives, Hierarchy, and Internal Labor Markets," *Journal of Political Economy*, June 1984, *92*, 486–507.

Malcolmson, James, "Rank Order Contracts for a Principal with Many Agents," *Review of Economic Studies*, October 1986, 807–18.

Milgrom, Paul, "Employment Contracts, Influence Costs, and Efficient Organizational Design," *Journal of Political Economy*, February 1988, *96*, 42–60.

Rosen, Sherwin, "The Optimal Distribution of Prizes in Single Elimination Tournaments," *American Economic Review*, June 1986, *76*, 311–23.

Shapiro, Carl, "Premiums for High Quality Products as Rents to Reputations," *Quarterly Journal of Economics*, November 1983, *98*, 659–79.

Shapiro, Carl and Joseph E. Stiglitz, "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, June 1984, *74*, 433–44.

Shapiro, Carl and Joseph E. Stiglitz, "Can Unemployment be Involuntary: Reply," *American Economic Review*, December 1985, *75*, 1215–17.

Smith, Adam, *The Wealth of Nations*. New York: Random House, 1937. Reprint of the 1776 original edition.

Williamson, Oliver, "Credible Commitments: Using Hostages to Support Exchange," *American Economic Review*, June 1983, *83*, 519–40.