INTERNATIONAL ECONOMIC REVIEW
Vol. 33, No. 3, August 1992

THE INSURANCE EFFECT OF GROUPS*

BY CANICE PRENDERGAST[1]

Many people belong to groups which partially identify their characteristics. This paper argues that when agents are risk averse, and where only group characteristics are visible rather than individual characteristics, the Pareto optimal construction of groups differs in systematic ways from standard notions of how groups should be formed. Two applications are considered: the assignment of workers to jobs, where it is shown the optimal contract does not involve maximizing output and accident insurance, where the optimal contract does not involve complete insurance.

1. INTRODUCTION

Many people belong to groups which serve to partially identify otherwise hidden talents or characteristics. For example, it is possible to find out that certain authors have written an article but their individual contributions are much more difficult to ascertain. Similarly, it may be easy to determine that a worker is a foreman, but much more difficult to tell how much better he is than other foremen. Or that somebody is employed, which tells us something about his talents, but not whether he is more talented than the average employed person. Each of these examples illustrates cases where we can see the "group" to which an agent belongs (author, foreman, employed) rather than an individual's characteristics.

Standard notions of efficiency tell us how these groups should be formed. Academics should write papers together only if their joint product is greater than if they operated separately. A worker should be employer as a foreman only if his productivity is higher as a foreman than elsewhere in the organization, and so on. The principal result of this paper is that when agents are risk averse, and where only groups of the type above are visible to some agents rather than individual characteristics, these assignment rules may not be Pareto optimal. Instead, for example, someone who is less competent as a foreman than as an operative may still be employed as a foreman and this yields higher welfare than employing that worker as an operative.

Let me explain the intuition by a simple example. Assume that workers who join a firm are uncertain of their ability. They are risk averse and seek insurance over that uncertainty. Untalented workers should be fired on standard efficiency grounds as they have higher productivity elsewhere. Assume that after a period of production, information on ability becomes available to the employer and the employee but that other employers must infer this information from whether the

* Manuscript received January 1991.

[1] Many thanks are due to Bengt Holmstrom, Al Klevorick, Jeff Borland, two anonymous referees and seminar participants at Nuffield College, Oxford, for many helpful comments. All errors are my own.

worker is employed or not. Thus whether an agent is employed or not may transmit information on ability.

Other firms can bid for workers. This causes the wages of talented employed workers to be bid up. But bidding wages up is harmful for risk sharing as workers seek insurance over their ability. However, the reason that wages are bid up is that firms are inferring information from the fact that workers are laid off. Then firms may not lay off some low quality workers as this is the only way that they can be insured over their ability. If the high ability workers cannot reveal themselves to be talented, then they will not quit and so they insure the low ability workers who remain employed, which is ex ante efficient. Hence the standard efficiency rule whereby workers are laid off if their productivity is higher elsewhere no longer holds.

The result needs three ingredients; (i) agents must be risk averse and seek insurance over their "type" (for example, their ability), (ii) assignment rules (such as whether you are a foreman, an author, or employed) must transmit some information on type and (iii) agents must be able to renegotiate contracts.[2] I believe that this idea is applicable in a wide range of environments. I formally model two applications, one for an accident insurance market (Section 1) and another for a labor market (Section 2).

In Section 1, I consider the following scenario. Assume that agents seek insurance over the possibility of an accident in each of two periods. As in Rothschild and Stiglitz (1976), I assume that there are two types of agents, high risk and low risk. However, unlike Rothschild and Stiglitz, I assume that agents do not know whether they are high or low risk ex ante, being unsure how accident prone they are relative to the population. More specifically, agents who have an accident in period one are more likely to have an accident in period two than those who have no accident in period one. This puts those who do not claim at an advantage in period two if they can renegotiate their contracts by signaling that they are low risk. More specifically, those who do not claim in period one can go to another insurance company and get a contract based on the fact that they have not previously claimed.

Insurance companies are assumed to observe potential customers' claims records; they cannot observe, however, whether agents have had accidents.[3] Consider the effect of making a claim in period one. This has two effects. First, it insures the agent against the accident as in any insurance market. Second, it reveals information that the agent is likely to be high risk, which implies that he is charged a higher premium in period two than if no claim was made.

The Pareto optimal contract insures the agent over two dimensions; the likelihood of an accident and his risk characteristics. However, if those who do not claim in period one get better terms in period two than those who claim, the insurance company does not insure the agent across time because those who do not claim can renegotiate the contract in their favor in period two. In other words, agents are not perfectly insured over their risk type. Assume that other insurance companies can

[2] In the example above, (i) is satisfied as workers seek insurance over ability, (ii) holds because employment status carries imperfect information on ability and (iii) is satisfied as other firms bid up wages.

[3] In other words, insurance companies get (possibly) imperfect information on risk characteristics.

observe the set of accidents for which the firm allows claims. I show that with this assumption, the firm will not pay out on some accidents in period one. This is shown to limit renegotiation opportunities for period two, which offers better insurance over risk type than perfect insurance in period one.

This result is shown to hold with two scenarios: (i) where the agent and current insurer can observe the agent's accident history, and (ii) where only the agent knows whether he has an accident. When the agent has private information, he must be offered inducement to claim, which is not true with (i). In both cases, the intuition holds that by claiming on a small first period accident, there is a second order gain to period one utility, but a first order loss in period two to being typed with the high risk group; hence, no agent claims on a small loss.

In Section 2, I use a similar insight to analyze the assignment of workers to tasks. Assume that workers are uncertain of their ability and seek insurance over this uncertainty, as in Harris and Holmstrom (1982). Further assume that other employers infer information on ability from the position held by the employee (i.e., imperfect information on ability). Then if talented workers hold different positions (or jobs) to low ability workers, and other firms can observe the jobs which workers carry out, the wages of the most talented workers will be bid up. This is harmful from a risk sharing perspective. I show that for a technology used by Waldman (1984), the optimal assignment of workers to jobs overpromotes workers so that some workers are promoted beyond their ability. This offers better insurance than assigning workers on the basis of comparative advantage, where a worker is assigned to the job he performs best, as standard notions of efficiency suggest.

Because the optimal assignment of workers to tasks involves overpromotion, this paper can be used as an explanation of the Peter Principle (Peter and Hull 1970), which holds that workers are often promoted beyond their capacity. This paper further addresses why workers who are found to be below the required standard for the job are not subsequently demoted, as job assignment is linked to wages by competitive pressures, implying a link between productive efficiency and insurance.

At a generic level, this paper suggests that in many dynamic environments, (i) agents can generate reputations and (ii) if they are risk averse, they may seek insurance over their reputations which may affect resource allocation. One application of this idea with empirical relevance is how firms should lay off workers. Standard marginal analysis suggests that a worker should be employed if his marginal productivity exceeds what can be produced elsewhere, taking account of all investment opportunities and moving costs. However, empirical evidence in Gibbons and Katz (1991) illustrates that workers laid off at an employer's discretion typically fare worse (in terms of length of subsequent unemployment and wages) than those laid off where firms had no discretion over who to lay off (such as due to a plant closure). They attribute this result to harmful information being transmitted on worker ability by discretionary layoff. For example, reemployment earnings for white collar workers are more than 6 percent lower for those laid off than for those displaced by plant closings.

These results suggest that harmful information is revealed to the market subsequent to a layoff. As a result, risk averse workers may seek insurance over this possibility by accepting contracts which offer more job security than would otherwise occur, at the expense of lower wages.

2. ACCIDENT INSURANCE

Consider an individual who seeks insurance over the possibility of an accident in each of two periods, one and two. More specifically, consider an agent who has an endowment of e in each of the two periods but may have an accident at a cost $c \in (0, \bar{c})$ in each period. The distribution of accidents is given by a commonly known atomless distribution F with density $f > 0$ over $(0, \bar{c})$. Agents are risk averse and have preferences over income y^t in period t given by $V(y^1, y^2) = U(y^1) + U(y^2)$, where $U' > 0$, $U'' < 0$. (Discounting makes no difference to the results.)

There are two types of individuals, high risk and low risk. I assume the following reduced form. If an agent has had an accident in period one, he is high risk and has a probability $\bar{p}$ of an accident in period two. If the agent has no accident in period one, the probability of an accident in period two is $\underline{p} < \bar{p}$. I assume that these probabilities are independent of the realized value of c in period one. Let the unconditional probability of a first period accident be p.

The agent has no information before period one as to his risk type, other than the information on the population. Whether an accident occurs is private information to the agent. To keep matters simple, I assume that the agent cannot lie about the cost of the accident to his insurance company—his only option is not to report it.

In the introduction, I stressed the importance of coarse information where only imperfect indicators of characteristics were available. In this case, I assume that other insurance companies cannot observe the agent's accident report[4] but it can observe whether the agent claimed for an accident in period one. I assume perfect competition among insurance companies.

The structure of the game is as follows. At the beginning of period one, the agent pays a premium, ρ^1, and receives a payment $r^1(c^1)$ if he has an accident of cost c^1. At the beginning of period two, the agent pays a premium $\rho^2(c^1)$ and receives a payment $r^2(c^1, c^2)$. The insurance company chooses the premiums and the payments for each period. For notational simplicity, let the r^i's be net of the premium. (If $r^1 = -\rho^1$ for some $c^1 > 0$, then the insurance company does not pay out on some accidents in period one.)

By suitable normalization let the average cost of an accident be 1. Consider the set of accidents for which the firm will pay out. The insurance company will always pay out on all accidents in the final period (period two) as no more information is relevant; all losses should be recoverable. Hence I need only consider the set of accidents for which payment is made in period one. It should be clear to see that the company will pay out on all accidents above some cost $\underline{c}$ but not below $\underline{c}$. (In other words, I can rule out situations where the company pays out on accidents of costs \$100 and \$200, but not on accidents of cost \$150.) The reason is simple. Firms may not pay out on some accidents so that the high risk can mimic the low risk.

[4] More specifically, the agent cannot prove that he did not have an accident.

However, this insurance effect of belonging to the low risk group is independent of the cost of the accident (as $\bar{p}$ is independent of c^1) but the cost of not claiming in period one is increasing in its cost. Hence the firm chooses one value, $\underline{c}$ in period one, above which it pays out. Standard models of insurance suggest that $\underline{c}$ should be zero in the absence of moral hazard and transactions costs. I now illustrate another reason for minimum claims.

The insurance company's objective is to maximize the agent's expected utility $E[U(y^1) + U(y^2)]$ subject to zero profits. The optimal contract must take into account that (a) claiming in period one reveals information on risk type, and (b) that agents can renegotiate their contracts by threatening to move to another insurance company after period one.

The choice of game form is important here because Rothschild and Stiglitz have shown that there are existence problems for models of this type. To overcome the existence problem, I assume the Wilson game form (Wilson 1977). Using this game form, equilibrium is found by imposing the following conditions: (a) the insurance company must make nonnegative profits on any contract offered and (b) there cannot exist another contract which, if offered, makes positive profits even when all contracts which lose money as a result of this entry are withdrawn.

The game is solved by backward induction. I begin in period two. Because other insurance companies can see the agent's claim in period one, and the claim reveals information on the agent's risk type (since those who have accidents are high risk), other insurance companies will make offers based on whether the agent has claimed.

I begin by considering the contract offered by the other companies to agents who have not claimed in period one. I have argued that the high risk agents may not claim in order to appear like low risk agents. An agent may not claim on an accident in period one because by doing so, he looks like other low risk agents. By appearing like all other agents, he may get a lower premium in period two. But this is only useful if the other companies do not screen customers, whereby low risk customers self select a different contract to the high risk customers. If screening contracts are used, the high risk agents do not benefit from not claiming in period one as they can be distinguished from the low risk agents in the second period. However, I now show that for small enough $\underline{c}$ the other insurance companies never offer a screening contract.

The Contract Offered by Other Companies. Let the proportion of low risk agents in the group which has not claimed in period one be λ. There are two types of possible equilibria: separating and pooling. The separating equilibrium induces low risk agents to take a different contract to the high risk agents. This implies that the high risk agents get a guaranteed utility of $U(e - \overline{p})$; i.e., they are perfectly insured. In order for separation to be possible, the low risk agent must not be tempted to take the high risk agent's contract. Insurance companies offer a premium ρ^2, and a payment (net of the premium) in the event of an accident, $r^2(c)$, to the low risk agents (where the superscript on c is dropped for simplicity). The equilibrium separating contract for the low risk agents, which must separate high risk from low risk agents must satisfy

$$\max_{\rho^2, r^2} U(e - \rho^2)(1 - \underline{p}) + \underline{p} \int U(e + r^2(c) - c)\, dF(c) \tag{1}$$

subject to

$$\rho^2(1 - \underline{p}) + \underline{p} \int r^2(c)\, dF(c) \geq 0 \tag{2}$$

and

$$U(e - \bar{p}) \geq (1 - \bar{p})U(e - \rho^2) + \bar{p} \int U(e + r^2(c) - c)\, dF(c). \tag{3}$$

This program maximizes the expected utility of the low risk agent subject to the budget constraint, (2), and the separating constraint, (3), where the high risk agent prefers $U(e - \bar{p})$ to his utility with the low risk agent's contract $(1 - \bar{p})U(e - \rho^2) + \bar{p}\int U(e - r^2(c) - c)dF(c)$. Deriving the separating equilibrium is a straightforward application of Rothschild and Stiglitz, who show that low risk agents must incur strictly positive costs to screen risk type.

The other type of contract possible is where both high and low risk agents choose the same contract. Then the optimal contract maximizes the utility of the low risk agent subject to the fact that there is a proportion $1 - \lambda$ high risk agents. The optimal pooling contract solves (1) subject to

$$\lambda(1 - \underline{p})\rho^2 + (1 - \lambda)(1 - \bar{p})\rho^2 \geq (\lambda \underline{p} + (1 - \lambda)\bar{p}) \int U(e + r^2(c) - c)\, dF(c). \tag{2'}$$

This problem is easily solved by pointwise maximization and is characterized by the first order condition

$$\frac{(1 - \underline{p})U'(e - \rho^2)}{(1 - \underline{p})\lambda + (1 - \lambda)(1 - \bar{p})} = \frac{\underline{p}U'(e + r(c) - c)}{\lambda \underline{p} + (1 - \lambda)\bar{p}}. \tag{4}$$

My interest is in showing that insurance companies will not pay out on sufficiently small losses. To do so, I need to show that if $\underline{c}$ is low, the pooling equilibrium is always used. Taking the limit of (4) as λ tends to 1 (i.e., as $\underline{c}$ tends to zero) gives optimal risk sharing, $U'(e - \rho^2) = U'(e + r(c) - c)$; in other words, the low risk agents bear infinitismal cost from the existence of a small number of high risk agents. However, the low risk agents must incur strictly positive costs of signaling their type under a separating equilibrium. Hence for small enough $\underline{c}$ other insurance companies offer pooling contracts. Since the principal result is that insurance companies do not reimburse agents for small accidents, I can legitimately assume that around $\underline{c}$, pooling contracts are used.

The Optimal Pooling Contract. In period two, let the high risk agents earn $U_h(\underline{c})$ and the low risk agents earn $U_l(\underline{c})$ from the pooling contract.[5] (The dependence on $\underline{c}$ comes from the fact that $\lambda = (1-p)/[pF(\underline{c}) + (1-p)]$ in (2′) above.) In the second period, the insurance company never offers more than $U_i(\underline{c})$ to agent i if he has not claimed in period one. This is because the company never subsidizes the "winners" in the lottery over risk types. Hence those who do not claim in period one earn U_h if they have an accident and U_l if they have no accident. Next consider those who claim in period one. They have revealed themselves to be high risk. As a result, they can be given perfect insurance. However rather than offer guaranteed utility of $U(e-\bar{p})$, the firm may subsidize these agents, by making an additional payment R, so that the high risk agents get a guaranteed income of $U(e-\bar{p}+R)$, where $R \geq 0$. In other words, the firm may compensate those who have an accident in period one by a flat fee in period two. R is independent of c because the cost of the accident in period one is irrelevant for prospects in period two and because utility is separable across periods.

The firm's program is then given by

$$\max_{\rho^1, r^1(c), R, \underline{c}} \quad (1-p)U(e-\rho^1) + p\int_0^{\underline{c}} U(e-c-\rho^1)\, dF(c) \tag{5}$$
$$+ p\int_{\underline{c}}^{\bar{c}} U(e-c-r^1(c))\, dF(c) + p(1-F(\underline{c}))U(e-\bar{p}+R)$$
$$+ (1-p)U_l(\underline{c}) + pF(\underline{c})U_h(\underline{c})$$

subject to

$$p(1-F(\underline{c}))R + p\int_{\underline{c}}^{\bar{c}} r^1(c)\, dF(c) \leq \rho^1[(1-p) + pF(\underline{c})] \tag{6}$$

and

$$R \geq 0. \tag{7}$$

Thus the firm maximizes the expected utility of the agent over the two periods subject to the budget constraint (6), and the lump sum payment constraint for the high risk agent, (7). The reservation constraint for those who do not claim in period one is already incorporated in U_h and U_l, so that renegotiation opportunities are catered for.

PROPOSITION 1. *The insurance company does not insure agents on small losses in period one.*

PROOF. The first order conditions to the program above are given by

[5] U_h and U_l differ when $\underline{c} > 0$ because the pooling contract offers incomplete insurance which affects high and low risk agents differently (see (4)).

$$\frac{\partial}{\partial r^1(c)}: \; U'(e - c - r^1(c)) = \delta \tag{8}$$

for all $c \geq \underline{c}$, where δ is a Lagrange multiplier on (5),

$$\frac{\partial}{\partial R}: \; U'(e - \bar{p} - R) = \lambda + \phi \tag{9}$$

where ϕ is a Lagrange multiplier on (7),

$$\frac{\partial}{\partial \rho}: \; \frac{(1-p)U'(e-\rho^1) + p\int_0^{\underline{c}} U'(e-c-\rho^1)\,dF(c)}{(1-p)+pF(\underline{c})} = \delta \tag{10}$$

and

$$\frac{\partial}{\partial \underline{c}}: \; p[U(e-\underline{c}-\rho^1) - U(e-\underline{c}-r^1(\underline{c}))] + p[U_h - U(e-\bar{p}+R)] + (1-p)\frac{\partial U_l}{\partial \underline{c}} + pF(\underline{c})\frac{\partial U_h}{\partial \underline{c}} = -\delta(pR + pr^1(c) + p\rho^1). \tag{11}$$

Note that $R > 0$; if $R = 0$, the agent earns $U(e - \bar{p})$ in period one. But first period utility exceeds $U(e - \bar{p})$ since the firm breaks even. Hence utility can be increased by increasing second period utility, conditional on a claim in period one, while reducing all payments equally in period one. Hence $R > 0$ so that $\phi = 0$. As a result, (11) can be simplified to

$$p[U(e-\underline{c}-\rho^1) - U(e-\underline{c}-r^1(\underline{c}))] + p[U_h - U(e-\bar{p}+R)] + f(\underline{c})^{-1}(1-p)\frac{\partial U_l}{\partial \underline{c}} + pF(\underline{c})\frac{\partial U_h}{\partial \underline{c}} = U'(e-\bar{p}+R)(pR + pr^1(c) + \rho^1 p). \tag{12}$$

To show that the insurance company operates a minimum claim, evaluate (12) at $\underline{c} = 0$; at $\underline{c} = 0$, there is perfect insurance for those who do not claim in period one. The pooling contract then offers

$$U\left(e - \frac{p}{p+(1-p)F(\underline{c})}\underline{p} - \frac{F(\underline{c})(1-p)}{p+(1-p)F(\underline{c})}\bar{p}\right) \to U(e-\underline{p}) \quad \text{as} \quad \underline{c} \to 0.$$

Similarly, at $\underline{c} = 0$,

$$\frac{\partial U_l}{\partial \underline{c}} = \frac{p}{1-p} f(0)(\underline{p} - \bar{p})U'(e - \underline{p}).$$

Substituting into (12) gives

$$(13) \qquad \frac{U(e-\underline{p}) - U(e-\bar{p}+R)}{U'(e-\underline{p})} + \underline{p} - \bar{p} + \frac{U'(e-\bar{p}+R)}{U'(e-\underline{p})} < 0$$

at $\underline{c} = 0$ so that utility rises as $\underline{c}$ increases from 0. Thus the firm uses a minimum claim. □

The intuition for this result is as follows. Agents realize that by claiming in period one that they reveal themselves to be high risk. As a result, the optimal insurance contract given agents' renegotiation opportunities is to reduce the benefit of having a no-claims record. Hence the insurance company does not offer payment on small accidents. Note that if information on risk characteristics were available to other firms then all accidents would be insured; those who are high risk in period two have a utility of $U(e - \bar{p})$ while all low risk agents receive utility of $U(e - \underline{p})$ and all accidents are claimed in period one. Hence this result relies on the coarseness of information on risk types.

The Moral Hazard Problem. There is an additional moral hazard problem to contend with here as agents can choose whether to claim for losses in period one. However, this does not change the result that firms do not pay out on small losses in period one. This is not surprising, as agents themselves have no incentive to claim on a small loss in period one, as they realize the costs of being labelled high risk for period two. Hence moral hazard for the agent operates in the same direction as the case where the insurer can observe accident history. This is proved in the Appendix.

This model is an example of the principle discussed in the introduction. In dynamic environments, agents can develop reputations, which can be harmful in expected terms if agents are risk averse. Hence the optimal contract should restrict opportunities for developing reputations, even at the cost of distorting otherwise optimal arrangements. In this case, the desire by agents not to reveal their risk type results in incomplete accident insurance. In the next section, I show how similar logic can affect the assignment of workers to jobs.

3. THE ASSIGNMENT OF WORKERS TO TASKS

There has been a lot of recent interest in how workers are assigned to jobs in organizations. A number of papers, originating with Waldman (1984), have taken the position that current employers are more likely to observe a worker's ability than prospective employers who can generally observe a worker's job and his wage. The basic result from this literature, which also includes Milgrom and Oster (1987) and Ricart i Costa (1988) is (a) that it is Pareto optimal to assign workers to the jobs they perform best and (b) firms generally underpromote workers because promoting a worker reveals information to other firms that the worker is talented so his wage is bid up. Firms then underpromote to stop information on ability being revealed. In this section, I show that both (a) and (b) can be reversed using the arguments described in the introduction. In particular, I show that if firms have commitment over contracts, the insurance effects considered above can imply (a) workers are overpromoted and (b) overpromoting yields higher welfare than assigning workers to the jobs they are best at.

Consider the following scenario. A group of workers join a firm in period one and are unsure of their ability. They can be employed for two periods. The workers are risk averse and seek insurance over this uncertainty, as in Harris and Holmstrom (1982) who consider wage dynamics caused by uncertainty over ability. The workers can be assigned to one of two jobs, job A or job B. Job B is better suited to more talented workers. After one period of production, the worker's employer can infer each worker's ability. Other employers can only infer the job carried out by the employees, as in Waldman (1984), Milgrom and Oster (1988) and Ricart i Costa (1988). Because the most talented carry out job B, the wages of job B workers are bid up, which is harmful from an insurance perspective.

The first best contract would assign workers to the jobs they perform best and pay all workers the same wage, thus providing perfect insurance and assigning workers to maximize output. But the unconstrained Pareto optimum is impossible because high ability workers can quit and work elsewhere. Because direct monetary transfers are impossible to insure workers, an alternative mechanism is required. I show that for the technology chosen, the optimal insurance scheme involves overpromotion, in that some workers who carry out job A better are given job B. In this way, workers are given a higher probability of promotion than if workers are allocated by comparative advantage. Crucially, high ability workers will not quit if they have difficulty verifying their ability to the market. Since the market can only observe the job carried out by a worker, the market cannot identify which workers are the overpromoted ones and so interpersonal insurance becomes possible.

The following model is used. A perfectly competitive firm signs contracts with a continuum of workers.[6] The firm and workers exist for two periods and a contract is signed at the beginning of period one. The contract specifies a wage for each job and the proportion of workers who will carry out each job in each period. Thus the firm can contract over the shape of its hierarchy.[7] There are only two jobs, A and B, where job B is better suited to high ability workers. Workers differ in terms of their ability $a \in \Omega$, where a is defined over a support $(\underline{a}, \bar{a})$ according to a distribution F with density $f > 0$. The technology is as follows; in job A output is $x > 0$, for all a while output in job B is a. This is a simplified version of the technology used in Waldman (1984).

I assume that ability is unknown to all parties when the worker joins the firm but that it becomes known to the current employer and the worker after period one. However, this information is not known to the general market (potential employers) who must infer the worker's ability from his wage and job. Unlike Ricart i Costa (1988), I assume that the worker cannot write output contingent contracts with the general market. Workers are risk averse and value a wage stream $\{w^1, w^2\}$ by $U(w^1) + U(w^2)$, where the superscripts refer to time periods. I assume that $U' >$

[6] It is not necessary to assume that there is a continuum of workers. In Prendergast (1989), I give an example where the same intuition holds in a firm with three workers.

[7] A similar assumption is made in the tournament literature.

0, $U'' < 0$. Discounting adds nothing to the story. Firms are risk neutral and are perfectly competitive.

The structure of the game is as follows. The firm offers a contract to the worker to which the market responds with a counteroffer. The worker accepts one of the two contracts. However, he can quit his contract after the first period at no cost. This concludes the bargaining structure, which derives from Waldman (1984). The worker retires after period two.

I make the following technological assumptions.

ASSUMPTION 1. *(i) $Ea < x$, (ii) $\bar{a} > x$.*

As a benchmark, consider the case where there is no asymmetry of information between the agents in period two, i.e., there is perfect information on ability in period two. Then the unique Nash equilibrium is where (a) all workers are given job A in period one and paid x, and (b) all those above ability x are promoted to job B in period two and paid a, their ability, while all those below ability x are paid x and employed in job A. Therefore, *when individual characteristics are observable*, the usual intuition on the assignment of workers to jobs holds, namely, workers should be employed where their productivity is highest. I now show that this intuition is false when other employers can only observe the job at which workers are employed.

This benchmark yields a simple definition of overpromotion; workers are said to be overpromoted if, in addition to all workers above ability x being promoted to job B, some workers below ability x are promoted also.

The contract specifies the number of workers in each job in each period. In period one, all workers carry out job A since $Ea < x$. It should also be clear that if the contract specifies that a proportion α of the workers should carry out job B, then the firm always puts the most talented workers in that job. Any equilibrium which involves a worker of ability α carrying out job A while another of ability $\alpha' < \alpha$ carries out job B can be improved upon by the firm by switching the two workers. Hence the firm need only choose a critical level of ability α above which workers are promoted in period two. It also follows that all workers in the same job at any point in time should earn the same wage, as this has optimal risk sharing properties.[8]

The implication of these observations is that if all workers above ability α are given job B in period two, then the market will offer their expected value $Ey = \int_{\alpha}^{\bar{a}} a \, dF(a)$.[9] The principal result of this section is given in Proposition 2.

PROPOSITION 2. *All Nash equilibria of the model imply overpromotion, i.e., if all workers above ability a^* are promoted in period two, then $a^* < x$.*

PROOF. The firm's objective is

[8] If two workers who held the same job earned wages of w and w', where $w \neq w'$, both workers would prefer the certain wage of $(w + w')/2$.

[9] More correctly, this should be $Ey = \max(x, \int_{\alpha}^{\bar{a}} a dF(a))$ but the firm never assigns workers to job B if their average productivity is lower than x.

$$\max_{a^*, w^1, w_A^2, w_B^2} \quad U(w^1) + F(a^*)U(w_A^2) + [1 - F(a^*)]U(w_B^2) \tag{14}$$

where w^1 is the wage paid in period one and w_i^2 is the wage paid in job i in period two. This is maximized subject to the job B worker's individual rationality constraint

$$w_B^2 \geq \int_{a^*}^{a} a \, dF(a) \tag{15}$$

and the budget constraint

$$w^1 + F(a^*)w_A^2 + [1 - F(a^*)]w_B^2 \leq x[1 + F(a^*)] + \int_{a^*}^{\bar{a}} a \, dF(a). \tag{16}$$

The solution to this program is given by (i) $w^1 = x$, (ii) $w_A^2 = x$, (iii) $w_B^2 = Ey$, and

$$x - a^* = x - w_B^2 - \frac{U(x) - U(w_B^2)}{U'(w_B^2)}. \tag{17}$$

A necessary condition for overpromotion is that

$$x - w_B^2 - \frac{U(x) - U(w_B^2)}{U'(w_B^2)} > 0, \quad \text{where} \quad w_B^2 = \int_{x}^{\bar{a}} a \, dF(a). \tag{18}$$

But for any increasing concave function, (18) is always satisfied by the First Mean Value Theorem. Hence all equilibria involve overpromotion. □

The intuition for this result is very simple. From the "productively efficient" point where all above x are promoted, there are second order losses to promoting slightly more workers. However there are first order gains to risk sharing so that overpromotion leads to higher expected utility to workers. Note that it is impossible to give direct insurance purely through the wage rate as good workers quit.

Note that this result can explain the Peter Principle (Peter and Hull 1970), which contends that workers are promoted beyond their competence. The Peter Principle holds that workers continue to be promoted until such point as they can no longer carry out their job competently. This paper also provides an explanation of why these workers are not demoted after their abilities are subsequently revealed. Here overpromotion is optimal; there are no efficiency gains from demoting the less able as the productivity gains from reassigning these workers are outweighed by the losses caused by worse insurance.[10]

REMARK 1. The result that workers are overpromoted does not necessarily extend to other technologies. Instead workers may be underpromoted. In this

[10] One strategy suggested in Peter and Hull (1970) to ameliorate the Peter Principle is to demote workers to their old jobs but to retain their title and salary. This would be consistent with the risk-sharing benefit of overpromotion, but relies on the market misunderstanding the worker's title.

example, workers are insured by including some low quality workers with the high quality workers in job B. In this way, those of ability slightly less than x benefit from being assigned the same job as those above ability x. However for other technologies, the optimal contract may imply including some high quality workers with the low quality workers in job A. In other words, instead of mixing some bad workers with the good, the firm may mix some good workers with the bad. In the example above, output in job A is x, independent of ability, so that the low ability workers do not benefit from the inclusion of high ability workers in that job. This is not true for other technologies. More specifically, if output in job A rises with ability, the low ability workers benefit from the inclusion of high ability workers in job A, so that underpromotion may be optimal. In Prendergast (1989), I show that in general there is an incentive to distort the assignment of workers from that which maximizes output.

REMARK 2. Another caveat is that the results are not robust to other bargaining games. Consider a case where the current employer makes the final offer. Then the uninformed party (the market) makes the second from final offer, w, for the worker. If the worker is worth more than w, the firm responds with a better offer. If he is worth less, the firm does not respond and the market loses money on the worker. As a result, the market can only lose money on any worker it hires. This phenomenon, which is known as the Winner's Curse problem, is well known in the auctions literature. This implies that the market will never make an offer above x, solving the renegotiation problem.[11]

REMARK 3. Another relevant issue is that workers' skills are not perfectly transferable across firms; instead, workers may match better with some firms than others. Consider an alternative to the technology above where I assume that productivity in the firms is as above, but "ability" in other firms is given by $\tau Ea + (1 - \tau)a$. Hence with probability τ, performance in another firm is uncorrelated with performance in the firm and with probability $(1 - \tau)$, performance is perfectly correlated.

Assume that the firm promotes all workers above ability α to job B. The market values those workers at $\max(x, \int_{\alpha}^{\bar{a}} a\, dF(a))$ and those not promoted are valued at $\max(x, \int_{\underline{a}}^{\alpha} a\, dF(a))$. The results described above continue to hold once τ is sufficiently small. If τ is small, then assignment to job B says little about how well the worker performs elsewhere, reducing renegotiation opportunities. Hence the result relies on correlation of talents across firms.

REMARK 4. I mentioned above that this result reverses some of the findings in the literature (that underpromotion occurs) by showing how overpromotion can occur. This occurs for two reasons; first, the firm has commitment power and second, workers want insurance. Both effects are absent from the paper this model most closely resembles, Waldman (1984). The principal result in Waldman is that workers are underpromoted. This result disappears if the firm has commitment power and a sufficient number of workers. In that case, the firm promotes the

[11] One way of reintroducing the issue of renegotiation is by assuming that workers quit for reasons exogenous to the wage offer. See Greenwald (1986).

number of workers which maximizes output. The insight of this model is that they may promote more for insurance reasons and that this yields higher welfare than promoting the output maximizing number of workers. However, if commitment power by firms is eliminated, then this model collapses to that in Waldman.

This model is part of a wider literature on how workers should be assigned to jobs. MacDonald (1984) considers how information arrives over time on worker's comparative advantage, and the effect that this has on wage dynamics. Waldman (1989) and O'Flaherty and Siow (1988) examine the use of up-or-out contracts where firms either promote workers after a probationary period or fire them. MacLeod and Malcomson (1989) examine a hierarchy with both adverse selection, where workers have private information on their ability, and moral hazard to examine the development of careers for workers of different abilities within organizations. The focus of this model is rather different; I ignore many of the learning and incentive issues addressed in these papers in order to stress the role of the labor market for hierarchies.

4. DISCUSSION

This paper gives two examples of how observing only imperfect information on agent's types can aid insurance and can result in policies which appear at odds with efficiency. At face value it would appear inefficient not to lay off workers if their productivity is higher elsewhere. Yet empirically those who are laid off at management's discretion seem to suffer a stigma so that a reluctance to lay off may in fact offer optimal insurance in the face of firms making inferences from firing decisions.[12] Similarly, promoting workers once they have senority certainly appears inefficient, but it may stop information being revealed as to who the most talented workers are, aiding insurance.[13]

I believe that the principal insight of the paper is that rules which at times appear inefficient (such as not insuring fully, overpromoting, reluctance to lay off etc.) may be more efficient than are sometimes perceived. There is no doubt that there are moral hazard problems which can be added to these models, such as negligence in the accident insurance case or shirking in labor markets. The have been ignored here to more simply illustrate the insurance issues of interest.

University of Chicago, U.S.A.

APPENDIX

I now consider the fact that agents may choose not to report accidents in period one. Let C' be the set of accidents for which the insurance company pays out in period one. Let $U_i(\lambda)$ be the utility of agent i offered by another insurance company in period two if there has been no claim in period one, where λ is the

[12] It might be thought that this problem can be solved by suitable use of severance payments. In Prendergast (1989) it is shown that this is not generally the case.

[13] For more details on this example, see Prendergast (1989).

proportion of high risk individuals in the set C'. The firm's objective is to maximize expected utility subject to the same constraints as in program (1) plus, for all $c \in C'$,

$$U(e - c - \rho^1) + U_h(\lambda(C')) \leq U(e - c + r^1(c)) + U(e - \bar{p} + R). \tag{A.1}$$

The expression on the left-hand side of (A.1) is the utility given that the agent does not claim, while the right-hand side gives the utility after a claim.

The proof of the proposition that the firm does not pay out on small losses is as follows. As in Proposition 1, the firm offers a contract where $e - c - r^1(c) = e - \bar{p} + r^2(c)$. Let this utility be $\underline{U}(c)$. Then since $\rho^1 > 0$, $U_1(C') > U(e - \rho^1)$ which implies that $U(e - \rho^1) < \underline{U}(c)$. Suppose that ε is included in C', where ε is arbitrarily close to zero. Then eliminate ε from C'. This causes a second order loss in utility from not claiming for the accident. If ε is excluded from C', the firm has additional revenue of $f(\varepsilon)(r^1(\varepsilon) + R)$. Let this revenue be used to reduce ρ^1 but leave r^1 unchanged. This causes a first order gain in utility (by providing better insurance over risk type), outweighing the second order loss from not claiming. Hence small accidents will not be claimed. □

REFERENCES

GIBBONS, R. AND L. KATZ, "Layoffs and Lemons," Working Paper, National Bureau of Economic Research, 1988.

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

HARRIS, M. AND B. HOLMSTROM, "A Theory of Wage Dynamics," *Review of Economic Studies* 49 (1982), 315–333.

MILGROM, P. AND S. OSTER, "Job Discrimination, Market Forces and the Invisibility Hypothesis," *Quarterly Journal of Economics* 102 (1988), 453–477.

MACDONALD, G., "A Market Equilibrium Theory of Job Assignment and Sequential Accumulation of Information," *American Economic Review* 72 (1982), 1038–1055.

MACLEOD, W. B. AND J. MALCOMSON, "Reputation and Hierarchy in Dynamic Models of Employment," *Journal of Political Economy* 96 (1988), 832–854.

O'FLAHERTY, B. AND A. SIOW, "On-the-job Screening, Up or Out Rules and Firm Growth," Discussion Paper No. 421, Columbia University, 1989.

PETER, L. AND R. HULL, *The Peter Principle* (New York: Mc. Morrow Williams, 1970).

PRENDERGAST, C., *Theories of Internal Labor Markets*, Ph.D. Dissertation, Yale University, 1989.

RICART I COSTA, J., "Managerial Task Assignment and Promotion," *Econometrica* 56 (1989), 49–66.

ROTHSCHILD, M. AND J. STIGLITZ, "Equilibrium in Competitive Insurance Markets: An Essay in the Economics of Imperfect Information," *Quarterly Journal of Economics* 90 (1976), 629–649.

WALDMAN, M., "Job Assignment, Signaling and Efficiency," *Rand Journal of Economics* 15 (1984), 255–270.

———, "Up or Out Contracts: A Signaling Perspective," Working Paper No. 170, Institute of Industrial Relations, University of California, Los Angeles, 1989.

WILSON, C., "A Model of Insurance Markets with Incomplete Information," *Journal of Economic Theory* 16 (1977), 167–207.