THE OPTIMAL LEVEL OF CORPORATE LIABILITY GIVEN THE LIMITED ABILITY OF CORPORATIONS TO PENALIZE THEIR EMPLOYEES

Steven Shavell

Discussion Paper No. 160

6/95

Harvard Law School
Cambridge, MA   02138

The Optimal Level of Corporate Liability
Given the Limited Ability of Corporations
To Penalize Their Employees

Steven Shavell*

Abstract

A central conclusion of the economic analysis of corporate liability is that the level of liability should generally equal harm. However, this result does not necessarily hold when the ability of corporations to impose penalties on employees for causing harm is limited (usually at most dismissal from their jobs), implying that employees' motives to prevent harm may be inadequate. Firms can partially remedy this problem by paying employees an above-market wage, because that will raise employees' desire to keep their jobs. But a firm's incentive to pay such a supernormal wage deviates from the socially desirable incentive to pay supernormal wages. This leads to the result that optimal level of liability can be either above or below harm.

*Harvard Law School.

The Optimal Level of Corporate Liability
Given the Limited Ability of Corporations
To Penalize Their Employees

Steven Shavell[*]

1. Introduction

A central point of the economic analysis of corporate
liability is that the level of liability should generally equal
harm. The rationale is twofold. If liable firms must pay
damages[1] equal to harm, then, first, firms will in principle be
led to take appropriate care to prevent harm; and, second,
product prices will tend to reflect the full social cost of
production, inducing consumers to make socially correct purchase
decisions.

Somewhat surprisingly, however, this familiar conclusion
about optimal damages does not necessarily hold when it is
recognized that the ability of corporations to impose financial
penalties on employees for causing harm is limited; the major
sanction suffered by a misbehaving employee is usually at most
dismissal from his job.[2] If employees have less at stake than

---

[1]As is conventional, the term "damages" refers to the level of liability,
the amount a liable party has to pay.

[2]Other sanctions are of course possible; an employee might be denied
pension benefits or even part of his assets (suppose that he were sued
personally by an injured party) as a consequence of his behavior. But
considering sanctions in addition to dismissal and loss of wages would not
alter the points to be made here, as they depend only on the assumption that
the employee might cause harm exceeding the sanction he would suffer.

the harm they might cause, their motive to prevent harm may be inadequate from a firm's perspective. Firms can partially remedy this problem by paying employees an above-market wage, because that will raise employees' desire to keep their jobs and thus to prevent accidents.[3] But, as will be discussed, a firm's incentive to pay a supernormal wage deviates from the socially desirable incentive to pay supernormal wages. This leads to the result that *optimal damages can be either above or below harm.*[4]

To amplify, suppose provisionally that corporate damages equal harm. Then it is true that a firm will want its employees to take care whenever that is socially desirable, whenever the cost of care is less than the expected reduction in harm that it brings about. But a firm faces a difficulty in influencing an employee to exercise proper care (assuming that the firm cannot directly control his level of care) if the employee's wage loss

---

[3]The payment of above-market wages to induce employees to exercise extra care to avoid accidents is similar to payment of "efficiency wages," above-market wages to induce employeees not to shirk on the job. On efficiency wages, see, for example, Shapiro and Stiglitz (1984), Weiss (1990), and Yellen (1984).

[4]The conclusion reached here contrasts with those of several previous articles considering optimal corporate damages: Newman and Wright (1990) demonstrated that optimal corporate damages equal harm under the implicit assumption that employees are able to pay for harm; and Polinsky and Shavell (1994) noted that optimal damages equal harm in a model allowing for employees to be unable to pay for harm, but explicitly omitted consideration of supernormal wages. (Several earlier articles, notably, Kornhauser (1982) and Sykes (1981), investigate models of corporate liability, but focus on issues apart from optimal damages. Also, Segerson and Tietenberg (1992) study corporate liability, but place emphasis on the question whether liability should be imposed on employees as well as on firms.) The conclusion obtained here is also different from that contained in an interesting article by Pitchford. Pitchford (1995) examines a model of vicarious liability on lenders for harms caused by judgment-proof firms and finds that damages equal to harm may not be optimal. The reason is that the higher are the damages imposed on lenders, the more they will charge firms for loans, and thus the less assets firms will have remaining; this will lower firms' incentives to reduce risk.

from dismissal is less than the harm he might cause. The wages of the employee who sits at the control console of a nuclear power plant will be far less than the hundreds of millions of dollars of injuries that he could engender through an error. Such an employee's degree of care may well be suboptimal, less than what his firm and society would wish.[5] To combat this difficulty, a firm may decide to pay an employee an above-market wage. If the employee receives a supernormal wage, he will have more to lose if he commits error and is dismissed, so he will take more care than if he were paid only the market wage. The gain to a firm from payment of a supernormal wage -- the reduction in its expected liability due to the greater level of care of the employee -- may outweigh the added wage cost to the firm. The market wage of the nuclear power plant operator may be $50,000, yet it may be worth the firm's while to pay him $70,000 because the marginal care he will exercise in order not to lose the $70,000 job may reduce the firm's expected liability by more than $20,000.

The incentive of a firm to utilize a supernormal wage to spur care from employees departs from the social incentive to use supernormal wages for two reasons. First, *for a firm, supernormal wages are an added expense, whereas for society, supernormal wages are costless -- they are mere transfer payments.* The extra $20,000 the power plant operator may be paid

---

[5]His level of care will definitely be suboptimal if he is risk-neutral, which is what is assumed for simplicity in the model that is investigated.

would not represent a social cost but rather a shift of purchasing power from the firm to an employee. It is only the operator's $50,000 market wage that measures his social cost, because this wage corresponds to his productivity in an alternative job. As supernormal wages are costly to a firm but not directly to society, there is a reason to believe that a firm will not raise supernormal wages as much as would be socially desirable to stimulate greater care from employees (perhaps it would be best for the power plant operator to receive $120,000 instead of $70,000).

Second, *supernormal wages are reflected in product prices and make product prices exceed social cost. This is of no moment to an individual firm even though it is socially detrimental because it undesirably discourages purchases.* If wages are $70,000 or some other supernormal level, this will be reflected in product prices, as would any cost to the firm. But, as just emphasized, the supernormal component of wages does not represent a social cost of production, so that product prices will tend to be too high, depressing purchases undesirably. This factor might make it socially advantageous for the supernormal wage to be lowered from the level the firm would establish if damages equal harm, and thus for society to induce firms to decrease supernormal wages by reducing damages to a level below harm.

In the next section, I investigate the above issues in a simple model of firm and employee behavior, and I then conclude with remarks on the interpretation of the analysis.

4

2. The Model

Firms hire identical employees to produce a product; both firms and employees are risk neutral.[6] Employees incur disutility from work. Employees also select a level of care; the higher is the level of care, the lower is the probability of an accident that causes harm to others; firms cannot observe employees' levels of care. Let

k = disutility of work per unit of output; $k > 0$;[7]

x = level of care of an employee, $x \geq 0$;

$c(x)$ = cost of care per unit of output; $c(0) = 0$, $c'(0) = 0$, $c''(0) > 0$;[8]

$p(x)$ = probability of an accident per unit of output:

$0 < p(x) < 1$; $p'(x) < 0$; $p''(x) > 0$;

h = harm if an accident occurs, $h > 0$.

Individuals obtain utility from consumption of the firms' product. Define

---

[6] Were I to allow parties to be risk averse, the qualitative nature of the conclusions would not be altered (but the analysis would be needlessly complicated). Notably, although a risk-averse employee would tend to take more care than a risk-neutral one, he would still take less than optimal care if his assets were sufficiently low in relation to harm and the firm's incentive to pay a supernormal wage would still deviate from the socially optimal incentive in the two ways described in the introduction.

[7] Alternatively, I could define k as the wage and the marginal social product of employees were they to work in another industry (the interpretation made in the introduction).

[8] The present formulation, with the cost of care being distinguished from the level of care, is used chiefly for expositional reasons: because $c'(0) = 0$, there will be an interior solution for the choice of care whenever an employee will suffer a positive loss in wages for causing an accident. Under the more usual formulation, where in effect $c(x) = x$, an employee's choice of care would be zero for all positive wage losses sufficiently small. To take this into account would not change the conclusions but would cause me to make tedious qualifications to the arguments.

q = quantity of output consumed per individual;

u(q) = utility of an individual; $u'(q) > 0$; $u''(q) < 0$.

Social welfare is defined to be the utility a representative individual derives from output minus the costs of production -- the disutility of work, the cost of care, and the expected harm associated with production. Thus, social welfare is given by

(1) $$W = u(q) - q[k + c(x) + p(x)h].$$

The socially optimal outcome is identified by the level of care x* and the quantity q* that maximize (1). It is apparent that x* minimizes $k + c(x) + p(x)h$, or equivalently, x* minimizes

(2) $$c(x) + p(x)h;$$

therefore x* is determined by the first-order condition

(3) $$-p'(x)h = c'(x).$$

More generally, let us denote by x*(h) the socially optimal x given h. The socially optimal q maximizes $u(q) - q[k + c(x*) + p(x*)h]$ and is determined by the condition

(4) $$u'(q) = k + c(x*) + p(x*)h.$$

This has the usual interpretation that marginal utility equals social marginal cost.

Firms are assumed to be strictly liable for accidents caused by their employees.[9] Let

---

[9] If firms were instead liable under the negligence rule -- meaning that firms are liable if and only if x < x* -- the analysis would become more elaborate but the thrust of the conclusions would be unaltered. The main differences would be as follows. First, the problem of inadequate employee care would not be as severe, because an actor's motive to take care under the negligence rule is sharper than under strict liability; see Shavell (1987), at 179-182. Second, use of supernormal wages would not necessarily raise price above social marginal cost, because if employees are non-negligent, firms escape liability for harm caused; see generally Shavell (1980).

d = damages that a firm must pay if there is an accident.

Firms select the wage that they pay their employees. The wage is assumed to depend only on whether an accident occurs (because firms cannot observe their employees' levels of care). Specifically, define

w_a = wage if there is an accident, $w_a \geq 0$;

w_n = wage if there is not an accident, $w_n \geq 0$.

(The latter wage can be thought of as a base wage from which a penalty is subtracted in the event of an accident.) It is assumed for simplicity that an employee has no assets, so that the wage cannot be negative.[10] A firm chooses wages w_a and w_n to minimize its expected costs per unit of output, including its liability costs, for this is necessary to maximize profits. Expected unit costs are

(5)
$$(1 - p(x))w_n + p(x)(w_a + d).$$

Firms minimize (5) over w_a and w_n subject to two constraints. First, an employee's expected net wage -- his expected wage net of his cost of care -- must be at least equal to his disutility of work,

(6)
$$(1 - p(x))w_n + p(x)w_a - c(x) \geq k.$$

If the left side of (6) exceeds k, an employee is said to earn a *supernormal wage*. Second, an employee chooses care x to maximize his expected net wage. Since an employee's expected net wage is

---

[10]Were the assumption instead that an employee has positive assets, the essential nature of the conclusions would not be altered. For instance, Proposition 1 would still hold; the only difference would be in the location of the thresholds d_1 and d_2.

the left side of (6), x is determined by

$$(7) \qquad\qquad -p'(x)(w_n - w_a) = c'(x),$$

which has the interpretation that the marginal expected

enhancement in the wage due to a reduction in the likelihood of

an accident equals the marginal cost of care. Because

minimization of (5) subject to (6) and (7) determines w_a, w_n, and

x as functions of d, we may write $w_a(d)$, $w_n(d)$, and $x(d)$, and we

have

PROPOSITION 1. (a) If damages d are less than or equal to a

threshold d_1: supernormal wages are not paid, the wage if an

accident does not occur exceeds the wage if an accident occurs by

d (that is, $w_n(d) - w_a(d) = d$), and the level of care minimizes

the cost of care plus expected liability (that is, $x(d) = x*(d)$).

(b) If damages are in the interval between d_1 and another,

higher threshold d_2: supernormal wages are not paid, the wage if

an accident does not occur is d_1, the wage if an accident occurs

is zero, and the level of care is $x*(d_1)$, which is less than

$x*(d)$.

(c) If damages exceed the threshold d_2: supernormal wages

are paid, the wage if an accident does not occur exceeds d_1, the

wage if an accident occurs is zero, and the level of care exceeds

$x*(d_1)$ but is less than $x*(d)$.

Notes. This result is illustrated in Figure 1 and is

explained as follows. (The proof is in an appendix.) A firm can

create incentives for an employee to take care, and thus lower

its expected liability, by reducing the wage if an accident

8

care x

$w_n(d) - w_a(d) = d$ $w_n(d) = d_1$ $w_n(d) > d_1$
                         $w_a(d) = 0$ $w_a(d) = 0$

$x^*(d)$

$x(d)$

$x(d) = x^*(d_1)$

$x(d) = x^*(d)$

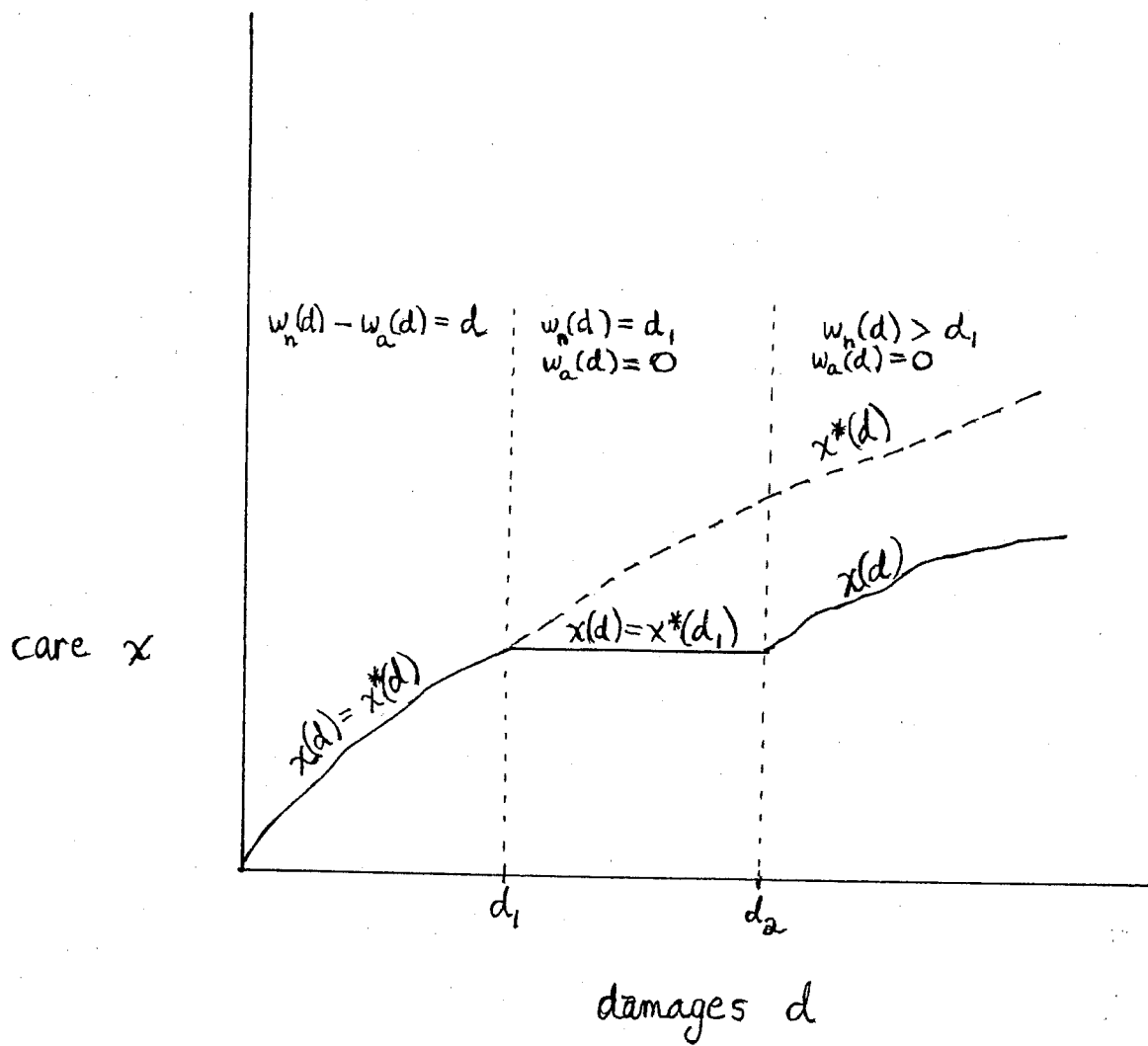d_1 d_2

damages d

FIGURE 1

occurs. In particular, if a firm sets wages so that they fall by d if an accident occurs, an employee will choose care as the firm would wish, so as to minimize the costs of care plus expected liability, $c(x) + p(x)d$; that is, the employee will choose $x^*(d)$. Therefore, a firm will want to choose wages in this way as long as the gap d in wages is not too large. If, however, d is sufficiently large, then it will be impossible for the firm to create a gap in wages of d, even if it sets w_a equal to zero, without making w_n so high that the expected net wage exceeds k. This explains why, if d is below a threshold d_1, the gap in wages will be d, and it suggests that if d is sufficiently high, the gap will not be d. The reason that there is a second threshold d_2 that must be surpassed before the firm begins to pay a supernormal wage is somewhat subtle and is best deferred to the proof.◄

Now let us describe the determination of product price and demand for the product. Firms are assumed to be in competition, so that price will equal minimized unit cost. That is, if we let

π = product price,

then

(8) $\pi = (1 - p(x))w_n + p(x)(w_a + d)$.

Because a consumer chooses q to maximize $u(q) - \pi q$, quantity purchased is determined by $u'(q) = \pi$, or

(9) $u'(q) = (1 - p(x))w_n + p(x)(w_a + d)$.

This condition implicitly determines q as a function of d, since x, w_n, and w_a are functions of d.

We may thus rewrite social welfare (1) as a function of d,

(10) $W(d) = u(q(d)) - q(d)[k + c(x(d)) + p(x(d))h]$.

Hence,

(11) $W'(d) = q'(d)\{u'(q(d)) - [k + c(x(d)) + p(x(d))h]\}$

$\qquad -q(d)x'(d)[c'(x(d)) + p'(x(d))h]$.

The first term in (11) is the effect of a change in d on social welfare due to a change in purchases. Note that the factor in braces, and thus the first term, will be zero whenever price equals social marginal cost. In particular, if supernormal wages are not paid when d = h, price will equal social marginal cost (from (6) and (8), it is clear that π will equal $k + c(x(d)) + p(x(d))h$). However, the factor in braces will be positive whenever price exceeds social marginal cost, so that the first term will be negative ($q'(d)$ is negative because a rise in d raises prices and lowers demand). This will be so if a supernormal wage is paid when d = h. The second term in (11) is the effect of a change in d on welfare due to a change in the level of care. This will be positive if the level of care is below x*(h). Let us next state how optimal damages and the optimal solution relate to the magnitude of harm.

PROPOSITION 2. (a) If harm h is less than or equal to the threshold d_1, optimal damages equal h. In this case, supernormal wages are not paid and the level of care and the quantity produced are socially optimal.

(b) If harm is in the interval $(d_1, d_2]$, optimal damages are either equal to h or exceed d_2. If optimal damages equal h, then

10

supernormal wages are not paid, the level of care is suboptimal, but the quantity produced is optimal given the level of care. If optimal damages exceed d_2, then supernormal wages are paid, the level of care is suboptimal, and the quantity produced is suboptimal.

(c) If harm exceeds d_2, optimal damages may be above or below harm. In either case, damages are greater than d_2, supernormal wages are paid, the level of care is suboptimal, and the quantity produced is suboptimal.

Notes. We can explain why optimal damages are as claimed (most other aspects of the proposition follow from knowledge of optimal damages) using the previous proposition and the interpretation of (11). (The proof of the proposition is again given in the appendix.) First, when $h \leq d_1$, we know from Proposition 1(a) that if $d = h$, firms will create a gap of h between the non-accident and accident wages, without use of supernormal wages, so that the optimal outcome will result. Hence, the optimal d must be h.

When h is in $(d_1, d_2]$, we know from Proposition 1(b) that if $d = h$, the wage gap will be d_1, the level of care will be $x*(d_1) < x*(h)$, and supernormal wages will not be paid. Moreover, we know, as illustrated in Figure 1, that the level of care will equal $x*(d_1)$ for all d in $(d_1, d_2]$. This means that $x'(d)$ is zero in the interval, so that the second term of (11) is zero. We also know from Proposition 1(b) that supernormal wages are not paid for d in the interval, so that in particular for $d = h$, the

11

first term of (11) is zero. Thus $W'(h) = 0$, suggesting that $d = h$ is a local maximum. It is possible, however, that setting d above d_2 is superior to $d = h$: raising d will lead to a higher x, improving social welfare, and this effect might dominate the opposing effect that price will exceed social cost, lowering purchases.

When h exceeds d_2, Proposition 1(c) tells us that if $d = h$, the wage gap will be less than h, the level of care will be less than $x^*(h)$, and supernormal wages will be paid. Because of the latter, the first term in (11) will be negative; raising d will increase the difference between price and social marginal cost, lowering social welfare. And because the level of care will be inadequate, the second term in (11) will be positive; raising d will increase care and increase social welfare. This suggests that $W'(h)$ could be positive or negative, and thus that the optimal d could be above or below h.

3. Concluding Comment

Although it has been shown that optimal damages may be different from harm, this result should be cautiously interpreted for a variety of reasons. First, it was implicitly assumed that the only instrument the state could use to affect social welfare was the level of damages. However, if the state is able to impose criminal sanctions on employees, their behavior can be improved.[11] Further, if the state is allowed to subsidize wages and product prices, it can achieve its objectives without

---

[11]This is the main point of Polinsky and Shavell (1993).

altering damages from their usual level equal to harm: the state can induce greater employee care by contributing to the non-accident wage; and the state can offset any undesired increase in price associated with the payment of supernormal wages with an appropriate product price subsidy.

Second, for the state to know how to adjust damages optimally as described in the analysis here, it has to possess a great deal of information: it must know the function $p(x)$, relating employee care to the accident probability, and also $u(x)$, embodying consumer demand. (By contrast, the state need only know the harm h to set damages equal to the usual level h.) Because the state will often have only imperfect information about $p(x)$ and $u(x)$, the expected benefits derived from setting d unequal to h will be lower than those in the analysis.

Third, in the analysis, there was only a single type of employee, and there were no actions that firms could take directly affecting safety. Realistically, however, there will be many types of employee, some of whom will have adequate incentives to take care, and there will also be equipment and other investment decisions that firms can make that directly promote risk reduction. In respect to these employees and risk-reducing decisions, use of damages unequal to harm will have undesirable effects. Setting damages above harm, for example, will lead some employees to take excessive, defensive precautions, and will induce firms to invest unduly in safety; these disadvantages may outweigh any benefits from altering the

13

behavior of employees whose care is inadequate.

In the end, therefore, the best course for society may well be to set damages equal to harm.

References

Kornhauser, Lewis A., "An Economic Analysis of the Choice Between Enterprise and Personal Liability for Accidents," *California Law Review*, Vol. 70, 1982, 1345-1392.

Newman, Harry A., and David W. Wright, "Strict Liability in a Principal-Agent Model," *International Review of Law and Economics*, Vol. 10, 1990, 219-231.

Pitchford, Rohan, "How Liable Should a Lender Be? The Case of Judgement-Proof Owner-Managed Firms", forthcoming, *American Economic Review*, 1995.

Polinsky, A. Mitchell, and Steven Shavell, "Should Employees be Subject to Fines and Imprisonment Given the Existence of Corporate Liability?" *International Review of Law and Economics*, Vol. 13, 1993, 239-257.

Segerson, Kathleen, and Tom Tietenberg, "The Structure of Penalties in Environmental Enforcement: An Economic Analysis," *Journal of Environmental Economics and Management*, Vol. 23, 1992, 179-200.

Shapiro, Carl, and Joseph E. Stiglitz, "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, Vol. 74, 1984, 433-444.

Shavell, Steven, *Economic Analysis of Accident Law*, 1987, Harvard University Press, Cambridge.

Shavell, Steven, "Strict Liability versus Negligence," *Journal of Legal Studies*, Vol. 9, 1980, 1-25.

Sykes, Alan O., Note, "An Efficiency Analysis of Vicarious Liability Under the Law of Agency," *Yale Law Journal*, Vol. 91, 1981, 168-206.

Weiss, Andrew, *Efficiency Wages*, 1990, Princeton University Press, Princeton.

Yellen, Janet L., "Efficiency Wage Models of Unemployment," *American Economic Review*, Vol. 74, 1984, 200-205.

Appendix

Proof of Proposition 1. Part (a): Let d_1 be defined implicitly by the conditions

(A1) $(1 - p(x))d_1 - c(x) = k$,

(A2) $-p'(x)d_1 = c'(x)$.

The interpretation of the conditions is that if $w_n = d_1$ and $w_a = 0$, an employee who chooses his x will be just willing to work. Note that d_1 is uniquely determined and must be positive.[12]

Now the following two claims are readily established: (i) If $d \le d_1$, there exist w_n and w_a such that $w_n - w_a = d$ and such that (6) holds with equality.[13] (ii) The problem of minimizing (5) subject only to (6) has as a solution $x*(d)$ and any w_n and w_a satisfying (6) with equality.[14] It follows from (i) and (ii) that for any $d \le d_1$, a firm will choose wages such that $w_n - w_a = d$ and that employees will choose $x*(d)$. Specifically, a firm's problem is to minimize (5) subject to (6) and (7). Thus, certainly if w_n, w_a, and x minimize (5) subject only to (6), and they also satisfy (7), they must solve the firm's problem. But (ii) describes solutions to minimization of (5) subject only to

---

[12]Uniqueness follows because the left side of (A1) is an employee's expected utility, and this is clearly increasing in w_n. Hence, there can be only one level of w_n, namely d_1, for which the left side of (A1) equals k.

[13]At d_1 the claim is true by definition (set $w_a = 0$). If $d < d_1$, let $w_n = d + t$ and $w_a = t$; the claim will be established if we can show (6) holds with equality for some $t \ge 0$. If $t = 0$, the left side of (6) is less than k (for the left side equals k if $d = d_1$, and the employee must be worse off if w_n is lower than d_1). Thus, if t is increased sufficiently, (6) will be satisfied.

[14]Condition (6) must hold with equality, for otherwise w_n and w_a could be lowered, reducing (5). Since (6) must hold with equality, we may substitute from (6) to obtain that (5) equals $k + c(x) + p(x)d$. This is minimized if and only if x is $x*(d)$.

(6), and (i) tells us that such a solution can be found that also satisfies (7).

Parts (b) and (c): I claim that $w_a = 0$ for any $d > d_1$. Assume otherwise, that $w_a > 0$, and suppose that a supernormal wage is paid. This leads to a contradiction, since then w_a and w_n could each be reduced by a positive ϵ (it is clear that w_n must also be positive because $w_n > w_a$ must be true; I omit details), maintaining an employee's choice of x, but reducing a firm's unit costs. Hence, if $w_a > 0$, it must be that no supernormal wage is paid. But this implies that $x < x*(d)$.[15] This in turn can be shown to lead to a contradiction because it means that a firm's unit costs could be lowered by an appropriate adjustment in wages.[16]

Because w_a can be taken to be zero, a firm's problem for $d > d_1$ is

(A3) $\min (1 - p(x))w_n + p(x)d$

over w_n such that

(A4) $(1 - p(x))w_n - c(x) \geq k$,

---

[15] If $x \geq x*(d)$, then $w_n - w_a \geq d$. Now lower w_a to zero and w_n to $w_n' = w_n - w_a$. This will preserve the employee's choice of x, but the employee will be worse off by w_a. Next raise w_n' until (6) holds with equality. This construction results in (A1) and (A2) being satisfied at a w_n' exceeding d, which exceeds d_1 -- contradicting the uniqueness of d_1.

[16] If $w_a > 0$, it is possible to reduce w_a by $\epsilon > 0$ for any small ϵ. This will raise x, for x is determined by $p'(x)(w_n - w_a - \epsilon) = c'(x)$. We can therefore write $x = x(\epsilon)$, where $x'(\epsilon) > 0$. Now raise both w_n and w_a by $t(\epsilon)$, so as to satisfy (6) with equality, noting that this does not change x since $w_n + t(\epsilon) - (w_a + t(\epsilon) - \epsilon) = w_n - w_a - \epsilon$. Unit costs of the firm as a function of ϵ are $(1 - p(x))(w_n + t(\epsilon)) + p(x)(w_a + t(\epsilon) - \epsilon) + p(x)d = k + c(x(\epsilon)) + p(x(\epsilon))d$. Differentiating with respect to ϵ gives $x'(\epsilon)[c'(x(\epsilon)) + p'(x(\epsilon))d]$. But at $\epsilon = 0$, $c'(x) + p'(x)d < 0$, since $x < x*(d)$ is the maintained hypothesis. Thus, unit costs can be lowered by raising ϵ from 0, a contradiction to the optimality of $w_a > 0$.

(A5) $-p'(x)w_n = c'(x)$.

Now (A4) and (A5) are satisfied if and only if $w_n \geq d_1$,[17] so that a firm's problem becomes

(A6) min $(1 - p(x))w_n + p(x)d$

over $w_n \geq d_1$, where $x = x(w_n)$ is implicitly determined by (A5). The derivative of (A6) with respect to w_n is

(A7) $p'(x)x'(w_n)(d - w_n) + (1 - p(x))$.

It is clear that if $d > d_1$ is sufficiently close to d_1, the first term in (A7) will be dominated by $(1 - p(x))$ for w_n in $[d_1,d]$, implying that (A7) will be positive for all $w_n \geq d_1$ (for the first term in (A7) is positive for $w_n \geq d$); thus, the firm's choice of w_n for such a d is the corner solution, $w_n = d_1$. It is clear as well that if d is sufficiently large, (A7) will be negative when evaluated at $w_n = d_1$, implying that the firm's choice of w_n will exceed d_1 and that supernormal wages are paid. Moreover, if for some d, the firm chooses $w_n > d_1$, that must also be true for any higher d.[18] Hence, the claimed $d_2 > d_1$ exists. Finally, when the firm chooses $w_n > d_1$, since (A7) must be zero, we know that $w_n < d$, which implies that $x < x*(d)$. We also note for future

---

[17]This follows because an employee's expected utility is increasing in w_n and because his expected utility equals k if $w_n = d_1$.

[18]Let the firm's choice of w_n be w_n' at d, so that in particular its unit costs are lower than if w_n is d_1. That is,
(*) $(1 - p(x(w_n')))w_n' + p(x(w_n'))d < (1 - p(x(d_1)))d_1 + p(x(d_1))d$.
This implies that for $d^\circ > d$,
(**) $(1 - p(x(w_n')))w_n' + p(x(w_n'))d^\circ < (1 - p(x(d_1)))d_1 + p(x(d_1))d^\circ$
because $p(x(w_n'))(d^\circ - d) < p(x(d_1))(d^\circ - d)$, as $x(d_1) < x(w_n')$. Now (**) implies that w_n' in particular is superior to d_1 when damages are d°, so that the firm's choice of w_n must exceed d_1.

reference that for $d \geq d_2$, $w_n'(d) > 0$,[19] so that $x'(d) > 0$, where $x(d) = x(w_n(d))$.

Proof of Proposition 2. Part (a): Suppose that $h \leq d_1$ and that $d = h$. Then by Proposition 1(a), $x(h) = x^*(h)$. Also, since no supernormal wage is paid, (6) and (8) imply that $\pi = k + c(x^*(h)) + p(x^*(h))h$, so that (4) and (9) imply that $q = q^*$. Hence, the outcome is socially optimal, so that $d = h$ must be socially optimal.

Part (b): Suppose that h is in $(d_1, d_2]$. Then, by Proposition 1(b), if $d = h$, no supernormal wage is paid. Hence, (6) and (8) imply that $\pi = k + c(x(h)) + p(x(h))h$. Thus, the first term in (11) is zero, so that $W'(h) = -q(h)x'(h)[c'(x(h)) + p'(x(h))h]$. But $x'(h) = 0$, so that $W'(h) = 0$. For $d < h$ but still in $(d_1, d_2]$, $\pi = k + c(x(d)) + p(x(d))d < k + c(x(d)) + p(x(d))h$, so that the first term in (11) is positive, and since the second is zero, $W'(d) > 0$. Hence, such d cannot be optimal (and a similar argument shows that $d \leq d_1$ cannot be optimal). For $d \geq h$ but still in $(d_1, d_2]$, the price is $\pi = k + c(x(d)) + p(x(d))d > k + c(x(d)) + p(x(d))h$, so that the first term in (11) is negative, and since the second is zero, $W'(d) < 0$, so that such d cannot be optimal. Hence, $d = h$ is a local optimum. However, $d = h$ may not be a global optimum, for $W'(d) > 0$ is possible for $d > d_2$. Specifically, for such d, the first term of (11) is still negative, but the second term is positive in a neighborhood of d_2

---

[19]If we differentiate implicitly the first order condition that (A7) equals zero, we find that the sign of $w_n'(d)$ is that of $-p'(x)x'(w_n)$, which is positive.

19

(for $x'(d) > 0$ for $d > d_2$ and $[c'(x(d)) + p'(x(d))h]$ is negative
in a neighborhood of d_2 because $x(d)$ will be close to $x*(d_1)$ which
is less than $x*(h))$. In this case where the optimal d exceeds
d_2, $x(d)$ must be less than $x*(h)$: otherwise, the second term in
(11) is less than or equal to zero, and since the first term is
negative, $W'(d)$ would be negative. Because supernormal wages are
paid and $d > h$, the quantity produced is suboptimal.

Part (c): Suppose that $h > d_2$ and $d = h$. Then the first
term of (11) will be negative because $\pi > k + c(x(h)) + p(x(h))h$.
Also, the second term of (11) will be positive because, by
Proposition 1(c), $x(h) < x*(h)$. Thus $W'(d)$ can be of either sign
and the optimal d can be above or below h. The optimal d must
exceed d_2, however, for $W'(d)$ is positive for $d < d_2$, and is also
positive in a neighborhood above d_2. Additionally, the level of
care must be less than $x*(h)$: otherwise, the second term in (11)
is less than or equal to zero, and since d must exceed h (this
follows because $x*(d) > x(d)$ and $x(d) \geq x*(h))$, the first term in
(11) must be negative, so that $W'(d)$ must be negative, a
contradiction. Because $x(d) < x*(h)$ at the optimal d, the second
term in (11) must be positive, implying that the first term must
be negative, meaning that the quantity produced is suboptimal.