working paper
department
of economics

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

WORKING PAPER
DEPARTMENT
OF ECONOMICS

INTEGRATING PUNISHMENT AND EFFICIENCY CONCERNS IN
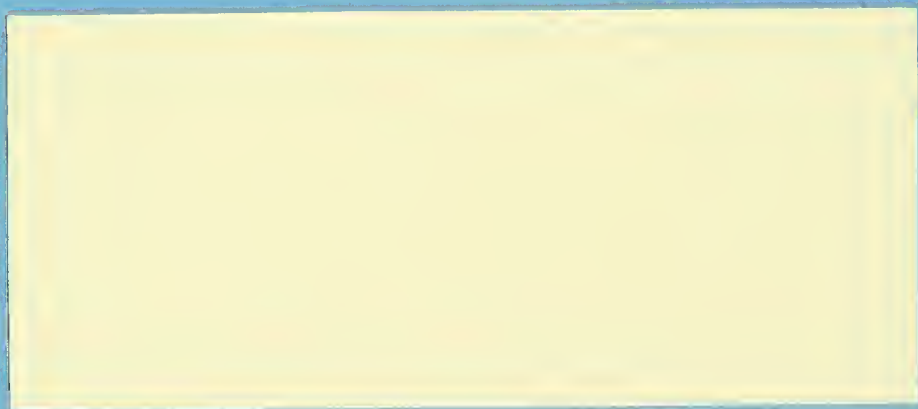*PUNITIVE DAMAGES FOR RECKLESS DISREGARD*
OF RISKS TO OTHERS

Peter A. Diamond

No. 97-19   October , 1997

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

50 MEMORIAL DRIVE
CAMBRIDGE, MASS. 02139

October, 1997

Integrating Punishment and Efficiency Concerns in Punitive Damages for Reckless Disregard of Risks to Others

Peter Diamond

Abstract

Punitive damages deter and punish. Using a social welfare function that incorporates both economic efficiency and a desire for retribution, this paper explores the effects of punitive damages in situations of reckless disregard that might be viewed as outrageous. If the defendant was making a rational decision that reflected all of social costs, any level of punitive damages lowers efficiency. If there are inadequacies of compensatory damages, the costs borne by the defendant are less than the full social costs. Then, punitive damages may be able to improve economic efficiency as well as providing retribution. Also considered is deterrence of nonrational reckless disregard (e.g., drunk driving) along with retribution.

Optimal punitive damages to balance both concerns is compared with the level which would 'be best considering only retribution, and the level which would be best considering only economic efficiency. In some situations, consideration of both deterrence and retribution argues for an intermediate level of punitive damages; in others it can call for higher or lower damages than either of the levels considering just a single issue.

Integrating Punishment and Efficiency Concerns in Punitive Damages for Reckless Disregard of Risks to Others

Peter Diamond[1]

Punitive damages are only supposed to be assessed in response to outrageous behavior. Mere occurrence of an accident, or occurrence of an accident after negligence is not supposed to trigger punitive damages. Both judicial and academic justifications for the use of punitive damages refer to deterrence and punishment as the two bases for their determination.[2][3] The social interest in deterrence is linked to a concern for economic efficiency, based on a presumption (which may or may not be

---

2 As an example of instructions to the jury, the following was given in Re the Exxon Valdez.
"Plaintiff has the burden of proving that punitive damages should be awarded by a preponderance of the evidence. You may award punitive damages only if you find that defendant's conduct
    (1) was malicious; or
    (2) manifested reckless of callous disregard for the rights of others.
Conduct is malicious if it is accompanied by ill will, or spite, or if it is for the purpose of injuring another.
In order for conduct to be in reckless or callous disregard of the rights of others, four factors must be present. First, a defendant must be subjectively conscious of a particular grave danger or risk of harm, and the danger or risk must be a foreseeable and probable effect of the conduct. Second, the particular danger or risk of which the defendant was subjectively conscious must in fact have eventuated. Third, a defendant must have disregarded the risk in deciding how to act. Fourth, a defendant's conduct in ignoring the danger or risk must have involved a gross deviation from the level of care which an ordinary person would use, having due regard to all the circumstances.
Reckless conduct is not the same as negligence. Negligence is the failure to use such care as a reasonable, prudent, and careful person would use under similar circumstances. Reckless conduct differs from negligence in that it requires a conscious choice of action, either with knowledge of serious danger to others or with knowledge of facts which would disclose the danger to any reasonable person."
Phase I Jury Instruction No. 28, Clerk's Docket No. 5309.
3 See, e. g., Ellis, 1982.

right in particular settings) that the outrageous behavior is economically inefficient, is based on inadequate financial disincentives, despite the presence of both compensatory damages and (possibly) civil and criminal punishments.[4] The social interest in punishment comes from a view that a balancing of outrageous behavior with punishment makes the outcome more socially desirable. A desire to punish per se, to make the punishment fit the crime, a just desert theory of damages has punishment as an end, not as a means to deterrence. We will refer to this motivation as retribution. These two sides of punitive damages are not separable. The assessment of punitive damages to have more deterrence is a form of punishment; the assessment of punitive damages as retribution is a further deterrent. This paper explores an evaluation of the effects of punitive damages incorporating both efficiency and retribution concerns.

One can dispute the appropriateness of using punitive damages for punishment (retribution) purposes, rather than relying only on fines and criminal law; one can question the appropriateness of a punishment orientation without the procedural protections of criminal law; and one can question the setting of punishment by a jury process with nothing in the way of guidelines for suitable punishments for different examples of outrageous behavior. Nevertheless, with continued use of punitive damages to punish, it seems worthwhile to explore the implications of accepting a jury's desire for retribution along with its interest in deterrence in the analysis of punitive damages.

Using an analytically convenient categorization, outrageous behavior comes in two types - malicious intent and reckless disregard for the risk to others.[5] The analytical distinction between these is that malicious intent involves a desire by the defendant to benefit directly and illegitimately at the expense of the plaintiff; the benefit is not available without inflicting a cost on the plaintiff. Battery and fraud are examples. A gain obtained in this way is then not counted as a social gain in considering social welfare. Reckless disregard, on the other hand, can be attributed to legitimate activities that involve a risk to others. While subjecting others to risk is an everyday occurrence, sometimes it is viewed as outrageous. This paper considers situations of reckless disregard, but does not consider situations of malicious intent.[6]

---

4 On punitive damages and deterrence, see Diamond, 1997a, Polinshy and Shavell, 1997.
5 Outrageous behavior without a malicious intent is referred to with a variety of (not fully interchangeable) terms such as reckless or callous disregard or reckless indifference to the rights of others, gross negligence, and legal malice.
6 My earlier paper (1997a) argued that in settings of malicious intent, jury instructions should be different from those given in situations with reckless disregard. I expect that that conclusion would be extended to consideration of punishment as well as deterrence.

This paper explores several situations where reckless disregard, without malicious intent, might be viewed as outrageous. First, behavior involving risk might be considered outrageous even though the defendant was making a rational decision that reflected all of social costs. While such a judgment seems unfair and is conducive to inefficiency, it might reflect public attitudes toward risks that are simply not consistent with efficiency considerations.[7] In this case, catering to such a public desire for retribution creates a tension between retribution and efficiency; any level of punitive damages in response to such an attitude lowers efficiency.

In Diamond (1997a), several models of reckless disregard that might be subjected to punitive damages were considered. If there are inadequacies of compensatory damages, rational risk taking based on the costs borne by the defendant might reflect costs which are less than the full social costs. If some examples of such behavior are judged outrageous, then we can consider retribution along with the need for deterrence when defendants bear only part of accident costs.

Another model of behavior that might be judged reckless disregard is where the information used by the defendant in decision-making was judged to be inadequate even though there was a rational decision as to how much information to gather. As argued in Diamond (1997a) this behavior implies efficiency as long as decision-makers are rational in seeking information and have appropriate prior beliefs. For this situation, we again have the finding that any punitive damages for retribution will lower economic efficiency. Again, one could question the legitimacy of retribution in this setting.

Another category of reckless disregard considered in the earlier paper was that of nonrational reckless disregard. Some drunk driving is an example, where some risks are simply ignored, even though their existence is known. The combination of deterrence of such behavior along with a desire for retribution is also analyzed.

In all of the cases, the analysis focuses on choosing punitive damages to balance a concern for economic efficiency with a desire for retribution after behavior seen as outrageous. To analyze this issue, the paper starts with two simpler concepts of levels of punitive damages. When considering only retribution there is some level of punitive damages, referred to as the ideal retributive level which would be the best level considering only issues of retribution. Similarly, when considering only deterrence, there is some level of punitive damages, referred to as the ideal

---

7 For example, Breyer (1993) has argued that inconsistencies in public attitudes toward risk have contributed to inefficient regulation of risks.

deterrence level which would be the best level considering only issues of economic efficiency. These two levels might coincide, or they might be different, with either one larger than the other. The paper considers setting punitive damages optimally to reflect both a desire for retribution and a concern for economic efficiency in terms of these two ideal levels, each of which reflects only one of these concerns. In some situations, consideration of both deterrence and retribution argues for an intermediate level of punitive damages; in others it can call for higher or lower damages than either of the levels considering one issue alone.

While one might argue that wealth is relevant for the level of retribution, the use of wealth in determining punitive damages creates inefficiency in the allocation of resources. The balancing of these concerns is also considered.

This paper begins by reviewing, in Section I, the relationship between compensatory damages and economic efficiency. Section II considers the formulation of a social welfare function that incorporates both economic efficiency and a desire for retribution. Section III considers the optimal balancing of deterrence and retribution in a situation where compensatory damages alone provide the correct level of deterrence, with mathematical analysis in Section IV, which can be skipped without loss of continuity. In Appendices A and B, the analysis is extended to consider the assumptions that the ideal retribution varies with the level of precaution and with the level of wealth. The nature of outrageous behavior warranting punishment is discussed in Section V. The analysis is then extended to situations where some accidents do not result in liability (Section VI, derivation in Section VII) and where some defendants are not rational in their choice of precaution (Section VIII, derivation in Section IX). An alternative interpretation of the model is in Section X. There are some concluding remarks.

I. Compensatory damages and efficiency

The widely recognized law-and-economics argument that compensatory damages will induce efficiency (in the absence of a need for incentives for victims), points out that if compensatory damages equal the monetary value of the harm to others, then a rational decisionmaker will weigh the value of the harm to others along with the net gain of any activity in deciding whether to engage in the activity.[8] For analytical convenience, this argument is examined here in a situation of strict liability, not negligence, since any defendant at serious risk for being found liable for punitive damages because of reckless disregard is very likely to be held negligent and so liable for compensatory damages if there is a negligence standard.

_____
8 See, e. g., Cooter and Ulen, 1997, 272-6.

For convenience of presentation, let us introduce some formal notation. While we could do the analysis in terms of actions that affect the probability of an accident, it is more convenient to consider the probability of an accident directly as the control variable of the defendant (within limits). The defendant can lower the probability of an accident by expending resources and changing behavior in a wide variety of ways. There are decreasing gains from pursuing such precautions, and the defendant will expend resources as long as the cost is less than the expected benefits from a lower accident probability.

We let p equal the probability of <u>avoiding</u> an accident, where p defines the level of precaution being taken by the defendant.[9] Let u[p] be the utility of the activity, net of costs of the activity, including the cost of precaution, but gross of any legal liability of the defendant.[10] Recognizing both the costs to the defendant of any accident and the cost of avoiding accidents, we assume that u[p] is first increasing, then decreasing in p. We also assume that it is impossible to have a zero probability of an accident while engaging in this activity, and that the cost of avoiding accidents rises without limit as the probability of avoiding all accidents increases toward its upper limit. The specific mathematical assumptions about u are detailed in the footnote, along with an example.[11] In Figure 1, we show an example of the assumed pattern of utility relative to precaution.

We denote by A the cost to the plaintiff in the event of an accident, including non-economic costs. Thus A is the amount of compensation required to restore the plaintiff to the position held

---

9 It is common in the law and economics literature to consider an explicit care or precaution variable and then to relate both the cost of care and the probability of an accident to this explicit care variable. Since both the cost of care and probability of an accident functions are monotonic, we can simplify the notation by using probabilities as control variables, since there is a direct functional link between the cost of care and the probability of an accident.

10 Throughout the paper, arguments of functions will be denoted by [], as in u[p]. Parentheses will be denoted by () and {}.

11 We assume that u[p] is strictly concave in p, $u''[p]<0$. We also assume that $u'''[p]<0$. For example, there might be a level of utility from the activity if there is no accident, an expected cost of accidents that is proportional to the probability of an accident, and a cost of avoiding accidents that is unbounded as the probability of an accident goes to zero. Then, u[p] might have the form: $u[p] = k_0 - k_1(1-p) - k_2/(1-p)$ for some positive constants k_i. Thus $k_0-k_1-k_2$ is utility of the activity (ignoring liabilty) if an accident is certain (p=0). With this utility function, we have the derivatives: $u'[p] = k_1 - k_2(1-p)^{-2}$; $u''[p] = -2k_2(1-p)^{-3}$; $u'''[p] = -6k_2(1-p)^{-4}$.
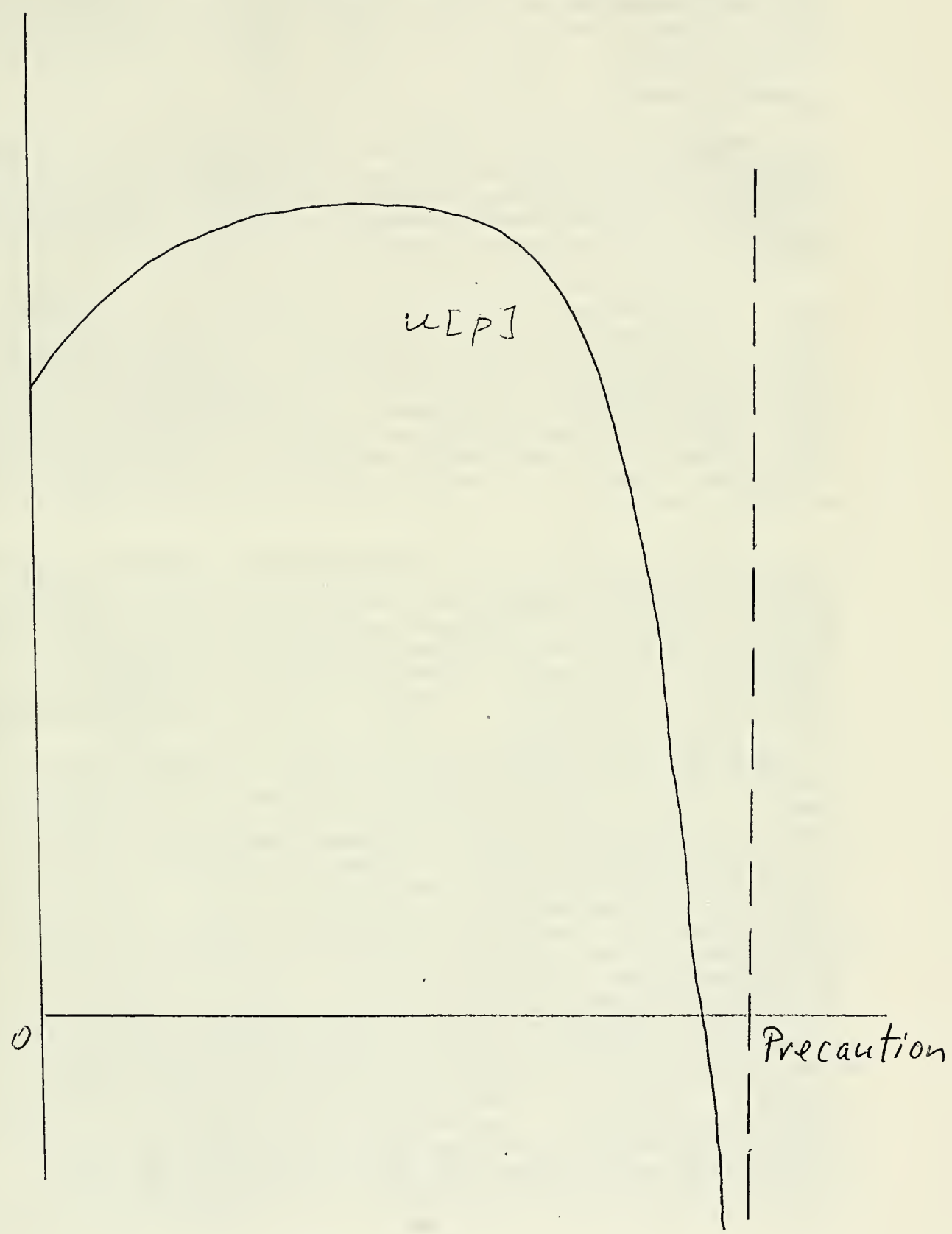
Figure 1

before the accident.[12] If the social evaluation of the choice of the
level of precaution is utilitarian, denoted in monetary terms, the
social value, W, of equilibrium with a chosen level of precaution
of p is

(1) $W[p] = u[p] - (1-p)A$.

This social evaluation recognizes the utility of the defendant, $u[p]$,
and the expected accident costs of the plaintiff, $(1-p)A$.[13][14] The
payment of damages by the defendant to the plaintiff is viewed as
a transfer without direct social significance. For convenience, the
legal costs are taken to be zero.[15]

        If the defendant is liable for the plaintiff's accident costs,
then the individual utility of the defendant, denoted $U[p]$, equals
$u[p]-(1-p)A$, which coincides with $W[p]$. In this setting, a rational
decisionmaker would indeed select the level of precaution that is
socially efficient. This analysis is structured for the situation of
an accident involving strangers (in the contractual sense), as with
environmental damage. Additional complications arise when there
is a contractual relationship and so one must examine the impact
on prices. Thus, this analysis does not directly apply to issues of
products liability.

        Implicit in this formulation of social evaluation is an absence
of externalities other than the possibility of an accident. If the
activity of the defendant has social values which are different
from the gain to the defendant less the accident costs, then this
further deviation of individual and social values must be
considered. For example, innovations in product and in
technology are generally viewed as having externalities since they
affect the opportunities·of others in ways that are not captured by

---

12 More generally, we could also allow accident costs, A, to vary
with the level of precaution; but this would not alter the shape of
the conclusions, given suitable assumptions on this interaction to
preserve the plausible description of the accident environment.
13 This approach ignores all issues of income distribution by
adding up individual utilities in dollar terms. For a defense of
this approach, see Shavell, 1981.
14 Note that this expression involves no concern about the
incentives for the plaintiff to avoid accidents or to lower the costs
of accidents that do occur.   '
15 This assumption makes the analysis easier to follow. Moreover,
the effects of legal costs have been studied in the literature. The
costs of the defendant are an additional deterrent to accident
generating behavior, although one that is a social cost of the
accident. The legal costs of the plaintiff are an additional social
cost of the accident. The effect of such costs on the analysis
would pay attention particularly to the frequency of litigation
with and without punitive damages and the advantages and
disadvantages of different amounts of litigation.

market transactions.[16] In addition, the undertaking of large
projects can involve consumer surplus, which again is not captured
by market transactions. Below, we will be concerned with
overdeterrence. To the extent that the activities deterred involve
positive externalities, then the social concern for overdeterrence
becomes larger. "Reputation costs" raise a similar issue. When
reactions to an accident do not accurately reflect true risks, then
the defendant's costs from an accident can be larger than the
social costs. Such reputation costs are thought to be particularly
relevant with consumer product risks. Conversely, when deterring
accident-generating behavior also deters other negative
externalities then there is less concern about overdeterrence. In
different particular situations one or the other of these concerns
may be larger.

     As has been noted in the literature, this argument for
efficiency with compensatory damages breaks down if the
probability of paying damages is less than the probability of an
accident, 1-p, or if the level of compensatory damages is less than
the costs inflicted on the plaintiff, A. Below, we return to these
issues. We also add to the analysis civil and criminal penalties
that also serve to deter and to punish.

II.   Social evaluation of punishment

     Court decisions involving punitive damages refer to a desire
to punish per se, to use punitive damages to inflict a cost on the
defendant, to have the punishment fit the crime. In this section
we present an approach to incorporating the desire for punishment
along with concern for economic efficiency.

     In recognizing both deterrence and a desire to punish, the
motivation for punitive damages parallels that of the criminal
law. For example, consider this statement by John Rawls (1955):

> "For our purposes we may say that there are two
> justifications of punishment. What we may call the
> retributive view is that punishment is justified on the
> grounds that the wrongdoing merits punishment. It is
> morally fitting that a person who does wrong should suffer
> in proportion to his wrongdoing. That a criminal should be
> punished follows from his guilt, and the severity of the
> appropriate punishment depends on the depravity of the
> act. The state of affairs where a wrongdoer suffers
> punishment is morally better than the state of affairs
> where he does not; and it is better irrespective of any of
> the consequences of punishing him.
>
>   What we may call the utilitarian view holds that on the
> principle that bygones are bygones and that only future

---

16 On the impact of liability on innovation, see Huber and Litan,
1991.

consequences are material to present decisions, punishment
is justifiable only by reference to the probable
consequences of maintaining it as one of the devices of the
social order. Wrongs committed in the past are, as such, not
relevant considerations for deciding what to do. If
punishment can be shown to promote effectively the
interest of society it is justifiable, otherwise it is not."
(Pages 4-5.)

That there is a tension between these two justifications for
punishment is well recognized in the literature. For example,
Chapter 1 in Packer (1968) is entitled "The Dilemma of
Punishment;" Goldman writes of the "Paradox of Punishment"
(1979). Some of the philosophical debate is about which of these
two justifications is appropriate; some, e. g., Packer, argue for the
relevance of both justifications in deciding when to punish. In
contrast with the qualitative relevance of both justifications,
Goldman identifies the dilemma in quantitative terms: that the
level of punishment for criminal activity that he considers just
(and so justified) is sufficiently low as a deterrent to crime as to
be unpalatable; conversely, that the level he thinks needed for
deterrence is too large to be just. The purpose of this section is to
integrate a concern for retribution into a formulation of social
welfare which also has a concern for efficiency in a form that
leads to a quantitative outcome.[17] In other words, if one considers
only deterrence or only retribution, it is likely that the levels of
punitive damages seen as appropriate for the two single purposes
would be different from each other. When this is the case, the
question might be how to choose between the two bases for setting
punitive damages, or as analyzed here, how to compromise
between the two levels suggested by the two bases. While it would
appear that Goldman would not approve of grasping both horns of
this dilemma simultaneously and smoothly trading off the costs of
improper deterrence and unjust retribution, that is the approach
taken here. The focus here is not on the appropriateness of
making this shotgun marriage of concerns, but on exploring the
implications of having such a marriage. Nor do we explore the
appropriateness of accepting a desire to punish; the starting place
is to incorporate both retribution and deterrence concerns. As put
by Wertheimer (1975): "It seems reasonable to want a punishment

---

17 There is a need to consider punishment along with utilitarian
concerns to make sense of having punishment vary monotonically
with the seriousness of crime. Optimal tax theory, which shows
the importance for taxes of elasticities of demand, implies that
monotonicity will not necessarily follow from utilitarian concerns.
This disagrees with assertions to the contrary, such as that of
Rawls (1955). In footnote 14 (Page 12-13) Rawls considers the
proportionality of punishment to offense: "if utilitarian
consideration are followed penalties will be proportional to
offenses in this sense: the order of offenses according to
seriousness can be paired off with the order of penalties according
to severity."

to "fit the crime," and it seems reasonable to use that punishment that will maximize utility." (Page 420.) Or, as put by Posner (1975): "since no rational society can ignore the costs of its public policies, they are issues to which economics has great relevance. The demand for justice is not independent of its price." (Page 778.)

In defining social welfare, we assume that the payment of punitive damages is a transfer that has no direct welfare implications as a transfer, but does as part of retribution (as well as affecting deterrence). That is, we assume that if an accident has occurred which is judged to have been caused by outrageous behavior, then there is a social desire to inflict a punishment. We refer to the level of punishment desired for retribution as ideal retribution and denote it by P^r. For the purposes of this analysis, we do not analyze in detail the issues that enter into the determination of the ideal retribution level; we merely assume that the ideal retribution is given at some level for this class of defendants and does not vary with the precaution taken. In the Appendices, we consider having ideal retribution vary with the precaution of a typical defendant and with the wealth of the defendant.

We assume that inflicting a level of punishment different from the ideal retribution involves a social cost. To integrate this social cost with economic efficiency concerns, we need to answer several questions. In considering the punishment of defendants, there are payments of punitive damages and of compensatory damages and often also of civil and criminal fines – do we consider the sum or only some of these payments to compare with the ideal retribution? We take the answer to be that the sum of all damages and fines should be compared with P^r, although for the analytical purposes of this paper, the distinction is not important, since we do not vary the cost of accidents or the level of fines.[18] We assume that P^r is at least as large as the accident cost caused by the defendant, A. If civil and criminal fines have been set to fully capture the desire for retribution, then P^r will be equal to the sum of accident costs and civil and criminal fines. We recognize the possibility that the desire for retribution exceeds the level of accident costs plus fines, leaving room for punitive damages to improve the satisfaction of the desire for retribution. We do not examine why legislated fines might be below such a level in some situations, nor the appropriateness of leaving the decision to punish by more than the legislated amount to a jury.

We write the cost of incorrect punishment as a function of the difference between ideal retribution and the sum of compensatory

_____

18 This view is not new. Morris (1931) argued that the sum of compensatory and punitive damages should be viewed as punishment and applauded the fact that West Virginia juries are instructed to think in these terms (page 1188). He also called for "reciprocal adjustment of penalties of the civil and criminal courts." (page 1197.)

damages, civil and criminal fines, and punitive damages, with a
social cost from incorrect punishment whenever the sum of
payments does not equal the ideal retribution. In units
comparable to utility, how should deviations from ideal
retribution be evaluated? We assume the existence of a function
giving this evaluation, without much consideration of its
properties.[19] There can be different causes of a deviation of
actual punishment from ideal retribution - causes that include
jury error in determining liability for punitive damages, jury
error in setting a level of punitive damages, or a conscious social
decision to deviate from ideal retribution because of possible
economic inefficiency from levying ideal retribution. We do not
consider having different functions for different causes, although
one could extend the analysis in that way.

Denoting by C the sum of compensatory damages and civil
and criminal fines paid, and by P the level of punitive damages
paid, we write as $V[C+P-P^r]$, the social cost, measured in units of
individual utility, from having levied a punishment different
from the ideal retribution, with $V[0]=0$. We assume that V is
decreasing in P for C+P below P^r and increasing in P for C+P
above P^r, with the rates of decrease and increase monotonically
changing. That is we assume that V has a positive second
derivative, with the first derivative negative for underpunishment
and positive for overpunishment.[20]

We define the social welfare function as the utility of the
plaintiff plus the utility of the defendant less this social cost V.[21]
In this situation, where damage payments are viewed as a transfer,
we are concerned with the utility of the defendant from taking
precaution, the expected cost of accidents, and the social cost of
improper punishment. Thus we write social welfare as

(2) $W[P] = u[p] - (1-p)A - (1-p)V[C+P-P^r]$.

With this formulation, concern about accidents involves only
actual accident costs, provided that the decision-makers
responsible for the accidents are properly punished. If they are
not properly punished, then there is an added cost to any accident,
a cost from inappropriate punishment, whether that punishment is
too large or too small. With this structure, we are only concerned

---

[19] For discussion of this issue in the context of taxation, see
Diamond, 1997b. There it is analyzed how taxation should differ
from ideal taxation to reflect the deadweight burdens of taxation
as well as the administrative costs of trying to get taxes "just
right."
[20] Since V is nonnegative, we also have $V'[0]=0$. We assume
$V''[C+P-P^r]>0$ for all $C+P-P^r$.
[21] This formulation does not recognize any social significance to
the overcompensation of the plaintiff if the sum of punitive and
compensatory damages exceeds accident costs.

with behavior that actually results in an accident, not behavior that might have resulted in an accident (no harm, no foul).[22]

One missing piece in this discussion is the lack of a formal connection between causing accidents and deserving punishment. As written down, every accident is viewed as warranting punishment. Moreover, the level of ideal retribution is viewed as independent of the level of precaution taken to avoid the accident.[23] This is peculiar. If the law is working as modelled in Section I, then everyone is basing their precaution decisions on the full level of costs that is deemed to be appropriate. In this context, it is unclear why there is any desire to punish beyond having the defendant pay the cost of the accident - there is no outrageously low level of precaution because of paying attention to too few costs. Yet, some juries may want to punish based merely on the occurrence of serious accidents, and hindsight bias may distort the evaluation of whether punishment is warranted by too little precaution.[24] If the legal system chooses to go along with this desire to punish, then the analysis can be interpreted as the appropriate balancing between respecting that desire of juries and lowering economic efficiency.

Alternatively, the legal system might require the jury to identify a basis for considering outrageous the behavior that resulted in an accident. This would allow the court to distinguish between malice and reckless disregard, a distinction called for in my previous paper (1997a) when considering only the issue of deterrence. Identifying a basis for punishment as a first step toward setting punitive damages seems an improvement over the

---

22 Implicit in equation (2) is the idea that when punishment is at the ideal level for retribution and if that punishment induces economic efficiency, then there is no further concern with the occurrence of the behavior that was deemed to be in need of punishment. One might well argue for further deterrence to decrease the occurrence of outrageous behavior in response to a further dislike of such behavior, even when that behavior is ideally punished. Alternatively, one might consider economic efficiency and desired retribution as the only suitable bases for social evaluation. The formulation above is usable with either interpretation by interpreting the accident cost as either the cost that is suitable for economic efficiency or that cost plus any additional concern for the behavior that led to the occurrence of the accident. The text will be based on the interpretation that A is actual accident costs, without any additional outrage adjustment, consistent with linking deterrence to economic efficiency and punishment to retribution. The mathematics can be interpreted as fitting the alternative interpretation.
23 More generally, we could write the ideal retribution as a function of the probability of an accident. In the text, we use the simple case in equation (2), leaving the more general case for Appendix A.
24 On hindsight bias see Fischhoff, 1982.

current undifferentiated jury instructions, that do not distinguish between malice and reckless disregard and do not guide the jury through explicit consideration of deterrence and retribution separately. These elements function differently in different situations, and will be analyzed below in several different situations that might be considered reckless disregard. But first we consider the simplest model to bring out the workings of the model before turning to more complex situations.

III. Balancing deterrence and retribution

We could combine the generalized social welfare function approach in equation (2) with any of the models of individual behavior that warrant deterrence analyzed in Diamond (1997a). It is helpful to begin with a simpler situation, where the defendant pays all of the accident costs, denoted by A, as well as paying civil or criminal penalties, if any, which are equal to C-A. Despite paying all of the accident costs, it is assumed that the defendant's behavior is viewed as outrageous and subjected to punitive damages at the level denoted by P. That is, we assume that the utility of the defendant is:

(3) $U[p] = u[p] - (1-p)(C + P)$.

The defendant has utility gross of legal costs of u[p] and pays compensatory damages of A, fines of C-A, and punitive damages of P if there is an accident. We make the same assumptions about u as above (detailed in footnote 11). The only decision for the defendant is the level of precaution, p, which we can consider as a proxy for a variety of different decisions including the level of activity, since accidents depend on both the level of activity and the precaution taken per unit of activity.[25] We write the level of precaution chosen to optimize U[p] in equation (3) as a function of total payments in the event of an accident, p[C+P]. With our assumptions on u, the level of precaution increases with damages, but at a decreasing rate. In Figure 2 we show the level of precaution as a function of punitive damages for a given level of compensatory damages and fines.

When there is no underdetection and no underassessment of compensatory damages, if there were no desire to punish, there would be no reason for punitive damages, whether fines were zero or positive. That is, inflicting punitive damages in this situation lowers economic efficiency, since defendants react to the risk of punitive damages. That is, to maximize the sum of the utilities of the defendant and the plaintiff, u[p]-(1-p)A, we would set fines equal to zero and punitive damages equal to zero. In contrast, to

---

25 That is, any incentive to pay higher (total) costs, to take more care than the expected cost minimizing level, raises the cost of doing business and so leads to an upward shift in the supply curve of the firm and so to higher prices and lower quantities supplied.
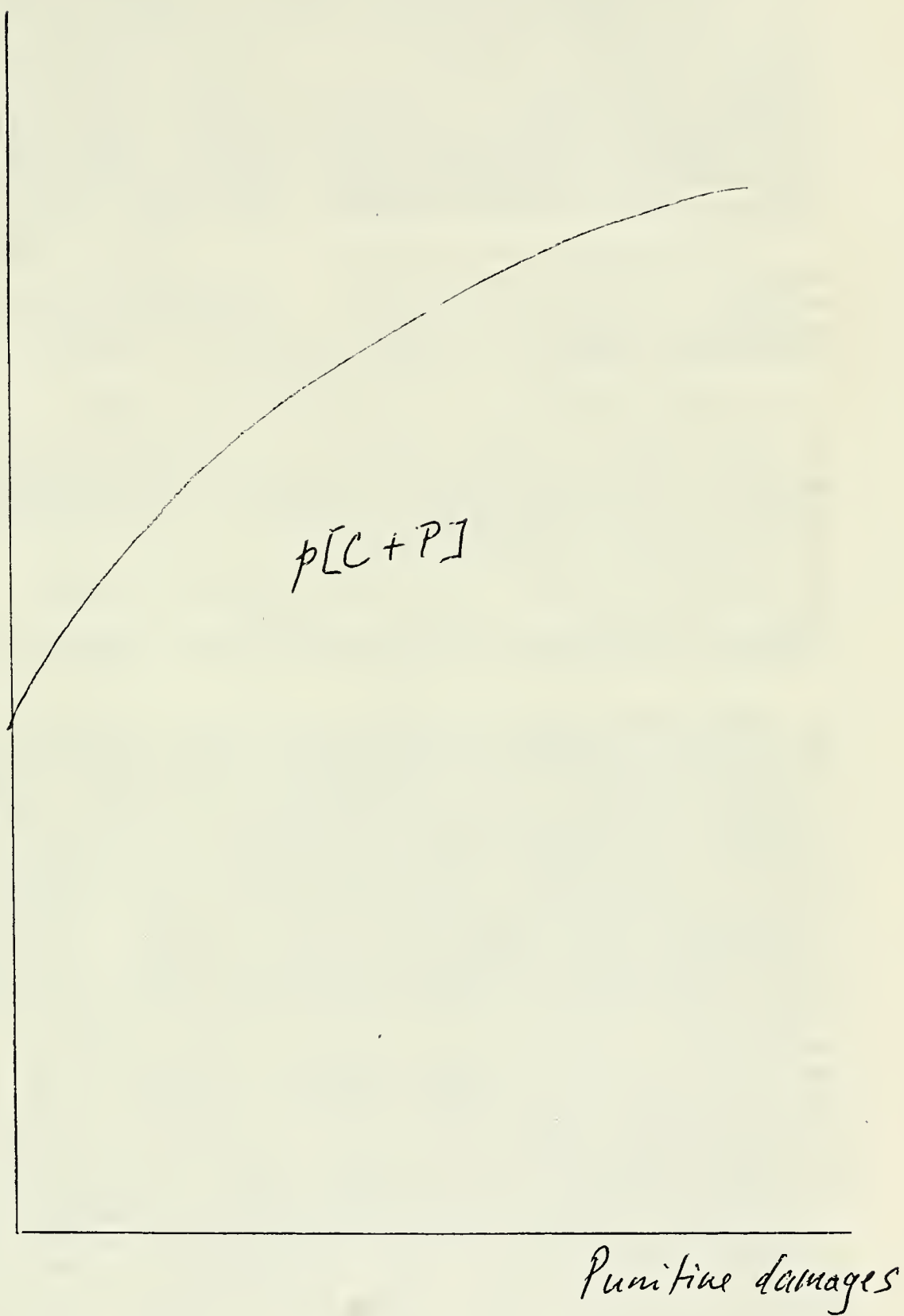
Precaution

$p[C + P]$

Punitive damages

Figure 2

minimize $V[C+P-P^r]$, we would set punitive damages equal to the ideal retribution level less compensatory damages and fines paid, P^r-C, if this difference is positive.[26] Assuming that ideal retribution exceeds fines plus accident costs, $P^r>C$, the setting of punitive damages to maximize (2), taking as given that the defendant sets precaution to maximize (3), results in a level of punitive damages that is less than the ideal retribution minus compensatory damages paid, minus fines paid, Pr-C. The proof of this point is in the next section. Thus, a jury that set punitive damages at the ideal retribution level would not maximize social welfare since it would overdeter, and so cause inefficiency.

To characterize the optimal level of punitive damages, we can calculate a first order condition for the maximization of social welfare, (2), taking as given that individuals set precaution to maximize utility, (3). Rewriting the social objective function to incorporate individual behavior, we have:

(4) $W[P] = u[p[C+P]] - (1-p[C+P])(A + V[C+P-P^r])$.

The derivation and formal analysis are in the next section, which can be skipped without loss of continuity. In the situation where punitive damages should be positive, the first order condition for optimal punitive damages is:

(5) $(1-p)V'[C+P-P^r] = \{V[C+P-P^r] - (P+C-A)\}p'[C+P]$.

This formulation brings out the two elements that are balanced in achieving an interior optimum.

When total damages plus fines are less than ideal retribution, an increase in punitive damages lowers the expected cost of inadequate punishment. The marginal gain from increased retribution equals the probability of an accident, $1-p$, times the marginal social disutility of inadequate punishment, V'. An increase in punitive damages increases precaution by the marginal amount p'. There are two consequences of the decrease in accident probability. One is that we have less frequent occurrence of inadequate retribution. The other is that we have more inefficiency from overdeterrence. The marginal efficiency cost of this overdeterrence is the extent to which punitive damages plus fines exceed the level just necessary to induce efficient deterrence. Since, in the situation being analyzed, there is no need for punitive damages or fines to induce efficient deterrence, the marginal cost of overdeterrence is the full amount of the punitive damages plus fines. That is, punitive damages and fines work like distortionary taxes.[27] The optimum occurs where the

_____

26 In this situation, if accident costs plus fines exceeds the ideal retribution level, $P^r<C$, then there should not be punitive damages.
27 Marginal individual return to precaution is $u'+C+P$. The economic side of social welfare has a marginal return to precaution that is $u'+A$. Thus, individuals take too much precaution relative to the standard for economic efficiency, with $P+C-A$ being equal to the amount that the private return to precaution exceeds the social return.

gain from having a more appropriate punishment when accidents
do occur is just balanced with these two effects from inducing
more precaution - a lowering of the economic efficiency of
precaution, and a lowering of the frequency of inadequately
punished defendants.

Analyzing the properties of optimal punitive damages, there
are two possibilities. One possibility is a corner condition where
punitive damages should be zero since fines are large enough. If
there is an optimum with positive punitive damages, then we have
the result that punitive damages are set so that they satisfy the
following inequalities:

(6) $V[C+P-P^r] < P+C-A < P^r-A.$

That is, at the optimum the social cost of inadequate punishment
after an accident, V, is less than the sum of punitive damages and
fines, which, in turn, is less than the excess of the ideal
retribution over compensatory damages. For example, consider a
case where accident costs (A) are 100 and there is a desire to
punish which doubles the size of accident costs, so that ideal
retribution (P^r) is 200. Then, punitive damages plus fines should
be set below 100, the level that would achieve ideal retribution,
(P^r-A). How much below this level would be optimal depends on
the shape of the cost of inadequate retribution. If punishing at
150 when the ideal is 200 generates a social cost of less than 50,
then we would satisfy the inequalities above. If the social cost of
too little retribution is larger than the sum of punitive damages
and fines, then the sum should be larger.

In the analysis above, the ideal retribution was based on the
occurrence of an accident. No adjustment was made either for the
typical level of precaution of all defendants, or the actual level of
precaution of this particular defendant. We pursue these two
ideas in Appendix A. One can have different views as to whether
ideal retribution should vary with the wealth of the defendant.
Without addressing that question, Appendix B explores how
optimal punishment should vary with wealth if one assumed that
ideal retribution is proportional to wealth. This situation is
chosen for analysis since it fits with arguments commonly made
by plaintiffs that call for punitive damages proportional to the
wealth of defendants.

IV. **Derivation of optimal punitive damages**

We first show that optimal punitive damages are less than P^r.
Note that precaution, p, is an increasing function of total damages
plus fines assessed, C+P (everything else held constant) which is
given by the first order condition for the maximization of (3).
given above:

(3) $U[p] = u[p] - (1-p)(C + P).$

Thus p[C+P] satisfies:

(7)   u'[p] + C + P = 0.

For later use, we differentiate this expression twice with respect to p.

(8)   u"p' + 1 = 0;
      u"'p' + u"p" = 0.

We have assumed that in the relevant range, u"'<0, and so p"<0.

   To set P to maximize social welfare, we substitute p[C+P], as defined implicitly in (7), into (4).

(9)   W[P] = u[p[C+P]] - (1-p[C+P])(A + V[C+P-P^r]).

Calculating the derivative of W, we have:

(10)   W'[P] = (u'[p[C+P]]+A+V[C+P-P^r])p'[C+P] -(1-p[C+P])V'[C+P-P^r]
            = (-C-P+A + V[C+P-P^r])p'[C+P] - (1-p[C+P])V'[C+P-P^r],

where we have used (7) to simplify (10). In the situation where punitive damages should be positive, the first order condition for optimal punitive damages is as given above:

(5)   (1-p)V'[C+P-P^r] = {V[C+P-P^r] - (P+C-A)}p'[C+P].

   We note that if there are no fines, C=A, and a desire for retribution beyond accident costs, P^r>A, then W[P] is increasing in P at P=0, since

(11)   W'[0] = (V[A-P^r])p'[A] - (1-p[A])V'[A-P^r] > 0.

With V positive, V' negative, and p' positive, we conclude that when C=A, W'[0]>0. When there are fines, C>A, then W'[0] may be either positive or negative, even if A<P^r.

Similarly, when ideal retribution exceeds accident costs plus fines, setting punitive damages equal to the difference, P=P^r-C, we have:

(12)   W'[P^r-C] = (A-P^r)p'[P^r] < 0,   '

where we have used the conditions V[0]=V'[0]=0. Since u[p]-(1-p)A is decreasing in P, we have lower values of W above P^r-C than at Pr-C. Thus the optimum, if not at the corner at zero, lies between 0 and Pr-C. Using the first order condition, (10), the condition p'>0, and the result that at the optimum V'<0, at the optimum we have, P+C-A>V[C+P-P^r].

V. Outrageous behavior

Punitive damages are supposed to be assessed only in response to outrageous behavior. Malicious intent is a natural source for outrage. This paper considers only settings of reckless disregard that are considered outrageous. Sometimes, reckless disregard is attributed to legitimate activities that involve a risk to others. While subjecting others to risk is an everyday occurrence, sometimes it is viewed as outrageous. Using models of defendant behavior in Diamond (1997a), we explore several situations that might be viewed as outrageous. In the situation analyzed in Section III above, behavior involving risk might be considered outrageous even though the defendant was making a rational decision that reflected all of social costs. While such a judgment seems unfair and is conducive to inefficiency, it might reflect public attitudes toward risk that are simply not consistent with efficiency considerations. For example, Breyer (1993) has argued that inconsistencies in public attitudes toward health and safety risks have contributed to inefficient regulation of risks. In this situation we have a tension between retribution and efficiency that is reflected in the modelling in Section III above. That is, any level of punitive damages in response to such an attitude lowers efficiency.

If there are inadequacies of compensatory damages, rational risk taking might be based on the costs borne by the defendant, costs which are less than the full social costs. If such behavior is judged outrageous, then we can combine retribution with a modification of the model above reflecting only partial bearing of accident costs. Such a model is explored in the next section.

Another model of behavior that might be judged reckless disregard is where the information used by the defendant in decision-making was judged to be inadequate even though there was a rational decision as to how much information to gather. As argued in Diamond (1997a) this behavior implies efficiency as long as decision-makers are rational in seeking information and have appropriate prior beliefs. For this situation, the analysis in Section III above again applies.

A different category of reckless disregard considered in the earlier paper was that of nonrational reckless disregard. Some drunk driving is an example of such behavior, where some risks are simply ignored, even though their existence is known. The combination of deterrence of such behavior along with a desire for retribution is analyzed below, starting in Section VIII.

VI. Underassessment of compensatory damages[28]

---

28 Some analyses of criminal law use a similar model to the one employed here. Thus a concern about appropriate punishment will move optimal punishment away from the solution where sanctions are maximized and enforcement is minimized.

As discussed in Section I, when defendants pay compensatory damages that equal accident costs after all accidents, there is economic efficiency. Conversely, when compensatory damages are assessed after some, but not all, accidents, or compensatory damages are less than accident costs, deterrence will be too small for economic efficiency unless fines are set just right. Assume that compensatory damages are assessed after a fraction, a, of accidents, with compensatory damages equal to the fraction b of accident costs. Assume that in all of the accidents that do result in compensatory damages, fines, F, and punitive damages, P, are assessed as well. We denote by C the sum of compensatory damages and fines paid, C= bA+F.

The behavior of defendants is described by the maximization of individual utility which is given by a modification of equation (3) to reflect the probability, a, of assessment of damages and the compensatory damages plus fines, C, which might differ from the accident costs:

(13) $U[p] = u[p] - a(1-p)(C + P)$.

Given our assumption of risk neutrality on the part of defendants, the precaution decision is being made in response to expected costs $a(C+P)$, and can be written as $p[a(C+P)]$.

We must modify the social welfare function (as given in equation (4)) to reflect the assumption that after some accidents no defendant is identified. There is no punishment after the fraction (1-a) of accidents, accidents that warrant retribution at the level P^r including the payment of compensatory damages and fines. Thus we write the concern for retribution in the social welfare function as a cost of $a(1-p)V[C+P-P^r] + (1-a)(1-p)V[-P^r]$. With a desire for retribution, $P^r>0$, accidents without liability lower social welfare by more than the same accidents would if the defendant paid damages, unless there is punishment well in excess of desired retribution. Note that the assumption that some accident generation is escaping punishment is not being viewed as a reason to lower the ideal retribution on those who are assessed damages.

The social welfare function, previously given in equation (2), now becomes:

(14) $W[P] = u[p] - (1-p)A - \{a(1-p)V[C+P-P^r] + (1-a)(1-p)V[-P^r]\}$.

When some accidents are caused by defendants who escape liability, the cost of inappropriate punishment changes from $\{V[C+P-P^r]\}$ to $\{aV[C+P-P^r] + (1-a)V[-P^r]\}$. That is, in addition to being concerned that defendants who are held liable are suitably punished, there is the additional concern that some people causing accidents escape all punishment, i.e., are not subjected to any liability. This additional concern increases the social gain from

deterring accidents. This will generate a payoff to deterrence beyond the level that induces economic efficiency even when that level would be viewed as a suitable level of retribution.

Note that this formulation includes both situations where there is a desire for retribution beyond the needs of deterrence and situations where, because of the poor workings of compensatory damages, ideal deterrence needs would call for a level of punitive damages beyond what would be ideal for retribution.

As above, we want to maximize social welfare, given in equation (14), subject to the constraint that defendants maximize utility, as given in equation (13). The first order condition for optimal punitive damages differs in several ways from that given above in equation (5). As derived in the next section, the first order condition is now written as:

$$(15) \quad (1-p)V'[C+P-P^r] = \{A(1-ab)-a(F+P)+aV[C+P-P^r]+(1-a)V[-P^r]\}p'.$$

This expression matches that in equation (5) above when a and b are both equal to one. Increasing punitive damages changes the level of inappropriate punishment (which appears on the left-hand side of equation (15)) and increases the level of precaution (which appears on the right-hand side). Both of these effects work only through the fraction, a, of accidents that result in liability, a multiplicative factor that cancels in equation (15). The marginal social value of increased punishment for those held liable is $(1-p)V'[C+P-P^r]$, which equals the probability of an accident times the marginal change in the social cost of inappropriate punishment. The marginal social value of increased precaution depends on the efficiency effect, which is the excess of the private over the social return to precaution, $\{a(F+P)-A(1-ab)\}$, [29] and on the avoidance of accidents that lead to inappropriate punishment, from C+P not equalling P^r for those who are held liable and from the lack of liability for the rest.

We have modeled two ways in which compensatory damages can be inadequate. Compensatory damages might be less than accident costs. However, fines are assumed to be assessed on all defendants who pays compensatory damages. Thus fines can perfectly offset the economic effects of compensatory damages that are too low.[30] Moreover, fines and compensatory damages enter the evaluation of the level of retribution in the same way. However, when some people who cause accidents escape all

_____

29 At the margin, individuals evaluate additional precaution by u'+a(C+P). Economic efficiency requires a marginal evaluation of u'+A. Thus the individual incentive exceeds the incentive for efficiency by a(F+P)-(1-ab)A. This can be positive or negative depending on the size of fines.
30 If a=1, then having F=(1-b)A removes the difference between private and social economic incentives for precaution.

liability, the situation is more complicated. While increasing fines can offset the inadequate ex ante incentive for precaution, it does not enter the evaluation of retribution in a perfectly offsetting way. In particular, because of inadequate retribution for those escaping liability, there remains a social gain from additional deterrence even if fines are set at a level that produces economic efficiency and happens to just match the desired level of retribution of those who pay the fines. To see this point, we examine the ideal deterrent and retributive levels of punishment and note that optimal punishment is larger than the smaller of these and larger than both of them when they are equal.

The ideal deterrence level of punitive damages is the level that would be set if there were no concern with punishment per se. The ideal deterrent to induce efficient precaution, including compensatory and punitive damages and fines would equate expected payments in damages and fines with expected accident costs to others. Thus, the ideal deterrent, denoted by P^d, satisfies:

(16) $P^d = A/a.$

Deterrence includes compensatory and punitive damages and fines paid. Thus the ideal level of punitive damages for deterrence is equal to P^d-C.[31]

A central question is whether the ideal retribution, P^r, is larger or smaller than the ideal deterrent, P^d. The probability of liability, a, is central to this comparison. As this probability gets small, P^d rises without limit, eventually exceeding the ideal retribution. As this probability goes to one, P^d goes to the level of accident costs caused, which we have assumed is no larger than the ideal retribution, on the grounds that people should at least pay for the accident costs caused. When there are no fines and compensatory damages are adequate, ideal punitive damages for deterrence tend to zero, becoming less than ideal punitive damages for retribution.

As shown in the next section, the optimal level of punitive damages is larger than the minimum of these two ideal punishments, less the other costs borne, P^d-C and P^r-C, and may even be larger than both of them. Indeed, if P^d and P^r happen to coincide, then the optimal level of punitive damages is larger than both. The pressure for more deterrence comes from the social disutility of the lack of punishment after accidents that result in no payment of damages. Larger punitive damages deter such accidents as long as there is a risk of being held liable for punitive damages. Thus, having potential defendants escape liability adds to the importance of deterrence.

We also might have the situation that the fines, F, are sufficiently large that the desire to punish the defendants who are

---

31 The ideal level of punitive damages for deterrence is derived and discussed in Diamond (1997a).

identified is satisfied by the fines, F, together with the compensatory damages paid, bA.

We can conclude that optimal punitive damages might be zero if C is large enough. Alternatively, when optimal punitive damages are positive, they are larger than the smaller of the ideal levels for just economic efficiency or retribution, and may be larger than both of them if they are not too far apart and it is particularly important to avoid the social cost from a lack of retribution for those who escape liability.

The structure of results here is strongly influenced by the assumption that all potential defendants have the same probability of being held liable. This assumption permits greater punishment when liable to substitute for a smaller probability of being held liable. With variation in the population in the probability of being held liable, the results would change. This issue, which is also important for the theory of criminal penalties, would be interesting to explore.

The analytical approach taken here could be a basis for jury instructions after a finding of liability for punitive damages in cases having this type of structure. Rather than simply asking juries to select a level of punitive damages to reflect both deterrence and punishment, the instructions could request that the jury consider each element separately. The instruction about deterrence would explain that the goal was efficient deterrence, not the deterrence of all accidents. The jury would be told to subtract compensatory damages and fines when moving from concerns about deterrence and punishment to levels of punitive damages. The instruction would tell the jury not to simply add the levels suitable for deterrence and retribution, but to average them, adjusted upward for any additional desire to deter because of a concern about the defendant possibly having escaped retribution. As we will see below, the distinction that is drawn in this paper between rational and nonrational disregard lends itself to guiding juries differently in these different situations.

VII. Derivation of optimal punitive damages

Precaution, p, is an increasing function of expected total damages assessed, $a(bA+F+P)=a(C+P)$, which is given by the first order condition for the maximization of (13). Thus $p[a(C+P)]$ satisfies:

$$(17) \quad u'[p] + a(C+P) = 0.$$

We continue to assume that $p'>0$ and $p''<0$.

As above, we substitute the level of precaution, $p[a(C+P)]$, as defined by maximization of (13), in the social welfare function (14), which we repeat here:

(14) $W[P] = u[p] - (1-p)A - (1-p)\{aV[C+P-P^r] + (1-a)V[-P^r]\}.$

This gives the objective function to be maximized:

(18) $W[P] = u[p[a(C+P]] - \{1-p[a(C+P)]\}\{A+aV[C+P-P^r]+(1-a)V[-P^r]\}.$

Calculating the derivative of W, we have:

(19) $W'[P] = \{u'+A+aV[C+P-P^r]+(1-a)V[-P^r]\}ap' - (1-p)aV'[C+P-P^r]$
         $= \{A(1-ab)-a(F+P)+aV[C+P-P^r]+(1-a)V[-P^r]\}ap'$
                   $- (1-p)aV'[C+P-P^r].$

where we have used (17) to simplify (19). This gives the first order conditon presented above:

(15) $(1-p)V'[C+P-P^r] = \{A(1-ab)-a(F+P)+aV[C+P-P^r]+(1-a)V[-P^r]\}p'.$

Let us examine $W'[P]$ evaluated at $P=P^d-C$ and $P=P^r-C$ when these values are positive.

(20) $W'[P^d-C] = \{aV[P^d-P^r]+(1-a)V[-P^r]\}ap' - (1-p)aV'[P^d-P^r].$

where we have used (16), $aP^d = A$. If $P^d<P^r$, $V'<0$ and so $W'[P^d-C] > 0$. If $P^d=P^r$, W' is still positive because of the gain from deterring accidents that do not result in liability and so escape the desire to have retribution. If $P^d>P^r$, W' may be either positive or negative, depending on the balance between the cost of overpunishing and the gain from deterring improperly punished accidents.

Considering $P=P^r-C$, we have:

(21) $W'[P^r-C] = \{A-aP^r+(1-a)V[-P^r]\}ap'.$

If $P^d>P^r$, implying that $A>aP^r$, we have $W'>0$. If $P^d=P^r$, W' is still positive because of the gain from deterring accidents that do not result in liability and so escape retribution. If $P^d>P^r$, W' may be either positive or negative, depending on the balance between the cost from overdeterring relative to economic efficiency and the gain from further deterring accidents that do not result in liability and so escape the desire to have retribution.

We might have the situation that the fines, F, are sufficiently large that the desire to punish the defendants who are identified is satisfied by the fines, F, together with the compensatory damages paid, bA. To examine this we consider $W'[0]$:

(22) $W'[0] = \{A(1-ab)-aF+aV[C-P^r]+(1-a)V[-P^r]\}ap' - (1-p)aV'[C-P^r].$

This might be negative for C large enough.

We can conclude that optimal punitive damages might be zero if C is large enough. Alternatively, when optimal punitive

damages are positive, they are larger than the smaller of the ideal
levels for just economic efficiency or retribution, and may be
larger than both of them if they are not too far apart and it is
particularly important to avoid the social cost from a lack of
retribution for those who escape liability.

VIII. Nonrational disregard

The assumption of rational decision-making gives a structure
to the analysis of the effects of legal rules, a structure that has
considerable relevance since many agents are trying to pursue
their self-interest in a reasonably coherent way. However, there
are situations where such a model does not adequately describe the
behavior of a particular defendant. A prime example might be
some drunk-driving accidents, where the drivers (at least when
sober) are aware of the risks of accident, and may not be
underestimating accident probabilities, but rather are proceeding
to drink anyway, recognizing that later driving behavior will not
be decided rationally.

A crisp model of this sort of behavior is developed in
Diamond (1997a). We refer to behavior that is not rational and is
considered reckless disregard as nonrational disregard. Assume
that when a defendant acts with nonrational disregard, the
probability of an accident that injures others is treated as if it
were zero. This results in some probability of avoiding an
accident which we denote p^n and the resulting utility (gross of
liability costs) of $u[p^n]$. We continue to assume that when the
defendant acts with attention to the consequences of his actions,
then the probability evaluations are correct, the decision-making is
rational.

We assume that with probability q the defendant acts in a
deliberate way, maximizing utility net of legal costs, as in the
analysis above. With probability 1-q the defendant acts with
nonrational disregard. That is, with probability 1-q, the
defendant has utility $u[p^n]$ and (true) probability of avoiding an
accident p^n. For simplicity, we assume that all accidents from this
activity do lead to liability for compensatory damages and that
the defendant bears all of the costs of the accident as
compensatory damages.

The central new issue to be modeled is the extent to which the
legal system can distinguish between accidents happening after
nonrational disregard and accidents happening despite a rational
and accurate evaluation by the defendant of the risks involved.
For example, driving after a little bit of alcohol consumption is
legal and, if the driving is careful, is not negligent, although even
a low level of alcohol may affect the probability of an accident.
That is, the drinking may not have resulted in negligent driving,
but may have reduced the probability of avoiding an accident.
Driving after more alcohol consumption raises the sort of

efficiency question examined here, but there is not a bright line separating the two sorts of decisions. In terms of the model, rational decision-makers sometimes drive after moderate drinking but do not drive after heavy drinking, while nonrational decision-makers drive after whatever amount of drinking. Accidents can occur in any of these circumstances and a jury may not be able to tell apart someone who wouldn't drive after heavy drinking from someone who would, as well as being unsure about exactly how much drinking occurred prior to the particular accident leading to the case being evaluated.

We assume that all accidents occurring after the care chosen with nonrational disregard result in punitive damages, while a fraction f of accidents that happen after an accurate evaluation of risks also result (erroneously) in punitive damages. This failure to perfectly discriminate circumstances is realistic and central to the concerns analyzed.[32] Because of the imposition of punitive damages on some occasions when there is an accurate evaluation, the incentives to take care are too large when there is accurate evaluation - leading to inefficiency.

In this situation, we also want to recognize the cost of incorrect punishment. We assume that it is desired to punish those who act with nonrational disregard in excess of having them bear the accident costs caused, but not to punish those who act with a rational evaluation of risks beyond having them bear the accident costs.

Returning to the formal model, we assume that all accidents result in compensatory damages that equal accident costs, A, and (for simplicity) that there are no civil and criminal fines. When doing an accurate evaluation, the defendant maximizes utility as given by

(23) $U[p] = u[p] - (1-p)(A + fP)$.

where f is the probability of being held liable for punitive damages after an accident, despite having chosen a level of precaution with an accurate evaluation of risks. Thus f is a measure of the mistakes made in the assessment of punitive damages. We denote by p^r the level of precaution chosen in this circumstance of rational decision-making. Note that the level of precaution with accurate evaluation, p^r, depends on the level of punitive damages, P, since there is a risk of punitive damages. In contrast, the level of precaution when behaving with nonrational disregard, p^n, does not depend on the level of punitive damages since the possibility of an accident (and so liability for both compensatory and punitive damages) is ignored. Thus we write the function $p^r[A+fP]$, but treat p^n as a scalar, not a function.

---

32 For a discussion of erroneous assessment of punitive damages, see Ellis, 1982.

The function $p^r[A+fP]$ is defined implicitly by the first order condition for individual utility maximization. Differentiating (23) with respect to p and setting the derivative equal to zero, we have the condition that implicitly defines p^r:

(24) $u'[p^r] + A + fP = 0.$

Since u[p] is concave in p, p^r is increasing in P as long as f>0. That is, an increase in punitive damages increases precaution. We also assume that the rate of increase in precaution with the level of punitive damages is decreasing (over the relevant range). Thus, we are assuming $p^{r'}>0$ and $p^{r''}<0$.[33] This is illustrated in Figure 3, along with p^n.

Because of the risk of punitive damages even with a rationally chosen level of precaution, punitive damages lower the efficiency of rational decisions. We have assumed that punitive damages do not affect nonrational decisions. If punitive damages did not affect the fraction of the population making rational decisions, we would be in a setting similar to that in Section III - if punitive damages were wanted for retribution, they would give greater deterrence than appropriate for economic efficiency. However, if we assume that punitive damages decrease the prevalence of nonrationnal decision-making, then we have two effects on efficiency - a gain from a higher prevalence of rational decision-making,[34] but a loss in the quality of rational decision-making. That is, the efficiency purpose of punitive damages in this setting is to direct attention to the possibility of an accident, and so its consequences, in order to reduce the prevalence of nonrational disregard. In other words, we are describing nonrational behavior as a two-step process. The first step is whether the risks of punitive damages are large enough for the defendant to behave appropriately - evaluating the risks at hand and making a suitable precaution decision. In the second step, if the defendant is not behaving appropriately, then we assume that there is no impact at all from punitive damages. In terms of the drunk-driving example, we assume that the risk of punitive damages affects the propensity to drink in a situation where the defendant might drive afterwards, but does not affect how the defendant does drive if the defendant gets drunk and then drives.

We write the fraction showing rational behavior as a function only of punitive damages, q[P]. In order to have an efficiency role for punitive damages, we assume that nonrational disregard is deterred by punitive damages, q'>0, but that the effect of a marginal dollar gets smaller as the level of punitive damages

---

[33] We can express these derivatives in terms of the utility function: $u''p^{r'} = -1$; $u''p^{r''} + u'''(p^{r'})^2 = 0$. The concavity of u implies that $p^{r'}>0$. A sufficient condition for $p^{r''}<0$ is $u'''<0$.
[34] Packer (1968) has argued that an important dimension of deterrence is that it also works through unconscious motivations (page 43).

Precaution

$p^r [A + f P]$

p^n
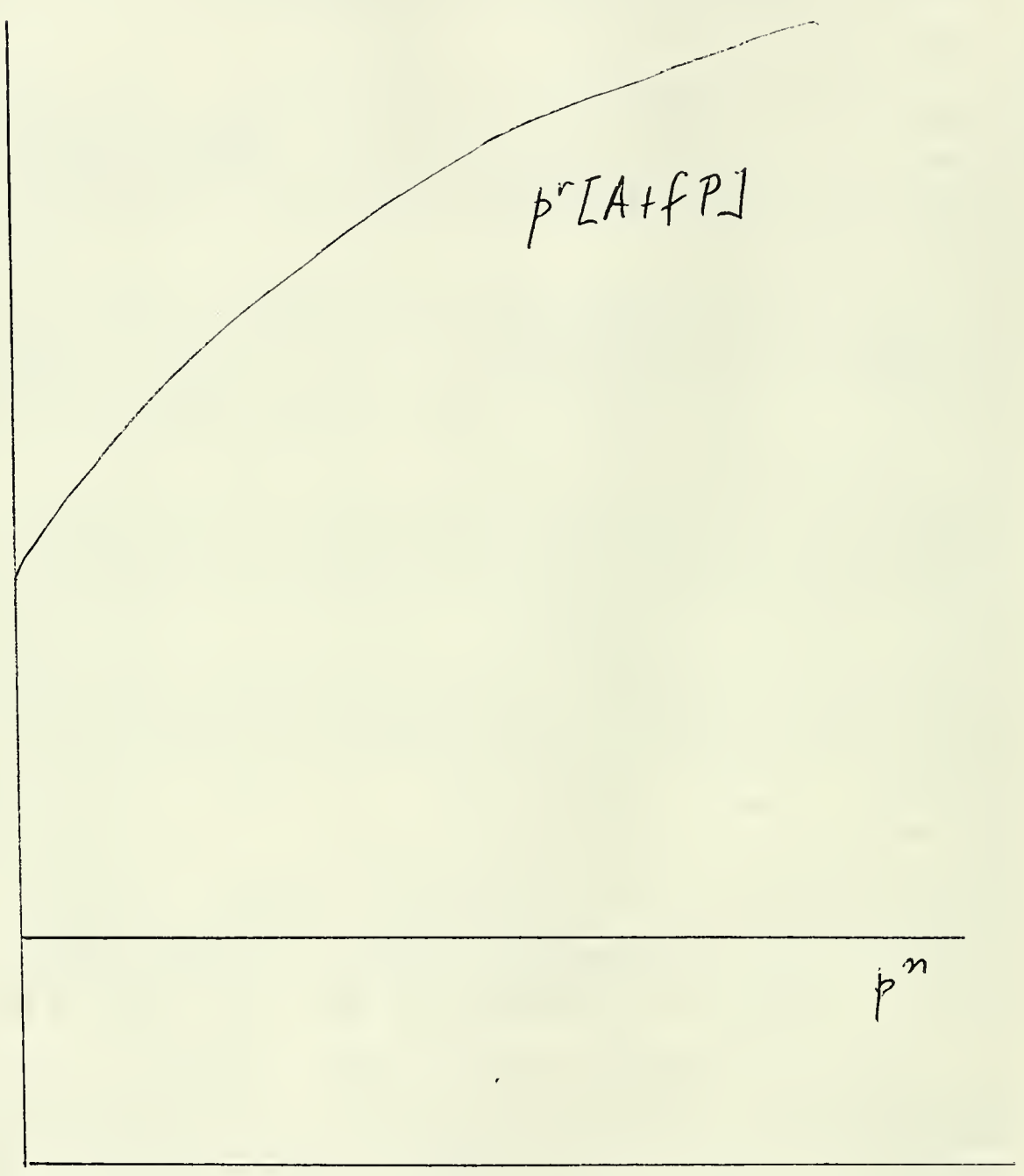
Punitive damages

Figure 3

increases, $q''<0$.[35] This is illustrated in Figure 4. Only with an impact on the fraction choosing rationally is there an efficiency role for punitive damages.[36]

We can now write social welfare as a function of the level of punitive damages by recognizing the two possibilities - of nonrational disregard and of accurate evaluation[37]:

$$(25) \quad W[P] = (1-q[P])\{u[p^n] - (1-p^n)(A+V[A+P-P^r])\}$$
$$+ q[P]\{u[p^r[A+fP]] - (1-p^r[A+fP])(A+fV[P])\}.$$

That is, social welfare adds the utilities from the activity less the expected accident costs, less the costs of inappropriate punishment. These elements are summed over both defendants showing nonrational disregard and defendants having accurate evaluations. Inappropriate punishment can happen to all of those with nonrational disregard who have accidents and the fraction f of those making a rational assessment who have accidents.

In order to have the first order condition in an intuitive form, we first define the economic gain from having the defendant make an accurate assessment of the risk, rather than showing nonrational disregard. This economic gain depends on the level of punitive damages since the level of punitive damages affects the rationally chosen level of precaution. We denote this economic gain by $G^e[P]$:

$$(26) \quad G^e[P] = \{u[p^r[A+fP]] - (1-p^r[A+fP])A\} - \{u[p^n] - (1-p^n)A\}.$$

The economic gain tells us the direct efficiency benefit from having a decision-maker be rational rather than not. This economic gain is smaller the larger the level of punitive damages since punitive damages lower the efficiency of rational decision-making; $G^{e\prime}<0$.

---

35 For example, the fraction showing nonrational disregard might shrink toward an irreducible minimum as P rises without limit. An example of a function with this property is $q[P] = k_0 - k_1/(1+P)$, for some positive constants k_i.
36 The function q[P] reflects behavioral responses to the attention drawn by punitive damages, not a rational calculation, since such rationality appears to be inconsistent with the concept of nonrational disregard. We assume that there is no direct resource cost to the change in the level of regard to this decision. That is, we assume that the individual either does or does not make a rational decision about the level of precaution. We ignore investments undertaken to affect the defendant's ability to make rational decisions, such as hiring a driver or participating in a program to learn better self-control.
37 We assume that, over the relevant range, W(P) is concave in P. Thus there will be a unique optimal level of punitive damages, given as the unique solution to the first order condition for maximization of W.
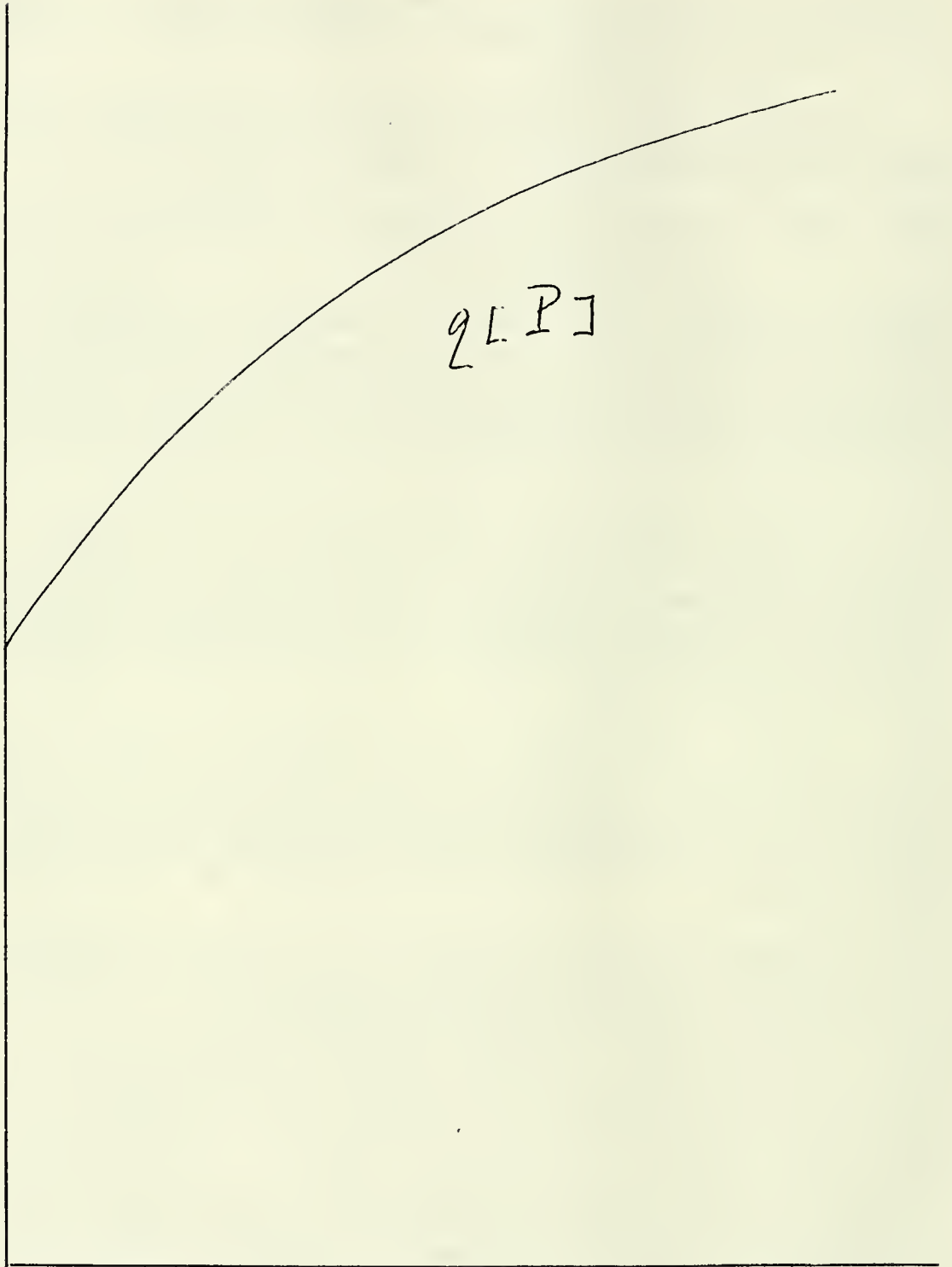
Figure 4

In addition to the economic advantage from having more people making rational assessments, there is an impact from the change in the level of inappropriate punishment. By assumption all of the nonrational who have accidents are punished, paying total damages A+P. With ideal retribution of P^r, there is a level of inappropriateness, $A+P-P^r$. The fraction f of those who make rational assessments and have accidents are wrongly punished by the full amount of punitive damages, P, since we have assumed that the ideal punishment is for them to pay only for the accident costs generated. If someone makes rational decisions, there is a social gain (or possibly a loss) from the difference in the extent of inaccurate punishment. We denote this net gain by $G^V[P]$:

(27) $G^V[P] = (1-p^r)fV[P] - (1-p^n)V[A+P-P^r]$.

That is, when someone is making a correct assessment, there is a probability $1-p^r$ of an accident, with the fraction f of these accidents resulting in punitive damages, which have a social cost V[P], since these people should not be punished. The difference between this and the same calculation for someone showing nonrational disregard is the gain from the change in inappropriate punishment. This gain expression may be positive or negative, depending on the difference in precaution levels and on the frequency of erroneous assessment of punitive damages. With an increase in punitive damages, the cost of punishing the rational might go up or down - they have fewer accidents, but the extent of inappropriate punishment is larger.

In addition to changing the number who are nonrational, a change in punitive damages has direct effects. The economic effect of raising punitive damages is to increase the precaution of potential defendants who are rational. Since these defendants would be taking the socially efficient level of precaution in the absence of punitive damages, any increase in punitive damages lowers the efficiency of the chosen level of precaution. The expected punitive damages, fP act just like a distortionary tax on this decision. The importance of this distortionary tax depends on the responsiveness of the precaution decision to financial incentives. Thus, as shown in the next section, the contribution to the social gain is $-(fP)fp^{r'}[A+fP]$, which is negative. In addition to the effect of more precaution on economic efficiency, there is an impact on the frequency of inappropriate punishment. More precaution affects economic efficiency and the frequency of inappropriate punishment. In addition to the effects of additional punitive damages that work through precaution, there is a direct effect on the extent of inappropriate punishment. This affects both those who are nonrational, by the marginal amount $(1-p^n)V'[A+P-P^r]$, and those who make accurate assessment, by the marginal amount $(1-p^r)fV'[P]$.

Bearing in mind this structure of the gains from avoiding nonrational disregard, the first order condition for the optimal level of punitive damages can be written as:

(28) $(1-p^n)V'[A+P-P^r] = q'[P]\{G^e[P]+G^v[P]\} + q[P]\{G^{e\prime}[P]+G^{v\prime}[P]\}$

The term on the left is the marginal gain from changing the accuracy of punishment of the nonrational. The terms on the right reflect the change in the numbers showing nonrational disregard and the change in the gains, both in efficiency and inappropriate punishment levels, from the level of rational decision-making that occurs in equilibrium.

With so much going on simultaneously in the model, there is little insight from examining the first order condition in general. Thus we confine analysis to a special setting. Considering just the efficiency impact of punitive damages, we can define the level of punishment that would be most efficient, P^d; that is the level of punishment that optimizes deterrence considering only economic gains.[38] Punitive damages considering only economic efficiency would be P^d-A. Let us examine the situation if this deterrence based measure is just equal to the level of punishment wanted for retribution: $P^d = P^r$. Let us denote the level of punishment by P^e when these happen to be equal. Evaluating the derivative of the social welfare funciton at this level of punishment we have:

(29) $W'[P^e-A] = - q[P^e-A](1-p^r)fV'[P^e-A]$
$+ \{q[P^e-A]p^{r\prime}-q'[P^e-A](1-p^r)\}fV[P^e-A].$

In general, this might be positive or negative - that is the optimal level of punitive damages might be larger or smaller than this ideal level, even when these two ideal levels coincide. There are two elements going on. Raising punitive damages from this level increases the inappropriate punishment of those wrongly assessed punitive damages. Raising punitive damages also changes the number of people wrongly assessed punitive damages. This change works through two effects - a larger number of rational decision-makers, but a smaller accident rate per decisionmaker. If these two impacts roughly balance, $qp^{r\prime} = (1-p^r)q'$, then optimal punitive damages should be smaller than the level desired for retribution even if that is the ideal level for deterrence purposes.

Among other elements, the balance of effects depends on the extent of the ability to distinguish between the rational and nonrational decision processes. Sometimes, as with drunk drivers, there will be direct evidence about decision processes and not just the outcome of an accident. In other settings, particularly corporate settings, it will sometimes be more difficult to distinguish between these decision processes. The presence of hindsight bias, discussed in Diamond (1997a), adds to the risk of inappropriate assessment of punitive damages. The impact of concern for punishment on the optimal level of punitive damages involves several elements. First, the level chosen just for deterrence might be larger or smaller than the ideal level of

---

38 For analysis of P^d, see Diamond, 1997a.

punishment. The larger the punitive damages the larger the concern for inappropriate application of punitive damages to those who were not reckless. But increased precaution, as a result of increased punitive damages, decreases the frequency of such inappropriate assessment. The gain in punishment from inducing less nonrational behavior may be positive or negative - the frequency of application of punitive damages is less for those making accurate assessment, but the cost of inappropriate punishment is larger per application. Thus the sign depends on the size of inappropriate assessment of punitive damages, f, when both types are inappropriately assessed.

The complexity in this section has kept us from simple conclusions, with the analysis showing the way that different important elements enter the analysis. With consideration of nonrational decision-making, guidelines for policy become more complex. With concern for both economic efficiency and retribution, policy considerations also become more complex. The possibility of jury error adds yet another layer to the needed analysis. Thus a jury setting punitive damages needs to consider the effect of punitive damages in discouraging nonrational behavior and the worth of such discouragement. The jury needs to be concerned with both overdeterring and overpunishing those at risk of being found in reckless disregard despite rational decision-making.

IX. Derivation of optimal punitive damages

To examine the optimal level of punitive damages, we maximize $W[P]$ with respect to P. First, we restate the social welfare function, recognizing that the ideal punitive damages for those taking adequate precaution is zero:

$$(30) \quad W[P] = (1-q[P])\{u[p^n] - (1-p^n)(A+V[A+P-P^r])\} + q[P]\{u[p^r[A+fP]] - (1-p^r[A+fP])(A+fV[P])\}.$$

It is convenient to state this in terms of the social gains from having someone do an accurate evaluation of risks rather than behave with nonrational disregard. We divide this gain into two parts - the economic gain and the punishment gain. Denoting the economic gain by $G^e[P]$:

$$(31) \quad G^e[P] = u[p^r[A+fP]] - (1-p^r[A+fP])A - \{u[p^n] - (1-p^n)A\}.$$

Denoting the punishment gain by $G^v[P]$, we have:

$$(32) \quad G^v[P] = (1-p^n)V[A+P-P^r] - (1-p^r[A+fP])fV[P].$$

Using these social gain functions, we can rewrite the social welfare function (30) in the more convenient form:

$$(33) \quad W[P] = u[p^n] - (1-p^n)(A+V[A+P-P^r]) + q[P]\{G^e[P]+G^v[P]\}$$

Differentiating (33) with respect to P, we have:

$$(34) \quad W'[P] = -(1-p^n)V'[A+P-P^r] + q'[P]\{G^e[P]+G^V[P]\} \\ + q[P]\{G^{e\prime}[P]+G^{V\prime}[P]\}$$

To analyze this, we consider the derivatives of the two gain functions. Differentiating (31) and using the individual first order condition and the properties of p^r, we have:

$$(35) \quad G^{e\prime}[P] = (u'[p^r[A+fP]] + A)fp^{r\prime}[A+fP] \\ = -(fP)fp^{r\prime}[A+fP] < 0.$$

The sign of $G^{e\prime\prime}$ is ambiguous. Differentiating (32) we have

$$(36) \quad G^{V\prime}[P] = (1-p^n)V'[A+P-P^r] - (1-p^r)fV'[P] + p^{r\prime}[A+fP]fV[P].$$

When $A+P=P^r$, the first term is zero, but the other two terms are of opposite signs. The sign of $A+P-P^r$ determines the sign of the first term.

We define P^d by considering the incentive that would optimize just economic efficiency. The value is defined by the first order condition for maximizing W, assuming that V and V' are uniformly zero.

$$(37) \quad 0 = q'[P^d-A]G^e[P^d-A] + q[P^d-A]G^{e\prime}[P^d-A].$$

If we then evaluate the derivative of W at this value of punitive damages, P^d-A, we have:

$$(38) \quad W'[P^d-A] = -(1-p^n)V'[P^d-P^r] + q'[P^d-A]G^V[P^d-A] \\ + q[P^d-A]G^{V\prime}[P^d-A]\}$$

If we evaluate the derivative of W at the ideal retributive level of punitive damages, P^r-A, we have:

$$(39) \quad W'[P^r-A] = q'[P^r-A]\{G^e[P^r-A]+G^V[P^r-A]\} \\ + q[P^r-A]\{G^{e\prime}[P^r-A]+G^{V\prime}[P^r-A]\} \\ = q'[P^r-A]\{G^e[P^r-A] - (1-p^r)fV[P^r-A]\} \\ + q[P^r-A]\{G^{e\prime}[P^r-A]-(1-p^r)fV'[P^r-A] +p^{r\prime}fV[P^r-A]\}.$$

If P^d equals P^r, both V and V' when evaluated at P^d-P^r are zero. With $P^d=P^r$, we have:

$$(40) \quad W'[P^r-A] = q'[P^r-A]G^V[P^r-A] + q[P^r-A]G^{V\prime}[P^r-A] \\ = \{q[P^r-A]p^{r\prime}-q'[P^r-A](1-p^r)\}fV[P^r-A] \\ - q[P^r-A](1-p^r)fV'[P^r-A].$$

The first term is ambiguous, reflecting the change in the number of inappropriately punished when punitive damages are increased - a change coming from two sources, an increase in the number of nonreckless at risk of improper punishment, but a decrease in

their accident rate. The second term is negative, reflecting the loss from increasing the punishment of the nonreckless.

X.   Another model of nonrational disregard and jury error

The model in Section VIII assumed some level of jury error about recklessness without modelling a basis for those errors. It seems useful to go one step deeper to derive the level of jury error from a model of what can be observed. We now construct an explicit model of jury error based on observability, one that is close to the model above for analytical purposes. This approach uses the idea that individual rational decision-making involves only stochastic control of one's later behavior.[39]

As before, we assume that nonrational decision-makers have a precaution level p^n. Rational agents sometimes behave in the same way that nonrational people behave all the time, and sometimes behave with higher precaution, which we denote by p^r. When people behave with higher precaution we will refer to that as being attentive. There is a cost of being attentive, of not being reckless, a fraction r of the time. To capture this cost, we write utility as the function $u^a[r]$. As with $u[p]$, we assume that $u^a[r]$ is first increasing and then decreasing, reflecting increasing costs as agents are more attentive (i. e., as r increases) and an infinite cost for not being reckless ever. The function u^a also includes all the elements that entered the utility gross of legal liability above.

The choice of a frequency, r, of being attentive results in an average precaution level p, which satisfies

(41) $p = (1-r)p^n + rp^r$.

Solving for r, we have

(42) $r = (p-p^n)/(p^r-p^n)$.

Thus we can write the utility associated with an average precaution level p as

(43) $u[p] = u^a[(p-p^n)/(p^r-p^n)]$.

We assume that juries observe the precaution level at the time of the accident, p^r or p^n, but not the attention level or the average precaution level. Thus all accidents by nonrational individuals (who are never attentive) result in punitive damages, while rational agents are subjected to punitive damages after accidents when they were not attentive. This happens in a fraction of their accidents, f, that satisfies

------

39 For examples of this type of modelling of behavior, see Diamond, 1974a and b.

(44) $f = (1-r)p^n / \{(1-r)p^n + rp^r\}$.

As in the model above, f is the fraction of accidents by rational decision-makers that results in punitive damages being assessed. This differs somewhat from the model above since f is now influenced by the choice of r, which can be replaced analytically by the monotonically related precaution level, p. Thus individual behavior can be written in the form of $u[p] - (1-p)(A+f[p]P)$. This is similar to the structure used above.

If we assume that the ideal punishment for the truly reckless is P^r, while the ideal punishment for the rational who have had a reasonable level of being inattentive is just that they pay accident costs, then we have written this model in a form close to that above. That is, f represents jury error in assessing punitive damages where punishment is not appropriate. With this structure of information, we have made jury error inevitable – since some fraction of accidents following appropriate precaution is indistinguishable to the jury from accidents by the reckless, either the jury must forgo punishing the reckless or the jury must sometimes punish the nonreckless. This tradeoff between type one and type two errors is an inevitable part of jury evaluations. Different standards for jury findings give different mixes of type one and type two errors, but the presence of a tradeoff is inevitable.

We do not explore the differences in models, using the formulation just to suggest how jury error and the avoidance of jury error can be incorporated into the analysis.

XI Conclusions

Punitive damages are meant to deter and to punish. The presence of this double purpose greatly complicates selection of a suitable level of damages. It seems implausible that juries, given little in the way of instructions for fulfilling this double purpose could reason through to a satisfactory level of damages on a consistent basis.[40] At present, judicial oversight of damage levels also has little guidance about the level of damages (as opposed to the process). This paper (and my previous paper, 1997a) argue for the importance of distinguishing among different bases for levying punitive damages and giving different guidance to both juries and judges in the different situations. Without such distinctions, it seems implausible that punitive damages could serve their social role.

The first distinction to draw is between situations of reckless disregard and those of malice. This paper did not consider situations of malice. In my previous paper (1997a) which

---

40 For a discussion of giving more guidance to the jury, see Owen (1994).

considered only the issue of deterrence, I argued that jury instructions should be different in situations of malice from those of reckless disregard. In particular, in the presence of malicious intent the court should examine the preferences of the defendant while selecting a suitable punitive amount to deter behavior that generates gains that are not suitable for inclusion in social accounting. The focus is precisely on the preferences (including any potential for ill-gotten gains) of the defendant. In contrast, reckless disregard of a risk is behavior that is viewed as outrageous as a result of inadequate attention to a cost falling on the plaintiff. The inadequate attention could come from a rational disregard of costs that the tort system will not assess or from a nonrational disregard of the costs that the tort system will assess. Thus, jury instructions in the case of reckless disregard should focus on the costs that are not adequately represented in the defendant's decision process. Presumably, consideration of retribution would reinforce the advantages of drawing distinctions since the basis for retribution is different with malice and reckless disregard.[41]

Current instructions do not guide the jury through explicit consideration of deterrence and retribution separately. Nor do they indicate how the two goals of deterrence and punishment, if thought about separately are then to be combined. This paper argues that the starting place for this combination is an average of what would be done separately for each of the goals, minus any civil or criminal fines that are paid. The first question to ask is whether the ideal level of retribution is larger or smaller than would be set if economic deterrence were the only goal. There is no simple answer to this question, with outcomes varying across situations.

In a situation where punitive damages are not needed for deterrence, accepting a desire to punish implies an average of the level desired for retribution and the zero level appropriate for deterrence. With more complicated needs for deterrence, additional factors may require damages which are not an average of the two ideal levels. The models above imply that these additional factors enter differently in situations where some accidents do not result in liability and in situations where all accidents result in suitable compensatory liability.

In my earlier paper, it was argued that the role of punitive damages in improving deterrence differed between situations of

---

41 Intentional harms have been distinguished from reckless disregard as a suitable basis for punitive damages. For example, Owen (1985): "If our accident law of torts should be abolished, as one day it probably should, a question arises whether there is any remaining place in the legal system for retribution or deterrence. I think there is. ... I believe that some form of punitive damages should be retained for intentional torts. [footnote omitted] The goals of retribution, corrective justice, and deterrence are most appropriate in redressing intentionally inflicted harm." (page 670).

rational disregard and situations of nonrational disregard. For the former, one wanted to consider the costs that did not enter into decisions because the defendant did not expect to pay them. For the latter the focus was on inducing the nonrational to behave more rationally. Moreover, it was argued that concerns about inappropriate deterrence were different, since recognizing when costs might not be faced was different from recognizing when rational decisions were absent.

This paper argued that recognizing the desire for retribution further emphasizes the difference between these settings. With some accidents avoiding punishment, there is an added reason for deterrence. In contrast, possibly punishing rational decisionmakers who faced all social costs was a reason for lowering punitive damage levels.

The concerns identified here should be turned into different jury instructions in different settings. It might be useful to research the thinking of juries with more ex post interviews. When juries use large punitive damages judgments to "send a message" is this a form of punishment or a strong deterrence motive? If, as I suspect, it is the latter, then there are two important questions to ask.[42] One is whether the punishment is excessive relative to retribution standards, and, along the lines of the analysis above, there should be some compromise between deterrence and retribution ideal levels. Second is whether the jury version of deterrence by extreme punishment, while possibly suitable in some settings of malicious intent, is a manifestation of hindsight bias in a setting of reckless disregard. If the jury fails to recognize both the costs of overdeterrence and the risk of accident despite appropriate care, then these punishments are likely to be too large.

This interaction of punishment with deterrence will be explored in a corporate setting of vicarious liability in a forthcoming paper (Diamond (1997c)).

---

42 For example, two of the jurors in the Exxon Valdez case have been quoted: "Our job was to make sure they'd never do it again." Nancy Provost, cited by Munk, 1994, p. 89.
"we wanted to make sure Exxon is never going to do this again." Spann, quoted by Natalie Phillips, Anchorage Daily News, January 22, 1995, back page.

Appendix A. Ideal punishment varying with precaution, individual precaution unobservable

In the analysis above, the ideal retribution was based on the occurrence of an accident. No adjustment was made either for the typical level of precaution of all defendants, or the actual level of precaution of this particular defendant. We pursue these two ideas in this section. For simplicity, we assume full bearing of accident costs and no civil or criminal penalties.

First we assume that the precaution of the particular defendant is not observable by the jury setting punitive damages. But we assume that the jury does know the typical level of precaution, and has an ideal retribution that varies with the typical level of precaution, denoted $P^r[p]$. Thus we rewrite social welfare as:

(45) $W[P] = u[p] - (1-p)A - (1-p)V[A+P-P^r[p]]$.

We continue to assume that social welfare is a concave function of punitive damages. Since the actual precaution of the particular defendant is not observable, the level of punitive damages, P, does not vary with the precaution of the particular defendant. Thus the behavior of the defendant, with precaution as a function of the sum of compensatory and punitive damages, $p[A+P]$, is the same as above.

When we now maximize social welfare, given in (45), assuming that precaution is chosen to maximize defendant utility, given in (3), we obtain a first order condition that has one more term than previously. In place of (5), we now have:

(46) $(1-p)V'[A+P-P^r](1-P^{r\prime}p') = (V[A+P-P^r] - P)p'[A+P]$.

Previously, the expression $(1-p)V'$ was multiplied by 1; now it is multiplied by $(1-P^{r\prime}p')$, which is more than one since $P^{r\prime}$ is negative, while p' is positive. That is, a higher level of punitive damages increases precaution; a higher level of precaution decreases the ideal punishment, thereby closing some of the gap between actual and ideal punitive damage levels. It remains true that the optimal level of punitive damages lies below the ideal level, in order to balance the deterrence and retribution goals, and above the social cost of inappropriate punishment, V.

To see the effect of this change in the social welfare function, we compare optimal punitive damages when ideal retribution does and does not vary with precaution. Let us denote the optimal level of punitive damages when ideal damages do not vary by P_0; that is P_0 is the solution to (5) when ideal damages are P^r. With this level of actual punitive damages, the equilibrium level of precaution, is $p[A+P_0]$. In order to make the comparison interesting, we want to select a function for ideal damages that passes through the same value; that is, that has $P^r[A+p[P_0]]$ equal

to P^r. Now we can ask whether it is better to raise or lower the level of punitive damages when the ideal retribution has changed from a scalar to this function of the precaution level. In other words we want to evaluate the derivative of W in (45), evaluated at the point where (5) holds. Evaluating the derivative of welfare at this level, the derivative is not equal to zero, but is equal to $(1-p)V'[A+P-P^r](P^{r'}p')$. This expression is positive, implying that it would be good to increase the level of punitive damages. In other words, when the ideal level of retribution decreases with the level of precaution, the optimal level of punitive damages is more than it would be if the ideal level did not vary, assuming that the ideal function passes through the previous equilibrium.

The analysis would become much more complicated if we assumed that precaution was observable. In this situation, the juries collectively are not selecting a typical level of punitive damages, P. Rather the juries collectively are now selecting a level of punitive damages that varies with the actual precaution of the defendant, P[p]. When choosing precaution, the defendant now pays attention to the fact that punitive damages will be lower in the event of an accident should a higher level of precaution be chosen. Thus the first order condition for precaution depends not only on the level of punitive damages, but also on how they vary with precaution. If all defendants were the same, it would be easy to manipulate the threatened increase in punitive damages in order to encourage precaution cheaply (from an efficiency point of view). Thus, to analyze this model in an interesting fashion, we would need to consider heterogeneous defendant characteristics that result in a distribution of actual choices of precaution. We would then be selecting punitive damages as a function of precaution in order to do this maximization. The formal analysis would use the techniques developed by Mirrlees (1971) in his analysis of income taxation. Such an analysis is sufficiently complex that it is not considered here.

Appendix B. Punishment and wealth

One can have different views as to whether ideal retribution should vary with the wealth of the defendant. The question we shall explore is how optimal retribution should vary with wealth if one assumes that ideal retribution varied with wealth. In particular we assume that ideal retribution is proportional to wealth. This situation is chosen for analysis because plaintiffs commonly make arguments that call for punitive damages proportional to the wealth of defendants. Thus, we assume that ideal punishment is written as $P^r Y$, where Y is the measure of wealth relevant for describing ideal retribution. There remains the issue of how to scale deviations from ideal retribution. The two simple situations are to evaluate these deviations in absolute terms and in proportional terms. If we scale deviations in proportional terms, then we would write V as a function of $[((A+P)/Y)-P^r]$. On the other hand, if we scale deviations in

absolute terms, then we would write V as a function of $[A+P-P^rY]$. Using these two formulations, we can reexamine the first order condition to derive how punitive damages should vary with the wealth of the defendant if one assumes proportionality in ideal punishment.

Since we are balancing two objectives, we can not fully satisfy either of them. Thus we are in the realm of second-best analysis. Analytical results in this realm are always complicated since the answer depends not only on variation in the objectives but also on how behavioral responses affect the implicit weighting of the factors in achieving an optimal balancing. That is, the responsiveness of precaution to punitive damages is a critical part of the balancing of objectives. Yet that responsiveness can have a complicated (indirect) relationship with wealth, opening up many possible patterns of optimal punitive damages. These complexities limit firm conclusions from theoretical analysis without more specific description of both the function V and the behavior of defendants.

For a given level of wealth, this problem is an example of the one analyzed above. Thus we carry over the results that $V[A+P-P^rY] < A+P < P^rY$. With deviations in proportional terms, it does not necessarily follow that optimal punitive damages should increase with the wealth of the defendant, even though ideal damages are assumed to be proportional to wealth. With deviations in absolute terms, then punitive damages should increase with the wealth of the defendant. Provided that the curvature of the precaution function is not too large, in both situations, when punitive damages should increase with wealth, they should do so less than in proportion to wealth. However a great curvature in the precaution being chosen by defendants of some wealth level might reverse this conclusion.[43] The derivations of these two results are in the next section.

Derivation of optimal punitive damages

We now extend the analysis in Section IV to the situation where the desired punishment is a fraction of wealth and deviations from this ideal are judged in relative terms, so that we rewrite the disutility of deviations from ideal punishment as $V[((A+P)/Y) - P^r]$. We consider deviations evaluated in absolute terms separately. Social welfare is the sum over individuals of different wealths of the social welfare coming from their own utility, the expected accident costs they cause and the social cost of inadequate punishment. With no further changes in the formulation, we rewrite the part of the social welfare function coming from a typical defendant with wealth Y as:

---

43 There would also be a, presumably small, efficiency cost from implicitly taxing wealth since the same activity results in higher damages for people who saved more or earn more.

(47) $W[P; Y] = u[p[A+P]] - (1-p[A+P])A - (1-p[A+P])V[((A+P)/Y) - P^r]$.

This differs from (2) only in the change in V. For a given level
of wealth, this problem is an example of the one analyzed above.
Thus we carry over the results that $V[((A+P)/Y) - P^r] < P < P^r Y$.
Recalculating the first order condition, (5) becomes:

(48) $\{u'[p] + A + V[((A+P)/Y) - P^r]\}p'[A+P] - (1-p)V'[((A+P)/Y) - P^r]/Y$
$= 0$.

To calculate the variation of optimal punitive damages with
respect to wealth, we differentiate (48) implicitly:

(49)
$$\frac{dP}{dY} = \frac{\{(A+P)V'p' - (1-p)(A+P)V''/Y - (1-p)V'\}/Y^2}{2V'p'/Y - (1-p)V''/Y^2 + u''(p')^2 + (V-P)p''}.$$

From the assumed concavity of W in P, the denominator is
negative. With $V'<0$, $p'>0$, and $V''>0$, the first two terms in the
numerator are also negative, although the third term is positive.
One way of examining these terms is in terms of the elasticity of
accident probability with respect to punitive damages, $-Pp'/(1-p)$.
If the elasticity is approximately one, so that accident probability
declines proportionally to punitive damages, then the first and
third terms roughly balance and the second term gives the
expected negative sign. Large elasticities are needed to reverse
the sign. Thus if P is increasing with Y, the numerator is smaller
in absolute value than the sum of the first two terms. In the
denominator, with $p''<0$ and $P>V$, the first three terms are
negative, but the fourth is positive.

Converting (49) into an elasticity, we have

(50)
$$\frac{Y}{(A+P)}\frac{dP}{dY} = \frac{(A+P)V'p'/Y - (1-p)(A+P)V''/Y^2 - (1-p)V'/Y}{(A+P)\{2V'p'/Y - (1-p)V''/Y^2 + u''(p')^2 + (V-P)p''\}}.$$

The positive sign of the third term in the numerator, the number
2 in the denominator and the negative sign of the third term in
the denominator, push toward an elasticity below one, but the
fourth term in the denominator pushes the other way. We can
conclude that when the curvature of the precaution function is
small, the elasticity is less than one. This can be expressed
alternatively in terms of u''' being small.

Turning to the situation where the desired punishment is a
fraction of wealth and deviations from this ideal are judged in
absolute terms, we rewrite the disutility of deviations from ideal
punishment as $V[A+P - P^r Y]$. We rewrite the part of the social
welfare function coming from a typical defendant with wealth Y
as:

(51) $W[P; Y] = u[p[A+P]] - (1-p[A+P])A - (1-p[A+P])V[A+P-P^r Y]$.

Recalculating the first order condition, (5) becomes:

(52) $\{u'[p] + A + V[A+P-P^r Y]\}p'[A+P] - (1-p)V'[A+P-P^r Y] = 0$.

To calculate the variation of optimal punitive damages with respect to wealth, we differentiate (52) implicitly:

(53)
$$\frac{dP}{dY} = \frac{P^r V' p' - (1-p)P^r V''}{2V'p' - (1-p)V'' + u''(p')^2 + (V-P)p''}.$$

The numerator is negative, so the derivative is positive. Converting (53) into an elasticity, we have

(54)
$$\frac{Y}{(A+P)}\frac{dP}{dY} = \frac{P^r Y\{V'p' - (1-p)V''\}}{(A+P)\{2V'p' - (1-p)V'' + u''(p')^2 + (V-P)p''\}}.$$

The term with u" in the denominator and the factor 2 in the denominator both contribute toward an elasticity below one. However, the term with p" pushes in the other direction as does the presence of $P^r Y/(A+P)$, which is greater than one. The elasticity will be less than one when the curvature of the precaution function is small and the punishment is close to ideal.

References

Breyer, Stephen, 1993, <u>Breaking the Vicious Circle: Toward Effective Risk Regulation</u>, Cambridge: Harvard University Press.

Diamond, Peter, 1974a, Single Activity Accidents, <u>Journal of Legal Studies</u> 3 (1), January, 107-162.

Diamond, Peter, 1974b, Accident Law and Resource Allocation, <u>Bell Journal of Economics and Management Science</u>, 5 (2), Autumn, 366-405.

Diamond, Peter, 1997a, Efficiency Effects of Punitive Damages.

Diamond, Peter, 1997b, Horizontal Equity, Ideal Taxation, and Optimal Taxation, forthcoming.

Diamond, Peter, 1997c, Vicarious Liability for Punitive Damages, forthcoming.

Ellis, Dorsey D., Jr., 1982, Fairness and Efficiency in the Law of Punitive Damages, 56 <u>Southern California Law Review</u>, 1, 1-78.

Fischhoff, Baruch, 1982, Debiasing, Ch. 31 in D. Kahneman, P. Slovic and A. Tversky (eds.), <u>Judgment under Uncertainty: Heuristics and Biases</u>, Cambridge: Cambridge University Press, 422-444.

Goldman, Alan H., 1979, The Paradox of Punishment, 9 <u>Philosophy and Public Affairs</u>.

Huber, Peter W., and Robert E. Litan, <u>The Liability Maze</u>, Washington D. C.: The Brookings Institution.

Mirrlees, J. A., 1971, An exploration in the theory of optimum income taxation, <u>Review of Economic Studies</u>, 38, 175-208.

Morris, Clarence, 1931, Punitive Damages in Tort Cases, <u>Harvard Law Review</u>, 44: 1173-1209.

Munk, Nina, 1994, "We're partying hearty!" <u>Forbes</u>, October 24, 84-90.

Owen, David G., 1985, Deterrence and Desert in Tort: A Comment, <u>California Law Review</u>, 73: 665-676.

Owen, David G., 1994, A Punitive Damages Overview: Functions, Problems and Reform, <u>Villanove Law Review</u>, 39: 363-413.

Packer, Herbert L., 1968, <u>The Limits of the Criminal Sanction</u>, Stanford: Stanford University Press.

Polinsky, A. Mitchell, and Steven Shavell, 1997, Punitive Damages: An Economic Analysis, Discussion Paper 212, Harvard Law School.

Posner, Richard A., 1975, The Economic Approach to Law, 53 Texas Law Review, 757-783.

Rawls, John, 1955, Two concepts of rules, The Philosophical Review, 64, 1, 3-32.

Shavell, Steven, 1981, A Note on Efficiency vs. Distributional Equity in Legal Rulemaking: Should Distributional Equity Matter Given Optimal Income Taxation?, American Economic Review, 71, 2, May, 414-18.

Wertheimer, Alan, 1975, Should the Punishment Fit the Crime? Social Theory and Practice, 3, 4, 403-423.