International Review of Law and Economics (1992) **12**, 345–355

A Note on Marginal Deterrence

STEVEN SHAVELL

Harvard Law School, Cambridge, MA 02138, USA

I. Introduction and Summary

The theory of deterrence has been concerned primarily with situations in which individuals consider whether to commit a *single* harmful act. For instance, a person may be deciding whether to discharge a pollutant into a lake. In some contexts, however, a person may be contemplating which of *several* harmful acts to commit—whether to discharge a pollutant into a lake or instead to discharge it onto the ground (where it might cause a different level of harm). In such contexts, the threat of sanctions plays a role in addition to the usual one of deterring individuals from committing harmful acts: it influences which harmful acts *undeterred* individuals choose to commit. Notably, undeterred individuals will have a reason to commit less rather than more harmful acts if expected sanctions rise with harm.

This tendency is sometimes said to reflect *marginal deterrence* because an individual will be deterred from committing a more harmful act owing to the difference, or margin, between the expected sanction for it and for a less harmful act. The term "marginal deterrence" seems to be due to Stigler (1970), but the notion has been well known from the time of some of the earliest writing on sanctions. See Beccaria (1770, 32), Montesquieu (1748, Book VI, Ch. 16, 161–62), and Bentham (1789, 171). Bentham, for example, states (citing an essentially identical passage of Montesquieu) that an object of punishment is "to induce a man to choose always the least mischievous of two offenses; therefore where two offenses come in competition, the punishment for the greater offense must be sufficient to induce a man to prefer the less."

A point of the present note, however, is that considerations of marginal deterrence are not a *raison d'être* for sanctions to rise with harm. Optimal sanctions rise with harm in models with marginal deterrence *only* if one makes a particular assumption about enforcement effort—that it is of a general nature (in a sense to be defined). But this assumption *also* implies that optimal sanctions rise with harm in the usual models without marginal deterrence. Still, as will be noted, marginal deterrence does have a more refined implication for optimal sanctions under the assumption of general enforcement effort.

To investigate marginal deterrence, I consider a simple model with monetary sanctions in which each person can do nothing or commit one of two harmful acts: either act 1, a low harm act, or act 2, a high harm act. Because individuals choose between

I thank Louis Kaplow, A. Mitchell Polinsky, and an anonymous referee for comments and the National Science Foundation (grant no. SES-8821400) for support.

two harmful acts, the model allows for marginal deterrence. This *two-act model* was introduced in Reinganum and Wilde (1986) and is studied further in Wilde (1989). I consider also a related *one-act model* in which some individuals choose whether to commit act 1, others choose whether to commit act 2, but none has the opportunity to choose between the two acts. Hence, there is no possibility for marginal deterrence in the one-act model. Comparison of these models will allow us to determine the influence on optimal enforcement of marginal deterrence, of the opportunity of individuals to choose between committing different harmful acts.[1]

As was indicated, the conclusions depend on the nature of enforcement effort. Suppose first that enforcement effort can be controlled independently for each harmful act, so that the probability of apprehension is *specific* to each act. Then in both models the optimal sanction for each act is the maximal sanction, the entire wealth of a person. The reason is well known and due essentially to Becker (1968): If the sanction for an act were less than maximal, the sanction could be raised and the probability of apprehension lowered so as to keep the expected sanction for the act constant; deterrence of the act would therefore be maintained, but enforcement resources conserved; hence, social welfare could be improved. Thus, in both models, optimal sanctions are equal, to wealth, for acts 1 and 2.[2] Optimal probabilities of apprehension, however, are generally different for the acts (higher for act 2 than for act 1 under certain assumptions).[3]

For optimal sanctions to be different for the two acts, enforcement cannot be specific to the act. Suppose, instead, that enforcement effort is of a *general* nature, affecting in the same way the probability of apprehension for committing different harmful acts; therefore, assume that the probability of apprehension for committing act 1 *equals* that for committing act 2.[4] Then the argument of Becker does not apply independently for each act; if the probability of apprehension is lowered for act 1, the probability is simultaneously lowered for act 2. (How, exactly, this alters the Becker argument is best understood from the analysis.) It is shown under this assumption that in *both* the one-act model and the two-act model, the sanction for act 1 is typically lower than that for act 2, which is maximal. (Hence the earlier statement that marginal deterrence is not in itself a reason for sanctions to rise with harm.) The explanation is that in both models, it is best for the *expected* sanction for act 1 to be

[1]The principal contribution of this note is that it compares the one-act model to the two-act model. To carry out the comparison, it is simpler to consider the case of monetary sanctions. Reinganum and Wilde (1986) and Wilde (1989) restrict attention to the two-act model and emphasize the case of non-monetary sanctions.

[2]That optimal sanctions are extreme in the two-act model when enforcement is specific is first observed in Reinganum and Wilde (1986).

[3]Consideration of marginal deterrence enters into the determination of the optimal probabilities of apprehension in the two-act model, but there is no necessary relationship between these optimal probabilities and those in the one-act model.

[4]In Shavell (1991), I analyze and contrast general and specific enforcement effort in a one-act model. The assumption of general enforcement effort is appropriate whenever, by virtue of his activity, an enforcement agent has the opportunity to apprehend those committing different types of violations. For example, a policeman on the beat will be able to apprehend both car thieves and burglars, whoever he happens to see committing a crime. However, the policeman will not necessarily apprehend thieves and burglars with the same probability; the assumption that general enforcement effort results in the same probability of apprehension for different acts is a simplifying one, the importance of which is noted in the concluding remarks.

lower than the expected sanction for act 2; and since the probability of apprehension for both acts must be the same, the optimal sanction for act 1 is less than that for act 2. It is also shown that in the one-act model, the optimal expected sanction for act 1 typically equals the harm done by act 1, whereas in the two-act model the optimal expected sanction for act 1 is *below* the harm done by act 1. (This is the more refined conclusion about the difference that marginal deterrence makes.)

The interpretation of the latter result reflects marginal deterrence: it is socially beneficial in the two-act model to lower the expected sanction for act 1 somewhat below the harm that act does, for this induces certain individuals who would have committed the more harmful act 2 to commit act 1 instead (even though some individuals who would not have committed any harmful act now commit act 1).

The model and the analysis is presented in Section II, and concluding remarks are offered in Section III.

II. The Model

The model is as described in the Introduction. Risk-neutral individuals may commit harmful acts, of which there are two: act 1, resulting in a low level of harm, and act 2, resulting in a high level of harm. If an individual commits a harmful act, he derives a benefit; otherwise, he does not. In one version of the model, the one-act model, half of the individuals choose whether or not to commit act 1, and half of the individuals choose whether or not to commit act 2.[5] In the other version of the model, the two-act model, each individual may choose whether to commit either act 1 or act 2. If an individual commits a harmful act and is apprehended, he will pay a money sanction. Specifically, let

h_i = harm due to act i; $i = 1, 2$; $0 < h_1 < h_2$;

b_i = benefit if an individual commits act i; $b_i \in [0,\bar{b}]$; $h_2 < \bar{b}$;

$f_i(b_i)$ = probability density of b_i; f_i is positive on $[0,\bar{b}]$;

w = wealth of each individual;

s_i = sanction for committing act i; $s_i \in [0,w]$.

The total population size is 1. In the two-act model, the benefits b_1 and b_2 of individuals are independently distributed.[6]

Social welfare equals the benefits individuals derive from their acts less the harm done less enforcement costs (to be described).

Observe that first-best behavior in the one-act model is for an individual to commit act i if and only if $b_i \geqslant h_i$.[7] In the two-act model, first-best behavior is for an individual to commit act i if and only if $b_i \geqslant h_i$ and $b_i - h_i \geqslant b_j - h_j$ $(i \neq j)$.

The one-act and two-act models will now be compared under the assumption that enforcement effort is specific to the act and then that enforcement effort is general.

[5] The fraction one-half is used for concreteness; it will be evident that none of the propositions depends on the assumption about the fraction of the population who may choose a particular act.

[6] It will be clear that the propositions to be established do not depend on this simplifying assumption.

[7] In the case where $b_i = h_i$, I adopt the convention that it is best for a person to commit the harmful act, and I make a similar assumption later that an individual will commit a harmful act if b_i equals the expected sanction.

A. Specific Enforcement

If enforcement effort is *specific* to the act, let

e_i = enforcement effort devoted to apprehending those who commit act i;

$p_i(e_i)$ = probability of apprehending someone who commits act i; $p_i'(e_i) > 0$.

One-act model. In the one-act model, a person will commit act i if and only if $b_i \geqslant p_i s_i$. Social welfare is therefore

$$.5\int_{p_1 s_1}^{\bar{b}} (b_1 - h_1) f_1(b_1) db_1 + .5\int_{p_2 s_2}^{\bar{b}} (b_2 - h_2) f_2(b_2) db_2 - (e_1 + e_2) \qquad (1)$$

The first term is associated with those who commit act 1, the second with those who commit act 2, and the third is enforcement effort. Exp. (1) is to be maximized over the s_i and e_i. Here and throughout this paper, * will denote optimal values of variables. The e_i^* is assumed to be positive (otherwise the enforcement problem is not of interest). The following proposition will be shown in the appendix (as are the other propositions).

Proposition 1. In the one-act model with specific enforcement, (a) optimal sanctions for the two acts are the same and equal to the maximal sanction, wealth. (b) The expected sanction for each act is less than the harm it causes. (c) The optimal probabilities of apprehension for the acts are generally different (and are determined by condition (A1) in the appendix).

Notes.[8] (1) It is possible that $p_1^* w > p_2^* w$. However, a sufficient condition for $p_1^* w < p_2^* w$ is that the functions p_1 and p_2 are equal and that the densities f_1 and f_2 are equal.

(2) The reason that $p_i^* w < h_i$ is that if $p_i w$ were equal to h_i, a reduction in e_i would allow a first-order savings in enforcement effort, but it would not result in a first-order loss due to underdeterrence, since those who would just be willing to commit act i would obtain benefits of approximately h_i.

Two-act model. In the two-act model, an individual will commit act 1 if $b_1 \geqslant p_1 s_1$ and $b_1 - p_1 s_1 \geqslant b_2 - p_2 s_2$; he will commit act 2 under a similar condition; and he will commit neither act if $b_i < p_i s_i$ for both i. Figure 1 illustrates the regions in which act 1, act 2, or neither will be committed. From Figure 1, it is apparent that the following expression gives social welfare:

$$\begin{aligned} &\int_{p_1 s_1}^{\bar{b}} \int_0^{b_1 - p_1 s_1 + p_2 s_2} (b_1 - h_1) f_2(b_2) db_2 f_1(b_1) db_1 \\ &+ \int_0^{p_1 s_1} \int_{p_2 s_2}^{\bar{b}} (b_2 - h_2) f_2(b_2) db_2 f_1(b_1) db_1 \\ &+ \int_{p_1 s_1}^{\bar{b}} \int_{b_1 - p_1 s_1 + p_2 s_2}^{\bar{b}} (b_2 - h_2) f_2(b_2) db_2 f_1(b_1) db_1 - (e_1 + e_2). \end{aligned} \qquad (2)$$

The first term is associated with those who commit act 1 and the second and third terms with those who commit act 2. We have

[8] Claims made in these and other notes that are not clear are justified in the appendix.

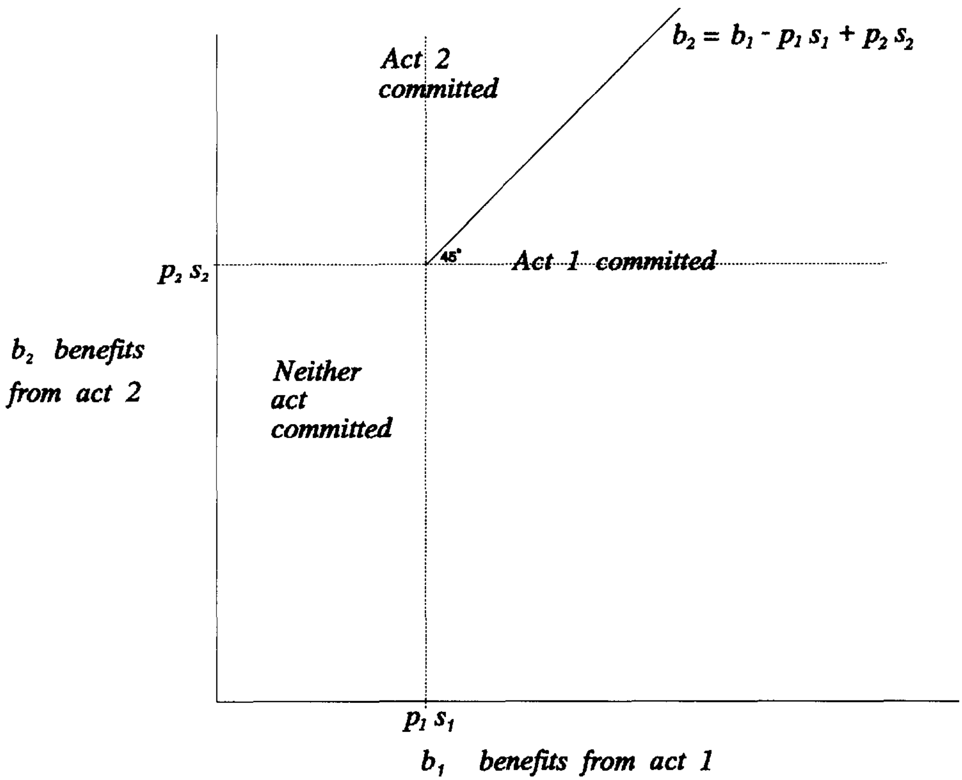

FIG. 1.

Proposition 2. In the two-act model with specific enforcement, (a) optimal sanctions for the two acts are the same and equal to the maximal sanction, wealth. (b) The expected sanction for each act is less than the harm it does. (c) The optimal probabilities of apprehension for the acts are generally different (and are determined by conditions (A2) and (A3) in the appendix).

Notes. (1) Again, it is possible that $p_1^* w > p_2^* w$, and a sufficient condition for $p_1^* w < p_2^* w$ is that the functions p_i are equal and that the densities f_i are equal.

(2) The conditions determining the p_i^* reflect marginal deterrence.

(3) $p_i^* w < h_i$ for essentially the reason applying in the one-act model, to save enforcement effort.

Comparison of the one-act and two-act models. Optimal enforcement is similar in the models. In both models, optimal sanctions are maximal for acts 1 and 2, and in both models the optimal expected sanction for each act is less than harm. The only difference is that determination of the optimal p_i in the two-act model implicitly involves considerations of marginal deterrence.[9]

[9]However, there does not appear to be any simple relationship between the optimal p_i in the one-act and two-act models.

B. General Enforcement

If enforcement is *general*, let

e = enforcement effort devoted to apprehending those who commit either act;
$p(e)$ = probability of apprehension; $p'(e) > 0$.

One-act model. In the one-act model, an individual will commit act i if and only if $b_i \geqslant ps_i$, so that social welfare is

$$.5\int_{ps_1}^{\bar{b}}(b_1 - h_1)f_1(b_1)db_1 + .5\int_{ps_2}^{\bar{b}}(b_2 - h_2)f_2(b_2)db_2 - e. \qquad (3)$$

The following is true, assuming that e^* is positive.

Proposition 3. In the one-act model with general enforcement, (a) the optimal sanction for the less harmful act is $s_1^* = h_1/p$, *so that the expected sanction equals the harm* h_1 *(unless* h_1/p^* *exceeds wealth, in which case* s_1^* *equals wealth). (b) The optimal sanction for the more harmful act equals wealth, and the expected sanction is less than the harm it causes. (c) The optimal probability of apprehension is determined by a condition given in the appendix (see (A5) and (A6)).*

Note. $p^*w < h_2$ in order to save enforcement effort, as explained in note (2) to Proposition 1. Because p^* is general, it cannot be lowered specifically for act 1; thus $p^*s_1 = h_1$ may well be optimal. Equivalently, were p^* such that $p^*w < h_1$, then p^*w might be so much less than h_2 as to cause a serious problem of underdeterrence of act 2.

Two-act model. By analogy to (2), it is evident that social welfare is

$$\begin{aligned}&\int_{ps_1}^{\bar{b}}\int_0^{b_1 - ps_1 + ps_2}(b_1 - h_1)f_2(b_2)db_2 f_1(b_1)db_1 \\ &+ \int_0^{ps_1}\int_{ps_2}^{\bar{b}}(b_2 - h_2)f_2(b_2)db_2 f_1(b_1)db_1 \\ &+ \int_{ps_1}^{\bar{b}}\int_{b_1 - ps_1 + ps_2}^{\bar{b}}(b_2 - h_2)f_2(b_2)db_2 f_1(b_1)db_1 - e.\end{aligned} \qquad (4)$$

Assuming that e^* is positive, we have

Proposition 4. In the two-act model with general enforcement, (a) the optimal sanction for the less harmful act is such that the expected sanction is less than the harm caused by the act. (b) The sanction for the more harmful act equals wealth, and the expected sanction is less than the harm due to the act. (c) The optimal probability of apprehension is determined by a condition given in the appendix (see (A7)).

Note. The reason that $p^*s_1^* < h_1$ reflects marginal deterrence. Specifically, assume that $p^*s_1 = h_1$ and consider the two effects of lowering s_1 slightly. The first effect has to do with marginal deterrence: some individuals who were just willing to commit act 2 will now just prefer to commit act 1. This will raise social welfare. An individual who was just willing to commit act 2 is someone for whom b_2 is approximately equal to p^*w. But p^*w we know is less than h_2 (in order to save enforcement costs), implying that the individual would reduce social welfare by the positive amount $h_2 - p^*w$ if he commits act 2. If the individual is now just willing to commit act 1, however, his benefit b_1 must be approximately h_1, so that he will not reduce social welfare if he commits act 1. Hence, by inducing individuals to commit act 1 rather than act 2, a loss in social welfare is avoided. The second effect of lowering s_1 slightly is that some individuals who would not have committed any act may now commit act 1. But this causes no reduction in social welfare since the benefits of the individuals must be approximately equal to h_1.

Comparison of the one-act and two-act models. In both models, the sanction for act 1 is lower than that for act 2, which is wealth (except in the case where the constraint $s_1 \leq w$ is binding). However, $p^*s_1^* < h_1$ in the two-act model, whereas in the one-act model $p^*s_1^* = h_1$,suggesting a tendency for s_1 to be lower in the two-act model. This is only a tendency because the optimal probabilities p^* are generally different in the two models.

III. Concluding Remarks

(a) A point that is made in the proofs to propositions 3 and 4 (see step (i)) deserves emphasis. Namely, it is observed that if the expected sanction can be set equal to harm for each act (that is, if the wealth constraint is not binding), then first-best behavior results in both the one-act model and the two-act model. In other words, when expected sanctions equal harm, not only do individuals decide correctly whether or not to commit single harmful acts—the usual type of deterrence is optimal—so do undeterred individuals decide optimally which harmful acts to commit—marginal deterrence also is optimal. This point is relevant in contexts where expected sanctions are, in fact, approximately equal to harm, or could be (because the probability of sanctions is high and the wealth of most individuals exceeds the needed sanctions).

(b) In the model with general enforcement effort, recall that it was assumed that the probability of apprehension was the same for each act. More realistically, the probability of apprehension, though determined by general enforcement effort, may vary with the act. If the probability of apprehension were to fall with the harmfulness of acts, the results obtained in both the one-act and two-act models that sanctions ought to rise with harm would be reinforced, but if the probability of apprehension were to rise with harm, the results might be reversed.

(c) Marginal deterrence is of relevance in two types of situation that seem worth distinguishing. The first is typified by the examples mentioned in the Introduction, where a person chooses whether to discharge a pollutant into a lake or onto the ground.

The other type of situation is where a person chooses whether to increase the harm he does when committing one harmful act by committing an additional harmful act, the classic example being the person who kidnaps and then decides whether to kill his hostage. In such a situation, consideration of marginal deterrence does *not* imply that the sanction for the more harmful act, murder, should exceed that for the less, kidnapping. To accomplish marginal deterrence, all that is necessary is that sanctions for committing multiple harmful acts be *cumulative*. As long as there is a sanction for murder that is added to the sanction for kidnapping, there will be a reason for the kidnapper not to commit the additional crime of murder; this will be true whether or not the sanction for murder alone is higher than that for kidnapping alone.

(d) The main results obtained here appear to apply where there is a continuum of harmful acts (the quantity of a pollutant that is discharged) and individuals in the multiple-act model may choose any act in the continuum. In such a case, if enforcement is specific, the simple argument given above implies that the optimal sanction for each act will be maximal (with enforcement effort varying among acts), and if enforcement is general, optimal sanctions will vary with acts (the schedule of sanctions being the solution to an optimal control theory problem).

(e) In the case where sanctions are non-monetary, I have not succeeded in obtaining an appealing characterization of the difference between the one-act and the two-act models, although one supposes that in some general sense the results should be similar to those discussed here.

References

Beccaria, Cesare. (1770). *An Essay on Crimes and Punishments.* Rept. ed. Albany: W.C. Little, 1872.

Becker, Gary. (1968). "Crime and Punishment: An Economic Approach." *Journal of Political Economy* **76:**169–217.

Bentham, Jeremy. (1789). *An Introduction to the Principles of Morals and Legislation,* in *The Utilitarians.* Rept. Garden City, N.Y.: Anchor Books, 1973.

Montesquieu, Charles. (1748). *The Spirit of Laws.* Rept. ed. Berkeley: University of California Press, 1977.

Reinganum, Jennifer, and Louis Wilde. (1986). "Nondeterrables and Marginal Deterrence Cannot Explain Nontrivial Sanctions," California Institute of Technology.

Shavell, Steven. (1989). "A Note on Optimal Deterrence When Individuals Choose among Harmful Acts," Discussion Paper No. 57, Program in Law and Economics, Harvard Law School.

Shavell, Steven. (1991). "Specific versus General Enforcement of Law." *Journal of Political Economy* **99:** 1088–1108.

Stigler, George. (1970). "The Optimum Enforcement of Laws." *Journal of Political Economy* **78:** 526–36.

Wilde, Louis. (1989). "Criminal Choice, Nonmonetary Sanctions and Marginal Deterrence: A Normative Analysis," California Institute of Technology.

Appendix

Proof of Proposition 1. If $s_i^* < w$, raise s_i to w and reduce e_i so that $p_i w = p_i^* s_i^*$. Hence, from (1), it is clear that the behavior of those who might commit act i will not be affected, but since e_i is lower, (1) is higher, contradicting the optimality of s_i^*. Thus, $s_i^* = w$.

From (1) it is clear that the first-order condition determining e_i^* is

$$-.5p_i'(e_i)wf_i(p_i w)(p_i w - h_i) = 1. \tag{A1}$$

It is evident from (A1) that in general e_1^* will be unequal to e_2^*, and that $p_i^* w < h_i$.

Notes to Proposition 1. With regard to the claim about the possibility that $p_1^* w > p_2^* w$, suppose, for example, that f_1 is very high in an interval $[k_1,h_1]$; that f_2 is very low in an interval $[k_2,h_2]$; and that $k_1 > k_2 > 0$. Then (A1) will be satisfied for $p_1 w$ in $[k_1,h_1]$, but (A1) cannot be satisfied for $p_2 w$ in $[k_2 w,h_2]$, so that $p_2^* w < p_1^* w$.

With regard to the other claim of Note (1), observe that (A1) is of the form $g(e_i,h_i) = 0$, which implicitly determines e_i as a function of h_i. Differentiating this with respect to h_i, one obtains $e_i'(h_i) = -g_2(e_i,h_i)/g_1(e_i,h_i)$. But the denominator is negative (the second order condition for e_i to be an optimum), so that sign $e_i' =$ sign $g_2 = .5p_i'wf_i(p_iw) > 0$. Hence, e_i is increasing in h_i. From this the claim follows.

Proof of Proposition 2. The argument in the previous proof shows that $s_i^* = w$. Using this fact and differentiating (2) with respect to the e_i, one obtains the first-order conditions

$$-p_1'(e_1)wf_1(p_1w)(p_1w - h_1)F_2(p_2w) + p_1'(e_1)w\int_{p_1w}^{\bar{b}}[(p_2w - h_2) - (p_1w - h_1)]f_2(b_1 - p_1w + p_2w)f_1(b_1)db_1 = 1 \tag{A2}$$

and

$$-p_2'(e_2)wf_2(p_2w)(p_2w - h_2)F_1(p_1w) + p_2'(e_2)w\int_{p_1w}^{\bar{b}}[(p_1w - h_1) - (p_2w - h_2)]f_2(b_1 - p_1w + p_2w)f_1(b_1)db_1 = 1, \tag{A3}$$

where the F_i are cumulative distribution functions. The e_1 and e_2 satisfying (A2) and (A3) will generally be different.

Assume that $p_1w \geq h_1$. Then since the first term in (A2) is non-positive, the second term must be positive, which implies that $p_2w - h_2 > p_1w - h_1$, so that $p_2w > h_2$. However, $p_2w > h_2$ means that the first term in (A3) is negative, so that the second term in (A3) is positive, which implies that $p_1w - h_1 > p_2w - h_2$. This is a contradiction. A symmetric argument shows that $p_2w \geq h_2$ leads to a contradiction. Hence, $p_iw < h_i$, as claimed.

Notes to Proposition 2. With regard to the claim of Note (1), observe first that if $p_1w > p_2w$, social welfare could be increased by reversing the e_i.[10] Hence, it must be that $p_1w \leq p_2w$. If $p_1w = p_2w$, however, then examination of (A2) and (A3) leads to a contradiction.[11]

With regard to Note (2), observe that (A2) and (A3) reflect considerations of marginal deterrence; the second term in each is associated with the effect of undeterred individuals switching from act 1 to 2 or from 2 to 1 as e_1 or e_2 is raised.

Proof of Proposition 3. The argument consists of several steps.

(i) Given any positive p, if s_i satisfying $ps_i = h_i$ is feasible, that is, if $h_i/p \leq w$, then $s_i = h_i/p$ is optimal; otherwise, $s_i = w$ is optimal: If $ps_i = h_i$, first-best behavior results, so $s_i = h_i/p$ is optimal if it is feasible. Otherwise, $s_i = w$ is optimal, as it will deter the greatest number of individuals who ought to be deterred from committing act i.[12]

[10] If, initially, $e_1 = a > b = e_2$, for some positive a and b, so that $p_1w = p(a)w > p_2w = p(b)w$, set $e_1 = b$ and $e_2 = a$, so that $p_1w = p(b)w < p_2w = p(a)w$. It is easy to verify (I omit details) that, since the densities f_i are equal and independently distributed, the total benefits derived by the set of individuals who commit acts are equal in the two situations. However, in the second situation more individuals commit act 1 and fewer act 2 than in the first situation. Thus, less harm is done in the second situation. Total enforcement effort is $a + b$ in both situations. Therefore, social welfare is higher in the second situation.

[11] If $p_1w = p_2w$ and the $f_i = f$ and the $p_i = p$, (A2) becomes $-p'wf(pw)(pw - h_1)F(pw) + p'w\int(h_1 - h_2)f(b_1)^2 db_1 = 1$ and (A3) becomes $-p'wf(pw)(pw - h_2)F(pw) - p'w\int(h_1 - h_2)f(b_1)^2db_1 = 1$. These are two equations of the form $a(pw - h_1) + b = 1$ and $a(pw - h_2) - b = 1$, where a is unequal to zero. Solving each for $1 - b$, we deduce that $pw - h_1 = -pw + h_2$, or that $pw = (h_1 + h_2)/2 > h_1$. This contradicts $pw < h_1$, which was shown in the proposition.

[12] The claim of this paragraph can be verified as well from differentiation of (3). The derivative of (3) with respect to s_1 (the derivative with respect to s_2 is analogous) is $-.5p(ps_1 - h_1)f_1(ps_1)$; this is zero if $ps_1 = h_1$ and is positive for s_1 such that $ps_1 < h_1$.

(ii) $s_2^* = w$: If $s_2^* < w$, then, by (i), $s_2^* = h_2/p$ and, since $h_1/p < h_2/p$, $s_1^* = h_1/p$. Hence, e and p can be lowered slightly and s_1 and s_2 both raised so that $ps_i = h_i$ still holds. Thus, the behavior of individuals will be unchanged, yet e will be lower, contradicting the optimality of s_2^*.

(iii) $p^*w < h_2$: If $p^*w > h_2$, then (i) implies $s_2^* = h_2/p^*$; but this means $s_2^* < w$, contradicting (ii). Hence, $p^*w \leqslant h_2$. Now the derivative of (3) with respect to e is

$$-.5p'(e)[s_1(ps_1 - h_1)f_1(ps_1) + s_2(ps_2 - h_2)f_2(ps_2)] - 1. \tag{A4}$$

If $pw = h_2$, then (i) implies that $ps_1 = h_1$, so (A4) reduces to -1, meaning that welfare can be raised by lowering e and p. Hence, p^*w must be less than h_2.

(iv) p^* is determined by

$$-.5p'(e)s_2(pw - h_2)f_2(ps_2) = 1 \tag{A5}$$

if $s_1^* < w$; and p^* is determined by

$$-.5p'(e)w(pw - h_1)f_1(pw) + w(pw - h_2)f_2(pw)] = 1 \tag{A6}$$

if $s_1^* = w$: This is clear from what was shown about the s_i^* and from substitution in (A4).

Proof of Proposition 4: The argument again consists of a series of steps.

(i) Given any positive p, if s_2 such that $ps_2 = h_2$ is feasible, that is, if $pw \geqslant h_2$, then $s_2^* = h_2/p$ and $s_1^* = h_1/p$: Under the assumption, $ps_i = h_i$, so that first-best behavior results. Hence, the s_i must be optimal.[13]

(ii) $p^*s_1^* \leqslant h_1$: If $p^*s_1^* > h_1$, then $p^*w < h_2$; for otherwise, by (i), $p^*s_1^* = h_1$, a contradiction. Thus, assume $p^*w < h_2$ and reduce s_1 so that $p^*s_1 = h_1$. Two changes in behavior occur. First, some individuals who had committed neither act are led to commit act 1. This raises social welfare, since an individual who commits act 1 must be one for whom $b_1 \geqslant h_1$. Second, some individuals who had committed act 2 commit act 1. This also raises social welfare. For if an individual chooses act 1 over 2, then $b_1 - h_1 > b_2 - p^*s_2$; but since $p^*w < h_2$, we know that $p^*s_2 < h_2$, so that $b_2 - p^*s_2 > b_2 - h_2$. Hence, the choice of act 1 indeed raises social welfare, a contradiction.

(iii) $s_2^* = w$: Let us show that if $s_2^* < w$, we are led to a contradiction in each of two possible cases: when $p^*w \geqslant h_2$, and when $p^*w < h_2$.

If $p^*w \geqslant h_2$, then by (i), $p^*s_i^* = h_i$. Raise s_2 to w and reduce e and p so that $pw = h_2$. With this p, raise s_1 also so that $ps_1 = h_1$. (This is possible, since $h_1 < h_2$.) Then behavior will not have changed, yet e is lower, so that welfare is higher, a contradiction.

If $p^*w < h_2$, raise s_2 to w and raise s_1 to the minimum of $s_1^* + (w - s_2^*)$, h_1/p^*, and w. (Since, by (ii), $s_1^* \leqslant h_1/p^*$, we know that s_1 is indeed less than or equal to the new s_1.) Then social welfare will increase. There are three possible types of change in behavior. First, an individual who had committed act 2 may decide not to commit either act. This must raise welfare, since for such an individual, $b_2 < p^*w < h_2$. Second, an individual who had committed act 1 may decide not to commit either act; this too must raise social welfare since for such an individual $b_1 < p^*s_1 \leqslant h_1$. Third, an individual who had committed act 2 may instead commit act 1. (This is possible since s_1 is raised by an amount less than or equal to $w - s_2^*$; and for that reason, no one would switch from act 1 to act 2.) For such an individual, $b_2 - p^*w \leqslant b_1 - s_1$, but $b_2 - h_2 <$

[13]The reader may also verify that the first-order conditions obtained by differentiating (4) with respect to the s_i are satisfied when $ps_i = h_i$.

$b_2 - p^*w$, since $p^*w < h_2$, and $b_1 - s_1 \leq b_1 - h_1$, by definition of s_1. Hence $b_2 - h_2 < b_1 - h_1$, meaning that social welfare is raised by the switch to act 1.

(iv) $p^*w < h_2$: If $p^*w > h_2$, then, by (i), $s_2^* = h_2/p^* < w$, which contradicts (iii). Hence, $p^*w \leq h_2$. Now the derivative of (4) with respect to e is (after cancellation) seen to be

$$p'(e)\{-s_1 f_1(ps_1)(ps_1 - h_1)F_2(pw) + (w - s_1)\int_{ps_1}^{\bar{b}}(ps_1 - h_1 - pw + h_2)f_2(b_1 - ps_1 + pw)f_1(b_1)db_1 - wf_2(pw)(pw - h_2)F_1(ps_1)\} - 1. \tag{A7}$$

If $p^*w = h_2$, (i) implies that $p^*s_1^* = h_1$, so that (A7) reduces to $-1 < 0$, and thus (4) is increased by lowering e. Hence, $p^*w < h_2$ must be true.

(v) $p^*s_1^* < h_1$: If $p^*w < h_1$, the claim is trivially true. Otherwise, by (ii), we need only rule out the possibility that $p^*s_1 = h_1$. The derivative of (4) with respect to s_1 is

$$-pf_1(ps_1)(ps_1 - h_1)F_2(pw) + p\int_{ps_1}^{\bar{b}}[(pw - h_2) - (ps_1 - h_1)]f_2(b_1 - ps_2 + pw)f_1(b_1)db_1. \tag{A8}$$

If $p^*s_1 = h_1$, then (A8) is negative, since, by (iv), $p^*w < h_2$. Hence, it must be beneficial to lower s_1, and $p^*s_1^* < h_1$ must hold.

(vi) p^* is determined by the condition that exp. (A7) $= 0$: This is evident, since (A7) is the derivative of (4) with respect to e.