

## **Deterrence and the Adjustment of Sentences During Imprisonment**

A. Mitchell Polinsky and Steven Shavell\*

*Abstract*: The prison time actually served by a convicted criminal depends to a significant degree on decisions made by the state during the course of imprisonment—notably, on whether to grant parole. We study a model of the adjustment of sentences assuming that the state’s objective is the optimal deterrence of crime. In the model, the state can lower or raise a criminal’s initial sentence on the basis of deterrence-relevant information obtained during imprisonment. Our focus on sentence adjustment as a means of promoting deterrence stands in contrast to the usual emphasis in sentence adjustment policy on avoiding recidivism.

*Key words*: Deterrence; imprisonment; sentence adjustment; sanctions; parole; recidivism

---

\* Stanford Law School and National Bureau of Economic Research; and Harvard Law School and National Bureau of Economic Research, respectively. We thank the John M. Olin Program in Law and Economics at Stanford Law School and the John M. Olin Center for Law, Economics, and Business at Harvard Law School for research support; Albert Choi, Nuno Garoupa, Oliver Hart, Louis Kaplow, Thomas Miceli, Murat Mungan, Kathryn Spier, Abraham Wickelgren, Diego Zambrano, and anonymous referees for comments; and Paul Riskind and Martin Sicilian for research assistance. [Email contacts: <polinsky@stanford.edu> and <shavell@law.harvard.edu>.]

## 1. Introduction

When an individual is convicted of a crime, a court's sentencing decision will not fully determine the length of his imprisonment. The time actually served by an offender will depend to a significant degree on decisions made by the state after he is incarcerated. Notably, prisoners are often granted parole<sup>1</sup> and also can benefit from reductions of sentences on account of good time and earned time credits.<sup>2</sup> In fact, the sentences of individuals held in state prisons are estimated to be lowered by 54 percent as a result of these adjustments.<sup>3</sup>

The contribution of this article is to study the general practice of altering sentences during imprisonment within a model of deterrence. Specifically, we examine how information about prisoners obtained during imprisonment can be employed to modify sentences so as to optimally deter crime.<sup>4</sup> Such information could concern, for example, a prisoner's gain from crime (say the extent of his money laundering as revealed by a ratting cell mate) or the disutility that he

---

<sup>1</sup> Nearly 80 percent of state prisoners are eventually released to parole supervision. See "Reentry Trends in the U.S. Releases from State Prison," Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice (available at <[https://www.bjs.gov/content/reentry/reentry\\_contents.cfm](https://www.bjs.gov/content/reentry/reentry_contents.cfm)>; page last revised July 8, 2020). As a result of the Sentencing Reform Act of 1984, federal prisoners are not now eligible for parole.

<sup>2</sup> Good time credits are awarded for obeying prison rules and earned time credits are granted for participating in self-improvement activities, such as vocational training and drug treatment. See generally National Conference of State Legislatures, "Good Time and Earned Time Policies for State Prison Inmates," updated January 2016 (available at <[https://docs.legis.wisconsin.gov/misc/lc/study/10\\_00\\_a\\_m\\_room\\_412\\_east\\_state\\_capitol/memono4g](https://docs.legis.wisconsin.gov/misc/lc/study/10_00_a_m_room_412_east_state_capitol/memono4g)>). See also LaFave et al. (2019, § 26.2(c)).

<sup>3</sup> See Kaeble (2018, p. 4, Table 3). This article from the Bureau of Justice Statistics reports that state prisoners serve 45.5 percent of their maximum sentence length before their first release. Federal prisoners, who constitute 13 percent of all prisoners, can benefit from good time credits and other sentence reductions, which lower their sentences on average by 12 percent. See Carson (2018, p. 3) and Motivans (2015, p. 39, Table 7.11).

<sup>4</sup> Although, to our knowledge, this issue has not been studied previously, several authors address related questions. Miceli (1994), Garoupa (1996), and Polinsky (2015) analyze how the modification of sentences during imprisonment can induce good conduct *within* prison. Pyne (2015) discusses a mechanism that leads prisoners to reveal their disutility from imprisonment, which can be used to foster deterrence. Also, Bernhardt, Mongrain, and Roberts (2012) and Kuziemko (2013), among many others, investigate the use of sentence adjustment to limit recidivism.

experiences from incarceration. It could also include whether a prisoner has a history of recalcitrance (suggesting that he is more difficult to deter) or whether he is temperate in character and follows prison rules (indicating that he is easier to deter).<sup>5</sup>

We begin in Section 2 by determining the optimal prison sentence in the standard model of crime, in which sentences are imposed at the time of conviction and are not adjusted afterwards.

In Sections 3 and 4 we analyze what we refer to as the *sentence adjustment model*, in which the state might modify a prisoner's sentence in the light of information obtained during his incarceration bearing on his incentive to commit crime. For concreteness, we treat a prisoner's gain from committing a crime as a stand-in for this type of information.

In Section 3 we assume that the state learns a prisoner's gain with a probability. We observe there that if the state obtains this information, the sentence should be lowered from its initial level to zero when the prisoner could not have been deterred, whereas the sentence should be set at a level sufficiently high to accomplish deterrence when he can be deterred.<sup>6</sup> We also note that social welfare in this version of the sentence adjustment model is greater than in the standard model.

In Section 4 we consider the sentence adjustment model when the information the state might obtain about a prisoner's gain is imperfect. In particular, we assume that the population of potential offenders is divided into distinct groups and that the state might observe a prisoner's group and thus learn the relevant conditional probability distribution of gains for that prisoner.

---

<sup>5</sup> For elaboration, see comment (d) in Section 5 below.

<sup>6</sup> Of course, because this policy will result in individuals who can be deterred choosing not to commit a crime, they will not be sentenced in fact.

We assess the social value of this information and characterize the optimal initial sentence and the optimal adjusted sentence if the state observes a prisoner's group.

To elaborate, if the probability of learning a prisoner's group is low, information obtained by the state about the prisoner might not have social value. But such information will have social value if the probability exceeds a threshold. The value of information then increases up to a second threshold probability at which the second-best optimum can be achieved.<sup>7</sup> Once this probability is reached, the optimal initial sentence and the optimal adjusted sentences are determined by simple formulas. Moreover, the groups can be divided into subsets, one in which it is optimal to lower the initial sentence when the state learns that a prisoner is a member of a group in this subset, and another in which it is optimal to raise the initial sentence.

In Section 5 we conclude with several comments on the applicability of the model, including whether the actual practice of downward-only sentence adjustment is socially desirable and whether potential offenders take sentence adjustment into account. We also remark on the failure of prison authorities and scholars to consider the role of deterrence in sentence adjustment policy. Instead, their virtually exclusive concern is with how the early release of prisoners affects recidivism. Yet the early release of prisoners should also affect deterrence.

## **2. The Standard Model**

We first consider the standard model of deterrence, in which criminals are sentenced before imprisonment and sentences are not adjusted afterwards.<sup>8</sup> In this model individuals

---

<sup>7</sup> The second-best optimum is the outcome that would be achievable if the state could observe each prisoner's group with certainty.

<sup>8</sup> See Becker (1968) and Polinsky and Shavell (2000).

choose whether to commit a harmful act and differ in the gains that they would obtain if they do so. Let

$h$  = harm caused by the act;  $h > 0$ ;

$g$  = gain that an individual would obtain from the act; and

$f(g)$  = probability density of  $g$  in  $[0, \bar{g}]$ ;  $f(g) > 0$ ;  $\bar{g} > 0$ ,

where  $f$  is assumed to be continuous and differentiable and  $F$  is the cumulative distribution function of  $f$ .<sup>9</sup> We assume that  $\bar{g} < h$ , which will mean that all acts are socially undesirable to commit. We also assume that the state cannot observe  $g$  but knows its probability density.

Individuals who commit the act are caught with a probability and sentenced to a term of imprisonment. Imprisonment imposes disutility on the criminal and results in the state bearing costs associated with the operation of prisons. Let

$p$  = probability that an individual who commits the harmful act is caught;  $p > 0$ ;

$s$  = prison sentence;  $s$  is in  $[0, \bar{s}]$ ;  $\bar{s} > 0$ ; and

$k$  = cost to the state per unit time of imprisonment;  $k > 0$ .

We treat  $p$  as fixed in order to focus on optimal sentencing policy. The sentence  $s$  is assumed to be equivalent to its disutility to a criminal. Hence, an individual will commit the harmful act if  $g > ps$  and otherwise will be deterred.<sup>10</sup> We assume that  $\bar{s} < \bar{g}$  to guarantee that it is not possible to deter all individuals from committing the harmful act; the case of full deterrence would not be of interest.

---

<sup>9</sup> The function  $f$  should be interpreted as the density of gains conditional on the information available to the state at the time of sentencing.

<sup>10</sup> We assume for convenience that the individual is deterred if  $g = ps$ .

Social welfare is the sum of individuals' gains from committing the act, less the harm caused, less the disutility of imprisonment, and less the cost to the state of imprisonment.<sup>11</sup>

$$W(s) = \int_{ps}^{\bar{g}} [g - h - ps(1 + k)]f(g)dg. \quad (1)$$

In other words, each individual who commits the act changes social welfare by the amount  $g - h - ps - psk = g - h - ps(1 + k)$ . The state's problem is to choose the sentence  $s$  to maximize social welfare (1). The solution,  $s^*$ , is assumed to be unique and can be characterized as follows.

*Proposition 1.* In the standard model of deterrence,

- (a) the optimal sentence  $s^*$  may be at either endpoint or in the interior of  $[0, \bar{s}]$ ;
- (b) if  $0 < s^* < \bar{s}$ , then  $s^*$  must satisfy the first-order condition (5); and
- (c) if  $s^* > 0$ , then a positive number of individuals will be imprisoned.

*Note.* The optimal sentence  $s^*$  reflects a tradeoff between the marginal social benefit of deterring more individuals as  $s$  is raised and the marginal social cost of imposing longer sentences on individuals who are not deterred. The optimal sentence might be zero because the marginal social cost of imprisonment might always outweigh the marginal social benefit of deterrence, or the optimal sentence might be maximal for the opposite reason. When  $s^*$  is positive, some individuals will always be imprisoned because a positive fraction of individuals are undeterrable (those in  $(ps^*, \bar{g}]$ ).

*Proof.* Part (a): The derivative of social welfare is

$$W'(s) = -p[ps - h - ps(1 + k)]f(ps) - \int_{ps}^{\bar{g}} p(1 + k)f(g)dg \quad (2)$$

---

<sup>11</sup> This measure of social welfare does not reflect the cost to the state of maintaining the probability of detection  $p$ . Were we to include that cost, it would be a constant term in social welfare.

$$= p[(h + psk)f(ps) - (1 - F(ps))(1 + k)],$$

where the first term is the marginal benefit from deterring individuals of type  $ps$  and the second term is the marginal cost from sentencing undeterred individuals for an additional unit of time.

We first show that  $s^*$  can be 0 by demonstrating that  $W'(s)$  can be negative for all  $s$  in  $[0, \bar{s}]$ . We do this by providing an example in which there is an upper bound on the first term in brackets in the second line of (2) that is strictly less than a lower bound of the second term in brackets. In particular, suppose that  $f(g)$  is uniform at height  $\hat{f} > 0$  between 0 and  $p\bar{s}$ .<sup>12</sup> Then  $(h + p\bar{s}k)\hat{f}$  is an upper bound on the first term. Clearly,  $(1 - F(p\bar{s}))(1 + k)$  is a lower bound for the second term. Thus,  $W'(s)$  will be negative for  $s$  in  $[0, \bar{s}]$  if

$$(h + p\bar{s}k)\hat{f} < (1 - F(p\bar{s}))(1 + k); \quad (3)$$

but this will be true for a sufficiently low  $\hat{f}$ .

To demonstrate that  $s^*$  can be an interior point, we claim that  $W'(0) > 0$  and  $W'(\bar{s}) < 0$  can both hold. We know that  $W'(0) = p[hf(0) - (1 + k)]$ , which will be positive if  $f(0)$  is sufficiently high. Similarly,  $W'(\bar{s}) = p\{[h + p\bar{s}k]f(p\bar{s}) - (1 - F(p\bar{s}))(1 + k)\}$ , which will be negative if  $f(p\bar{s})$  is sufficiently low.

To prove that  $s^*$  can be  $\bar{s}$ , we show that  $W'(s)$  can be positive for all  $s$  in  $[0, \bar{s}]$ . Let  $f(g)$  be uniform over  $[0, \bar{g}]$ , and thus  $f(g) = 1/\bar{g}$ . Then  $h(1/\bar{g})$  is a lower bound for the first term in brackets of the second line of (2). Clearly,  $(1 + k)$  is an upper bound for the second term in brackets. Thus,  $W'(s)$  will be positive for  $s$  in  $[0, \bar{s}]$  if

---

<sup>12</sup> Our argument does not depend on  $f(g)$  for  $g > p\bar{s}$ .

$$h(1/\bar{g}) > (1 + k), \quad (4)$$

which will hold for a sufficiently high  $h$ .

Part (b): If  $s^*$  is an interior solution, it must satisfy the first-order condition from (2), namely

$$[h + psk]f(ps) = (1 - F(ps))(1 + k). \quad (5)$$

Part (c): Individuals who will not be deterred include those for whom  $g$  is in  $(ps^*, \bar{g}]$ . This group has positive mass since  $ps^* < \bar{g}$ . The fraction  $p$  of these individuals will be imprisoned for the positive length of time  $s^*$ .  $\square$

We will denote the maximum level of social welfare in the standard model,  $W(s^*)$ , by  $W^*$ .

### 3. The Sentence Adjustment Model When the State Might Obtain Perfect Information about Criminals During Imprisonment

We now modify the standard model by assuming that after a criminal is initially sentenced and imprisoned, the state will learn what his gain  $g$  was with some probability. In that event, the state can change his sentence. Let

$s_o$  = initial sentence;  $s_o$  is in  $[0, \bar{s}]$ ;

$q$  = probability that the state learns  $g$  when a criminal is in prison;  $q$  is in  $[0, 1]$ ;<sup>13</sup> and

$s(g)$  = adjusted sentence if the state learns  $g$ ;  $s(g)$  is in  $[0, \bar{s}]$ .

A person of type  $g$  will commit the harmful act when

$$g > p[(1 - q)s_o + qs(g)], \quad (6)$$

---

<sup>13</sup> For simplicity, we assume that  $q$  does not depend on the length of the initial sentence. If we were to assume that  $q$  rises with  $s_o$  and that  $q = 0$  if  $s_o = 0$ , the essential character of our results would not be affected.



for if the person is caught and sentenced to  $s_o$ , the probability is  $(1 - q)$  that the state will not observe  $g$  and thus that the sentence will remain  $s_o$  and the probability is  $q$  that the state will observe  $g$  and alter the sentence to  $s(g)$ .

The state's problem is to choose  $s_o$  and the function  $s(\cdot)$  to maximize social welfare, which is

$$W(s_o, s(\cdot)) = \int_0^{\bar{g}} \frac{[g - h - p[(1 - q)s_o + qs(g)](1 + k)]f(g)dg}{p[(1 - q)s_o + qs(g)]} \quad (7)$$

The solution is as follows.

*Proposition 2.* In the sentence adjustment model when the state obtains perfect information about criminals during imprisonment with probability  $q$ ,

- (a) the optimal initial sentence  $s_o^*$  may be at either endpoint or in the interior of  $[0, \bar{s}]$ ;
- (b) if  $0 < s_o^* < \bar{s}$ , then  $s_o^*$  must satisfy first-order condition (12);
- (c) the optimal adjusted sentence  $s^*(g)$  accomplishes deterrence for all types of individuals  $g$  who can be deterred, including a low range of  $g$  over which  $s^*(g)$  is zero; and  $s^*(g)$  is zero for all types  $g$  who cannot be deterred:  $s^*(g) = 0$  for  $g$  in  $[0, p(1 - q)s_o]$ ;  $s^*(g) = [(g/p) - (1 - q)s_o]/q$  and rises from 0 to  $\bar{s}$  for  $g$  in  $(p(1 - q)s_o, p[(1 - q)s_o + q\bar{s}]]$ ; and  $s^*(g) = 0$  for  $g$  in  $(p[(1 - q)s_o + q\bar{s}], \bar{g}]$ ;
- (d) if  $s_o^* > 0$ , then a positive number of individuals will be imprisoned; and
- (e) some individuals will be deterred in the optimal solution.

*Notes.* (i) It is clear that if  $s(g)$  can be chosen so as to deter individuals of any type  $g$ , it will be desirable to do so because the harmful act is socially undesirable. (For low values of  $g$ , deterrence does not require  $s(g)$  to be positive since the initial sentence  $s_o$  is imposed with a positive probability.) Conversely, if deterrence is not possible,  $s(g)$  should be zero because

nothing would be accomplished by imposing a costly sentence.<sup>14</sup> This explains part (c). It also explains part (e) because it will always be possible to deter some individuals by employing a positive  $s^*(g)$ .

(ii) The explanations of parts (a), (b), and (d) regarding  $s_o^*$  parallel the explanation of the results regarding  $s^*$  in Proposition 1.

*Proof.* Part (c): If an individual is deterred, there will be no change in social welfare, whereas if he is not deterred, the change in social welfare will be  $g - h - p[(1 - q)s_o + qs(g)](1 + k)$ , which is negative. Therefore, it is desirable to deter any individual who can be deterred. We know from (6) that a person can be deterred if and only if  $g \leq p[(1 - q)s_o + q\bar{s}]$ . It follows that for individuals with  $g$  in  $[0, p(1 - q)s_o]$ , deterrence will be achieved if  $s(g) = 0$ . For individuals with  $g$  in  $(p(1 - q)s_o, p[(1 - q)s_o + q\bar{s}]]$ , it can be verified from (6) that deterrence will just be achieved if  $s(g) = [(g/p) - (1 - q)s_o]/q$ ; in this interval,  $s(g)$  rises from 0 to  $\bar{s}$ . For individuals with  $g$  in  $(p[(1 - q)s_o + q\bar{s}], \bar{g}]$ , who cannot be deterred, the change in social welfare is  $g - h - p[(1 - q)s_o + qs(g)](1 + k)$ , which is maximized when  $s(g) = 0$ . Note that  $s^*(g)$  is not unique in the first two intervals because a higher  $s(g)$  would also deter.

Part (e): Some individuals can be deterred, those for whom  $g \leq p[(1 - q)s_o + q\bar{s}]$ , and from part (c) we know that they will be deterred.

Part (a): Given part (c), social welfare (7) can be written as

$$W(s_o) = \int_{p[(1-q)s_o + q\bar{s}]}^{\bar{g}} [g - h - p(1-q)s_o(1+k)]f(g)dg, \quad (8)$$

so that

---

<sup>14</sup> The points that if deterrence is possible, sentences should be chosen to accomplish that objective and if deterrence is not possible, sentences should be zero, are stressed in Shavell (1985, pp. 1241-1243) and Shavell (1987, p. 587). If one were to consider goals of the criminal justice system other than deterrence—notably incapacitation—it could be socially desirable to impose positive sentences on undeterrable individuals.

$$W'(s_o) = -p(1-q)\{p[(1-q)s_o + q\bar{s}] - h - p(1-q)s_o(1+k)\}f(p[(1-q)s_o + q\bar{s}]) \quad (9)$$

$$- \int_{p[(1-q)s_o + q\bar{s}]}^{\bar{g}} p(1-q)(1+k)f(g)dg$$

$$= p(1-q)[(h - pq\bar{s} + p(1-q)s_o k)f(p[(1-q)s_o + q\bar{s}]) - (1 - F(p[(1-q)s_o + q\bar{s}]))(1+k)].$$

We first show that  $s_o^*$  can be 0 by demonstrating that  $W'(s_o)$  can be negative for all  $s$  in  $[0, \bar{s}]$ .

The proof parallels that of the corresponding result in Proposition 1. Again, assume that  $f(g)$  is uniform at height  $\hat{f}$  between 0 and  $p\bar{s}$ . Then  $(h - pq\bar{s} + p(1-q)\bar{s}k)\hat{f}$  is an upper bound for the first term in brackets in the last line of (9) and  $(1 - F(p\bar{s}))(1+k)$  is a lower bound for the second term. Thus,  $W'(s_o)$  will be negative for  $s_o$  in  $[0, \bar{s}]$  if

$$(h - pq\bar{s} + p(1-q)\bar{s}k)\hat{f} < (1 - F(p\bar{s}))(1+k), \quad (10)$$

which will hold for a sufficiently low  $\hat{f}$ .

To demonstrate that  $s_o^*$  can be an interior point, we claim that  $W'(0) > 0$  and  $W'(\bar{s}) < 0$  can hold. We know that  $W'(0) = p(1-q)[(h - pq\bar{s})f(q\bar{s}) - (1 - F(q\bar{s}))(1+k)]$ , which will be positive if  $f(q\bar{s})$  is sufficiently high. Similarly,  $W'(\bar{s}) = p(1-q)[(h - pq\bar{s} + p(1-q)\bar{s}k)f(p\bar{s}) - (1 - F(p\bar{s}))(1+k)]$ , which will be negative if  $f(p\bar{s})$  is sufficiently low.

To prove that  $s_o^*$  can be  $\bar{s}$ , we show that  $W'(s_o)$  can be positive for all  $s_o$  in  $[0, \bar{s}]$ . Let  $f(g)$  be uniform over  $[0, \bar{g}]$ , and thus  $f(g) = 1/\bar{g}$ . Then  $(h - pq\bar{s})(1/\bar{g})$  is a lower bound for the first term in brackets of the second line of (9) and  $(1+k)$  is an upper bound for the second term in brackets. Thus,  $W'(s_o)$  will be positive for  $s_o$  in  $[0, \bar{s}]$  if

$$(h - pq\bar{s})(1/\bar{g}) > (1+k), \quad (11)$$

which will hold for a sufficiently high  $h$ .

Part (b): If  $s_o^*$  is an interior solution, it must satisfy the first-order condition from (9), namely

$$[(h - pq\bar{s} + p(1 - q)s_o k)]f(p[(1 - q)s_o + q\bar{s}]) = (1 - F(p[(1 - q)s_o + q\bar{s}]))(1 + k). \quad (12)$$

Part (d): Individuals who will not be deterred include those for whom  $g$  is in  $(p\bar{s}, \bar{g}]$ . This group has positive mass since  $p\bar{s} < \bar{g}$ . The fraction  $pq$  of these individuals will be imprisoned for the positive length of time  $s_o^*$ .  $\square$

The next result concerns social welfare under the sentence adjustment model and compares that model to the standard model.

*Proposition 3.* Under the optimal solution to the sentence adjustment model when the state obtains perfect information about criminals during imprisonment with probability  $q$ ,

- (a) social welfare is higher than under the optimal solution to the standard model;
- (b) social welfare is increasing in the probability  $q$  that the state will observe the criminal's gain  $g$ ; and
- (c) some use of sentences is always desirable (specifically,  $s^*(g)$  must be positive for a range of  $g$ ), whereas in the standard model any use of sentences might be undesirable ( $s^*$  might be zero).

*Notes.* (i) The sentence adjustment model results in two welfare advantages relative to the standard model. First, it leads to lower sentencing costs. Specifically, suppose that sentences in the sentence adjustment model are initially chosen so as to duplicate the sentence in the standard model—that is,  $s_o = s(g) = s^*$ —in which case the same individuals will be deterred in both models, those for whom  $g \leq ps^*$ . However, sentencing costs can be reduced by lowering  $s(g)$  to 0 for all individuals for whom  $g > ps^*$ , that is, for those who are not deterred. Second, the sentence adjustment model accomplishes greater deterrence without the bearing of additional sentencing

costs when  $s^* < \bar{s}$ . Suppose again that the sentence adjustment model duplicates the standard model and consider individuals whose  $g$  is slightly greater than  $ps^*$  and who therefore would not be deterred in the standard model. These individuals can be deterred in the sentence adjustment model by raising  $s(g)$  above  $s^*$ .

(ii) The two welfare advantages of the sentence adjustment model depend on the observation of an offender's gain  $g$ . Therefore, if the probability  $q$  of observing  $g$  rises, welfare will rise.

(iii) In the standard model any use of sentences might be undesirable because the deterrence benefits of a positive sentence might always be outweighed by the resulting sentencing costs. In contrast, under the sentence adjustment model, it is always possible to employ positive sentences to some degree to accomplish deterrence without incurring sentencing costs. For example, set  $s_o = 0$  and  $s(g) = \bar{s}$  for  $g$  in  $[0, pq\bar{s}]$  and  $s(g) = 0$  for  $g$  in  $(pq\bar{s}, \bar{g}]$ . Then individuals with  $g$  in the first interval will be deterred and no others will face sentences.

*Proof.* Part (a): In the standard model, first suppose that  $s^* > 0$ . Then some individuals will be deterred, those whose gains are less than or equal to  $ps^*$ , and some will commit the harmful act, those whose gains exceed  $ps^*$  (there are such individuals given our assumption that  $\bar{g} > \bar{s}$ ). Fraction  $p$  of the latter group will be caught and sentenced to a term of length  $s^*$ . Social welfare can be improved in the sentence adjustment model by setting  $s_o = s^*$ ,  $s(g) = s^*$  for  $g \leq ps^*$ , and  $s(g) = 0$  for  $g > ps^*$ . This system of sentences will also deter individuals whose gains are less than or equal to  $ps^*$  and not deter individuals whose gains are higher, but it will result in lower sentencing costs for the latter group because, with the probability  $q$  of observing the gains of those who are caught, the sentence will be zero, lowering sentencing costs and raising social welfare.

Now suppose  $s^* = 0$ , in which case everyone will commit the harmful act. This outcome can be improved upon in the sentence adjustment model by setting  $s_o = 0$ ,  $s(g) = \bar{s}$  for  $g \leq pq\bar{s}$ , and  $s(g) = 0$  for  $g > pq\bar{s}$ . In this way, some individuals will be deterred, those with gains less than or equal to  $pq\bar{s}$ , thereby raising social welfare since their gains are less than the harm (given our assumption that  $\bar{g} < h$ ). Everyone else will commit the harmful act and not be sentenced, just as in the standard model when  $s^* = 0$ .

Part (b): We prove this result by showing that social welfare (8) rises with  $q$  holding  $s_o$  constant. (Social welfare would be at least as high if  $s_o$  were chosen optimally as a function of  $q$ .) When  $q$  rises, the integrand increases, augmenting social welfare given the lower limit of integration. The lower limit of integration remains the same if  $s_o = \bar{s}$  and grows if  $s_o < \bar{s}$ . Since the integrand is negative, social welfare must be higher.

Part (c): We know from Proposition 2(c) that in the sentence adjustment model  $s^*(g)$  is positive over some positive interval of  $g$ . And we know from Proposition 1(a) that in the standard model  $s^* = 0$  is possible.  $\square$

#### **4. The Sentence Adjustment Model When the State Might Obtain Imperfect Information about Criminals During Imprisonment**

We next consider the sentence adjustment model assuming that the state might obtain imperfect information about a criminal's gain  $g$  when he is in prison. In particular, we suppose that with probability  $q$  the state learns only the group of individuals to which a criminal belongs, where there are  $n \geq 2$  mutually exclusive groups making up the population of potential

offenders.<sup>15</sup> Let  $f(g|j)$  be the conditional density of  $g$  for group  $j$  and let  $r(j) > 0$  be the probability of that group.<sup>16</sup> The unconditional density  $f(g)$  is the sum of the  $r(j)f(g|j)$  over  $j$ . Each individual in the population is assumed to know his group.<sup>17</sup>

In this section the adjusted sentence must be a function of the group to which a prisoner belongs. Let

$s(j)$  = adjusted sentence if the state observes that a prisoner is in group  $j$ .

Hence, an individual who is a member of group  $j$  will commit the harmful act if

$$g > p[(1 - q)s_o + qs(j)] \quad (13)$$

and will be deterred otherwise.

The state's problem is to choose  $s_o$  and the function  $s(\cdot)$  to maximize social welfare, which is the weighted sum of social welfare for the  $n$  groups:

$$W(s_o, s(\cdot)) = \sum_{j=1}^n r(j)W_j(s_o, s(j)), \quad (14)$$

where

$$W_j(s_o, s(j)) = \int_{p[(1-q)s_o + qs(j)]}^{\bar{g}} \{g - h - p[(1-q)s_o + qs(j)](1+k)\} f(g|j) dg. \quad (15)$$

---

<sup>15</sup> The assumption that the groups are mutually exclusive is not restrictive. Suppose that there could be overlap among the groups and let  $q$  be the probability that the state learns the groups to which an individual belongs. Then the number of groups can be expanded and redefined in such a way that they are mutually exclusive and amenable to the application of our model. To explain how this can be done, consider the case in which there are initially two groups, 1 and 2, with some individuals in group 1 alone, some in group 2 alone, and some in groups 1 and 2. One can then treat these three sets of individuals as being in newly-defined groups 1, 2, and 3, respectively, and perform the analysis in this section with respect to these three mutually exclusive groups. An analogous redefinition of groups that might be overlapping can be done for any initial number of groups.

<sup>16</sup> As in the standard model, we assume that  $f(g|j) > 0$  in  $[0, \bar{g}]$ .

<sup>17</sup> For example, an individual would know whether he is in a group whose members generally comply with rules or break them.

To explain (15), a person in group  $j$  will commit the harmful act when (13) holds, in which case the expression in braces is his contribution to social welfare—his gain less the harm he causes and less the expected private and public cost of his sentence.

Note from (15) that  $W_j(s_o, s(j))$  depends on its arguments only through the expected sentence  $(1 - q)s_o + qs(j)$ . We can therefore write  $W_j((1 - q)s_o + qs(j))$  instead of  $W_j(s_o, s(j))$ . The notation  $W_j((1 - q)s_o + qs(j))$  is helpful because it shows that  $W_j(s_o, s(j))$  may be interpreted as social welfare in the standard model when the sentence is  $(1 - q)s_o + qs(j)$  and the density of  $g$  is  $f(g|j)$ ; see (1).

It will be convenient to refer to the maximum value of (14) given  $q$  as

$W^*(q)$  = maximum level of social welfare in the sentence adjustment  
model with imperfect information given  $q$ .

Consider now the state's problem if it were not able to observe any individual's group, that is, if  $q = 0$ . Then, because the  $s(j)$  would never be employed, (14) would reduce to

$$W(s_o, s(\cdot)) = \int_{ps_o}^{\bar{g}} [g - h - ps_o(1 + k)]f(g)dg, \quad (16)$$

where  $f(g) = \sum_{j=1}^n r(j)f(g|j)$ . Observe that (16) is social welfare in the standard model (1) when  $s$  is replaced by  $s_o$ . Accordingly, the maximum of (16) over  $s_o$  is  $W^*$ , the maximum level of social welfare in the standard model, that is,  $W^*(0) = W^*$ .

Next, consider the state's problem if it were able to observe each individual's group with certainty, when  $q = 1$ . Then, because  $s_o$  would never be employed, (14) would reduce to

$$W(s_o, s(\cdot)) = \sum_{j=1}^n r(j) \int_{ps(j)}^{\bar{g}} [g - h - ps(j)(1 + k)]f(g|j)dg. \quad (17)$$



It is clear that for each  $j$ , the  $s(j)$  that maximizes (17) is the same as the  $s$  that would maximize social welfare  $W(s)$  in the standard model if the density of gains were  $f(g|j)$ ; see (1). Hence, we will denote the sentence for group  $j$  that maximizes (17) by

$s_j^*$  = the optimal sentence for group  $j$  in the standard model,

which we will sometimes refer to as the optimal standard sentence for group  $j$ . We assume that the  $s_j^*$  are unique, that not all  $s_j^*$  are equal, and, without loss of generality, that the groups are labeled such that  $s_1^* \leq s_2^* \leq \dots \leq s_n^*$ . We will denote the value of (17) evaluated at the  $s_j^*$  by

$W^{**}$  = second-best level of social welfare in the sentence adjustment

model with imperfect information.

$W^{**}$  is said to be second-best because it is the achievable level of social welfare when information about an individual's group  $j$  is perfect. Thus,  $W^*(1) = W^{**}$ .

We next define two threshold levels of  $q$  to which we will refer below. Let

$\hat{q}$  = the least upper bound of the set of  $q$  for which  $W^*(q)$  equals  $W^*$ ; and

$\tilde{q} = (s_n^* - s_1^*)/\bar{s}$ .

Note that  $\hat{q}$  exists because the set in question is nonempty (it includes  $q = 0$  since  $W^*(0) = W^*$ ) and because the set is bounded (by 1). Observe also that  $\tilde{q} > 0$  because our assumption that not all the  $s_j^*$  are equal implies that  $s_n^* > s_1^*$ . Moreover,  $\tilde{q}$  could be as high as 1, for  $s_n^* = \bar{s}$  and  $s_1^* = 0$  is possible.

We now state our primary results concerning the present version of the sentence adjustment model.

*Proposition 4.* In the sentence adjustment model when the state obtains imperfect information about criminals during imprisonment with probability  $q$ , optimal social welfare  $W^*(q)$  equals  $W^*$ , the optimal level of social welfare in the standard model, at  $q = 0$  and rises to

$W^{**}$ , the second-best level of social welfare, at  $q = \tilde{q}$ . Whenever  $W^*(q)$  exceeds  $W^*$ , sentence adjustment is desirable. Specifically,

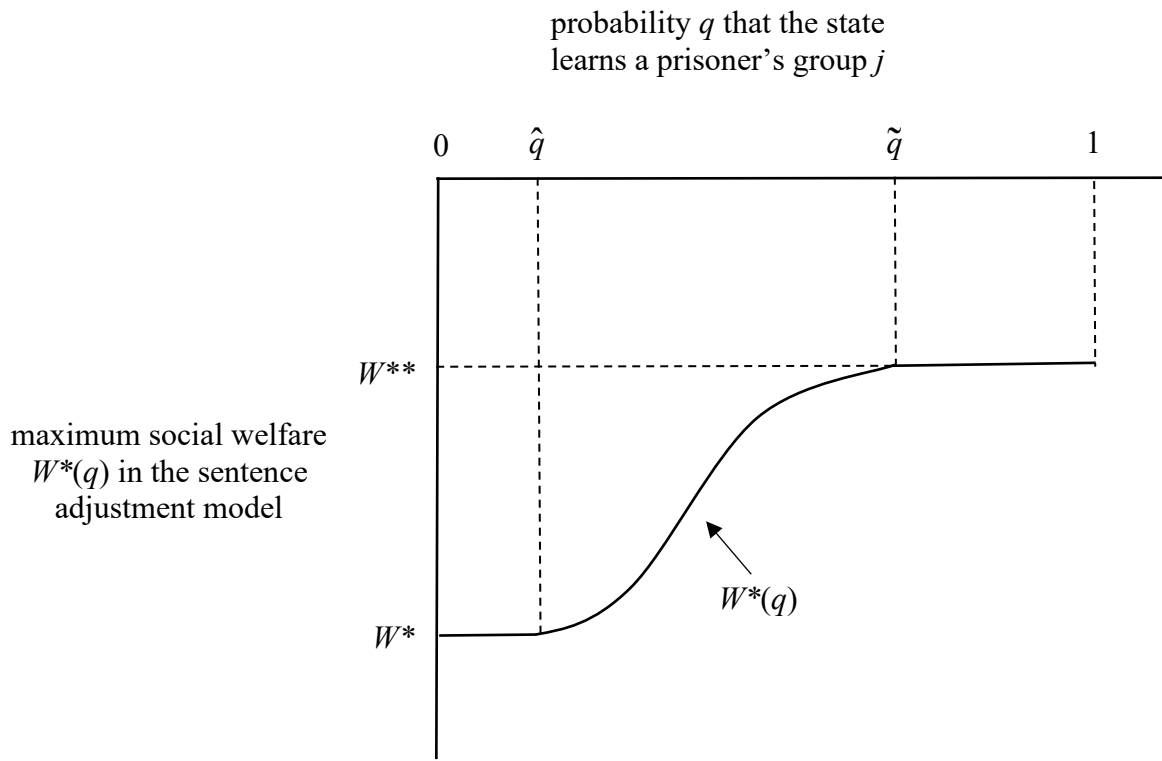
- (a)  $W^* < W^{**}$ ;
- (b)  $W^*(q)$  is non-decreasing in  $q$ ;
- (c)  $W^*(q) = W^*$  for  $q$  in  $[0, \hat{q}]$ , where  $0 \leq \hat{q} < 1$ ;
- (d)  $\hat{q} < \tilde{q}$ ;
- (e)  $W^* < W^*(q) < W^{**}$  for  $q$  in  $(\hat{q}, \tilde{q})$ ; and
- (f)  $W^*(q) = W^{**}$  for  $q$  in  $[\tilde{q}, 1]$ .

*Notes.* (i) An example of the claims of the proposition is represented in Figure 1 below. Note from the figure that social welfare is negative; that is because the gains of individuals from committing the harmful act are less than the harm and because imprisonment is socially costly.

(ii) The claim of part (a), that  $W^*$  is less than  $W^{**}$ , follows from the observation that  $W^*$  results from applying the same  $s^*$  to all groups, whereas  $W^{**}$  is determined by applying a different  $s(j)^*$  to each group with certainty.

(iii) The proof of part (b) is a demonstration that the optimal outcome at any  $q$  can be duplicated at any  $q' > q$  and therefore that  $W^*(q)$  could not be decreasing in  $q$ .

(iv) With regard to part (c), note that the interpretation of the claim that  $\hat{q}$  can be positive is that there may exist an initial interval of  $q$  over which information about a criminal's group cannot be used to enhance social welfare. This possibility contrasts with the result in some other



**Figure 1**

incentive contexts that informative signals about individuals always have value.<sup>18</sup> Our expectation, however, is that here  $\hat{q}$  would normally be 0, in which case any information about criminals would be beneficial.

(v) To prove part (d), we first note from part (c) that  $W^*(\hat{q}) = W^*$ . Then we show that  $W^*(\tilde{q}) = W^{**}$ , which exceeds  $W^*$  by part (a). Hence,  $W^*(\hat{q}) < W^*(\tilde{q})$ . Because, by (b),  $W^*(q)$  is non-decreasing, we conclude that  $\tilde{q}$  must exceed  $\hat{q}$ .

(vi) With regard to part (e), we know from part (c) that  $W^*(\hat{q}) = W^*$  and from the proof of part (d) that  $W^*(\tilde{q}) = W^{**}$ . That  $W^*(q)$  increases over this interval reflects the increased opportunity for tailoring of adjusted sentences with higher  $q$ .

(vii) An important aspect of the claim in part (f) is that the second-best level of social welfare  $W^{**}$  can be obtained at  $\tilde{q}$  when it is less than 1. This is possible because the  $s(j)$  can be chosen to compensate for the probability  $(1 - \tilde{q})$  of no sentence adjustment, resulting in an expected sentence of  $s_j^*$ . For  $q > \tilde{q}$ ,  $W^*(q)$  cannot be increased because  $W^{**}$  is second-best optimal. We will elaborate on this intuition below when we discuss Proposition 5.

*Proof.* See the Appendix.

---

<sup>18</sup> Notably, this is true of signals about a risk-averse agent's effort in a standard principal-agent model. See Hölmstrom (1979) and Shavell (1979). The reason that information is always worthwhile employing there is that monetary transfers are used to generate beneficial incentives. Such transfers do not generate risk-bearing costs in the small (the first-order effect of the initial imposition of risk on a risk-averse individual is zero). In our case, however, information might not be worthwhile employing because beneficial incentives can only be generated through the use of imprisonment, which is socially costly to impose, even in the small.

To illustrate the main results of Proposition 4 consider two equal-sized groups,  $r(1) = r(2) = .5$ . The parameter values of the example, other than the probability  $q$  of obtaining information about prisoners, are presented in Table 1 and the densities of the gains for the two groups are described in Figure 2.<sup>19</sup>

Given a probability of detection of .8 and a maximum sentence of 90, the maximum expected sentence is 72. Thus, in each group there are some individuals who are deterrable: in group 1, those with gains between 10 and 50; and in group 2 those with gains between 30 and 70. Note that individuals in group 1 are generally easier to deter than those in group 2.<sup>20</sup> Also, in both groups there are some individuals who are undeterrable: those with gains between 80 and 100. The presence of undeterrable individuals implies that there will be a social cost from sentencing.

In this example the second-best outcome would be achieved if it were possible to impose the optimal sentence in the standard model of  $s_1^* = 60$  for group 1 and the optimal sentence in the standard model of  $s_2^* = 85$  for group 2. Given these sentences, social welfare would be  $W^{**} = -52$ .<sup>21</sup> If the two groups could not be differentiated, the optimal sentence in the standard model for the two groups combined would be  $s^* = 80$ , resulting in social welfare of  $W^* = -58$ .

---

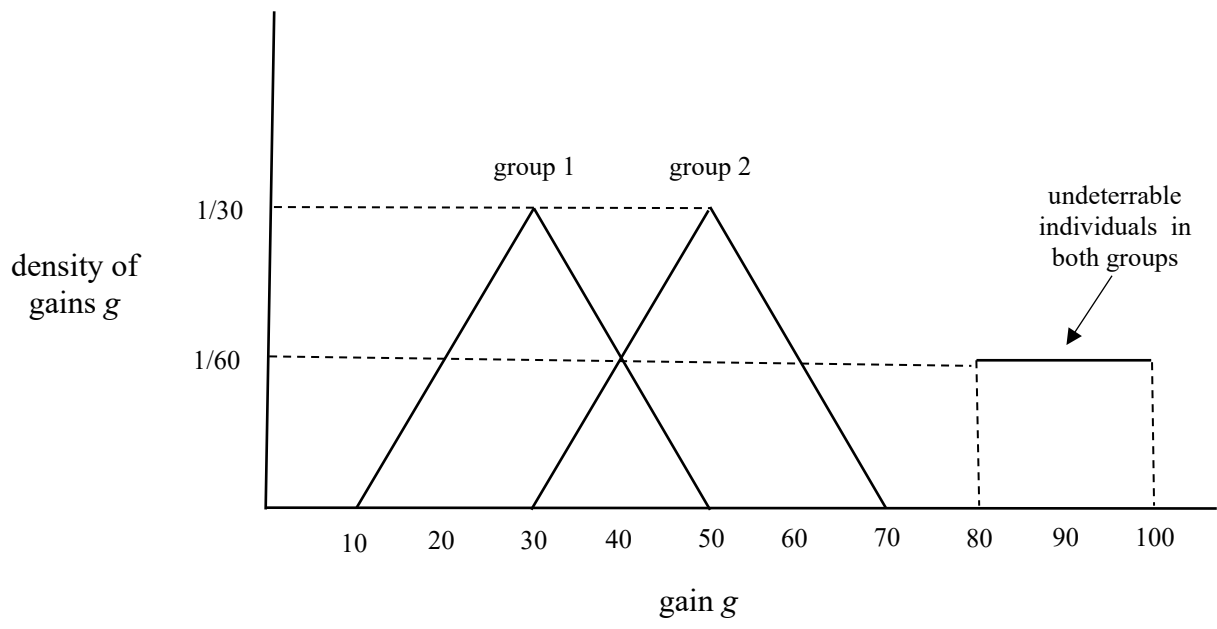
<sup>19</sup> The density functions in Figure 2 are not continuously differentiable and positive at all points, contrary to the assumptions in our model. It will be obvious, however, that density functions that are very close to those in Figure 2 that are continuously differentiable and positive at all points would generate the same qualitative results as our example. We note too that the results reported below are based in part on spreadsheet calculations in which the gains of individuals have discrete values between 1 and 100.

<sup>20</sup> More individuals in group 1 are deterred by any expected sentence between 10 and 70.

<sup>21</sup> Recall that social welfare is negative in our model (see note (i) to Proposition 4).

100	maximum gain of individuals $\bar{g}$
125	harm $h$
.8	probability of catching offenders $p$
90	maximum sentence $\bar{s}$
1	cost to the state per unit of imprisonment $k$

**Table 1**



**Figure 2**

Table 2 presents the results in the example when information about prisoners is used to optimally adjust sentences, given various probabilities  $q$  of observing their group. For each value of  $q$ , the table provides the optimal initial sentence  $s_o^*$ , the optimal adjusted sentences  $s(j)^*$ , the resulting expected sentences  $(1 - q)s_o^* + qs(j)^*$ , and maximum social welfare  $W^*(q)$ . In this example,  $\hat{q} = 0.00$ , meaning that information about prisoners is valuable at all positive levels of  $q$ ; and  $\tilde{q} = (85 - 60)/90 = .28$ , so that the second-best outcome can be achieved by adjusting sentences optimally for any  $q$  at or above this level. In Table 2, social welfare steadily increases with  $q$  for the values of  $q$  below .28 and then is constant at  $W^{**} = -52$  for the values of  $q$  above .28.

The next proposition discusses how the second-best optimum can be attained for  $q$  sufficiently high.

*Proposition 5.* As demonstrated in Proposition 4, the second-best level of social welfare  $W^{**}$  can be achieved if and only if the probability  $q$  of obtaining imperfect information about a prisoner is in  $[\tilde{q}, 1]$ . Assuming that  $s_l^* > 0$  and  $s_n^* < \bar{s}$ , the following sentences will result in this outcome when  $q$  is in  $[\tilde{q}, 1]$ .

- (a) the optimal initial sentence  $s_o^*$  is  $s_l^*/(1 - \tilde{q})$ ; and
- (b) the optimal adjusted sentence  $s(j)^*$  for group  $j$  is  $[s_j^* - (1 - q)s_o^*]/q$ .

Moreover,

- (c)  $s(1)^* \leq s(2)^* \leq \dots \leq s(n)^*$ , where at least one of these inequalities is strict;
- (d) for all  $q$  in  $[\tilde{q}, 1]$ , the set of groups  $j$  can be arranged into the same three subsets: a non-empty low subset  $\{1, \dots, j_L\}$  in which  $s(j)^* < s_o^*$ ; a possibly empty intermediate subset



<i>probability of learning a prisoner's group, <math>q</math></i>	<i>optimal initial sentence, <math>s_o^*</math></i>	<i>optimal adjusted sentence for group 1, <math>s(1)^*</math></i>	<i>optimal adjusted sentence for group 2, <math>s(2)^*</math></i>	<i>expected sentence for group 1</i>	<i>expected sentence for group 2</i>	<i>maximum social welfare given <math>q</math>, <math>W^*(q)</math></i>
.00	80	—	—	64	64	−58
.10	79	0	89	57	64	−56
.20	78	0	88	50	64	−54
.30	83	6	89	48	68	−52
.40	83	25	88	48	68	−52
1.00	—	60	85	48	68	−52

Notes: (a) The results reported in this table are rounded to the nearest integer. (b) The solutions when  $q = .30$  and  $q = .40$  are derived from the formulas provided in Proposition 5 below. (c) If  $q = .00$ ,  $s(1)^*$  and  $s(2)^*$  are irrelevant because the two groups are never differentiated. (d) If  $q = 1.00$ ,  $s_o^*$  is irrelevant because the two groups are always differentiated.

**Table 2**

$\{j_{L+1}, \dots, j_{H-1}\}$  in which  $s(j)^* = s_o^*$ ; and a non-empty high subset  $\{j_H, \dots, n\}$  in which  $s(j)^* > s_o^*$ ; and

(e) for groups in the low subset,  $s(j)^*$  approaches  $s_j^*$  from below as  $q$  rises from  $\tilde{q}$  and equals  $s_j^*$  at  $q = 1$ ; for groups in the intermediate subset,  $s(j)^* = s_o^*$  for all  $q$  in  $[\tilde{q}, 1]$ ; and for groups in the high subset,  $s(j)^*$  approaches  $s_j^*$  from above and equals  $s_j^*$  at  $q = 1$ .

*Notes.* (i) Given the formulas in parts (a) and (b),  $s_o^*$  and  $s(j)^*$  can be shown to be feasible and to result in the expected sentence for each group  $j$  of  $s_j^*$ . Hence, the outcome will be second-best.

These formulas will be seen to reflect natural intuitions about how sentences should be adjusted in the light of information about offenders acquired during imprisonment. The existence of explicit formulas for optimal sentences obviously depends on the specific structure of imperfect information in our model. Were we to have employed the most general model of imperfect information—involving a continuous range of signals with no constraints on the posterior distributions given signals—simple formulas analogous to ours would not exist.

(ii) The formulas for  $s_o^*$  and  $s(j)^*$  that result in the second-best optimum when  $q$  is at least  $\tilde{q}$  require only knowledge of the optimal standard sentences  $s_j^*$  for the individual groups. In contrast, calculation of the optimal sentences when  $q$  is lower than  $\tilde{q}$  generally depends on the entire distributions of gains  $f(g|j)$  of the various groups (not just on the  $s_j^*$ ) and is therefore more complicated.

(iii) Part (c) follows from the formula for  $s(j)^*$  in part (b).

(iv) It can be shown that parts (d) and (e) follow from the formulas in parts (a) and (b).

Note that if a group is in the low subset, the optimal adjusted sentence,  $s(j)^*$ , is not only less than the optimal initial sentence,  $s_o^*$ , but also below the optimal standard sentence for that group,  $s_j^*$ .

That the adjusted sentence thus overshoots  $s_j^*$  in this sense is necessary in order for the expected sentence for the group,  $(1 - q)s_o^* + qs(j)^*$ , to equal  $s_j^*$ . In essence,  $s(j)^*$  must overshoot  $s_j^*$  because the probability  $q$  of observing a prisoner's group is less than 1. As  $q$  rises, the need to overshoot  $s_j^*$  diminishes, which explains why  $s(j)^*$  converges toward  $s_j^*$  as  $q$  approaches 1. Parallel observations apply if a group is in the high subset, and this logic also explains why no adjustment is needed if a group is in the intermediate subset.

(v) We note that the solution that we have discussed that achieves the second-best level of social welfare is not necessarily unique.<sup>22</sup>

(vi) We did not discuss the possibilities that  $s_l^* = 0$  and/or  $s_n^* = \bar{s}$  in the present proposition because they are of limited interest.<sup>23</sup>

*Proof.* See the Appendix.

To illustrate Proposition 5, consider an example with four groups for which the optimal sentences in the standard model  $s_j^*$  are 20, 40, 60, and 80, and the maximum sentence  $\bar{s}$  is 100. Then  $\tilde{q}$  is  $(80 - 20)/100 = 0.60$ . Using the formulas in parts (a) and (b), Table 3 lists the optimal initial sentence  $s_o^*$  and the optimal adjusted sentences  $s(j)^*$  for values of  $q$  at  $\tilde{q} = .60$  and at .70, .80, and .90 (the table also lists the  $s_j^*$  in the row labeled 1.00). As can be seen, for each value of  $q$  the relationships in part (c) are satisfied strictly. Additionally, it is apparent that regardless of the value of  $q$ , groups 1 and 2 are in the low subset discussed in part (d), that the intermediate

---

<sup>22</sup> For example, it is clear that in a region sufficiently close to  $q = 1$ , the second-best optimum can be achieved with  $s_o = 0$  and  $s(j) = s_j^*/q$ .

<sup>23</sup> If  $s_l^* = 0$  and  $s_n^* = \bar{s}$ , then  $\tilde{q} = (s_n^* - s_l^*)/\bar{s} = 1$ , in which case the interval  $[\tilde{q}, 1]$  becomes the point 1. At  $q = 1$  information about groups is perfect, and it is clear that the second-best optimum can be achieved with  $s(j) = s_j^*$  and that  $s_o$  can take on any value because it is never employed. If just  $s_l^* = 0$ , the low subset referred to in part (d) could not exist, and if just  $s_n^* = \bar{s}$ , the claim about  $s(j)^*$  approaching  $s_j^*$  from above could not apply for group  $n$ .

$q$	$s_o^*$	$s(1)^*$	$s(2)^*$	$s(3)^*$	$s(4)^*$
.60	50	0	33	67	100
.70	50	7	36	64	93
.80	50	13	38	63	88
.90	50	17	39	61	83
1.00	—	20	40	60	80

Notes: (a) The results reported in this table are rounded to the nearest integer. (b) If  $q = 1.00$ ,  $s_o^*$  is irrelevant because the four groups are always differentiated; in this case  $s(j)^* = s_j^*$ .

**Table 3**

subset is empty, and that groups 3 and 4 are in the high subset. It is also clear that the properties described in part (e) are illustrated in Table 3; for instance, with respect to group 1 in the low subset,  $s(1)^*$  rises from 0 to 17 as  $q$  rises from 0.60 to 0.90, approaching  $s_I^*$  of 20 from below.

## 5. Concluding Comments

(a) *Sentence adjustment in practice is only in a downward direction.* Although in our model sentences can be raised or lowered during imprisonment, in reality sentences either cannot be changed or can only be lowered. Specifically, sentences are either determinate—fixed at the outset and unmodifiable—or indeterminate—specified as a range, such as five to ten years, with a parole board deciding whether a prisoner serves less than the maximum term.<sup>24</sup>

(b) *The social desirability of downward-only sentence adjustment.* A regime in which sentences can only be adjusted downward, that is, an indeterminate sentencing regime,<sup>25</sup> can increase social welfare. This is because the information obtained about a prisoner might suggest that his sentence should be less than the fixed sentence that would be optimal in the standard

---

<sup>24</sup> See generally Clear, Reisig, and Cole (2019, pp. 78-79). In the United States and in other countries indeterminate sentencing is used more frequently than determinate sentencing. See Lawrence (2015, p. 4) and Aharonson (2013, pp. 164-175).

<sup>25</sup> To see that an indeterminate sentencing regime can be interpreted as a regime in which sentences can only be adjusted downward from an initial level, let  $s_L$  and  $s_U$  be the lower and upper bounds of an indeterminate sentence. Then if a parole board does not grant a reduction from  $s_U$ , the prisoner will serve a term of  $s_U$ ; otherwise, the prisoner will serve a shorter term but one not less than  $s_L$ . Hence, an indeterminate sentencing scheme is equivalent to a sentencing scheme in our model in which the initial sentence is  $s_U$  and the adjusted sentences  $s(j)$  are constrained to be between  $s_L$  and  $s_U$ .

model.<sup>26</sup> However, in a range of circumstances, downward-only sentence adjustment cannot raise social welfare as much as unconstrained sentence adjustment.<sup>27</sup>

(c) *Do potential offenders take sentence adjustment into account?* Substantial evidence exists that individuals are responsive to the threat of criminal sanctions.<sup>28</sup> This could only be the case if individuals view the prospect of sanctions as serious. Given that the level of sentences is significantly altered during imprisonment—as noted in the introduction sentences are reduced on

---

<sup>26</sup> For instance, if social welfare in the standard model for each group is concave in that group's expected sentence, social welfare can be increased through a regime of downward-only sentence adjustment. We sketch the argument here for the imperfect information model. We first claim that  $s^* > s_l^*$ . If this were not true, then one possibility is that  $s^* = s_l^*$ . But that would contradict the optimality of  $s^*$ . In particular, raising  $s$  above  $s_l^*$  at the margin would increase group  $n$ 's social welfare in the standard model due to the concavity assumption (because  $s_l^* < s_n^*$ ) and also that of any other group for which  $s_j^* > s_l^*$ ; but doing so would not have a first-order effect on the social welfare of any group for which  $s_j^* = s_l^*$ . Hence, raising  $s$  marginally must increase social welfare of the population as a whole (it is the sum of the groups). This would contradict the optimality of  $s^* = s_l^*$ . The other possibility is that  $s^* < s_l^*$ . Then marginally raising  $s$  above  $s^*$  would raise social welfare in the standard model of all groups due to the concavity assumption, again contradicting the optimality of  $s^*$ .

Given that  $s^* > s_l^*$ , social welfare with downward-only sentence adjustment will be higher than social welfare in the absence of sentence adjustment. To demonstrate this, let  $s_o = s^*$  and  $s(j) = s^*$  for all  $j > 1$  and let  $s(1)$  be marginally less than  $s^*$ . That will not affect the social welfare in the standard model of any group  $j$  above 1 and will raise social welfare of group 1 given that  $s_l^* < s^*$  and the concavity assumption. Hence social welfare of the entire population will have risen.

<sup>27</sup> We show here that for a range of the probability  $q$  of observing information, our optimal sentence adjustment scheme is preferable to any downward-only scheme under the assumption that  $s_n^* < \bar{s}$ . To prove this, we first demonstrate the superiority of the optimal sentence adjustment scheme at  $q = \tilde{q} = (s_n^* - s_l^*)/\bar{s}$ . We know from part (f) of Proposition 4 that the optimal sentence adjustment scheme achieves the second-best level of social welfare  $W^{**}$  at  $\tilde{q}$ . Hence, under this sentencing scheme the expected sentence for each group  $j$  equals  $s_j^*$ . We now show that at  $q = \tilde{q}$ , a downward-only sentence scheme cannot result in both an expected sentence of  $s_n^*$  for group  $n$  and  $s_l^*$  for group 1; a fortiori, any downward-only scheme must be inferior to the optimal sentence adjustment scheme at  $\tilde{q}$ . Let  $s_l(1)$  be the sentence imposed under a downward-only sentencing regime if the state learns that a prisoner is in group 1. Thus, the expected sentence for individuals in group 1 will be  $(1 - \tilde{q})s_U + \tilde{q}s_l(1)$ . Observe also that for a downward-only regime to achieve an expected sentence of  $s_n^*$  for individuals in group  $n$ , it must be that  $s_U \geq s_n^*$  (otherwise the expected sentence would be a weighted average of two numbers less than  $s_n^*$ ). We also know that  $s_l(1) \geq s_L \geq 0$ . Hence, we have

$$(1 - \tilde{q})s_U + \tilde{q}s_l(1) \geq (1 - \tilde{q})s_n^* = [(\bar{s} - s_n^* + s_l^*)/\bar{s}]s_n^* > s_l^*,$$

where the last inequality holds because it can be written as  $(\bar{s} - s_n^*)s_n^* > (\bar{s} - s_n^*)s_l^*$ , which is satisfied because  $\bar{s} - s_n^* > 0$  and  $s_n^* > s_l^*$ . In other words, if the expected sentence for individuals in group  $n$  equals  $s_n^*$ , the expected sentence for individuals in group 1 must exceed  $s_l^*$ , which implies that the downward-only sentence adjustment regime is inferior to the optimal sentence adjustment regime at  $\tilde{q}$ .

To complete the argument for the claim in this footnote, we note by continuity that there will be a range of  $q$  above  $\tilde{q}$  over which no downward-only sentencing regime can achieve the second-best outcome. But part (f) of Proposition 4 shows that under our sentence adjustment model, the second-best outcome can be obtained in this region.

<sup>28</sup> See, for example, Andenaes (1966, pp. 960-973), Levitt and Miles (2007, pp. 466-474), Durlauf and Nagin (2012, pp. 47-71), and Chalfin and McCrary (2017, pp. 13-32).

average by 54 percent—it is reasonable to suppose that potential offenders would factor into their calculation of expected sentences the general effect of these adjustments. This belief is reinforced by the fact that the institution of parole is an element of our popular culture, as reflected in books, television, and the news.

(d) *Can the state acquire information about prisoners that bears on deterrence?* It is plausible that the state can obtain information during imprisonment that is pertinent to deterrence. Suppose that it is learned that a prisoner was accorded special status in his gang for having committed a murder. This could indicate that a longer sentence would be appropriate for purposes of deterrence because of the higher gain from his crime that the prisoner enjoyed relative to what the court had known at the time of sentencing.<sup>29</sup> Or suppose that it becomes apparent that a prisoner does not experience great disutility from imprisonment. This knowledge could also suggest the need for a lengthier sentence to properly deter.

Conversely, suppose that it is observed that a prisoner has generally followed prison rules, has made an effort to obtain job training, and is remorseful for his crime. This kind of record might indicate that a shorter sentence than would otherwise apply would be appropriate on grounds of deterrence.<sup>30</sup> Or suppose that it is found that a prisoner is mentally ill. Such a discovery might suggest that his responsiveness to the prospect of punishment is low or

---

<sup>29</sup> It could also be that a shorter sentence would be better from the perspective of optimal deterrence—for instance, if the majority of gang members who commit murder obtain such high status benefits that they are impossible to deter.

<sup>30</sup> The personality type of such a prisoner suggests that his criminal act may have been motivated by unusual circumstances and that ordinarily individuals with his set of characteristics could be deterred by a shorter sentence. The reason for lowering the sentence is not that the particular individual in prison would have been deterred with a shorter sentence, for obviously he would not have. Rather, it is that a shorter sentence is more desirable in expectation for individuals of this type.

nonexistent, in which case a shorter sentence could be desirable from the perspective of deterrence.

It should be noted that the information obtained by the state prior to an offender's imprisonment is usually the result of a plea bargain and only rarely the result of a trial.<sup>31</sup> Thus, the additional information acquired during imprisonment is greater than if trials were common.

(e) *The failure of parole boards and prison authorities as well as scholars to consider the effects of sentence adjustment on deterrence.* As we observed in the introduction, sentences are reduced during imprisonment through parole and good time and earned time credits. But as we also stated, these reductions are not ordinarily motivated by considerations of deterrence. When parole boards decide on sentence reduction, they focus on whether a prisoner is likely to commit another crime if he is released, that is, whether he will become a recidivist.<sup>32</sup> Moreover, good time credits are employed mainly as a reward for good behavior within prison; and earned time credits are used to encourage prisoners to participate in self-improvement activities.<sup>33</sup> Additionally, discussions of sentence adjustment by scholars are concerned principally with issues of recidivism and other goals unrelated to deterrence.<sup>34</sup>

This lack of attention to the effects of sentence adjustment on deterrence can obviously undermine deterrence. Suppose a person is sentenced to a ten-year prison term for robbery but is

---

<sup>31</sup> See Subramanian et al. (2020, pp. 1-2), where it is observed (note 3) that only approximately two percent of criminal cases in state and Federal courts go to trial.

<sup>32</sup> See generally Cohen (2018, § 4:30), who observes that “the most basic criteria for release on parole” are “whether there is a reasonable probability that a prison inmate, if placed on parole, will be able to live and conduct himself or herself as a respectable, law-abiding person, and whether release will be compatible with the offender’s own welfare and the welfare and safety of society.”

<sup>33</sup> See note 2 above.

<sup>34</sup> For example, within legal and criminological scholarship see Cullen (2013), Klingele (2010), and Rhine, Petersilia, and Reitz (2017); and within economic scholarship see Bernhardt, Mongrain, and Roberts (2012) and Kuziemko (2013).



released after five years because a parole board concludes that, given his exemplary behavior in prison, he would no longer be a risk to society.<sup>35</sup> Clearly, other individuals contemplating committing robbery could expect that they too would serve a sentence much shorter than ten years if they would behave well in prison.<sup>36</sup>

It is noteworthy that the detrimental consequences of ignoring the effects of sentence reductions on deterrence apply even if prison authorities cannot acquire deterrence-relevant information about prisoners. Thus, even if parole boards can only obtain information relevant to recidivism, they should take into account that the early release of prisoners will lessen the deterrence of crime.<sup>37</sup>

(f) *Sentence adjustment should be greater the lower the probability of obtaining information about prisoners.* The formula in Proposition 5 for the optimal adjusted sentence implies that the magnitude of the adjustment is greater the lower the probability of acquiring information about a prisoner (see note (iv) following that proposition). Thus, if it was learned only by sheer chance that a prisoner's crime had been committed intentionally, the sentence

---

<sup>35</sup> According to Kaebler (2018, p. 4, Table 3), the mean prison sentence for robbery is 9.0 years, whereas prisoners serve on average 5.2 years.

<sup>36</sup> One might think that a decline in deterrence due to sentence reductions by parole boards could be offset by raising initial sentences. Although deterrence could be raised *on average* in this manner, increasing an initial sentence would not achieve desirable deterrence because it would elevate deterrence for all groups of individuals, regardless of their optimal levels of deterrence. Consider, for instance, a group for whom the optimal level of deterrence is very low, say for simplicity zero. Then the higher the initial sentence, the greater will be overdeterrence of individuals in this group.

<sup>37</sup> Consider an example similar to that in the previous paragraph: all individuals in the population of potential offenders are identical and would be deterred from robbery by a ten-year sentence, but not by a five-year sentence. Assume too that none of these individuals would commit another crime after five years of imprisonment. Because individuals are identical, nothing relevant to deterrence can be learned about them during imprisonment. Yet a parole board whose goal is to release prisoners who would not commit further crimes would release all of these prisoners after five years, resulting in all potential offenders committing robbery.

should be increased more than if this fact was likely to have been obtained. A higher adjustment is required to maintain the expected sentence at its appropriate level.

(g) *The similarity between a court's use of information obtained at trial and a prison authority's use of information obtained during imprisonment.* The problem facing a trial court concerned with the optimal deterrence of crime is analogous to that studied in our model. Specifically, the court will acquire information during a trial that it can use to adjust the sentence that it otherwise would have chosen based on pre-trial information about the defendant.<sup>38</sup>

---

<sup>38</sup> Formally, the court can be imagined to impose a sentence  $s_o$  if it does not acquire information at trial beyond its knowledge of the probability density  $f(g)$  and to impose a sentence  $s(j)$  if it learns at trial that the defendant is in group  $j$ .

## References

- Aharonson, Ely, "Determinate Sentencing and American Exceptionalism: The Underpinnings and Effects of Cross-National Differences in the Regulation of Sentencing Discretion," *Law and Contemporary Problems*, Vol. 76, No. 1 (2013), pp. 161-187.
- Andenaes, Johannes, "The General Preventive Effects of Punishment," *University of Pennsylvania Law Review*, Vol. 114, No. 7 (May 1966), pp. 949-983.
- Bernhardt, Dan, Steeve Mongrain, and Joanne Roberts, "Rehabilitated or Not: An Informational Theory of Parole Decisions," *Journal of Law, Economics, and Organization*, Vol. 28, No. 2 (June 2012), pp. 186-210.
- Carson, E. Ann, "Prisoners in 2016," Bureau of Justice Statistics, U.S. Department of Justice, NCJ 251149, January 2018 (updated August 7, 2018).
- Chalfin, Aaron, and Justin McCrary, "Criminal Deterrence: A Review of the Literature," *Journal of Economic Literature*, Vol. 55, No. 1 (March 2017), pp. 5-48.
- Clear, Todd R., Michael D. Reisig, and George F. Cole, *American Corrections* (Boston: Cengage Learning, Twelfth Edition, 2019).
- Cohen, Neil P., *The Law of Probation and Parole, Second Edition* (June 2018 update).
- Cullen, Francis T., "Rehabilitation: Beyond Nothing Works," in Michael Tonry (ed.), *Crime and Justice: A Review of Research*, Vol. 42 (Chicago: University of Chicago Press, 2013).
- Durlauf, Steven N., and Daniel S. Nagin, "The Deterrent Effect of Imprisonment," in Philip J. Cook, Jens Ludwig, and Justin McCrary (Eds.), *Controlling Crime: Strategies and Tradeoffs* (Chicago: University of Chicago Press, 2012), pp. 43-94.
- Garoupa, Nuno, "Prison, Parole, and the Criminal Justice Act 1991," *Research in Law and Economics*, Vol. 18 (1996), pp. 125-142.

- Hölmstrom, Bengt, "Moral Hazard and Observability," *Bell Journal of Economics*, Vol. 10, No. 1 (Spring 1979), pp. 74-91.
- Kaeble, Danielle, "Time Served in State Prison, 2016," Bureau of Justice Statistics Bulletin, U.S. Department of Justice, NCJ 252205, November 2018.
- Klinge, Cecelia, "Changing the Sentence without Hiding the Truth: Judicial Sentence Modification as a Promising Method of Early Release," *William & Mary Law Review*, Vol. 52 (2010), pp. 465-536.
- Kuziemko, Ilyana, "How Should Inmates be Released from Prison? An Assessment of Parole versus Fixed-Sentence Regimes," *Quarterly Journal of Economics*, Vol. 128 (2013), pp. 371-424.
- LaFare, Wayne R., Jerold H. Israel, Nancy J. King, and Orin S. Kerr, *Criminal Procedure* (St. Paul, MN: Thomson Reuters, [Fourth]4<sup>th</sup> Edition, updated December 2019).
- Lawrence, Alison, "Making Sense of Sentencing: State Systems and Policies," National Conference of State Legislatures, June 2015 (available at <<https://www.ncsl.org/documents/cj/sentencing.pdf>>).
- Levitt, Steven D., and Thomas J. Miles, "Empirical Study of Criminal Punishment," in A. Mitchell Polinsky and Steven Shavell (Eds.), *Handbook of Law and Economics*, Volume 1 (Elsevier Science, 2007), pp. 403-454.
- Miceli, Thomas J., "Prison and Parole: Minimizing the Cost of Non-Monetary Sanctions as Deterrents," *Research in Law and Economics*, Vol. 16 (1994), pp. 197-211.
- Motivans, Mark, "Federal Justice Statistics, 2012—Statistical Tables," Bureau of Justice Statistics, U.S. Department of Justice, NCJ248470, January 2015.

- Polinsky, A. Mitchell, “Deterrence and the Optimality of Rewarding Prisoners for Good Behavior,” *International Review of Law and Economics*, Vol. 44 (October 2015), pp. 1-7.
- Polinsky, A. Mitchell, and Steven Shavell, “The Economic Theory of Public Enforcement of Law,” *Journal of Economic Literature*, Vol. 38, No. 1 (March 2000), pp. 45-76.
- Pyne, Derek, “Can Early Release Both Reduce Prison Costs and Increase Deterrence?,” *Economics Letters*, Vol. 135 (2015), pp. 69-71.
- Rhine, Edward E., Joan Petersilia, and Kevin R. Reitz, “The Future of Parole Release,” *Crime and Justice*, Vol. 46 (2017), pp. 279-338.
- Shavell, Steven, “Risk Sharing and Incentives in the Principal and Agent Relationship,” *Bell Journal of Economics*, Vol. 10, No. 1 (Spring 1979), pp. 55-73.
- Shavell, Steven, “Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent,” *Columbia Law Review*, Vol. 85, No. 6 (1985), pp. 1232-1262.
- Shavell, Steven, “The Optimal Use of Nonmonetary Sanctions as a Deterrent,” *American Economic Review*, Vol. 77, No. 4 (1987), pp. 584-592.
- Subramanian, Ram, Léon Digard, Melvin Washington II, and Stephanie Sorage, *In the Shadows: A Review of the Research on Plea Bargaining* (New York: Vera Institute of Justice, 2020).

## Appendix

*Proof of Proposition 4.* Part (a): Let  $s^*$  be the optimal sentence in the standard model, resulting in social welfare  $W^*$ . Given the assumption that not all of the  $s_j^*$  are equal, let groups  $i$  and  $k$  be such that  $s_i^* \neq s_k^*$ . Since  $s^*$  cannot be equal to both of these sentences, suppose it is unequal to  $s_i^*$ . When  $q = 1$  and the state can observe each person's group with certainty, set  $s(j)$  equal to  $s^*$  for all groups other than group  $i$  and set  $s(i)$  equal to  $s_i^*$ . The resulting level of social welfare for group  $i$  will be  $W_i(s_i^*)$ , which is higher than it had been under the standard model, when it was  $W_i(s^*)$ . The level of social welfare for all other groups  $j$  will be unchanged, as their expected sentence will still be  $s^*$ . Hence,  $W^{**}$  must exceed  $W^*$ .

Part (b): For any  $q < 1$ , let  $s_o(q)$  denote the initial sentence and  $s(j, q)$  the adjusted sentence for group  $j$ . Hence, the expected sentence for group  $j$  will be

$$E(j, q) = (1 - q)s_o(q) + qs(j, q). \quad (\text{A1})$$

Consider any  $q' > q$ . We will show that we can retain the initial sentence  $s_o(q)$  and find a new feasible adjusted sentence  $s(j, q')$  for group  $j$  such that  $E(j, q') = E(j, q)$ , that is, satisfying

$$(1 - q')s_o(q) + q's(j, q') = (1 - q)s_o(q) + qs(j, q). \quad (\text{A2})$$

If (A2) holds, then the behavior of individuals in group  $j$  will be the same under  $q'$  and  $s(j, q')$  as it was under  $q$  and  $s(j, q)$ . Thus, social welfare will be the same, implying that optimal social welfare  $W^*(q')$  must be at least as high as  $W^*(q)$ , establishing part (b).

To determine the  $s(j, q')$  that satisfies (A2), solve this equation for  $s(j, q')$  to obtain  $s(j, q') = [(q' - q)s_o(q) + qs(j, q)]/q'$ . Clearly  $s(j, q')$  is non-negative since  $q' > q$ . Moreover,  $s(j, q') \leq \bar{s}$  because both  $s_o(q)$  and  $s(j, q)$  are bounded by  $\bar{s}$  and  $[(q' - q)\bar{s} + q\bar{s}]/q' = \bar{s}$ . Thus,  $s(j, q')$  is feasible.

Part (c): To show that  $\hat{q} = 0$  is possible, we need to demonstrate that  $W^*(q) > W^*$  may hold for all  $q > 0$ . This will be so in the following example. Let there be two groups 1 and 2, for which we denote social welfare in the standard model (see (1)) by  $W_1(s)$  and  $W_2(s)$ . Assume that  $s_1^* < \bar{s}$  and is unique, that  $s_2^* = \bar{s}$ , and that  $W_2'(s) > 0$  for  $s$  in  $[0, \bar{s}]$ .<sup>39</sup> Further, for any  $\lambda$  in  $(0, 1)$ , consider the population comprised of a fraction  $\lambda$  of individuals from group 1 and the remaining fraction  $(1 - \lambda)$  from group 2. Let  $W_\lambda(s)$  denote social welfare in the standard model given  $s$  for this combined population. We now demonstrate in several steps that, for suitably high  $\lambda$ ,  $W^*(q) > W^*$  for all  $q > 0$ .

Step (i). We will prove that for any  $\varepsilon > 0$ , there is a  $\lambda^\circ < 1$  sufficiently close to 1 such that, for all  $\lambda$  in  $(\lambda^\circ, 1)$ , the difference between the level of social welfare for group 1,  $W_1(s)$ , and the level of social welfare for the combined population,  $W_\lambda(s)$ , is less than  $\varepsilon$  over the entire range of  $s$ , namely  $[0, \bar{s}]$ . (That is,  $W_\lambda(s)$  uniformly converges to  $W_1(s)$  on  $[0, \bar{s}]$  as  $\lambda$  approaches 1.) The difference between  $W_1(s)$  and  $W_\lambda(s)$  at any  $s$  is the following absolute value:  $|W_1(s) - [\lambda W_1(s) + (1 - \lambda)W_2(s)]| = (1 - \lambda)|W_1(s) - W_2(s)|$ . The maximum of this difference over  $s$  in  $[0, \bar{s}]$  occurs at some  $s_m$  that maximizes  $|(W_1(s) - W_2(s))|$ , so the maximum difference between the two levels of social welfare is  $(1 - \lambda)|(W_1(s_m) - W_2(s_m))|$ . Hence, if we want this difference to be less than  $\varepsilon$ , we need only guarantee that  $(1 - \lambda)|(W_1(s_m) - W_2(s_m))| < \varepsilon$ , or that

$$\lambda > [|(W_1(s_m) - W_2(s_m))| - \varepsilon] / |(W_1(s_m) - W_2(s_m))|. \quad (\text{A3})$$

The right-hand side of (A3) provides the value of  $\lambda^\circ$  required in this step, and it is clear that it is less than 1 since  $\varepsilon > 0$ .

---

<sup>39</sup> If for  $s$  in  $[0, \bar{s}]$  the marginal benefit of raising  $s$  exceeds the corresponding marginal cost, then  $W_2'(s)$  will be positive for  $s$  in  $[0, \bar{s}]$ . This will obviously hold if harm  $h$  is high enough; see (2).

Step (ii). We now demonstrate that for all  $\lambda$  in  $(\lambda^o, 1)$ , where  $\lambda^o$  is the right-hand side of (A3), any  $s$  that maximizes  $W_\lambda(s)$  must be less than  $\bar{s}$ . Select

$$\varepsilon < .5[W_I(s_I^*) - W_I(\bar{s})], \quad (\text{A4})$$

and note that the right-hand side of this inequality is positive because we assumed that  $s_I^*$  is unique. By step (i) and (A3), we know that for this  $\varepsilon$ , there exists a  $\lambda^o$  such that for all  $\lambda$  in  $(\lambda^o, 1)$ , the difference between  $W_I(s)$  and  $W_\lambda(s)$  is less than  $\varepsilon$  for all  $s$ . It follows in particular that  $W_\lambda(s_I^*) > W_I(s_I^*) - \varepsilon$  and  $W_\lambda(\bar{s}) < W_I(\bar{s}) + \varepsilon$ . Hence,  $W_\lambda(s_I^*)$  will be greater than  $W_\lambda(\bar{s})$  if  $W_I(s_I^*) - \varepsilon > W_I(\bar{s}) + \varepsilon$ , which is equivalent to (A4). Accordingly,  $\bar{s}$  cannot maximize  $W_\lambda(s)$ .

Step (iii). Last, we show that for any  $\lambda$  in  $(\lambda^o, 1)$  as described in step (ii),  $W^*(q) > W^*$  for any  $q > 0$ . This will imply that  $\hat{q} = 0$  because it will mean that sentence adjustment raises social welfare for any positive  $q$ . In the absence of sentence adjustment, social welfare will be  $W_\lambda(s^*)$ , where we know from step (ii) that  $s^* < \bar{s}$ . Social welfare in the sentence adjustment model will also be  $W_\lambda(s^*)$  if we set  $s_o = s^*$  and  $s(1) = s(2) = s^*$ . Yet we can achieve a higher level of social welfare for any  $q$  by choosing any  $s(2) > s^*$  and leaving  $s_o$  and  $s(1)$  unchanged: this change will not affect social welfare in group 1 but it will raise social welfare in group 2 because it will increase the expected sentence for that group, and  $W_2'(s)$  was assumed to be positive for all feasible  $s$ .

Next we confirm that  $\hat{q} > 0$  is possible. To do so we again employ an example with two groups and now assume that  $s_I^* = 0$ , that  $W_I'(s) < 0$  for  $s$  in  $[0, \bar{s}]$ ,<sup>40</sup> that  $s_2^* > 0$  and is unique,

---

<sup>40</sup>  $W_I(s)$  will always be declining if, for every  $s$ , the marginal benefit of raising  $s$  is less than the corresponding marginal cost. It is clear from (2) that the marginal benefit can be bounded from above by any number because  $h + psk$  is bounded by  $h + p\bar{s}k$  and  $f(ps)$  can be less than any  $\varepsilon$  in the interval  $[0, p\bar{s}]$  since  $f(ps)$  can be unboundedly high in the interval  $(p\bar{s}, \bar{g}]$ . The marginal cost is at least as high as  $(1 - F(p\bar{s}))(1 + k)$ .



and that  $W_2'(0) < 0$ , implying that  $W_2(s)$  is declining in some positive interval  $[0, s_\delta]$ . We also again consider a population composed of a fraction  $\lambda$  in  $(0, 1)$  of individuals from group 1 and the rest from group 2; and we continue to denote social welfare in the standard model given  $s$  for this population by  $W_\lambda(s)$ .

We will prove that if  $\lambda$  is sufficiently high— $\lambda$  is in  $(\lambda^\circ, 1)$  for a specified  $\lambda^\circ < 1$ —and if  $q$  is sufficiently low— $q$  is in  $(0, q^\circ)$  for a specified  $q^\circ > 0$ —then  $s(1)^* = s(2)^* = 0$ . Hence, for any such  $\lambda$  and  $q$ , social welfare achievable in the sentence adjustment model can be obtained in the standard model with  $s = s_o^*$ . This means that information has no value in a positive interval of  $q$ . Therefore, the least upper bound of the set of  $q$  for which information has no value,  $\hat{q}$ , must be positive. The argument will be made in steps.

Step (iv). We first show that  $s(1)^*$  must be 0 regardless of  $\lambda$  and  $q$ . To demonstrate this, write social welfare as  $\lambda W_1((1 - q)s_o + qs(1)) + (1 - \lambda)W_2((1 - q)s_o + qs(2))$ . Since  $s(1)$  appears only in the first term and since we assumed that  $W_1'(s) < 0$  for all  $s$ , it follows that  $s(1)^* = 0$ . Thus, social welfare in the example is

$$W(s_o, s(\cdot)) = \lambda W_1((1 - q)s_o) + (1 - \lambda)W_2((1 - q)s_o + qs(2)). \quad (\text{A5})$$

Step (v). We next claim that for any  $\varepsilon > 0$ , there is a  $\lambda^\circ < 1$  such that if  $\lambda$  is in  $(\lambda^\circ, 1)$ , then  $W_1(0) - W_1((1 - q)s_o^*) < \varepsilon$  holds for any  $q$  in  $(0, 1)$ . To prove this, assume otherwise. Then there must be an  $\varepsilon > 0$  such that the asserted  $\lambda^\circ$  does not exist. This means that for any  $\lambda^\circ < 1$ , there must be a  $\lambda$  in  $(\lambda^\circ, 1)$  such that  $W_1(0) - W_1((1 - q)s_o^*) \geq \varepsilon$  for some  $q$  in  $(0, 1)$ , or that

$$W_1((1 - q)s_o^*) \leq W_1(0) - \varepsilon \quad (\text{A6})$$

for some such  $q$ . From (A5) and (A6), it follows that

$$\begin{aligned} W(s_o^*, s(j)^*) &= \lambda W_1((1 - q)s_o^*) + (1 - \lambda)W_2((1 - q)s_o^* + qs(2)^*) \\ &\leq \lambda[W_1(0) - \varepsilon] + (1 - \lambda)W_2((1 - q)s_o^* + qs(2)^*) \end{aligned} \quad (\text{A7})$$

$$\leq \lambda[W_1(0) - \varepsilon] + (1 - \lambda)W_2(s_2^*).$$

Because the first line in (A7) is the maximum achievable level of social welfare, it must be at least as large as that with any feasible choices of  $s_o$  and the  $s(j)$ . In particular, it must be greater than or equal to  $\lambda W_1(0) + (1 - \lambda)W_2(0)$ , the level of social welfare if  $s_o = s(1) = s(2) = 0$ . Thus, from (A7), it follows that

$$\lambda W_1(0) + (1 - \lambda)W_2(0) \leq W(s_o^*, s(j)^*) \leq \lambda[W_1(0) - \varepsilon] + (1 - \lambda)W_2(s_2^*), \quad (\text{A8})$$

or that

$$\lambda W_1(0) + (1 - \lambda)W_2(0) \leq \lambda[W_1(0) - \varepsilon] + (1 - \lambda)W_2(s_2^*). \quad (\text{A9})$$

This inequality can be written as

$$[(1 - \lambda)/\lambda][W_2(s_2^*) - W_2(0)] \geq \varepsilon, \quad (\text{A10})$$

which clearly cannot hold if  $\lambda$  is sufficiently close to 1. This contradiction proves our claim about the existence of  $\lambda^\circ$ .

Step (vi). We now show that there is a  $\lambda^\circ < 1$  and a  $q' > 0$  such that if  $\lambda > \lambda^\circ$  and  $q \leq q'$ , then  $s_o^* < s_\delta$ . Let  $\varepsilon$  equal  $W_1(0) - W_1(.5s_\delta) > 0$ . From step (v), we know that for this  $\varepsilon$  there exists a  $\lambda^\circ$  such that if  $\lambda$  is in  $(\lambda^\circ, 1)$ , then

$$W_1(0) - W_1((1 - q)s_o^*) < W_1(0) - W_1(.5s_\delta), \quad (\text{A11})$$

or  $W_1(.5s_\delta) < W_1((1 - q)s_o^*)$ . Since  $W_1'(s) < 0$  for all  $s_1$ , this implies that  $.5s_\delta > (1 - q)s_o^*$  or  $s_o^* < .5s_\delta/(1 - q)$ . If  $q = q' = .5$ , the right-hand side of this last inequality is  $s_\delta$ .

Step (vii). Finally, we demonstrate that  $s(2)^* = 0$  for  $\lambda$  exceeding the  $\lambda^\circ$  identified in step (vi) and for  $q$  less than a  $q^\circ$  to be determined here. By step (vi), we know that  $s_o^* < s_\delta$  for  $\lambda > \lambda^\circ$  and  $q \leq q'$ . Now consider a  $q''$  sufficiently low that  $(1 - q'')s_o^* + q''\bar{s} < s_\delta$ . This implies that  $(1 - q'')s_o^* + q''s(2) < s_\delta$  and, for any  $q \leq q''$ , that  $(1 - q)s_o^* + qs(2) < s_\delta$ . In other words, for  $q \leq q''$ , the expected sentence for group 2,  $(1 - q)s_o^* + qs(2)$ , is less than  $s_\delta$ . But given the assumption

that  $W_2'(s) < 0$  in  $[0, s_\delta]$ , it follows that  $s(2)^* = 0$ . If  $q^o$  is set equal to  $\min(q', q'')$ , then the claim of this step is demonstrated. That completes the proof that  $s(1)^*$  and  $s(2)^*$  equal 0 in our example and thus that  $\hat{q} > 0$  is possible.

We next demonstrate that  $W^*(q) = W^*$  for  $q$  in  $[0, \hat{q}]$  when  $\hat{q} > 0$  (if  $\hat{q} = 0$ , the assertion is self-evident). We first claim that  $W^*(q) = W^*$  for  $q$  in  $[0, \hat{q})$ . If this were not true, then there would be a  $q' < \hat{q}$  such that  $W^*(q') > W^*$ . Since  $W^*(q)$  is non-decreasing in  $q$  (part (b) of the proposition), it must be that  $W^*(q) \geq W^*(q') > W^*$  for  $q$  in  $[q', \hat{q})$ . But this contradicts the assumption that  $\hat{q}$  is the least upper bound of the set of  $q$  for which  $W^*(q)$  equals  $W^*$ , since the least upper bound would have to be at or below  $q'$ . Thus,  $W^*(q) = W^*$  for  $q$  in  $[0, \hat{q})$ . Since  $W^*(q)$  is continuous in  $q$ , it must also be that  $W^*(\hat{q}) = W^*$ .

Finally, we know that  $\hat{q} < 1$ , since  $W^*(\hat{q}) = W^*$ , whereas  $W^*(1) = W^{**} > W^*$ .

Part (d): We know from having just shown that  $\hat{q} < 1$  that if  $\tilde{q} = 1$ , then  $\hat{q} < \tilde{q}$ . Hence, to complete the proof of the claim, we can restrict attention to the case of  $\tilde{q} < 1$ .

We first show that if  $q = \tilde{q}$ ,

$$s_o = s_l^*/(1 - \tilde{q}), \tag{A12}$$

and

$$s(j) = [s_j^* - (1 - \tilde{q})s_o]/\tilde{q}, \tag{A13}$$

then these sentences will be feasible and the expected sentence for individuals in each group  $j$  will be  $s_j^*$ . Hence,  $W^*(\tilde{q}) = W^{**}$ .

To verify these claims, note that the expected sentence for individuals in group  $j$  will be

$$(1 - \tilde{q})s_o + \tilde{q}s(j) = (1 - \tilde{q})s_o + \tilde{q}\{[s_j^* - (1 - \tilde{q})s_o]/\tilde{q}\} = s_j^*. \quad (\text{A14})$$

It remains to show that  $s_o$  and the  $s(j)$  are within  $[0, \bar{s}]$ . With regard to  $s_o$ , observe that

$$\begin{aligned} s_o &= s_l^*/(1 - \tilde{q}) = s_l^*/(1 - [(s_n^* - s_l^*)/\bar{s}]) \\ &= s_l^*/([\bar{s} - (s_n^* - s_l^*)]/\bar{s}) = s_l^*\bar{s}/(\bar{s} - s_n^* + s_l^*), \end{aligned} \quad (\text{A15})$$

which is non-negative because the numerator is non-negative and the denominator of the last expression is positive (because we have assumed that  $\tilde{q} < 1$ ). Furthermore, this expression is bounded by  $\bar{s}$  because  $s_n^* \leq \bar{s}$ . Hence,  $s_o$  is feasible.

With regard to the feasibility of the  $s(j)$ , note that

$$[s_j^* - (1 - \tilde{q})s_o]/\tilde{q} = [s_j^* - s_l^*]/\tilde{q} = [s_j^* - s_l^*]\bar{s}/(s_n^* - s_l^*). \quad (\text{A16})$$

The numerator of the last expression is non-negative and the denominator is positive. Hence,  $s(j)$  is non-negative. Moreover, the last expression is bounded by  $\bar{s}$  since  $[s_j^* - s_l^*]/(s_n^* - s_l^*)$  is less than or equal to 1. Thus  $s(j)$  is feasible.

We have now proved that  $W^*(\tilde{q}) = W^{**}$ . By part (c), we know that  $W^*(\hat{q}) = W^*$  and by part (a) that  $W^{**} > W^*$ . Hence  $\tilde{q}$  and  $\hat{q}$  must be unequal. And by part (b),  $W^*(q)$  cannot be decreasing in  $q$ . Hence,  $\hat{q} < \tilde{q}$  must hold.

Part (e): To prove that  $W^*(q) > W^*$  for  $q$  in  $(\hat{q}, \tilde{q})$ , suppose otherwise. Then  $W^*(q)$  must equal  $W^*$  for some  $q' > \hat{q}$ . But that would contradict the definition of  $\hat{q}$ , that it is the least upper bound of the set of  $q$  for which  $W^*(q)$  equals  $W^*$ .

It remains to show that  $W^*(q) < W^{**}$  for  $q$  in  $(\hat{q}, \tilde{q})$ . If this claim is not true, there must be a  $q' < \tilde{q}$  such that  $W^*(q') = W^{**}$ . For that to hold, there would have to be feasible  $s_o$  and  $s(j)$  such

that the expected sentence for each group  $j$  equals  $s_j^*$ . But we will demonstrate in several steps that such sentences cannot exist for  $q$  in  $(\hat{q}, \tilde{q})$ .

Step (i). We first prove that if  $s_o > s_j^*/(1 - q)$ , there does not exist a feasible  $s(j)$  such that

$$(1 - q)s_o + qs(j) = s_j^* \quad (\text{A17})$$

can be satisfied. This is so because if  $s_o > s_j^*/(1 - q)$ , then  $(1 - q)s_o > s_j^*$ , implying that (A17) could hold only if  $s(j)$  was negative and thus infeasible.

Step (ii). We also claim that if  $s_o < (s_j^* - q\bar{s})/(1 - q)$ , there does not exist a feasible  $s(j)$  such that (A17) can be satisfied. If  $s_o$  is this low, then  $(1 - q)s_o < s_j^* - q\bar{s}$ , which means that the left-hand side of (A17) is less than  $s_j^* - q\bar{s} + qs(j)$ . But this term is bounded by  $s_j^* - q\bar{s} + q\bar{s} = s_j^*$  because feasibility requires that  $s(j) \leq \bar{s}$ . Hence, the left-hand side of (A17) must be below  $s_j^*$  and thus (A17) cannot be satisfied for feasible  $s(j)$ .

Step (iii). We conclude by showing that a necessary condition for a second-best optimum is that  $q \geq \tilde{q}$ . We do this by demonstrating that if  $q < \tilde{q}$ , it is not possible to choose  $s_o$ ,  $s(1)$ , and  $s(n)$  such that the expected sentences for groups 1 and  $n$  equal  $s_1^*$  and  $s_n^*$ , respectively. Steps (i) and (ii) imply that in order for there to be feasible  $s(1)$  and  $s(n)$  to achieve these expected sentences,  $s_o$  must be contained in both  $[(s_1^* - q\bar{s})/(1 - q), s_1^*/(1 - q)]$  and  $[(s_n^* - q\bar{s})/(1 - q), s_n^*/(1 - q)]$ . Observe that since  $s_1^* < s_n^*$ , the first region is to the left of the second region. Thus, in order for there to be an  $s_o$  that is in both regions, it must be that  $s_1^*/(1 - q) \geq (s_n^* - q\bar{s})/(1 - q)$  or, equivalently, that  $s_1^* \geq s_n^* - q\bar{s}$ . But this implies that  $q \geq (s_n^* - s_1^*)/\bar{s} = \tilde{q}$ . Thus, for  $q < \tilde{q}$ , it must be that  $W^*(q) < W^{**}$ .

Part (f). We know from part (b) that  $W^*(q)$  is non-decreasing in  $q$  and from the proof of part (d) that  $W^*(\tilde{q}) = W^{**}$ . Moreover,  $W^*(1) = W^{**}$ . Thus,  $W^*(q)$  must equal  $W^{**}$  for  $q$  in  $[\tilde{q}, 1]$ .  $\square$

*Proof of Proposition 5.* Parts (a) and (b): First observe that given the formulas in parts (a) and (b), the expected sentence for individuals in group  $j$  is  $s_j^*$ :

$$\begin{aligned}
& (1 - q)s_o^* + qs(j)^* & (A18) \\
& = (1 - q)\{s_l^*/(1 - \tilde{q})\} + q\{[s_j^* - (1 - q)s_o^*]/q\} \\
& = (1 - q)\{s_l^*/(1 - \tilde{q})\} + q\{[s_j^* - (1 - q)[s_l^*/(1 - \tilde{q})]]/q\} \\
& = (1 - q)\{s_l^*/(1 - \tilde{q})\} + \{s_j^* - (1 - q)[s_l^*/(1 - \tilde{q})]\} = s_j^*.
\end{aligned}$$

We demonstrated in the proof of part (d) of Proposition 4 that  $s_o^*$  is feasible. We now verify that the  $s(j)^*$  are feasible for  $q$  in  $[\tilde{q}, 1]$ . We first show that  $s(j)^* \geq 0$ , or equivalently, that  $s_j^* - (1 - q)s_o^* \geq 0$ . Substituting for  $s_o^*$  implies that we must prove that  $s_j^* - [(1 - q)s_l^*/(1 - \tilde{q})] \geq 0$ . Since  $(1 - q)/(1 - \tilde{q}) \leq 1$  for  $q$  in  $[\tilde{q}, 1]$ , we have  $s_j^* - [(1 - q)s_l^*/(1 - \tilde{q})] \geq s_j^* - s_l^* \geq 0$ .

We next verify that  $s(j)^* \leq \bar{s}$ , or  $[s_j^* - (1 - q)s_o^*]/q \leq \bar{s}$  for  $q$  in  $[\tilde{q}, 1]$ . Since the  $s(j)^*$  are weakly increasing in  $j$ , it is sufficient to show that  $[s_n^* - (1 - q)s_o^*]/q \leq \bar{s}$ , or

$$s_n^* - (1 - q)s_o^* \leq q\bar{s} \quad (A19)$$

Since  $s_o^* = s_l^*/(1 - \tilde{q})$  and  $\tilde{q} = (s_n^* - s_l^*)/\bar{s}$ ,  $s_o^*$  can be written as  $s_l^*\bar{s}/(\bar{s} - s_n^* + s_l^*)$ .

Hence, (A19) becomes

$$s_n^* - [(1 - q)s_l^*\bar{s}/(\bar{s} - s_n^* + s_l^*)] \leq q\bar{s}. \quad (A20)$$

By multiplying both sides of (A20) by  $\bar{s} - s_n^* + s_l^*$  and solving the resulting expression for  $q$ , one obtains after simplification  $q \geq (s_n^* - s_l^*)/\bar{s} = \tilde{q}$ . Since the premise of the present proposition is that  $q$  is in  $[\tilde{q}, 1]$ , we have established that the  $s(j)^*$  do not exceed  $\bar{s}$ .

Part (c): This follows immediately from the formula for  $s(j)^*$  in part (b) since  $s_1^* \leq s_2^* \leq \dots \leq s_n^*$  and it is assumed that not all  $s_j^*$  are equal.

Part (d): We first demonstrate that  $s_l^* < s_o^*$ . This follows from the formula in part (a) if  $1 - \tilde{q} < 1$  or  $\tilde{q} > 0$ , which holds because  $s_n^* > s_l^*$ . We next show that  $s_o^* < s_n^*$ , that is, that  $s_l^*/(1 - \tilde{q}) < s_n^*$ , or  $s_l^*\bar{s}/[\bar{s} - (s_n^* - s_l^*)] < s_n^*$ , or  $s_l^*\bar{s} < s_n^*[\bar{s} - (s_n^* - s_l^*)]$ . Expanding the right-hand side of the last inequality and rearranging terms, we obtain  $s_l^*\bar{s} - s_l^*s_n^* < s_n^*\bar{s} - s_n^{*2}$  or  $s_l^*(\bar{s} - s_n^*) < s_n^*(\bar{s} - s_n^*)$ , which holds.

Because  $s_l^* < s_o^* < s_n^*$ , the  $n$  groups can be divided into three possible subsets: a nonempty low subset  $\{1, \dots, j_L\}$ , including group 1, in which  $s_j^* < s_o^*$ ; a possibly empty intermediate subset  $\{j_{L+1}, \dots, j_{H-1}\}$  in which  $s_j^* = s_o^*$ ; and a nonempty high subset  $\{j_H, \dots, n\}$ , including group  $n$ , in which  $s_j^* > s_o^*$ .

Now we will show that these three subsets, defined in terms of the relationship between the  $s_j^*$  and  $s_o^*$ , are the three subsets claimed to exist, defined in terms of the relationship between the  $s(j)^*$  and  $s_o^*$ .

To this end, recall from (A18) that for  $q$  in  $[\tilde{q}, 1]$ ,

$$(1 - q)s_o^{*+} + qs(j)^* = s_j^*. \quad (\text{A21})$$

Consider a group  $j$  in the low subset, for which  $s_j^* < s_o^*$ . Since  $s_j^* < s_o^*$ , it is clear that (A21) cannot hold unless  $s(j)^* \leq s_j^*$ . Hence, for this subset,  $s(j)^* \leq s_j^* < s_o^*$ . For a group  $j$  in the intermediate subset, since  $s_j^* = s_o^*$ , it is apparent that (A21) requires that  $s(j)^* = s_j^*$ , and thus that

$s(j)^* = s_j^* = s_o^*$ . And for a group  $j$  in the high subset, since  $s_j^* > s_o^*$ , (A21) cannot hold unless  $s(j)^* \geq s_j^*$ , and therefore  $s(j)^* \geq s_j^* > s_o^*$ . This establishes part (d).

Part (e): For a group  $j$  in the low subset, we know from the proof of part (d) that  $s(j)^* \leq s_j^* < s_o^*$  for  $q$  in  $[\tilde{q}, 1]$ . We also have that

$$s_j^* - s(j)^* = s_j^* - \{[s_j^* - (1 - q)s_o^*]/q\} = s_j^* - (s_j^*/q) + [(1 - q)/q]s_o^* \quad (\text{A22})$$

The last expression in (A22) reduces to  $[(1 - q)/q](s_o^* - s_j^*)$ , which is positive for  $q < 1$ . The derivative of  $[(1 - q)/q]$  with respect to  $q$  is  $-1/q^2 < 0$ . Therefore the difference between  $s_j^*$  and  $s(j)^*$  declines as  $q$  increases, and it is clear that  $s(j)^* = s_j^*$  at  $q = 1$ . Hence, we have proved the claim about the low subset.

For the intermediate subset, we know from the proof of part (d) that  $s(j)^* = s_j^* = s_o^*$ . Since  $s_o^*$  and  $s_j^*$  are not functions of  $q$ ,  $s(j)^*$  must equal  $s_o^*$  regardless of the relevant  $q$ .

For the high subset, the proof of the claim is analogous to that of the proof for the low subset.  $\square$