MEMORANDUM

No 06/99

SHOULD THE STANDARD OF EVIDENCE BE REDUCED FOR WHITE COLLAR CRIME?

By
Tone Ognedal

ISSN: 0801-1117

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.sv.uio.no/sosoek/
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no/
e-mail: frisch@frisch.uio.no

List of the last 10 Memoranda:

No 24	by Torbjoern Haegeland and Tor Jakob Klette: Do Higher Wages Reflect Higher Productivity? Education, Gender and Experience Premiums in a Matched Plant-Worker Data Set. 37 p.
No 25	by Karl Ove Moene and Michael Wallerstein: Inequality and Redistribution. 41 p.
No 26	by Steinar Holden and Karl Ove Moene: Measures to improve Wage Formation in Sweden. 28 p.
No 27	by Frode Johansen and Tor Jakob Klette: Wage and Employment Effects of Payroll Taxes and Investment Subsidies. 27 p.
No 28	by Hilde Bojer: Equivalence Scales and Intra-household Distribution. 12 p.
No 01	By Gabriela Mundaca and Jon Strand: Speculative attacks in the exchange market with a band policy: A sequential game analysis. 45 p.
No 02	By Pedro P.Barros and Tore Nilssen: Industrial Policy and Firm Heterogeneity. 18 p.
No 03	By Steinar Strøm: The Economics of Screenin Programs. 8 p.
No 04	By Kai Leitemo and Øistein Røisland: Choosing a Monetary Policy Regime: Effects on the Traded and Non-Traded Sectors
No 05	by Eivind Bjøntegård: The Composite Mean Regression as a Tool in Production Studies. 9 p.

A complete list of this memo-series is available in a PDF® format at:
http://www..sv.uio.no/sosoek/memo/

SHOULD THE STANDARD OF EVIDENCE BE REDUCED FOR WHITE COLLAR CRIMES?

by

Tone Ognedal
Department of Economics
University of Oslo

Mailing adress: Department of Economics, Box 1095, 0317 Oslo, Norway
E-Mail: tone.ognedal@econ.uio.no

I am grateful to Tor Jakob Klette, Karl Ove Moene and Atle Seierstad for useful comments, and to Gaute Erichsen for assistance.

ABSTRACT

When uncertain evidence may lead to legal errors, a reduced standard of proof affects the incentive to violate the law. With noisy evidence, an agent may be convicted even if he obeys the law (type I error) or acquitted even if he violates the law (type II error). If the standard of proof is reduced, the expected penalty for minor crimes increases relative to more serious crimes and relative to law compliance. Agents who initially commited minor law violations will become more law compliant, but agents who initially committed violations above a certain level will choose even more serious violations. Thus, the net effect on law violation is ambiguous. The increased probability of type I errors induces excessive law compliance among those who initially complied with the law. The focus is on white collar crime, defined as illegal conduct of an otherwise legal activity.

1.INTRODUCTION

Despite increased efforts over the last decades to investigate and prosecute white collar crimes, it is widely held that the probability of being convicted remains low. One main reason is that white collar crimes, defined as illegal conduct of an otherwise legal economic activity (Träskman, 1975), is often difficult to identify. It may for instance be difficult to judge whether a deviant price pattern over a period is caused by illegal manipulation or by pure coincidences. Thus, even when substantial resources are used to investigate and prosecute, the circumstantial evidence often does not meet the standard required for conviction. If a low probability of conviction cannot be compensated by a high punishment, for example because the defendant is a corporation with limited liability, the result may be that the expected penalty becomes too low to deter crime.

One may ask if a lower standard of evidence would deter white collar crime more efficiently. One reason why the answer is not obvious is the tradeoff between convicting innocent defendants (type I error) and acquitting guilty ones (type II error). A high standard of proof, and thereby a high standard of evidence, keeps the number of type I errors low, but at the cost of more type II errors. An agent is convicted only if the observed evidence exceeds the legal limit with some margin, and as a consequence, agents who commit minor violations can expect to escape conviction. If the standard of evidence is reduced, the probability of being convicted for minor violations increases. The expected penalty increases for minor violations relative to both more severe violations and law compliance. While the marginal cost of minor violations goes up, the marginal cost of severe violations goes down. Moreover, the already law abiding agents may choose excessive law compliance in response to the increased chance of wrongful convictions. Even if a reduced standard of evidence reduces the number of violations, it may therefore increase the total social cost of crime since violations become more severe and overcompliance more excessive.

In most countries, however, the formal standard of proof is not supposed to change with changes in the policy towards crime. In the US, the law requires that the defendant must be found guilty "beyond reasonable doubt" in criminal cases and liable with "preponderance of evidence" in civil cases. Yet, even if the formulation of these standards remain unaltered, the practice varies (Hemmens, Scarborough and del Carmen, 1997). For example, the US antidrug policy of the 1980's reduced the amount of proof necessary to convict for drug related crimes, according to Halsted (1997). Some policy changes have the same effect as a reduction in the standard of proof. For example, if tax evasion is easy to hide in incorrect accountings, one solution is to make incorrect accounting punishable, independent of the cause. The probability of being convicted therefore increases for an agent who has violated the tax law, but also for one who has only misinterpreted the law. Finally, a large amount of law enforcement takes place outside the court and with more arbitrary standards of proof. For example, health and safety regulations are to a large extent enforced by government agencies that can withdraw licences, close a store temporarily or impose fines without a conviction in court. For these adverse decision, the agencies do not have to prove violation "beyond reasonable doubt".(1)

Schrag and Scotchmer (1994) have studied the relation between the standard of evidence and crime, but in a model where the agents face exogenous opportunities for illegal actions. The severity of a violation is exogenous to the agent and overcompliance is not possible. As a consequence, the cost of crime to the agent is simply the difference in expected penalty between a guilty and an innocent defendant. In this framework there is a unique standard of evidence that maximizes the cost of comitting a crime. If the courts use a standard of evidence above this level, a reduction in the standard of evidence will increase deterrence and reduce the number of crimes. The social costs of crime are reduced when the number of crimes are reduced. My paper indicates that this simple relation between the standard of evidence and the cost of crime does not hold if the severity of a violation is chosen by the agent.

Before I move to varying degrees of severity I start in section 2 by exploring a model where the agents have a discrete choice between violating the law or not, as in Schrag and Scotchmer's model. My framework differ from theirs in other respects. For example, Schrag and Scotchmer assume that the court knows that a crime is comitted, but is uncertain about who did it. They investigate the effects of using character evidence as apriori information in the court's decision. The most relevant assumption for white collar crime is that the uncertainty relates to whether or not the law is violated. When this is the case the most relevant apriori evidence is statistical information about crime rates. The main result in Schrag and Scotchmer, that there is a deterrence maximizing standard of evidence, still holds. Yet, the policy implications differ.

In section 3 the agents choice of law compliance is a continous variable. One example is the expected level of a harmful emission, which is illegal above a certain level. I demonstrate how a lower standard of proof leads to a flatter marginal cost of crime curve, and consequently to a higher variance in the agents choices. Section 4 concludes the paper.

2. TO VIOLATE OR NOT

In this section, I analyse the effects of a reduced standard of proof on agents who choose between violating the law or not. The severity of the violation is exogenous. Yet the authorities cannot know with certainty whether an agent has violated the law or not, but observe a variable that is more likely to be high if the agent has broken the law than if he has not. The observation, x, is presented as evidence in court. Let the cumulative distribution function for x be G(x) if the agent is guilty and N(x) if he is not. $N(x) \geq G(x)$ is assumed. g(x) and n(x) are the corresponding probability density functions. The supports of the density functions g(x) and n(x) overlap, since the court cannot always infer from an observation whether or not the agent has violated the law. I assume that the probability density functions satisfy the socalled monotone likelihood ratio property, i.e. g(x)/n(x) increases with x. This implies that if the court uses Bayesian updating it will take a higher x value as a stronger evidence of guilt. The probability that x is observed, i.e. that an agent is checked, is independent of whether the agent has broken the law or not.

I do not address the questions of how the evidence can be improved by the judicial process, as discussed in Rubinfeld and Sappington (1987). In their model, the adversarial process is valuable because it gives the defendants incentive to signal their innocence or guilt. The quality of the evidence therefore varies with the punishment and the standard of proof. In my model, the quality of the evidence is exogenous, independent of the judicial process. Moreover, I ignore pure legal mistakes by the court and only discuss errors that occur because the evidence is uncertain. The probabilities of the two types of errors are therefore endogenously determined by the distribution of the evidence and the decision rule used by the court. Polinsky and Shavell (1989), in contrast, discuss the effects of legal errors that include pure mistakes by the court such as misinterpretation of the evidence and misunderstanding of the law. As a consequence, they treat the probabilities of the two types of legal errors as exogenous variables that may vary independently of one another.

In addition to the evidence x the court has apriori information about the probability that the defendant has violated the law. Apriori information is defined as the information about the defendant that he regards as independent of his choice to violate the law or not, such as crime rates[(2)]. This type of apriori information, often referred to as "naked statistical evidence", may still be regarded as controversial in some jurisdictions. On the one hand, hardly any court would allow convictions on the basis of naked statistical evidence alone. For example, the fact that one knows that sixty percent of the agents must be guilty as charged is not sufficient to win a civil suit against any member of the group[(3)]. On the other hand, in both US and european countries, statistical evidence is now routinely used as a supplement to evidence about the individual case (see e.g. Fienberg and Straf (1991) and Matthews (1994)). Let q be the court's apriori probability that the defendant is guilty, that is their assessment of guilt before they have heard the evidence x. q can in principle take any value between 0 and 1.

The court's reasoning about the defendants guilt is modelled as if they used Bayesian inference analysis.[(4)] Bayesian analysis is now widely accepted as an appropriate technique for assessing evidence in legal proceedings, and sometimes explicitly used in expert testimonies in court. Moreover, even if Bayesian analysis is not rigorously applied, it may be a reasonable representation of how beliefs are formed. Using Bayes rule, the probability that an agent is guilty given the evidence x and the apriori probability of guilt q is

$$Pr(guilty \mid x) = \frac{g(x)q}{g(x)q + n(x)(1-q)} \equiv \Phi(x;q) \quad \textbf{(1)}$$

An agent is convicted for violation of the law if the probability that he is guilty exceeds the standard of proof s, i.e. if

$$\Phi(x;q) \geq s \quad \textbf{(2)}$$

$\Phi(x;q)$ increases with x, since g(x)/n(x) increases with x. This implies the agent is convicted if and only if x exceeds a critical level X given by $\Phi(X,q)=s$. Inserting for $\Phi(x;q)$ from (1), this condition can be written as

$$\frac{g(X)}{n(X)} = \frac{s}{1-s}\frac{1-q}{q} \quad \textbf{(3)}$$

X is the standard of evidence used by the court. Since g(X)/n(X) increases with X, the standard of evidence is higher the higher the standard of proof s is and the lower the apriori probability of guilt is. Hence, we can write

$$X = X(q,s) \quad X_q > 0 \quad X_s < 0 \quad \textbf{(4)}$$

Since the agent is convicted when $x \geq X$ is observed, the probability of being convicted is 1-G(X) for a guilty agent and 1-N(X) for an innocent one. With a penalty P for law violation, the expected penalty is (1-G(X))P for a guilty agent and (1-N(X))P for an innocent one. A reduction in the standard of evidence increases the probability of conviction, and thereby also the expected penalty, for both guilty and innocent agents. The marginal expected cost of violating the law is the difference between the expected penalty for a guilty and an innocent agent, hereafter referred to as the cost of crime. The cost of crime can be written as

$$C(X) = P\,[N(X) - G(X)] \qquad \textbf{(5)}$$

The effect on the cost of crime from a reduction in the standard of evidence X is

$$-C'(X) = P\,[g(X) - n(X)] \qquad \textbf{(6)}$$

The effect of a reduction in the standard of evidence depends on the effect on g(X)-n(X), which is the difference between the increased probability of conviction for an innocent and a guilty defendant.This gives us the following results:

Proposition 1

If the standard of evidence is reduced, the cost of crime is increased if the apriori probability of guilt is lower than the standard of proof (q<s) and reduced if the apriori probability of guilt exceeds the standard of proof (q>s). The cost of crime is maximized when the standard of proof is set equal to the apriori probability of guilt.

Proof:

Let X^m denote the standard of evidence determined by $g(X^m)=n(X^m)$, i.e. where $C'(X^m)=0$. Since g(x)/n(x) is strictly decreasing in x, it follows from (5) that $C'(X)>0$ for $X<X^m$ and $C'(X)<0$ for $X>X^m$. Consequently, X^m is the standard of evidence that maximizes the cost of crime. It follows from (3) that $X=X^m$ if s=q, since this gives $g(X^m)/n(X^m)=1$. Since g(x)/n(x) increases with x, $X>X^m$ if q<s and $X\leq X^m$ if $q\geq s$. Hence, $-C'(X)<0$ if q>s and $-C'(X)>0$ if q<s.

When $q<s$, a defendant is considered innocent based on the apriori information. He is therefore convicted if and only if the evidence x is sufficiently more likely if the defendant is guilty than if he is not, which implies that the critical level of evidence must satisfy $g(X)>n(X)$. When the critical level of evidence is reduced, the probability of conviction therefore increases more for a guilty agent than for an innocent one. As a result, the expected cost of crime increases. The opposite is the case when $q \geq s$.

Under the current rules, the standard of proof used in court is above 1/2 even in civil cases. Since the crime rate is far below 50 percent for most violations, $s>q$ is normally satisfied and a reduction in the standard of proof will increase the cost of crime. To maximize the cost of crime the standard of proof s should be set equal to the crime rate q. This would imply that a defendant might be convicted even if he is innocent with "preponderance of evidence". Moreover, the policy has the paradoxical implication that the standard of proof should be lower the lower the crime rate is. A crime minimizing policy would therefore hardly be acceptable in court.

One may interpret the case where $q \geq s$ as a reversal of the burden of proof, since the defendant is considered guilty before the evidence about his case is presented. Even though it seems unlikely that a court will openly reverse the burden of proof, it is argued that they sometimes do, for instance in drug-crimes and certain white collar crimes. A reversal of the burden of proof may also occur for law enforcement outside the court, for example in decisions made by regulatory agencies, where the standard of proof may be lower than in court. The result above suggests that even though a reversal of the burden of proof increases the probability of convicting guilty defendants, it also encourages more crime.

Equilibrium crime rates

An agent i will violate the law if his marginal gain from crime exceeds the marginal cost of crime. Since the critical level of evidence X is a function of q and s, as given by (4), it follows from (5) that we can write the cost of crime as a function of q and s

$$C(X(q;s)) \equiv c(q,s) \quad \textbf{(7)}$$

where $c_q(q;s) = C'(X)X_q$ is positive for q<s and negative for q>s. I assume that the marginal gain from crime r may differ between the agents. Let F(r) be the cumulative distribution of r over the interval $[r_L,r_H]$ such that $F(r_L)=0$ and $F(r_H)=1$. Let Q denote the crime rate, i.e. the fraction of agents that violates the law. Q is then given by 1-F(c). The crime rate Q is reduced when the cost of crime is increased since F(c) is an increasing function of c. Since c is a function of q and s, as given by (7), we can write Q as a function of q and s,

$$Q = 1 - F(c(q,s)) \equiv Q(q,s) \quad \textbf{(8)}$$

This gives us $Q_q(q;s) = - F'(c)c_q(q;s)$. The sign of Q_q is the same as the sign of $-c_q$. Hence, $Q_q<0$ for q<s and Q_q for q>s as illustrated on figure 1. In equilibrium, the fraction q that the court believes is criminal should equal the actual crime rate Q. The equilibrium crime rate q^* is therefore given by $q^*=Q(q^*;s)$, where s is exogenous. The equilibrium is located where Q as a function of q intersects with the 45^o-line, as shown on figure 1.

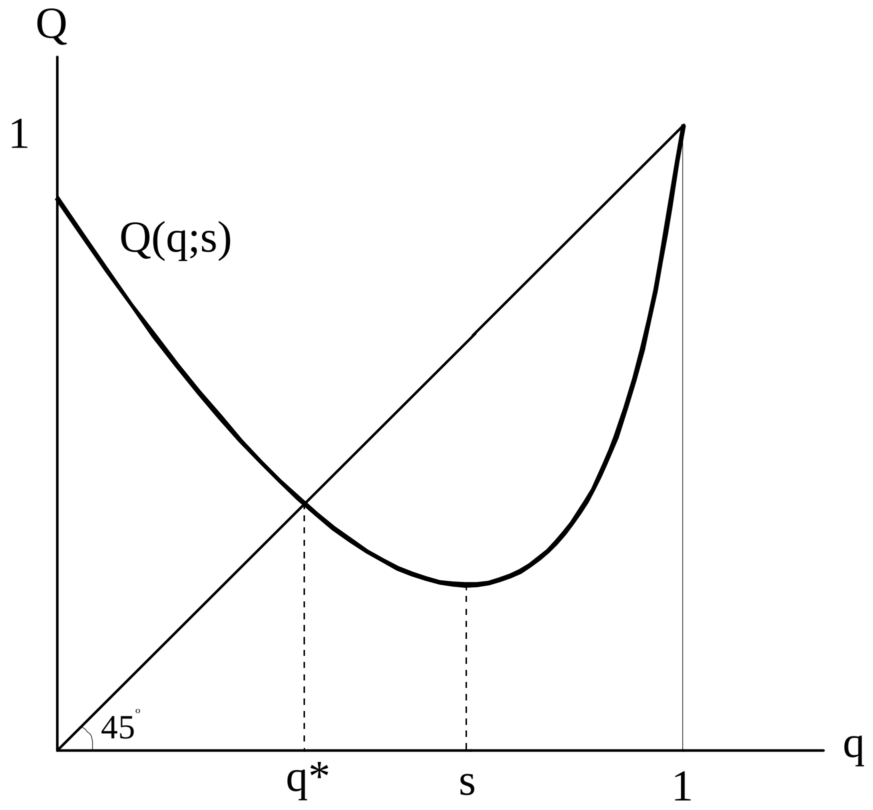

Figure 1

It follows from proposition 1 that a decrease in the standard of proof s shifts Q downwards for all q<s and upwards for q>s. The effects on the equilibrium crime rate depends on whether the equilibrium crime rate q^* is above or below s. In the most likely case, where $q^*<s$, the equilibrium crime rate is reduced when the standard of proof is reduced.

3. DEGREES OF LAW VIOLATION

For many types of laws and regulations, the severity of the violations is obviously important. Moreover, overcompliance with the law to avoid being misjudged may represent a social waste. As a consequence, an enforcement policy should not be measured by its effect on the fraction of violations only, but also by its effects on the severity of the violations among the violators and the degree of law compliance among the law obedient agents. Accordingly, I model the agent's choice of law compliance as a continous variable rather than a discrete choice between violating the law or not.

To gain in focus, I apply the model to an example of environmental crime. The production of firms in an industry is potentially harmful and the law requires that expected emission θ from a firm must not exceed L. The firms can reduce the expected emission by using resources on monitoring and other safety procedures, but the actual emission may be higher or lower than the expected level. Hence, even a firm with a high safety standard (θ low) may be unfortunate and cause considerable emissions and a firm with a low safety standard (θ high) may be lucky and cause no emission. The expected emission can take values between 0 and 1. The cost of keeping a certain expected safety level is expressed as $S(\theta;z)$, where z is a cost parameter that may differ between firms. $S_\theta<0$ and $S_{\theta\theta}>0$ is assumed. The authorities can only observe the actual emission x[(5)]. To simplify I assume that $x=\theta+\varepsilon$, where ε is a stochastic variable with probability density function $f(\varepsilon)$ and $\varepsilon\in[\![-\infty,\infty]\!]$. $f(\varepsilon)$ is symmetric and $E\varepsilon=0$. In the appendix I demonstrate that the results also hold with probability density functions that are not symmetric, such as the binominal or the poisson distribution. As in section 2, the probability of being investigated is independent of both actual and expected emission.

Since the authorities cannot observe the expected level of emission θ directly, their decision to convict or acquit a firm for violating the law must be based on the observed emission x and eventual apriori information of the distribution of θ. Let the probability density function a(θ) represent the apriori information of θ. If no apriori information is available or admissible it may be reasonable to assume that all θ-values between 0 and 1 are equally likely apriori, that is to set a(θ) equal to 1. If a firm is convicted for violation of the law, it is given a penalty P that increases with the observed deviation from the standard, i.e. P=P(x-L) and P'(x-L)>0.

Using Bayes rule, the probability that a firm has violated the limit Θ if an emission x is observed is

$$Pr(q \geq L \mid x) = \frac{\int_{\Theta}^{1} f(x-q)a(q)dq}{\int_{0}^{1} f(x-q)a(q)dq} \equiv G(L;x) \qquad \textbf{(9)}$$

A firm is convicted and punished if and only if this probability is above the standard of proof s. I assume that the distribution of the evidence is such that the court's assessed probability of violation G(L;x) is higher the higher the observed emission x is. When G(L;x) is a strictly increasing function of x, a firm is convicted if and only if the observation x exceeds a critical level X given by

$$G(L;X) = s \qquad \textbf{(10)}$$

Since G(L;X) increases in X, it is immediate from (10) that the standard of evidence X is reduced if the standard of proof s i reduced. Since the apriori information is now a distribution of the expected emission θ, the relation between the standard of evidence and the apriori information is more complicated than in the previous section where the apriori information could be represented by a probability of guilt.

With an apriori distribution that is sufficiently biased, such that θ is believed to be high, the standard of evidence X may be lower than L. This implies that a defendant can be convicted even if the observed emission is below the legal limit. Since this outcome hardly occurs in practice and since the main conclusions do not depend on whether the critical evidence level is above or below the legal limit, I only discuss the case where the critical level of evidence X exceeds the legal limit L.

A firm's choice of safety level is a tradeoff between the cost of keeping a low expected level of emission S(θ;z) and the expected penalty for high emissions. The expected penalty from choosing θ when the standard of evidence is X is

$$C(q\,;X) = \int_X^\infty f(x - q\,)P(x - L)dx \qquad \textbf{(11)}$$

The firm chooses the level of θ that minimizes its total costs S(θ;z)+C(θ,X). This gives us the first order condition

$$-S_q(q\,;z) = C_q(q\,;X) \qquad \textbf{(12)}$$

where C'_θ can be derived from (11), which gives

$$C_q(q\,;X) = -\int_X^\infty f'(x - q\,)P(x - L)dx > 0 \qquad \textbf{(13)}$$

A increase in the expected emission increases the probability of a large emission. The expected penalty is then increased for two reasons: First, the probability of being convicted is higher. Second, a large punishment is more likely since a large emission is less likely.

Let θ^* denote the firm's optimal choice of expected emission. The effect on the optimal expected emission from a reduction in the standard of evidence is found by differentiating (12) with respect to X. This gives us

$$\frac{d\,q^*}{-\,dX} = \frac{C_{qx}}{[\,S_{qq} + C_{qq}\,]} \qquad \textbf{(14)}$$

Since the nominator $[S_{\theta\theta}+C_{\theta\theta}]$ is positive, the sign of $-d\theta^*/dX$ equals the sign of $C_{\theta x}$, the change in the marginal expected penalty for θ when the standard of evidence is changed. From (13) we get

$$C_{qX} = f\,'(X - q\,)P(X - L) \qquad \textbf{(15)}$$

$f\,'(X-\theta)$ is positive for $\theta>X$ and negative for $\theta<X$. Hence, $C_{\theta X}$ is positive if $\theta>X$ and negative if $\theta<X$. It follows that $-d\theta^*/dX$ is negative if $\theta^*<X$ and positive if $\theta^*>X$. This gives us the following results

Proposition 2

In firms with an initial expected level of emission above the standard of evidence X the expected level of emission θ^* is increased if the standard of evidence is reduced. In firms with an initial expected level of emission below X the optimal expected level of emission is reduced if the standard of evidence is reduced.

When the expected emission is below the standard of evidence X, as in figure 2a, an increase in θ makes it more likely to get outcomes above X. With a lower standard of evidence more of these outcomes will be punished, which increases the marginal cost of θ. With an initial standard of evidence equal to X_a an increase in expected emission from θ_L to θ_H increases the probability of conviction with the shaded area A. If the standard of evidence is reduced to X_a' the increased probability of conviction is A+B, and therefore higher. In figure 2b the critical level of evidence is below the expected harm. While an increase in expected harm increases the probability of the high x-values that lead to conviction, it also lowers the probability of the low x-values that lead to conviction. If the standard of evidence is reduced, lower x-values are punished and consequently it becomes more valuable to reduce the probability of these outcomes. Hence,by reducing the probability of the low, punishable x-values, an increase in θ is now less costly. If the expected emission is increased from θ_L to θ_H , the probability of conviction increases with A÷B when the standard of evidence is X_b. If X is reduced to X_b', the increase in the probability of conviction is only A÷B÷C.

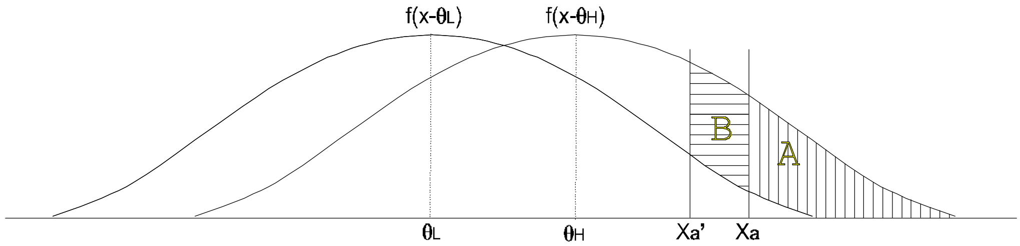

Figure 2a

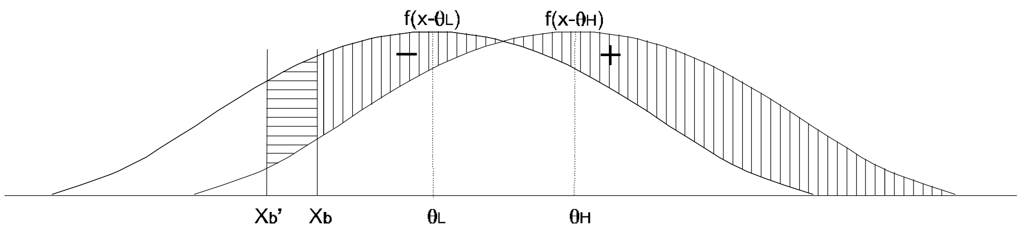

Figure 2b

Figure 3 below illustrates the optimal choice of expected emission for two firms, I and II, with different cost parameters, i.e. z-values. Firm I has lower marginal cost of reducing expected emission that firm II, and as a consequence firm I's optimal expected emission θ_I^* is lower than the optinal expected emission θ_{II}^* of firm II. Let us assume that $\theta_I<X<\theta_{II}$. When X is reduced, it follows from (15) that the marginal cost function $C_\theta(\theta;X)$ shifts upwards for $\theta<X$ and downwards for $\theta>X$. As a result, the optimal expected level of emission is reduced for firm I, from θ_I^1 to θ_I^2, but increased for firm II, from θ_{II}^1 to θ_{II}^2.

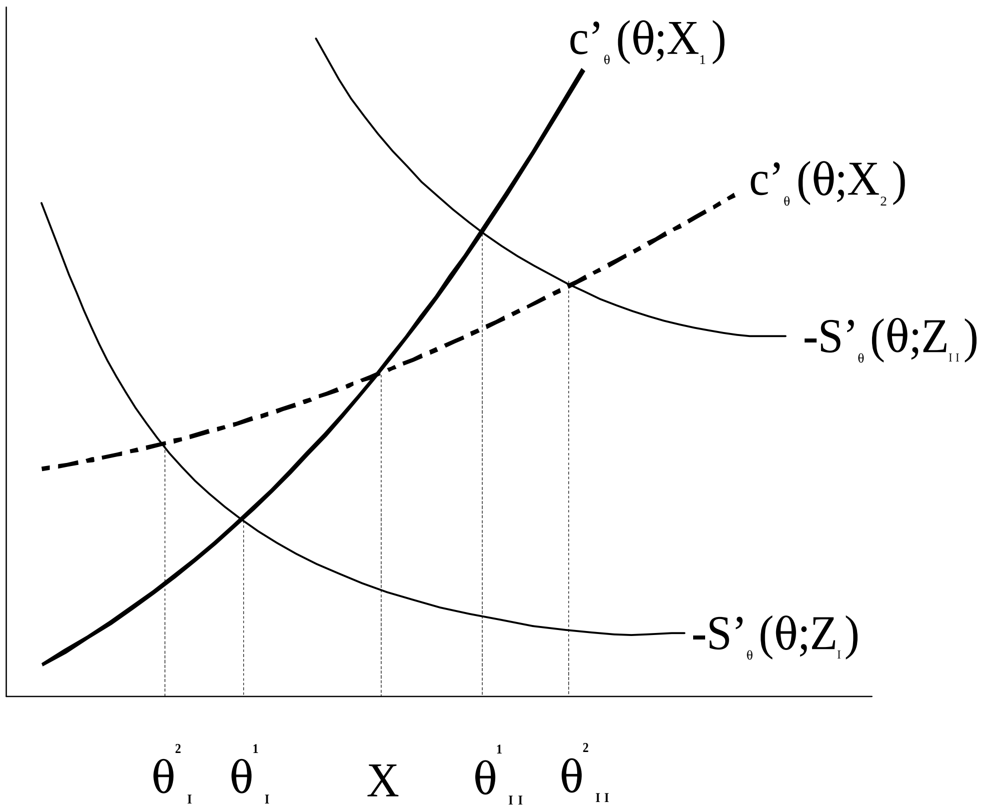

Figure 3

To find the net result on crime from a reduction in the standard of proof it is useful to distinguish between three groups of agents: Group 1 consists of those who already obey the law, i.e. they choose $\theta<L$. The second group consists of those who commit offenses that are so minor that they expect to be acquitted because of insufficient evidence, i.e. $L<\theta<X$. The third group consists of those who commit so serious violations that they expect a conviction if they are detected,i.e. $\theta>X$. The first group reduces their expected emission further to avoid a conviction by mistake, which means that they overcomply. The third group chooses even more severe law violations. Both changes are costly to society. The positive effect of an increased probability of conviction is the reduction in the minor law violations comitted by the second group.

4.CONCLUDING DISCUSSION

If the probability of conviction is increased by reducing the standard of proof, the expected penalty for law violations increases. As long as the courts make legal errors, however, the net effect on crime is ambiguous. When the standard of proof is reduced, the expected penalty for minor law violations increases relative to both law obedience and more severe violations. Since the marginal expected penalty for comitting a minor violation instead of no violation increases, the number and the size of minor violations goes down, as intended. The policy has two negative side effects, however. First, a reduction in the standard of proof reduces the difference in expected penalty between a minor and a severe violation. As a result, the severe violations become even more severe. Second, when the courts make legal errors even law obedient agents may be convicted. When the standard of proof is reduced, outcomes that initially would lead to acquittal now lead to conviction. As a result, the law obedient agent overcomply to avoid more mistakes.

To induce the agents to choose the optimal level of law compliance, the increase in expected penalty for an increase in the severity of violation should equal the marginal social cost. Ideally, the penalty function P(.) could be designed to obtain the optimal expected penalty function, regardless of the probability of conviction. If the probability of conviction is kept low to minimize legal errors, this can be compensated for by a high punishment. In practice, however, there are several restraints on the use of punishments. First, most legal systems require that the punishment must be reasonable given the crime, and given the punishment for other crimes. Hence, it would hardly be acceptable that a minor tax law violation was given the same punishment as a murder even if this should be optimal because the probability of being convicted for tax law violations is low. Second, when the defendant is a firm, as may often be the case for white collar crime, the penalty is limited by the firm's assets.

Since the cost of complying with legal standards may vary between agents, it may seem optimal that the standards should differ accordingly. In environmental law, the standard is sometimes formulated such that it is possible to use higher standards the lower the agent's marginal cost of reducing pollution is. Negligence rules such as "due care" is one example of a flexible standard. Different legal standards for different agents would not change the conclusions, however. As long as the marginal expected penalty of increasing the regulated activity differ from the marginal social cost, an agent may choose an activity level that differs from the social optimal one.

APPENDIX A

Let p denote the probability that there is an accidental emission in a production plant. p is determined by the firm's choice of safety procedures. Hence, the number of accidents in n operation days is binominally distributed with mean np. The firm's choice of p therefore correspond to the choice of expected emission θ in section 3. The law requires that the probability of an accident should not exceed q, i.e. the expected number of accidents should not exceed qn (which corresponds to L in section 3). The authorities cannot observe p directly, however, but observe the number of accidents x. The firm is convicted if the observed number of accident exceeds the standard of evidence X, where X>qn.

The expected penalty from choosing p is

$$c(p;X) = \int_X^n \binom{n}{x} p^x (1-p)^{n-x} P(x-L)dx \tag{A.1}$$

The effect of an reduction in the standard of evidence on the marginal cost of p is then

$$C_{pX} = -\binom{n}{X} P(X-L)\, p^X (1-p)^{n-X} \left[\frac{X}{p} - \frac{n-X}{1-p} \right] \tag{A.2}$$

C_{pX} corresponds to $C_{\theta X}$ in section 3. Since X>L, it follows from (A.2) that

$-C_{pX}<0$ if np>X and $-C_{pX}>0$ if np<X, which is the same results as in proposition 2.

If n is large and p is close to zero, the binominal distribution can be approximated by the poisson distribution. Hence, it is easily verified that the conclusions also hold in this case.

NOTES

(1) Although my analysis applies to all types of law enforcement, I use the terms such as "guilt and "penalty" from criminal law for an easier exposition.

(2) Tribe (1971) argues that it is not appropriate to classify some evidence as "apriori information", as if it was unquestionable facts in contrast to the uncertain evidence about the individual case.

(3) One famous illustration of the problem with the use of statistical information in court is the socalled Gatecrashers paradox: A concert manager knows that only 4 out of 10 in the audience paid their tickets, the rest broke in. If this information is admitted as the only evidence in a civil suit against someone in the audience, the defendant will be held liable since he is guilty with "preponderance of evidence". Few courts would accept statistical information as sufficient for conviction. A testimony from an eyewitness that is correct in only 6 out of 10 observations might be considered sufficient for conviction, however (Cooter and Ulen (1996)).

(4) Other methods for assessing the defendant's guilt, such as standard hypothesis testing, would lead to a similar decision rule for the court and therefore also to similar conclusions.

(5) An formulation that is closer to the analysis in section 2 would be to assume that the firm controlled the actual harm θ, but that the authorities made errors in their observations of the harm. The source of the uncertainty of the observations does not affect the main result from the analysis, however.

REFERENCES

Cooter,R. and T.Ulen (1996), "Law and Economics", 2nd ed., Addison-Wesley, New York.

Fienberg S.E. and M.L.Straf (1991), "Statistical evidence in the US courts: an appraisal", Journal of Royal Statistical Society Association, vol.154.

Halsted,J.B (1997), "The anti-drug policies of the 1980s: have they increased the likelihood for both wrongful convictions and sentencing disparities?", Criminal Justice Policy Review, vol.6, no.3.

Hemmens C., K.E.Scarborough and R.V del Carmen (1997), "Grave doubts about "reasonable doubt": confusion in state and federal courts", Journal of Criminal Justice, vol.25, no.3.

Matthews,R. (1994), "Improving the odds on justice?", New Scientist, april 1994.

Polinsky,A.M. and S.Shavell (1989), "Legal error, litigation and the incentive to obey the law", Journal of Law, Economics and Organization, vol.5

Rubinfeld,D.L. and Sappington,D.E.M. (1987), "Efficient awards and standards of proof in judicial proceedings", RAND Journal of Economics, vol.81, no.2

Schrag,J. and S.Scotchmer (1994), "Crime and prejudice: The use of character evidence in criminal trials", Journal of Law, Economics and Organization, vol.10, no.2.

Tribe,L.H. (1971), "Trial by mathematics: precision and ritual in the legal process", Harvard Law Review, vol.84, no.6.

Träskmann,P.O.(1975), Arbetsmaterial. NU B 1977:25 Straffesystemer i Norden.